Universiteit Leiden
The Netherlands

Statistical learning for complex data to enable precision medicine strategies

Zwep, L.B.

Citation

Zwep, L. B. (2023, April 12). *Statistical learning for complex data to enable precision medicine strategies*. Retrieved from https://hdl.handle.net/1887/3590763

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3590763

Note: To cite this publication please use the final published version (if applicable).

Statistical learning for complex data to enable precision medicine strategies

Laura B. Zwep

Cover design: Tiva Pam

Printed by: JouwBoekDrukkerij.nl - U2pi BV, Den Haag

ISBN: 978 94 9329 958 0

© L.B. Zwep, 2023

All rights reserved. No part of this book may be reproduced in any form or by any means without permission of the author.

Statistical learning for complex data to enable precision medicine strategies

PROEFSCRHIFT

ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof.dr.ir. H. Bijl,

volgens besluit van het college voor promoties

te verdedigen op woensdag 12 april 2023

klokke 13:45 uur

door

Laura Bertien Zwep

geboren te Utrecht, Nederland

in 1994

Promotor:	Prof. dr. J.J. Meulman
Co-promotor:	Dr. J.G.C van Hasselt
Promotiecommissie:	Prof. dr. H. Irth
	Prof. dr. J.A. Bouwstra
	Prof. dr. C.A.J. Knibbe
	Prof. dr. F. Mentré (Université Paris Cité)
	Dr. M. van Smeden (UMC Utrecht)

The research described in this thesis was partially funded by the Data Science Research Programme (Leiden University) and was performed at the Systems Pharmacology and Pharmacy division of the Leiden Academic Centre for Drug Research (LACDR), Leiden University (Leiden, The Netherlands).

Table of contents

1 General Introduction and Outline 7

I Statistical approaches in pharmaceutical research

2 Beyond the Randomized Clinical Trial: Innovative Data Science to Close the Pediatric Evidence Gap 17

II Biomarker discovery

3 Identification of high-dimensional omics-derived predictors for tumor growth dynamics using machine learning and pharmacometric modeling 39

4 Longitudinal metabolomics of community-acquired pneumonia 63

III Real world data

5 Identification of antibiotic collateral sensitivity and resistance interactions in population surveillance data 81

6 Inference of collateral sensitivity effects in large scale antimicrobial resistance surveillance data 97

7 Virtual patient simulation using copula modeling 119

IV General discussion and summary

8 General discussion and summary 137

9 Nederlandse Samenvatting 149

Appendices

Curriculum vitae and list of publications 158

Affiliations of authors 160

Acknowledgements 161

Chapter 1

General Introduction and Outline

1.1 Introduction

This thesis addresses the use of advanced statistical learning techniques to characterize complex and large scale data in biomedical and pharmacological research to enable development of precision medicine strategies.

1.1.1 Treatment variability

Addressing individual variability in patients' responses to drug treatment is of crucial importance to provide adequate treatment for each patient. To optimize treatment outcomes, individualization of treatment strategies is needed. This is often referred to as precision medicine, an approach based on individuals, rather than on average population effects. Lack of individualization can lead to insufficient treatment efficacy if underdosed, and adverse drug effects if overdosed.

Variation in treatment response can be due to between-patient variability and within-patient variability. Between-patient variability occurs when patients react differently to the same treatment. By quantifying this variability, patient- or disease-specific predictors can be identified to inform design of precision treatment strategies. Within-patient variability concerns the change of treatment response within a patient, due to changes in progression or adaptation of the disease during treatment. Understanding and explaining these types of variability can improve early medical decision making, through monitoring of patient response biomarkers.

The explanation of treatment response variation in drug research is part of the discipline of (clinical) pharmacology. At its core, pharmacological research concerns characterization of the dynamics of drug exposure in the body, or pharmacokinetics (PK), and the dynamics of corresponding drug effects as measured through biomarkers, or pharmacodynamics (PD). The PK describes the way the drug is moving through and changing in the body, where aspects such as absorption, concentration at the target site and clearance of a drug play a central role. The PD is the effect a drug has on a patient, mostly clinical outcomes such as blood pressure for hypertension drugs or tumor size for anti-cancer drugs. Variability in treatment response between patients can often be attributed to patient-specific factors which affect PK or PD relationships. Quantitatively capturing PK-PD relationships and identifying factors associated with inter-individual variability through mathematical and statistical models, commonly referred to as pharmacometrics, has developed as an important tool to aid in the design of individualized treatment strategies.

An important topic in pharmacology is the occurrence and development of treatment resistance. Treatment resistance can occur when a population of targeted cells or pathogens adapts to the administered drug treatment. Many different mechanisms of treatment resistance can arise due to evolutionary processes, selection pressure and rapid cell division (zur Wiesch et al., 2011). Similar treatment resistance mechanisms are observed in both oncology and infectious diseases (Groenendijk & Bernards, 2014; zur Wiesch et al., 2011).

In oncology, treatment response variability is an important factor in treatment

failure. Within patients, tumors are shown to develop resistance, which contributes to high treatment failure (Sun & Hu, 2018). Understanding of resistance development has improved cancer treatment, but there is still a lot of unexplained variability between patients (Sun & Hu, 2018; Yin et al., 2019). Understanding underlying factors for resistance development and predictive biomarkers for treatment response could improve cancer treatment outcomes.

In infectious diseases, resistance to antimicrobial drugs represent a global health challenge (Talebi Bezmin Abadi et al., 2019; World Health Organization, 2014). Pathogens can develop resistance against multiple antimicrobial treatments, turning simple infections into serious health threats. Alternative treatment strategies could help prevent the development of resistance and even reduce resistance (Maltas & Wood, 2019). One such strategy is the use of collateral sensitivity, a phenomenon where resistance to one antibiotic reduces the resistance to a second antibiotic. Collateral sensitivity is one strategy which is of interest to design treatment strategies which suppress the risk of resistance (Aulin et al., 2021; Pál et al., 2015; Roemhild & Andersson, 2021).

1.1.2 Patient- and disease associated factors

Knowledge of underlying factors of treatment response variability, such as patient and disease characteristics, is pivotal to develop strategies which can improve treatment outcomes. Insight into the factors contributing to variation in treatment response can help to predict the treatment response in different patients, enabling precision treatment strategies. Data to support deriving such insights are increasingly available from clinical studies and from routine patient care (Morrato et al., 2007).

The variability in treatment response and PK and PD of drugs between patients is large. Different patient covariates can explain parts of this variance; these covariates are for example age and body weight, but also include measurable biological factors, known as biomarkers, which are concentrations of molecules, or other physiological measures that can indicate underlying biological processes at a molecular or cellular level (Depledge et al., 1993; Strimbu & Tavel, 2010). Most pharmacological studies characterize time-dependent trends in the patients with regards to drug concentrations, treatment response and biomarker levels. These trends can be determined by measuring biomarkers reflecting different aspects of the patient's physiological characteristics. Next to the dependence structures introduced by longitudinal measurements, most covariates are also interdependent. These covariates vary with each other, often due to physical properties or biological processes, such as height being related to weight physically.

Molecular profiling 'omics' technologies for characterization of DNA, RNA, proteins, and metabolites are increasingly used to characterize biological samples from patients and during drug research (Nice, 2018). Omics data are often high-dimensional, having more variables than patients ($p \gg n$), due to the possibility to measure hundreds (metabolomics), thousands (transcriptomics) or even millions (genomics) of variables. These large sets of omics data allow for thorough characterization of pa-

tients, but also pose a challenge in terms of data analysis, due to this high-dimensional nature, and that they can be measured over time.

The increasing use of electronic health record databases has provided new opportunities for using routine health care data collected from patients in scientific research (Currie & MacDonald, 2000; Swift et al., 2018). These real world data are used to monitor patients and their treatment response in the clinic, and to make decisions about treatments and dosing schedules. Improved data availability, due to developments in data management and sharing, creates opportunities for studying patient characteristics that can predict treatment response, enabling more individualized dosing regimens in the clinic.

Overall, the complexity of these pharmacological, molecular and health care data requires the use of appropriate statistical techniques that are able to address important biomedical questions require appropriate handling of the associated heterogeneous, high dimensional, and longitudinal data.

1.1.3 Statistical methods and pharmacometrics

Complex data, such as longitudinal and high-dimensional data, require different data analysis methods. Several methods have been developed in the fields of pharmacometrics and statistics with the aim to detect covariates and biomarkers that can explain the treatment response variability and estimating their effect size.

Longitudinal data allow for studying treatment responses over time, but pose a challenge for data analysis. Measurements within a patient often are typically more similar than measurements between patients, violating the assumption of independent residuals, which is assumed in standard regression models. Mixed effect models have been developed to include the dependency structure between different measurements, enabling the characterization of the inter patient variability (McCulloch & Searle, 2000). With patient characteristics and biomarkers, part of this inter-patient variability can be explained in order to better predict outcomes for specific patients.

Pharmacometrics concerns the modeling and prediction of PK and PD measures using longitudinal data analysis methods. Through, mostly nonlinear, mixed effect modeling, random effects are estimated which represent the individual variability, thereby quantifying how diverse the response to certain drugs is over different patients, and predicting the drug effects in the population. Pharmacometric models can then be used for simulations to predict treatment responses and variability in different patient populations. These simulations take into account the unexplained between-patient variability, as well as covariates used to explain part of the difference between patients (Mould & Upton, 2012, 2013; Upton & Mould, 2014).

Next to variability between measurements, modeling interdependence between covariates also poses a challenge. Pharmacometric models often include covariates that are interdependent. To simulate different (special) patient populations, simulation of realistic sets of patient covariates is crucial, but this requires an accurate estimation of the dependence between covariates (Smania & Jonsson, 2021).

A third data analysis challenge is posed by high-dimensional data, such as most

omics data, where standard linear regression and more complex nonlinear mixed effect models are not applicable anymore, because the parameters of a model cannot be uniquely estimated (Johnstone & Titterington, 2009). One way of circumventing this problem is to use shrinkage, where a penalty is placed on the size of the parameters (e.g., regression weights), which is a technique developed within the field of statistics. The two most common shrinkage methods for linear regression are Ridge regression (Hoerl & Kennard, 1970), which penalizes the sum of the squares of the parameter values, effectively shrinking large parameter values more, and the lasso (Tibshirani, 1996), which penalizes the sum of the absolute parameter values, which shrinks some parameters to zero. So the lasso selects the most relevant parameters, which are estimated to be non-zero. In both cases, a shrinkage parameter is used to determine how strong the penalty is. Another way to analyze high-dimensional data is by using dimension reduction techniques, such as principal component analysis and proximity mapping, where variability in the high-dimensional data is summarized into much less dimensions (Heiser et al., 2020).

Although methods for high-dimensional data, longitudinal data and other complex data have been extensively developed and used, combining different data analysis methods to study treatment response variability still remains a challenge. The combination of pharmacometric approaches and statistical methods, and the application of different statistical methods in pharmacological research, can potentially improve our understanding of treatment variability and allow for the optimization of treatment and dosing regimens for individual patients.

1.2 Scope

In this thesis, we studied the use of advanced statistical techniques for the analysis of biomedical datasets to enable development precision medicine strategies, with a particular focus on pharmacological applications. With an increase in data complexity, techniques from different disciplines need to be integrated to answer research questions regarding precision treatment and antibiotic resistance. This thesis first describes this increasing data complexity and different data science techniques in more detail (Section I). Next, the thesis aims to integrate statistical techniques for analyzing high-dimensional data and pharmacometric methods to facilitate omics biomarker research (Section II) and, finally, different statistical methods are used to build tools for the pharmacological studies in clinical pathogens and populations (Section III). Thus, the thesis contains the following sections.

Section I: Data science in pharmaceutical research

In Chapter 2, we discuss the use of different data types to enhance clinical pharmacological research. These complex data require the use of different data analysis techniques and could provide insights that are hard to obtain from randomized clinical controlled trials.

Section II: High-dimensional biomarker discovery

Section II focusses on detection of biomarkers in high-dimensional omics data, using methods from statistics and data science. First, in Chapter 3, we use the lasso in combination with a pharmacometric model for tumor growth dynamics to identify potential biomarkers for treatment response and resistance development. In Chapter 4, we focus on biomarker detection to monitor the clinical course of bacterial infections in patients with community acquired pneumonia (CAP), for early decision making concerning monitoring disease progression. To detect possible biomarkers for disease progression and treatment response, longitudinal, high-dimensional metabolomics data are analyzed with dimension reduction through PCA, to explore different biochemical metabolic classes and their roles in the changes over time.

Section III: Real world data

In Section III, we develop tools to study antibiotic resistance and patient characteristics in clinical routine health care data, to support translation of concepts studied in vitro and in silico to be researched in clinical pathogen and patient data. Chapter 5 describes a method for detection of collateral sensitivity in large clinical data on antibiotic susceptibility. Using this method, Chapter 6 explores collateral sensitivity in different bacterial species and over different antibiotic classes. In Chapter 7, the statistical concept of copulas is used as a method for simulation of virtual patients for pharmacometric research. Copulas are multivariate density functions that can be used to estimate the joint density of multiple variables. We evaluate its use for estimation joint densities and subsequent simulation of patient's covariates used in pharmacometric models.

Section IV: General discussion and summary

In Chapter 8 we discuss the findings in this thesis and the future prospects for the use of statistics and pharmacometrics to evaluate treatment response variability. We discuss overall themes that are of relevance for successful research in precision medicine.

References

Aulin, L. B. S., Liakopoulos, A., van der Graaf, P. H., Rozen, D. E., & van Hasselt, J. G. C. (2021, Sep). Design principles of collateral sensitivity-based dosing strategies. *Nature Communications*, *12*(1), 5691. Retrieved from http://dx.doi.org/10.1038/s41467-021-25927-3https://www.nature.com/articles/s41467-021-25927-3 doi: 10.1038/s41467-021-25927-3

Currie, C. J., & MacDonald, T. M. (2000). Use of routine healthcare data in safe and cost-effective drug use. *Drug Safety*, *22*(2), 97–102. Retrieved from https://doi.org/10.2165%2F00002018-200022020-00002 doi: 10.2165/00002018-200022020-00002

Depledge, M. H., Amaral-Mendes, J. J., Daniel, B., Halbrook, R. S., Kloepper-Sams, P., Moore, M. N., & Peakall, D. B. (1993). The conceptual basis of the biomarker approach. In *Biomarkers* (pp. 15–29). Springer Berlin Heidelberg. Retrieved from https://doi.org/10.1007%2F978-3-642-84631-1_2 doi: 10.1007/978-3-642-84631-1_2

Groenendijk, F. H., & Bernards, R. (2014, May). Drug resistance to targeted therapies: Déjà vu all over again. *Molecular Oncology*, *8*(6), 1067–1083. Retrieved from https://doi.org/10.1016%2Fj.molonc.2014.05.004 doi: 10.1016/j.molonc.2014.05.004

Heiser, W. J., Busing, F. M. T. A., & Meulman, J. J. (2020). Mapping Networks and Trees with Multidimensional Scaling of Proximities. In (pp. 385–407). Retrieved from http://link.springer.com/10.1007/978-981-15-2700-5_24 doi: 10.1007/978-981-15-2700-5_24

Hoerl, A. E., & Kennard, R. W. (1970, Feb). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. Retrieved from https://doi.org/10.1080%2F00401706.1970.10488634 doi: 10.1080/00401706.1970.10488634

Johnstone, I. M., & Titterington, D. M. (2009, Nov). Statistical challenges of high-dimensional data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *367*(1906), 4237–4253. Retrieved from https://doi.org/10.1098%2Frsta.2009.0159 doi: 10.1098/rsta.2009.0159

Maltas, J., & Wood, K. B. (2019, Oct). Pervasive and diverse collateral sensitivity profiles inform optimal strategies to limit antibiotic resistance. *PLOS Biology*, *17*(10), e3000515. Retrieved from https://doi.org/10.1371%2Fjournal.pbio.3000515 doi: 10.1371/journal.pbio.3000515

McCulloch, C. E., & Searle, S. R. (2000). *Generalized, linear, and mixed models*. Wiley. Retrieved from https://doi.org/10.1002%2F0471722073 doi: 10.1002/0471722073

Morrato, E. H., Elias, M., & Gericke, C. A. (2007, Dec). Using population-based routine data for evidence-based health policy decisions: lessons from three examples of setting and evaluating national health policy in australia, the UK and the USA. *Journal of Public Health*, *29*(4), 463–471. Retrieved from https://doi.org/10.1093%2Fpubmed%2Ffdm065 doi: 10.1093/pubmed/fdm065

Mould, D., & Upton, R. (2012, Sep). Basic concepts in population modeling, simulation, and model-based drug development. *CPT: Pharmacometrics & Systems Pharmacology*, *1*(9), 6. Retrieved from https://doi.org/10.1038%2Fpsp.2012.4 doi: 10.1038/psp.2012.4

Mould, D., & Upton, R. (2013, Apr). Basic concepts in population modeling, simulation, and model-based drug development-part 2: Introduction to pharmacokinetic modeling methods. *CPT: Pharmacometrics & Systems Pharmacology*, *2*(4), 38. Retrieved from https://doi.org/10.1038%2Fpsp.2013.14 doi: 10.1038/psp.2013.14

Nice, E. C. (2018, Jul). Challenges for omics technologies in the implementation of personalized medicine. *Expert Review of Precision Medicine and Drug Development*, *3*(4), 229–231. Retrieved from https://doi.org/10.1080%2F23808993.2018.1505429 doi: 10.1080/23808993.2018.1505429

Pál, C., Papp, B., & Lázár, V. (2015). Collateral sensitivity of antibiotic-resistant microbes. *Trends in Microbiology*, *23*(7), 401–407. doi: 10.1016/j.tim.2015.02.009

Roemhild, R., & Andersson, D. I. (2021, Jan). Mechanisms and therapeutic potential of collateral sensitivity to antibiotics. *PLOS Pathogens*, *17*(1), e1009172. Retrieved from https://doi.org/10.1371/journal.ppat.1009172.g001https://dx.plos.org/10.1371/journal.ppat.1009172 doi: 10.1371/journal.ppat.1009172

Smania, G., & Jonsson, E. N. (2021, Apr). Conditional distribution modeling as an alternative method for covariates simulation: Comparison with joint multivariate normal and bootstrap techniques. *CPT: Pharmacometrics & Systems Pharmacology*, *10*(4), 330–339. Retrieved from https://doi.org/10.1002%2Fpsp4.12613 doi: 10.1002/psp4.12613

Strimbu, K., & Tavel, J. A. (2010, Nov). What are biomarkers? *Current Opinion in HIV and AIDS*, *5*(6), 463–466. Retrieved from http://journals.lww.com/01222929-201011000-00003 doi: 10.1097/COH.0b013e32833ed177

Sun, X., & Hu, B. (2018, Nov). Mathematical modeling and computational prediction of cancer drug resistance. *Briefings in Bioinformatics*, *19*(6), 1382–1399. Retrieved from https://academic.oup.com/bib/article/19/6/1382/3886023 doi: 10.1093/bib/bbx065

Swift, B., Jain, L., White, C., Chandrasekaran, V., Bhandari, A., Hughes, D. A., & Jadhav, P. R. (2018, Sep). Innovation at the Intersection of Clinical Trials and Real-World Data Science to Advance Patient Care. *Clinical and Translational Science*, *11*(5), 450–460. Retrieved from http://doi.wiley.com/10.1111/cts.12559https://onlinelibrary.wiley.com/doi/10.1111/cts.12559 doi: 10.1111/cts.12559

Talebi Bezmin Abadi, A., Rizvanov, A. A., Haertlé, T., & Blatt, N. L. (2019, Dec). World Health Organization Report: Current Crisis of Antibiotic Resistance. *BioNanoScience*, *9*(4), 778–788. Retrieved from http://link.springer.com/10.1007/s12668-019-00658-4 doi: 10.1007/s12668-019-00658-4

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Royal Statistical Society*, *58*(1), 267–288. Retrieved from www.jstor.org/stable/2346178

Upton, R. N., & Mould, D. R. (2014). Basic Concepts in Population Modeling, Simulation, and Model-Based Drug Development: Part 3—Introduction to Pharmacodynamic Modeling Methods. *CPT: Pharma-*

cometrics & Systems Pharmacology, *3*(1), e88. Retrieved from http://doi.wiley.com/10.1038/psp.2013.71 doi: 10.1038/psp.2013.71

World Health Organization. (2014). *Antimicrobial resistance: global report on surveillance* (Tech. Rep.). Retrieved from https://apps.who.int/iris/bitstream/handle/10665/193736/9789241509763_eng.pdf

Yin, A., Moes, D. J. A., Hasselt, J. G., Swen, J. J., & Guchelaar, H. (2019, Oct). A Review of Mathematical Models for Tumor Dynamics and Treatment Resistance Evolution of Solid Tumors. *CPT: Pharmacometrics & Systems Pharmacology*, *8*(10), 720–737. Retrieved from https://onlinelibrary.wiley.com/doi/10.1002/psp4.12450 doi: 10.1002/psp4.12450

zur Wiesch, P. A., Kouyos, R., Engelstädter, J., Regoes, R. R., & Bonhoeffer, S. (2011, Mar). Population biological principles of drug-resistance evolution in infectious diseases. *The Lancet Infectious Diseases*, *11*(3), 236–247. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S1473309910702644 doi: 10.1016/S1473-3099(10)70264-4

Chapter 2

Beyond the Randomized Clinical Trial: Innovative Data Science to Close the Pediatric Evidence Gap

Authors

Sebastiaan C. Goulooze

Laura B. Zwep

Julia E. Vogt

Elke H. J. Krekels

Thomas Hankemeier

John N. van den Anker

Catherijne A. J. Knibbe

Clinical Pharmacology & Therapeutics 2020; 107(4), 786–795

Abstract

Despite the application of advanced statistical and pharmacometric approaches to pediatric trial data, a large pediatric evidence gap still remains. Here, we discuss how to collect more data from children by using real-world data from electronic health records, mobile applications, wearables, and social media. The large datasets collected with these approaches enable and may demand the use of artificial intelligence and machine learning to allow the data to be analyzed for decision making. Applications of this approach are presented, which include the prediction of future clinical complications, medical image analysis, identification of new pediatric end points and biomarkers, the prediction of treatment nonresponders, and the prediction of placebo-responders for trial enrichment. Finally, we discuss how to bring machine learning from science to pediatric clinical practice. We conclude that advantage should be taken of the current opportunities offered by innovations in data science and machine learning to close the pediatric evidence gap.

2.1 Introduction

Historically, the evidence basis of pediatric treatments has lagged behind those in adult patients. A key aspect of this is the lack of pediatric data, which originates from the logistic, ethical, and legal challenges of performing clinical investigations in children (Brussee et al., 2016). Additionally, the pediatric population is more heterogeneous than the adult population, with maturational differences in pharmacokinetics, pharmacodynamics, and disease etiology across the pediatric age range from preterm neonates to adolescents (Brussee et al., 2016). Consequently, data collected in children within a narrow age range might still leave us with limited information regarding the treatment of children outside the studied age range. Finally, similar to other patient populations, optimal treatment will also differ for individuals within the same age group, for instance, because of obesity, genetic polymorphisms, or disease severity, and should be improved with more personalized treatment approaches (Allegaert et al., 2017).

To date, academic hospitals and industry perform clinical studies and randomized clinical trials (RCTs) on current and new drugs in children. Many academic studies focus on commonly used drugs in hospitalized patients, as the in-patient situation facilitates the collection of data. Generally, to minimize the study burden on pediatric subjects, the frequency and amount of data collection is limited and often not standardized. For example, to limit the number of venous samples, drug concentrations in plasma might be quantified in scavenged samples that were taken as part of standard of care (Krekels et al., 2017). Population pharmacometric modeling approaches have been successfully used to deal with these unbalanced data to better understand pediatric pharmacology (Brussee et al., 2016; Krekels et al., 2017). More recently, we have seen an increased use of mechanistic or physiologically based models, which leverage prior knowledge regarding the physiological changes in organ weight, blood flow,

and protein expression during a child's life (Allegaert et al., 2017; Mehrotra et al., 2016; Rostami-Hodjegan, 2012). An important aspect of such models is their improved predictive performance when used to extrapolate from adults to children (Danhof, 2015).

These pharmacometric modeling approaches are now, despite limited data, being used with success to support neonatal and pediatric drug development as well as dosing of commonly used off-label drugs (Brussee et al., 2016; Mehrotra et al., 2016; Barker et al., 2018). However, recent failures of RCTs in children have taught us that there is more to these studies than confirming model-based predictions (Momper et al., 2015). These failures have been attributed to different reasons, such as an increased placebo effect in children, different disease etiology compared with adults, and inadequate dose selection (Momper et al., 2015). Another important cause is the failure to recruit sufficient patients, which can force investigators to costly increases of the study duration or even premature termination of a study due to low feasibility of recruiting the target sample size (Joseph et al., 2015; Denhoff et al., 2015). Failed drug trials—and the general lack of pediatric clinical trials being performed particularly in primary health care—contribute to the high prevalence of off-label drug use in children, especially in the first years of life (Yackey et al., 2019). It is clear that despite the advances in approaches to data collection and analysis, a large need for additional research in pediatrics still remains.

To tackle the limitations of conventional clinical research, we need to move beyond the RCTs and their analysis with traditional statistical and advanced pharmacometric techniques. In this narrative review, we will discuss novel approaches to collecting data in pediatric patients to get more information from both clinical trials and real-world data. In addition, we will discuss how large datasets that are derived from new data collection approaches enable, and may demand, the use of innovative data science approaches, such as machine learning. Finally, we will discuss both applications and challenges to the widespread use of machine learning in pediatric medicine. Together, these innovations have the potential to greatly support our ability to generate high-quality evidence to guide optimal pediatric clinical care, thereby closing the pediatric evidence gap.

2.2 Advances in pediatric data collection

Improving our capacity for pediatric data collection is necessary for closing the pediatric evidence gap. Pediatric (randomized) clinical studies are costly and time-consuming to perform, and a sole reliance on these studies may limit our capacity for medical research in children. These studies are generally site-centric, meaning that most data is collected in a hospital or physical study site. Figure 2.1 illustrates how the capacity for pediatric data collection can be increased by moving beyond site-centric pediatric studies toward real-world data and new techniques for patient-centric data collection (Swift et al., 2018). Below we elaborate on the different opportunities and challenges (ethical and privacy) of these advances in data collection in pediatrics.

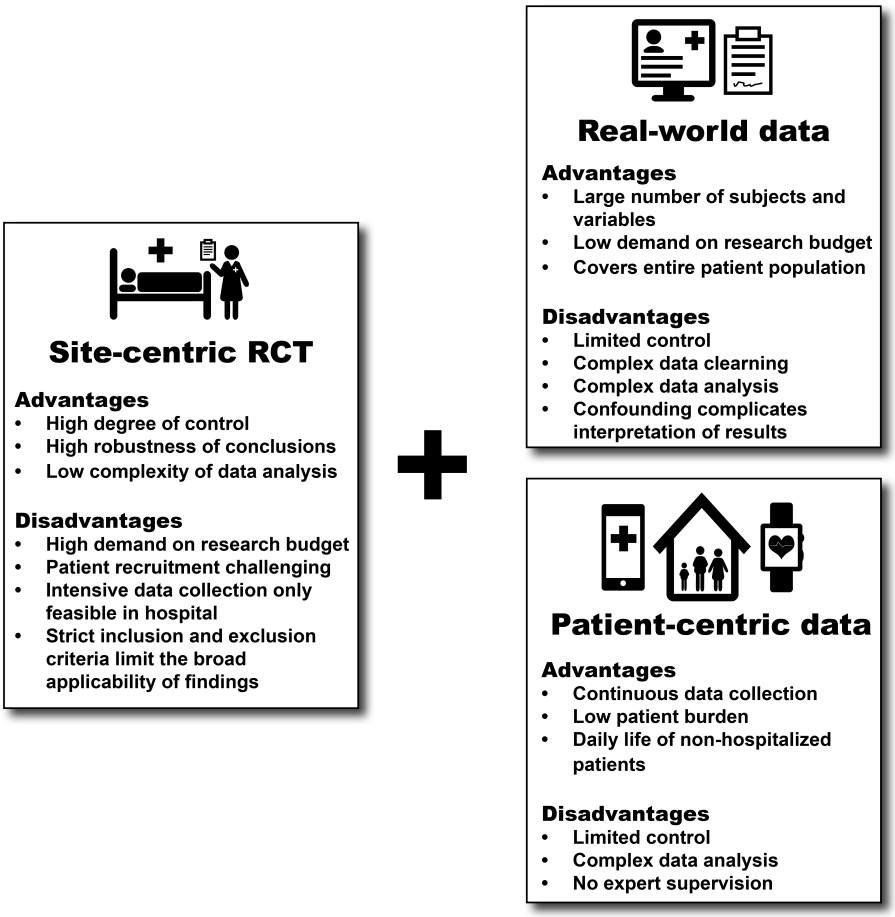


Figure 2.1: Innovative pediatric data collection beyond the site-centric randomized clinical trial (RCT). Data from these pediatric clinical studies can be supplemented by increased use of real-world pediatric data from electronic health records. Additional information can be obtained without increasing patient burden by using patient-centric data collection tools, such as mobile applications, wearables, and social media data. Site-centric RCTs refer to studies in which data collection is limited to one or more hospitals or physical study sites. Patient-centric data refers to data collected from the patient at home or during other parts of their daily routine.

2.2.1 Real-world data collection

The collection of real-world data through electronic health records (EHRs) has sharply increased in the last decade, which opens up an unprecedented potential for data collection with more subjects, more variables, and lower costs (Goldstein et al., 2016). The use of EHR data for research purposes comes with its own set of challenges, due to the large amount of data and variables to be analyzed. Machine learning techniques are often required to maximize the information extracted from EHRs. In addition to

large amounts of structured data, a part of the information in EHRs is hidden in clinical or laboratory notes, which complicates data analysis when this information is required to answer a particular question (Swift et al., 2018). To extract information from such notes into structured data, techniques like natural-language processing may provide a great opportunity for answering pediatric research questions (Nadkarni et al., 2011; Savova et al., 2016). These techniques enable analyses that would be impossible to perform on the text data itself when it would be too time-consuming to do a manual extraction of the relevant features from the text data. For example, in radiology, natural-language processing was used to automatically notate whether a certain condition or finding is mentioned within the text of the report (Pons et al., 2016). In another example, Liang and others used natural-language processing to allow the use of unstructured information from EHRs for the development of a deep-learning model for automatic pediatric diagnoses that surpassed the accuracy of junior, but not senior, physicians (H. Liang et al., 2019). Finally, the data extracted using natural-language processing might be required to identify patients eligible for inclusion in cohorts for observational research (Savova et al., 2016).

Although effectiveness research with real-world data can be problematic due to the difficulty in controlling for confounding variables and nonrandomized treatment decisions, real-world data offer many other opportunities (Swift et al., 2018; Eichler et al., 2018; Miksad et al., 2019). First, real-world data might be used to generate or select hypotheses on the most effective treatment that can then be tested in an RCT. Alternatively, real-world data might be used to confirm that the findings in a well-controlled RCT also apply to the wider, more heterogeneous pediatric population or establish that some subpopulations require additional research (Eichler et al., 2018). Additionally, real-world data can also be used to better characterize patients outside clinical studies as natural history cohorts that can subsequently be used as an external control to replace placebo arms in pediatric trials (Miksad et al., 2019). Although externally controlled studies require additional considerations to deal with potential biases compared with traditional RCTs, this approach might provide an opportunity for performing studies in cases where sufficiently powered RCTs are difficult to perform due to rarity of the indication or reluctance of parents to consent to a placebo-controlled trial (Miksad et al., 2019; Dejardin et al., 2017; Food and Drug Administration (FDA), 2019a). Finally, real-world data may be more suitable than RCTs for answering drug safety questions regarding rare adverse effects or adverse effects that present themselves years after the initial drug exposure (Eichler et al., 2018; McMahon & Pan, 2018).

To deliver the best medical practice tomorrow, it is important that we harness the full potential of the data collected today. At the moment, data in EHRs are still primarily collected for medical practice and may sometimes be ill suited for secondary use as research data. This is compounded by the fact that physicians are primarily responsible for treating patients and not for generating high-quality research data (Eichler et al., 2018). In a learning healthcare system, real-world data are not only collected to treat the individual patient but also readily usable to improve clinical practice by contributing to the generation of knowledge and innovations (Eichler et

al., 2018; C. P. Friedman et al., 2016). Examples of initiatives include the PEDSnet learning healthcare system, a large clinical data research network that currently holds data of over 6 million children from 2009 onward and has enabled the generating of real-world evidence in a variety of clinical settings, including obesity, leukemia, and long-term safety of (maternal) drug use (Forrest et al., 2014). In addition, important are initiatives like the European EHR4CR project (Moor et al., 2015) that support the integration of data from different EHR systems, as this allows the creation of larger datasets, and the external validation of findings in datasets from different sites (Goldstein et al., 2016; Eichler et al., 2018).

2.2.2 Patient-centric data collection

In addition to data from site-centric RCTs, in which most data is collected in one or more physical study sites (McMahon & Pan, 2018), the collection of patient-centric data has the potential to increase the capacity for data collection (Figure 2.1) (McMahon & Pan, 2018; C. P. Friedman et al., 2016). Patient-centric data refers to data collected from the patient at home or during other parts of their daily routine. Depending on the context, data could be collected using mobile applications, wearables, and social media. A specific advantage of patient-centric data is the increase of study data without increasing the study burden associated with additional study visits that may, in the case of children, affect their parents or caregivers as well. The opportunities of patient-centric data collection are particularly important for studying chronic diseases in children that do not require hospitalization or frequent hospital check-ups as part of their treatment. Another potential application would be the long-term follow-up of previously hospitalized patients.

Mobile applications. In its simplest form, a mobile application might be an electronic diary, designed to collect self-reported outcomes, which can be reported by children when they are beyond a certain age or by the parents in case of younger children. Compared with a paper diary, electronic diaries are reported to improve compliance with alerts and to reduce the risk of errors during data entry (Izmailova et al., 2018). In other cases, the primary aim of the application is to promote healthy behavior in the child through motivation or education, for example, in applications that help older children with self-management of asthma or type 1 diabetes (Majeed-Ariss et al., 2015). The interactions by the child and/or their parents with these applications may offer great opportunities for data collection.

Wearables. The use of wearables creates the possibility of continuous data collection in an at-home setting, which supports characterizing the intra-individual and inter-individual variability in disease and drug response, as well as quantifying exposure–response relationships for drugs in the pediatric population (Kothare et al., 2018). The latter is especially true if the clinical outcome or a surrogate end point can be quantified at home. Similar to mobile applications, the wearable itself might not only be

used to collect data but also to motivate desirable behavior. For example, Hooke et al. evaluated the use of activity trackers to promote physical activity in children with acute lymphoblastic leukemia in an effort to reduce treatment-induced fatigue (Hooke et al., 2016).

Wearables can also include biochemical sensors to noninvasively measure electrolytes, metabolites, and proteins in an at-home setting. Wearables worn on the skin can be used to measure analytes directly in sweat, but can also noninvasively extract analytes, such as proteins and glucose from the skin's interstitial fluid (Kim et al., 2019). Although many analytes of interest cannot yet be measured using wearable sensors, future developments in this area will likely expand the applicability of these techniques for patient-centric collection of pediatric biochemical data.

Although these wearables may provide great opportunities for data collection in otherwise difficult to study patient populations, like children, it is important to note that the field of clinical application of wearables is still in its infancy when we consider its clinical utility, even for adult patients (Khozin & Coravos, 2019). There are a variety of challenges that need to be met in scientific, logistic, ethical, and privacy aspects, as covered extensively by a recent review by Izmailova et al (Izmailova et al., 2018). For example, commercially available wearables frequently do not report the raw data, but only the summary or secondary data that has been processed with undisclosed and proprietary algorithms. This complicates the interpretation of wearable data, especially when collecting data from multiple types of wearables with differing terminology and data standards. For the pediatric application of wearables, additional validation will be required to ensure devices are also fit-for-purpose for children of a particular age group, and whether the data measured with these devices have the same relevance for the clinical outcome. Finally, the use of wearables by study participants might affect their behavior (e.g., they might walk more when wearing a wearable that tracks their daily step count), which could be a problem depending on the research question and design of the study. Despite these challenges, their ability for continuous data collection at low burden to the patient could provide a great opportunity in the effort to fill the pediatric evidence gap, especially if the link can be made to clinical outcomes and biomarkers.

Social media data. The use of social media has increased dramatically over the last decade. It has been reported that children who use medication might use these platforms to share experiences that are not communicated to their healthcare practitioner (Dreisbach et al., 2019). As such, social media might contain information useful to pediatric pharmacovigilance that is not available elsewhere. Recent studies explored patient reports of adverse effects on social media platforms, such as Twitter (Patel et al., 2018) and patient fora (Marshall et al., 2015). This information was explored by counting how many times different adverse effects were mentioned in combination with a certain drug. These studies could serve as a method for signal detection of rare adverse effects, or to supplement information on known adverse effects that are underestimated in children.

At the moment, the use of social media data for pharmacovigilance is still in its

infancy. In a recent study from the IMI project WEB-RADR, natural-language processing techniques that were used to automatically label social media posts with drugs and adverse effects combinations were only correct in about 40% of the cases (van Stekelenborg et al., 2019). Using these imprecise techniques, the authors found no indication that posts on general social media platforms like Facebook and Twitter would have an added value to traditional methods of pharmacovigilance. Another challenge identified in the WEB-RADR project is that some drugs are hardly discussed in social media posts, thus having little to no potential for advancing pharmacovigilance (van Stekelenborg et al., 2019). The use of social media posts in pharmacovigilance might be more beneficial with further advances in natural-language processing and by directing research efforts toward patient fora, which would carry a higher percentage of relevant posts than general social media platforms.

2.3 Ethical and Privacy Aspects of Pediatric Data Collection

Innovations in data collection will support our ability to effectively treat pediatric patients in the future, especially when the collected data is Findable, Accessible, Interoperable, Reusable to allow secondary analyses to be performed by the broader research community. These benefits need to be balanced with the right to privacy of the patients whose data are used in this research. Maintaining and further developing ethical and data security standards are crucial to ensure ongoing support by patients and their parents of data collection for research purposes (Shaw et al., 2019). Maintaining data security is particularly challenging for patient-centric data collection where sensitive data are collected on a mobile phone or wearable, as data leaks could occur when the device is lost or during data transfer from the device to the central database.

Appropriate security measures need to be in place to minimize the risk of violating the patient's privacy. In this respect, the removal of identifying information can contribute to maintaining privacy when using data for research purposes. However, when the research question requires that data from different databases are linked, some form of patient identifier might be needed to do this (Currie, 2013). A potential solution to this issue is to add a small amount of noise to the data to ensure patients cannot be identified (Currie, 2013). Another interesting approach is to "share the answers, not the data." In this case, a data analysis or model might be run on the data, and only the aggregated results are returned to the researchers.

The issue of consent is particularly complex for pediatrics. Depending on the age of the child, (written) informed consent might be obtained from the parents, the child, or both. However, in the case of reuse of the data, there are questions that remain unanswered (Taylor et al., 2017). Can the parental informed consent be considered to be valid for reuse of the data years later, even if the patient has since reached adolescence or adulthood? It is recognized that retrospectively obtaining informed consent for large datasets of observational real-world data could likely result in lengthy

and costly procedures, which would limit their use in practice (Currie, 2013). However, for some observational analyses of de-identified data, the need for informed consent can be waived by institutional review board, if appropriate privacy measures are taken (Currie, 2013).

2.4 Machine Learning for Evidence Generation

Innovations in pediatric data collections provide great opportunities for research and hold great promise in closing the pediatric evidence gap, but this promise can only be fulfilled if these data are used effectively to address clinically relevant questions. To do so is challenging due to the size and complexity of datasets collected with these novel techniques. Collecting new types of data will, therefore, go hand in hand with the increasing use of artificial intelligence and machine learning in pediatrics.

The term machine learning is often used interchangeably with the term artificial intelligence (AI). AI is an area in the discipline of computer science that aims to create intelligently perceiving, reasoning, and acting machines. A subset of AI is machine learning, which encompasses a wide range of advanced data analyses techniques. Depending on techniques used, machine learning algorithms can predict both numerical outcomes (e.g., a disease severity score) or class labels (e.g., healthy vs. diseased).

With respect to the different classes of machine learning techniques, linear models are an easy to interpret class of machine learning techniques for the analysis of structured data (Figure 2.2). Linear regression, which is the most common linear modeling technique, can be used for both prediction and hypothesis testing, but is not suitable when there are many variables in the dataset. In those cases, penalized regression techniques can be used, which have a penalty term to constrain overfitting. Examples of such techniques include lasso (Tibshirani, 1996) and ridge regression (Hoerl & Kennard, 1970). A second class of machine learning techniques are tree-based models, such as Classification and Regression Trees (Breiman et al., 1984; J. H. Friedman, 2001) and random forests (Breiman, 2001a). Depending on the specific type of technique, the output of a tree-based model might be a form of a decision tree, which can still be relatively well explained. A third class of machine learning techniques is deep learning or deep neural networks. Deep learning has been used extensively for image analysis and text mining outside the medical world and has recently started to be used on medical images and EHRs (Figure 2.2) (Miotto et al., 2017). Complex deep-learning models can have a good predictive performance when dealing with unstructured data due to flexibility of such models (Figure 2.2). However, deep-learning models are often difficult to explain, as it is generally difficult to understand how the input data leads to the model prediction.

Of note, it is important to recognize that machine learning will supplement, and not replace, traditional statistics in pediatric research. The use of traditional statistical tests or linear models might be more appropriate if the primary goal of the analysis is not to obtain a prediction model (Breiman, 2001b; Donoho, 2017). This includes situations when the goal of the analysis is hypothesis testing ("Does the treatment

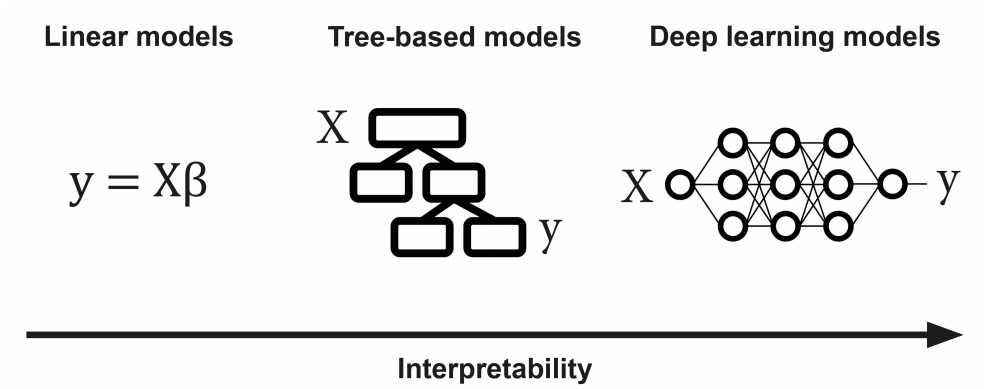


Figure 2.2: Explicability of the various machine learning techniques. On the far left, linear models have a clear explanation, but require that the data are structured. Linear regression, the most common linear modeling technique, can be used for both prediction and hypothesis testing, but is not suitable when there are many variables in the dataset. On the opposite end of the spectrum, deep-learning models are generally difficult to interpret and explain, and not suitable for hypothesis testing. However, due to the flexibility of deep-learning models, they are able to handle complex and unstructured data, such as image and text data. Depending on the data (structured or unstructured) and the goal of the analysis (raw predictive performance or testing hypotheses), different techniques will be most appropriate.

work better than placebo?") or estimation of treatment effect ("What effect does the treatment have on the outcome?"). However, there are various clinical problems in which the predictions made by machine learning can contribute to closing the pediatric evidence gap, as will be illustrated with examples in the next section.

2.5 Applications of Machine Learning in Pediatrics

The opportunities offered by the various machine learning techniques can benefit pediatric practice in a variety of ways. In this section, we will discuss different applications of machine learning in pediatrics, including: the prediction of future clinical complications, medical image analysis, identification of new pediatric end points and biomarkers, prediction of treatment nonresponders, and the prediction of placebo-responders to allow pediatric trial enrichment.

2.5.1 Predicting future clinical complications

The ability to predict clinical complications in the future can be used to deliver more personalized medicine in pediatrics. For this purpose, machine learning plays a crucial role due to its improved potential predictive performance compared with traditional statistical methods, especially when the data are unstructured or otherwise complex. Children, who are predicted to be at high risk for a certain event, can subsequently be monitored and treated more intensively. In recent research, new algorithms have been explored to make good predictions using data from previous

studies or real-world data. Box 1 shows three case studies in which machine learning techniques were used to make predictions about future clinical complications, such as childhood obesity (Dugan et al., 2015), late onset sepsis (Mani et al., 2014), and neonatal hyperbilirubinemia (Daunhawer et al., 2019).

Box 1 Prediction of clinical complications in pediatrics using machine learning

Case study 1. Childhood obesity Dugan et al. (2015) explored predictors of childhood obesity, with the aim of eventually being able to provide targeted obesity prevention for high-risk children. The answers on a dynamic questionnaire and measurements of clinical staff were mined from over 7000 children below the age of 2 years. These features were used to predict the prevalence of obesity after their second birthday. Using tree-based machine learning, an accurate model predicting childhood obesity was obtained, which included predictors like pre-existing obesity, ethnicity, height and maternal depression.

Case study 2. Neonatal sepsis Mani et al. (2014) evaluated the usefulness of different classification algorithms to predict late onset sepsis in neonates, using early results of laboratory tests and nursing observations. The best classification algorithm surpassed the clinician in both the sensitivity and specificity of predicting neonatal sepsis. After validation, clinical implementation could allow earlier treatment of sepsis while reducing the number of patients unnecessarily treated with antibiotics.

Case study 3. Neonatal hyperbilirubinemia Daunhawer et al. (2019) used machine learning techniques to predict neonatal hyperbilirubinemia. An ensemble classifier combining the logistic regression lasso and random forests was able to predict accurately whether a neonate would undergo phototherapy treatment in the next 48 hours. The predictions were made using clinical variables, such as birth weight and health information about the mother. This model could support a more personalized bilirubin monitoring approach, with more intensive monitoring of high-risk patients.

2.5.2 Medical image analysis

Deep-learning models have been particularly effective in image analysis, mainly in radiology (Yamashita et al., 2018). A deep-learning model can learn to classify images as healthy or diseased or can notate the areas in the image that correspond to organs or other anatomic structures. For example, a deep-learning model was able to identify the segmentation of white matter, gray matter, and cerebrospinal fluid in the brains of babies (Zhang et al., 2015). The automation of these tasks with a deep-learning model can reduce the time spent on an image by limiting the radiologist's task to checking and adjusting the lines drawn by the algorithm. In another example, a deep-learning model was able to identify the skeletal maturity of children by assessing hand radiographs (Larson et al., 2018). Another common application is the detection of malignant tumors in medical images, which could serve as a second opinion to detect malignancies that might have been missed by the radiologist (Suzuki et al.,

2005; M. Liang et al., 2016).

In addition to increasing efficiency, deep-learning models could also extract information from image data that is not included in the radiologist report. This would include features that are too complex and time-consuming to extract manually or features that are not currently being used in clinical decision making (Hosny et al., 2018). With automated extraction of additional information from medical images, deep learning-based image analysis can be used to perform research on imaging-based pediatric biomarkers that would not be feasible with manual image analysis.

2.5.3 Identifying end points and biomarkers in pediatrics

The development and validation of pharmacodynamic end points for children is recognized as an important methodological step in closing the evidence gap of pediatric medicine (Kelly et al., 2018). Having suitable disease-specific pharmacodynamic end points for children is essential for demonstrating efficacy and for establishing the exposure-response relationship of drugs needed for pediatric drug labeling. Additionally, these measures of patient disease severity or well-being can guide treatment decisions in clinical practice. For this, the efficacy and safety end points used in adults may not be fit-for-purpose across the pediatric age range: the clinical end point might not occur until later in life, might not be directly measurable, or the clinical presentation of the disease might differ too much from any adult counterpart (Kelly et al., 2018).

Machine learning can be used in biomarker and end-point discovery by performing variable selection and dimension-reduction when there are multiple variables considered to be potentially relevant for pediatric outcome. For example, Hartley et al. used electroencephalography data to derive a summary measure for nociceptive brain activity in infants (Hartley et al., 2017). In this example, the electroencephalography-based measure of pain was learned from the context (i.e., by comparing the response profiles after non-noxious or noxious stimulation). In another example, a supervised learning approach was used to derive a measure of iatrogenic withdrawal severity in children by combined analysis of nurse's expert opinion of the child's withdrawal severity and the observed withdrawal symptoms (Goulooze et al., 2019). Finally, machine learning may be used to identify early biomarkers that correspond to long-term clinical end points or quality of life (Bera et al., 2019). For example, a machine learning tool is currently being developed to analyze cough sound data as a digital biomarker of acute respiratory disease in children (Coravos et al., 2019).

2.5.4 Predicting treatment responders

Machine learning techniques can also be used to identify nonresponders (i.e., children who are unlikely to respond to a particular treatment). The clinical benefit lies in avoiding therapy that might give adverse effects at low chance of beneficial effects, as well as reducing the need for trial-and-error approaches for treatment personalization (Doherty et al., 2018).

For the adults, machine learning techniques have been used to predict nonresponders to drug treatment in different settings, including oncology, immunology, and postoperative pain (Doherty et al., 2018; Gram et al., 2016; Huang et al., 2018). Depending on the similarity of disease between adults and children, and the explicability and the biological plausibility of the machine learning model, models developed in adults might be also applicable in the pediatric setting after validation. In other cases, the pediatric pathophysiology might be too different or the disease might be absent in adults. In this case, efforts would be warranted to develop new machine learning models to predict drug response in children, so that they can also benefit from these innovations.

2.5.5 Predicting placebo responders to improve trial success

Prospective (randomized) clinical trials remain the gold standard to get drugs registered for the pediatric population. However, some of these RCTs fail to demonstrate efficacy in children (Momper et al., 2015). These failures have been attributed to a numbers of reasons, one of which is the high placebo response observed in indications such as depression, migraine, and bipolar disorder (Momper et al., 2015). A high placebo response would limit the ability of a trial to demonstrate efficacy or would require a very large sample size to do so. Additionally, it has been shown that younger children tend to have a stronger placebo response than older children (Weimer et al., 2013). This would make it especially difficult to demonstrate efficacy in younger children, which is problematic considering that the off-label drug use is highest in children in the first year of life (Yackey et al., 2019).

One way to limit the impact of placebo response on trial outcomes would be to identify baseline predictors of placebo response so that trials can be enriched prerandomization with subjects that are less likely to respond strongly to placebo (Momper et al., 2015). This strategy has been used in pediatric trials, resulting, for example, in the successful application for a pediatric indication of rizatriptan for acute treatment of migraine (Sun et al., 2013). For adults, it has been proposed that machine learning techniques may have better predictive power when using multiple variables to predict placebo response, as was demonstrated for depression in a geriatric population (Zilcha-Mano et al., 2018). The use of machine learning techniques to reduce the placebo response in pediatric trials might, therefore, increase the success rate of pediatric drug trials and support pediatric drug labeling.

2.6 Bringing Machine Learning to Pediatric Practice

Whereas promising, more work needs to be done before the machine learning applications mentioned in the previous section are ready for widespread clinical use in children. Methods for predicting placebo response need to be developed for different therapeutic indications and prove their worth in practice by increasing the success of pediatric registration trials (Figure 2.3, left column). Biomarkers and end points

suggested by machine learning need to be validated and supported by the relevant stakeholders (Figure 2.3, middle column). When this is the case, having better pediatric end points and biomarkers will impact not only pediatric practice, but also pediatric research. Considerable work is also required to bring a machine learning model to the clinic as a medical decision support tool, as this requires extensive external validation of the model, the development of a user-friendly software tool, and assessment of the impact of the use of this tool in clinical practice (Figure 2.3, right column). Below, we will discuss the issue of validation of machine learning models for clinical use and the particular challenges of implementing medical decision support tools in pediatric clinical practice.

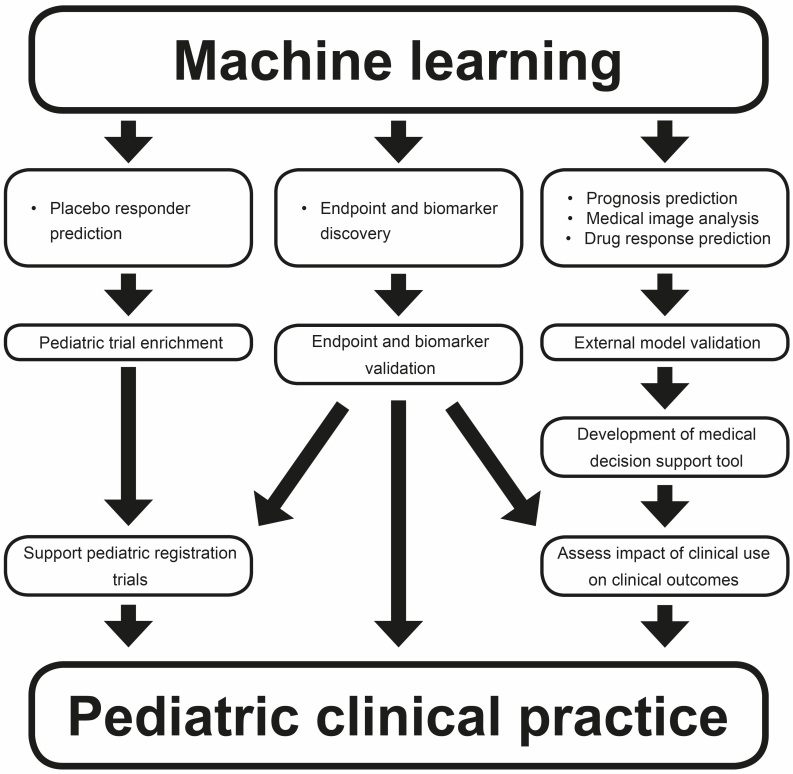


Figure 2.3: How applications of machine learning in pediatrics can support pediatric clinical practice.

2.6.1 Validation for clinical use in pediatrics

Machine learning models enable us to use complex data to achieve improved predictions of health and disease in children compared with traditional methods. However, it is important that the trained model does not "overfit" the data. An overfitted machine learning model has good predictive performance in the dataset it was trained on, but

poor performance when predicting for new cases. Validation of the model on an independent test data set is, therefore, essential to ensure the scientific quality and the clinical utility of the model (Figure 2.3, right column). Obtaining suitable datasets for this external validation can be challenging, especially in pediatric research, which underlines the importance of efforts to promote data sharing and the use of real-world data for research purposes (Ince et al., 2009).

Considering the heterogeneity of the pediatric population aged 0–18 years, it is also important to consider that a model might have a good predictive performance for children in a particular age group, but a poor performance for others (e.g., preterm neonates vs. term neonates). This risk is particularly high if certain age groups are underrepresented or absent in the dataset used to develop the model (Vayena et al., 2018). Transparency about the validity of the model and for which pediatric population this validity has been shown is, therefore, crucial.

Finally, it is important to recognize that even externally validated model predictions are not guaranteed to improve patient outcome when used in clinical practice. Some have, therefore, proposed that the clinical use of models as medical decision support tools should be supported by studies that demonstrate their impact on relevant clinical end points (Figure 2.3, right column) (Darcy et al., 2016). Considering the added difficulty to perform such trials in children, we argue that it is important to consider the need for such trials on a case-by-case basis, depending on the potential risk and benefits of the use (and nonuse) of machine learning tools in clinical decision making. In cases where dedicated pediatric trials are not feasible, modeling and simulation workflows used in pharmacometrics might be used to assess the likely clinical benefit-risk ratio of decision support tools by integrating available data from adult and pediatric patients (Bellanti et al., 2015).

2.6.2 Implementation of medical decision support tools

Implementation of findings from machine learning studies into pediatric clinical practice will not happen without focused efforts and close involvement of the various stakeholders. Currently, the widespread clinical implementation of scientific evidence is a lengthy process (> 15 years on average) and only achieved in about half of the cases (Bauer et al., 2015). Wittmeier et al. have argued in favor of systematic stepwise approaches to bring scientific knowledge to pediatric clinical practice. An important aspect of this is to engage in activities that have been shown to successfully support implementation, such as educational outreach and meetings, use of local opinion leaders, computerized reminders, audit, and feedback (Wittmeier et al., 2015). For the implementation of machine learning as a medical decision support tool in pediatrics, there are additional challenges to overcome (Figure 2.3, right column) (Shaw et al., 2019).

Because the predictions or classifications of machine learning tools can incorporate information of multiple variables, they are not as readily integrated in clinical guidelines as knowledge that relies on a single variable (e.g., age or bodyweight) for decision making. Therefore, software packages might be needed so that physicians

can easily use models in medical decision making (Figure 2.3, right column). It is important to stress that such software packages should be quick and simple to use and ideally linked to the EHR system so that there is no need for error-prone data entry of a large number of variables by the clinician.

The need to integrate machine learning tools into software packages does complicate their implementation, as they can be classified by the US Food and Drug Administration (FDA) as a medical device if the physician is not able to independently evaluate the basis of the recommendation (Food and Drug Administration (FDA), 2019b). With complex machine learning models, this is likely the case. Many software packages that provide recommendations based on models obtained with machine learning would, therefore, require lengthy regulatory approval procedures before they can be used in clinical practice.

In addition to being easy to use, the advice of the model should be explicable by the clinician. Here lies a key challenge for machine learning tools, especially for techniques like neural networks, which provide more "black box" predictions (Zorc et al., 2019). The integration of such black box predictions in clinical decision making is problematic because it means a departure from the paradigm of evidence-based medicine (Adkins, 2017). Additionally, shared decision making between the patient and physician also requires that decisions supported by machine learning tools can also be explained (Vayena et al., 2018; Zorc et al., 2019). Therefore, explicability for both the physician and the patient is likely a requirement for meaningful contributions to the decision process. Ongoing efforts to improve the explicability of complex machine learning models are, therefore, crucial to support their clinical acceptance and implementation (Cabitza et al., 2017).

2.7 Conclusions

Innovations in data collection and analysis could revolutionize many aspects of medical science and clinical practice in the upcoming decades. With the increased use of real-world data within a learning healthcare system and patient-centric data collection there is a potential to significantly expand our capacity for pediatric data collection. There are many useful potential applications of the predictive performance of machine learning models, and future work may integrate these applications with mechanistic modeling to improve understanding of the underlying biology. In addition, even though efforts are required to bring these innovations to the clinic, it is crucial that we capitalize on this opportunity to close the pediatric evidence gap.

References

Adkins, D. E. (2017, Feb). Machine learning and electronic health records: A paradigm shift. *American Journal of Psychiatry*, *174*(2), 93–94. Retrieved from https://doi.org/10.1176%2Fappi.ajp.2016.16101169 doi: 10.1176/appi.ajp.2016.16101169

Allegaert, K., Simons, S. H., Tibboel, D., Krekels, E. H., Knibbe, C. A., & van den Anker, J. N. (2017, Nov). Non-maturational covariates for dynamic systems pharmacology models in neonates, infants, and

children: Filling the gaps beyond developmental pharmacology. *European Journal of Pharmaceutical Sciences*, *109*, S27–S31. Retrieved from https://doi.org/10.1016%2Fj.ejps.2017.05.023 doi: 10.1016/j.ejps.2017.05.023

Barker, C. I. S., Standing, J. F., Kelly, L. E., Faught, L. H., Needham, A. C., Rieder, M. J., ... Offringa, M. (2018, Apr). Pharmacokinetic studies in children: recommendations for practice and research. *Archives of Disease in Childhood*, archdischild–2017–314506. Retrieved from https://doi.org/10.1136%2Farchdischild-2017-314506 doi: 10.1136/archdischild-2017-314506

Bauer, M. S., Damschroder, L., Hagedorn, H., Smith, J., & Kilbourne, A. M. (2015, Sep). An introduction to implementation science for the non-specialist. *BMC Psychology*, *3*(1). Retrieved from https://doi.org/10.1186%2Fs40359-015-0089-9 doi: 10.1186/s40359-015-0089-9

Bellanti, F., van Wijk, R. C., Danhof, M., & Pasqua, O. D. (2015, Jul). Integration of PKPD relationships into benefit-risk analysis. *British Journal of Clinical Pharmacology*, *80*(5), 979–991. Retrieved from https://doi.org/10.1111%2Fbcp.12674 doi: 10.1111/bcp.12674

Bera, K., Schalper, K. A., Rimm, D. L., Velcheti, V., & Madabhushi, A. (2019, Aug). Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nature Reviews Clinical Oncology*, *16*(11), 703–715. Retrieved from https://doi.org/10.1038%2Fs41571-019-0252-y doi: 10.1038/s41571-019-0252-y

Breiman, L. (2001a, Oct 01). Random forests. *Machine Learning*, *45*(1), 5-32. Retrieved from https://doi.org/10.1023/A:1010933404324 doi: 10.1023/A:1010933404324

Breiman, L. (2001b, Aug). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3). Retrieved from https://doi.org/10.1214%2Fss%2F1009213726 doi: 10.1214/ss/1009213726

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and Regression Trees. In (p. 368).

Brussee, J. M., Calvier, E. A. M., Krekels, E. H. J., Välitalo, P. A. J., Tibboel, D., Allegaert, K., & Knibbe, C. A. J. (2016, Jun). Children in clinical trials: towards evidence-based pediatric pharmacotherapy using pharmacokinetic-pharmacodynamic modeling. *Expert Review of Clinical Pharmacology*, *9*(9), 1235–1244. Retrieved from https://doi.org/10.1080%2F17512433.2016.1198256 doi: 10.1080/17512433.2016.1198256

Cabitza, F., Rasoini, R., & Gensini, G. F. (2017, Aug). Unintended consequences of machine learning in medicine. *JAMA*, *318*(6), 517. Retrieved from https://doi.org/10.1001%2Fjama.2017.7797 doi: 10.1001/jama.2017.7797

Coravos, A., Khozin, S., & Mandl, K. D. (2019, Mar). Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *npj Digital Medicine*, *2*(1). Retrieved from https://doi.org/10.1038%2Fs41746-019-0090-4 doi: 10.1038/s41746-019-0090-4

Currie, J. (2013, Apr). "big data" versus "big brother": On the appropriate use of large-scale data collections in pediatrics. *Pediatrics*, *131*(Supplement_2), S127–S132. Retrieved from https://doi.org/10.1542%2Fpeds.2013-0252c doi: 10.1542/peds.2013-0252c

Danhof, M. (2015, Aug). Kinetics of drug action in disease states: towards physiology-based pharmacodynamic (PBPD) models. *Journal of Pharmacokinetics and Pharmacodynamics*, *42*(5), 447–462. Retrieved from https://doi.org/10.1007%2Fs10928-015-9437-x doi: 10.1007/s10928-015-9437-x

Darcy, A. M., Louie, A. K., & Roberts, L. W. (2016, Feb). Machine learning and the profession of medicine. *JAMA*, *315*(6), 551. Retrieved from https://doi.org/10.1001%2Fjama.2015.18421 doi: 10.1001/jama.2015.18421

Daunhawer, I., Kasser, S., Koch, G., Sieber, L., Cakal, H., Tütsch, J., ... Vogt, J. E. (2019, Mar). Enhanced early prediction of clinically relevant neonatal hyperbilirubinemia with machine learning. *Pediatric Research*, *86*(1), 122–127. Retrieved from https://doi.org/10.1038%2Fs41390-019-0384-x doi: 10.1038/s41390-019-0384-x

Dejardin, D., Delmar, P., Warne, C., Patel, K., van Rosmalen, J., & Lesaffre, E. (2017, Dec). Use of a historical control group in a noninferiority trial assessing a new antibacterial treatment: A case study and discussion of practical implementation aspects. *Pharmaceutical Statistics*, *17*(2), 169–181. Retrieved from https://doi.org/10.1002%2Fpst.1843 doi: 10.1002/pst.1843

Denhoff, E. R., Milliren, C. E., de Ferranti, S. D., Steltz, S. K., & Osganian, S. K. (2015, Oct). Factors associated with clinical research recruitment in a pediatric academic medical center—a web-based survey. *PLOS ONE*, *10*(10), e0140768. Retrieved from https://doi.org/10.1371%2Fjournal.pone.0140768 doi: 10.1371/journal.pone.0140768

Doherty, M. K., Ding, T., Koumpouras, C., Telesco, S. E., Monast, C., Das, A., ... Schloss, P. D. (2018, May). Fecal microbiota signatures are associated with response to ustekinumab therapy among crohn's disease patients. *mBio*, *9*(2). Retrieved from https://doi.org/10.1128%2Fmbio.02120-17 doi: 10.1128/mbio.02120-17

Donoho, D. (2017, Oct). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745–766. Retrieved from https://doi.org/10.1080%2F10618600.2017.1384734 doi: 10.1080/10618600.2017.1384734

Dreisbach, C., Koleck, T. A., Bourne, P. E., & Bakken, S. (2019, May). A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *International Journal of Medical Informatics*, *125*, 37–46. Retrieved from https://doi.org/10.1016%2Fj.ijmedinf.2019.02.008 doi: 10.1016/j.ijmedinf.2019.02.008

Dugan, T. M., Mukhopadhyay, S., Carroll, A., & Downs, S. (2015). Machine Learning Techniques for Prediction of Early Childhood Obesity. *Applied Clinical Informatics*, *06*(03), 506–520. doi: 10.4338/aci-2015-03-ra-0036

Eichler, H.-G., Bloechl-Daum, B., Broich, K., Kyrle, P. A., Oderkirk, J., Rasi, G., ... Paris, V. (2018, Oct). Data rich, information poor: Can we use electronic health records to create a learning healthcare system for pharmaceuticals? *Clinical Pharmacology & Therapeutics*, *105*(4), 912–922. Retrieved from https://doi.org/10.1002%2Fcpt.1226 doi: 10.1002/cpt.1226

Food and Drug Administration (FDA). (2019a). *Clinical Decision Support Software. Draft Guidance for Industry and Food and Drug Administration Staff* (Tech. Rep.). Silver Spring, MD: Food and Drug Administration. Retrieved from http://www.fda.gov/downloads/MedicalDevices/.../UCM263366.pdf

Food and Drug Administration (FDA). (2019b). *Rare Diseases: Natural History Studies for Drug Development* (Tech. Rep. No. March 2019). Silver Spring, MD: Food and Drug Administration. Retrieved from https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

Forrest, C. B., Margolis, P. A., Bailey, L. C., Marsolo, K., Beccaro, M. A. D., Finkelstein, J. A., ... Kahn, M. G. (2014, Jul). PEDSnet: a national pediatric learning health system. *Journal of the American Medical Informatics Association*, *21*(4), 602–606. Retrieved from https://doi.org/10.1136%2Famiajnl-2014-002743 doi: 10.1136/amiajnl-2014-002743

Friedman, C. P., Allee, N. J., Delaney, B. C., Flynn, A. J., Silverstein, J. C., Sullivan, K., & Young, K. A. (2016, Nov). The science of learning health systems: Foundations for a new journal. *Learning Health Systems*, *1*(1), e10020. Retrieved from https://doi.org/10.1002%2Flrh2.10020 doi: 10.1002/lrh2.10020

Friedman, J. H. (2001, Oct). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. Retrieved from https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full doi: 10.1214/aos/1013203451

Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. A. (2016, May). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *Journal of the American Medical Informatics Association*, *24*(1), 198–208. Retrieved from https://doi.org/10.1093%2Fjamia%2Focw042 doi: 10.1093/jamia/ocw042

Goulooze, S. C., Ista, E., van Dijk, M., Hankemeier, T., Tibboel, D., Knibbe, C. A., & Krekels, E. H. (2019, Oct). Supervised multidimensional item response theory modeling of pediatric iatrogenic withdrawal symptoms. *CPT: Pharmacometrics & Systems Pharmacology*, *8*(12), 904–912. Retrieved from https://doi.org/10.1002%2Fpsp4.12469 doi: 10.1002/psp4.12469

Gram, M., Erlenwein, J., Petzke, F., Falla, D., Przemeck, M., Emons, M., ... Drewes, A. (2016, Jul). Prediction of postoperative opioid analgesia using clinical-experimental parameters and electroencephalography. *European Journal of Pain*, *21*(2), 264–277. Retrieved from https://doi.org/10.1002%2Fejp.921 doi: 10.1002/ejp.921

Hartley, C., Duff, E. P., Green, G., Mellado, G. S., Worley, A., Rogers, R., & Slater, R. (2017, May). Nociceptive brain activity as a measure of analgesic efficacy in infants. *Science Translational Medicine*, *9*(388). Retrieved from https://doi.org/10.1126%2Fscitranslmed.aah6122 doi: 10.1126/scitranslmed.aah6122

Hoerl, A. E., & Kennard, R. W. (1970, Feb). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. Retrieved from https://doi.org/10.1080%2F00401706.1970.10488634 doi: 10.1080/00401706.1970.10488634

Hooke, M. C., Gilchrist, L., Tanner, L., Hart, N., & Withycombe, J. S. (2016, Jan). Use of a fitness tracker to promote physical activity in children with acute lymphoblastic leukemia. *Pediatric Blood & Cancer*, *63*(4), 684–689. Retrieved from https://doi.org/10.1002%2Fpbc.25860 doi: 10.1002/pbc.25860

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018, May). Artificial intelligence in radiology. *Nature Reviews Cancer*, *18*(8), 500–510. Retrieved from https://doi.org/10.1038%2Fs41568-018-0016-5 doi: 10.1038/s41568-018-0016-5

Huang, C., Clayton, E. A., Matyunina, L. V., McDonald, L. D., Benigno, B. B., Vannberg, F., & McDonald, J. F. (2018, Nov). Machine learning predicts individual cancer patient responses to therapeutic drugs

with high accuracy. *Scientific Reports*, *8*(1). Retrieved from https://doi.org/10.1038%2Fs41598-018-34753-5 doi: 10.1038/s41598-018-34753-5

Ince, I., de Wildt, S. N., Tibboel, D., Danhof, M., & Knibbe, C. A. (2009, Mar). Tailor-made drug treatment for children. *Drug Discovery Today*, *14*(5-6), 316–320. Retrieved from https://doi.org/10.1016%2Fj.drudis.2008.11.004 doi: 10.1016/j.drudis.2008.11.004

Izmailova, E. S., Wagner, J. A., & Perakslis, E. D. (2018, Apr). Wearable devices in clinical trials: Hype and hypothesis. *Clinical Pharmacology & Therapeutics*, *104*(1), 42–52. Retrieved from https://doi.org/10.1002%2Fcpt.966 doi: 10.1002/cpt.966

Joseph, P. D., Craig, J. C., & Caldwell, P. H. (2015, Feb). Clinical trials in children. *British Journal of Clinical Pharmacology*, *79*(3), 357–369. Retrieved from https://doi.org/10.1111%2Fbcp.12305 doi: 10.1111/bcp.12305

Kelly, L. E., Sinha, Y., Barker, C. I. S., Standing, J. F., & Offringa, M. (2018, Apr). Useful pharmacodynamic endpoints in children: selection, measurement, and next steps. *Pediatric Research*, *83*(6), 1095–1103. Retrieved from https://doi.org/10.1038%2Fpr.2018.38 doi: 10.1038/pr.2018.38

Khozin, S., & Coravos, A. (2019, Apr). Decentralized trials in the age of real-world evidence and inclusivity in clinical investigations. *Clinical Pharmacology & Therapeutics*, *106*(1), 25–27. Retrieved from https://doi.org/10.1002%2Fcpt.1441 doi: 10.1002/cpt.1441

Kim, J., Campbell, A. S., de Ávila, B. E.-F., & Wang, J. (2019, Feb). Wearable biosensors for healthcare monitoring. *Nature Biotechnology*, *37*(4), 389–406. Retrieved from https://doi.org/10.1038%2Fs41587-019-0045-y doi: 10.1038/s41587-019-0045-y

Kothare, P. A., Jadhav, P. R., Gupta, P., Harrelson, J. C., & Dickmann, L. (2018, Jun). Harnessing the potential of emerging digital health and biological sampling technologies for clinical drug development: Promise to reality. *Clinical Pharmacology & Therapeutics*, *104*(6), 1125–1135. Retrieved from https://doi.org/10.1002%2Fcpt.1100 doi: 10.1002/cpt.1100

Krekels, E. H., van Hasselt, J. C., van den Anker, J. N., Allegaert, K., Tibboel, D., & Knibbe, C. A. (2017, Nov). Evidence-based drug treatment for special patient populations through model-based approaches. *European Journal of Pharmaceutical Sciences*, *109*, S22–S26. Retrieved from https://doi.org/10.1016%2Fj.ejps.2017.05.022 doi: 10.1016/j.ejps.2017.05.022

Larson, D. B., Chen, M. C., Lungren, M. P., Halabi, S. S., Stence, N. V., & Langlotz, C. P. (2018, Apr). Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. *Radiology*, *287*(1), 313–322. Retrieved from https://doi.org/10.1148%2Fradiol.2017170236 doi: 10.1148/radiol.2017170236

Liang, H., Tsui, B. Y., Ni, H., Valentim, C. C. S., Baxter, S. L., Liu, G., ... Xia, H. (2019, Feb). Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nature Medicine*, *25*(3), 433–438. Retrieved from https://doi.org/10.1038%2Fs41591-018-0335-9 doi: 10.1038/s41591-018-0335-9

Liang, M., Tang, W., Xu, D. M., Jirapatnakul, A. C., Reeves, A. P., Henschke, C. I., & Yankelevitz, D. (2016, Oct). Low-dose CT screening for lung cancer: Computer-aided detection of missed lung cancers. *Radiology*, *281*(1), 279–288. Retrieved from https://doi.org/10.1148%2Fradiol.2016150063 doi: 10.1148/radiol.2016150063

Majeed-Ariss, R., Baildam, E., Campbell, M., Chieng, A., Fallon, D., Hall, A., ... Swallow, V. (2015, Dec). Apps and adolescents: A systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. *Journal of Medical Internet Research*, *17*(12), e287. Retrieved from https://doi.org/10.2196%2Fjmir.5043 doi: 10.2196/jmir.5043

Mani, S., Ozdas, A., Aliferis, C., Varol, H. A., Chen, Q., Carnevale, R., ... Weitkamp, J.-H. (2014, Mar). Medical decision support using machine learning for early detection of late-onset neonatal sepsis. *Journal of the American Medical Informatics Association*, *21*(2), 326–336. Retrieved from https://doi.org/10.1136%2Famiajnl-2013-001854 doi: 10.1136/amiajnl-2013-001854

Marshall, S. A., Yang, C. C., Ping, Q., Zhao, M., Avis, N. E., & Ip, E. H. (2015, Oct). Symptom clusters in women with breast cancer: an analysis of data from social media and a research study. *Quality of Life Research*, *25*(3), 547–557. Retrieved from https://doi.org/10.1007%2Fs11136-015-1156-7 doi: 10.1007/s11136-015-1156-7

McMahon, A. W., & Pan, G. D. (2018, Jun). Assessing drug safety in children — the role of real-world data. *New England Journal of Medicine*, *378*(23), 2155–2157. Retrieved from https://doi.org/10.1056%2Fnejmp1802197 doi: 10.1056/nejmp1802197

Mehrotra, N., Bhattaram, A., Earp, J. C., Florian, J., Krudys, K., Lee, J. E., ... Sinha, V. (2016, Apr). Role of quantitative clinical pharmacology in pediatric approval and labeling. *Drug Metabolism and Disposition*, *44*(7), 924–933. Retrieved from https://doi.org/10.1124%2Fdmd.116.069559 doi: 10.1124/dmd.116.069559

Miksad, R. A., Samant, M. K., Sarkar, S., & Abernethy, A. P. (2019, May). Small but mighty: The use of real-world evidence to inform precision medicine. *Clinical Pharmacology & Therapeutics*, *106*(1), 87–90. Retrieved from https://doi.org/10.1002%2Fcpt.1466 doi: 10.1002/cpt.1466

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2017, May). Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics*, *19*(6), 1236–1246. Retrieved from https://doi.org/10.1093%2Fbib%2Fbbx044 doi: 10.1093/bib/bbx044

Momper, J., Mulugeta, Y., & Burckart, G. (2015, Jun). Failed pediatric drug development trials. *Clinical Pharmacology & Therapeutics*, *98*(3), 245–251. Retrieved from https://doi.org/10.1002%2Fcpt.142 doi: 10.1002/cpt.142

Moor, G. D., Sundgren, M., Kalra, D., Schmidt, A., Dugas, M., Claerhout, B., ... Coorevits, P. (2015, Feb). Using electronic health records for clinical research: The case of the EHR4cr project. *Journal of Biomedical Informatics*, *53*, 162–173. Retrieved from https://doi.org/10.1016%2Fj.jbi.2014.10.006 doi: 10.1016/j.jbi.2014.10.006

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011, Sep). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, *18*(5), 544–551. Retrieved from https://doi.org/10.1136%2Famiajnl-2011-000464 doi: 10.1136/amiajnl-2011-000464

Patel, R., Belousov, M., Jani, M., Dasgupta, N., Winokur, C., Nenadic, G., & Dixon, W. G. (2018, Feb). Frequent discussion of insomnia and weight gain with glucocorticoid therapy: an analysis of twitter posts. *npj Digital Medicine*, *1*(1). Retrieved from https://doi.org/10.1038%2Fs41746-017-0007-z doi: 10.1038/s41746-017-0007-z

Pons, E., Braun, L. M. M., Hunink, M. G. M., & Kors, J. A. (2016, May). Natural language processing in radiology: A systematic review. *Radiology*, *279*(2), 329–343. Retrieved from https://doi.org/10.1148%2Fradiol.16142770 doi: 10.1148/radiol.16142770

Rostami-Hodjegan, A. (2012, May). Physiologically based pharmacokinetics joined with in vitro–in vivo extrapolation of ADME: A marriage under the arch of systems pharmacology. *Clinical Pharmacology & Therapeutics*, *92*(1), 50–61. Retrieved from https://doi.org/10.1038%2Fclpt.2012.65 doi: 10.1038/clpt.2012.65

Savova, G., Pestian, J., Connolly, B., Miller, T., Ni, Y., & Dexheimer, J. W. (2016). Natural language processing: Applications in pediatric research. In *Translational bioinformatics* (pp. 231–250). Springer Singapore. Retrieved from https://doi.org/10.1007%2F978-981-10-1104-7_12 doi: 10.1007/978-981-10-1104-7_12

Shaw, J., Rudzicz, F., Jamieson, T., & Goldfarb, A. (2019, Jul). Artificial intelligence and the implementation challenge. *Journal of Medical Internet Research*, *21*(7), e13659. Retrieved from https://doi.org/10.2196%2F13659 doi: 10.2196/13659

Sun, H., Bastings, E., Temeck, J., Smith, P. B., Men, A., Tandon, V., ... Rodriguez, W. (2013, Mar). Migraine therapeutics in adolescents. *JAMA Pediatrics*, *167*(3), 243. Retrieved from https://doi.org/10.1001%2Fjamapediatrics.2013.872 doi: 10.1001/jamapediatrics.2013.872

Suzuki, K., Li, F., Sone, S., & Doi, K. (2005, Sep). Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network. *IEEE Transactions on Medical Imaging*, *24*(9), 1138–1150. Retrieved from https://doi.org/10.1109%2Ftmi.2005.852048 doi: 10.1109/tmi.2005.852048

Swift, B., Jain, L., White, C., Chandrasekaran, V., Bhandari, A., Hughes, D. A., & Jadhav, P. R. (2018, May). Innovation at the intersection of clinical trials and real-world data science to advance patient care. *Clinical and Translational Science*, *11*(5), 450–460. Retrieved from https://doi.org/10.1111%2Fcts.12559 doi: 10.1111/cts.12559

Taylor, M. J., Dove, E. S., Laurie, G., & Townend, D. (2017, Nov). When can the child speak for herself? the limits of parental consent in data protection law for health research. *Medical Law Review*, *26*(3), 369–391. Retrieved from https://doi.org/10.1093%2Fmedlaw%2Ffwx052 doi: 10.1093/medlaw/fwx052

Tibshirani, R. (1996, Jan). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. Retrieved from https://doi.org/10.1111%2Fj.2517-6161.1996.tb02080.x doi: 10.1111/j.2517-6161.1996.tb02080.x

van Stekelenborg, J., Ellenius, J., Maskell, S., Bergvall, T., Caster, O., Dasgupta, N., ... Pirmohamed, M. (2019, Aug). Recommendations for the use of social media in pharmacovigilance: Lessons from IMI WEB-RADR. *Drug Safety*, *42*(12), 1393–1407. Retrieved from https://doi.org/10.1007%2Fs40264-019-00858-7 doi: 10.1007/s40264-019-00858-7

Vayena, E., Blasimme, A., & Cohen, I. G. (2018, Nov). Machine learning in medicine: Addressing ethical challenges. *PLOS Medicine*, *15*(11), e1002689. Retrieved from https://doi.org/10.1371%2Fjournal.pmed.1002689 doi: 10.1371/journal.pmed.1002689

Weimer, K., Gulewitsch, M. D., Schlarb, A. A., Schwille-Kiuntke, J., Klosterhalfen, S., & Enck, P. (2013, Apr). Placebo effects in children: a review. *Pediatric Research*, *74*(1), 96–102. Retrieved from https://

doi.org/10.1038%2Fpr.2013.66 doi: 10.1038/pr.2013.66

Wittmeier, K. D. M., Klassen, T. P., & Sibley, K. M. (2015, Apr). Implementation science in pediatric health care. *JAMA Pediatrics*, *169*(4), 307. Retrieved from https://doi.org/10.1001%2Fjamapediatrics.2015.8 doi: 10.1001/jamapediatrics.2015.8

Yackey, K., Stukus, K., Cohen, D., Kline, D., Zhao, S., & Stanley, R. (2019, Mar). Off-label medication prescribing patterns in pediatrics: An update. *Hospital Pediatrics*, *9*(3), 186–193. Retrieved from https://doi.org/10.1542%2Fhpeds.2018-0168 doi: 10.1542/hpeds.2018-0168

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018, Jun). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, *9*(4), 611–629. Retrieved from https://doi.org/10.1007%2Fs13244-018-0639-9 doi: 10.1007/s13244-018-0639-9

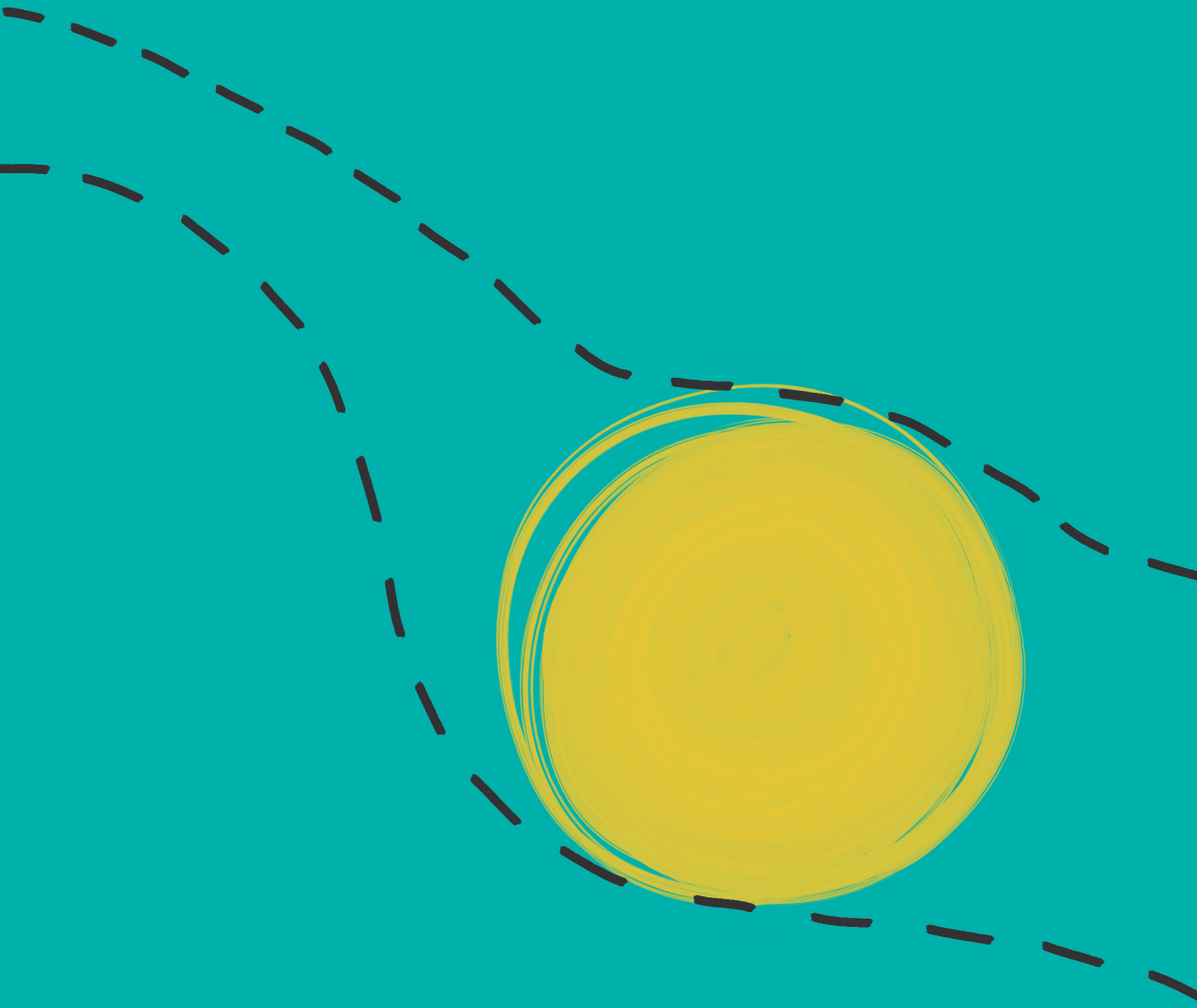
Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., & Shen, D. (2015, Mar). Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *NeuroImage*, *108*, 214–224. Retrieved from https://doi.org/10.1016%2Fj.neuroimage.2014.12.061 doi: 10.1016/j.neuroimage.2014.12.061

Zilcha-Mano, S., Roose, S. P., Brown, P. J., & Rutherford, B. R. (2018, Jun). A machine learning approach to identifying placebo responders in late-life depression trials. *The American Journal of Geriatric Psychiatry*, *26*(6), 669–677. Retrieved from https://doi.org/10.1016%2Fj.jagp.2018.01.001 doi: 10.1016/j.jagp.2018.01.001

Zorc, J. J., Chamberlain, J. M., & Bajaj, L. (2019, Jul). Machine learning at the clinical bedside—the ghost in the machine. *JAMA Pediatrics*, *173*(7), 622. Retrieved from https://doi.org/10.1001%2Fjamapediatrics.2019.1075 doi: 10.1001/jamapediatrics.2019.1075

Acknowledgement

The authors thank Margot van Ark for assisting in the preparation of the figures.



Chapter 3

Identification of high-dimensional omics-derived predictors for tumor growth dynamics using machine learning and pharmacometric modeling

Authors

Laura B. Zwep

Kevin L. W. Duisters

Martijn Jansen

Tingjie Guo

Jacqueline J. Meulman

Parth J. Upadhyay

J. G. Coen van Hasselt

CPT: Pharmacometrics & Systems Pharmacology 2021; 10(4), 350–361

Abstract

Pharmacometric modeling can capture tumor growth inhibition (TGI) dynamics and variability. These approaches do not usually consider covariates in high-dimensional settings, whereas high-dimensional molecular profiling technologies ("omics") are being increasingly considered for prediction of anticancer drug treatment response. Machine learning (ML) approaches have been applied to identify high-dimensional omics predictors for treatment outcome. Here, we aimed to combine TGI modeling and ML approaches for two distinct aims: omics-based prediction of tumor growth profiles and identification of pathways associated with treatment response and resistance. We propose a two-step approach combining ML using least absolute shrinkage and selection operator (lasso) regression with pharmacometric modeling. We demonstrate our workflow using a previously published dataset consisting of 4706 tumor growth profiles of patient-derived xenograft (PDX) models treated with a variety of mono- and combination regimens. Pharmacometric TGI models were fit to the tumor growth profiles. The obtained empirical Bayes estimates-derived TGI parameter values were regressed using the lasso on high-dimensional genomic copy number variation data, which contained over 20,000 variables. The predictive model was able to decrease median prediction error by 4% as compared with a model without any genomic information. A total of 74 pathways were identified as related to treatment response or resistance development by lasso, of which part was verified by literature. In conclusion, we demonstrate how the combined use of ML and pharmacometric modeling can be used to gain pharmacological understanding in genomic factors driving variation in treatment response.

3.1 Introduction

Pharmacometric modeling of tumor growth inhibition (TGI) dynamics is extensively used to model the longitudinal response of tumor size in response to drug treatment in preclinical animal models or patients. Pharmacometric TGI models have increasingly been used to characterize drug-exposure response relationships using semi-mechanistic parameters related to, for instance, direct treatment effects or resistance to personalize drug treatment (Ribba et al., 2014; Bender et al., 2014). Using TGI models, interindividual variation in tumor growth rate, treatment efficacy, and treatment resistance can be quantified and related to patient-specific characteristics (Ribba et al., 2014; Rodriguez-Brenes et al., 2013). In recent years, TGI models have been integrated with time-to-event models to predict clinical outcomes, such as overall survival, which allow prediction of clinical outcomes based on the patient-specific tumor growth dynamics parameters (Claret et al., 2009; van Hasselt et al., 2015b, 2015a).

The use of high-dimensional molecular profiling technologies, including next-generation sequencing, to develop personalized treatment schedules is rapidly developing. In particular in oncology, the use of "omics" technologies to characterize tumor-

specific molecular differences to predict variation in treatment response is of great interest (Shlien & Malkin, 2009). Although both omics and TGI modeling are of relevance toward personalized treatment strategies, pharmacometric TGI models are not frequently directly applied to high-dimensional covariates. In pharmacometric modeling, stepwise covariate inclusion approaches are still the most commonly used approach to include covariates, which is unsuitable for testing of covariates in a high-dimensional setting.

Current analyses of high-dimensional "omics" datasets predicting treatment response are mostly performed using machine learning (ML) methodologies, such as sparse regression models, random forests, and deep learning, to obtain predictive signatures of treatment response (Degenhardt et al., 2017; Nicolò et al., 2020; Xie et al., 2019). The majority of studies with ML approaches are based on either dichotomous survival response or clinical response metrics, such as based on the Response Evaluation Criteria in Solid Tumors (RECIST) system (Lathrop & Kaklamani, 2018; Eisenhauer et al., 2009), wherein the observed dynamic tumor disease progression profile is reduced into a limited number of categories. These simplified categorical treatment response metrics lack biological or pharmacological relevance, because factors, such as resistance and direct treatment effects, are merged (Chadeau-Hyam et al., 2013).

A commonly used ML method is the least absolute shrinkage and selection operator (lasso), which is a linear regression method with ℓ_1 regularization that can be used for high-dimensional analysis, and results in variable selection (Tibshirani, 1996). Although ML approaches, such as sparse regression models using the lasso (Bertrand et al., 2008, 2015; Ribbing et al., 2007; Haem et al., 2017), have been implemented in pharmacometric modeling, they are computationally expensive due to the combination of nonlinearity and estimation of random effects, which often lead to convergence problems. The implementations of the lasso involve alternating algorithms, which alternate between estimating the random effects and the lasso optimization, so although lasso is rather efficient, iterating through multiple random effect estimation steps can severely reduce computational efficiency. This can lead to long computation times and poor convergence rates, especially in high-dimensional settings.

In this study, we propose a two-step approach combining ML, using lasso regression, with pharmacometric modeling. We demonstrate our approach using a large dataset consisting of longitudinal tumor growth profiles of patient-derived xenograft (PDX) models treated with a variety of mono- and combination regimens (Gao et al., 2015). We develop pharmacometric tumor growth models quantifying intertumor variation in growth rates, drug effect, and resistance, after which we implement ML-based lasso models to address the following aims: (1) to predict longitudinal tumor growth profiles based on omics-derived predictors using a multivariate lasso model; and (2) to identify biological pathways associated with interindividual variation in treatment response or resistance development using a group lasso regression model (Figure 3.1).

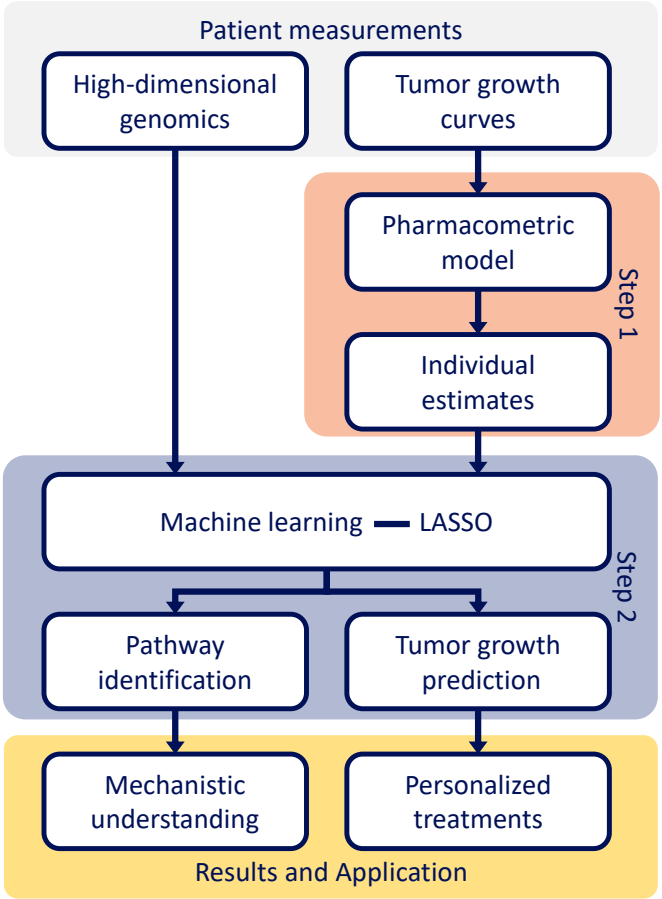


Figure 3.1: A schematic visualization of the proposed two-step approach. First, the tumor growth curves were modeled to obtain tumor growth parameter estimates, second, the individual estimates tumor growth parameter estimates were regressed on copy number variations (genomics) by different least absolute shrinkage and selection operator (lasso) techniques. The group lasso was applied to obtain biological pathways. The multivariate lasso was applied to predict the tumor growth parameter values, which were then inserted into the tumor growth inhibition model equations to obtain predictions of the tumor growth curves.

3.2 Methods

3.2.1 Data

Data from a large scale preclinical study in PDX mice models were used (Gao et al., 2015). This dataset consisted of over 4000 PDX experiments, which were derived from a total of 277 patients, where multiple PDX experiments were derived from the same tumor. The PDX experiments from one tumor were all treated with different anticancer agents as mono treatment or combination treatment, or left untreated (e.g., natural growth experiments). There was a total of 62 unique treatments and for every

tumor treated one PDX was left untreated, leading to an incomplete design with multiple PDX experiments per treatment. Tumor volume was measured daily. For each unique tumor, at the start of the treatment, genomic data based on gene copy number variations (CNVs) were obtained, yielding a total of 23,852 CNVs. We included data for 174 unique tumors and 55 unique treatments, corresponding with 3244 tumor-treatment combinations. This selection was based on availability of CNV data and adequate fit (see section below). The analysis was conducted separately for every treatment, so the number of observations differed per analysis, ranging from 17 to 171 observations (Table S3.1).

3.2.2 Tumor growth inhibition model

A TGI model was fitted to the longitudinal tumor volume measurements using the non-linear regression modeling software NONMEM (Beal & Sheiner, 1980), with first order conditional estimation with interaction(Claret et al., 2009). The TGI model captured the longitudinal tumor volume measurements, per PDX, through estimation of three parameters: growth rate (k_g), treatment efficacy (k_d), and time-dependent resistance development (k_r) in an ordinary differential equation (Equation 3.1).

$$\frac{dV_i(t)}{dt} = k_{g,i} \cdot V_i(t) - k_{d,i} \cdot e^{-k_{r,i} \cdot t} \cdot V_i(t) \tag{3.1}$$

with tumor volume $V(t)$ at time t and tumor growth model parameters k_g, k_d and k_r. Random effects with a log-normal distribution, were added to all fixed effect TGI parameters as following: $k_{g,i} = k_g \cdot exp\ (\eta_{k_g i})$.

To fit the TGI model, we first estimated individual value for k_g separately for every tumor using the untreated PDX data (Equation 3.2).

$$\frac{dV_i(t)}{dt} = k_{g,i} \times V_i(t) \tag{3.2}$$

The empirical Bayes estimates (EBEs) of k_g were extracted and included as data in the TGI model. EBEs in NONMEM is the estimation of the posterior individual random effects ($\hat{\eta}_i$), based on the empirically obtained prior distribution of η and the individual data, as previously described (Sheiner et al., 1972). The residual error was modeled with both an additive and proportional error.

We observed that not all tumor growth curves showed time-dependent resistance development (e.g., regrowth), so both a full TGI model and a reduced model, without a term for resistance, were fitted, effectively allowing k_r to become zero. For every PDX, a likelihood ratio test was conducted to evaluate whether inclusion of k_r added significantly to the model fit (at significance level 0.05). A second criterium was added to only select the full model if the k_r was estimated to be smaller than 1.0, because the term $k_d \cdot e^{k_r \cdot t}$ goes to zero very fast with t for larger k_r, effectively making k_d unidentifiable.

To evaluate the model fit separately for each treatment, we plotted the conditional weighted residuals per treatment, which represent the goodness of fit for the TGI

models (Nguyen et al., 2017). Due to the large number of tumor growth curves, treatments with curves with a bad model fit were removed from further analysis. The EBEs of k_d and k_r were extracted for the treatments with good model fit.

3.2.3 Tumor growth profile prediction

The multivariate outcome the log-transformed k_g, log-transformed k_d, and k_r was regressed on the genomic CNV data within every treatment using a multivariate lasso (Equation 3.3) (Simon et al., 2013). The multivariate lasso, similarly to the standard lasso, minimized the loss function to estimate the linear parameters β is mainly due to outcome Y and parameter β, which are, in this case, both matrices containing a column for every outcome. The penalty term is the root of the summed square error over the vector β_j.

$$\hat{\beta}_{MVlasso} = \arg\min_{\beta}(||Y - X\beta||_2^2 + \lambda \sum_{j=1}^{p} ||\beta_j||_2) \tag{3.3}$$

The lasso hyperparameter λ, which determines the size and number of non-zero parameters, was chosen through 10-fold cross-validation, to identify the λ which minimized the prediction error in terms of mean squared error. This minimizing λ differed per treatment. The treatments where the minimizing λ did not outperform the null model, which estimated no non-zero coefficients for the CNVs, were removed from further analysis, both in prediction of the tumor growth curves and the pathway selection. In a second analysis, only the log-transformed k_g, log-transformed k_d were regressed on the CNV data. Prediction errors were evaluated both on the scale of the predicted parameter values and on the scale of the predicted tumor growth curves.

The individual TGI parameter values predicted from the lasso were extracted. The ordinary differential equation (Equation 3.1 was solved for these predicted parameter values to bring the predictions back on the longitudinal tumor volume scale. For robustness, the cross-validation step was repeated twenty times over different cross-validation splits and the predicted curves were averaged over the twenty repetitions.

A measure of prediction error was defined on tumor curve scale through comparing the curves from the estimated parameters from the TGI model to the curves with the predicted parameters from lasso. The prediction error was defined as the absolute fraction of the area between the predicted and the estimated curves (ABC) over the area under the estimated curve (AUC), called the scaled ABC (sABC, Equation 3.4).

$$sABC_i(\tau) = \frac{\int_0^{\tau} |\tilde{V}_i(t) - \hat{V}_i(t)| dt}{\int_0^{\tau} \tilde{V}_i(t) dt} \tag{3.4}$$

for individual i with volume $\tilde{V}_i(t)$ estimated from the TGI model fit (IPRED) and volume $\hat{V}_i(t)$ predicted from the multivariate lasso. The area is considered until some cut-off τ, which in our study was set to 56 days (two months). The sABC was used because

it is a one-dimensional and interpretable error measure. The sABC metric allowed for the comparison of the two functions produced by the TGI model fit and the lasso parameter value prediction. The sABC of the lasso with CNVs was compared with the sABC of the null model, to see whether the CNVs added predictive power.

3.2.4 Pathway selection

To gain biological insight gained beyond selection of individual genes contributing to the predictive performance of treatment efficacy and time-dependent resistance development, the log-transformed k_d and k_r were separately regressed on the CNVs through pathway analysis using overlapping group lasso (Yuan & Lin, 2006; Jacob et al., 2009). The overlapping grouped lasso uses a combination of the lasso and the ℓ_2 norm, a square root of the sum of squares of the coefficients, which is also used for RIDGE regression (Hoerl & Kennard, 1970), to select variables on a group level (Equation 3.5). Each of the G groups contain a set of indices $\Im_g$, including all parameter indices of the β's in group g. The size of the group is denoted as $|\Im_g|$, which is used to scale the penalty to account for the different group sizes.

$$\hat{\beta}_{grouplasso} = \arg\min_{\beta}(||Y - X\beta||_2^2 + \lambda \sum_{g=1}^{G} \sqrt{|\Im_g|}||\beta_{\Im_g}||_2) \tag{3.5}$$

The groups were defined as the pathways from the WikiPathways ontology, which contains a comprehensive overview of biological pathways and processes (Kuleshov et al., 2016; Chen et al., 2013; Kutmon et al., 2015). A total of 5,998 CNVs was grouped to one or more pathways.

Again, ten-fold cross-validation was used to identify the λ, which determined how many pathways were selected. While utilizing a combination of ℓ_1 and ℓ_2 penalties, there is only one hyperparameter in the group lasso (Yuan & Lin, 2006). Subsequently, part of the discovered correlations between pathways and treatment response was researched in literature for validation. This analysis was conducted in R (R Core Team, 2020) (version 3.6.3) using the library grpregOverlap (`https://github.com/YaohuiZeng/grpregOverlap`).

3.2.5 Code availability

All scripts and models use for the analysis are available on github (`https://github.com/vanhasseltlab/PDX`).

3.3 Results

3.3.1 Tumor growth inhibition model development

The TGI model was fitted to the PDX tumor growth curves, separately for every treatment. For three treatments, no model was converged, these were left out of the anal-

ysis. The model fit was evaluated through the conditional weighted residuals (Figure S3.1) and the visual inspection of the PDX fits (Figure S3.2). The visual inspection showed the tumor dynamics for treatment TAS266 were not captured. Combination therapy LFW527 and binimetinib showed skewed residual distributions. The two treatments were discarded for further analysis. The model fit for the other treatments was sufficient.

All individual parameter estimates (EBEs) were extracted from the TGI model (Figure 3.2a). Figure 3.2b shows how the values of the parameter estimates affect the curve. The percentage of PDX experiments with non-zero time-dependent resistance development was 12.6%. The TGI model for the chosen treatments showed sufficient fits for the next step parameter values prediction step.

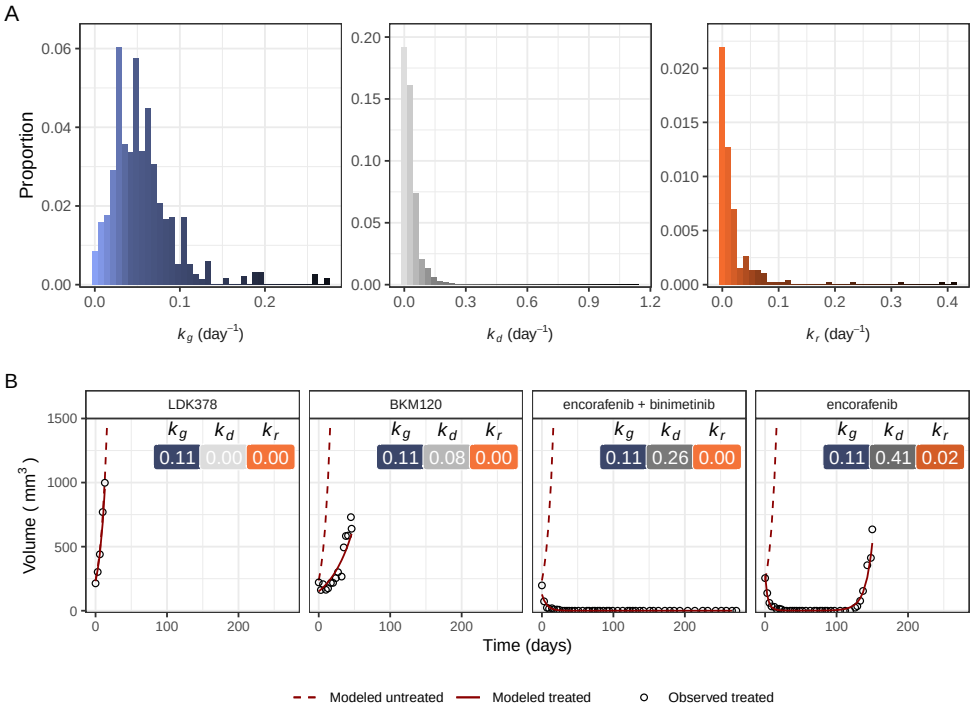


Figure 3.2: Results of the tumor growth inhibition (TGI) model estimation. A) The distributions of the individual, estimated TGI parameters. B) Selected tumor growth profiles showing how k_d and k_r vary for different treatments, with from left to right a very ineffective treatment, a slightly effective treatment, a very effective treatment and a very effective treatment with time-dependent resistance development.

The effect of shrinkage of the individual prediction values (η), often referred to as eta-shrinkage, was evaluated in Figure S3.2. The fit of the individual growth curves was not influenced by shrinkage. Since the tumor volumes were densely sampled over time, with an averages of 0.3 samples per day, 50 days follow-up time and 14 measurements per experiment, we did not expect problems with shrinkage.

3.3.2 Prediction of tumor growth profiles using genomic predictors

The estimated individual TGI parameters k_r, log-transformed k_d and log-transformed k_g were simultaneously predicted by the multivariate lasso. The prediction errors for the k_g, k_d and k_r were calculated as root mean square error. Although the variation between the treatments was high (Figure S3a), overall the RMSE was high. For k_g, the RMSE was 0.035, while a mean estimated k_g of 0.0564. For both the k_d and the k_r the RMSEs, 0.044 and 0.049 respectively, were actually higher than the mean estimated k_d (0.033) and the k_r (0.0564), indicating a bad prediction of the tumor growth dynamics from CNVs. A multivariate lasso with only the log-transformed k_d and log-transformed k_g was also fitted. These two lasso models were compared based on the prediction error of the log-transformed k_d and log-transformed k_g and the sABC error measure (Figure S3), where the model without predicting the k_r seemed to fit better, especially in the case of combination therapies BYL719 and cetuximab, and BKM120 and LJC049, which was used for consecutive analysis.

Out of 52 treatments, 33 treatments were detected with a better prediction than the null model, based on the average MSE over the cross-validation replications. For the other treatments, the predictive ability was not improved by adding CNVs as predictors to the lasso regression. The log k_g and log k_d were transformed back to their original scale and the parameters were used to solve the ordinary differential equation (Eq. 1) from the model.

The predictive performance of the lasso for predicting the TGI parameter values was evaluated by comparing the curves from the predicted estimates to the curves from the TGI model fit, since the estimated curves were already shown to fit the data well. The predictions and estimations are functions instead of measures, so the scaled area between the curves was calculated as error. The overall median sABC is 0.456, which can be interpreted as the area between the predicted and estimated curve, is less than half the area below the estimated curve (Figure 3.3a). The sABC distributions for the different treatments were shown (Figure 3.3b). A lower sABC shows a lower prediction error. 23.6% of the curves has an sABC below 0.2, so the difference between the curves is less than 20% of the AUC of the estimated curve. The treatment LFA102 has a median sABC of only 0.153, indicating a good prediction. The worst predictions are in the treatment LGH447 with a median sABC of 0.867. Compared to the null model, the lasso reduced the sABC by a median decrease of 3.8%. This shows low predictive ability of the CNVs to predict tumor growth curves.

3.3.3 Identification of pathways associated with treatment efficacy and resistance

The TGI parameter values of k_d and k_r were regressed on CNVs grouped in pathways using the overlapping group lasso. The group lasso selected the pathways with predictive power for the 33 treatments where predictiveness was shown in the curve prediction step. Out of the 472 pathways from WikiPathways (Kutmon et al., 2015), 71

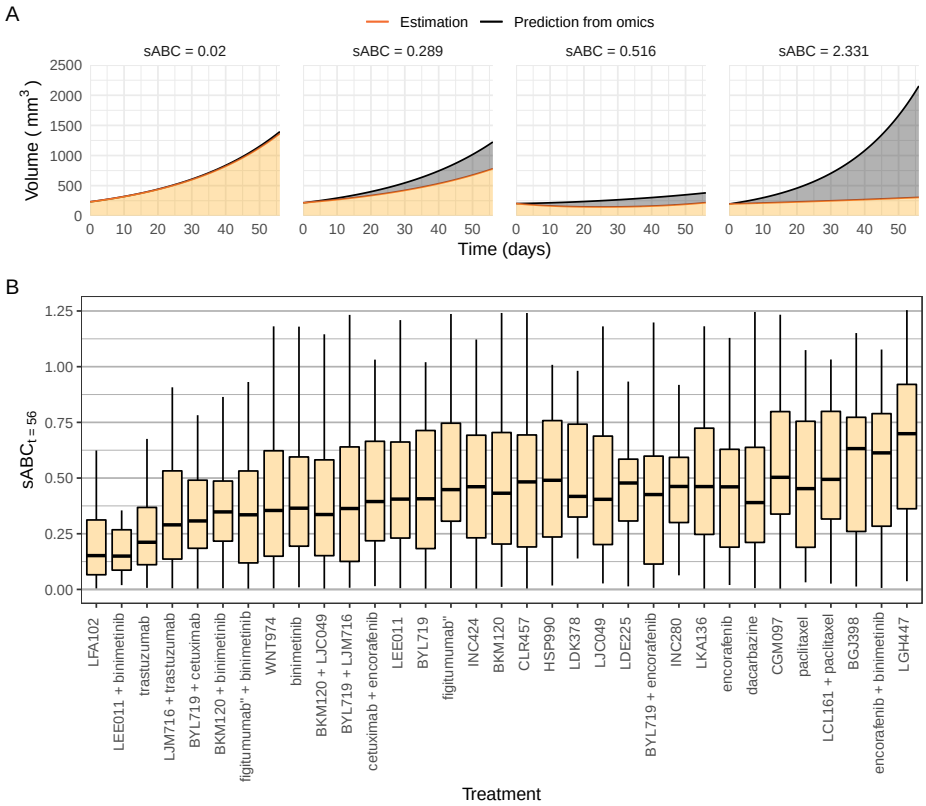


Figure 3.3: Predicted curves from the multivariate least absolute shrinkage and selection operator. A) Tumor growth curves visualized with the area under the estimated curve (orange) and between the estimated and predicted curves (grey) and the error (in scaled area between the predicted and the estimated curves [sABC]). From left to right show a very good prediction to a very poor prediction. B) The distributions of the individual patient-derived xenograft sABCs for the different treatments given by the interquartile range. Outliers are not included in the plot.

different pathways were selected for one or more of 19 different treatments, with a total of 118 detected pathway-treatment response correlations (Figure 3.4, Figure 3.5). The pathways were specifically correlated to either treatment efficacy or resistance development. More pathways were identified for k_d than k_r, due to smaller variation in k_r.

For paclitaxel, trastuzumab, encorafenib and figitumumab, the FDA approved drugs administered as monotherapy, we compared identified pathways with literature reports to evaluate their biological validity. We identified for 14 pathways for these four drugs, of which nine could be confirmed in literature (Table 3.1), where we confirm previously described mechanisms were detected through our method.

Figure 3.4: Selected pathways obtained by the group least absolute shrinkage and selection operator for the treatment efficacy (k_d = black), time- dependent resistance development (k_r = orange) or both (blue) over the different treatments. The distribution of pathways found for different treatments (top).

3.4 Discussion

In order to utilize high-dimensional omics data to further advance treatment response prediction and understanding, we developed a two-step approach combining a machine learning method with pharmacometric modeling.

We showed how CNVs can contribute to prediction of variability in tumor growth dynamics. This approach establishes a practical framework to enable personalized treatment selection or even dose optimization. Even though we have applied our approach to preclinical PDX data, TGI models have been widely used for modeling of clinical tumor size measurements to which our approach can be applied. Pharmacometric models including TGI models are typically based on ODE models, which is why we have chosen to formulate our model as ODEs and not using an analytical expression. Importantly, the use of a TGI model enables further integration with either clinical outcome prediction models (Claret et al., 2013) or it can be integrated with PK-PD models for TGI to refine dosing regimens to optimally suppress tumor growth. We expect this approach can also be implemented for the analysis of clinical tumor growth data.

In this study, we have set a cut-off of 56 days to evaluate the ability to back-predict tumor growth profiles; however, the predictions can also be extrapolated over a longer time-span, depending on the nature of available omics-data or specific disease or treatment characteristics. In terms of this sABC, the CNVs did however not show great improvement of predictive ability as compared to a null model. This was

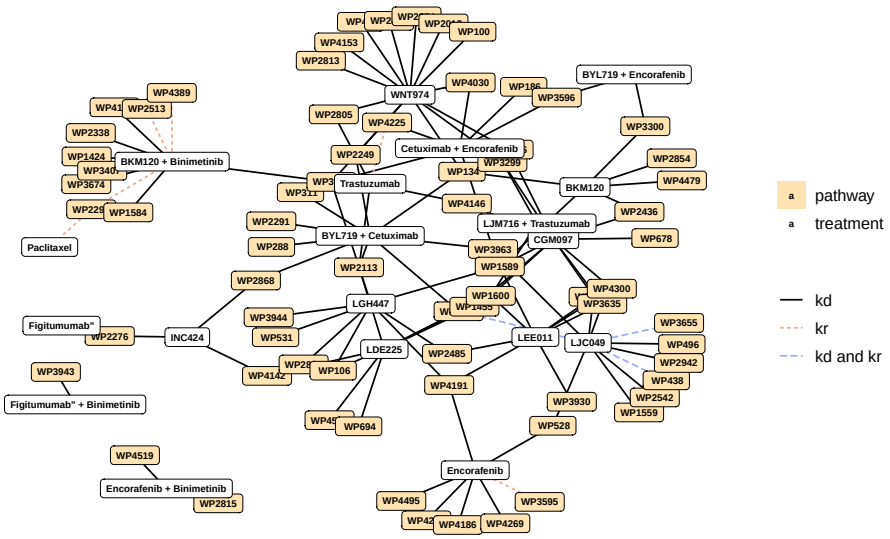


Figure 3.5: Overview of overlapping pathways between the different treatments. The nodes are the treatments (white background) and pathways (orange background) and the edges indicate which tumor growth inhibition parameter links the pathway to the treatment.

already visible in the large prediction errors on the tumor growth parameter values. The predictive ability was evaluated instead of the model fit in order to study more generalizable results. A large proportion of variance can often be explained by omics data, but the high-dimensional nature of the data makes it hard or impossible to distinguish between noise and structural differences.

To identify biological factors predictive for either treatment efficacy or resistance development we used a group lasso, grouping individual gene-associated CNVs to known biological pathways. We have used the Wikipathways ontology for grouping pathways, although other pathway databases can be used in a similar fashion. The pathway group lasso yields a set of pathways predictive of the outcomes treatment efficacy and time-dependent resistance development. Out of 14 identified pathways predictive of treatment efficacy and resistance development, nine pathways were confirmed by literature search. This is an indication of how omics pathway analysis for dynamic tumor growth responses could be a useful tool for validating pathway associations with factors responsible for treatment response, as well as discovering new correlations with pathways. Such a pathway-oriented approach has been previously proposed, but not in context with TGI or pharmacometric modeling (Silver et al., 2012).

In this study, we have used two versions of lasso regression for two distinct aims: variable selection and prediction. We selected the use of the lasso over other ML approaches due to its intrinsic property of variable selection.14 The selection for variables in high-dimensional data is not well accommodated in many algorithms, while the lasso inherently shrinks noise variables to zero. The lasso can achieve high sensitivity, but it can suffer from low specificity, this, however, is not considered as much of a problem in exploratory analyses.

The use of the group lasso allows for direct pathway selection based on omics data, which is computationally efficient and interpretable (Silver et al., 2012). The variable selection performance of the lasso has been investigated previously, and has been shown to perform competitively or comparatively better than other methods (Lenters et al., 2017; Hastie et al., 2017; Zheng & Liu, 2011).

The multivariate lasso was used to simultaneously predict the three model parameter values. A limitation of the multivariate lasso used in this study is that it does not take into account the dependence between the outcome variables, while the tumor growth model parameters are expected to be correlated. A second limitation was shown by the comparison between the predictions with and without adding k_r to the multivariate outcome. Prediction of one parameter can restrict the prediction of another parameter. We expect this problem can be overcome by better modeling of the joint and marginal distributions of the multivariate outcome.

The lasso has been previously implemented in pharmacometric nonlinear mixed effect models.16–18 These direct implementations have the advantage of informing the lasso directly within the longitudinal modeling. Models with a very high number of variables, however, become computationally hard. To our knowledge, these lasso implementations have not been successfully applied to very high-dimensional data, where the number of variables (p) was an order of magnitude larger than the number of observations (n), either due to convergence problems, or exploding computation

times. The two-step method is more dependent on the fit of the first model and the accuracy of the EBEs. Our method is more feasible in high-dimension, since the steps of the complex longitudinal model estimation and the high-dimensional predictors are separated.

The two-step approach can directly use other ML algorithms besides the lasso. Algorithms such as Random Forests and Gradient Boosting are able to capture non-linearity more easily, and can be used to improve model prediction accuracy. There is still a challenge in modeling multiple outcomes at the same time, such as the k_g, k_d and k_r in our study, but multivariate outcome modeling extensions have been made in in other high-dimensional methods, such as random forests (Segal & Xiao, 2011), which can be also used to predict tumor growth parameter values, as in the second step of our approach.

In summary, we demonstrated how combining machine learning and pharmacometric modeling can be used to gain pharmacological understanding of factors driving variation in treatment response, and to enable omics-based personalized treatment regimens.

Table 3.1: Pathway-treatment correlations found in literature. Scientific literature indicating previous findings on the pathways correlated to treatment efficacy and resistance, for the treatments paclitaxel, trastuzumab and encorafenib.

Treatment	Pathway	Response type	Literature	Relation
Paclitaxel	RalA downstream regulated genes (WP2290)	k_r	Ganapathy et al. (2016)	Paclitaxel is a mitotic inhibitor by stabilizing the microtubule and RalA has been previously shown to disrupt microtubule formation and inducing mitotic catastrophe.
Trastuzumab	Synthesis and degradation of ketone bodies (WP311)	k_d	Jobard et al. (2017)	Ketone production was shown to be increased with effective trastuzumab treatment.
	Macrophage markers (WP4146)	k_d	Shi et al. (2015)	Trastuzumab interacts with Fc$upgamma$ receptors on macrophages for the killing of HER2 cancer cells
	Pyrimidine metabolism and related diseases (WP4225)	k_d	Ghosh et al. (2009); Liu et al. (2019)	The pyrimidine metabolism pathway has been found in previous studies to correlate with drug response to Trastuzumab, based on pathway enrichment analysis in transcriptomics and metabolomics studies
	Caloric restriction and aging (WP4191)	k_d	Chappell et al. (2011)	There is a connection between Raf/MEK inhibitors and aging
	Somatroph axis (GH) and its relationship to dietary restriction and aging (WP4186)	k_d	Chappell et al. (2011)	There is a connection between Raf/MEK inhibitors and aging
Encorafenib	mir-124 predicted interactions with cell cycle and differentiation (WP3595)	k_r	Ross et al. (2018)	Resistance to Encorafenib has been shown to be correlated to cell cycle and differentiation
	Ethanol metabolism resulting in production of ROS by CYP2E1 (WP4269)	k_d	Friedlander and Cajulis (2019)	Ethanol metabolism resulting in production of ROS by CYP2E1 was found to have a connection to the development of melanoma, thus might be related to drug efficacy of encorafenib in melanoma treatment.
	IL-10 Anti-inflammatory Signaling Pathway (WP4495)	k_d	Sloane et al. (2017); Sumimoto et al. (2006)	IL-10 has been researched in the context of overexpression of Raf in cancer patients, showing that IL-10 is an immunosuppressive factor that is decreased by MEK inhibitors

References

Beal, S., & Sheiner, L. (1980, May). The NONMEM system. *The American Statistician*, *34*(2), 118. Retrieved from https://doi.org/10.2307%2F2684123 doi: 10.2307/2684123

Bender, B. C., Schindler, E., & Friberg, L. E. (2014, Dec). Population pharmacokinetic–pharmacodynamic modelling in oncology: a tool for predicting clinical response. *British Journal of Clinical Pharmacology*, *79*(1), 56–71. Retrieved from https://doi.org/10.1111%2Fbcp.12258 doi: 10.1111/bcp.12258

Bertrand, J., Comets, E., & Mentré, F. (2008, Nov). Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters. *Journal of Biopharmaceutical Statistics*, *18*(6), 1084–1102. Retrieved from https://doi.org/10.1080%2F10543400802369012 doi: 10.1080/10543400802369012

Bertrand, J., Iorio, M. D., & Balding, D. J. (2015, May). Integrating dynamic mixed-effect modelling and penalized regression to explore genetic association with pharmacokinetics. *Pharmacogenetics and Genomics*, *25*(5), 231–238. Retrieved from https://doi.org/10.1097%2Ffpc.0000000000000127 doi: 10.1097/fpc.0000000000000127

Chadeau-Hyam, M., Campanella, G., Jombart, T., Bottolo, L., Portengen, L., Vineis, P., ... Vermeulen, R. C. (2013, Aug). Deciphering the complex: Methodological overview of statistical models to derive OMICS-based biomarkers. *Environmental and Molecular Mutagenesis*, *54*(7), 542–557. Retrieved from https://doi.org/10.1002%2Fem.21797 doi: 10.1002/em.21797

Chappell, W. H., Steelman, L. S., Long, J. M., Kempf, R. C., Abrams, S. L., Franklin, R. A., ... McCubrey, J. A. (2011, Mar). Ras/raf/MEK/ERK and PI3k/PTEN/akt/mTOR inhibitors: Rationale and importance to inhibiting these pathways in human health. *Oncotarget*, *2*(3), 135–164. Retrieved from https://doi.org/10.18632%2Foncotarget.240 doi: 10.18632/oncotarget.240

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., ... Ma'ayan, A. (2013, Apr). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*(1). Retrieved from https://doi.org/10.1186%2F1471-2105-14-128 doi: 10.1186/1471-2105-14-128

Claret, L., Girard, P., Hoff, P. M., Cutsem, E. V., Zuideveld, K. P., Jorga, K., ... Bruno, R. (2009, Sep). Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *Journal of Clinical Oncology*, *27*(25), 4103–4108. Retrieved from https://doi.org/10.1200%2Fjco.2008.21.0807 doi: 10.1200/jco.2008.21.0807

Claret, L., Gupta, M., Han, K., Joshi, A., Sarapa, N., He, J., ... Bruno, R. (2013, Jun). Evaluation of tumor-size response metrics to predict overall survival in western and chinese patients with first-line metastatic colorectal cancer. *Journal of Clinical Oncology*, *31*(17), 2110–2114. Retrieved from https://doi.org/10.1200%2Fjco.2012.45.0973 doi: 10.1200/jco.2012.45.0973

Degenhardt, F., Seifert, S., & Szymczak, S. (2017, Oct). Evaluation of variable selection methods for random forests and omics data sets. *Briefings in Bioinformatics*, *20*(2), 492–503. Retrieved from https://doi.org/10.1093%2Fbib%2Fbbx124 doi: 10.1093/bib/bbx124

Eisenhauer, E., Therasse, P., Bogaerts, J., Schwartz, L., Sargent, D., Ford, R., ... Verweij, J. (2009, Jan). New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer*, *45*(2), 228–247. Retrieved from https://doi.org/10.1016%2Fj.ejca.2008.10.026 doi: 10.1016/j.ejca.2008.10.026

Friedlander, P., & Cajulis, C. B. (2019, Aug). *Melanoma*. Wiley. Retrieved from https://doi.org/10.1002%2F9781119189596.ch18 doi: 10.1002/9781119189596.ch18

Ganapathy, S., Fagman, J. B., Shen, L., Yu, T., Zhou, X., Dai, W., ... Chen, C. (2016, Oct). Ral a, via activating the mitotic checkpoint, sensitizes cells lacking a functional nf1 to apoptosis in the absence of protein kinase c. *Oncotarget*, *7*(51), 84326–84337. Retrieved from https://doi.org/10.18632%2Foncotarget.12607 doi: 10.18632/oncotarget.12607

Gao, H., Korn, J. M., Ferretti, S., Monahan, J. E., Wang, Y., Singh, M., ... Sellers, W. R. (2015, Oct). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nature Medicine*, *21*(11), 1318–1325. Retrieved from https://doi.org/10.1038%2Fnm.3954 doi: 10.1038/nm.3954

Ghosh, R., Narasanna, A., Gonazalez-Angulo, A., Mills, G., & Arteaga, C. (2009, Dec). Differential signaling by ErbB receptor (HER) dimers: Implications for response to anti-HER2 therapies in breast cancer. *Cancer Research*, *69*(24_Supplement), 705–705. Retrieved from https://doi.org/10.1158%2F0008-5472.sabcs-09-705 doi: 10.1158/0008-5472.sabcs-09-705

Haem, E., Harling, K., Ayatollahi, S. M. T., Zare, N., & Karlsson, M. O. (2017, Jan). Adjusted adaptive lasso for covariate model-building in nonlinear mixed-effect pharmacokinetic models. *Journal of Pharmacokinetics and Pharmacodynamics*, *44*(1), 55–66. Retrieved from https://doi.org/10.1007%2Fs10928-017-9504-6 doi: 10.1007/s10928-017-9504-6

Hastie, T., Tibshirani, R., & Tibshirani, R. J. (2017, Jul). Extended Comparisons of Best Subset Selection, Forward Stepwise Selection, and the Lasso. *arXiv*, 1–19. Retrieved from http://arxiv.org/abs/1707.08692

Hoerl, A. E., & Kennard, R. W. (1970, Feb). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. Retrieved from https://doi.org/10.1080%2F00401706.1970.10488634 doi: 10.1080/00401706.1970.10488634

Jacob, L., Obozinski, G., & Vert, J.-P. (2009). Group lasso with overlap and graph lasso. In *Proceedings of the 26th annual international conference on machine learning - ICML '09*. ACM Press. Retrieved from https://doi.org/10.1145%2F1553374.1553431 doi: 10.1145/1553374.1553431

Jobard, E., Trédan, O., Bachelot, T., Vigneron, A. M., Aït-Oukhatar, C. M., Arnedos, M., ... Elena-Herrmann, B. (2017, Jun). Longitudinal serum metabolomics evaluation of trastuzumab and everolimus combination as pre-operative treatment for HER-2 positive breast cancer patients. *Oncotarget*, *8*(48), 83570–83584. Retrieved from https://doi.org/10.18632%2Foncotarget.18784 doi: 10.18632/oncotarget.18784

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., ... Ma'ayan, A. (2016, May). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(W1), W90–W97. Retrieved from https://doi.org/10.1093%2Fnar%2Fgkw377 doi: 10.1093/nar/gkw377

Kutmon, M., Riutta, A., Nunes, N., Hanspers, K., Willighagen, E. L., Bohler, A., ... Pico, A. R. (2015, Oct). WikiPathways: capturing the full diversity of pathway knowledge. *Nucleic Acids Research*, *44*(D1), D488–D494. Retrieved from https://doi.org/10.1093%2Fnar%2Fgkv1024 doi: 10.1093/nar/gkv1024

Lathrop, K., & Kaklamani, V. (2018, Dec). The response evaluation criteria in solid tumors (RECIST). In *Predictive biomarkers in oncology* (pp. 501–511). Springer International Publishing. Retrieved from https://doi.org/10.1007%2F978-3-319-95228-4_46 doi: 10.1007/978-3-319-95228-4_46

Lenters, V., Vermeulen, R., & Portengen, L. (2017, Sep). Performance of variable selection methods for assessing the health effects of correlated exposures in case–control studies. *Occupational and Environmental Medicine*, *75*(7), 522–529. Retrieved from https://doi.org/10.1136%2Foemed-2016-104231 doi: 10.1136/oemed-2016-104231

Liu, W., Wang, Q., & Chang, J. (2019). Global metabolomic profiling of trastuzumab resistant gastric cancer cells reveals major metabolic pathways and metabolic signatures based on UHPLC-q exactive-MS/MS. *RSC Advances*, *9*(70), 41192–41208. Retrieved from https://doi.org/10.1039%2Fc9ra06607a doi: 10.1039/c9ra06607a

Nguyen, T. H. T., Mouksassi, M.-S., Holford, N., Al-Huniti, N., Freedman, I., Hooker, A. C., ... and, F. M. (2017, Feb). Model evaluation of continuous data pharmacometric models: Metrics and graphics. *CPT: Pharmacometrics & Systems Pharmacology*, *6*(2), 87–109. Retrieved from https://doi.org/10.1002%2Fpsp4.12161 doi: 10.1002/psp4.12161

Nicolò, C., Périer, C., Prague, M., Bellera, C., MacGrogan, G., Saut, O., & Benzekry, S. (2020, Nov). Machine learning and mechanistic modeling for prediction of metastatic relapse in early-stage breast cancer. *JCO Clinical Cancer Informatics*(4), 259–274. Retrieved from https://doi.org/10.1200%2Fcci.19.00133 doi: 10.1200/cci.19.00133

R Core Team. (2020). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.r-project.org/

Ribba, B., Holford, N., Magni, P., Trocóniz, I., Gueorguieva, I., Girard, P., ... Friberg, L. (2014, May). A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis. *CPT: Pharmacometrics & Systems Pharmacology*, *3*(5), 113. Retrieved from https://doi.org/10.1038%2Fpsp.2014.12 doi: 10.1038/psp.2014.12

Ribbing, J., Nyberg, J., Caster, O., & Jonsson, E. N. (2007, May). The lasso—a novel method for predictive covariate model building in nonlinear mixed effects models. *Journal of Pharmacokinetics and Pharmacodynamics*, *34*(4), 485–517. Retrieved from https://doi.org/10.1007%2Fs10928-007-9057-1 doi: 10.1007/s10928-007-9057-1

Rodriguez-Brenes, I. A., Komarova, N. L., & Wodarz, D. (2013, Oct). Tumor growth dynamics: insights into evolutionary processes. *Trends in Ecology & Evolution*, *28*(10), 597–604. Retrieved from https://doi.org/10.1016%2Fj.tree.2013.05.020 doi: 10.1016/j.tree.2013.05.020

Ross, K. C., Chin, K. F., Kim, D., Marion, C. D., Yen, T. J., & Bhattacharjee, V. (2018, Jan). Methotrexate sensitizes drug-resistant metastatic melanoma cells to braf v600e inhibitors dabrafenib and encorafenib. *Oncotarget*, *9*(17), 13324–13336. Retrieved from https://doi.org/10.18632%2Foncotarget.24341 doi: 10.18632/oncotarget.24341

Segal, M., & Xiao, Y. (2011, Jan). Multivariate random forests. *WIREs Data Mining and Knowledge Discovery*, *1*(1), 80–87. Retrieved from https://doi.org/10.1002%2Fwidm.12 doi: 10.1002/widm.12

Sheiner, L. B., Rosenberg, B., & Melmon, K. L. (1972, Oct). Modelling of individual pharmacokinetics for computer-aided drug dosage. *Computers and Biomedical Research*, *5*(5), 441–459. Retrieved from https://doi.org/10.1016%2F0010-4809%2872%2990051-1 doi: 10.1016/0010-4809(72)90051-1

Shi, Y., Fan, X., Deng, H., Brezski, R. J., Rycyzyn, M., Jordan, R. E., ... An, Z. (2015, Mar). Trastuzumab triggers phagocytic killing of high HER2 cancer cells in vitro and in vivo by interaction with fcgamma receptors on macrophages. *The Journal of Immunology*, *194*(9), 4379–4386. Retrieved from https://doi.org/10.4049%2Fjimmunol.1402891 doi: 10.4049/jimmunol.1402891

Shlien, A., & Malkin, D. (2009). Copy number variations and cancer. *Genome Medicine*, *1*(6), 62. Retrieved from https://doi.org/10.1186%2Fgm62 doi: 10.1186/gm62

Silver, M., Montana, G., & In, A. D. N. (2012, Jan). Fast identification of biological pathways associated with a quantitative trait using group lasso with overlaps. *Statistical Applications in Genetics and Molecular Biology*, *11*(1), 1–43. Retrieved from https://doi.org/10.2202%2F1544-6115.1755 doi: 10.2202/1544-6115.1755

Simon, N., Friedman, J., & Hastie, T. (2013, Nov). A Blockwise Descent Algorithm for Group-penalized Multiresponse and Multinomial Regression. *arXiv*. Retrieved from http://arxiv.org/abs/1311.6529 doi: 10.48550/arXiv.1311.6529

Sloane, R. A. S., Gopalakrishnan, V., Reddy, S. M., Zhang, X., Reuben, A., & Wargo, J. A. (2017, May). Interaction of molecular alterations with immune response in melanoma. *Cancer*, *123*(S11), 2130–2142. Retrieved from https://doi.org/10.1002%2Fcncr.30681 doi: 10.1002/cncr.30681

Sumimoto, H., Imabayashi, F., Iwata, T., & Kawakami, Y. (2006, Jun). The BRAF–MAPK signaling pathway is essential for cancer-immune evasion in human melanoma cells. *Journal of Experimental Medicine*, *203*(7), 1651–1656. Retrieved from https://doi.org/10.1084%2Fjem.20051848 doi: 10.1084/jem.20051848

Tibshirani, R. (1996, Jan). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. Retrieved from https://doi.org/10.1111%2Fj.2517-6161.1996.tb02080.x doi: 10.1111/j.2517-6161.1996.tb02080.x

van Hasselt, J., Gupta, A., Hussein, Z., Beijnen, J., Schellens, J., & Huitema, A. (2015a, Jun). Disease progression/clinical outcome model for castration-resistant prostate cancer in patients treated with eribulin. *CPT: Pharmacometrics & Systems Pharmacology*, *4*(7), 386–395. Retrieved from https://doi.org/10.1002%2Fpsp4.49 doi: 10.1002/psp4.49

van Hasselt, J., Gupta, A., Hussein, Z., Beijnen, J., Schellens, J., & Huitema, A. (2015b, Jun). Integrated simulation framework for toxicity, dose intensity, disease progression, and cost effectiveness for castration-resistant prostate cancer treatment with eribulin. *CPT: Pharmacometrics & Systems Pharmacology*, *4*(7), 374–385. Retrieved from https://doi.org/10.1002%2Fpsp4.48 doi: 10.1002/psp4.48

Xie, G., Dong, C., Kong, Y., Zhong, J., Li, M., & Wang, K. (2019, Mar). Group lasso regularized deep learning for cancer prognosis from multi-omics and clinical features. *Genes*, *10*(3), 240. Retrieved from https://doi.org/10.3390%2Fgenes10030240 doi: 10.3390/genes10030240

Yuan, M., & Lin, Y. (2006, Feb). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(1), 49–67. Retrieved from https://doi.org/10.1111%2Fj.1467-9868.2005.00532.x doi: 10.1111/j.1467-9868.2005.00532.x

Zheng, S., & Liu, W. (2011, Nov). An experimental comparison of gene selection by lasso and dantzig selector for cancer classification. *Computers in Biology and Medicine*, *41*(11), 1033–1040. Retrieved from https://doi.org/10.1016%2Fj.compbiomed.2011.08.011 doi: 10.1016/j.compbiomed.2011.08.011

Supplementary material

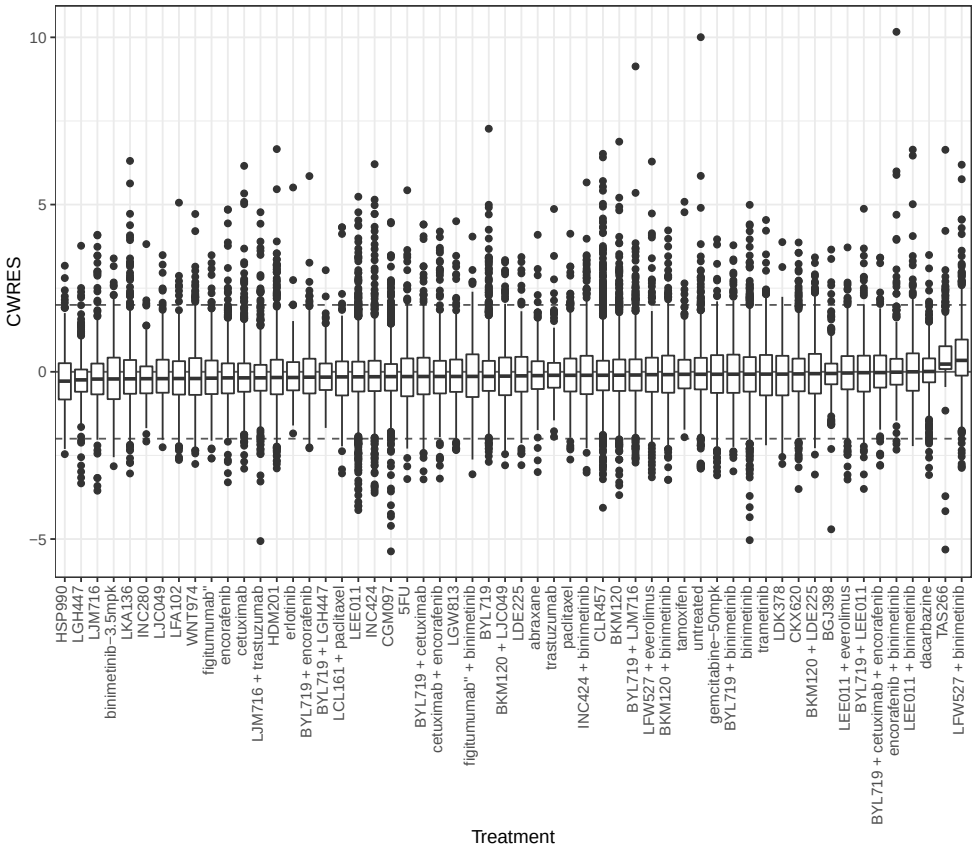


Figure S3.1: The conditional weighted residuals over the different treatments.

Table S3.1: Number of observations for every treatment, that were used in the second step of the analysis for the prediction of growth curves and the pathway selection.

Treatment	Number of PDXs
5FU	41
abraxane	33
BGJ398	62
binimetinib	166
binimetinib-3.5mpk	34
BKM120	168
BKM120 + binimetinib	56
BKM120 + LDE225	31
BKM120 + LJC049	39
BYL719	138
BYL719 + binimetinib	40
BYL719 + cetuximab	39
BYL719 + cetuximab + encorafenib	40
BYL719 + encorafenib	39
BYL719 + LEE011	36
BYL719 + LGH447	24
BYL719 + LJM716	133
cetuximab	64
cetuximab + encorafenib	39
CGM097	129
CKX620	64
CLR457	156
dacarbazine	30
encorafenib	72
encorafenib + binimetinib	32
erlotinib	22
figitumumab”	36
figitumumab” + binimetinib	35
gemcitabine-50mpk	32
HDM201	134
HSP990	22
INC280	23
INC424	69
INC424 + binimetinib	32
LCL161 + paclitaxel	23
LDE225	30
LDK378	31
LEE011	166
LEE011 + binimetinib	17
LEE011 + everolimus	37
LFA102	37
LFW527 + binimetinib	44
LFW527 + everolimus	29
LGH447	25
LGW813	32
LJC049	39
LJM716	38
LJM716 + trastuzumab	37
LKA136	111
paclitaxel	61
tamoxifen	38
TAS266	32
trametinib	34
trastuzumab	37
untreated	171
WNT974	65

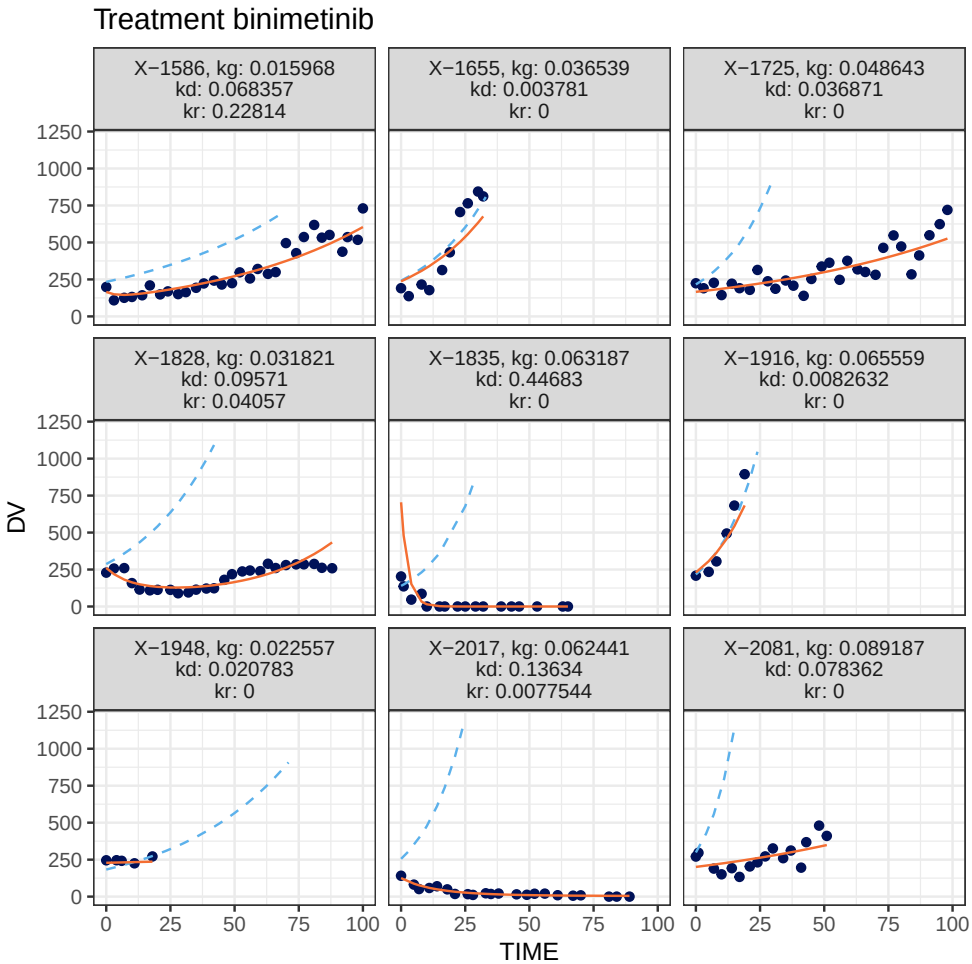


Figure S3.2: Set of figures with the non−linear mixed effect model fits for a subset of PDXs. Within every treatment, the volume (DV) is plotted against time (TIME) for every PDX. The modeled natural growth curves are shown (blue dashed line) and the model (red solid line) fit to the observed values (dark blue points). Tumor growth curves from treatment TAS266 show a bad model fit. Figures with model fits for all PDXs can be found online:

https://ascpt.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fpsp4.12603&file=psp412603-sup-0002-FigS2.pdf

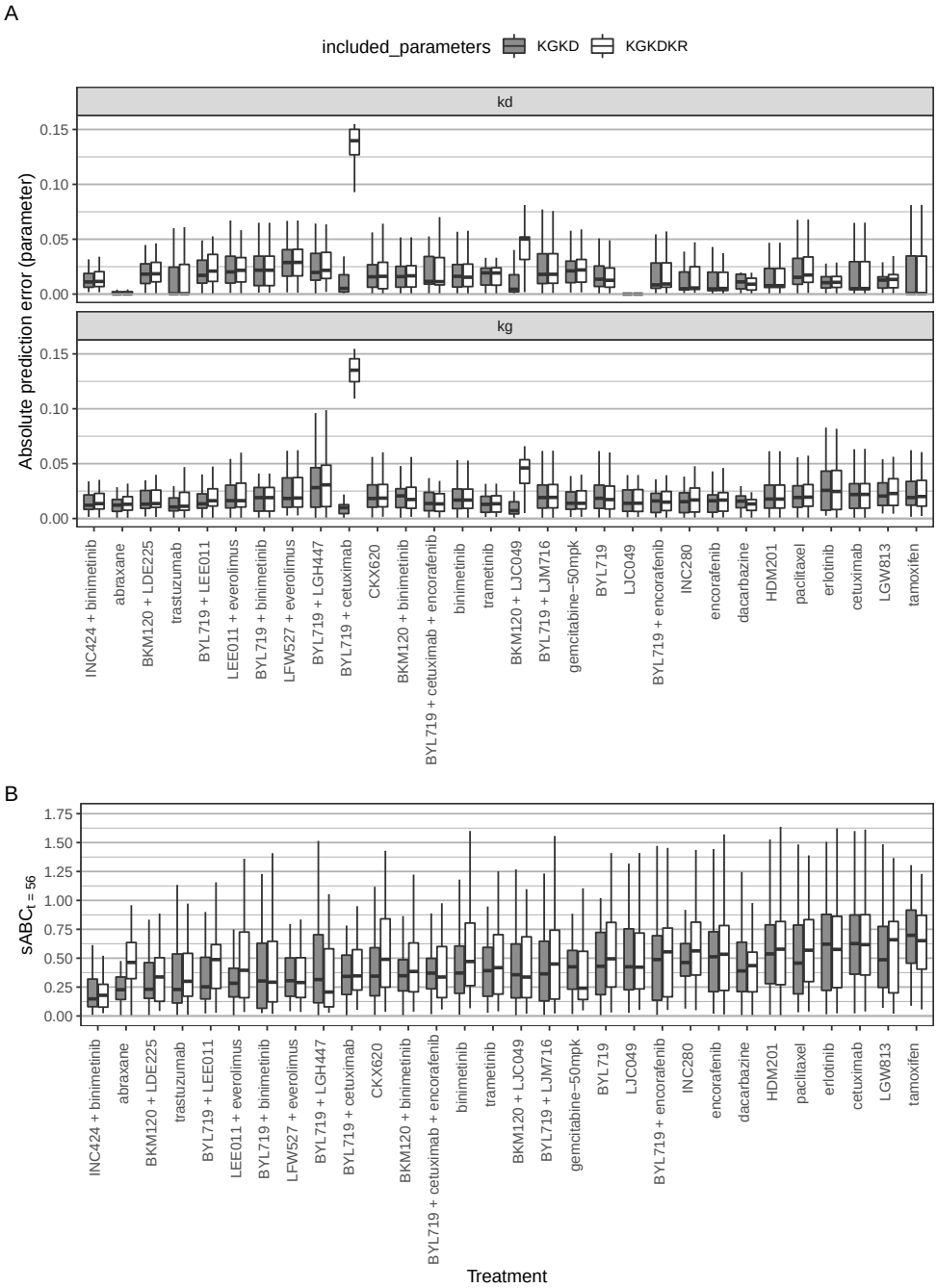


Figure S3.3: The prediction errors for the two models with (white) and without (grey) inclusion of k_r in the multivariate lasso for all treatments where the k_r was estimated non-zero in the non-linear mixed effect estimation. A) the absolute prediction error of the parameter values. B) the sABC for the same treatments. Both the prediction errors and the sABC are generally higher for the lasso without inclusion of k_r.

Chapter 4

Longitudinal metabolomics of community-acquired pneumonia

Authors

Laura B. Zwep*

Ilona den Hartog*

Thomas Hankemeier

Jacqueline J. Meulman

Ewoudt M. W. van de Garde

J. G. Coen van Hasselt

*Contributed equally

Abstract

Longitudinal biomarkers in patients with community-acquired pneumonia (CAP) may help monitoring of disease progression and treatment response. The metabolic host response could be a source of such biomarkers since it closely represents the current state of the patient. To this end, we performed longitudinal metabolic profiling for a comprehensive range of metabolites in patients with CAP. Previously collected serum samples from 25 patients with CAP with a confirmed Streptococcus pneumoniae infections were used. Samples were collected at multiple time points after hospital admission and up to 30 days after admission. A wide range of metabolites was measured, including amines, acylcarnitines, organic acids, and lipids. The associations between metabolites and C-reactive protein (CRP), procalcitonin, the CURB disease severity score (CURB) at admission, and total length of stay were examined. Distinct longitudinal profiles of metabolite profiles were identified, in particular for cholesteryl esters, diacyl-phosphatidylethanolamine, diacylglycerols, lysophosphatidylcholines, sphingomyelin, and triglycerides. Positive correlations were found between CRP and Phosphatidylcholine (PC) (34:1) (cor = 0.63) and negative correlations were found for CRP and nine lysophosphocholines (cor = 0.57 to 0.74). The CURB disease severity score was negatively associated with six metabolites, including acylcarnitines (tau = 0.64 to 0.58). Finally, we found negative correlations between the length of stay and six triglycerides (TGs), especially TGs (60:3) and (58:2) (cor = 0.63 and 0.61). In conclusion, the identified metabolites may provide inside into biological mechanisms underlying disease severity and may be of interest as potential biomarker to monitor treatment response.

4.1 Introduction

Community-acquired pneumonia (CAP) is a lower respiratory tract infection with a high incidence and is associated with the hospitalization of approximately one million adults per year (Battleman et al., 2002). The most common cause of CAP is *Streptococcus pneumoniae* (Meijvis et al., 2011). In hospitalized CAP patients, there is a need to monitor the antibiotic treatment response to optimize the treatment strategy (Pletz et al., 2022). In addition, there is a need for guidance on decisions about earlier termination of antibiotic treatment to minimize the risk of antimicrobial resistance. Monitoring of treatment response is currently achieved through observation of clinical symptoms and with inflammatory markers such as C reactive protein (CPR) and procalcitonin (PCT) (Aulin et al., 2021; Karakioulaki & Stolz, 2019). In particular, PCT is relevant for informing early treatment termination decisions but lacks predictive performance for CAP prognosis (Guo et al., 2018). Therefore, there is a need for biomarkers that give early insights into the clinical course of CAP.

Biomarkers that reflect the current physiological state of the patient have the potential to accurately monitor and predict the treatment response in CAP patients. Because the metabolome closely represents this physiological state, metabolomics-

techniques may enable discovery of relevant novel biomarkers. Indeed, for CAP and sepsis, the potential for metabolomics-based biomarkers measured at a static time point has been demonstrated (Seymour et al., 2013). However, the longitudinal monitoring of metabolic changes within patients may allow for an improved characterization of treatment response (Kohler et al., 2017). For example, CAP patients show a change in lysophosphatidylcholines that mirrors the transition from acute illness to recovery after starting antibiotic treatment (Müller et al., 2019). Further systematic characterization of longitudinal metabolic changes in CAP patients may thus be of relevance for identification of metabolic biomarkers that can predict and monitor the treatment response in these patients.

To this end, in this study, we aimed to comprehensively characterize the change of longitudinal metabolite profiles in hospitalized CAP patients with a confirmed *S. pneumoniae* infection using metabolomics, and relate these changes to disease severity, inflammation markers, and treatment response outcomes.

4.2 Materials and methods

4.2.1 Patient cohort

In this study, we utilized serum samples from 25 hospitalized CAP patients with an *S. pneumoniae* infection. These samples were previously collected as part of a larger clinical study that was performed between November 2007 and September 2010 (Meijvis et al., 2011). We selected samples from patients that had a confirmed infection with *S. pneumoniae*, while we excluded patients with a mixed infection or multiple pathogens. All patients that died within the study time were removed (one patient). Samples were collected at five times: on the day of admission (day 0), and days 1, 2, 4, and 30 after admission. CRP and creatinine were measured in the hospital setting at the same time points as blood samples were obtained. Not all time points were available for each patient, resulting in 115 samples over the 25 patients.

On the day of admission, disease severity was determined using the CURB score, which is a scoring system based on confusion, blood urea > 7 mmol/l, respiratory rate (RR) ≥ 30/min; systolic BP < 90 mmHg or diastolic BP ≤ 60 mmHg (Neill et al., 1996). A score of two or higher is classified as severe CAP.

4.2.2 Bio-analytical procedures

Serum samples were analyzed using five targeted LCMS methods and one targeted GCMS method by the Biomedical Metabolomics Facility of Leiden University, Leiden, The Netherlands, as described previously (den Hartog et al., 2021). A total of 369 unique metabolites was measured as relative levels, of which 6 metabolites were removed due to high missingness ($\leq$20%), resulting in 363 metabolites being evaluated in data analysis. Biochemically-selected sums and ratios of metabolites were calculated and added to the data (Table S4.1).

PCT was measured in the same serum samples used for the metabolomics analysis. PCT analysis was performed using the human procalcitonin CLIA kit from Abbexa (abx190129). Samples were measured in duplicate if sample volumes were sufficient (95% of samples).

4.2.3 Data analysis

The metabolite levels were scaled through log-transformation and standardization. To explore the variability of the high-dimensional metabolomics dataset, the dimension reduction method principal component analysis (PCA) was used. The PCA was used on the scaled metabolomics data over the different time points, with the metabolites as variables and each observation being a sample from a patient for a specific time point (Ham et al., 1997). As part of PCA, missing values were imputed through multiple imputation using expectation maximization (EM-PCA), which iteratively calculates the principal components and imputes the missing values (Josse et al., 2011).

To evaluate how much of the variation in the metabolites could be explained by the change over time, the first two principal components were related to time using a polynomial regression model. The importance of the metabolites to explain the variation between the patients over time was evaluated by evaluating the squared variable loadings. Specifically, the squared variable loadings within and between biochemical metabolite classes were evaluated to study similarities within classes and see which biochemical classes vary more between the patients.

To characterize the metabolic time profiles and profiles of current inflammation

Table 4.1: Patient characteristics

	CAP patients (N=25)
Age (years)	
Median [Min, Max]	67.0 [18.0, 98.0]
Sex	
Male	12 (48.0%)
Female	13 (52.0%)
CURB score	
Median [Min, Max]	1.00 [0, 3.00]
Duration of symptoms before admission (days)	
Median [Min, Max]	3.00 [1.00, 14.0]
Missing	15 (60.0%)
Antibiotic treatment before admission	
No	8 (32.0%)
Yes	2 (8.0%)
Missing	15 (60.0%)
Length of stay (days)	
Median [Min, Max]	7.50 [2.50, 24.5]

markers for different patients, we estimated the correlations between the scaled metabolite levels and the CRP, PCT and creatinine over time. Next, we evaluated which metabolites could be of interest for the prediction of the clinical course, by estimating the correlations between the scaled metabolite levels and a clinical disease severity marker (CURB score (Neill et al., 1996)) at hospital admission, and the outcome length of stay (LOS) in the hospital. Since the CURB and LOS are static values, while the metabolites changed over time, the correlations between these outcomes and the change in metabolite levels from baseline ($m_{t=k} - m_{t=0}$) at each time point (k) were calculated. The metabolites with the largest correlations were further evaluated in literature research to assess their biological function.

All analyses were performed in R. The scripts and data used for the analyses were deposited on GitHub (`github.com/vanhasseltlab/LongitudinalMetabolomicsCAP`).

4.3 Results

4.3.1 Metabolite time profiles

Metabolic profiling was performed for 25 patients and resulted in 363 metabolite levels on five time points. The patient characteristics are displayed in Table 4.1. Comorbidities in patients included kidney disease ($n = 1$), cardiovascular disease ($n = 4$), malignancy ($n = 2$), COPD ($n = 1$, $n_{missing} = 15$), diabetes ($n = 3$, $n_{missing} = 15$). No patients were using corticosteroids before admission ($n_{missing} = 15$).

Metabolite profiles within all CAP patients shifted over time, as shown in the PCA over all time points (Figure 4.1). The close relationship between metabolite levels and time is reflected in the results from the polynomial regression model which showed that 45% of the metabolite variation captured in these first two principal components

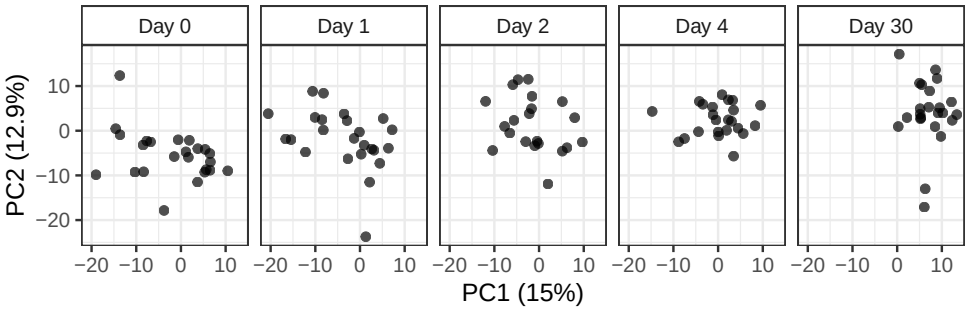


Figure 4.1: Patient metabolite profiles over time, in PCA scores. Every point represents the scores of an individual patient at a certain time point, in two dimensions based on the metabolite values. The panels show a trend over time of the metabolite profiles. Abbreviations: PC: principal component.

could be explained by time.

The metabolites that were targeted in this study were categorized into different biochemical classes. Metabolites from different biochemical classes showed distinct contributions to the total variation between the patients over time as was expressed in the variable loadings and directionality of the principal components (Figure 4.2). The squared PCA loadings represent the weight that the different metabolites in the biochemical class have in explaining the variation between patients over time. Of the variation in principal component one and two, 48% was explained by metabolites of the classes of cholesteryl esters, LPC's, sphingomyelins, diacylglycerols, and triglycerides (Figure 4.2A). The metabolites were categorized in classes based on their biochemistry and not based on their biological functions. The PCA results showed that metabolites that are categorized in the same class do not necessarily behave similarly (Figure 4.2B). For example, amino acids behave very differently from each other. Metabolites that do behave similarly in their biochemical class are for example triglycerides and sphingomyelins.

For each patient, the metabolic time profiles were shown as the two first components from the PCA (Figure 4.3, Figure S4.1). Generally, a shift from low to high principal component values was seen over time, corresponding to the shift in metabolite levels for the different metabolites (Figure 4.2B). The large variability in the time profiles, indicates a large interpatient variability in metabolic levels and changes over time.

4.3.2 Inflammation marker associations

To explore associations between metabolite profiles and inflammation, the metabolite values were compared to currently used inflammation biomarkers. Correlations were found between CRP and PCT and several metabolites. For example, phosphocholine (PC) (34:1) showed a positive correlation with CRP (cor = 0.63). Several individual lysophosphocholines (LPCs) and the sum of all LPCs showed a negative correlation with CRP (cor = -0.57 to -0.74, Figure 4.4A). PC (34:1) was found to decrease over time and several LPCs showed an increase over time, thereby mirroring the clinical disease progression (Figure 4.4B). Positive correlations with CRP and PCT were reported for the short-chain acylcarnitines (SCACs) tiglylcarnitine, 2 methylbutyroylcarnitine, and isovalerylcarnitine (cor with PCT = 0.61, 0.58, and 0.57; cor with CRP = 0.54, 0.64, and 0.51, respectively). Negative correlations were seen between the long-chain acylcarnitine (LCAC) stearoylcarnitine and CRP (cor = 0.62). This trend for decreasing SCACs over time is also represented by the positive correlation of CRP and PCT with the sum of all SCACs (cor = 0.55 and 0.53, respectively).

Correlations between metabolite levels and creatinine, a marker of renal failure, were also found. The same trends were seen for creatinine as for CRP and PCT (Figure S4.2). Also, strong positive correlations were found between creatine and 1-Methylhistidine, SDMA, inositol, homoserine, methionine sulfone, and octanoylcarnitine (cor > 0.7)

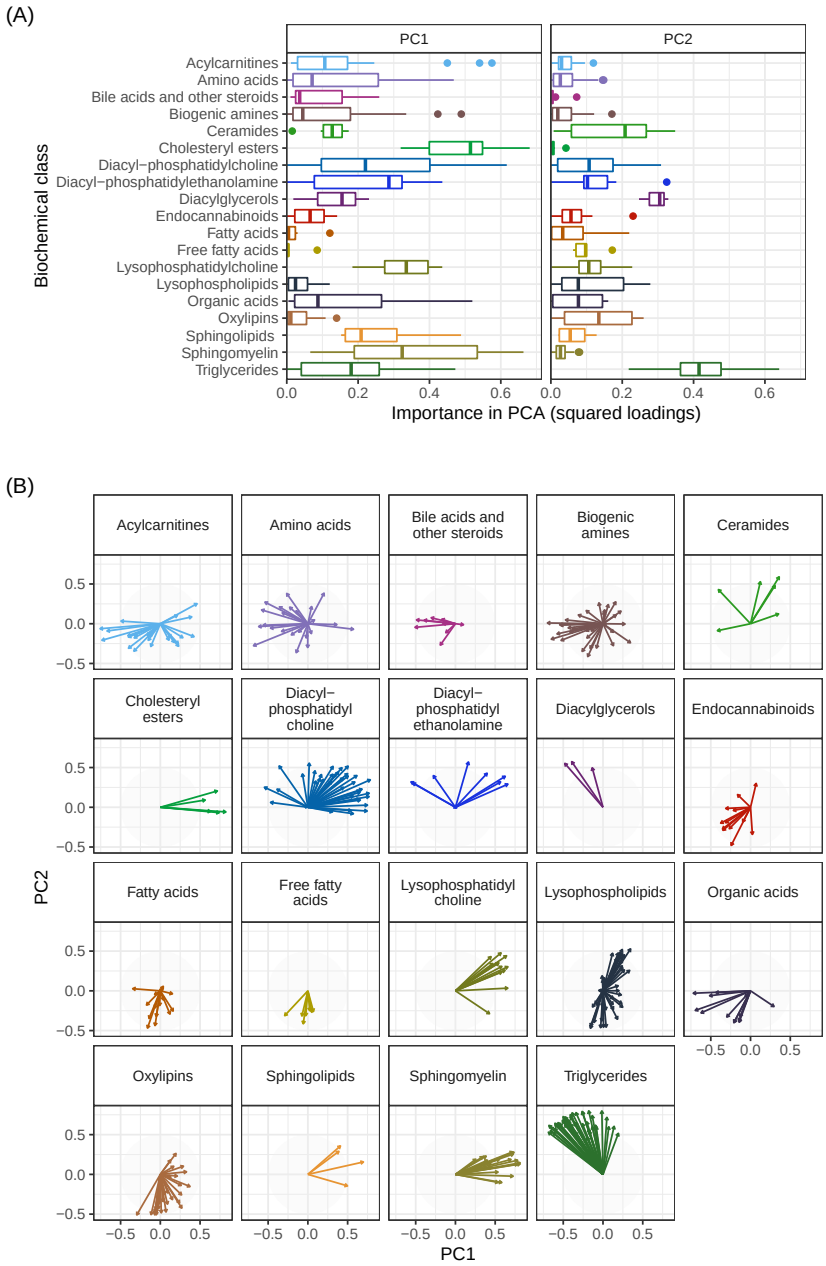


Figure 4.2: Metabolite contributions to the two dimensions of the PCA as variable loadings. A) The importance of each biochemical class for the different principal components (PCs), expressed by their squared metabolite loadings. Each box represents the squared loadings of the metabolites within a metabolic class. High squared loadings indicate a larger contribution to explaining the variation between patients. B) The loading plots for each biochemical metabolite class. The arrows indicate the importance (length) and direction of the metabolites in the principal component space. For example, high PC1 values correspond to high metabolite levels for metabolites with right pointing arrows, and low metabolite levels for metabolites with left pointing arrows. Arrows with a similar direction have similar metabolite patterns. Abbreviations: PC: principal component.

4.3.3 Disease severity score associations

To identify possible metabolic biomarkers for indication of disease severity, associations between the CURB disease severity score at admission and the change in metabolite levels on from day 0 to days 1, 2, 4, and 30 were evaluated (Figure S4.2). Negative associations were found between the CURB score and the change of metabolite levels (m) between day 0 and day 30 ($m_{t=30} - m_{t=0}$) of tiglylcarnitine, isovalerylcarnitine, 3 hydroxyisovaeric acid, carnitine, N6,N6,N6 trimethyl lysine, and isobutyryl carnitine (tau = 0.64 to 0.58, Figure 4.5). Patients with higher CURB scores showed decreasing levels of these metabolites.

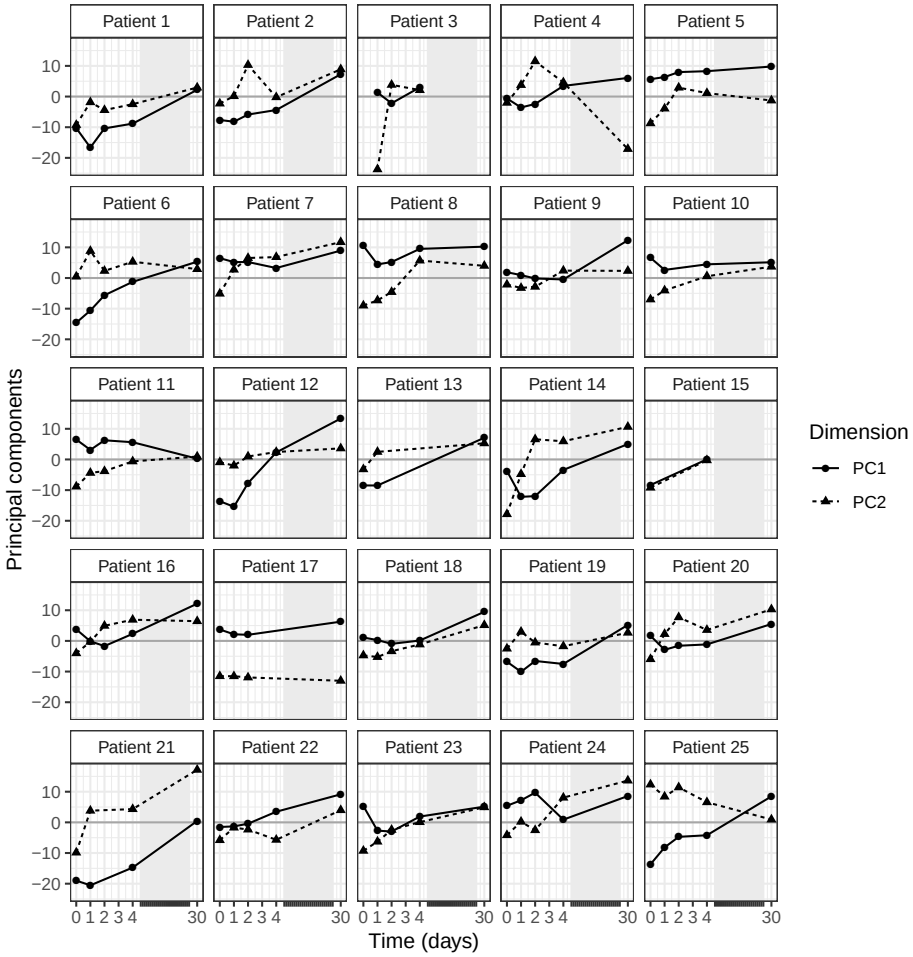


Figure 4.3: Individual metabolite profiles over time, expressed in PCA scores. The lines PC1 (solid) and PC2 (dashed), indicate the change in the corresponding principal component over time. Changes in PC values correspond to changes in metabolite levels according to their respective loadings. Abbreviations: PC: principal component.

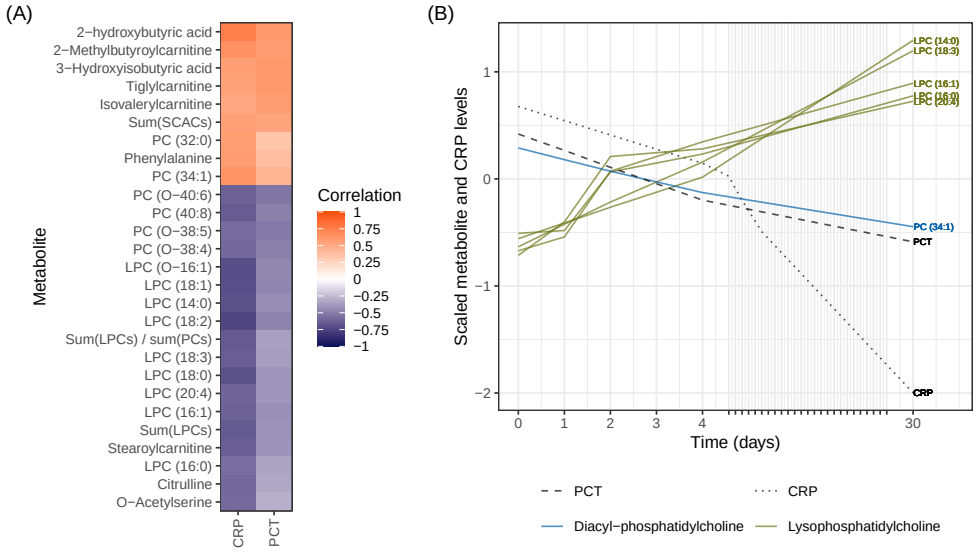


Figure 4.4: Correlations between inflammation markers CRP and PCT, and metabolites. A) The correlations between metabolites and CRP or PCT. Metabolites with a correlation >0.55 or <-0.55 for at least one marker are shown. A positive correlation (orange) indicates that a higher CRP or PCT level corresponds to an increase of that metabolite over time, while a negative correlation (blue) indicates a decrease over time for patients with a higher CRP or PCT level. B) Average CRP, PCT, PC (34:1), and LPC levels over time over all patients. Metabolite and CRP data were scaled. Abbreviations: see the abbreviation list.

4.3.4 Hospital length of stay associations

We evaluated the association between metabolites and clinical outcomes using the length of stay (LOS) as a potential surrogate endpoint. The strongest negative correlations to LOS were reported for the metabolite change over the first two days of admission ($m_{t=2} - m_{t=0}$, Figure 4.6), especially for the triglycerides (TGs) (60:3) and (58:2) (cor = 0.63 and 0.61 respectively). The correlations of these metabolites to LOS were much stronger than to CRP and PCT (cor = 0.08 and 0.25 respectively). Positive correlations were most pronounced when analyzing the metabolite change from the day of admission to day 30 ($m_{t=30} - m_{t=0}$). In the case of fatty acid (FA) (22:1) the day after admission ($m_{t=1} - m_{t=0}$) was the most strongly positively correlated to the LOS (cor = 0.58).

4.4 Discussion

In this study, we characterized the dynamics of the serum metabolites and their biochemical metabolite classes in pneumococcal CAP patients. We found that a large part of the variation in the metabolite values could be explained due to the changes over time within the patients. Several groups of metabolites were found to correlate with inflammation markers, CURB score, and length of hospital stay.

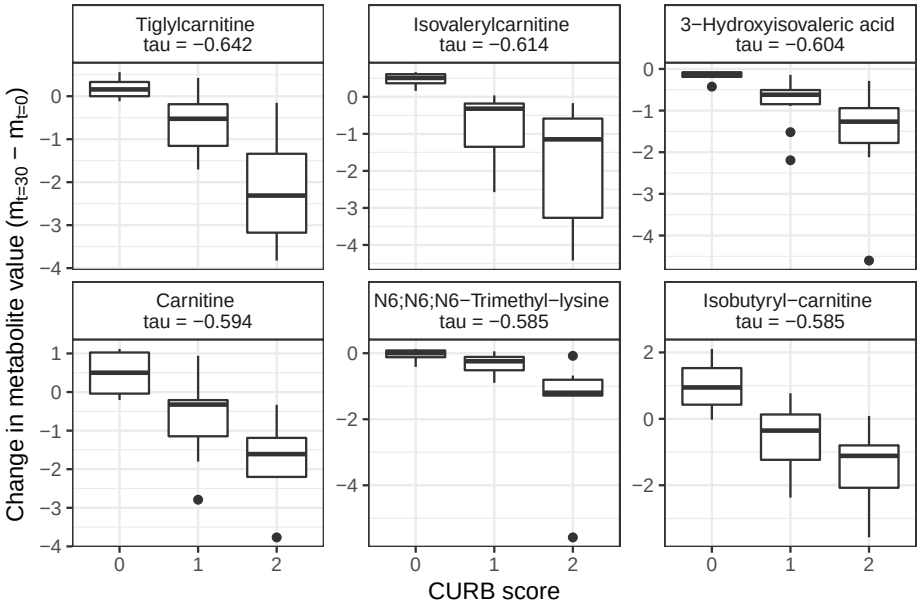


Figure 4.5: The correlation between the CURB score and six metabolites with highest associations. The change in metabolite level is the difference between the scaled metabolite level at day 30 and scaled metabolite level at admission (y-axis). These six metabolites all show a negative correlation with the CURB score (tau). This means, for patients with a CURB score of 0 the metabolite change between day 30 and day 0 is positive, so their metabolite levels were increasing over time. For patients with a CURB score of 2, the metabolite levels decreased over time.

The length of stay in the hospital was negatively correlated with the triglycerides, TG (60:3) and TG (58:2). Since these TGs are not highly correlated to CRP, PCT, or the CURB score, they explain a part of the variability of the disease progression that has not been studied before. Decreasing levels of TG (60:3) or (58:2) could be predictive for length of hospital stay. These results may not be specific for patients with *S. pneumoniae* infections. Triglycerides have not been found in metabolomics studies to etiological diagnosis of CAP, indicating its use for multiple infections, not just for pneumococcal CAP (den Hartog et al., 2021). TGs are also known to vary with diet, which could explain a negative correlation to disease severity (Parks, 2001).

PC (34:1) and LPCs (14:0), (16:0), (16:1), (18:0), (18:1), (18:2), (18:3) and (20:4) correlated to inflammatory markers, which also corresponds to previous findings (Banoei et al., 2020; Müller et al., 2019). PC (34:1), a ligand of nuclear receptor PPARα30, showed a positive correlation with CRP, which was previously associated with an anti-inflammatory response (Colombo et al., 2018). LPC (14:0) has been recently identified as a biomarker for disease severity in CAP patients (Nan et al., 2022). These metabolites could be of interest as treatment response biomarkers, not only in pneumococcal CAP patients, but also in other infections, because CRP and PCT are clinically used for many infections (Saleh et al., 2019). The CURB score was negatively associated with six metabolites, including some acylcarnitines. One of these acylcarnitines, tigylcar-

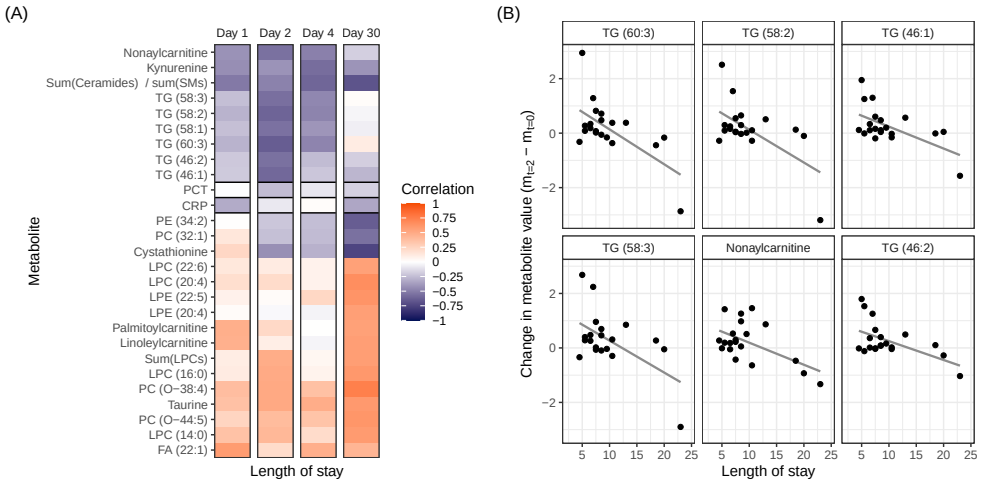


Figure 4.6: Associations between metabolites and length of stay. A) The correlations between the LOS and metabolite change from baseline at days 1, 2, 4, and 30 after admission ($m_{t=k} - m_{t=0}$). CRP and PCT are added as a reference. A positive correlation (orange) indicates that a longer stay in the hospital corresponds to an increase of that metabolite over time, while a negative correlation (blue) indicates a decrease over time for patients with longer stay. B) Metabolite levels over time for individual patients for metabolites with large negative correlations (cor < -0.55) over the first two days after admission. Abbreviations: see the abbreviation list.

4

nitine, has previously been found to be increased in non-survivors of CAP and could be considered a marker for disease severity (Banoei et al., 2020). Isovalerylcarnitine and isobutyrylcarnitine have, to our knowledge, not been studied as disease severity marker before, but may show a comparable performance to tigylcarnitine as their direction on the first principal component is similar.

We showed which biochemical metabolite classes explain most of the variation between individuals and over time. Triglycerides and LPCs were important for explaining the variation over time in the principal component analysis (PCA) and correlated with LOS and inflammatory markers. Within the biochemical classes, not all metabolites showed similar patterns, indicating that metabolites in some biochemical classes behave similarly during the infection, while metabolites in other classes behave differently (Figure 4.2B). The amino acids behave very differently, which could be expected since they are involved in a wide variety of biological functions (Wu, 2009). The longitudinal analysis of the metabolomics data enabled us to gain insight into acute and longer-term changes in the metabolome during the clinical course of CAP. The differences in metabolite levels are largely explained by changes over time, which could not have been evaluated without longitudinal data. Further studies using a longitudinal approach in this field could tackle long-existing issues in determining the appropriate empirical antibiotic therapy and guiding early targeted small spectrum antibiotic treatment or discontinuation.

This study was conducted in a well-characterized set 25 CAP patients with *S. pneumoniae* infections. The addition of patients with other causes of CAP is of in-

terest to compare metabolic time profiles for different treatment strategies based on the causative pathogen. Early recognition of a pathogen-drug mismatch using metabolomics could make antibiotic therapies more targeted and shorter. This study shows that mainly TGs, LPCs, PCs, and acylcarnitines are of interest for the disease severity and the length of stay for patients with CAP. By focusing on these metabolite classes, the number of metabolites that has to be measured for every patient can be reduced.

References

Aulin, L. B., de Lange, D. W., Saleh, M. A., van der Graaf, P. H., Völler, S., & van Hasselt, J. C. (2021). Biomarker-guided individualization of antibiotic therapy. *Clinical Pharmacology & Therapeutics*, *110*(2), 346-360. Retrieved from https://ascpt.onlinelibrary.wiley.com/doi/abs/10.1002/cpt.2194 doi: https://doi.org/10.1002/cpt.2194

Banoei, M. M., Vogel, H. J., Weljie, A. M., Yende, S., Angus, D. C., & Winston, B. W. (2020, Jul). Plasma lipid profiling for the prognosis of 90-day mortality, in-hospital mortality, ICU admission, and severity in bacterial community-acquired pneumonia (CAP). *Critical Care*, *24*(1). Retrieved from https://doi.org/10.1186%2Fs13054-020-03147-3 doi: 10.1186/s13054-020-03147-3

Battleman, D. S., Callahan, M., & Thaler, H. T. (2002, Mar). Rapid antibiotic delivery and appropriate antibiotic selection reduce length of hospital stay of patients with community-acquired pneumonia. *Archives of Internal Medicine*, *162*(6), 682. Retrieved from https://doi.org/10.1001%2Farchinte.162.6.682 doi: 10.1001/archinte.162.6.682

Colombo, S., Melo, T., Martínez-López, M., Carrasco, M. J., Domingues, M. R., Pérez-Sala, D., & Domingues, P. (2018, Aug). Phospholipidome of endothelial cells shows a different adaptation response upon oxidative, glycative and lipoxidative stress. *Scientific Reports*, *8*(1). Retrieved from https://doi.org/10.1038%2Fs41598-018-30695-0 doi: 10.1038/s41598-018-30695-0

den Hartog, I., Zwep, L. B., Vestjens, S. M. T., Harms, A. C., Voorn, G. P., de Lange, D. W., ... van Hasselt, J. G. C. (2021, Jun). Metabolomic profiling of microbial disease etiology in community-acquired pneumonia. *PLOS ONE*, *16*(6), e0252378. Retrieved from https://doi.org/10.1371%2Fjournal.pone.0252378 doi: 10.1371/journal.pone.0252378

Guo, S., Mao, X., & Liang, M. (2018, Oct). The moderate predictive value of serial serum CRP and PCT levels for the prognosis of hospitalized community-acquired pneumonia. *Respiratory Research*, *19*(1). Retrieved from https://doi.org/10.1186%2Fs12931-018-0877-x doi: 10.1186/s12931-018-0877-x

Ham, T. V. D., Meulman, J. J., Strien, D. C. V., & Engeland, H. V. (1997, Apr). Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, *170*(4), 363–368. Retrieved from https://doi.org/10.1192%2Fbjp.170.4.363 doi: 10.1192/bjp.170.4.363

Josse, J., Pagès, J., & Husson, F. (2011, Mar). Multiple imputation in principal component analysis. *Advances in Data Analysis and Classification*, *5*(3), 231–246. Retrieved from https://doi.org/10.1007%2Fs11634-011-0086-7 doi: 10.1007/s11634-011-0086-7

Karakioulaki, M., & Stolz, D. (2019, Apr). Biomarkers in pneumonia—beyond procalcitonin. *International Journal of Molecular Sciences*, *20*(8), 2004. Retrieved from https://doi.org/10.3390%2Fijms20082004 doi: 10.3390/ijms20082004

Kohler, I., Hankemeier, T., van der Graaf, P. H., Knibbe, C. A., & van Hasselt, J. C. (2017). Integrating clinical metabolomics-based biomarker discovery and clinical pharmacology to enable precision medicine. *European Journal of Pharmaceutical Sciences*, *109*, S15-S21. Retrieved from https://www.sciencedirect.com/science/article/pii/S0928098717302464 (Special issue in honour of Professor Meindert Danhof) doi: https://doi.org/10.1016/j.ejps.2017.05.018

Meijvis, S. C., Hardeman, H., Remmelts, H. H., Heijligenberg, R., Rijkers, G. T., van Velzen-Blad, H., ... Biesma, D. H. (2011, Jun). Dexamethasone and length of hospital stay in patients with community-acquired pneumonia: a randomised, double-blind, placebo-controlled trial. *The Lancet*, *377*(9782), 2023–2030. Retrieved from https://doi.org/10.1016%2Fs0140-6736%2811%2960607-7 doi: 10.1016/s0140-6736(11)60607-7

Müller, D. C., Kauppi, A., Edin, A., Gylfe, Å., Sjöstedt, A. B., & Johansson, A. (2019, May). Phospholipid levels in blood during community-acquired pneumonia. *PLOS ONE*, *14*(5), e0216379. Retrieved from https://doi.org/10.1371%2Fjournal.pone.0216379 doi: 10.1371/journal.pone.0216379

Nan, W., Xiong, F., Zheng, H., Li, C., Lou, C., Lei, X., ... Li, Y. (2022). Myristoyl lysophosphatidylcholine is a biomarker and potential therapeutic target for community-acquired pneumonia. *Redox Biology*, *58*, 102556. Retrieved from https://www.sciencedirect.com/science/article/pii/S2213231722003287 doi: https://doi.org/10.1016/j.redox.2022.102556

Neill, A., Martin, I., Weir, R., Anderson, R., Chereshsky, A., Epton, M., ... others (1996). Community acquired pneumonia: aetiology and usefulness of severity criteria on admission. *Thorax*, *51*(10), 1010–1016.

Parks, E. J. (2001, Oct). Effect of dietary carbohydrate on triglyceride metabolism in humans. *The Journal of Nutrition*, *131*(10), 2772S–2774S. Retrieved from https://doi.org/10.1093%2Fjn%2F131.10.2772s doi: 10.1093/jn/131.10.2772s

Pletz, M. W., Jensen, A. V., Bahrs, C., Davenport, C., Rupp, J., Witzenrath, M., ... Rohde, G. (2022, Sep). Unmet needs in pneumonia research: a comprehensive approach by the CAPNETZ study group. *Respiratory Research*, *23*(1). Retrieved from https://doi.org/10.1186%2Fs12931-022-02117-3 doi: 10.1186/s12931-022-02117-3

Saleh, M. A., van de Garde, E. M., & van Hasselt, J. C. (2019). Host-response biomarkers for the diagnosis of bacterial respiratory tract infections. *Clinical Chemistry and Laboratory Medicine (CCLM)*, *57*(4), 442–451. Retrieved 2022-12-05, from https://doi.org/10.1515/cclm-2018-0682 doi: doi:10.1515/cclm-2018-0682

Seymour, C. W., Yende, S., Scott, M. J., Pribis, J., Mohney, R. P., Bell, L. N., ... Angus, D. C. (2013, May). Metabolomics in pneumonia and sepsis: an analysis of the GenIMS cohort study. *Intensive Care Medicine*, *39*(8), 1423–1434. Retrieved from https://doi.org/10.1007%2Fs00134-013-2935-7 doi: 10.1007/s00134-013-2935-7

Wu, G. (2009, Mar). Amino acids: metabolism, functions, and nutrition. *Amino Acids*, *37*(1), 1–17. Retrieved from https://doi.org/10.1007%2Fs00726-009-0269-0 doi: 10.1007/s00726-009-0269-0

Supplementary material

Table S4.1: Metabolite ratios and sums

Metabolite sum or ratio name in R	Metabolite sum or ratio formula
BCAA_sum	isoleucine + leucine + valine
TCA_cycle_sum	Citric acid + lactic acid + malic acid + fumaric acid
urea_cycle_sum	Citrulline + arginine + ornithine + fumaric acid
lc_Carnitines_sum	Myristoilcarnitine + Hexadecenoylcarntine + Palmitoyl-carnitine + Stearoylcarnitine + Dodecenoylcarnitine + Tetradecenoylcarnitine + Linoleylcarnitine + Oleylcarnitine + Tetradecadienylcarntine
mc_Carnitines_sum	Hexanoylcarnitine + Octanoylcarnitine + Octenoylcarnitine + Decanoylcarnitine + Lauroylcarnitine + Nonaylcarnitine + Pimeylcarnitine + Decenoylcarnitine
sc_Carnitines_sum	Acetylcarnitine + Propionylcarnitine + Isobutyrylcarnitine + Butyrylcarnitine + Tiglylcarnitine + Methylbutyroylcarnitine + Isovalerylcarnitine
Cer_sum	Cer(d18:1/22:1) + Cer. (d18:1/24.1. + Cer(d18:1/24:0) + Cer(d18:1/16:0) + Cer(d18:1/23:0) + Cer(d18:1/24:0)
SM_sum	Sphingomyelin (d18:1/14:0) + (d18:1/15:0) + (d18:1/16:0) + (d18:1/16:1) + (d18:1/17:0) + (d18:1/18:0) + (d18:1/18:1) + (d18:1/18:2) + (d18:1/20:0) + (d18:1/20:1) + (d18:1/21:0) + (d18:1/22:0) + (d18:1/22:1) + (d18:1/23:0) + (d18:1/ 23:1) + (d18:0/24:0) + (d18:0/24:1) + (d18:0/24:2) + (d18:0/25:0) + (d18:0/25:1)
LPC_sum	Lysophosphatidylcholine (14:0) + (16:0) + (16:1) + (18:0) + (18:1) + (18:2) + (18:3) + (20:4) + (20:5) + (22:6) + (O-16:1) + (O-18:1)
PC_sum	Diacyl-phosphatidylcholine (32:0) + (32:1) + (32:2) + (34:1) + (34:2) + (34:3) + (34:4) + (36:1) + (36:2) + (36:3) + (36:4) + (36:5) + (36:6) + (38:2) + (38:3) + (38:4) + (38:5) + (38:6) + (38:7) + (40:4) + (40:5) + (40:6) + (40:7) + (40:8) + (O-34:1) + (O-34:2) + (O-34:3) + (O-36:2) + (O-36:3) + (O-36:4) + (O-36:5) + (O-36:6) + (O-38:4) + (O-38:5) + (O-38:6) + (O-38:7) + (O-40:6) + (O-42:6) + (O-44:5)
HT5_Trp_ratio	Serotonine / Tryptophan
ADMA_Arg_ratio	ADMA / Arginine
SDMA_Arg_ratio	SDMA / Arginine
Carnitine_sum_lc_Carnitines_ratio	Carnitine / LCAC sum
Carnitine_sum_mc_Carnitines_ratio	Carnitine / MCAC sum
Carnitine_sum_sc_Carnitines_ratio	Carnitine / SCAC sum
DCA_CA_ratio	DCA / CA
FA_14.1_14.0	FA (14:1) / FA (14:0)
FA_16.1_16.0	FA (16:1) / FA(16:0)
Gln_Glu	Glutamine / Glutamic acid
Kyn_Trp	Kynurenine / Tryptophan
sum_BCAA_sum_Phe_Tyr_ratio	BCAA sum / (Phenylalanine + Tyrosine)
sum_CER_sum_SM_ratio	Cer sum / SM sum
sum_LPC_sum_PC_ratio	LPC sum / PC sum

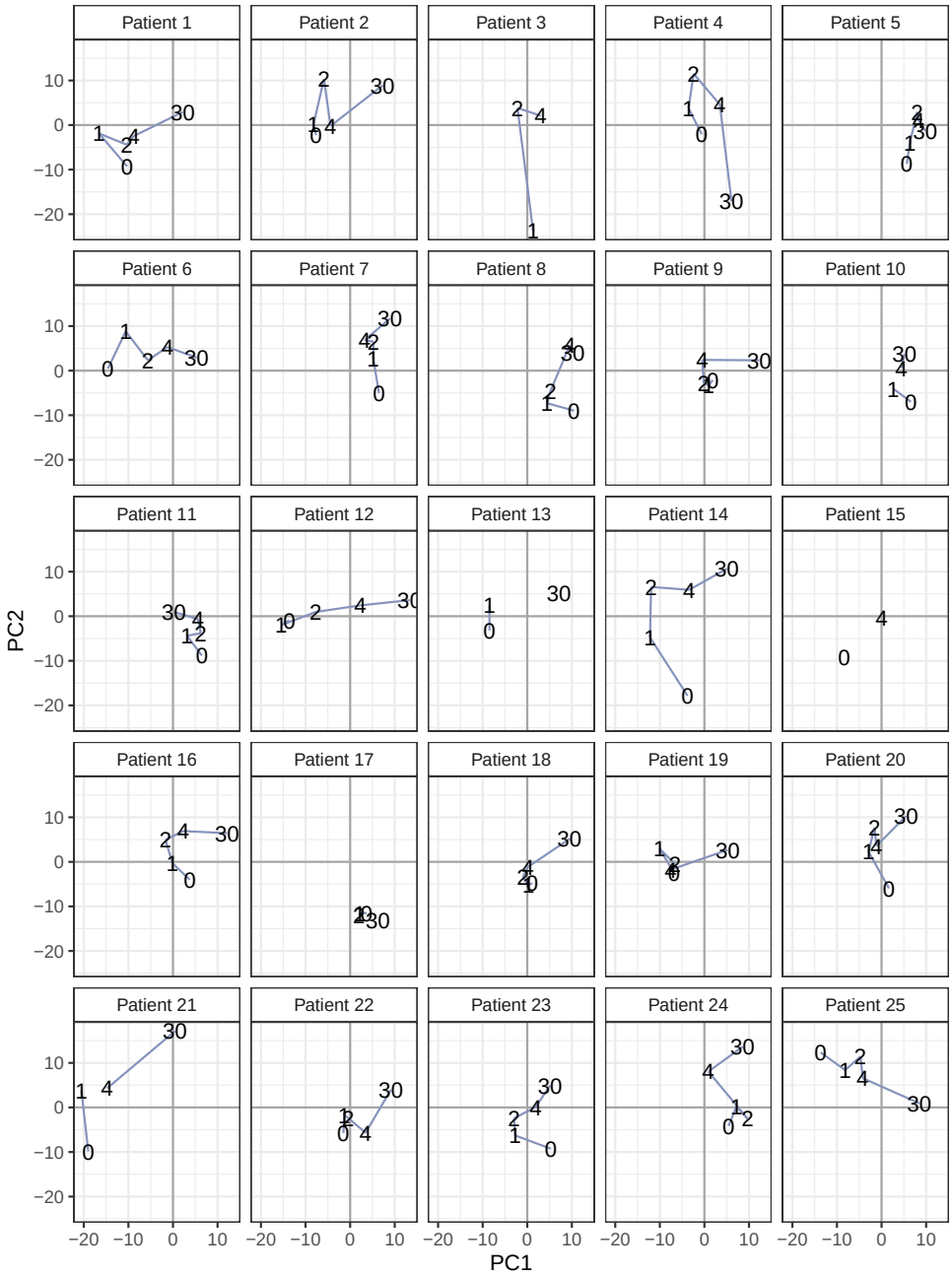


Figure S4.1: PCA score plots for each patient. For each patient, the time points are labelled and connected with lines. Abbreviations: PC: principal component.

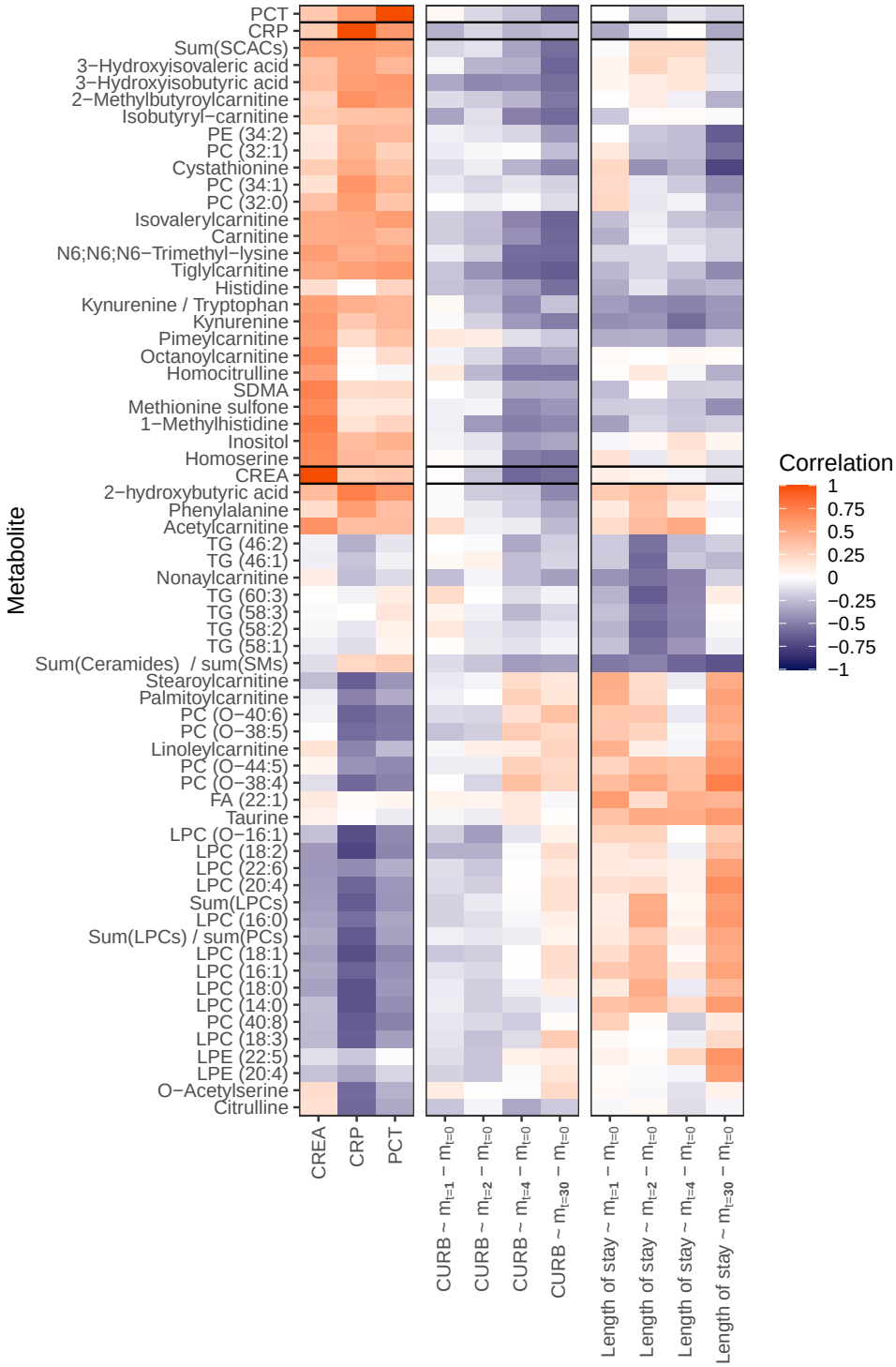


Figure S4.2: The correlations between metabolites and creatinine, CRP, and PCT over time; and the correlations of the CURB score and length of stay with a change of the metabolites between day k and day 0, where the change in metabolite levels is denoted by $m_{t=k} - m_{t=0}$.

Chapter 5

Identification of antibiotic collateral sensitivity and resistance interactions in population surveillance data

Authors

Laura B. Zwep

Yob Haakman

Kevin L. W. Duisters

Jacqueline J. Meulman

Apostolos Liakopoulos

J. G. Coen van Hasselt

JAC-Antimicrobial Resistance 2021; 3(4), 1–9

Abstract

Background Collateral effects of antibiotic resistance occur when resistance to one antibiotic agent leads to increased resistance or increased sensitivity to a second agent, known respectively as collateral resistance (CR) and collateral sensitivity (CS). Collateral effects are relevant to limit impact of antibiotic resistance in design of antibiotic treatments. However, methods to detect antibiotic collateral effects in clinical population surveillance data of antibiotic resistance are lacking.
Objectives To develop a methodology to quantify collateral effect directionality and effect size from large-scale antimicrobial resistance population surveillance data.
Methods We propose a methodology to quantify and test collateral effects in clinical surveillance data based on a conditional t-test. Our methodology was evaluated using MIC data for 419 *Escherichia coli* strains, containing MIC data for 20 antibiotics, which were obtained from the Pathosystems Resource Integration Center (PATRIC) database.
Results We demonstrate that the proposed approach identifies several antibiotic combinations that show symmetrical or non-symmetrical CR and CS. For several of these combinations, collateral effects were previously confirmed in experimental studies. We furthermore provide insight into the power of our method for multiple collateral effect sizes and MIC distributions.
Conclusions Our proposed approach is of relevance as a tool for analysis of large-scale population surveillance studies to provide broad systematic identification of collateral effects related to antibiotic resistance, and is made available to the community as an R package. This method can help mapping CS and CR, which could guide combination therapy and prescribing in the future.

5.1 Introduction

The treatment of bacterial infections increasingly relies on antibiotic combination therapy (Tamma et al., 2012). Although physiological interactions, i.e. synergy and antagonism, between pairs of antibiotics have been explored and exploited for such combination therapies (Eliopoulos & Moellering, 1982), evolutionary interactions resulting in collateral effects have only recently started to attract attention (Baym et al., 2016). Negative evolutionary interactions between antibiotics, known as collateral sensitivity (CS), occur when the emergence of resistance to an antibiotic is accompanied by increased sensitivity to a second antibiotic. On the contrary, positive evolutionary interactions, known as collateral resistance (CR), result in increased resistance to the second antibiotic (Pál et al., 2015).

The broad systematic identification of CR can be clinically important to avoid evolutionary unfavourable antibiotic combinations in empirical treatment (Amsalu et al., 2020), whereas CS can enable the design of antibiotic combination treatment strategies to suppress resistance (Imamovic & Sommer, 2013; Imamovic et al., 2018). Although perturbations of gene expression networks that subsequently affect the vul-

nerability of bacterial cells to chemicals have been proposed as the main mechanism underlying CS, the mechanistic details of CS remain elusive (Pál et al., 2015). For antibiotic pairs that show a CS relationship we may see either a unidirectional or reciprocal relationship, where the latter is most suitable to design such resistance suppressing cycling strategies (Imamovic & Sommer, 2013; Maltas & Wood, 2019; Barbosa et al., 2017; Podnecky et al., 2018; Liakopoulos et al., 2022). CS has been primarily identified and studied in controlled experimental evolution studies, mostly utilizing laboratory strains (Wright, 2007; Aulin et al., 2020).

In the experimental setting, collateral effects, and CS in particular, are determined by measuring the MIC against multiple antibiotics before and after desensitization, i.e. through development of resistance to a chosen antibiotic by experimental evolution. The fold change in MIC, the ratio of the MIC after and before desensitization, is then used to quantify a collateral response

The clinical relevance of CS effects remains unclear, due to a lack of studies that characterize collateral effects and CS in particular in clinically isolated bacterial pathogens. Unlike experimentally evolved laboratory strains, clinical bacterial isolates are associated with extensive genetic variability, and a parental wild-type strain is lacking to readily determine collateral effects such as is done experimentally (Turnidge et al., 2006).

Increasing availability of large-scale clinical antimicrobial susceptibility surveillance data (Wattam et al., 2016; World Health Organization, 2014) may offer an opportunity to address this knowledge gap. Such datasets include MIC values for commonly used antibiotics in clinically isolated pathogens. Recently, it was shown how dichotomous resistance values, i.e. a classification of sensitivity or resistance, can be used to estimate collateral effects of antibiotics in clinical population data (Obolski et al., 2016). However, in order to address questions about the therapeutic relevance of CS, it is of specific importance to be able to infer directionality and effect size of collateral effects from available clinical MIC data, which is not possible when only dichotomous MIC values are considered.

Here, we propose a methodology to systematically identify and quantify collateral effects from clinical MIC surveillance data, by comparing two MIC distributions conditional on the resistance to an antibiotic. Specifically, we develop a goodness-of-fit measure for estimating a non-causal collateral effect using the easily interpretable conditional t-test, which allows quantification of collateral effect directionality and effect size (Figure 5.1). We apply our method to a large public dataset with MIC measurements for multiple antibiotics in clinical *Escherichia coli* isolates to identify possible collateral effects.

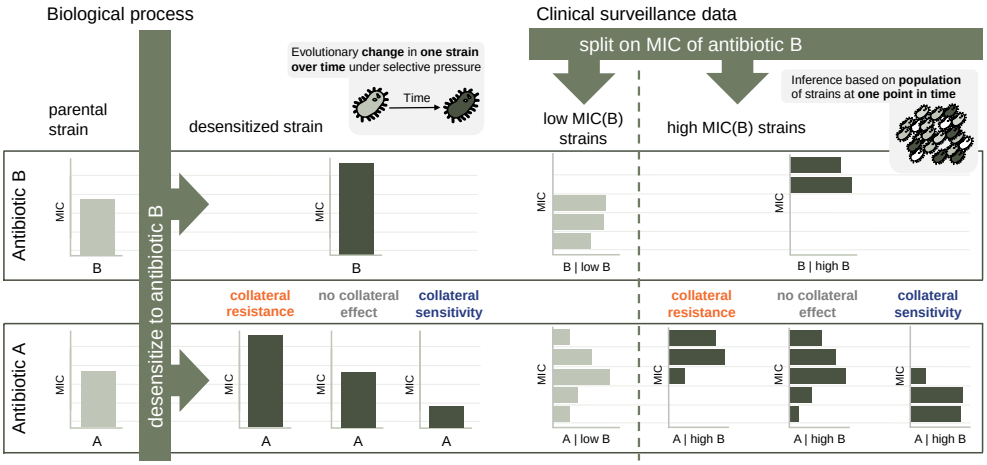


Figure 5.1: Collateral effects in the biological process and in population-based inference. Left panel: MICs from the parental strain change due to desensitization to antibiotic B. The MIC for B increases; the effect on the MIC for A indicates whether B has a collateral effect on antibiotic A.Right panel: The parental strain is unknown, so population of strains is divided into a high MIC(B) and low MIC(B) group; instead of MICs for individual strains, the MIC graphs show histograms of strains. Collateral effects are measured by comparing the conditional distribution for high and low MIC for antibiotic B.

5.2 Methods

5.2.1 Data pre-processing

Antibiotic MIC data of clinically isolated *E. coli* strains were obtained from the Pathosystems Resource Integration Center (PATRIC) (Wattam et al., 2016) using the command line interface, resulting in a dataset of 60 antibiotics and 495 strains. Antibiotics with MIC measurements for at least 200 strains were included in the study, resulting in 20 antibiotics (Table S5.1). Only strains with data for two or more antibiotics were used for further analysis, resulting in 419 eligible strains (Figure 5.2a). The MICs were measured on a 2-fold concentration scale. The data can be considered discrete, with only 26 unique values and 95% of the measurements falling within 9 of these values. The MICs were $\log_2$ transformed [$\log_2$(MIC)] to put them on a linear scale.

5.2.2 Collateral effect identification

To reflect the biological process of collateral effects, a pair of antibiotics (A and B) was tested by splitting the population of strains in two groups, one with high and one with low MIC for antibiotic B (Figure 5.1). For computing the collateral effects, only complete pairwise observations were considered (Figure 5.2b). Next, strains were dichotomized on a dichotomization criterion τ, based on their MIC value for antibiotic B. After dichotomization, the $\log_2$ fold change (FC) was calculated as the difference between the mean $\log_2(\text{MIC})$ for A given high MIC for B ($\text{MIC}_{A|B=\text{high}}$) and the mean

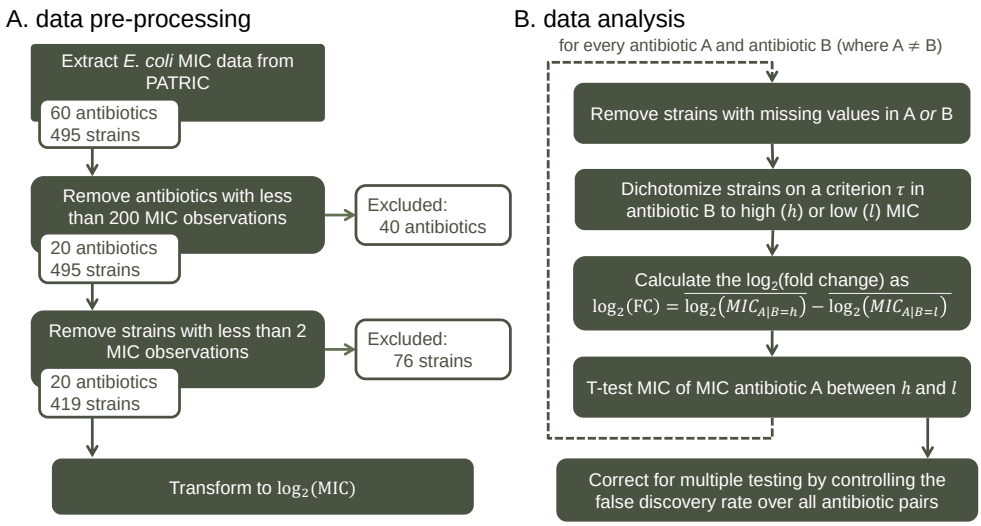


Figure 5.2: Data pre-processing steps (a) and data analysis strategy (b) schematic overview.

$\log_2(\text{MIC})$ for A given low MIC for B ($\text{MIC}_{A|B=\text{low}}$) (Equation 5.1).

$$\log_2 FC = \text{mean}(\log_2(\text{MIC}_{A|B=\text{high}})) - \text{mean}(\log_2(\text{MIC}_{A|B=\text{low}})) \tag{5.1}$$

where a $\log_2$ FC > 0 indicates a CR effect (or an MDR phenotype due to the co-presence of distinct antibiotic resistance mechanisms) and a $\log_2$ FC < 0 indicates a CS effect. Collateral effects between two antibiotics were tested in two directions: the collateral effect of B on A and the collateral effect of A on B. To test the fold change, the mean difference between the two groups was compared with an independent sample t-test on the log scale. Let

$$\mu_{A|B=\text{high}} = \text{mean}(\log_2(MIC_{A|B=\text{high}}))$$

then the tested hypotheses can be formulated as

$$H_0 : \mu_{A|B=\text{high}} = \mu_{A|B=\text{low}}$$

$$H_{CS} : \mu_{A|B=\text{high}} < \mu_{A|B=\text{low}}$$

$$H_{CR} : \mu_{A|B=\text{high}} > \mu_{A|B=\text{low}}$$

The grouping in B depends on the dichotomization criterion τ. We chose the τ in a way to make the two groups (high and low) most equally sized, to maximize the power of detecting collateral effects. In continuous data, this is naturally the median. The dichotomization does not depend on antibiotic A. To study the effect of different values of s, we evaluated the results for multiple values. The number of options for τ was very limited due to the discrete nature of the MIC observations. Importantly, the

MIC for antibiotic A should not be dichotomized, to preserve statistical power and retrieve continuous effect sizes. Between two sets of strains with either a high or low MIC for an antibiotic B, the MIC of antibiotic A is expected to be similar; when this is not the case, this might indicate a collateral effect.

The difference between the mean $\log_2(\text{MIC})$ for all the combinations of the 20 antibiotics was tested in two directions, which resulted in 380 statistical tests. A two-sided t-test was used to test both CS and CR. The power, the probability of detecting true effects, was evaluated analytically over different effect sizes of collateral effects in MIC data based on sample sizes, disbalance between the sizes of the high and low MIC(B)groups, and the standard deviation. The values over which the power was calculated were between the ranges found in the extracted data from PATRIC.

The corresponding p-values were calculated and adjusted by controlling the false discovery rate (FDR) with the Benjamini–Yekutieli procedure, which corrects for the dependency between the tests (Benjamini & Yekutieli, 2001). The allowed FDR was set to 0.05. The difference between the means was visualized for the significant results in a heatmap. The distributions of the MICs with the most significant differences in mean for CR and CS were also visualized as histograms.

The data analysis was done in the statistical scripting language R (R Core Team, 2020), using the ggplot2 (Wickham, 2008) package for visualization. The code used in this study is available on github (`github.com/vanhasseltlab/CollateralEffect_MICmethod`). The functions for testing collateral effects and producing figures are made available in the R package collatRal (`github.com/vanhasseltlab/collatRal`).

5.3 Results

5.3.1 Detection of collateral effects

A total of 419 *E. coli* strains and 20 antibiotics was included in our analyses after exclusion of antibiotics with low sample size (Figure 5.2a, Table S5.1). We identified 14 CS responses and 178 potential CR responses at an FDR of 0.05 with dichotomization at the median (Figure 5.3). The top five largest and smallest t-statistics, i.e. the most significant collateral responses, are summarized in Table 5.1, which show a very low p-value and FDR-adjusted p-value (q-value). The largest CS response is ertapenem on cefazolin, with a mean $\log_2$ FC of -1.95 (Figure 5.4a), which corresponds to a fold change of 0.26. The opposite direction (Figure 5.4b), the effect of cefazolin on ertapenem, also shows a significant CS response, but with a smaller effect size (-0.86 log_2 FC, $q < 0.05$). Cefazolin was associated with multiple CS responses with different antibiotics. The most significant CR response is that of meropenem on ertapenem with a mean $\log_2$ FC of 3.66 (Figure 5.4c), corresponding to a 12.6-fold change.

We chose the median to define the dichotomization criterion, but due to the discrete nature of the data, many values are equal to the median. Including the strains with a median MIC value in the low or high MIC group was not arbitrary, since it can change the group sizes substantially. The median strains were included in the small-

est group to make the sample sizes as equal as possible. For example, in Figure 5.4a, depicting the effect of ertapenem on cefazolin, the median of ertapenem was -1, and the strains with this median value were included in the B = low group to make the groups most equal in size [hence the high group contained $\log_2(\text{MIC}) \leq 0$]. This equal splitting was found to improve power (next section).

5.3.2 Power for identification of collateral effects

To understand which collateral effect sizes are detectable using our method, the power to detect collateral effects was calculated for different sample sizes, different group disbalances and different standard deviations, based on the values we found in the PATRIC data. The power to detect different effects greatly depends on the total sam-

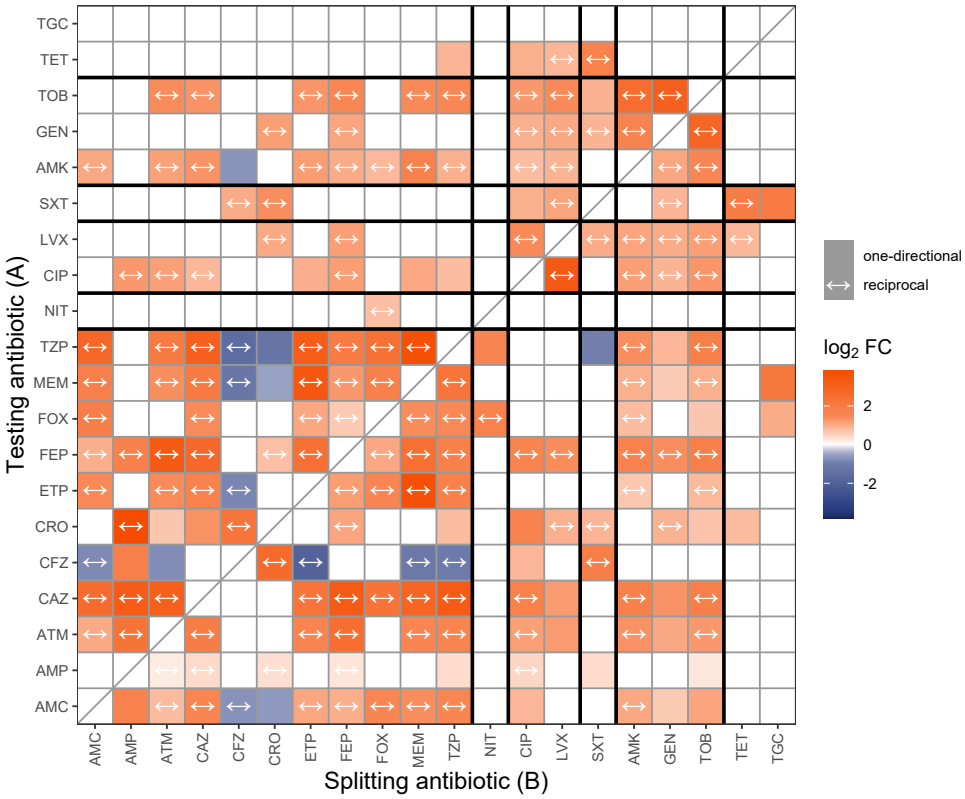


Figure 5.3: Heatmap of collateral responses identified based on clinical surveillance MIC data, for different antibiotic combinations. The effect size (blue/orange color) $\log_2$ FC is the mean $\log_2$(MIC) shift between the groups low and high MIC for splitting antibiotic B. On the y-axis the tested antibiotics that were split on antibiotic B (x-axis) and the bold lines separate the different antibiotic classes. A darker color indicates a larger effect, with orange showing CR and blue showing CS. White squares indicate no significant effect at the 0.05 FDR threshold. Reciprocality is denoted by a two-sided arrow.

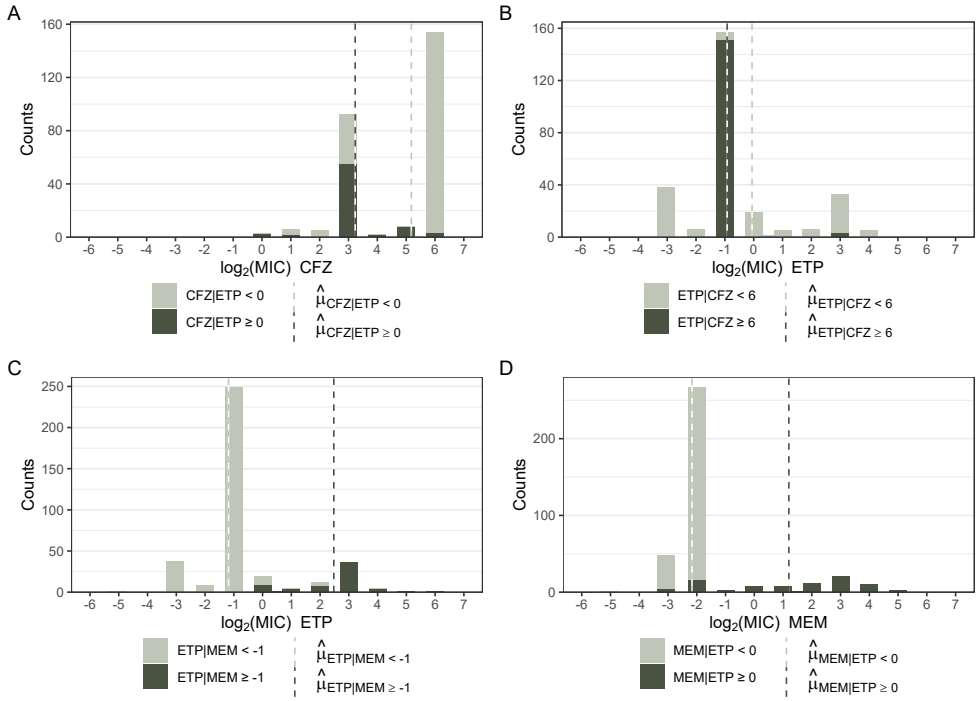


Figure 5.4: Log_2 transformed MIC [$\log_2(\text{MIC})$] distributions of the most significant CS effect (A) and the most significant CR effect (C). The frequency is shown on the y-axis for the different values of the $\log_2(\text{MIC})$. The different colors denote high (dark) or low (light) MIC to the splitting antibiotic. The dashed lines show the estimated means for the two groups. A) $\log_2(\text{MIC})$ distribution of cefazolin (CFZ) split by MIC of ertapenem (ETP) and B) the distribution of ertapenem under influence of cefazolin, which also has a significant CS response, showing a reciprocal effect. C) Log_2(MIC) distribution of ertapenem under influence of meropenem, which has a significant CR response (3.66 $\log_2$ FC) and D) the distribution of meropenem under influence of ertapenem, which also has a similarly sized CR response (3.37 $\log_2$ FC). This is a reciprocal collateral response.

ple size, with 2000 samples allowing the detection of a $\log_2$ FC of size 0.25 with power over 80%, while for smaller sample sizes, such as 200 total samples, the detectable effect size is 0.75 (Figure 5.5a). A larger disbalance of the two groups, high MIC and low MIC for antibiotic B, which is often seen in our data, does affect the power, but substantially only if the disbalance is greater than 25/75 (Figure 5.5b). Finally, the standard deviation of the MIC values has a large effect on how large the power to detect an effect is (Figure 5.5c), with the effect of the standard deviation on the power being inversely proportional to the effect of the effect size.

5.3.3 The effect of the dichotomization criterion

For estimating the collateral effects, we dichotomize the sample based on the MIC of one of the two antibiotics, at a chosen criterion (τ). We studied whether this choice influences the results and found that the value of τ does impact our results (Figure 5.6).

Table 5.1: Top 5 results of collateral sensitivity and resistance responses in the *E. coli* data, with lowest and highest value for the t-statistic respectively. The effect size is the difference in mean $\log_2$ fold change for the effect of splitting antibiotic B on testing antibiotic A. The number of observations in each group are $n_{B=\text{high}}$ and $n_{B=\text{low}}$ respectively.

Testing Antibiotic (A)	Splitting antibiotic (B)	T-statistic	p-value	$n_{B=\text{high}}$	$n_{B=\text{low}}$	Difference in mean ($\log_2$ FC)	q-value
Collateral sensitivity							
CFZ	ETP	-10.04	$2.43 \cdot 10^{20}$	69	201	-1.95	$2.41 \cdot 10^{18}$
TZP	CFZ	-6.13	$4.09 \cdot 10^{09}$	113	110	-1.61	$1.25 \cdot 10^{07}$
TZP	CRO	-5.58	$5.67 \cdot 10^{08}$	130	156	-1.32	$1.46 \cdot 10^{06}$
MEM	CFZ	-5.54	$6.68 \cdot 10^{08}$	154	140	-1.23	$1.69 \cdot 10^{06}$
CFZ	MEM	-5.13	$5.29 \cdot 10^{07}$	64	230	-1.16	$1.21 \cdot 10^{05}$
Collateral resistance							
ETP	MEM	26.73	$5.82 \cdot 10^{89}$	62	316	3.66	$1.44 \cdot 10^{85}$
MEM	ETP	23.85	$3.00 \cdot 10^{77}$	81	297	3.37	$3.71 \cdot 10^{74}$
TOB	GEN	21.99	$6.54 \cdot 10^{67}$	117	219	3	$5.40 \cdot 10^{64}$
CFZ	CRO	21.09	$6.61 \cdot 10^{60}$	137	147	2.6	$4.09 \cdot 10^{57}$
GEN	TOB	19.74	$5.05 \cdot 10^{58}$	146	190	2.82	$2.50 \cdot 10^{55}$

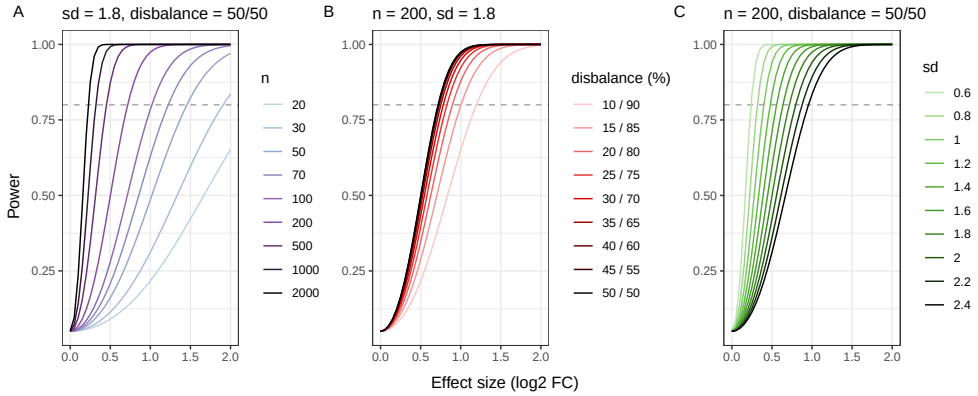


Figure 5.5: Properties of the collateral effects detection. Power analysis for t-test of collateral effects as the mean $\log_2$ FC. The power against effect size. A) Power for different sample sizes (n), lighter colors show lower n. Power increases greatly with n. B) Power for different disbalances in the group sample sizes. The most equal groups are divided 50/50, while 10/90 shows a large disbalance in sample sizes, which shows decreased power. C) The power for different standard deviations of the MIC distributions.

The group size equality (transparency of the lines) is a measure for how equal the two groups are, where an equality of 1 indicates a 50/50 split for the groups. The lines themselves show how the test statistics change over different values of τ, showing a dependence of collateral effect size and sometimes even direction (CR or CS) on MIC for some of the antibiotics.

The estimated collateral effect is very stable over dichotomization criteria for e.g. amoxicillin/clavulanic acid, piperacillin/tazobactam and tobramycin. Less stable are for example cefazolin and ciprofloxacin. For cefazolin however, the different effect sizes are found where the sample sizes of the group are very different, indicating that the effect is driven by a small part of the data. In case of ciprofloxacin, the change of estimated effect size over the different τ is not driven by a small part of the data.

5.4 discussion

Our analysis demonstrates that empirical determination of collateral effects is possible from MIC population surveillance data, by quantifying shifts in conditional MIC distributions. Importantly, our approach enables detection of collateral responses including directionality and effect size. We demonstrated the utility of our method to a set of *E. coli* MIC data, identifying CS and potential CR responses from available clinical surveillance data.

Collateral effects are experimentally established to show directionality, that is, the effect size between two antibiotics in terms of their collateral effects is not symmetrical. Thus, a collateral effect of antibiotic A on antibiotic B can differ from the effect of B on antibiotic A. Our method can be employed to detect both one-directional and two-directional (reciprocal) collateral responses. This is in contrast to other statistical methods, such as the odds ratio used in a previous study (Obolski et al., 2016) or a correlation, where both directions yield the same statistic. A correlation will thus either identify two-directional responses or not detect a collateral effect at all, which increases both the number of false discoveries and the number of false rejections, as compared with our method.

The collateral effect metric proposed, using fold change, has a clear interpretation. The effect size is the mean difference in $\log_2(\text{MIC})$s, or $\log_2$ FC, between a group with high MIC for an antibiotic B as compared with the group with a low MIC for the same antibiotic. A fold change value enables interpretation of the clinical relevance of detected MIC changes. This is of relevance, since, especially with large sample sizes, statistical significance does not always imply clinical relevance (Kieser et al., 2012). In addition, the fold change measure is comparable to the experimentally determined fold change measures in experimental evolution studies that determine collateral effects (Imamovic & Sommer, 2013; Imamovic et al., 2018; Maltas & Wood, 2019; Barbosa et al., 2017).

The choice of dichotomization criterion is not arbitrary, since it can affect the effect size estimation, such as for ciprofloxacin (Figure 5.6). The type of the identified collateral effects may even vary depending on the selected dichotomization criterion,

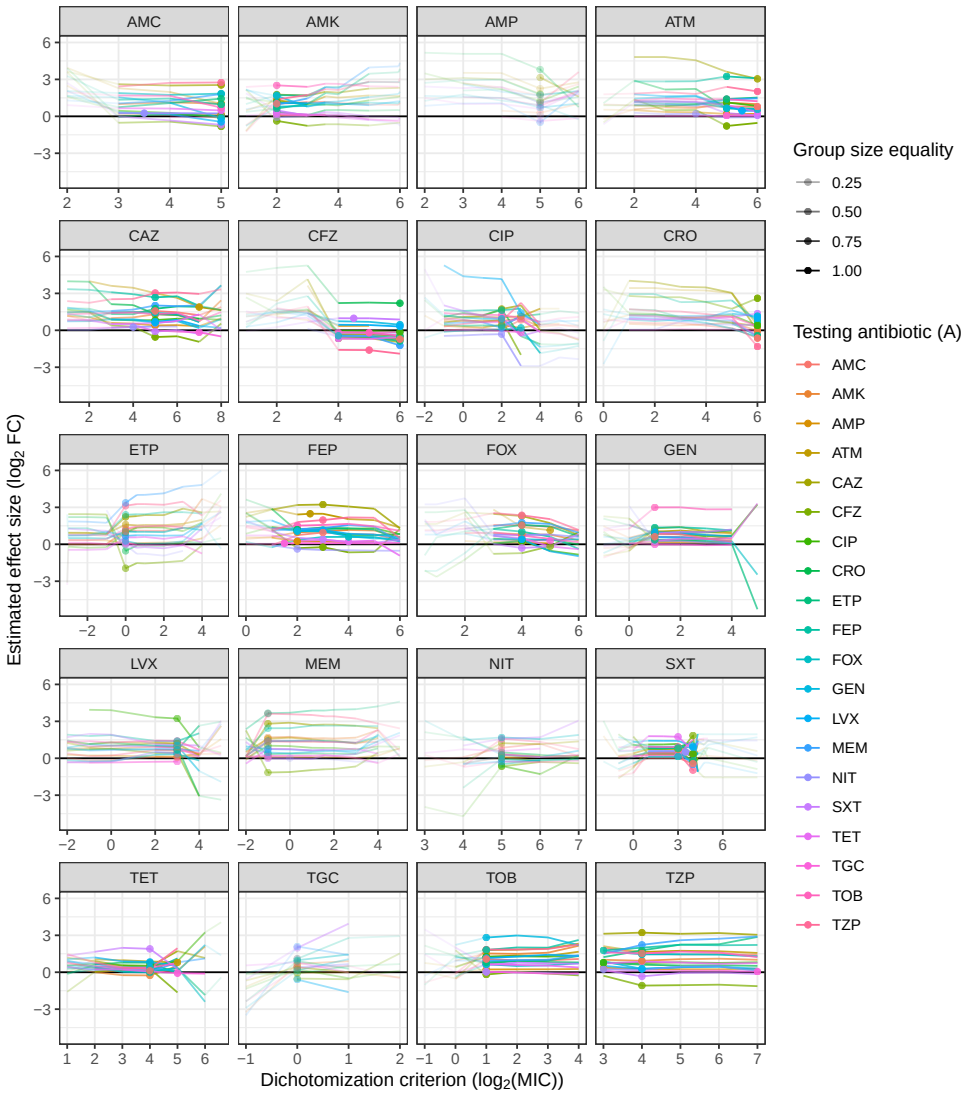


Figure 5.6: The effect of changing the dichotomization criterion (x-axis) on the estimated $\log_2$ FC (y-axis). Every panel denotes the antibiotic that is dichotomized (antibiotic B). The colors show the different antibiotics that were tested based on this dichotomization (antibiotic A). The dots denote the value of the $\log_2$FC found in the main analysis, based on equal split dichotomization. A more transparent line shows a larger inequality in group sizes, where the effect estimation is driven by only a small part of the total sample. Straight horizontal lines denote stable effect size estimation over different dichotomization criteria, while crooked lines show unstable estimation.

which is, for example, the case for ceftriaxone (A) on ciprofloxacin (B). For the choice of dichotomization criterion, three aspects need to be considered. Firstly, the data limit the possible choices of τ. There need to be enough data at both sides of the split to be able to use this test. Secondly, we showed that the power for detecting effects

is highest when the two groups are similar in size. To this end, a dichotomization on the equal group sizes yields the highest power and we chose this option for the study. Thirdly, the change of the estimated effect sizes over different dichotomization criteria indicates a dependence between collateral effect and MIC value, which could have an underlying biological cause. The choice of τ could therefore also be decided based on which MIC value you want to know about collateral effects. For instance, in Figure 5.4a a CS response is shown in strains with a $\log_2$ MIC for cefoxitin larger or equal to 0, indicating that the decrease in sensitivity for cefoxitin is correlated with an increase in sensitivity for cefazolin, above this threshold. Whether the same collateral effect is also detected at other dichotomization criteria can be tested by changing this value.

For our proof-of-concept analysis of our *E. coli* dataset we identified 192 collateral responses using the equal group splitting for the dichotomization, of which 14 were therapeutically interesting CS responses. Suggestive CR responses were extensively more prevalent than CS responses. This was in line with expectations, given the presence of antibiotics from the same class for which CR is likely to occur. For example, reciprocal CR responses were detected between the aminoglycosides amikacin, gentamicin and tobramycin, and the fluoroquinolones ciprofloxacin and levofloxacin. In accordance with previous studies showing CS between β-lactams (Rosenkilde et al., 2019), we identified reciprocal CS effects between cefazolin and the β-lactams meropenem and ertapenem, as well as between cefazolin and the β-lactam/β-lactamase inhibitor combinations amoxicillin/clavulanic acid and piperacillin/tazobactam. In contrast to the CS repeatedly found in the literature between ciprofloxacin and gentamycin (Imamovic et al., 2018; Barbosa et al., 2017; Podnecky et al., 2018; Liakopoulos et al., 2022), our analysis showed a CR response for this antibiotic combination.

The MIC data deposited in PATRIC originate from a variety of studies and MIC determination methods, which leads to large variation and possible error in MIC values. Therefore, the data were used as proof of concept. Large datasets that include MICs determined in a more consistent manner may be able to further improve the performance of our method. In larger datasets, our method could also be used to identify multidrug effects, by splitting the high and low groups conditioned on more than one antibiotic.

This statistical method does not set out to identify causal relationships, so while use of an antibiotic may lead to resistance development, it is not possible to show that the CS or resistance to another antibiotic has been caused by the use of the first antibiotic. Also, due to the observational nature of the data, we are unable to discriminate between CR responses and the occurrence of an MDR phenotype due to the co-presence of distinct resistance mechanisms to each of the individual antibiotic agents.

The occurrence of CS has been previously suggested as a phenomenon that can be utilized to design dosing schedules that prevent emergence of antimicrobial resistance and prolong the efficacy of the existing antimicrobial agents (Imamovic & Sommer, 2013), but so far it has been mostly studied in reference laboratory strains

of a limited number of bacterial species. Although our study suggests the occurrence of CS in clinical populations, it remains unclear to what extent these effects actually occur in clinically occurring pathogens. The developed methodology can be directly applied to clinical datasets of antimicrobial susceptibility that are widely available via national and international surveillance programmes, to both estimate the effect sizes and occurrence of collateral effects and to provide further insight into the clinical relevance of CS effects. These effects can be evaluated for specific antibiotic combinations and pathogen species, which may guide the design of CS-based dosing strategies of high clinical relevance and the selection of empirical treatment (Aulin et al., 2021). In addition, the quantification of CR in antimicrobial susceptibility surveillance datasets is of interest to identify antibiotic combinations that should be avoided as these could potentially lead to increased risk of treatment failure and the spread of antimicrobial resistance.

We conclude that the proposed methodology is relevant for identification of collateral responses based on clinical surveillance data. We implemented the functions for the method in the R package collatRal to make the method accessible to other researchers. Our method can be applied to larger surveillance datasets that also include MIC data for additional antibiotics and for other clinically relevant bacterial species. Identified collateral effects, and in particular CS, can provide important guidance for combination therapy and in the further design of CS-based dosing strategies that aim to suppress antibiotic resistance.

References

Amsalu, A., Sapula, S. A., Lopes, M. D. B., Hart, B. J., Nguyen, A. H., Drigo, B., ... Venter, H. (2020, Oct). Efflux pump-driven antibiotic and biocide cross-resistance in pseudomonas aeruginosa isolated from different ecological niches: A case study in the development of multidrug resistance in environmental hotspots. *Microorganisms*, *8*(11), 1647. Retrieved from https://doi.org/10.3390%2Fmicroorganisms8111647 doi: 10.3390/microorganisms8111647

Aulin, L. B. S., Koumans, C. I. M., Haakman, Y., van Os, W., Kraakman, M. E. M., Gooskens, J., ... van Hasselt, J. G. C. (2020, Nov). Distinct evolution of colistin resistance associated with experimental resistance evolution models in klebsiella pneumoniae. *Journal of Antimicrobial Chemotherapy*, *76*(2), 533–535. Retrieved from https://doi.org/10.1093%2Fjac%2Fdkaa450 doi: 10.1093/jac/dkaa450

Aulin, L. B. S., Liakopoulos, A., van der Graaf, P. H., Rozen, D. E., & van Hasselt, J. G. C. (2021, Sep). Design principles of collateral sensitivity-based dosing strategies. *Nature Communications*, *12*(1). Retrieved from https://doi.org/10.1038%2Fs41467-021-25927-3 doi: 10.1038/s41467-021-25927-3

Barbosa, C., Trebosc, V., Kemmer, C., Rosenstiel, P., Beardmore, R., Schulenburg, H., & Jansen, G. (2017, May). Alternative evolutionary paths to bacterial antibiotic resistance cause distinct collateral effects. *Molecular Biology and Evolution*, *34*(9), 2229–2244. Retrieved from https://doi.org/10.1093%2Fmolbev%2Fmsx158 doi: 10.1093/molbev/msx158

Baym, M., Stone, L. K., & Kishony, R. (2016, Jan). Multidrug evolutionary strategies to reverse antibiotic resistance. *Science*, *351*(6268). Retrieved from https://doi.org/10.1126%2Fscience.aad3292 doi: 10.1126/science.aad3292

Benjamini, Y., & Yekutieli, D. (2001, Aug). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, *29*(4). Retrieved from https://doi.org/10.1214%2Faos%2F1013699998 doi: 10.1214/aos/1013699998

Eliopoulos, G. M., & Moellering, R. C. (1982, Mar). Antibiotic synergism and antimicrobial combinations in clinical infections george. *Clinical Infectious Diseases*, *4*(2), 282–293. Retrieved from https://doi.org/10.1093%2Fclinids%2F4.2.282 doi: 10.1093/clinids/4.2.282

Imamovic, L., Ellabaan, M. M. H., Machado, A. M. D., Citterio, L., Wulff, T., Molin, S., ... Sommer, M. O. A. (2018, Jan). Drug-driven phenotypic convergence supports rational treatment strategies of chronic

infections. *Cell*, *172*(1-2), 121–134.e14. Retrieved from https://doi.org/10.1016%2Fj.cell.2017.12.012 doi: 10.1016/j.cell.2017.12.012

Imamovic, L., & Sommer, M. O. A. (2013, Sep). Use of collateral sensitivity networks to design drug cycling protocols that avoid resistance development. *Science Translational Medicine*, *5*(204). Retrieved from https://doi.org/10.1126%2Fscitranslmed.3006609 doi: 10.1126/scitranslmed.3006609

Kieser, M., Friede, T., & Gondan, M. (2012, Sep). Assessment of statistical significance and clinical relevance. *Statistics in Medicine*, *32*(10), 1707–1719. Retrieved from https://doi.org/10.1002%2Fsim.5634 doi: 10.1002/sim.5634

Liakopoulos, A., Aulin, L. B. S., Buffoni, M., Fragkiskou, E., Coen van Hasselt, J. G., & Rozen, D. E. (2022, May). Allele-specific collateral and fitness effects determine the dynamics of fluoroquinolone resistance evolution. *Proceedings of the National Academy of Sciences*, *119*(18), 1–12. Retrieved from https://pnas.org/doi/full/10.1073/pnas.2121768119 doi: 10.1073/pnas.2121768119

Maltas, J., & Wood, K. B. (2019, Oct). Pervasive and diverse collateral sensitivity profiles inform optimal strategies to limit antibiotic resistance. *PLOS Biology*, *17*(10), e3000515. Retrieved from https://doi.org/10.1371%2Fjournal.pbio.3000515 doi: 10.1371/journal.pbio.3000515

Obolski, U., Dellus-Gur, E., Stein, G. Y., & Hadany, L. (2016, Jun). Antibiotic cross-resistance in the lab and resistance co-occurrence in the clinic: Discrepancies and implications in e. coli. *Infection, Genetics and Evolution*, *40*, 155–161. Retrieved from https://doi.org/10.1016%2Fj.meegid.2016.02.017 doi: 10.1016/j.meegid.2016.02.017

Pál, C., Papp, B., & Lázár, V. (2015, Jul). Collateral sensitivity of antibiotic-resistant microbes. *Trends in Microbiology*, *23*(7), 401–407. Retrieved from https://doi.org/10.1016%2Fj.tim.2015.02.009 doi: 10.1016/j.tim.2015.02.009

Podnecky, N. L., Fredheim, E. G. A., Kloos, J., Sørum, V., Primicerio, R., Roberts, A. P., ... Johnsen, P. J. (2018, Sep). Conserved collateral antibiotic susceptibility networks in diverse clinical strains of escherichia coli. *Nature Communications*, *9*(1). Retrieved from https://doi.org/10.1038%2Fs41467-018-06143-y doi: 10.1038/s41467-018-06143-y

R Core Team. (2020). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.r-project.org/

Rosenkilde, C. E. H., Munck, C., Porse, A., Linkevicius, M., Andersson, D. I., & Sommer, M. O. A. (2019, Feb). Collateral sensitivity constrains resistance evolution of the CTX-m-15 beta-lactamase. *Nature Communications*, *10*(1). Retrieved from https://doi.org/10.1038%2Fs41467-019-08529-y doi: 10.1038/s41467-019-08529-y

Tamma, P. D., Cosgrove, S. E., & Maragakis, L. L. (2012, Jul). Combination therapy for treatment of infections with gram-negative bacteria. *Clinical Microbiology Reviews*, *25*(3), 450–470. Retrieved from https://doi.org/10.1128%2Fcmr.05041-11 doi: 10.1128/cmr.05041-11

Turnidge, J., Kahlmeter, G., & Kronvall, G. (2006, May). Statistical characterisation of bacterial wild-type MIC value distributions and the determination of epidemiological cut-off values. *Clinical Microbiology and Infection*, *12*(5), 418–425. Retrieved from https://doi.org/10.1111%2Fj.1469-0691.2006.01377.x doi: 10.1111/j.1469-0691.2006.01377.x

Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., ... Stevens, R. L. (2016, Nov). Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Research*, *45*(D1), D535–D542. Retrieved from https://doi.org/10.1093%2Fnar%2Fgkw1017 doi: 10.1093/nar/gkw1017

Wickham, H. (2008). *Elegant Graphics for Data Analysis: ggplot2*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

World Health Organization. (2014). *Antimicrobial resistance: global report on surveillance* (Tech. Rep.). Retrieved from https://apps.who.int/iris/bitstream/handle/10665/193736/9789241509763_eng.pdf

Wright, G. D. (2007, Mar). The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature Reviews Microbiology*, *5*(3), 175–186. Retrieved from https://doi.org/10.1038%2Fnrmicro1614 doi: 10.1038/nrmicro1614

Acknowledgement

We would like to thank Parth J. Upadhyay for critically reviewing the R code of the data analysis and the collatRal package.

Supplementary material

Table S5.1: Antibiotics and their abbreviations used in analysis of *E. coli* strains, and the number of strains used in the final analysis.

Antibiotic	Abbreviation	Number of strains	Class	Target(s)
Amoxicillin/clavulanic acid	AMC	256	Beta-lactams	Cell wall synthesis
Amikacin	AMK	385	Aminoglycosides	Protein synthesis, 30S
Ampicillin	AMP	405	Beta-lactams	Cell wall synthesis
Aztreonam	ATM	250	Beta-lactams	Cell wall synthesis
Ceftazidime	CAZ	255	Beta-lactams	Cell wall synthesis
Cefazolin	CFZ	296	Beta-lactams	Cell wall synthesis
Ciprofloxacin	CIP	412	Fluoroquinolones	DNA replication
Ceftriaxone	CRO	398	Beta-lactams	Cell wall synthesis
Ertapenem	ETP	381	Beta-lactams	Cell wall synthesis
Cefepime	FEP	384	Beta-lactams	Cell wall synthesis
Cefoxitin	FOX	233	Beta-lactams	Cell wall synthesis
Gentamicin	GEN	410	Aminoglycosides	Protein synthesis, 30S
Levofloxacin	LVX	284	Fluoroquinolones	DNA replication
Meropenem	MEM	412	Beta-lactams	Cell wall synthesis
Nitrofurantoin	NIT	210	Nitrofuran	Citric acid cycle; DNA, RNA, and protein synthesis
Trimethoprim/sulfamethoxazole	SXT	313	Antifolate	Folic acid synthesis
Tetracycline	TET	248	Tetracyclines	Protein synthesis, 30S
Tigecycline	TGC	206	Tetracyclines	Protein synthesis, 30S
Tobramycin	TOB	338	Aminoglycosides	Protein synthesis, 30S
Piperacillin/tazobactam	TZP	291	Beta-lactams	Cell wall synthesis

Chapter 6

Inference of collateral sensitivity effects in large scale antimicrobial resistance surveillance data

Authors

Laura B. Zwep

Gabriel Forn Cuni

Apostolos Liakopoulos

Annelot F. Schoffelen

Stephan Beisken

Jacqueline Meulman

J. G. Coen van Hasselt

Abstract

Collateral sensitivity (CS) occurs when bacteria develop resistance to one antibiotic leading to an increase in antibiotic sensitivity towards a second antibiotic. The phenomenon of CS has been proposed to be of interest for design of antibiotic treatment strategies to prevent emergence of resistance. Limited knowledge is available concerning the occurrence of CS across species and strains. Here, we aimed to comprehensively characterize the occurrence of CS in large scale population surveillance data of antimicrobial resistance for multiple species, strains, and antibiotics.

We combined multiple databases with minimal inhibitory concentration (MIC) values for >2.7 million MIC measurements for 20 bacterial species and 94 antibiotics. We applied a novel methodology to infer CS effect size. The similarity of CS effects between species and patterns over different antibiotic classes and for higher order collateral effects were explored.

A large set of CS effects was identified, with CS effects identified in the majority of species studied. We find that most antibiotic-specific CS effects are species-specific. CS effects were more commonly observed between different antibiotic classes, than within. Finally we find that specific combinations of resistance towards to antibiotics can be associated with specific CS effect towards a third antibiotic. All identified CS relationships have been incorporated in a web-application to allow in-depth exploration of our findings.

Our systematic analysis highlights the occurrence of CS across a large panel of strains. These results can help prioritize the selection of CS-based antibiotic treatment strategies for further experimental and clinical studies investigating antibiotic combination strategies to suppress resistance.

6.1 Introduction

Antimicrobial resistance (AMR) is an increasing global health threat, rendering commonly used antibiotics increasingly infective to treat bacterial infections. To this end, strategies which can reduce the risk for emergence of AMR are urgently needed. The occurrence of collateral sensitivity (CS) of AMR has been proposed as one strategy to support this goal (Maltas & Wood, 2019; Pál et al., 2015). CS occurs when bacteria develop resistance to one antibiotic which leads as a collateral effect to increased antibiotic sensitivity towards a second antibiotic. The phenomenon of CS may be used to design combinatory antibiotics strategies which can reduce the risk of developing AMR during treatment (Baym et al., 2016).

The potential clinical application of CS-based treatments is currently limited because of insufficient knowledge regarding its occurrence across bacterial species and strains. CS has so far mainly been studied in well controlled in vitro experiments, mainly conducted using a limited selection of bacterial species, strains, and antibiotics (Imamovic et al., 2018; Lázár et al., 2018; Liakopoulos et al., 2022). The majority of experimental studies have focused on a limited number of laboratory strains. Little

is known to what extent CS effects for specific antibiotics or their classes may be conserved or vary between different bacterial species but also how CS effects may be conserved across strains for specific species, and which of these CS are reciprocal. In addition, so far, studies have focused on situations where only the collateral effects for a single type of resistance. In clinical reality however, strains may develop resistance against multiple antibiotics. It is unclear if such higher-order resistance patterns affect the observed collateral responses towards other antibiotics.

Population surveillance of antibiotic susceptibility and resistance in clinically isolated pathogens represents a crucially important tool to monitor AMR in health care across countries, and such data is increasingly available for research purpose (van der Kuil et al., 2017; Johnson, 2015) Typically, such surveillance data consists of minimal inhibitory concentrations (MICs) for different antibiotics. Recently, we developed a method to infer CS effect sizes from individual-level MIC data, and is very suitable to profile for CS effects in MIC-based antimicrobial population surveillance data (Zwep et al., 2021).

In this study we apply our method to infer CS from large scale MIC datasets obtained through population surveillance and other public resources to provide a comprehensive overview of the occurrence of CS affects for clinically used antibiotics and commonly occurring bacterial strains and species for pairwise as well as higher-order CS effects.

6.2 Methods

6.2.1 Data resources

Individual-level MIC data for bacterial pathogens were acquired from four large resources: (i) publicly available MIC data deposited in National Center for Biotechnology Information (NCBI) (Sayers et al., 2020) and (ii) the Pathosystems Resource Integration Center (PATRIC) (Davis et al., 2019), (iii) proprietary MIC data from ARESdb data provided by the company OpGen (OpGen, 2022), and (iv) population surveillance data from the Dutch National Institute for Public Health and the Environment (RIVM) (van der Kuil et al., 2017). All four sources included minimal inhibitory concentrations (MICs) for different bacterial strains, over a wide range of antibiotics and antimicrobial resistance profiles.

6.2.2 Data preparation

The NCBI and PATRIC resources contained overlapping data and were merged to prevent duplicate inclusion of strains. Antibiotics with less than 100 observations for RIVM and NIH & PATRIC were removed, as well as strains with MICs for less than two antibiotics. The ARESdb data were complete, so no strains or antibiotics were removed. The data from the four sources were merged into a single data set, which was used in further analysis.

6.2.3 Collateral effect quantification and testing

All pairs of two antibiotics (antibiotic pairs) were tested for collateral effects using a previously described method developed by our group which quantifies empirical fold change ($\log_2$ FC) and tests the statistical significance of this outcome (Zwep et al., 2021). This method uses MICs to calculate the collateral effect by dichotomizing the MIC data on one antibiotic (B) to a group with high resistance B_r and a group with low resistance or sensitivity $\neg B_r$, and testing the difference between the means of the MIC distributions of a second antibiotic (A), such that

$$\log_2 \text{ FC}_{A|B_r} = \overline{\log_2 \text{ (MIC}_{A|B_r})} - \overline{\log_2 \text{ (MIC}_{A|\neg B_r})}$$

where B is the dichotomizing antibiotic and A is the antibiotic of which the MIC is evaluated. The bar denotes the sample mean. The following hypotheses were tested:

$$H_0 : \mu_{A|B_r} = \mu_{A|\neg B_r}$$

$$H_{CS} : \mu_{A|B_r} < \mu_{A|\neg B_r}$$

$$H_{CR} : \mu_{A|B_r} > \mu_{A|\neg B_r}$$

where $\mu_{A|B_r} = \overline{\log_2 \text{ (MIC}_{A|B_r})}$.

All combinations of antibiotics were tested in two directions, i.e., the effect of an antibiotic A on the antibiotic B and the effect of antibiotic B on antibiotic A. The dichotomization based on the first antibiotic was done so both groups were as close to equal sizes, by choosing the median as dichotomization criterion.

6.2.4 Multiple testing correction

After quantification and testing the collateral effects, the p-values were adjusted for multiple testing using the Benjamini-Yekutieli correction for false discovery rate (Benjamini & Yekutieli, 2001). In the context of collateral sensitivity, an important factor to take into account is the effect size. A small significant collateral effect, might be clinically irrelevant, so clinical relevance was also taken into account when deciding on which collateral effects are of interest. Clinical relevance was defined here by two factors: the size of the estimated fold change and the equality of the group sizes. Small effect sizes ($-0.5 < \log 2\text{FC} < 0.5$) were excluded and effects were only considered relevant if both groups consisted of at least 5% of the total number of strains. Lastly, only antibiotic pairs with more than 100 complete pairwise observations were evaluated (Figure 6.1).

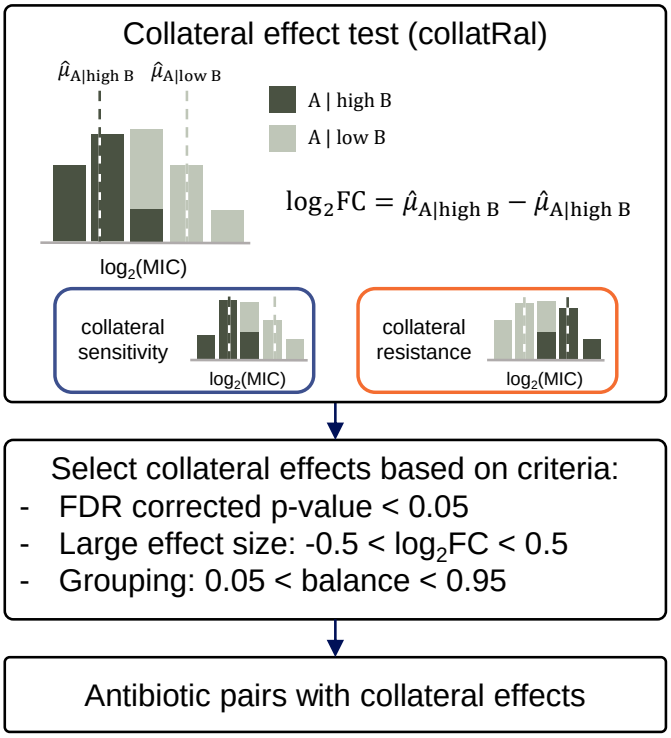


Figure 6.1: Data analysis workflow. First, a collateral effect test was used for each pair of antibiotics (antibiotic A and antibiotic B), within each species. The difference between the means of the $\log_2$-transformed MIC distributions is the estimated $\log_2$ FC. Next, results were filtered based on three criteria, for statistical significance and clinical relevance. Finally resulting in a set of antibiotic pairs with collateral effect.

6.2.5 Collateral effect networks

To evaluate all found collateral effects, a shiny dashboard was developed to explore subsets of species and antibiotics with network mapping of all collateral effect. The antibiotics were shown as nodes and the antibiotic effects as arrows, indicating the directionality, with options to adjust the thresholds for the clinical relevance: the minimal absolute effect size and the minimal balance.

6.2.6 Similarity between species and antibiotics

To evaluate the similarity in CS effects between the different species, the Euclidean distances between the CS responses were mapped using a proximity mapping program (Heiser et al., 2020). The Euclidean distances were calculated between each pair of species, based on the collateral effect size. Only the antibiotic pairs where at least one of the two species showed a negative collateral effect were considered, to remove the influence of collateral resistance effects. Proximity mapping was used to project the distances in two dimensions. We also explored which CS effects were

prevalent over multiple species.

We evaluated if collateral effects between certain antibiotic classes were more prevalent, by comparing CS between antibiotic pairs with different antibiotic classes. Antibiotics within the same antibiotic class were expected to appear less often, than between different antibiotic classes. This was evaluated both over all species and within each species.

6.2.7 Three-way CS interactions

To assess more complex interactions, we analyzed a three-way collateral effects. The strains were dichotomized based on a group with resistance against two different antibiotics, a dual resistance, and a group with sensitivity to at least one of the antibiotics. For each set of three antibiotics (A, B and C), the MICs of antibiotic A were compared between the group with resistance against both antibiotic B and antibiotic C (B_r&C_r) and the group without dual resistance ($\neg(B_r$&$C_r)$). The collateral effects were calculated in the same way as for the pairwise collateral effects, by comparing the mean $\log_2$ (MIC between the two groups and testing the hypotheses

$$H_0 : \mu_{A|B_r\&C_r} = \mu_{A|\neg(B_r\&C_r)})$$

$$H_{CS} : \mu_{A|B_r\&C_r} < \mu_{A|\neg(B_r\&C_r)})$$

We compared the results of these three-way collateral effects $\log_2$ FC$_{A|B_r\&C_r}$ to the two corresponding pairwise collateral effects ($\log_2$ FC$_{A|B_r}$ and $\log_2$ FC$_{A|C_r}$) to assess the interaction effect of B&C on A.

6.3 Results

6.3.1 Database characteristics

A pooled dataset of individual MIC data derived from multiple databases consisting of up to 2.7 million measurements for 20 bacterial species and 94 antibiotics, was used for the final analysis (Figure 6.2A. A total of 2066 unique antibiotic pairs was tested, where some were tested in multiple species and almost all were tested in two directions. The species with the largest number of strains (n) and antibiotics (d) are *E. coli* ($n = 2,740,266$, $d = 63$), *K. pneumoniae* ($n = 435,987$, $d = 54$), P. aeruginosa ($n = 403,095$, $d = 42$) and S. aureus ($n = 934,659$, $d = 49$). Not all strains contain measurements for every antibiotic. Within a species, each antibiotic contained a discreet distribution of MICs, such as for example for *C. coli* (Figure 6.2B). All used abbreviations for antibiotics can be found in Table S6.1.

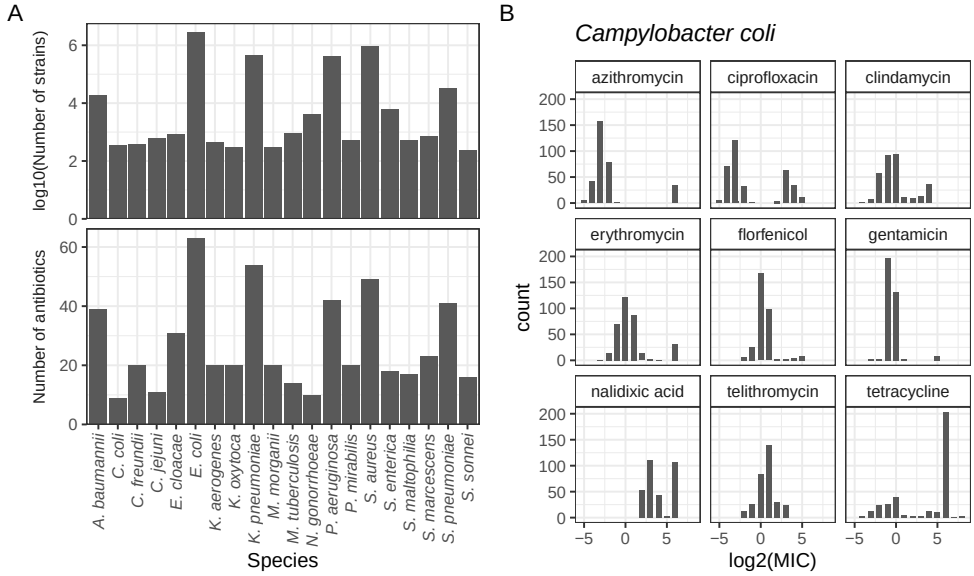


Figure 6.2: Description of the MIC data used for the estimation of collateral effects. A) The number of strains per species (top) and the number of unique antibiotics per species (bottom). B) Example of the MIC data as log transformed MIC distributions for the 9 different antibiotics available for *Campylobacter coli*.

6.3.2 Collateral sensitivity quantification

After testing all combinations, a total of 385 antibiotic pairs with collateral sensitivity effects was detected over the different species. Of these, 120 showed reciprocal CS (Figure 6.3). Most of the non-reciprocal effects showed negative effect sizes, but did not meet the threshold of minimal effect size (< 0.5), balance (< 0.1) or statistical significance (q-value < 0.05) (Figure 6.3A). CS was found for 15 out of 20 studied species. No CS effects were not found in *K. oxytoca*, *M. morganii*, *P. mirabilis*, *S. enterica* and *S. maltophilia* (Figure 6.3B). Overall, CS effects were found across all studied antibiotic classes, and for the majority of antibiotics (Figure 6.3C). The collateral effect networks for each antibiotic pair and species can be explored in the developed Shiny dashboard at `collateralviz.lacdr.leidenuniv.nl` (Figure 6.4).

6.3.3 Differences and similarities in CS between species

With a large number of species and antibiotic pairs tested, we compared how similar the CS effects were between the different species across all antibiotic pairs. Because of the large number of collateral effects tested, the differences between the collateral effects were used to map the species in a two-dimensional plain, using proximity mapping (Figure 6.5). Some species were very dissimilar such as *S. pneumonia* and *K. oxycota*, which are at opposite sides of the map.

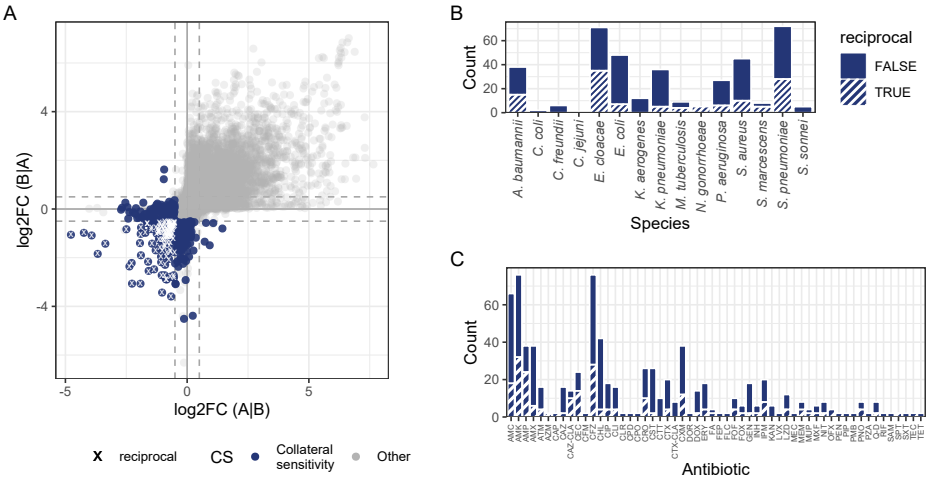


Figure 6.3: Overview of estimated collateral effects and reciprocality across species. A) Estimated effect sizes ($\log_2$ fold change (FC) of the MIC) of reciprocal and non-reciprocal collateral sensitivity responses and other responses, no collateral effect and collateral resistance. Antibiotic A and B represent an antibiotic pair, for which the effect size in both directions is estimated, A given dichotomization on B $(A|B)$ and, B given dichotomization on A $(B|A)$. Dashed lines denote the threshold of relevant effect sizes (abs($\log_2$ FC $= 0.5$). B) Number of collateral sensitivity effects for each species and whether they are reciprocal (striped) or not (blue). C) Number of collateral sensitivity effects for each antibiotic. Note that every collateral sensitivity effect contributes to two antibiotics.

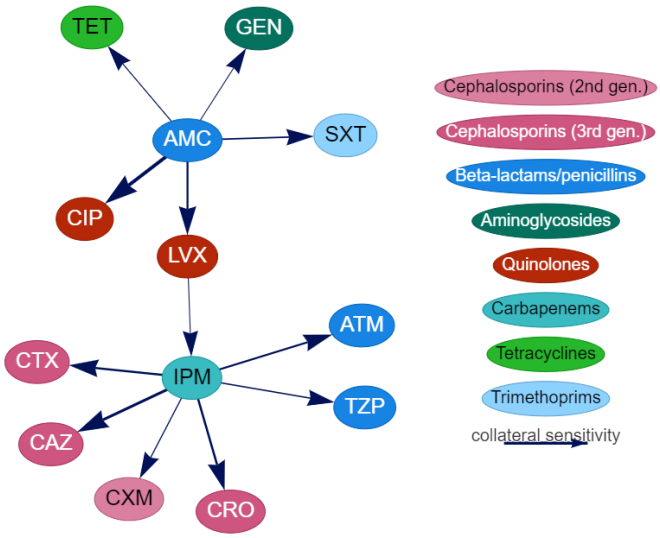


Figure 6.4: Collateral sensitivity network, example for *Klebsiella aerogenes*. Arrows denote the CS effect and direction, colors denote the different antibiotic classes (from Shiny dashboard).

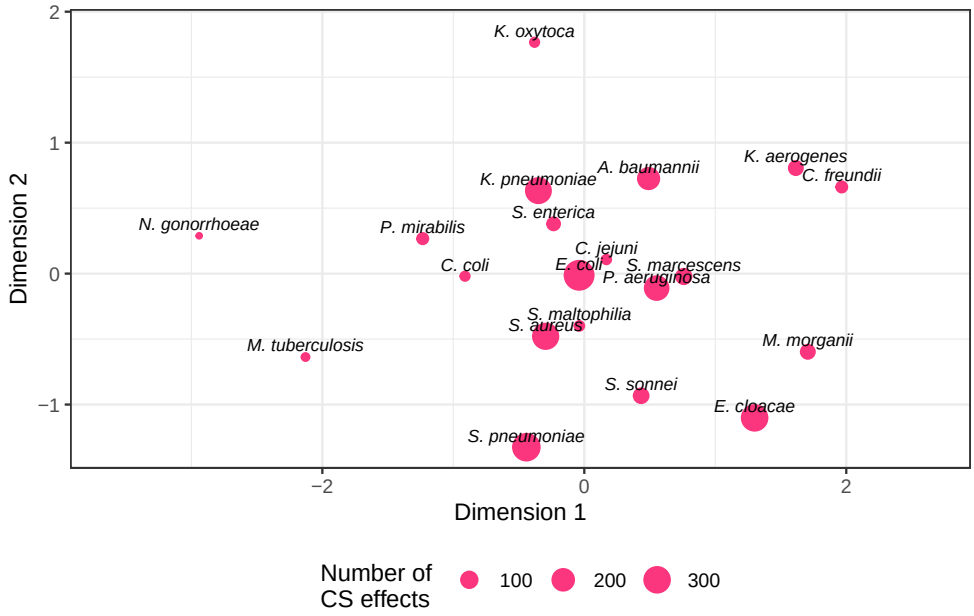


Figure 6.5: Similarity in CS responses between different species. Proximity mapping of different species based on collateral sensitivity effect estimates. Species with similar CS estimates are mapped in close proximity, while species with very different collateral sensitivity are mapped far apart. The size of the nodes shows on how many CS effects the distance was based.

When we focus on similarities in collateral effects at the level of specific antibiotic pairs, we find that certain pairs are more commonly associated with CS effects (Figure 6.6). In total, 22 antibiotic pairs were found for which consistent CS was identified over at least three species. In some pairs we only detected CS (e.g., MEM on LZD), while for most pairs the type of the collateral effect varied depending on the species. Interestingly, all these more consistent CS effects were found between different antibiotic classes.

6.3.4 Antibiotic classes with more CS effects

We explored how CS effects were distributed within- and between antibiotic classes. For the CS responses that were detected in multiple species (Figure 6.6), cephalosporins (CFZ, CXM, CTX and CRO) mostly showed CS with aminoglycosides (AMK, TOB and GEN) and carbapenems (MEM and IMP), and penicillins (AMP, AMX) had more CS with colistin (CST) and the class 'other antibacterials' (SXT, CHL). There were no CS effects consistent over at least three species (Figure 6.6) within the same antibiotic class, although there were CS effects between the different classes of beta-lactams: carbapenems, cephalosporins and penicillins.

Overall, across all species, no distinct patterns of CS were seen (Figure S6.1). When

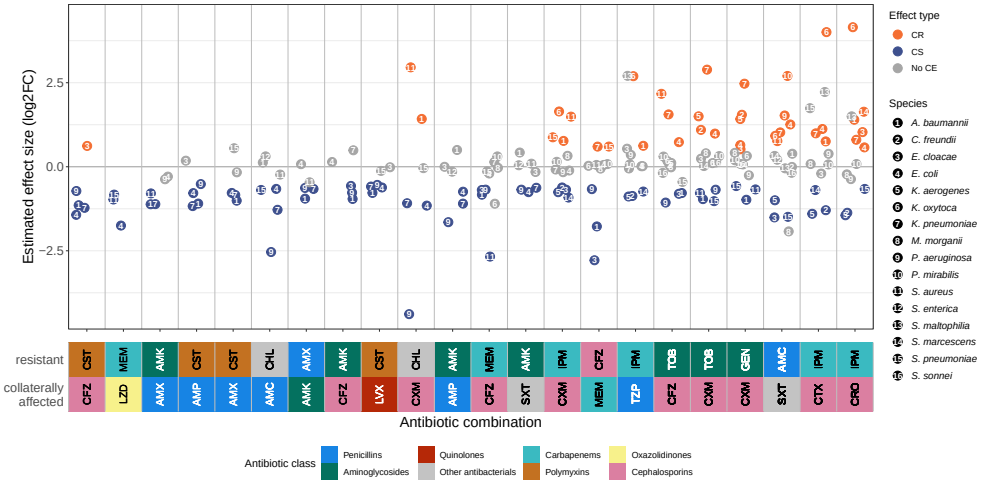


Figure 6.6: Effect sizes of antibiotic pairs with consistent collateral sensitivity (CS) effects over different species. Collateral sensitivity was considered 'consistent' if it was detected in at least three species. The x-axis label colors denote the different antibiotic classes between which these consistent CS effects were detected.

looking at differences between and within antibiotic classes, as expected, CS within the same class was either not found at all (quinolones, tetracyclines, aminoglycosides, glycopeptides) or found in a lower rate than between-class CS (penicillins, macrolides) (Figure 6.7, Figure S6.2). For cephalosporins the proportion of CS between- and within-classes was similar. Specific inspection of this class indicates that such within-class CS occurred mostly between different generations of cephalosporin antibiotics.

Within each species, different patterns of CS could be seen. For example in *S. pneumonia*, most CS effects are from the cephalosporins to other classes, such as the quinolones (Figure 6.8). There are also many CS effects between penicillins and other antibiotic classes.

6.3.5 Three-way interaction CS effects

To evaluate the effect of three-way interaction effects of antibiotics B&C on antibiotic A were estimated. In general, a large number of such CS effects was found, but the proportion of CS effects found did not differ between the antibiotic pairs and the sets of three antibiotics, with dual resistances (Figure S6.3). As expected, we did find an increased proportion of collateral resistance effects for most species, associated with the occurrence of multidrug resistance. Some combinations of dual resistances led to identification of additional antibiotics that show CS compared to the CS associated antibiotics conditioned on only single-antibiotic resistance. Some dual resistances (B&C) led to CS for at least 50% of the tested antibiotics (A) (Figure 6.9). Of these dual resistances, especially the antibiotic pairs gentamycin and imipenem (GEN & IMP) and ampicillin and imipenem (AMP & IMP) were interesting. For *K. aerogenes*,

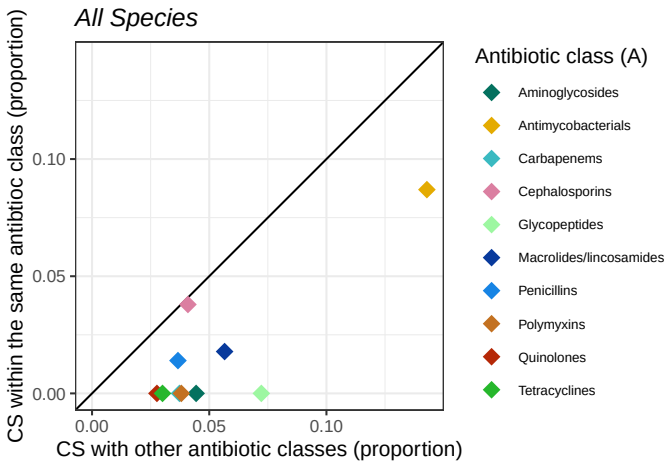


Figure 6.7: Collateral sensitivity (CS) between antibiotic classes. Over all species, the proportion of CS effects detected within and between antibiotic classes. On the x-axis the proportion of detected CS effects with other antibiotic classes. On the y-axis the proportion of detected CS effects within the same antibiotic class. The line denotes no difference between CS detected within and between antibiotic classes. All antibiotic classes in the lower right triangle more often show CS with other classes than within their own antibiotic class.

these dual resistances led to collateral sensitivity effects that were larger than the effects of the individual antibiotics on the third antibiotic (A). The difference between the two-way CS effects and the three-way effects indicate whether there is an interaction effect of the two conditioning antibiotics on the tested antibiotic.

6.4 Discussion

In this study we explored collateral sensitivity (CS) effects in clinical strains and quantified these effects in a wide variety of antibiotics and bacterial species. The combined data from the four databases used enabled comprehensive inference of expected CS effects, identifying a large number of reciprocal and unidirectional CS effects across species and antibiotics.

The databases used for this research contain a large set of different strains and MIC measurement techniques. The strains in ARESdb were very homogeneously measured and contains only clinical strains and the database is more specifically focused on strains with AMR profiles. PATRIC and NIH contained both laboratory and clinical strains and multiple techniques, such as the E-test and a disk diffusion assay. These techniques can yield different results, but due to the large numbers of strains, and since the MIC techniques were the same within a strain, this is expected to be averaged out. The RIVM database only contains Dutch samples, which can skew the

analysis to more commonly detected AMR and CS in the Netherlands.

From the detected collateral effects, some combinations were also found in experimental studies, such as amikacin (AMK) and ampicillin (AMP) in *E. coli* (Imamovic & Sommer, 2013) and small CS effect between ciprofloxacin (CIP) and gentamicin (GEN) in *S. pneumoniae* (Liakopoulos et al., 2022), but other findings from experimental studies were not detected in our screening, for example the CS between nitrofurantoin (NIT) and tigecycline (TGC) (Roemhild et al., 2020). It is not strange to find different results between experimental studies and this large clinical screening, since results from lab strains and environments are not directly translatable to the clinical setting.

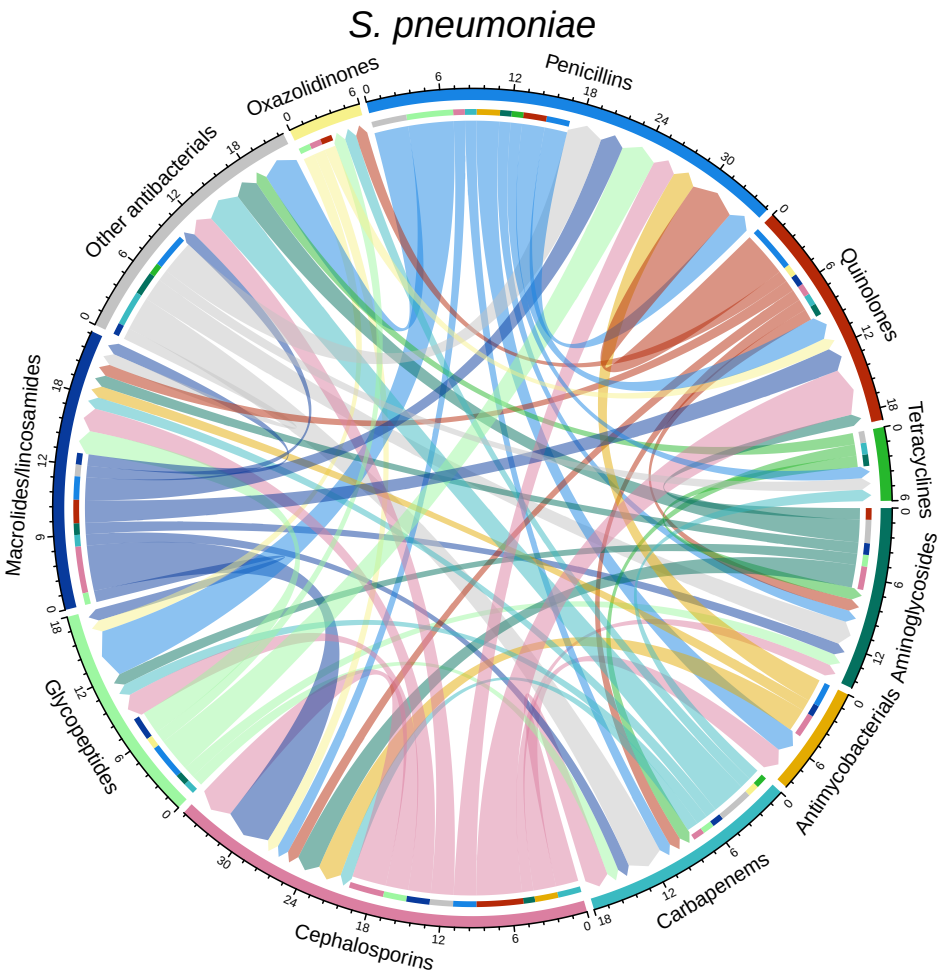


Figure 6.8: Number of CS effects between different classes of antibiotics. The arrows point towards the antibiotic classes where antibiotic sensitivity was increased.

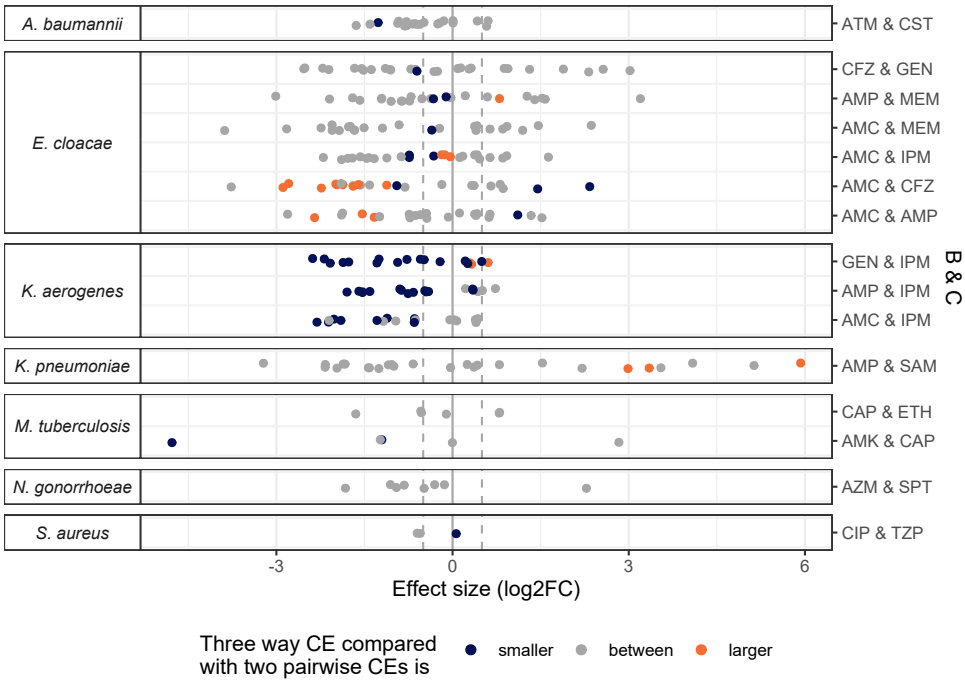


Figure 6.9: Collateral effects for the 15 three-way interactions with most collateral sensitivity responses, where resistance to a pair of antibiotic (B&C) increases the sensitivity to a third antibiotic (A). The color shows whether the three-way effect is smaller than ($A|B$&$C < A|B$ & $A|B$&$C < A|C$, blue), in between (grey) or larger ($A|B$&$C > A|B$ & $A|B$&$C > A|C$, orange) than the individual two-way collateral effects. With a smaller effect (blue) indicating an interaction between two antibiotics that makes the third antibiotic even more sensitive, such as for GEN & IMP and AMP & IMP in *K. aerogenes*.

Although many collateral sensitivity effects were found in most species, only few of them were found to be consistent over multiple species, showing the large diversity between species and indicating a need for species specific research for CS effects. We did show more similarity between certain species, such as *S. maltophilia* and *S. aureus*, which may allow for translation of CS effects between these species. Further apart from most species, as maybe expected, was for example *M. tuberculosis*, which is a less typical species, physiologically different from most other studied pathogens (Gagneux et al., 2006).

Most CS effects were detected between different antibiotic classes, however, no apparent patterns could be distinguished on which antibiotic classes more often show CS between each other, which is expected due to their different mechanisms of action (Lázár et al., 2013). There were, however, multiple CS responses found between different types of beta-lactams, but this has been shown before to occur also in experimental setting (Rosenkilde et al., 2019). Among the more consistent CS effects, a few trends could be spotted, like more CS between cephalosporins, aminoglycosides and carbapenems, and penicillins showed CS with colistin (CST) and trimetho-

prim/sulfamethoxazole and chloramphenicol, however these were not more prevalent over all detected CS effects. Especially aminoglycosides have been previously been identified as CS inducing antibiotics (Lázár et al., 2013).

Many CS effects showed no reciprocality, meaning there was only an effect of one antibiotic on another and not of the other on the one. This is also what is often found in experimental research (Imamovic & Sommer, 2013). Non-reciprocal CS effect have been shown not to necessarily hamper the usefulness of drug cycling therapies (Aulin et al., 2021). This indicates the relevance of discovering not only reciprocal, but also non-reciprocal CS effects. One study recently studied observational collateral effects in a large database with clinical strains, where the authors identified apparent CS effects in different species, based on a measure of mutual information (Beckley & Wright, 2021). This mutual information criterium can indicate whether there is a negative (CS) or positive (collateral resistance) association found between two antibiotics, but cannot take directionality in account, where the method used in our study, does estimate an effect size for each direction (Zwep et al., 2021).

The statistical method used in this study is not able to detect causal relations between antibiotics, so all found results are associative, rather than causal collateral effects, which also makes it impossible to distinguish between multidrug resistance and collateral resistance (Zwep et al., 2021). Next to this, the collateral effect estimate can depend on the dichotomization criterion chosen, which in this study was chosen to create the most equal group sizes of B_r and $\neg B_r$. Collateral sensitivity seems prevalent in clinical strains, but in order to translate these findings to the clinic, experimental validation is needed, both to validate these findings and to better understand the mechanisms behind CS responses. A better understanding of the mechanisms can facilitate discovery of CS responses in antibiotic pairs that are not researched. One way of gaining a better mechanistic understanding is by studying whole genome sequences of bacterial strains to discover genetic differences between resistant and sensitive strains to find the genes that might be involved in collateral effects (Roemhild et al., 2020).

To explore more complex interactions, we explored three-way interactions between antibiotics, based on whether a group with resistance to two antibiotics had a lower resistance to a third antibiotic. Due to the discrete nature of the data, not all splits were based on both antibiotics. Interestingly, some combinations of three antibiotics showed collateral sensitivity, where the separate pair-wise combinations did not show such large CS effect or even a collateral resistance effect, such as gentamycin and imipenem (GEN & IMP) and ampicillin and imipenem (AMP & IMP) showed in K. aerogenes. This indicates why focusing on multiple drug interactions can uncover more complex CS responses that might be useful especially in chronic infections, where drug combinations and cycling regimens often include more than two antibiotics. Analyzing the multidimensional collateral effects, however, vastly increases the number of combinations of antibiotics that are tested. With the data that were available in this study, a three-way interaction was feasible, but when going to four- or five-dimensional data, the dimension becomes exponentially larger.

6.5 Conclusion

Our study showed that CS commonly occurs in clinically relevant bacterial pathogenic species, strains and antibiotics, with limited consistency between species and antibiotic classes. Our findings may guide prioritization of CS-based antibiotic treatment strategies that reduce the risk for AMR.

References

Aulin, L. B. S., Liakopoulos, A., van der Graaf, P. H., Rozen, D. E., & van Hasselt, J. G. C. (2021, Sep). Design principles of collateral sensitivity-based dosing strategies. *Nature Communications*, *12*(1), 5691. Retrieved from http://dx.doi.org/10.1038/s41467-021-25927-3https://www.nature.com/articles/s41467-021-25927-3 doi: 10.1038/s41467-021-25927-3

Baym, M., Stone, L. K., & Kishony, R. (2016, Jan). Multidrug evolutionary strategies to reverse antibiotic resistance. *Science*, *351*(6268). Retrieved from https://doi.org/10.1126%2Fscience.aad3292 doi: 10.1126/science.aad3292

Beckley, A. M., & Wright, E. S. (2021, Oct). Identification of antibiotic pairs that evade concurrent resistance via a retrospective analysis of antimicrobial susceptibility test results. *The Lancet Microbe*, *2*(10), e545–e554. Retrieved from https://doi.org/10.1016%2Fs2666-5247%2821%2900118-x doi: 10.1016/s2666-5247(21)00118-x

Benjamini, Y., & Yekutieli, D. (2001, Aug). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, *29*(4). Retrieved from https://doi.org/10.1214%2Faos%2F1013699998 doi: 10.1214/aos/1013699998

Davis, J. J., Wattam, A. R., Aziz, R. K., Brettin, T., Butler, R., Butler, R. M., ... Stevens, R. (2019, Oct). The PATRIC bioinformatics resource center: expanding data and analysis capabilities. *Nucleic Acids Research*. Retrieved from https://doi.org/10.1093%2Fnar%2Fgkz943 doi: 10.1093/nar/gkz943

Gagneux, S., Long, C. D., Small, P. M., Van, T., Schoolnik, G. K., & Bohannan, B. J. M. (2006, Jun). The competitive cost of antibiotic resistance in mycobacterium tuberculosis. *Science*, *312*(5782), 1944–1946. Retrieved from https://doi.org/10.1126%2Fscience.1124410 doi: 10.1126/science.1124410

Heiser, W. J., Busing, F. M. T. A., & Meulman, J. J. (2020). Mapping networks and trees with multidimensional scaling of proximities. In *Advanced studies in behaviormetrics and data science* (pp. 385–407). Springer Singapore. Retrieved from https://doi.org/10.1007%2F978-981-15-2700-5_24 doi: 10.1007/978-981-15-2700-5_24

Imamovic, L., Ellabaan, M. M. H., Machado, A. M. D., Citterio, L., Wulff, T., Molin, S., ... Sommer, M. O. A. (2018, Jan). Drug-driven phenotypic convergence supports rational treatment strategies of chronic infections. *Cell*, *172*(1-2), 121–134.e14. Retrieved from https://doi.org/10.1016%2Fj.cell.2017.12.012 doi: 10.1016/j.cell.2017.12.012

Imamovic, L., & Sommer, M. O. A. (2013, Sep). Use of collateral sensitivity networks to design drug cycling protocols that avoid resistance development. *Science Translational Medicine*, *5*(204). Retrieved from https://doi.org/10.1126%2Fscitranslmed.3006609 doi: 10.1126/scitranslmed.3006609

Johnson, A. P. (2015, Jun). Surveillance of antibiotic resistance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1670), 20140080. Retrieved from https://doi.org/10.1098%2Frstb.2014.0080 doi: 10.1098/rstb.2014.0080

Lázár, V., Martins, A., Spohn, R., Daruka, L., Grézal, G., Fekete, G., ... Pál, C. (2018, Jun). Antibiotic-resistant bacteria show widespread collateral sensitivity to antimicrobial peptides. *Nature Microbiology*, *3*(6), 718–731. Retrieved from http://dx.doi.org/10.1038/s41564-018-0164-0http://www.nature.com/articles/s41564-018-0164-0 doi: 10.1038/s41564-018-0164-0

Lázár, V., Singh, G. P., Spohn, R., Nagy, I., Horváth, B., Hrtyan, M., ... Pál, C. (2013, Jan). Bacterial evolution of antibiotic hypersensitivity. *Molecular Systems Biology*, *9*(1), 700. Retrieved from https://doi.org/10.1038%2Fmsb.2013.57 doi: 10.1038/msb.2013.57

Liakopoulos, A., Aulin, L. B. S., Buffoni, M., Fragkiskou, E., van Hasselt, J. G. C., & Rozen, D. E. (2022, Apr). Allele-specific collateral and fitness effects determine the dynamics of fluoroquinolone resistance evolution. *Proceedings of the National Academy of Sciences*, *119*(18). Retrieved from https://doi.org/10.1073%2Fpnas.2121768119 doi: 10.1073/pnas.2121768119

Maltas, J., & Wood, K. B. (2019, Oct). Pervasive and diverse collateral sensitivity profiles inform optimal strategies to limit antibiotic resistance. *PLOS Biology*, *17*(10), e3000515. Retrieved from https://doi.org/10.1371%2Fjournal.pbio.3000515 doi: 10.1371/journal.pbio.3000515

OpGen. (2022). *ARESdb*. Retrieved from https://www.opgen.com/ares/ares-products/aresdb/

Pál, C., Papp, B., & Lázár, V. (2015, Jul). Collateral sensitivity of antibiotic-resistant microbes. *Trends in Microbiology*, *23*(7), 401–407. Retrieved from https://doi.org/10.1016%2Fj.tim.2015.02.009 doi: 10.1016/j.tim.2015.02.009

Roemhild, R., Linkevicius, M., & Andersson, D. I. (2020, Jan). Molecular mechanisms of collateral sensitivity to the antibiotic nitrofurantoin. *PLOS Biology*, *18*(1), e3000612. Retrieved from https://doi.org/10.1371%2Fjournal.pbio.3000612 doi: 10.1371/journal.pbio.3000612

Rosenkilde, C. E. H., Munck, C., Porse, A., Linkevicius, M., Andersson, D. I., & Sommer, M. O. A. (2019, Feb). Collateral sensitivity constrains resistance evolution of the CTX-m-15 beta-lactamase. *Nature Communications*, *10*(1). Retrieved from https://doi.org/10.1038%2Fs41467-019-08529-y doi: 10.1038/s41467-019-08529-y

Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., ... Sherry, S. T. (2020, Oct). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *49*(D1), D10–D17. Retrieved from https://doi.org/10.1093%2Fnar%2Fgkaa892 doi: 10.1093/nar/gkaa892

van der Kuil, W. A., Schoffelen, A. F., de Greeff, S. C., Thijsen, S. F., Alblas, H. J., Notermans, D. W., ... and, T. L. (2017, Nov). National laboratory-based surveillance system for antimicrobial resistance: a successful tool to support the control of antimicrobial resistance in the netherlands. *Eurosurveillance*, *22*(46). Retrieved from https://doi.org/10.2807%2F1560-7917.es.2017.22.46.17-00062 doi: 10.2807/1560-7917.es.2017.22.46.17-00062

Zwep, L. B., Haakman, Y., Duisters, K. L. W., Meulman, J. J., Liakopoulos, A., & van Hasselt, J. G. C. (2021, Sep). Identification of antibiotic collateral sensitivity and resistance interactions in population surveillance data. *JAC-Antimicrobial Resistance*, *3*(4). Retrieved from https://doi.org/10.1093%2Fjacamr%2Fdlab175 doi: 10.1093/jacamr/dlab175

Supplementary material

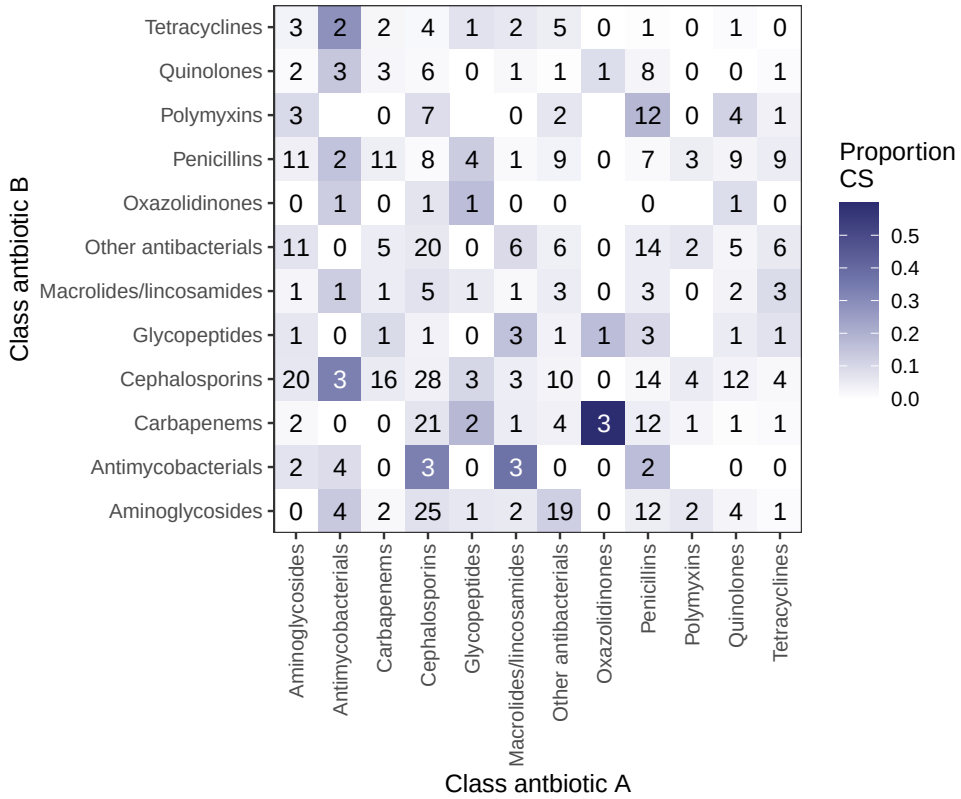


Figure S6.1: The number of CS responses found in at least one species for each combination of antibiotic classes. The color indicates the proportion of CS effects within all tested antibiotic pairs for each class and the numbers show the actual number of CS effects for that combination.

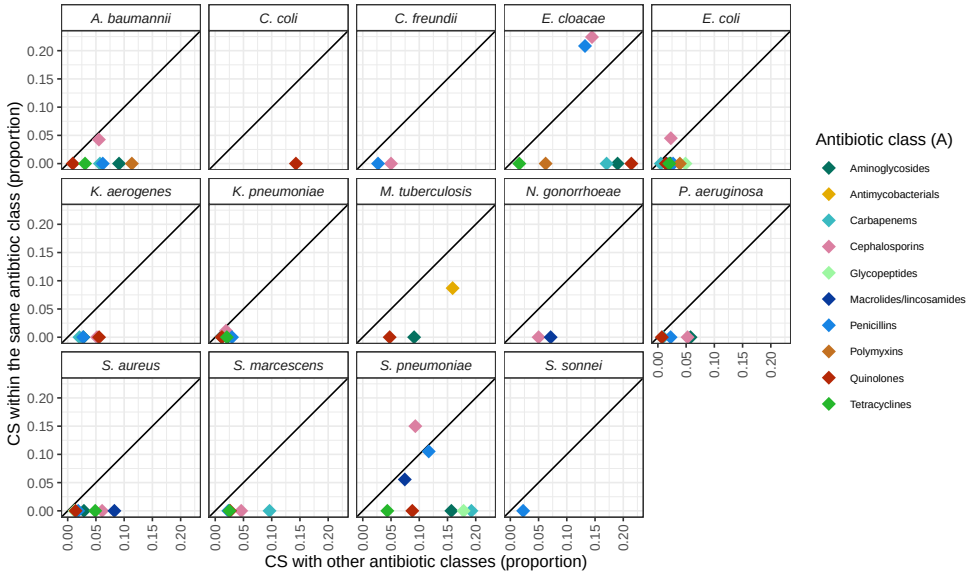


Figure S6.2: Collateral sensitivity (CS) between antibiotic classes. The proportion of CS effects detected within and between antibiotic classes for each species (C. jejuni was excluded, because there were no collateral effects estimated within the antibiotic class). On the x-axis the proportion of detected CS effects with other antibiotic classes. On the y-axis the proportion of detected CS effects within the same antibiotic class. The line denotes no difference between CS detected within and between antibiotic classes. All antibiotic classes in the lower right triangle more often show CS with other classes than within their own antibiotic class.

Table S6.1: All studied antibiotics, abbreviations, and antibiotic classes

Abbreviation	Antibiotic	Antibiotic class
AMA	4-aminosalicylic acid	Antimycobacterials
AMK	amikacin	Aminoglycosides
AMX	amoxicillin	Beta-lactams/penicillins
AMC	amoxicillin/clavulanic acid	Beta-lactams/penicillins
AMP	ampicillin	Beta-lactams/penicillins
SAM	ampicillin/sulbactam	Beta-lactams/penicillins
AZM	azithromycin	Macrolides/lincosamides
ATM	aztreonam	Beta-lactams/penicillins
CAP	capreomycin	Antimycobacterials
CEC	cefaclor	Cephalosporins (2nd gen.)
CFZ	cefazolin	Cephalosporins (1st gen.)
FEP	cefepime	Cephalosporins (4th gen.)
CFM	cefixime	Cephalosporins (3rd gen.)
CTX	cefotaxime	Cephalosporins (3rd gen.)
CTX-CLA	cefotaxime/clavulanic acid	Cephalosporins (3rd gen.)
CTT	cefotetan	Cephalosporins (2nd gen.)
CTF	cefotiam	Cephalosporins (2nd gen.)

FOX	cefoxitin	Cephalosporins (2nd gen.)
CPO	cefpirome	Cephalosporins (4th gen.)
CPD	cefpodoxime	Cephalosporins (3rd gen.)
CPX	cefpodoxime proxetil	Cephalosporins (3rd gen.)
CPT	ceftaroline	Cephalosporins (5th gen.)
CAZ	ceftazidime	Cephalosporins (3rd gen.)
CAZ-CLA	ceftazidime/clavulanic acid	Cephalosporins (3rd gen.)
CTF	ceftiofur	Cephalosporins (3rd gen.)
CEI	ceftolozane/enzyme inhibitor	Cephalosporins (5th gen.)
CRO	ceftriaxone	Cephalosporins (3rd gen.)
CXM	cefuroxime	Cephalosporins (2nd gen.)
CXA	cefuroxime axetil	Cephalosporins (2nd gen.)
LEX	cephalexin	Cephalosporins (1st gen.)
CEF	cephalothin	Cephalosporins (1st gen.)
CHL	chloramphenicol	Amphenicols
CIP	ciprofloxacin	Quinolones
CLR	clarithromycin	Macrolides/lincosamides
CLI	clindamycin	Macrolides/lincosamides
CST	colistin	Polymyxins
CYC	cycloserine	Oxazolidinones
DAP	daptomycin	Other antibacterials
DOR	doripenem	Carbapenems
DOX	doxycycline	Tetracyclines
ETP	ertapenem	Carbapenems
ERY	erythromycin	Macrolides/lincosamides
ETH	ethambutol	Antimycobacterials
ETI1	ethionamide	Antimycobacterials
FLR	florfenicol	Other antibacterials
FLC	flucloxacillin	Beta-lactams/penicillins
FOF	fosfomycin	Other antibacterials
FRM	framycetin	Aminoglycosides
FA	fusidic acid	Other antibacterials
GAT	gatifloxacin	Quinolones
GEN	gentamicin	Aminoglycosides
IPM	imipenem	Carbapenems
INH	isoniazid	Antimycobacterials
KAN	kanamycin	Aminoglycosides
LVX	levofloxacin	Quinolones
LZD	linezolid	Oxazolidinones
MEC	mecillinam (amdinocillin)	Beta-lactams/penicillins
MEM	meropenem	Carbapenems
MIN	minocycline	Tetracyclines
MXF	moxifloxacin	Quinolones
MUP	mupirocin	Other antibacterials
NAL	nalidixic acid	Quinolones
NET	netilmicin	Aminoglycosides
NIT	nitrofurantoin	Other antibacterials
NOR	norfloxacin	Quinolones
OFX	ofloxacin	Quinolones
OXA	oxacillin	Beta-lactams/penicillins
BPE	benzylpenicillin	Beta-lactams/penicillins
PNO	penicillin/novobiocin	Beta-lactams/penicillins

PIP	piperacillin	Beta-lactams/penicillins
TZP	piperacillin/tazobactam	Beta-lactams/penicillins
PMB	polymyxin b	Polymyxins
PRI	pristinamycin	Macrolides/lincosamides
PZA	pyrazinamide	Antimycobacterials
Q-D	quinupristin/dalfopristin	Macrolides/lincosamides
RFB	rifabutin	Antimycobacterials
RIF	rifampicin	Antimycobacterials
SPX	sparfloxacin	Quinolones
SPT	spectinomycin	Other antibacterials
STR	streptoduocin	Aminoglycosides
STR	streptomycin	Aminoglycosides
SMX	sulfamethoxazole	Trimethoprims
SXZ	sulfisoxazole	Other antibacterials
TEC	teicoplanin	Glycopeptides
TEL	telithromycin	Macrolides/lincosamides
TMC	temocillin	Beta-lactams/penicillins
TET	tetracycline	Tetracyclines
TIC	ticarcillin	Beta-lactams/penicillins
TIM	ticarcillin/clavulanic acid	Beta-lactams/penicillins
TGC	tigecycline	Tetracyclines
TOB	tobramycin	Aminoglycosides
TMP	trimethoprim	Trimethoprims
SXT	trimethoprim/sulfamethoxazole	Trimethoprims
VAN	vancomycin	Glycopeptides
AMA	para-aminosalicylic acid	Antimycobacterials
AMC	amoxicillin/clavulanate	Beta-lactams/penicillins
SAM	amipicillin/sulbactam	Beta-lactams/penicillins
CTX-CLA	cefotaxime/clavulanate	Cephalosporins (3rd gen.)
CAZ-CLA	ceftazidime/clavulanate	Cephalosporins (3rd gen.)
CRO	ceftriazone	Cephalosporins (3rd gen.)
CXM-S	cefuroxime/sodium	Cephalosporins (2nd gen.)
CEF	cefalotin	Cephalosporins (1st gen.)
CEF	cephalotin	Cephalosporins (1st gen.)
ETI1	ethiomide	Antimycobacterials
GEN	gentamycin	Aminoglycosides
PEN	penicillin	Beta-lactams/penicillins
PMB	polymyxin	Polymyxins
PMB	polymyxin B	Polymyxins
PZA	pyrazimide	Antimycobacterials
Q-D	synercid	Macrolides/lincosamides
RIF	rifampin	Antimycobacterials
SXT	sulfamethoxazole/trimethoprim	Trimethoprims
TIM	ticarcillin/clavulanate	Beta-lactams/penicillins

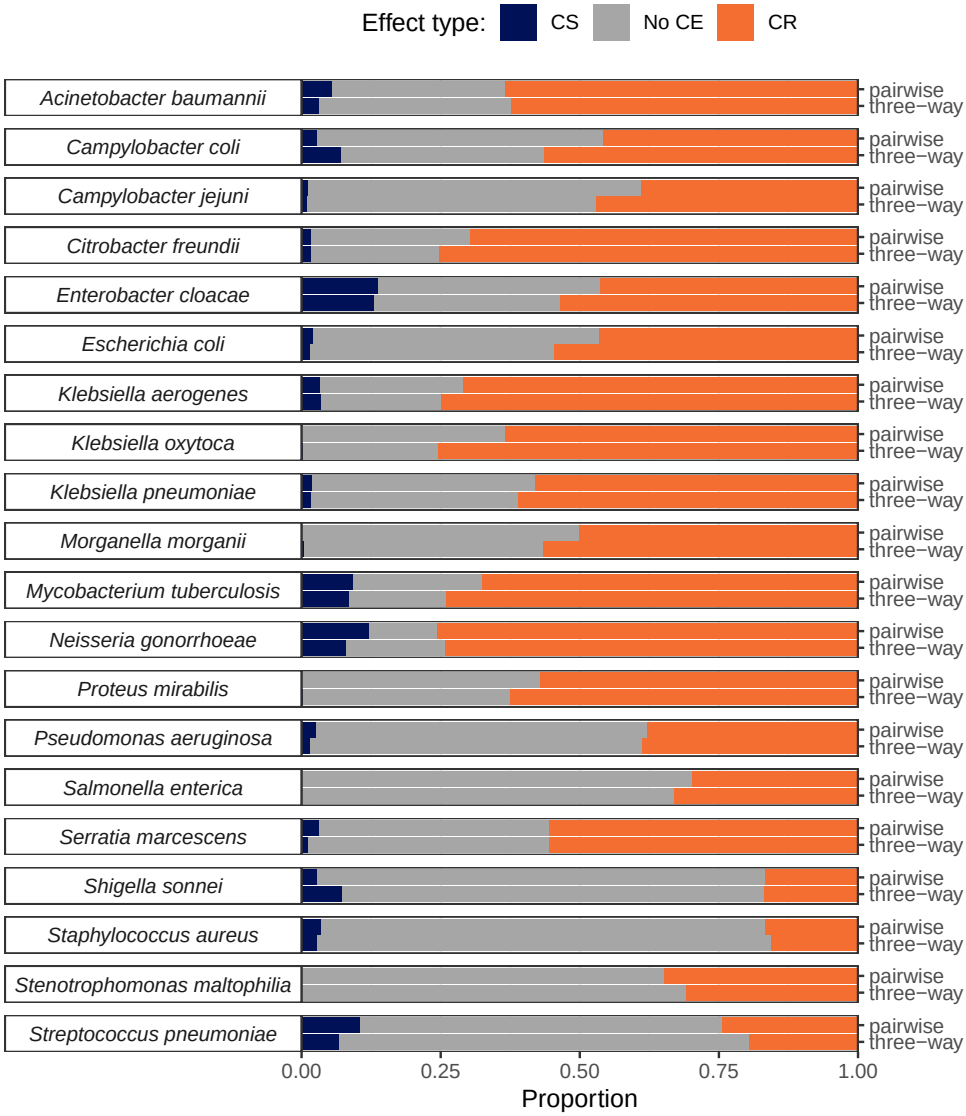


Figure S6.3: The proportion of collateral effects detected in the three-way analysis as compared to the pairwise analysis for each species. Abbreviations: CS; collateral sensitivity, no CE; no collateral effect, CR; collateral resistance

Chapter 7

Virtual patient simulation using copula modeling

Authors

Laura B. Zwep

Tingjie Guo

Thomas Nagler

Catherijne A. J. Knibbe

Jacqueline J. Meulman

J. G. Coen van Hasselt

Abstract

Clinical trial design and dosing optimization strategies are increasingly relying on model-based approaches in pharmacometrics and quantitative systems pharmacology (QSP), which incorporate patient characteristics to simulate the expected pharmacokinetic (PK) or pharmacodynamic (PD) response in cohorts of virtual patients. To this end, the individual-level patient characteristics, or covariates, are used as input for such simulations should accurately reflect the values seen in real patient populations. Current methods to achieve this goal either make unrealistic assumptions about the correlation between patient's covariates, or require direct access to actual data sets with individual-level patient data, which may often be limited by data sharing limitations. Here, we propose and evaluate the use of copulas to address current shortcomings in simulation of patient-associated covariates for virtual patient simulations for model-based dose and trial optimization in clinical pharmacology. Copulas are multivariate distribution functions that can capture joint distributions, including the correlation, of covariate sets. We compare the performance of copulas to alternative simulation strategies and we demonstrate their utility to a number of case studies. Our analyses demonstrate that copulas can reproduce realistic patient characteristics, both in terms of individual covariates and the dependence structure between different covariates, outperforming alternative methods, in particular when aiming to reproduce high-dimensional covariate sets. In conclusion, copulas represent a versatile and generalizable approach for virtual patient simulation which preserve relationships between covariates, and offer an open science strategy to facilitate re-use of patient data sets.

7.1 Introduction

Model-based approaches in pharmacometrics and quantitative systems pharmacology (QSP) (Bonate, 2000, 2001; Chelliah et al., 2020) have become of pivotal importance for the optimization of drug treatment strategies or clinical trial designs (Holford et al., 2010; Langenhorst et al., 2020). These model-based approaches typically simulate the expected pharmacokinetic (PK) and/or pharmacodynamic (PD) response and the associated inter-individual variability for a cohort of virtual patients. Here, the inter-individual variability in the PK or PD response is often in part captured by patient-specific characteristics such as age, weight, organ function biomarkers, or specific genetic polymorphisms, incorporated in quantitative PK-PD or QSP models. The increasing public availability of quantitative PK-PD or QSP models for many important therapeutics thus offers extensive opportunities for the clinical pharmacology community to perform virtual patient simulations. These simulations may aid in design of (stratified) dosing strategies in particular for new (special) patient population populations (De Cock et al., 2016), such as pediatric (Illamola et al., 2016; J. G. C. van Hasselt, Allegaert, et al., 2014; Vinks et al., 2015) or pregnant patients (J. G. van Hasselt et al., 2014), or, to evaluate different potential trial designs in specific types of

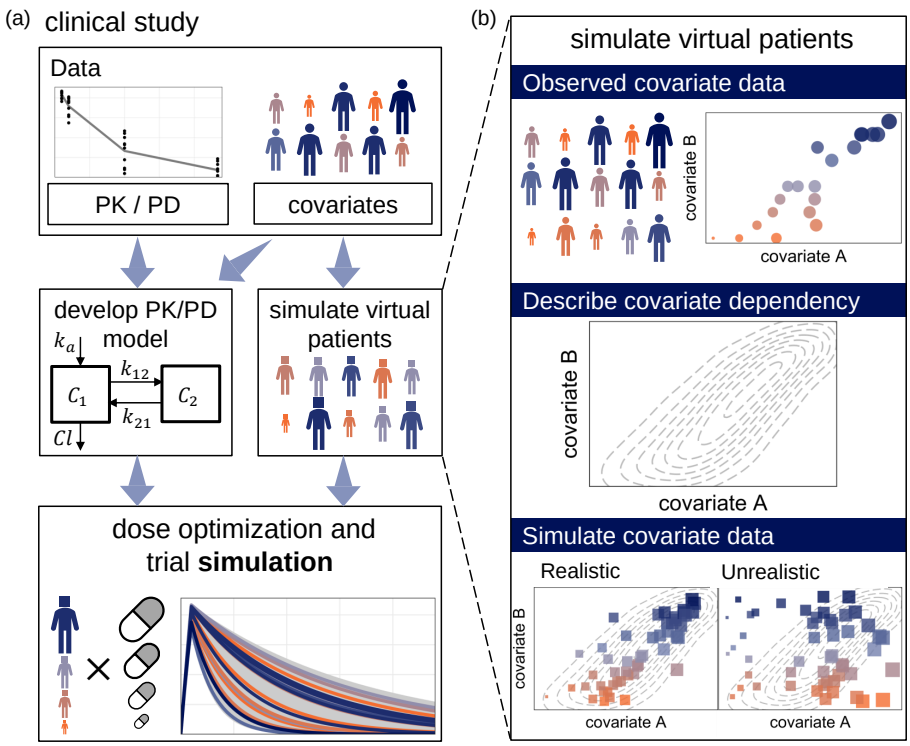


Figure 7.1: Pharmacometric workflow. (a) In order to optimize dosing for new medication or special patient populations, pharmacometric models, such as (PK/)PD models, are used to simulate new patient dosing regimens. Next to the developed pharmacometric model, simulation studies require covariate simulation. (b) An important challenge for covariate simulation is sampling realistic patients, where the dependency between covariates is preserved.

patients or treatments(J. G. C. van Hasselt, Allegaert, et al., 2014; J. G. C. van Hasselt, van Eijkelenburg, et al., 2014; Yoneyama et al., 2017) (Figure 7.1a).

A key requirement to enable simulation of realistic virtual patients is to produce realistic sets of patient-associated characteristics or covariates used in the model. Such covariates can include demographics (e.g. body weight, sex, age), organ function measures (e.g. renal or hepatic function), pharmacodynamic endpoints (cardiovascular readouts, biochemical biomarkers), and increasingly also high-dimensional pharmacogenomic data. Importantly, such covariates may have various distributions including an intricate dependency structure (i.e., correlation) that must be accounted for in virtual patient simulation to produce realistic patient-profiles (Figure 7.1b). Not considering such correlations leads to an inflation of the variability in covariates and hence unrealistic virtual patients. For example, a patient of 95 years old, with a high body weight and a very good kidney function is a combination that is not expected to actually exist. Various data analytical strategies are available to generate sets of

realistic patient covariates for virtual patient simulation. These strategies are either based on methods that require direct access to the appropriate individual patient-level covariate data, which may often not be available, or on methods that characterize the covariate distributions.

Covariate generation methods that utilize available patient-level covariate data include resampling methods such as the bootstrap (Efron, 1979), which preserve the dependence structure of the patient covariates by directly resampling from the observed data. However, these methods are only able to simulate patients that are already present in the data set and require a large enough number of patients to be included. These shortcomings were addressed by a recently proposed imputation method using conditional distributions (CD) (Smania & Jonsson, 2021), although this method remains dependent on access to patient-level data. Distribution-based simulation methods for virtual patient simulation do not require patient-level data access. Although initially distributions are often derived from patient-level data, subsequent use of these distributional models to generate sets of patient-level covariates is independent of access to such data. The most straightforward strategy is to capture the marginal density of covariates in univariate parametric distributions with associated means and variances for each covariate, and to subsequently draw random samples from these distributions. However, such an approach assumes that covariates are fully independent and do not show any correlation. Alternatively, multivariate normal distributions (MVND) (Tannenbaum et al., 2006) do capture the correlation structure (Teutonico et al., 2015), but make strong assumptions regarding the (multivariate normal) distributional shape, which is commonly violated. Thus, depending on the distribution of the covariates of interest this again can lead to unrealistic sets of virtual patient covariates.

Copulas are multivariate distribution functions that can capture the joint distribution, including the dependence structure for sets of covariates, and are thus of interest as a distribution-based approach for generating realistic sets of covariates. They address shortcomings of alternative distribution-based methods while not requiring access to patient-level data (Czado, 2019; Nagler & Czado, 2016; Sklar, 1973). In this study, we aim to evaluate and demonstrate the utility of copulas as a novel strategy to support realistic virtual patient simulation in the context of the field of clinical pharmacology. We first compare the performance of copula models in comparison to existing methods including the bootstrap, CD, MVND, and marginal distribution. We then demonstrate the application of copulas in three case studies focusing on pharmacokinetic simulations, time-varying covariates, and higher-dimensional covariates.

7.2 Methods

7.2.1 Data

Three different datasets of combined patient characteristics were used in this study to evaluate the performance and explore different applications. The first data set

contains a special patient population of pediatric patients (Cock et al., 2014) with 445 neonates and young children admitted to the ICU, with twelve measured covariates, including body weight, serum creatinine level (SCr) and age. These data were used to evaluate the simulation performance (Data set 1). A second data set on pregnancy data (Patel et al., 2013) with 123 subjects, with biomarkers measured over time, was used to simulate longitudinal covariate profiles (Data set 2). Lastly, MIMIC (Johnson et al., 2022), a large observational dataset with ICU patients, was used to evaluate the correlation structure between a large set of 30 variables for >53,000 patients (Data set 3).

7.2.2 Copula estimation and simulation

Vine copulas were used to estimate the joint density between all covariates. Kernel density estimation was used to estimate the marginal density of each covariate. Using the probability integral function, the covariates were transformed to a uniform scale, with values on the [0,1] domain (Nagler & Vatter, 2020). Based on the correlations between the covariates, a vine structure was chosen, where the most correlated covariates were placed closer to each other in the vine structure. For each bivariate copula, a set of parametric distributions was fit and the best fitting distributions were chosen by minimizing the AIC. Vine copulas with different distributions were fit using the R library rvinecopulib (Nagler & Vatter, 2021). The resulting copula density was used to simulate covariates with uniform marginal densities. The earlier estimated marginal densities were used to transform these covariates back to their original scale, yielding the simulated covariate sets for virtual patients. All analyses were performed in R (`https://github.com/vanhasseltlab/copula_vps`).

7.2.3 Evaluation of simulation performance

To evaluate how well copulas can be used for simulation of covariate sets, we calculated the performance of copula simulations on the pediatric data (Cock et al., 2014) (Data set 1). The estimation and simulation were performed in two differently sized covariate sets, with the same subjects, but a different number of covariates: one simulation on three covariates, age, SCr and body weight, and one on twelve covariates. The distribution of the simulated population was compared with the distribution of the observed population in terms of the mean and standard deviation for each covariate and correlation between each combination of covariates. A relative error was computed for each of these statistics (S) as

$$\text{Relative error} = \frac{\hat{S} - S}{S}$$

where $\hat{S}$ denotes the statistic of the simulated population. The simulations were repeated 100 times.

The copula results were compared to four other simulation methods, of which two method are based on patient-level data and two methods are based on characteriza-

tion of the covariate distribution. Bootstrap simulations were conducted by resampling full rows from the original data with replacement (Efron, 1979). The conditional distribution (CD) approach, which uses a multiple imputation algorithm to iteratively impute covariate values for virtual patients, was used as implemented by the developers of the method (Smania & Jonsson, 2021). The standard multiple imputation method 'predictive mean matching' was used, corresponding to their paper. The distribution-based methods used were the multivariate normal distribution (MVND) and marginal distributions (MDs), through maximum likelihood estimation. The best fitting multivariate normal distribution was fitted. The univariate MDs of each covariate was estimated using a kernel density estimation method (Nagler, 2017; Nagler & Vatter, 2020). Covariate values were sampled from the respective density functions.

7.2.4 Applications

Pharmacokinetic simulation of vancomycin in pediatric patients

For the proposed copula approach, the effect of preserving the dependence structure in covariate simulation methods was evaluated on PK predictions in pediatric patients. To this end, for Data set 1, the performance of the use of body weight and SCr from the three-covariate copula and the MDs simulation was compared in a population PK one-compartmental model for vancomycin (Grimsley & Thomson, 1999).

$$\frac{dA}{dt} = 0 - \frac{Cl}{V} \cdot A$$

$$Cl = \frac{3.56 \cdot WT}{SCr}$$

$$V = 0.669 \cdot WT$$

This PK model was used to calculate the PK curves from the original pediatric covariate data (Data set 1) and the simulated covariate data from the three-covariate copula and MDs simulations. These PK profiles were compared using the AUC of the first 24 hours after dosing. The correlation between the AUC and the covariates, SCr and body weight, was evaluated to identify whether this correlation was recovered between the covariates and the PK curve.

Time-varying covariates in pregnancy data

One of the possible applications of using copulas is the simulation of time-varying covariates. Using Data set 2 with six time-varying covariates (y) over the gestational age (t) during pregnancy (Patel et al., 2013), including albumin concentration, bilirubin concentration, lymphocytes, neutrophils, platelets and SCr, we fit a copula to simulate time varying covariates in a two-step procedure. First, we fitted a second degree mixed effects polynomial regression model on the temporal data for each covariate j and extracted three individual parameters for each patient i, the intercept ($\beta_{0j} + b_{0ji}$),

the linear term ($\beta_{1j} + b_{1ji}$) and the quadratic term ($\beta_{2j} + b_{2ji}$), resulting in a total of 18 dimensions.

$$\hat{y}_{ij}(t) = \beta_{0j} + b_{0ji} + \beta_{1j} \cdot t + b_{1ji} \cdot t + \beta_{2j} \cdot t^2 + b_{2ji} \cdot t^2$$
$$b_{0ji} \sim N(0, \sigma_0)$$
$$b_{1ji} \sim N(0, \sigma_1)$$
$$b_{2ji} \sim N(0, \sigma_2)$$

For example, yielding for albumin concentration:

$$\widehat{\text{Albumin conc}}_i(t) = 44.1 + b_{0ji} + 0.269 \cdot t + b_{1ji} \cdot t + 0.0017 \cdot t^2 + b_{2ji} \cdot t^2$$

$$b_{0ji} \sim N(0, 1.86)$$
$$b_{1ji} \sim N(0, 0.105)$$
$$b_{2ji} \sim N(0, 0.00224)$$

Second, instead of fitting a copula directly on the longitudinal covariates, the copula was fitted on the set of individual parameter estimates, yielding the six new sets of intercepts, linear and quadratic terms for each simulated patient. To create time-dependent covariates, the curves for each patient were retrieved from the simulated parameter sets. The performance of the copula simulation was evaluated by comparing the time-curves estimated from the copula simulated time curves with those estimated on the original pregnancy data. The performance was evaluated both in terms of the simulated individual parameters as the calculated time-curves. Next to simulation with the copula, the time-varying covariates were simulated in a similar two-step approach with MDs, to compare the differences between the MDs and copula.

Covariate distributions in large ICU data

To characterize the joint distributions in a large dataset, copula simulation was used to characterize and simulate from the MIMIC database (Data set 3) (Johnson et al., 2022). A copula model was fit to a large dataset of 30 available patient-associated covariates with primary focus on clinical laboratory measurements from >53,000 ICU patients. There were many values missing over the covariates and subjects. To estimate the copula on missing data, for each combination of covariates needed for a node in the vine copula structure, the complete observations were used. This simulation was used to demonstrate how copulas can be used to characterize the underlying dependency structure of these covariates and evaluate the correlations.

7.3 Results

7.3.1 Evaluation of simulation performance

The performance of the copulas was assessed on two differently sized datasets, one with three covariates and one with twelve covariates (Data set 1). First, for a set of three covariates, copulas show a low relative error of 0.02, 0.08 and 0.04 for the in terms of correlations between age and body weight, age and SCr, and body weight and SCr respectively (Figure 7.2a). Second, for the twelve-covariate simulations, the copula simulation slightly underestimates the covariances with a median error of 0.05 over all covariate combinations (Figure 7.2b).

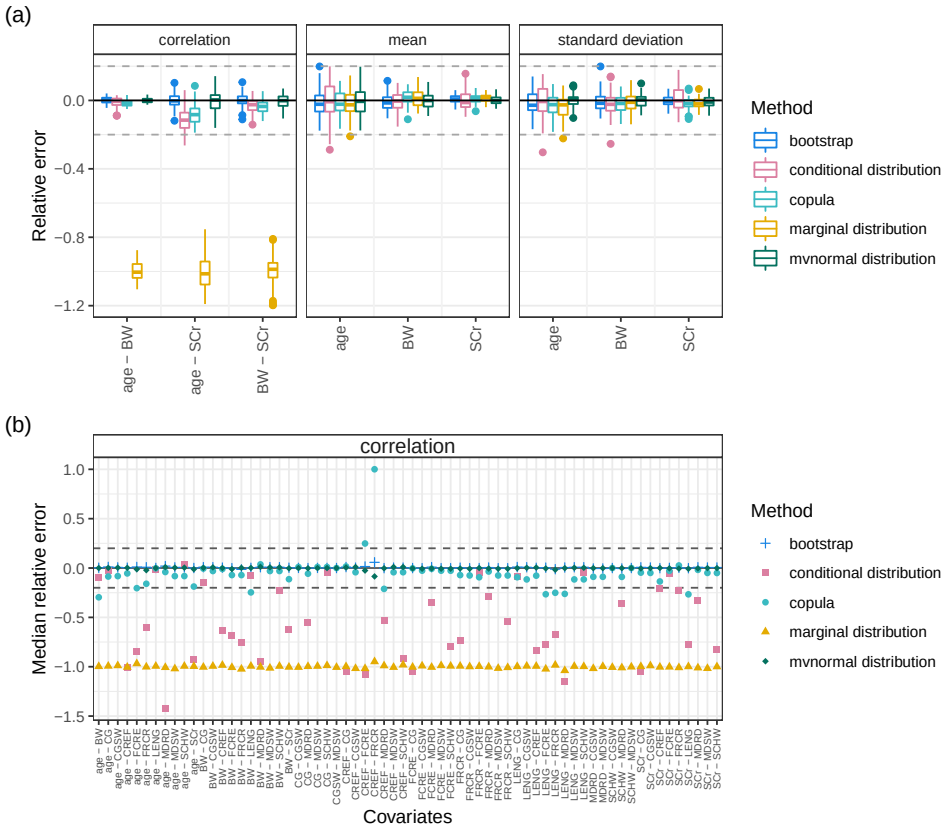


Figure 7.2: Relative error over 100 simulations as compared to the statistics of the observed population for five different simulation methods. (a) Boxplots of the correlation, mean and standard deviation of three covariates. (b) Median relative error of a large covariate simulation for the correlations of each combination of twelve covariates.

The performance of copulas was compared to four other simulation methods. For the three-covariate simulation the copula yielded similar results to the conditional distributions, which has relative errors of 0.01, 0.12 and 0.03 (Figure 7.2a), but for the

twelve-covariate simulations, the CD simulations show a large median underestimation with a relative error of 0.60 (Figure 7.2b). The bootstrap shows the best performance, since it can fully keep the dependence structure intact, both in the three-covariate (Figure 7.2a) and the twelve-covariate simulation (Figure 2b). The MDs was unable to capture any correlation, which is seen in the relative error of around -1.0 for each covariate combination. The MVND shows a good performance in the estimates for correlation, mean and standard deviation, but a visual check of the density plots shows a non normal distribution of the covariates, which is not well covered by the simulated density (Figure S7.1).

Overall, copulas performed closest to the bootstrap, which can fully capture the dependence, but it was not able to capture all covariate combinations equally well, such as a large overestimation of the combination CREF and FRCR. The twelve-covariate model showed a weakness in the conditional distributions, which the copulas did not show and although the MVND shows very good summary metrics, the distributions themselves perform worse than the copula (Figure S7.1).

7.3.2 Applications

Pharmacokinetic simulation of vancomycin in pediatric patients

The effect of ignoring the correlation between covariates on PK simulations was evaluated by comparing the PK curves from the copula simulations with those from the MDs simulation. Covariate sets simulated for SCr and body weight from Data set 1 were used to predict PK profiles and compute subsequent AUCs. The AUCs from the copula and the MDs simulations did not show differences in summary statistics such as the median and quartiles (Figure 7.3a). However, when comparing the correlations between the covariates and the AUC, we found that the original correlation between the AUC and body weight (r = 0.67) was lost in the MDs simulations (r = -0.07), whereas the copula preserved their dependence (r = 0.66) (Figure 7.3b). If the dependence between variables is not taken into account, this can lead to unrealistic virtual patients, such as individuals with a high body weight having a high AUC.

Time-varying covariates

To evaluate how well copulas can be used to simulate time-varying covariates, a two-step simulation method was used to simulate patients, with and without taking the dependency into account, by simulating from a copula and MDs respectively. For the time-varying covariates in the pregnancy data (Data set 2), polynomial linear regression curves were fitted for each covariate, resulting in polynomial equations. The individual parameters were estimated, resulting in a set of 18 parameter estimations for all subjects. A set of virtual patients was simulated from the estimated individual parameters. The correlations between the individual parameters from the simulated patients were on average close to the correlations between the estimated parameters of the observed data. The simulated individual parameters were used to generate time-varying covariate values, by calculating the curves from the intercept and the

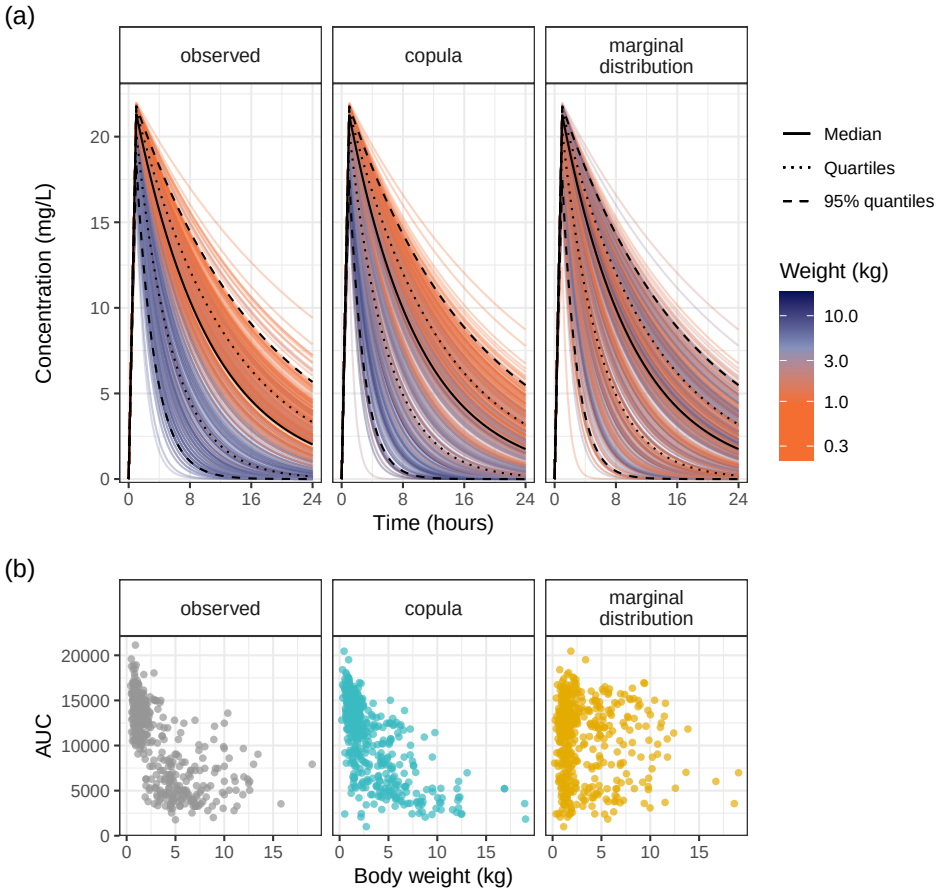


Figure 7.3: (a) Pharmacokinetic (PK) curves calculated for the observed population and the virtual patient populations from copula simulations and the marginal densities (MDs). The median and quantiles show a similar pattern between all three sets, however the weight is randomly distributed over the PK profiles for the simulation with MDs. (b) Scatter plot of area under the PK curve (AUC) against body weight.

linear and quadratic terms. Polynomial regression coefficients were simulated in a realistic domain, while simulating from a MDs led to more extreme polynomial curves, with a five times higher error on the standard deviation of the AUC (Figure 7.4). This shows how covariate values can be inflated when simulating independent covariates.

Covariate distributions in large ICU data

To establish the use of copula for simulation in a larger data set, a simulation was conducted based on 30 covariates from the MIMIC database (Data set 3). Copula estimation and simulation was feasible on this large dataset, showing how copulas can be useful for simulation for extensive pharmacometric models. The higher dimension did increase the underestimation of the correlations to a relative error of

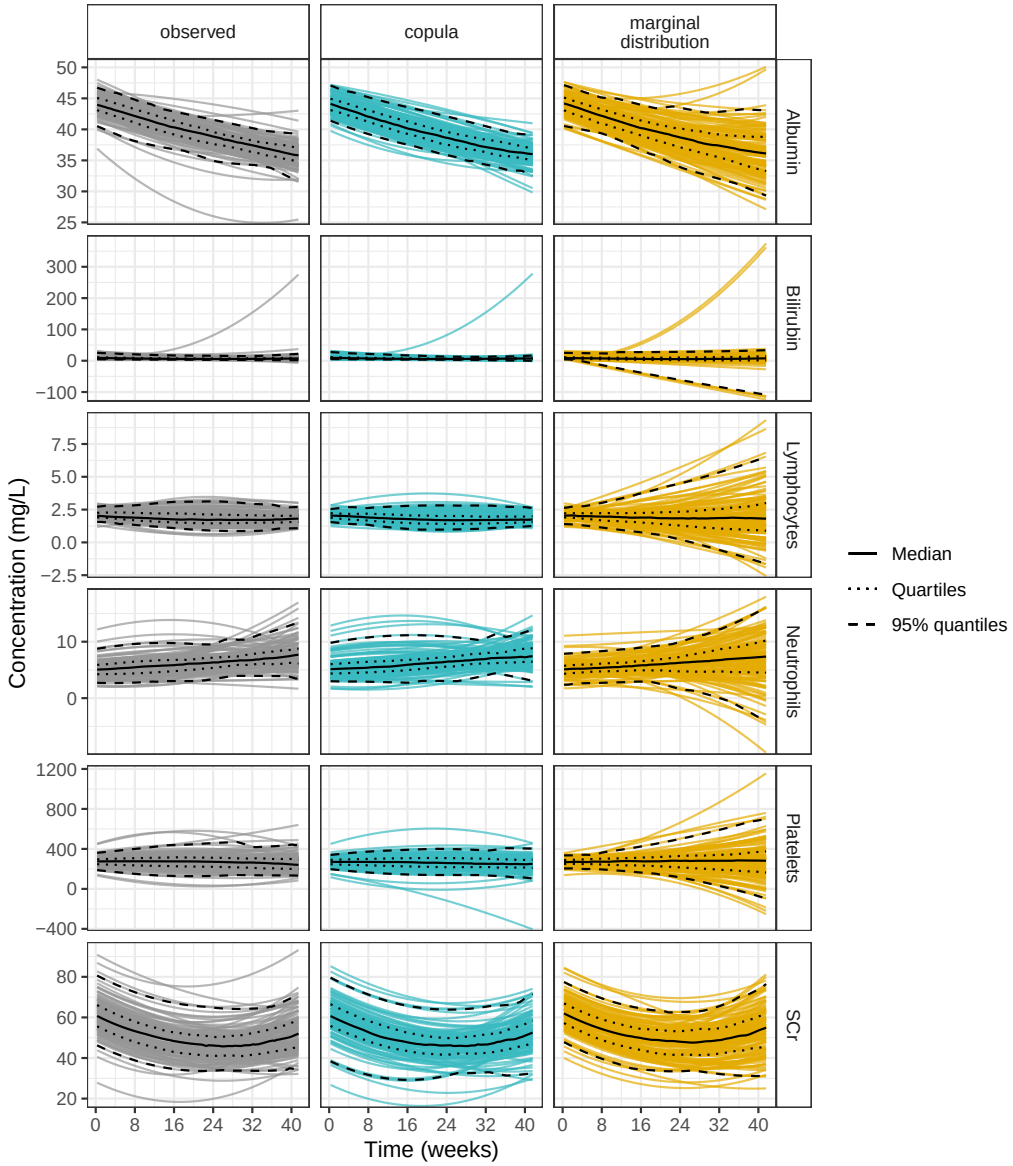


Figure 7.4: Polynomial curves for the six biomarkers from pregnancy data. In gray the estimated curves from the observed data. The copula (turquoise) shows very similar patterns, while the marginal distribution (yellow) shows extreme values, especially at the end of the curve.

0.11, which was slightly worse compared to the estimation in the lower dimensional twelve- and three-covariate data sets. Some covariates show interesting dependency structures, which can be evaluated and be used in covariate selection decision making (Figure 7.5). The results from the larger data set also show that through the use of copulas, it is feasible to share hospital data distributions.

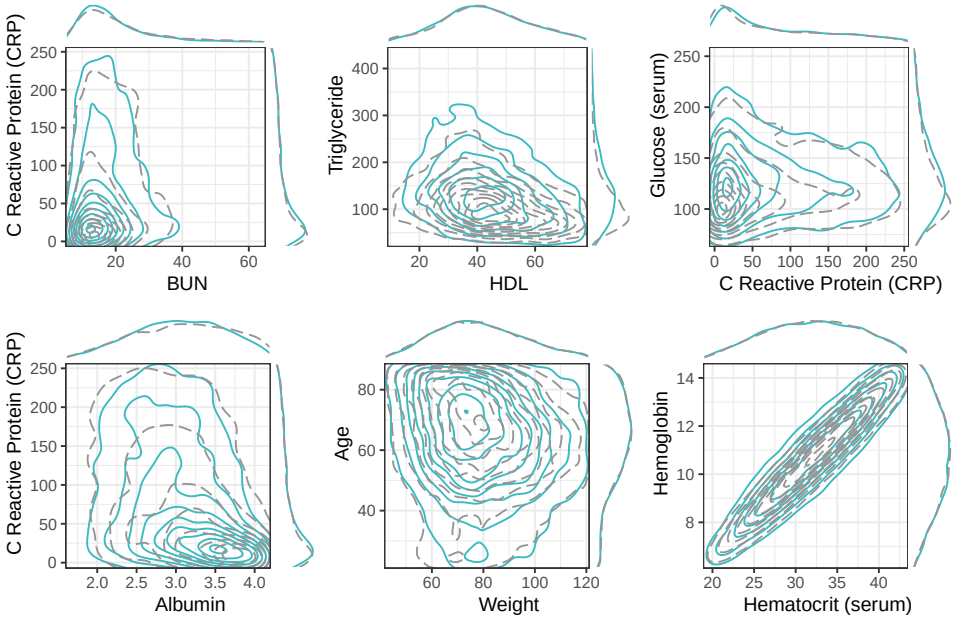


Figure 7.5: Set of selected covariate combinations with the densities of the observed population (gray dashed line) and the simulated population from a copula (blue solid lines), with marginal densities on the top and right sides of each plot. More overlap between the lines shows a better correspondence between the observed and simulated patient covariates.

7.4 Discussion

We showed a competitive or superior performance of copula simulations compared to other simulation methods, and we demonstrated multiple applications for covariate simulations using copulas. Copulas were able to preserve the correlations between covariates in lower and higher dimensional datasets. Preserving the dependence structure in copula simulations allows for simulating covariate sets for realistic PK predictions, time-varying covariates, and in a large scale data set, i.e., the MIMIC data, thus making it a suitable method for virtual patient covariate simulations in a variety of settings. Copula simulation has apparent benefits over currently used methods, since these either neglect the dependence structure among the covariates, or rely on real patient data in simulation.

We evaluated the performance of copulas compared to other simulation methods. While performing well in lower dimensions, we observed increasing underestimation in higher dimensions for CD, making the method less suitable for simulations in higher dimension. The MVND showed very promising results in terms of capturing the correlation (Figure 7.2). However, this is an inherent feature of how the MVND is estimated, which is based on the mean, standard deviation, and covariance. It does, on the other hand, not capture the actual shape of the distribution when covariates are not normally

distributed (Figure S7.1). Although the bootstrap can fully preserve the dependence structure between covariates, it cannot be used for simulation when actual data are unavailable. Additionally, due to the resampling nature of the bootstrap, one cannot simulate covariate values for virtual patients beyond which are present in the actual data set, which may result in simulating an unbalanced virtual patient population. The application of MDs was shown to simulate unrealistic patients, in the three situations studied.

Preserving the dependence between covariates is required for simulation of realistic patients in terms of PK predictions in the pediatrics vancomycin model, used in this study. The copula was able to preserve the relationship between the body weight and the AUC, which is of high clinical relevance. This feature of copulas provides a significant insight into how PK may differ between subgroups of patients. It allows one to optimize the dose for a particular patient group or to study the differences between patients groups. We found that PK at the population level is not affected by the method used for virtual patient simulation (Figure 7.3). The impact of preserving the dependence structure can differ per model, as can be seen in simulating the time-dependent covariates in the analysis of the pregnancy data. Here, polynomial regression coefficients need to be simulated in a realistic domain, in order to preserve the structure of the data, both on the individual and population level. Simulating from a marginal distributions lead to extreme polynomial curves.

Access to real individual-level patient data is often hampered by personal data protection regulations, which is a significant obstacle for community-driven design of optimized treatment strategies and trial designs (Conrado et al., 2017). Although copulas are mostly estimated on data, resulting copulas can be easily shared without sharing patient data, allowing one to use established copulas for virtual patient simulation (Gambs et al., 2021). Using copulas both opens opportunities for better replication and comparison studies, and copulas can facilitate in simulation platforms for sharing patient characteristics. The sharing of models has become more common in the pharmacometrics community, for example through platforms for model sharing, such as DDMoRe. However, models often require covariate input. Copulas can be used to set up a large scale covariate simulation platform, which can accompany the shared models to allow the clinical pharmacology community to simulate clinical trials and dosing regimens for (special) populations, even when there is no patient-level data available (Figure 7.6).

This paper did not address simulation of categorical variables. Discrete, ordered categorical and binary covariates can be captured as a copula, by using rank-based distributions (Czado & Nagler, 2022), however the copula method is not able to deal with unordered categorical variables in a natural way (Faugeras, 2017).

Regardless of the method of simulation, further research would also require looking into the underestimation of the correlation by the different simulation techniques, since there are limits to the full characterization of the joint distribution. Visualization of the simulation through density plots, allows to investigate how severe the discrepancy between the observed and population and the copula is and whether it seems clinically relevant. This can be evaluated on the level of the covariates, but also by

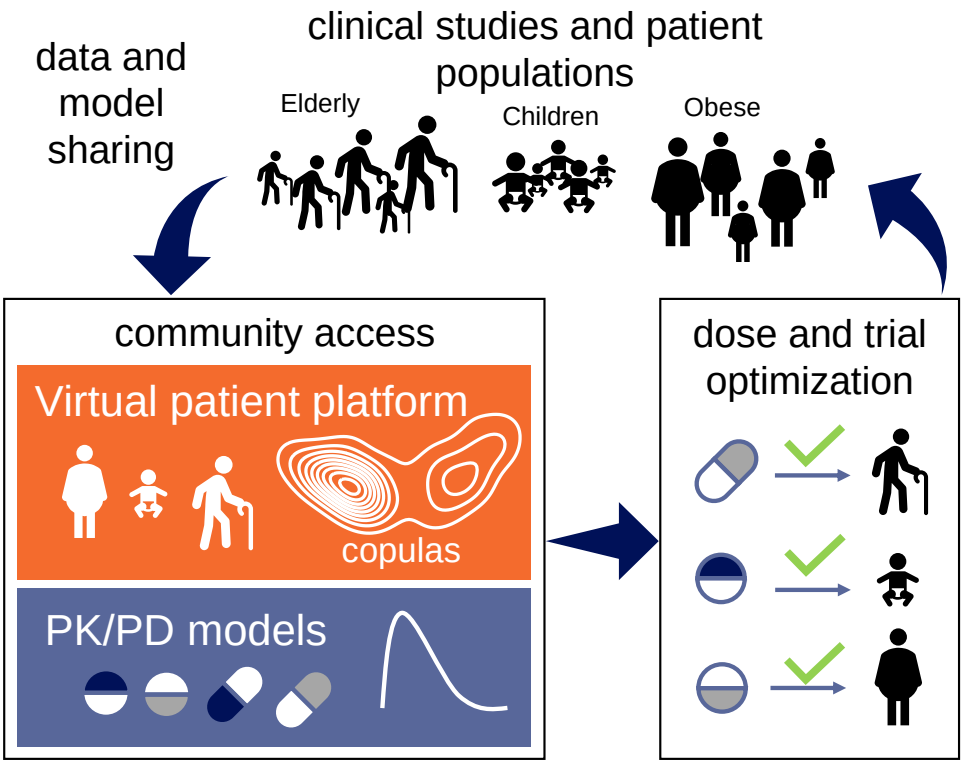


Figure 7.6: Community access pharmacometrics research pipeline. Data and pharmacometric models from (special) patient populations can be shared with the clinical pharmacology community. Through copulas, covariate sets can be simulated, which, when used in PK/PD models, can aid treatment and dosing optimization, ultimately improving treatment for the patients.

looking at the outcomes of pharmacometric models (Nguyen et al., 2017).

In summary, copulas represent an attractive approach to capture multivariate covariate distributions, which can be readily implemented for pharmacometric simulations. Importantly, the distribution-based nature of copula's has the distinct advantage that access to original individual-level datasets is not required when applied for virtual patient simulation, in contrast to resampling-based strategies. To this end, copula models can address hurdles in accessing real clinical data by developing open access simulation models for distinct (special) patient populations, which can be readily shared with the community and support clinical trial simulations and treatment optimization.

References

Bonate, P. L. (2000). Clinical trial simulation in drug development. *Pharmaceutical Research*, *17*(3), 252–256. doi: 10.1023/A:1007548719885

Bonate, P. L. (2001). A Brief Introduction to Monte Carlo Simulation. *Clinical Pharmacokinetics*, *40*(1), 15–22. Retrieved from http://link.springer.com/10.2165/00003088-200140010-00002 doi: 10.2165/00003088-200140010-00002

Chelliah, V., Lazarou, G., Bhatnagar, S., Gibbs, J. P., Nijsen, M., Ray, A., ... Kierzek, A. M. (2020, Aug). Quantitative systems pharmacology approaches for immuno-oncology: Adding virtual patients to the development paradigm. *Clinical Pharmacology & Therapeutics*, *109*(3), 605–618. Retrieved from https://doi.org/10.1002%2Fcpt.1987 doi: 10.1002/cpt.1987

Cock, R. F. W. D., Allegaert, K., Brussee, J. M., Sherwin, C. M. T., Mulla, H., de Hoog, M., ... Knibbe, C. A. J. (2014, May). Simultaneous pharmacokinetic modeling of gentamicin, tobramycin and vancomycin clearance from neonates to adults: Towards a semi-physiological function for maturation in glomerular filtration. *Pharmaceutical Research*, *31*(10), 2643–2654. Retrieved from https://doi.org/10.1007%2Fs11095-014-1361-z doi: 10.1007/s11095-014-1361-z

Conrado, D. J., Karlsson, M. O., Romero, K., Sarr, C., & Wilkins, J. J. (2017, Nov). Open innovation: Towards sharing of data, models and workflows. *European Journal of Pharmaceutical Sciences*, *109*, S65–S71. Retrieved from https://doi.org/10.1016%2Fj.ejps.2017.06.035 doi: 10.1016/j.ejps.2017.06.035

Czado, C. (2019). Simulating regular vine copulas and distributions. In *Analyzing dependent data with vine copulas* (pp. 123–144). Springer International Publishing. Retrieved from https://doi.org/10.1007%2F978-3-030-13785-4_6 doi: 10.1007/978-3-030-13785-4_6

Czado, C., & Nagler, T. (2022, Mar). Vine Copula Based Modeling. *Annual Review of Statistics and Its Application*, *9*(1), 453–477. Retrieved from https://www.annualreviews.org/doi/10.1146/annurev-statistics-040220-101153 doi: 10.1146/annurev-statistics-040220-101153

De Cock, P. A. J. G., Mulla, H., Desmet, S., De Somer, F., McWhinney, B. C., Ungerer, J. P. J., ... De Paepe, P. (2016, Dec). Population pharmacokinetics of cefazolin before, during and after cardiopulmonary bypass to optimize dosing regimens for children undergoing cardiac surgery. *Journal of Antimicrobial Chemotherapy*, *72*(3), dkw496. Retrieved from https://academic.oup.com/jac/article-lookup/doi/10.1093/jac/dkw496 doi: 10.1093/jac/dkw496

Efron, B. (1979, Jan). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, *7*(1). Retrieved from https://doi.org/10.1214%2Faos%2F1176344552 doi: 10.1214/aos/1176344552

Faugeras, O. P. (2017, Jan). Inference for copula modeling of discrete data: a cautionary tale and some facts. *Dependence Modeling*, *5*(1), 121–132. Retrieved from https://www.degruyter.com/document/doi/10.1515/demo-2017-0008/html doi: 10.1515/demo-2017-0008

Gambs, S., Ladouceur, F., Laurent, A., & Roy-Gaumond, A. (2021, Apr). Growing synthetic data through differentially-private vine copulas. *Proceedings on Privacy Enhancing Technologies*, *2021*(3), 122–141. Retrieved from https://doi.org/10.2478%2Fpopets-2021-0040 doi: 10.2478/popets-2021-0040

Grimsley, C., & Thomson, A. H. (1999, Nov). Pharmacokinetics and dose requirements of vancomycin in neonates. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *81*(3), F221–F227. Retrieved from https://doi.org/10.1136%2Ffn.81.3.f221 doi: 10.1136/fn.81.3.f221

Holford, N., Ma, S. C., & Ploeger, B. A. (2010, Jul). Clinical trial simulation: A review. *Clinical Pharmacology & Therapeutics*, *88*(2), 166–182. Retrieved from https://doi.org/10.1038%2Fclpt.2010.114 doi: 10.1038/clpt.2010.114

Illamola, S. M., Colom, H., & Hasselt, J. G. C. (2016, Sep). Evaluating renal function and age as predictors of amikacin clearance in neonates: model based analysis and optimal dosing strategies. *British Journal of Clinical Pharmacology*, *82*(3), 793–805. Retrieved from https://onlinelibrary.wiley.com/doi/10.1111/bcp.13016 doi: 10.1111/bcp.13016

Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2022). *MIMIC-IV*. PhysioNet. Retrieved from https://physionet.org/content/mimiciv/1.0/ doi: 10.13026/7vcr-e114

Langenhorst, J. B., Dorlo, T. P., Kesteren, C., Maarseveen, E. M., Nierkens, S., Witte, M. A., ... Huitema, A. D. (2020, Apr). Clinical trial simulation to optimize trial design for fludarabine dosing strategies in allogeneic hematopoietic cell transplantation. *CPT: Pharmacometrics & Systems Pharmacology*. Retrieved from https://doi.org/10.1002%2Fpsp4.12486 doi: 10.1002/psp4.12486

Nagler, T. (2017, May). Asymptotic analysis of the jittering kernel density estimator. *Mathematical Methods of Statistics*, *27*(1), 32–46. Retrieved from http://arxiv.org/abs/1705.05431 doi: 10.3103/S1066530718010027

Nagler, T., & Czado, C. (2016, Oct). Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *Journal of Multivariate Analysis*, *151*, 69–89. Retrieved from https://doi.org/10.1016%2Fj.jmva.2016.07.003 doi: 10.1016/j.jmva.2016.07.003

Nagler, T., & Vatter, T. (2020). kde1d: Univariate Kernel Density Estimation [Computer software manual]. Retrieved from https://cran.r-project.org/package=kde1d

Nagler, T., & Vatter, T. (2021). rvinecopulib: High Performance Algorithms for Vine Copula Modeling [Computer software manual]. Retrieved from https://cran.r-project.org/package=rvinecopulib

Nguyen, T. H., Mouksassi, M. S., Holford, N., Al-Huniti, N., Freedman, I., Hooker, A. C., ... Mentre, F. (2017). Model evaluation of continuous data pharmacometric models: Metrics and graphics. *CPT: Pharmacometrics and Systems Pharmacology*, *6*(2), 87–109. doi: 10.1002/psp4.12161

Patel, J. P., Green, B., Patel, R. K., Marsh, M. S., Davies, J. G., & Arya, R. (2013, Sep). Population pharmacokinetics of enoxaparin during the antenatal period. *Circulation*, *128*(13), 1462–1469. Retrieved from https://doi.org/10.1161%2Fcirculationaha.113.003198 doi: 10.1161/circulationaha.113.003198

Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika*, *9*(6), 449–460.

Smania, G., & Jonsson, E. N. (2021, Apr). Conditional distribution modeling as an alternative method for covariates simulation: Comparison with joint multivariate normal and bootstrap techniques. *CPT: Pharmacometrics & Systems Pharmacology*, *10*(4), 330–339. Retrieved from https://doi.org/10.1002%2Fpsp4.12613 doi: 10.1002/psp4.12613

Tannenbaum, S. J., Holford, N. H. G., Lee, H., Peck, C. C., & Mould, D. R. (2006, Oct). Simulation of correlated continuous and categorical variables using a single multivariate distribution. *Journal of Pharmacokinetics and Pharmacodynamics*, *33*(6), 773–794. Retrieved from https://doi.org/10.1007%2Fs10928-006-9033-1 doi: 10.1007/s10928-006-9033-1

Teutonico, D., Musuamba, F., Maas, H. J., Facius, A., Yang, S., Danhof, M., & Pasqua, O. D. (2015, May). Generating virtual patients by multivariate and discrete re-sampling techniques. *Pharmaceutical Research*, *32*(10), 3228–3237. Retrieved from https://doi.org/10.1007%2Fs11095-015-1699-x doi: 10.1007/s11095-015-1699-x

van Hasselt, J. G., van Calsteren, K., Heyns, L., Han, S., Mhallem Gziri, M., Schellens, J. H., ... Amant, F. (2014). Optimizing anticancer drug treatment in pregnant cancer patients: pharmacokinetic analysis of gestation-induced changes for doxorubicin, epirubicin, docetaxel and paclitaxel. *Annals of Oncology*, *25*(10), 2059–2065. Retrieved from https://doi.org/10.1093/annonc/mdu140 doi: 10.1093/annonc/mdu140

van Hasselt, J. G. C., Allegaert, K., van Calsteren, K., Beijnen, J. H., Schellens, J. H. M., & Huitema, A. D. R. (2014). Semiphysiological versus Empirical Modelling of the Population Pharmacokinetics of Free and Total Cefazolin during Pregnancy. *BioMed Research International*, *2014*, 1–9. Retrieved from http://www.hindawi.com/journals/bmri/2014/897216/ doi: 10.1155/2014/897216

van Hasselt, J. G. C., van Eijkelenburg, N. K. A., Beijnen, J. H., Schellens, J. H., & Huitema, A. D. R. (2014, Dec). Design of a drug-drug interaction study of vincristine with azole antifungals in pediatric cancer patients using clinical trial simulation. *Pediatric Blood & Cancer*, *61*(12), 2223–2229. Retrieved from https://onlinelibrary.wiley.com/doi/10.1002/pbc.25198 doi: 10.1002/pbc.25198

Vinks, A., Emoto, C., & Fukuda, T. (2015, Jul). Modeling and simulation in pediatric drug therapy: Application of pharmacometrics to define the right dose for children. *Clinical Pharmacology & Therapeutics*, *98*(3), 298–308. Retrieved from https://doi.org/10.1002%2Fcpt.169 doi: 10.1002/cpt.169

Yoneyama, K., Schmitt, C., Kotani, N., Levy, G. G., Kasai, R., Iida, S., ... Kawanishi, T. (2017, Dec). A pharmacometric approach to substitute for a conventional dose-finding study in rare diseases: Example of phase III dose selection for emicizumab in hemophilia a. *Clinical Pharmacokinetics*, *57*(9), 1123–1134. Retrieved from https://doi.org/10.1007%2Fs40262-017-0616-3 doi: 10.1007/s40262-017-0616-3

Supplementary material

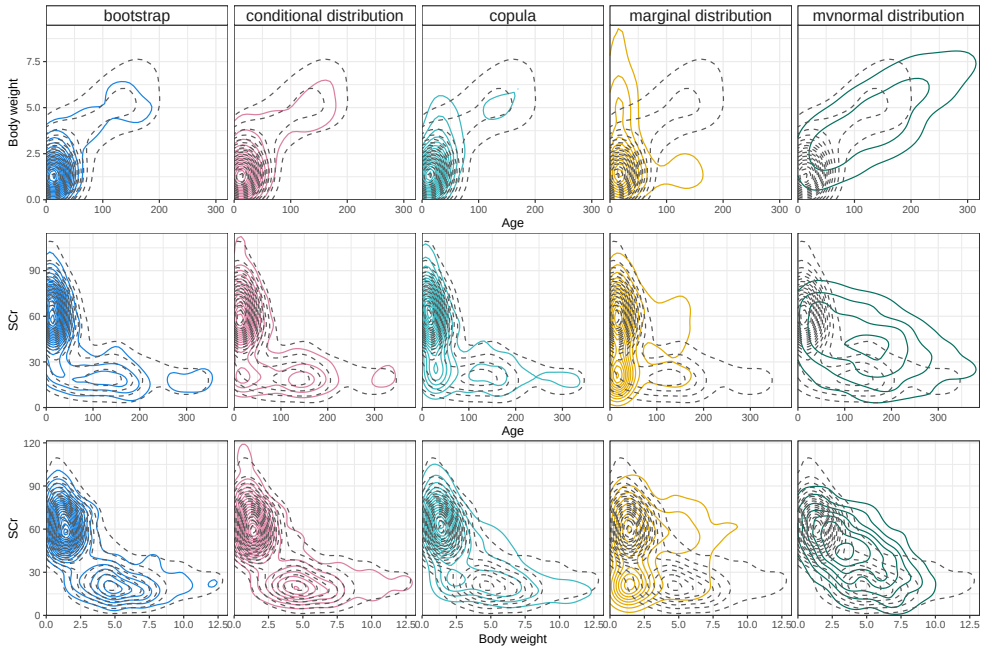


Figure S7.1: Densities of the three covariate simulations. Grey dashed lines show the observed joint density for each pair of covariates. The solid lines represent the joint density of a simulated population for each of the five simulation methods: bootstrap (blue), conditional distributions (pink), copula (turquoise), marginal distribution (yellow) and multivariate normal distribution (green).

Chapter 8

General discussion and summary

General discussion and summary

Explaining treatment response variability between and within patients can support treatment and dosing optimization, to improve treatment of individual patients. This thesis discussed multiple aspects of treatment variability and the associated statistical learning techniques which can be used to explain and/or predict part of that variability. Even though in recent times the availability of several high-throughput measurement technologies has created many new opportunities to develop improved treatment strategies, deriving actionable insights from such data remains a challenge (Section I). To this end, the use of longitudinal and high-dimensional data analysis techniques is needed to explore omics data for explaining treatment response and clinical course (Section II), and to answer clinical questions from routine healthcare data from hospitals and research institutes (Section III).

8.1 Data science in pharmaceutical research

To gain knowledge about the treatment response, clinical trials are a golden standard, but not all factors and not all patient populations can be included in clinical trials. Inclusion of data collected as part of routine health check-ups or from wearables and home devices could improve treatment decision making (Morrato et al., 2007; Swift et al., 2018). In Chapter 2, we described the additional types of data that could facilitate clinical decision making, through placebo-responder prediction, endpoint and biomarker discovery and prognosis and drug response prediction, and which opportunities and pitfalls these data introduce. We focused on pediatric patients, due to the difficulties in recruiting patients and the large individual variability, which call for research complementary to clinical trials (Brussee et al., 2016).

Real world data cannot be analyzed in the same way as classical clinical trial data, which calls for adaptation or extension of statistical methods used for pharmacological research. Machine learning methods, mostly statistical learning techniques, are considered for these types of data. This poses an additional challenge, because an important part of clinical studies is the interpretability of methods used, due to the need to be able to explain the choice of clinical decisions (Knoppers & Thorogood, 2017).

8.2 High-dimensional biomarker discovery

Due to the developments in biochemical measurement techniques, the molecular make-up of a human can be measured, for example in serum blood samples. Data about molecules such as metabolites, RNA and DNA, generally referred to collectively as 'omics' data, can nowadays be measured, greatly increasing the precision with which patients can be described (Pearson, 2016). Although these omics data give abundantly more measurements, discovery of influential or predictive biomarkers (Depledge et al., 1993) from omics data adds a number of data analysis challenges,

especially in the case of studying changes over time. Working with both longitudinal and high-dimensional data makes current available methods that often focus on only one of the two, hard to use directly.

In Chapter 3, we identified biomarkers from high-dimensional genomics data for the tumor treatment response in patient-derived xenografts, using a novel two-step approach (Zwep, Haakman, et al., 2021). The data were retrieved from published research (Gao et al., 2015), where the tumor growth of mice models was measured over time, and genomic data, in the form of copy number variations (CNV's), were measured at baseline. In the first step, we used a mathematical tumor growth inhibition model to describe the longitudinal tumor growth curves. This model characterizes a tumor growth curve with three parameters: the tumor growth (k_g), the drug effect (k_d) and the drug resistance development (k_r). These parameters were estimated based on the data, on which both a population effect and the individual effects (empirical Bayes estimates) were estimated. The individual parameter estimates were used in a second step, as outcomes in lasso regression, where they were related to the high-dimensional genomic data. Using cross-validation, we found a 4% median decrease in prediction error, by including genomic data. We were able to detect genomic effects on a pathway level, by using a pathway-informed group-lasso.

High-dimensional data pose a challenge, due to easily overfitting on the data. Although tumor growth inhibition models are readily available, pharmacometric estimation techniques currently are not able to estimate high-dimensional covariate effects for prediction and using the proposed two-step approach circumvents the computational difficulty. A two-step approach can, however, cause inflated errors: if error is introduced in the first step, this erroneous estimate is used in the second step. So assessing errors in both steps is important to reduce this risk.

In Chapter 4, longitudinal metabolomics data were studied to explore potential biomarkers for clinical course, the combination of treatment response and disease development, in hospitalized patients with community acquired pneumonia (CAP). We applied dimension reduction through principal component analysis, to explore patterns of metabolites over time and how different biochemical classes of metabolites relate to the clinical course. We calculated correlations between metabolites and two measures of clinical course: the CURB score, a score indicating how sick the patient is when entering the hospital, and the length of stay in the hospital, indicating how much time the patient took to recover. Metabolite patterns clearly changed over time within the patients, showing how important studying longitudinal metabolomics data in patients with CAP is. Several biochemical classes were identified that were correlated to the clinical course, such as the triglyceride and the lysophosphatidylcholine classes.

8.2.1 A path towards omics-related treatment individualization

Characterizing patients on a molecular level can potentially improve understanding of different treatment response and clinical course. However, these type of high-dimensional data pose a challenge in data analysis. The problem of sparsity in high-

dimensional data can be tackled using different statistical methods. Pharmacometrics was also developed to deal with sparsity, in terms of number of patients and time points, to study dose response in sparse data (Pillai et al., 2005). In Chapter 3 we utilized pathway knowledge to reduce the dimension of possible solutions, by penalizing on a pathway level. Utilizing prior knowledge, such as understanding of dose response curves, or in the present case, relations among genes, can reduce the dimensions that are irrelevant for the research question (van Nee et al., 2021).

Many omics technologies are currently shown to be promising, but most are not used in practice for the individualization of drug treatment. The development of metabolomics measurement technology allows for a very low-level characterization of the patient's physiological processes (Beger et al., 2016). However, both the measurement of metabolites and the high variability of concentrations, make it hard to distinguish between-individual variance from within-individual variance. In Chapter 4, we address this issue by analyzing changes over time, instead of measurements at one time point.

8.3 Real world data

Routine healthcare data are often to monitor patients and their treatment response, and a growing part of these data are stored and accessible for researchers. These real world data can improve our understanding of topics in pharmaceutical science.

Antibiotic resistance poses a threat to global health. How big of a threat and how it is developing is continuously monitored through surveillance of antibiotic resistance in hospitals (van der Kuil et al., 2017). Due to increased monitoring of antibiotic resistance, minimal inhibitory concentrations (MIC) of clinical bacterial strains are often measured, which could be used for detection of collateral sensitivity (CS). CS occurs when one drug can reduce resistance against a second drug, and could be useful in combatting or overcoming infections with resistant pathogens (Aulin et al., 2021). Although the phenomenon has been detected in the lab, knowledge of its prevalence in clinical practice is currently very limited.

In Chapter 5, we proposed a method for quantification of collateral effects in routine healthcare MIC data (Zwep, Haakman, et al., 2021). The proposed log2 fold change is an interpretable measure also used in experimental research, allowing easy comparison between experimental and observational results and enabling directionality between two antibiotics. This measure was used in Chapter 6 to quantify CS in large MIC data from different data sources, indicating CS is occurring in clinical practice, but it is very hard to find specific patterns over different antibiotic classes or species, and this lack of generalizability makes it hard to use collateral sensitivity in clinical studies and in practice (Nichol et al., 2019).

Next to healthcare data on pathogens, patient-level data are also collected during routine patient care. These data contain different types of covariates, such as biomarker concentrations, demographics, and other patient characteristics (Currie & MacDonald, 2000). In pharmacometric modeling, these characteristics are often used

to explain and predict inter-patient variability in pharmacokinetics and pharmacodynamics, with the possibility to extent to special patient populations. Pharmacometric simulations require the simulation of these patient covariates, but sharing these sensitive patient data between hospitals and research groups is often difficult, due to the protection of the patients' privacy.

In Chapter 7, we proposed the use of copulas as a suitable method for virtual patient simulation. Copulas are multivariate distribution functions that can capture joint distributions and provide a flexible way to describe and simulate patient covariate sets from these densities. Most covariates are not independent of each other, so modeling this dependency adequately is required for the simulation of realistic virtual patients and a joint distribution function captures this dependency between different covariates. Our study showed copulas are able to simulate realistic patient populations using copulas (Zwep et al., 2022). Realistic virtual patients are required to simulate different patient populations, enabling extrapolation of found results to specific patient populations of interest.

The copulas need to be estimated on the basis of data that are not always available to researchers. However, data collected by hospitals and research institutes can be used to estimate the copulas and these copulas can be shared with researchers to enable studying the population, without granting access to the underlying patient level data. This way, covariates of patient populations can be shared between researchers and hospitals, without concern about the privacy of the patients.

8.4 Perspectives and conclusions

8.4.1 Integration of statistics and pharmacometrics

Pharmacometrics and statistics, although two very related fields, have developed separately throughout large parts of their history. Pharmacokinetic and pharmacodynamic models are based on concentration and effect over time profiles, respectively. Development and studies of these models generally require knowledge about or data of concentrations, system-parameters and drug-specific parameters. Modeling is usually done using nonlinear mixed effect models, a statistical framework, very common in pharmacometrics, but not in many other fields of statistics, where (generalized) linear mixed effect models are more commonly used (McCulloch & Searle, 2000; Pillai et al., 2005).

Nonlinear mixed effect models are very useful for describing the usually nonlinear relations between time and drug concentrations and/or effects. They offer a solution to the longitudinal and unbalanced nature of the data to estimate parameter values and predict treatment responses (Pillai et al., 2005). Recently, the inclusion of high-dimensional data as potential explanatory variables in pharmacometric research has compelled the pharmacometrics community to involve other statistical methods, which are able to deal with this high-dimensionality.

Next to statistical learning methods for high-dimensional data analysis, such as

regularization methods and dimension reduction, artificial intelligence methods, such as (deep) neural networks, have recently gained a lot of popularity in pharmacometrics (Chaturvedula et al., 2019; Janssen et al., 2022; McComb et al., 2021). However, actual implementation of both statistical learning and artificial intelligence approaches remains a challenge (Knights & Ramanathan, 2016).

In this thesis we proposed a two-step method for identifying high-dimensional biomarkers in a tumor growth model, showing how to incorporate nonlinear mixed effect models with the lasso method (Zwep, Duisters, et al., 2021). Next to this combination of pharmacometrics and statistics, we also explored the use of copulas to facilitate virtual patient covariate sets. Both projects aimed to improve current pharmacometric practices. The goal is not to implement new methods of data science, the goal is to better predict clinical outcomes and to be able to optimize treatments. Integrating statistical methods and pharmacometrics is not a goal on its own, although it sometimes seems to be treated that way, introducing and using overcomplicated machine learning methods, while other techniques are readily available (Volovici et al., 2022).

Although integration of statistics and pharmacometrics is a natural way of expanding the types of research and data analysis possible, to improve treatment optimization, more research is needed to find the ways of integrating these two, while keeping the end goal in mind. (van der Kuil et al., 2017)

8.4.2 The importance of interpretability in pharmacology

When data and data analysis become more complex, sometimes interpreting predictions, metrics and underlying parameter values also increases in complexity. This affects the way science is conducted in different ways: in terms of effect size interpretation, understanding of clinical decision making, and biological understanding of the system.

Interpretable measures, such as the collateral sensitivity measure proposed in Chapter 5, facilitate the translation between experimental results and clinical observations, and can help to understand clinical relevance through interpreting the effect size (Zwep, Haakman, et al., 2021).

In Chapter 2, we briefly discussed the importance of interpretable clinical decision making, due to the need for physicians to understand the reasoning behind their decisions (Goulooze et al., 2020). Understanding the underlying decision making process of clinical advise is important for healthcare professionals to understand whether they should follow the advice, or treat differently in a specific case. An emerging field concerned with this problem is explainable artificial intelligence, where 'black box algorithms', models without interpretable model and parameter values, are extended with a method to show what the predictions are based on (Doshi-Velez & Kim, 2018; Xu et al., 2019). However, current discussions are going on about the usefulness of these explanations in individual clinical decisions, deeming current techniques unviable for clinical decision making (Ghassemi et al., 2021).

Interpretability can facilitate in the understanding of the underlying biological sys-

tems. By understanding which factors and processes drive a treatment or disease effect, it is possible to study new drugs. Especially omics research focusses on describing the biological system on a molecular level, to track changes causing or caused by the disease in order to counteract them (Perakakis et al., 2018). Biological understanding can be improved by studying sets of molecules, but can also be retrieved through looking at biological pathways or biochemical classes of molecules, giving a more high-level understanding of the biological processes.

Improved biological understanding allows more accurate modeling of biological systems, which is what is done in translational pharmacometrics through physiological understanding of the processes in drug responses. If a model captures the physiological processes well, it becomes possible to extrapolate predictions to new medication or populations (Agoram et al., 2007; Musante et al., 2016; Pérez-Nueno, 2015).

8.4.3 Generalizing results

Knowledge of mechanisms and patient responses becomes useful if the results of a study can be generalized outside the region of the data analyzed. In first instance, exploratory research is needed to obtain insight in which variables might be of importance for treatment responses, such as the metabolites that might be of interest for the clinical course in longitudinal CAP patients (Chapter 4). However, generalization requires a different framework (Leek & Peng, 2015). Two distinct aims for generalization have been formulated in Chapter 3: the understanding of biological mechanisms contributing to variable treatment responses and the prediction of tumor growth inhibition for different treatments. These two aims can more generally be described as inference and prediction respectively, and are not fully separable in terms of research aims (Bzdok & Ioannidis, 2019).

In inference, the effect size is of interest, for example the correlation between variables or the difference between groups. Inference includes methods of parameter estimation and hypothesis testing to quantify effects and test whether these effects are expected to be 'real'. By having a model of the world that captures the underlying data generating process, it is possible to infer a mechanism in a more general population, such as the effect of a treatment on a clinical outcome (Bzdok & Ioannidis, 2019). Hypothesis testing is based on this principle. With the increase of complexity and size of data(sets), more variables and correlations are of interest at once. Evaluating the generalizability of the found results requires an extra step in hypothesis testing. In Chapter 5 and Chapter 6, we evaluated large numbers of tests, by estimating and testing all combinations of antibiotics. By using a multiple testing correction, we controlled the probability of false hypothesis rejections.

Prediction is another aim in research, where not the effect size, but a specific patient response is of interest. Generalizability is ensured in prediction by validation, through comparison between observed values and predicted values in a sample that has not been used to make the underlying prediction. In prediction, the high-dimensional setting often causes overfitting, being able to predict the outcomes for the data that produced the underlying model really well, but not being able to predict

newly observed data. In Chapter 3, lasso regression was used both for inference and prediction (Tibshirani, 1996). The lasso uses a hyperparameter to determine shrinkage, which can be tuned through cross-validation, to reduce the risk of overfitting.

Despite the use of multiple testing correction or cross-validation based hyperparameter tuning, validation is important to the progress of science in both estimation of treatment effects and prediction of patient's individual treatment responses (Ghosh & Poisson, 2009). The research cycle of pharmacometrics is based on this principle, by modeling pharmacological processes, based on previous knowledge and data, and validating the predictive performance with new data. When the pharmacological model is established, it can be used for patient predictions and even population extrapolation (Marshall et al., 2016).

In the field of precision medicine and - more specifically - omics research, generalization is hard to achieve, because of the increased dimensionality and the explorative nature of omics research. It is important to involve experts and robust data analysis strategies to avoid overfitting and to gain a good understanding of pharmacology and biology (Buyse et al., 2010; Wilkinson et al., 2020).

8.4.4 Data sharing opportunities

The adaptation of statistical tools often relies on the use of available data sources, which our research on biomarkers for tumor growth inhibition (Chapter 3) and collateral sensitivity (Chapter 5, Chapter 6) is also based on. There is a call to increase data sharing to improve and accelerate research (Hulsen, 2020). Data sharing is a particularly difficult topic within healthcare, due to the intricate privacy issues and the laws protecting privacy rights (Knoppers & Thorogood, 2017). Developing ways to share data, while preserving privacy is an active field of research (Bonomi et al., 2020; Sweeney, 2002). One way to share information, is by sharing summary measures of cohorts, instead of the patient level data. In Chapter 7, we proposed to do this by using copulas to share information on a population level, while preserving the privacy of individual patients (Gambs et al., 2021). More data sharing can support precision medicine develop, but these data should be handled with caution, and methods like the copula could support this aim.

8.5 Conclusions

The use of statistical learning methods in precision medicine supports unraveling treatment resonse variability, using different types of data, such as high-dimensional omics data, but also routine healthcare data. Integration of statistical learning methods facilitates further pharmacological research, but care needs to be taken to keep a clear pharmacological interpretation of the results.

References

Agoram, B. M., Martin, S. W., & van der Graaf, P. H. (2007, Dec). The role of mechanism-based pharmacokinetic–pharmacodynamic (PK–PD) modelling in translational research of biologics. *Drug Discovery Today*, *12*(23-24), 1018–1024. Retrieved from https://doi.org/10.1016%2Fj.drudis.2007.10.002 doi: 10.1016/j.drudis.2007.10.002

Aulin, L. B. S., Liakopoulos, A., van der Graaf, P. H., Rozen, D. E., & van Hasselt, J. G. C. (2021, Sep). Design principles of collateral sensitivity-based dosing strategies. *Nature Communications*, *12*(1). Retrieved from https://doi.org/10.1038%2Fs41467-021-25927-3 doi: 10.1038/s41467-021-25927-3

Beger, R. D., Dunn, W., Schmidt, M. A., Gross, S. S., Kirwan, J. A., Cascante, M., ... Kaddurah-Daouk, R. (2016, Sep). Metabolomics enables precision medicine: "A White Paper, Community Perspective". *Metabolomics*, *12*(9), 149. Retrieved from http://link.springer.com/10.1007/s11306-016-1094-6 doi: 10.1007/s11306-016-1094-6

Bonomi, L., Huang, Y., & Ohno-Machado, L. (2020, Jun). Privacy challenges and research opportunities for genomic data sharing. *Nature Genetics*, *52*(7), 646–654. Retrieved from https://doi.org/10.1038%2Fs41588-020-0651-0 doi: 10.1038/s41588-020-0651-0

Brussee, J. M., Calvier, E. A. M., Krekels, E. H. J., Välitalo, P. A. J., Tibboel, D., Allegaert, K., & Knibbe, C. A. J. (2016, Jun). Children in clinical trials: towards evidence-based pediatric pharmacotherapy using pharmacokinetic-pharmacodynamic modeling. *Expert Review of Clinical Pharmacology*, *9*(9), 1235–1244. Retrieved from https://doi.org/10.1080%2F17512433.2016.1198256 doi: 10.1080/17512433.2016.1198256

Buyse, M., Sargent, D. J., Grothey, A., Matheson, A., & de Gramont, A. (2010, Apr). Biomarkers and surrogate end points—the challenge of statistical validation. *Nature Reviews Clinical Oncology*, *7*(6), 309–317. Retrieved from https://doi.org/10.1038%2Fnrclinonc.2010.43 doi: 10.1038/nrclinonc.2010.43

Bzdok, D., & Ioannidis, J. P. (2019, Apr). Exploration, inference, and prediction in neuroscience and biomedicine. *Trends in Neurosciences*, *42*(4), 251–262. Retrieved from https://doi.org/10.1016%2Fj.tins.2019.02.001 doi: 10.1016/j.tins.2019.02.001

Chaturvedula, A., Calad-Thomson, S., Liu, C., Sale, M., Gattu, N., & Goyal, N. (2019, Jun). Artificial intelligence and pharmacometrics: Time to embrace, capitalize, and advance? *CPT: Pharmacometrics & Systems Pharmacology*, *8*(7), 440–443. Retrieved from https://doi.org/10.1002%2Fpsp4.12418 doi: 10.1002/psp4.12418

Currie, C. J., & MacDonald, T. M. (2000). Use of routine healthcare data in safe and cost-effective drug use. *Drug Safety*, *22*(2), 97–102. Retrieved from https://doi.org/10.2165%2F00002018-200022020-00002 doi: 10.2165/00002018-200022020-00002

Depledge, M. H., Amaral-Mendes, J. J., Daniel, B., Halbrook, R. S., Kloepper-Sams, P., Moore, M. N., & Peakall, D. B. (1993). The conceptual basis of the biomarker approach. In *Biomarkers* (pp. 15–29). Springer Berlin Heidelberg. Retrieved from https://doi.org/10.1007%2F978-3-642-84631-1_2 doi: 10.1007/978-3-642-84631-1_2

Doshi-Velez, F., & Kim, B. (2018). Considerations for evaluation and generalization in interpretable machine learning. In *The springer series on challenges in machine learning* (pp. 3–17). Springer International Publishing. Retrieved from https://doi.org/10.1007%2F978-3-319-98131-4_1 doi: 10.1007/978-3-319-98131-4_1

Gambs, S., Ladouceur, F., Laurent, A., & Roy-Gaumond, A. (2021, Apr). Growing synthetic data through differentially-private vine copulas. *Proceedings on Privacy Enhancing Technologies*, *2021*(3), 122–141. Retrieved from https://doi.org/10.2478%2Fpopets-2021-0040 doi: 10.2478/popets-2021-0040

Gao, H., Korn, J. M., Ferretti, S., Monahan, J. E., Wang, Y., Singh, M., ... Sellers, W. R. (2015, Oct). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nature Medicine*, *21*(11), 1318–1325. Retrieved from https://doi.org/10.1038%2Fnm.3954 doi: 10.1038/nm.3954

Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021, Nov). The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health*, *3*(11), e745–e750. Retrieved from https://doi.org/10.1016%2Fs2589-7500%2821%2900208-9 doi: 10.1016/s2589-7500(21)00208-9

Ghosh, D., & Poisson, L. M. (2009, Jan). "omics" data and levels of evidence for biomarker discovery. *Genomics*, *93*(1), 13–16. Retrieved from https://doi.org/10.1016%2Fj.ygeno.2008.07.006 doi: 10.1016/j.ygeno.2008.07.006

Goulooze, S. C., Zwep, L. B., Vogt, J. E., Krekels, E. H., Hankemeier, T., Anker, J. N., & Knibbe, C. A. (2020, Apr). Beyond the Randomized Clinical Trial: Innovative Data Science to Close the Pediatric Evidence Gap.

Clinical Pharmacology & Therapeutics, *107*(4), 786–795. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/cpt.1744https://onlinelibrary.wiley.com/doi/10.1002/cpt.1744 doi: 10.1002/cpt.1744

Hulsen, T. (2020, Apr). Sharing is caring—data sharing initiatives in healthcare. *International Journal of Environmental Research and Public Health*, *17*(9), 3046. Retrieved from https://doi.org/10.3390%2Fijerph17093046 doi: 10.3390/ijerph17093046

Janssen, A., Bennis, F. C., & Mathôt, R. A. A. (2022, Aug). Adoption of machine learning in pharmacometrics: An overview of recent implementations and their considerations. *Pharmaceutics*, *14*(9), 1814. Retrieved from https://doi.org/10.3390%2Fpharmaceutics14091814 doi: 10.3390/pharmaceutics14091814

Knights, J., & Ramanathan, M. (2016). Detecting pharmacokinetic and pharmacodynamic covariates from high-dimensional data. In *Systems pharmacology and pharmacodynamics* (pp. 277–301). Springer International Publishing. Retrieved from https://doi.org/10.1007%2F978-3-319-44534-2_13 doi: 10.1007/978-3-319-44534-2_13

Knoppers, B. M., & Thorogood, A. M. (2017, Aug). Ethics and big data in health. *Current Opinion in Systems Biology*, *4*, 53–57. Retrieved from https://doi.org/10.1016%2Fj.coisb.2017.07.001 doi: 10.1016/j.coisb.2017.07.001

Leek, J. T., & Peng, R. D. (2015, Mar). What is the question? *Science*, *347*(6228), 1314–1315. Retrieved from https://doi.org/10.1126%2Fscience.aaa6146 doi: 10.1126/science.aaa6146

Marshall, S., Burghaus, R., Cosson, V., Cheung, S., Chenel, M., DellaPasqua, O., ... Visser, S. (2016, Mar). Good practices in model-informed drug discovery and development: Practice, application, and documentation. *CPT: Pharmacometrics & Systems Pharmacology*, *5*(3), 93–122. Retrieved from https://doi.org/10.1002%2Fpsp4.12049 doi: 10.1002/psp4.12049

McComb, M., Bies, R., & Ramanathan, M. (2021, Mar). Machine learning in pharmacometrics: Opportunities and challenges. *British Journal of Clinical Pharmacology*, *88*(4), 1482–1499. Retrieved from https://doi.org/10.1111%2Fbcp.14801 doi: 10.1111/bcp.14801

McCulloch, C. E., & Searle, S. R. (2000). *Generalized, linear, and mixed models*. Wiley. Retrieved from https://doi.org/10.1002%2F0471722073 doi: 10.1002/0471722073

Morrato, E. H., Elias, M., & Gericke, C. A. (2007, Dec). Using population-based routine data for evidence-based health policy decisions: lessons from three examples of setting and evaluating national health policy in australia, the UK and the USA. *Journal of Public Health*, *29*(4), 463–471. Retrieved from https://doi.org/10.1093%2Fpubmed%2Ffdm065 doi: 10.1093/pubmed/fdm065

Musante, C., Ramanujan, S., Schmidt, B., Ghobrial, O., Lu, J., & Heatherington, A. (2016, Nov). Quantitative systems pharmacology: A case for disease models. *Clinical Pharmacology & Therapeutics*, *101*(1), 24–27. Retrieved from https://doi.org/10.1002%2Fcpt.528 doi: 10.1002/cpt.528

Nichol, D., Rutter, J., Bryant, C., Hujer, A. M., Lek, S., Adams, M. D., ... Scott, J. G. (2019, Jan). Antibiotic collateral sensitivity is contingent on the repeatability of evolution. *Nature Communications*, *10*(1). Retrieved from https://doi.org/10.1038%2Fs41467-018-08098-6 doi: 10.1038/s41467-018-08098-6

Pearson, E. R. (2016, May). Personalized medicine in diabetes: the role of 'omics' and biomarkers. *Diabetic Medicine*, *33*(6), 712–717. Retrieved from https://doi.org/10.1111%2Fdme.13075 doi: 10.1111/dme.13075

Perakakis, N., Yazdani, A., Karniadakis, G. E., & Mantzoros, C. (2018, Oct). Omics, big data and machine learning as tools to propel understanding of biological mechanisms and to discover novel diagnostics and therapeutics. *Metabolism*, *87*, A1–A9. Retrieved from https://doi.org/10.1016%2Fj.metabol.2018.08.002 doi: 10.1016/j.metabol.2018.08.002

Pérez-Nueno, V. I. (2015, Aug). Using quantitative systems pharmacology for novel drug discovery. *Expert Opinion on Drug Discovery*, *10*(12), 1315–1331. Retrieved from https://doi.org/10.1517%2F17460441.2015.1082543 doi: 10.1517/17460441.2015.1082543

Pillai, G. C., Mentré, F., & Steimer, J.-L. (2005, Apr). Non-linear mixed effects modeling – from methodology and software development to driving implementation in drug development science. *Journal of Pharmacokinetics and Pharmacodynamics*, *32*(2), 161–183. Retrieved from https://doi.org/10.1007%2Fs10928-005-0062-y doi: 10.1007/s10928-005-0062-y

Sweeney, L. (2002, Oct). k-Anonymity: a Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(05), 557–570. Retrieved from https://www.worldscientific.com/doi/abs/10.1142/S0218488502001648 doi: 10.1142/S0218488502001648

Swift, B., Jain, L., White, C., Chandrasekaran, V., Bhandari, A., Hughes, D. A., & Jadhav, P. R. (2018, May). Innovation at the intersection of clinical trials and real-world data science to advance patient care. *Clinical and Translational Science*, *11*(5), 450–460. Retrieved from https://doi.org/10.1111%2Fcts.12559 doi: 10.1111/cts.12559

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Royal Statistical Society*, *58*(1), 267–288. Retrieved from www.jstor.org/stable/2346178

van der Kuil, W. A., Schoffelen, A. F., de Greeff, S. C., Thijsen, S. F., Alblas, H. J., Notermans, D. W., ... and, T. L. (2017, Nov). National laboratory-based surveillance system for antimicrobial resistance: a successful tool to support the control of antimicrobial resistance in the netherlands. *Eurosurveillance*, *22*(46). Retrieved from https://doi.org/10.2807%2F1560-7917.es.2017.22.46.17-00062 doi: 10.2807/1560-7917.es.2017.22.46.17-00062

van Nee, M. M., Wessels, L. F., & van de Wiel, M. A. (2021, Aug). Flexible co-data learning for high-dimensional prediction. *Statistics in Medicine*, *40*(26), 5910–5925. Retrieved from https://doi.org/10.1002%2Fsim.9162 doi: 10.1002/sim.9162

Volovici, V., Syn, N. L., Ercole, A., Zhao, J. J., & Liu, N. (2022, Sep). Steps to avoid overuse and misuse of machine learning in clinical research. *Nature Medicine*, *28*(10), 1996–1999. Retrieved from https://doi.org/10.1038%2Fs41591-022-01961-6 doi: 10.1038/s41591-022-01961-6

Wilkinson, J., Arnold, K. F., Murray, E. J., van Smeden, M., Carr, K., Sippy, R., ... Tennant, P. W. G. (2020, Dec). Time to reality check the promises of machine learning-powered precision medicine. *The Lancet Digital Health*, *2*(12), e677–e680. Retrieved from https://doi.org/10.1016%2Fs2589-7500%2820%2930200-4 doi: 10.1016/s2589-7500(20)30200-4

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., & Zhu, J. (2019). Explainable AI: A brief survey on history, research areas, approaches and challenges. In *Natural language processing and chinese computing* (pp. 563–574). Springer International Publishing. Retrieved from https://doi.org/10.1007%2F978-3-030-32236-6_51 doi: 10.1007/978-3-030-32236-6_51

Zwep, L. B., Duisters, K. L. W., Jansen, M., Guo, T., Meulman, J. J., Upadhyay, P. J., & Hasselt, J. G. C. (2021, Apr). Identification of high-dimensional omics-derived predictors for tumor growth dynamics using machine learning and pharmacometric modeling. *CPT: Pharmacometrics & Systems Pharmacology*, *10*(4), 350–361. Retrieved from https://doi.org/10.1002%2Fpsp4.12603 doi: 10.1002/psp4.12603

Zwep, L. B., Guo, T., Nagler, T., Knibbe, C. A., Meulman, J. J., & van Hasselt, J. C. (2022). Virtual patient simulation using copula modeling. *[in preparation]*.

Zwep, L. B., Haakman, Y., Duisters, K. L. W., Meulman, J. J., Liakopoulos, A., & van Hasselt, J. G. C. (2021, Sep). Identification of antibiotic collateral sensitivity and resistance interactions in population surveillance data. *JAC-Antimicrobial Resistance*, *3*(4). Retrieved from https://doi.org/10.1093%2Fjacamr%2Fdlab175 doi: 10.1093/jacamr/dlab175

Chapter 9

Nederlandse Samenvatting

Samenvatting

Om de behandeling met geneesmiddelen van individuele patiënten te verbeteren is het cruciaal om de variabiliteit tussen patiënten en hun reactie op verschillende geneesmiddelen en doseringsschema's, beter te begrijpen. In dit proefschrift hebben wij onderzoek gedaan naar verschillende factoren en statistische methodes om dergelijke variabiliteit tussen patiënten kunnen voorspellen. In de afgelopen jaren zijn er grote ontwikkelingen geweest op het gebied van meettechnologie waarmee zeer efficiënt genetische of biochemische profielen kunnen worden bepaald in patiëntmateriaal, die mogelijk als zogenaamde biomarker bij kunnen dragen aan het voorspellen van variatie tussen patiënten. Dergelijke profielen zijn hoog-dimensionaal: voor elk patiëntmonster kunnen er vele honderden tot duizenden unieke genetische of biochemische waarden bepaald worden. Bovendien worden deze hoog-dimensionale gegevens ook steeds meer longitudinaal (over de tijd) gemeten in patiënten om zo verdere inzichten in verloop van ziekte en therapieresponse te kunnen verkrijgen. Echter, het afleiden van bruikbare inzichten uit dergelijke gegevens is vooralsnog een uitdaging (Sectie 1). Om tot deze inzichten te komen, zijn statische methoden voor longitudinale en hoog-dimensionale gegevens nodig. Met deze methoden kan gezocht worden naar biomarkers die de variabiliteit tussen patiënten verklaren, in termen van reactie op de behandeling en klinisch verloop van de ziekte (Sectie 2). Daarnaast zijn deze methoden nodig om klinische vragen te beantwoorden met behulp van gegevens uit de medische praktijk, die in de reguliere gezondheidszorg en onderzoeksinstituten verkregen worden (Sectie 3).

9.1 Verschillende gegevens in farmaceutisch onderzoek

Het doel van klinische studies is om kennis te vergaren over de response van patiënten op een behandeling. Echter, niet alle soorten patiënten, medicatie en bijwerkingen worden door middel van klinische studies onderzocht, dit vanwege de uitdagingen en kosten die komen kijken, bij bijvoorbeeld kleine populaties of weinig voorkomende effecten. De beslissingen over dit soort behandelingen en patiëntengroepen kunnen in dat geval genomen worden met behulp van gegevens uit de gezondheidszorg en metingen die thuis kunnen worden gedaan (Morrato et al., 2007; Swift et al., 2018). In Hoofdstuk 2, beschreven we verschillende gegevenstypen die gebruikt kunnen worden voor het maken van klinische beslissingen. Ook beschreven we welke mogelijkheden en uitdagingen deze gegevenstypen met zich meebrengen. Hierbij ligt de focus op geneesmiddeleffecten in kinderen, omdat klinische studies daar vaak belemmerd worden door problemen met het rekruteren van patiënten (Brussee et al., 2016).

Gegevens uit de medische praktijk kunnen niet op dezelfde manier geanalyseerd worden als gegevens uit klinische studies. Dat zorgt ervoor dat er voor het analyseren van deze gegevens aangepaste of uitgebreide statistische methoden, zoals

machine learning methoden, nodig zijn. Het gebruik van zulke methoden levert een extra uitdaging op, omdat de uitkomsten van dit soort methoden vaak moeilijk te interpreteren zijn, en het bij het maken van medische beslissingen sterk van belang is om uit te kunnen leggen waarom deze beslissingen worden genomen (Knoppers & Thorogood, 2017).

9.2 Zoektocht naar biomarkers

De vooruitgang in meettechnieken voor het bepalen van moleculaire waarden maakt het mogelijk om de moleculaire samenstelling van een mens preciezer te meten, bijvoorbeeld uit bloed- of urinemonsters. Het is mogelijk om gegevens over moleculen, zoals metabolieten, RNA en DNA, te meten om een patiënt met grote precisie te beschrijven (Pearson, 2016). Deze zogeheten 'omics' gegevens maken het mogelijk om veel meer factoren te bestuderen dan voorheen, maar dat maakt de analyse voor het detecteren van voorspellende biomarkers ook een stuk complexer (Depledge et al., 1993), zeker in het geval van tijdgerelateerde veranderingen. De statistische technieken die op dit moment beschikbaar zijn, zijn vaak niet toepasbaar op gegevens die zowel hoog-dimensionaal als longitudinaal zijn. De meeste methoden zijn ontworpen om maar met één van deze twee aspecten om te kunnen gaan.

In Hoofdstuk 3, identificeerden we biomarkers voor het volgen van de reactie op verschillende kankerbehandelingen uit hoog-dimensionale DNA gegevens, die gemeten waren in muismodellen met kankercellen getransplanteerd uit patiënten. De gebruikte gegevens komen uit een gepubliceerde studie (Gao et al., 2015), waarbij DNA eigenschappen van kankercellen werden bepaald en vervolgens de tumorgroei over de tijd werd gemeten in muizen die behandeld waren met verschillende geneesmiddelen. In onze analyse hebben wij als eerste stap een longitudinaal tumorgroei inhibitie model ontwikkeld om de tumorgroei te beschrijven met behulp van drie karakteristieken: de groeisnelheid van de tumor (k_g), de sterkte van het geneesmiddeleffect (k_d) en de ontwikkeling van resistentie tegen het geneesmiddel (k_r). Met het model worden zowel populatie parameters als de individuele parameters (empirische Bayes schatters) bepaald. In de tweede stap werden de schattingen van deze individuele parameters als uitkomst in een zogenaamde lasso regressie gebruikt en werden ze gerelateerd aan de hoog-dimensionale DNA gegevens. We vonden door middel van een kruisvalidatie dat een deel van de variantie in geneesmiddelresponse verklaard kon worden met behulp van de DNA gegevens. Het gebruik van een biologisch geïnformeerde groep-lasso maakte het mogelijk om DNA effecten op het niveau van de biologisch processen te onderzoeken.

Het modelleren van hoog-dimensionale gegevens heeft als risico dat er zogenaamde overfitting kan optreden, waarbij de hoeveelheid variabelen ervoor kan zorgen dat de correlerende effecten niet geschat kunnen worden tussen de optredende ruis. Dit probleem kan ook voorkomen in de context van farmacometrische modellen, dus ook bij het schatten van parameters uit hoog-dimensionale DNA gegevens. Het voorspellen van de tumorgroei in de twee beschreven stappen omzeilt deze beperk-

ing. Een mogelijk risico van de door ons ontwikkelde twee-stapmethode voor biomarkeridentificatie is dat fouten in modelvoorspelling in de eerste stap kunnen worden uitvergroot in de tweede stap. Om het risico hierop te verkleinen is het dus van belang om de fout in beiden stappen aandachtig te bestuderen.

In Hoofdstuk 4 hebben we potentiële biomarkers voor het klinisch verloop, de combinatie tussen de reactie op de behandeling en het ziekteverloop, onderzocht in longitudinale metabolomische gegevens van in het ziekenhuis opgenomen patiënten met een thuis opgelopen longontsteking (community acquired pneumonia, CAP). We gebruikten de dimensie reductie methode principale component analyse om patronen van metabolieten over de tijd te verkennen en de relatie tussen de biochemische klassen van de metabolieten en het klinisch verloop te bestuderen. We berekenden de correlaties tussen de metabolieten en twee maten voor het klinisch verloop: de CURB score, een score die aangeeft hoe ziek de patiënt is bij aankomst in het ziekenhuis, en het aantal dagen van het verblijf in het ziekenhuis, welke aangeeft hoelang het duurde voor een patiënt om te herstellen. De patronen van de metabolieten veranderden duidelijk over de tijd, wat aangeeft hoe belangrijk het is om longitudinale metabolomische gegevens te bestuderen in patiënten met CAP. We identificeerden verschillende biochemische metabolietklassen die gerelateerd zijn aan het klinisch ziekteverloop, zoals de triglyceriden en de lyso-fosfatidylcholine klassen.

9.3 Gegevens uit de medische praktijk

In de gezondheidszorg worden routinematig zorggegevens van patiënten verzameld, om de hun gezondheid en hun reactie op de behandeling te volgen. Deze routinematig verzamelde zorggegevens worden tegenwoordig voor langere tijd opgeslagen, en vervolgens vaak beschikbaar gemaakt voor onderzoek. Dit soort gegevens, uit de medische praktijk, kunnen bijdragen aan het begrip van verschillende onderwerpen in de farmaceutische wetenschap.

Antibiotica-resistentie vormt een bedreiging voor de gezondheid van mensen wereldwijd. De mate van de dreiging en de ontwikkeling hiervan worden gemonitord door het bepalen van de mate van antibiotica-resistentie in patiënten opgenomen in ziekenhuizen (van der Kuil et al., 2017). Voor het verzamelen van deze gegevens wordt meestal de minimaal inhiberende concentraties (MIC) van de uit patiënten geïsoleerde bacteriële stammen bepaald en opgeslagen. Collaterale sensitiviteit (CS) is een proces waarbij in een populatie van pathogenen, de blootstelling aan één medicijn de resistentie tegen een tweede medicijn vermindert. Dit mechanisme kan bruikbaar kan zijn voor het bestrijden van resistente pathogenen (Aulin et al., 2021). De gemeten MIC waarden in de populatie kunnen gebruikt worden om deze CS te detecteren. Dit fenomeen is al meermaals gedetecteerd in proeven in een laboratoriumomgeving, maar de kennis over het voorkomen van CS in de klinische context is zeer beperkt.

In Hoofdstuk 5 hebben we een methode voorgesteld om collaterale effecten te kwantificeren in MIC gegevens (Zwep et al., 2021). Deze methode kwantificeert de

effecten als log_2 fold change, wat een interpreteerbare maat is die ook in laboratoriumproeven gebruikt wordt. Dit maakt het mogelijk om een directe vergelijking tussen laboratoriumproeven en de klinische praktijk te maken in de grootte van het effect. Deze maat werd gebruikt in Hoofdstuk 6 voor het kwantificeren van CS in grootschalige MIC bewakingsgegevens uit verschillende gegevensbronnen. Hieruit kwam naar voren dat CS leek voor te komen in de klinische praktijk, maar het was moeilijk om specifieke patronen over de verschillende soorten en antibiotica te detecteren. Dit gebrek aan generaliseerbaarheid maakt het lastig om CS in de klinische praktijk toe te passen (Nichol et al., 2019).

Routinematig verzamelde zorggegevens van patiënten bevatten verschillende typen covariaten, zoals biomarker concentraties, demografische gegevens en andere karakteristieken van de patiënten (Currie & MacDonald, 2000). In farmacometrische modellen worden deze karakteristieken vaak gebruikt om variabiliteit tussen patiënten te verklaren en voorspellen, wat mogelijkheden biedt tot het uitbreiden van de voorspellingen naar speciale patiëntengroepen. Farmacometrische simulaties vereisen het simuleren van deze covariaten, maar het delen van deze gevoelige patiëntengegevens tussen zorgverleners en onderzoeksgroepen is vaak niet mogelijk, vanwege het beschermen van de privacy van de patiënten.

In Hoofdstuk 7 stellen wij het gebruik van zogenaamde copulas voor als een geschikte methode voor het simuleren van virtuele patiënten. Copulas zijn multivariate verdelingsfuncties die gedeelde verdelingen kunnen beschrijven en bieden daarmee een flexibele manier om covariaten sets van patiënten te beschrijven en te simuleren. De gedeelde verdelingsfunctie van meerdere covariaten beschrijft de afhankelijkheid tussen de verschillende covariaten. De meeste covariaten zijn niet onafhankelijk van elkaar, dus het vatten van deze afhankelijkheid is cruciaal voor het simuleren van realistische virtuele patiënten. Onze studie liet zien dat we met copulas in staat zijn om realistische patiëntengroepen te simuleren (Zwep et al., 2022). Het is daarmee mogelijk om verschillende patiëntengroepen te simuleren, wat het extrapoleren van de resultaten uit modellen naar specifieke patiëntengroepen mogelijk te maakt.

De copulas worden geschat aan de hand van gegevens die niet altijd beschikbaar zijn voor onderzoekers, maar als de copulas geschat worden bij de onderzoeksgroep die de gegevens in bezit heeft kunnen de copulas gedeeld worden met andere onderzoekers, zodat zij de populatie kunnen simuleren en bestuderen zonder de oorspronkelijke gegevens in bezit te krijgen. Op deze manier kunnen covariaten van patiënten gedeeld worden tussen onderzoeksgroepen, zonder de privacy van de patiënt te schaden.

9.4 Conclusies

Om geneesmiddelgebruik te personaliseren is het nodig om de variabiliteit in de reacties van patiënten op een behandeling te ontrafelen. Hiervoor is het gebruik van statistische methoden die het gebruik van verschillende soorten gegevens, zoals

hoog-dimensionale omics, maar ook routinematig verzamelde zorggegevens, van essentieel belang. Integratie van state-of-the-art statistische methoden met farmacometrische modellen faciliteert farmacologisch onderzoek, door het mogelijk maken van gebruik van meerdere soorten gegevens, waarbij de interpreteerbaarheid van de farmacologische modellen behouden blijft.

Referenties

Aulin, L. B. S., Liakopoulos, A., van der Graaf, P. H., Rozen, D. E., & van Hasselt, J. G. C. (2021, Sep). Design principles of collateral sensitivity-based dosing strategies. *Nature Communications*, *12*(1). Retrieved from https://doi.org/10.1038%2Fs41467-021-25927-3 doi: 10.1038/s41467-021-25927-3

Brussee, J. M., Calvier, E. A. M., Krekels, E. H. J., Välitalo, P. A. J., Tibboel, D., Allegaert, K., & Knibbe, C. A. J. (2016, Jun). Children in clinical trials: towards evidence-based pediatric pharmacotherapy using pharmacokinetic-pharmacodynamic modeling. *Expert Review of Clinical Pharmacology*, *9*(9), 1235–1244. Retrieved from https://doi.org/10.1080%2F17512433.2016.1198256 doi: 10.1080/17512433.2016.1198256

Currie, C. J., & MacDonald, T. M. (2000). Use of routine healthcare data in safe and cost-effective drug use. *Drug Safety*, *22*(2), 97–102. Retrieved from https://doi.org/10.2165%2F00002018-200022020-00002 doi: 10.2165/00002018-200022020-00002

Depledge, M. H., Amaral-Mendes, J. J., Daniel, B., Halbrook, R. S., Kloepper-Sams, P., Moore, M. N., & Peakall, D. B. (1993). The conceptual basis of the biomarker approach. In *Biomarkers* (pp. 15–29). Springer Berlin Heidelberg. Retrieved from https://doi.org/10.1007%2F978-3-642-84631-1_2 doi: 10.1007/978-3-642-84631-1_2

Gao, H., Korn, J. M., Ferretti, S., Monahan, J. E., Wang, Y., Singh, M., ... Sellers, W. R. (2015, Oct). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nature Medicine*, *21*(11), 1318–1325. Retrieved from https://doi.org/10.1038%2Fnm.3954 doi: 10.1038/nm.3954

Knoppers, B. M., & Thorogood, A. M. (2017, Aug). Ethics and big data in health. *Current Opinion in Systems Biology*, *4*, 53–57. Retrieved from https://doi.org/10.1016%2Fj.coisb.2017.07.001 doi: 10.1016/j.coisb.2017.07.001

Morrato, E. H., Elias, M., & Gericke, C. A. (2007, Dec). Using population-based routine data for evidence-based health policy decisions: lessons from three examples of setting and evaluating national health policy in australia, the UK and the USA. *Journal of Public Health*, *29*(4), 463–471. Retrieved from https://doi.org/10.1093%2Fpubmed%2Ffdm065 doi: 10.1093/pubmed/fdm065

Nichol, D., Rutter, J., Bryant, C., Hujer, A. M., Lek, S., Adams, M. D., ... Scott, J. G. (2019, Jan). Antibiotic collateral sensitivity is contingent on the repeatability of evolution. *Nature Communications*, *10*(1). Retrieved from https://doi.org/10.1038%2Fs41467-018-08098-6 doi: 10.1038/s41467-018-08098-6

Pearson, E. R. (2016, May). Personalized medicine in diabetes: the role of 'omics' and biomarkers. *Diabetic Medicine*, *33*(6), 712–717. Retrieved from https://doi.org/10.1111%2Fdme.13075 doi: 10.1111/dme.13075

Swift, B., Jain, L., White, C., Chandrasekaran, V., Bhandari, A., Hughes, D. A., & Jadhav, P. R. (2018, May). Innovation at the intersection of clinical trials and real-world data science to advance patient care. *Clinical and Translational Science*, *11*(5), 450–460. Retrieved from https://doi.org/10.1111%2Fcts.12559 doi: 10.1111/cts.12559

van der Kuil, W. A., Schoffelen, A. F., de Greeff, S. C., Thijsen, S. F., Alblas, H. J., Notermans, D. W., ... and, T. L. (2017, Nov). National laboratory-based surveillance system for antimicrobial resistance: a successful tool to support the control of antimicrobial resistance in the netherlands. *Eurosurveillance*, *22*(46). Retrieved from https://doi.org/10.2807%2F1560-7917.es.2017.22.46.17-00062 doi: 10.2807/1560-7917.es.2017.22.46.17-00062

Zwep, L. B., Guo, T., Nagler, T., Knibbe, C. A., Meulman, J. J., & van Hasselt, J. C. (2022). Virtual patient simulation using copula modeling. *[in preparation]*.

Zwep, L. B., Haakman, Y., Duisters, K. L. W., Meulman, J. J., Liakopoulos, A., & van Hasselt, J. G. C. (2021, Sep). Identification of antibiotic collateral sensitivity and resistance interactions in population surveillance data. *JAC-Antimicrobial Resistance*, *3*(4). Retrieved from https://doi.org/10.1093%2Fjacamr%2Fdlab175 doi: 10.1093/jacamr/dlab175

Appendices

Curriculum vitae

Laura Zwep was born on January 12, 1994, in Utrecht, the Netherlands. Her academic education started with a Bachelor degree in Biology at Leiden University (2012 - 2015). After that she completed a Master degree (cum laude) in Statistical Science for the Life and Behavioural Sciences at Leiden University (2015 - 2017). During her Master program she performed research projects at the Biometris department at Wageningen University and the department of Epidemiology at the Erasmus Medical Center, Rotterdam.

Laura started her PhD at Leiden University in 2017 under the supervision of prof. dr. Jacqueline. Meulman and dr. Coen van Hasselt, working at the Leiden Academic Centre for Drug Research (LACDR). During her PhD she developed her interests in teaching and was involved in numerous student projects and lectured and contributed to several Bachelor and Master. Laura has presented her work at several international conferences, including oral presentation at the International Biometric Conference (2020) and the Population Approach Group Europe Meeting (2022). Since March 2023, Laura has started as assistant professor at the division of Systems Pharmacology and Pharmacy at LACDR and Leiden University, contributing to teaching and research at the interface of statistics and pharmacometrics.

List of publications

den Hartog, I., Zwep, L. B., Vestjens, S. M. T., Harms, A. C., Voorn, G. P., de Lange, D. W., ... van Hasselt, J. G. C. (2021, Jun). Metabolomic profiling of microbial disease etiology in community-acquired pneumonia. *PLOS ONE*, *16*(6), e0252378. Retrieved from http://dx.doi.org/10.1371/journal.pone.0252378 doi: 10.1371/journal.pone.0252378

Goulooze, S. C., Zwep, L. B., Vogt, J. E., Krekels, E. H., Hankemeier, T., Anker, J. N., & Knibbe, C. A. (2020, Apr). Beyond the Randomized Clinical Trial: Innovative Data Science to Close the Pediatric Evidence Gap. *Clinical Pharmacology & Therapeutics*, *107*(4), 786–795. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/cpt.1744 doi: 10.1002/cpt.1744

Guo, T., van Hest, R. M., Zwep, L. B., Roggeveen, L. F., Fleuren, L. M., Bosman, R. J., ... van Hasselt, J. G. C. (2020, Sep). Optimizing Predictive Performance of Bayesian Forecasting for Vancomycin Concentration in Intensive Care Patients. *Pharmaceutical Research*, *37*(9), 171. Retrieved from http://link.springer.com/10.1007/s11095-020-02908-7 (Publisher: Pharmaceutical Research) doi: 10.1007/s11095-020-02908-7

Malosetti, M., Zwep, L. B., Forrest, K., van Eeuwijk, F. A., & Dieters, M. (2020, Dec). Lessons from a GWAS study of a wheat pre-breeding program: pyramiding resistance alleles to Fusarium crown rot. *Theoretical and Applied Genetics*. Retrieved from https://doi.org/10.1007/s00122-020-03740-8 (Publisher: Springer Berlin Heidelberg) doi: 10.1007/s00122-020-03740-8

Zwep, L. B., Duisters, K. L. W., Jansen, M., Guo, T., Meulman, J. J., Upadhyay, P. J., & Hasselt, J. G. C. (2021, Apr). Identification of high□dimensional omics□derived predictors for tumor growth dynamics using machine learning and pharmacometric modeling. *CPT: Pharmacometrics & Systems Pharmacology*, *10*(4), 350–361. Retrieved 2021-05-18, from https://onlinelibrary.wiley.com/doi/10.1002/psp4.12603 (Publisher: Wiley) doi: 10.1002/psp4.12603

Zwep, L. B., Haakman, Y., Duisters, K. L. W., Meulman, J. J., Liakopoulos, A., & van Hasselt, J. G. C. (2021, Sep). Identification of antibiotic collateral sensitivity and resistance interactions in population surveillance data. *JAC-Antimicrobial Resistance*, *3*(4), 1–9. Retrieved from https://academic.oup.com/jacamr/article/doi/10.1093/jacamr/dlab175/6445223 doi: 10.1093/jacamr/dlab175

Affiliations of authors

John N. van den Anker
Children's National Health System, Washington D.C., USA
University of Basel Children's Hospital, Basel, Switzerland

Stephan Beisken
Ares Genetics, Vienna, Austria

Kevin L. W. Duisters
Mathematical Institute, Leiden University, Leiden, The Netherlands

Gabriel Forn Cuni
Institute of Biology, Leiden University, Leiden, The Netherlands

Ewoudt M. W. van de Garde
St. Antonius Hospital, Nieuwegein, The Netherlands
Department of Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

Tingjie Guo
Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands

Yob Haakman
Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands

Thomas Hankemeier
Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands

Ilona den Hartog
Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands

J. G. Coen van Hasselt
Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands

Martijn Jansen
Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands

Catherijne A. J. Knibbe
Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands

Elke H. J. Krekels
Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands
Department of Clinical Pharmacy, St. Antonius Hospital, Nieuwegein, The Netherlands

Apostolos Liakopoulos
Harvard Medical School, Harvard University, Boston, MA, USA

Jacqueline J. Meulman
LUXs Data Science, Leiden, The Netherlands
Department of Statistics, Stanford University, Stanford, CA, United States

Thomas Nagler
Department of Statistics, Ludwig Maximilian University of Munich, Munich, Germany

Annelot F. Schoffelen
RIVM, Eindhoven, The Netherlands

Parth J. Upadhyay
Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands

Julia E. Vogt
Department of Computer Science, ETH Zurich, Zurich, Switzerland
Swiss Institute of Bioinformatics, Lausanne, Switzerland

Acknowledgements

This thesis, and the enjoyable time I had as a PhD student, would not have been possible without the many people that were involved over the last five years. First, I would like to thank my promotor Jacqueline, who has always encouraged me, both by positively supporting me and by being very critical towards my scientific work. She is a great example on how to be a scientist with loyalty and (scientific) integrity. Second, I would like to thank Coen, my supervisor, who has provided me with many great research projects, and has many brilliant ideas, both about science and about the way the academic world could be improved. It has been a pleasure being supervised by him. I want to thank Kevin for guiding me through the first half of my PhD, by always taking me serious and sharing and developing ideas for our research.

I have many wonderful colleagues, who made my time as a PhD student really amazing. I was always able to get joy and support from my colleagues. I want to thank my paranymphs: Linda, who is extremely clever, very enthusiastic and always available to discuss science with me, and Ilona, who is an excellent organizer, always brings joy to the people around her, and provides the voice of reason whenever I am in a difficult situation. You are both very dear to me, we really had, and still have, an awesome time together.

I want to thank my office mates: Angie, Maik and Helle, with whom I spend a lot of time working hard and chatting about anything and everything. I couldn't have asked for a better direct working environment. Thank you also to the people who hopped by my office regularly such as Berfin, Laura and Annika, who kept me off my work sometimes, but always made my work and breaks more enjoyable. I want to thank the people from my cohort, Aline, Jinqiu, Feiyan, Yu, Lu and Mohammed, who rode the PhD rollercoaster together with me and made it a fun and interesting ride.

I want to thank all my collaborators who I got to work with, and especially Parth, Gabriel, Thomas and Tingjie. I want to thank Solomiia for creating and coordinating the course together, and also being a good friend.

I would like to thank my other colleagues Rob, Bas, Sinzi, Swantje, Elke, Carla, Chun, Marinda, An Duc, Cathy, Yuchen, Suruchi, Sebastian, Jorn, Dirk-Jan, my colleagues from the DSO, my football team Mathletico, and all others I got the pleasure to spend my time with. I would also like to thank the students I supervised, that have all been inspiring and a lot of fun to collaborate to with.

Lastly, I would like to thank people outside of my work. Of course Patrick, with whom I build a home and who has always helped me maintain a healthy work life balance. Thank you to my family and Patrick's family for the amazing support and love. Thank you to my friends and my football team for generally increasing my level of happiness.