Integrating Nonlinear Mixed Effects and Physiologically–Based Modeling Approaches for the Analysis of Repeated Measurement Studies

With Applications in Quantitative Pharmacology and Quantitative Psycholinguistics

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)
an der Mathematisch–Naturwissenschaftlichen Fakultät
der Universität Potsdam

von Alexander Solms

2017

Published online at the
Institutional Repository of the University of Potsdam:
URN urn:nbn:de:kobv:517-opus4-397070
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-397070

Abstract

During the drug discovery & development process, several phases encompassing a number of preclinical and clinical studies have to be successfully passed to demonstrate safety and efficacy of a new drug candidate. As part of these studies, the characterization of the drug's pharmacokinetics (PK) is an important aspect, since the PK is assumed to strongly impact safety and efficacy. To this end, drug concentrations are measured repeatedly over time in a study population. The objectives of such studies are to describe the typical PK time-course and the associated variability between subjects. Furthermore, underlying sources significantly contributing to this variability, e.g. the use of comedication, should be identified. The most commonly used statistical framework to analyse repeated measurement data is the nonlinear mixed effect (NLME) approach. At the same time, ample knowledge about the drug's properties already exists and has been accumulating during the discovery & development process: Before any drug is tested in humans, detailed knowledge about the PK in different animal species has to be collected. This drug-specific knowledge and general knowledge about the species' physiology is exploited in mechanistic physiological based PK (PBPK) modeling approaches —it is, however, ignored in the classical NLME modeling approach.

Mechanistic physiological based models aim to incorporate relevant and known physiological processes which contribute to the overlying process of interest. In comparison to data–driven models they are usually more complex from a mathematical perspective. For example, in many situations, the number of model parameters outrange the number of measurements and thus reliable parameter estimation becomes more complex and partly impossible. As a consequence, the integration of powerful mathematical estimation approaches like the NLME modeling approach—which is widely used in data-driven modeling—and the mechanistic modeling approach is not well established; the observed data is rather used as a confirming instead of a model informing and building input.

Another aggravating circumstance of an integrated approach is the inaccessibility to the details of the NLME methodology so that these approaches can be adapted to the specifics and needs of mechanistic modeling. Despite the fact that the NLME modeling approach exists for several decades, details of the mathematical methodology is scattered around a wide range of literature and a comprehensive, rigorous derivation is lacking. Available literature usually only covers selected parts of the mathematical methodology. Sometimes, important steps are not described or are only heuristically motivated, e.g. the iterative algorithm to finally determine the parameter estimates.

Thus, in the present thesis the mathematical methodology of NLME modeling is systemically described and complemented to a comprehensive description, comprising the common theme from ideas and motivation to the final parameter estimation. Therein, new insights for the interpretation of different approximation methods used in the context of the NLME modeling approach are given and illustrated; furthermore, similarities and differences between them are outlined. Based on these findings, an expectation-maximization (EM) algorithm to determine estimates of a NLME model is described.

Using the EM algorithm and the lumping methodology by Pilari and Huisinga [2010], a new approach on how PBPK and NLME modeling can be combined is presented and ex-

emplified for the antibiotic levofloxacin. Therein, the lumping identifies which processes are informed by the available data and the respective model reduction improves the robustness in parameter estimation. Furthermore, it is shown how *apriori* known factors influencing the variability and *apriori* known unexplained variability is incorporated to further mechanistically drive the model development. Concludingly, correlation between parameters and between covariates is automatically accounted for due to the mechanistic derivation of the lumping and the covariate relationships.

A useful feature of PBPK models compared to classical data-driven PK models is in the possibility to predict drug concentration within all organs and tissue in the body. Thus, the resulting PBPK model for levofloxacin is used to predict drug concentrations and their variability within soft tissues which are the site of action for levofloxacin. These predictions are compared with data of muscle and adipose tissue obtained by microdialysis, which is an invasive technique to measure a proportion of drug in the tissue, allowing to approximate the concentrations in the interstitial fluid of tissues. Because, so far, comparing human *in vivo* tissue PK and PBPK predictions are not established, a new conceptual framework is derived. The comparison of PBPK model predictions and microdialysis measurements shows an adequate agreement and reveals further strengths of the presented new approach.

We demonstrated how mechanistic PBPK models, which are usually developed in the early stage of drug development, can be used as basis for model building in the analysis of later stages, i.e. in clinical studies. As a consequence, the extensively collected and accumulated knowledge about species and drug are utilized and updated with specific volunteer or patient data. The NLME approach combined with mechanistic modeling reveals new insights for the mechanistic model, for example identification and quantification of variability in mechanistic processes. This represents a further contribution to the learn & confirm paradigm across different stages of drug development.

Finally, the applicability of mechanism–driven model development is demonstrated on an example from the field of Quantitative Psycholinguistics to analyse repeated eye movement data. Our approach gives new insight into the interpretation of these experiments and the processes behind.

Acknowledgments

First of all, I would like to express my sincere gratitude to my supervisor Wilhelm Huisinga for not only introducing me into the exciting field of Pharmacometrics and offering me the possibility of writing this thesis, but also for his inspiring enthusiam for research, his valuable input & feedback, and his endless support, guidance and patience throughout the last 6 years.

My gratitude also goes to Charlotte Kloft for numerous stimulating meetings and discussion, her valuable thoughts & input from a different angle, her commitment to drive interdisciplinary exchange, and her warm way.

I also want to express my deep thanks to Reinhold Kliegl for opening new perspectives and giving insight into a fascinating discipline, as well as for his contagious enthusiasm.

Many thanks also to the participants of the levofloxacin study and to Professor Zeitlinger for gratefully providing the levofloxacin dataset.

The Graduate Research Training Program PharMetrX is gratefully acknowledged for the financial support and the unique environment created by the excellent interdisciplinary character, the close cooperation between academia and industry and great colleagues. In particular, I am very grateful to Andre Schäftlein, Valerie Nock, Stephan Menz, Sebastian Wicha, Iris Minichmayr, Christoph Hethey, Bart Ploeger and Hauke Rüehs for their priceless companionship, for sharing their views and ideas and providing valuable comments and suggestions.

And finally, I want to thank Natascha and my family for their continuous encouragement and all kinds of support whenever necessary.

Contents

Acronyms

AUC area under the concentration–time curve.

B:P blood–to–plasma partition coefficient.

BMI body–mass–index.

BSA body–surface–area.

BSV between subject variability.

c.d.f. cumulative distribution function.

CFU colony forming unit.

CLT central limit theorem.

CV coefficient of variation.

DirMult Dirichlet–Multinomial distribution.

EBE empirical Bayes estimate.

EM expectation-maximization.

FDA Food and Drug Administration.

FO first-order.

FOCE first–order conditional estimation.

fuP fraction unbound in plasma.

GOF goodness–of–fit.

i.i.d. independent and identically distributed.

i.v. intravenous.

ISF interstitial space fluid.

LBW lean–body–weight.

LLN law of large numbers.

Acronyms

LRT likelihood ratio test.

MCMC Markov-Chain Monte-Carlo.

MIC minimum inhibitory concentration.

ML maximum likelihood.

MLE maximum likelihood estimation.

MSE mean squarred error.

NLME nonlinear mixed effects.

ODE ordinary differential equation.

OFV objective funtion value.

p.d.f. probability density function.

PBPK physiologically–based pharmacokinetics.

PD pharmacodynamics.

PK pharmacokinetics.

PK/PD pharmacokinetic–pharmacodynamic.

RR relative recovery.

RSE relative standard error.

RSME root mean squarred error.

SA stochastic approximation.

SAEM stochastic approximation expectation–maximization.

SE standard error.

VPC visual predictive check.

VWP visual word paradigm.

WSV within subject variability.

1. Introduction

An essential objective of clinical pharmacology studies is to characterize the pharmacokinetics (PK) of the drug of interest—i.e. the drug concentration time profile within the body—in the respective patient population. The PK are of interest because it is assumed that they are closely linked to the drug effects, the so–called pharmacodynamics (PD) of the drug. To study the PK, the drug is administered and the drug concentration is measured repeatedly over time, usually in plasma, in several patients. These type of observations are denoted as repeated measurement data—or in case of PK as population PK data. Based on these repeated measurement data, the typical profile as well as the variability of the concentration time course within the patient population are characterized. The variability between patients, denoted as between subject variability (BSV), is important to finally define a dose which ensures that all patients are treated efficaciously.

For example, for antibiotics it is important that the drug levels at the site of infection, e.g. lung or soft tissue, are sufficient to successfully kill bacteria and avoid emergence of resistance. However, these concentrations are difficult to measure directly in patients. Since measuring plasma concentrations is a well–established process and can be included in daily clinical routine, one approach is to take the plasma concentrations as surrogate for the drug concentrations at the target. However, the predictability of this surrogate is usually not known and under debate, e.g. see Müller et al. [2004].

An alternative is given by the invasive microdialysis approach, where a proportion of the local drug concentration is measured and extrapolated to the concentration within the tissue space of interest. This approach requires significantly more effort on the patient site, as well as on the site of the physician or study nurse conducting the measurements. Thus, it is not suitable to be conducted in larger clinical studies or during daily clinical routine. Additionally, the relationship between measured local drug concentrations and tissue concentration has to be determined. This determination itself is rather complex and potentially associated with relatively large uncertainty. Due to the complexity of this approach, microdialysis studies usually only include small number of patients. As a consequence, the BSV in the patient population as well as the average local concentration can only be determined with a relatively large uncertainty.

Another alternative to predict the tissue concentrations is given by the *in silico* physiologically–based pharmacokinetics (PBPK) modeling approach. This method combines prior knowledge about the physiology and the tissue compositions of the species with the physicochemical properties of the drug. In general, a PBPK model is based on *in vitro* experiments and preclinical *in vivo* data. However, in the PBPK modeling approach it is unclear how population PK data can be used to inform and optimize the PBPK model and how BSV can be incorporated.

In summary, with all three approaches similar issues can be addressed with different strengths and limitations. Interestingly, they nicely complement each other in some aspects: for example, while repeated measurement plasma PK data is very efficient for characterizing

BSV of the PK, its use for predicting tissue concentrations is limited, on the other hand the tissue predictions determined by the microdialysis approach are affected with relatively large uncertainty and due to the small study size the characterization of BSV is difficult, in theory the PBPK modeling approach can be used to predict tissue concentrations, but the respective predictions are rarely verified with clinical data, moreover, it is still a matter of debate how BSV can be incorporated and in which magnitude.

For the antibiotic levofloxacin four studies were conducted by Prof. Markus Zeitlinger from the Medical University of Vienna, Austria, where levofloxacin PK profiles were densely measured in 24 subjects in plasma, muscle and adipose interstitial space fluid (ISF), the latter two were obtained by microdialysis (Zeitlinger et al. [2003, 2007]; Bellmann et al. [2004]).

Based on these studies, our idea was to combine the PK data obtained in plasma and ISF and PBPK modeling into an integrative analysis to predict the target concentrations and the respective BSV. Thereby, the strategy was to use literature information about human physiology and the repeated measurement data obtained in plasma to build a mechanistic population PBPK model, including known sources of BSV and estimate the unexplained magnitude of BSV. In a next step, the link between the *in silico* predicted tissue concentrations and the *in vivo* microdialysis measurements should be examined for the comparison of both approaches.

Population PBPK modeling

Typically, population PK data is analyzed using nonlinear mixed effects (NLME) modeling. A NLME model consists of three main building blocks. One part, the structural model, is a mathematical function describing the relationship between the dependent variables—i.e. the measurements, and the independent variables—e.g. in PK in particular time and dose. The structural model is parametrized with so–called fixed and random effects. The fixed effects parameters are assumed to be constant over the whole population, while the random effects parameters are random variables describing the variations between the subjects of the study population. Thus, fixed effects relate to the typical profile, while random effects describe the variability within the population. The second block consists of the choice of the random effects and their stochastic distribution denoted as BSV model. Often a covariate model describing the relationship between covariates and fixed effects parameter is included with the objective to explain parts of the observed BSV and thereby reduce the unexplained BSV. The term nonlinear in NLME arises from the fact that the structural model is a nonlinear function w.r.t. the random effects parameters. The third block of a NLME model is the stochastic model which describes the link between the structural model and the measurements, often called residual error model or within subject variability (WSV) model. If the underlying data is sufficient and no distributional assumption regarding the random effects can be made, so–called non–parametric methods can be applied. In practice these approaches are rarely used, particularly in life sciences the assumption of normally or log–normally distributed parameter within a population are well accepted, see Karlsson et al. [1998]; Limpert et al. [2001].

In the analysis of population PK data usually an empirically motivated classical compartmental model is chosen as structural model. Classical compartmental models are described by a small system of (one to four) ordinary differential equations (ODEs), parametrized by

a small set of parameters. For example, Fish and Chow [1997] reported that the PK of levofloxacin were best described by a two compartment system (i.e. two ODEs) parametrized by four parameters. The more complex PBPK models are as well described by a system of ODEs which are parametrized by a large set of parameters, typically ≥ 200 parameters. Only a small part of these parameters are drug related and usually assumed to be constant in the population. The majority are related to physiology, e.g. body composition and regional blood flows. These physiological parameters are expected to vary between patients, but details of the respective BSV are rarely known.

Tsamandouras et al. [2013] discuss parameter estimation in the context of PBPK modeling in general. They conclude that a full PBPK model is over-parametrized and classical estimation methods are not practicable. To deal with this issue, their suggestion is to use Bayesian methods or model reduction techniques. Likewise, Leil [2014] discuss the challenges with available approaches to characterize variability and uncertainty in a PBPK context. He states that combining NLME modeling and mechanism–based modeling is inapplicable due to the complexity of mechanistic models and the associated number of parameters. He also proposes to use a Bayesian approach to deal with this issue.

At the same time, first attempts combining Bayesian methods and population PBPK modeling have already been made by Krauss et al. [2013, 2015] and Tsamandouras et al. [2015]. All authors report similar challenges: difficulties of defining prior distributions, potential correlation between parameters and computational effort to derive the posterior distribution—which corresponds to the estimation/fitting results in a Baysian context. Using a Bayesian approach requires that for all parameters of interest—including as well the parameters describing the BSV—prior distributions have to be chosen and defined. For example, Krauss et al. [2015] developed a population PBPK model for theophylline based on PK data of 12 subjects and estimated posterior distributions for a subset of 41 parameters, 38 of them modeled with a random effect. This resulted in a computation time of approximately 76 days—even with the use of parallel computing.

Unsurprisingly, it is a huge challenge to find appropriate prior information, particularly, stochastic properties of these information for many parameters. As a consequence, partly, information for prior distributions is missing or incomplete and simplifying assumptions have to be made. For example, Tsamandouras et al. [2015] conducted a population PBPK analysis using a Bayesian approach and stated that prior information about BSV was not available. Moreover, they had to fix some model parameters to reduce the dimensionality of the estimation, in particular for parameters with large uncertainty w.r.t. to the prior knowledge. In addition, their results show that only for few parameters the estimated parameter values—defined as mode of the posterior distribution, differs to the mode of the prior distribution—which can be interpreted as initial value. This indicates that either the prior information was very precise or—more likely—that these parameter were not informed by the available data.

In conclusion, because of the large number of parameters, the complexity of the physiological processes within a PBPK model and because in general PK data is not sufficient to inform all physiological processes, non–Bayesian frequentist NLME modeling are rarely combined with mechanistic PBPK modeling in the analysis of PK data. However, because prior distributions for physiological and physicochemical parameter of the drugs are rarely known and due to the computational effort, Bayesian approaches are very challenging and

rarely applied neither.

An alternative is to focus only on the most relevant parameters and on those which can be informed by the available data. Identifying these can for example be done using a lumping approach. This approach lumps tissues and organs with similar kinetics together and thereby reduces the PBPK model dimensions. Following the approach by Pilari and Huisinga [2010], the parameters of the lumped model are composed of the original PBPK model parameters, i.e. the extensive prior knowledge about physiology and drug, as well as included covariate effects can still be integrated and used. Because physiological and physicochemical parameter can affect several of the lumped parameters in intricate ways, the composition of the lumped parameter is rather complex. Concludingly, correlation between parameters and between covariates is automatically accounted for due to the mechanistic derivation of the lumping and the covariate relationships.

The challenge is to develop an approach where lumped PBPK and NLME modeling are integrated for the analysis of the population PK data. Such an approach could be applied to inform and optimize the PBPK model, i.e. to adjust relevant parameters, estimate missing parameters and quantify the unexplained BSV and WSV.

The NLME approach

Several NLME modeling software tools tailored for the analysis of pharmacological data are available, e.g. NONMEM, Phoenix NLME and Monolix. These tools are very powerful for the analysis of population PK data combined with classical compartmental analysis. However, the lumped PBPK model cannot easily be represented as simple compartmental model with classical apparent PK parameters. Using th lumping the PBPK model parameter are dynamically transformed into the apparent PK parameter on the basis of a physiological database and a set of convoluted functions, e.g. functions predicting distribution properties of each tissue. Not all NLME software tools allow such flexibility or at least not with reasonable expense. For the implementation of a PBPK model and the lumping approach, mathematical software like Matlab or R offer an optimal environment. For both software packages NLME estimation methods are available. For rather standard problems the respective developers have demonstrated that the estimation performs reliably. However, for non–standard problems these functions do not work satisfyingly. For example, Girard and Mentre [2005] and Plan et al. [2012] compared different algorithms and showed that R–nlme performed worst in terms of successful convergence and bias in the parameter estimates.

Therefore, we implemented a robust NLME estimation algorithm such that the proposed mechanistic NLME modeling approach can be applied independently of pure PBPK modeling software, such as PK–Sim or Simcyp Simulator and NLME software tools.

To analyze repeated measurement studies using NLME modeling the maximum likelihood (ML) principle is typically applied. Due to the nonlinearity of the structural model w.r.t. the random effects, for NLME there usually does not exist a closed-form expression of the likelihood function and algorithms based on deterministic/analytical approximations or stochastic approaches have to be used to determine the ML estimates.

Several authors investigated the pros and cons of the different approaches, e.g. see Liu and Wang [2016]; Bauer et al. [2007]; Gibiansky et al. [2012]; Plan et al. [2012]. The similar conclusion of all authors is that stochastic algorithms work better for complex model, and/or sparse data situations, while for simple structured model and dense data the deterministic

approaches are usually faster and perform sufficiently well.

Typically, mechanistic PK modeling is applied in the analysis of rather small studies with dense sampling for all subjects, e.g. preclinical studies or clinical phase I studies dedicated to investigate PK and/or interactions with comedication or food. The lumping determines the relationship between PBPK model parameters and apparent PK parameters. Thus, even if this relationship might be rather complex, the lumped model itself is given by a classical compartmental model. Therefore, in view of the designated application—the levofloxacin dataset consists of a small population with dense PK sampling, we used an estimation method based on a deterministic approximation approach.

Howsoever, the hereafter presented mechanism-driven modeling approach is independent of the choice of the NLME software and estimation algorithm. In particular, in case mechanistic modeling is applied on sparse data or the applied NLME model is rather complex, e.g. in terms of random effects or structural model, stochastic approximations method should be considered as an alternative.

The first deterministic NLME estimation methods in PK were introduced in a series of publications Sheiner and Beal [1980, 1981, 1983]; Sheiner [1984]; Beal and Sheiner [1982]. Thereby, the derivation delivered by Sheiner and Beal is rather heuristically motivated than mathematically founded. About 25 years later, Wang [2007] provided a mathematical derivation of these approximation methods. However, Wang [2007] only derived the approximations in case of an univariate random variable. Furthermore, not all details in the derivation are adequately justified.

Concludingly, in Part I, the most commonly applied deterministic methods in the analysis of population PK data are introduced and the mathematical methodology of NLME modeling is systemically and comprehensively described. In addition, a new and clear interpretation of these methods is presented, which appears to be useful to illustrate the corresponding similarities and dissimilarities. To finally use these approximation in the context of parameter estimation, iterative algorithms have to be applied. A popular alternative to these approaches is given by the expectation-maximization (EM) algorithm, because the computation of the likelihood function as mentioned above is avoided.

The available EM algorithms can be separated into either based on deterministic or stochastic approximations, such as the stochastic approximation expectation–maximization (SAEM) algorithm Kuhn and Lavielle [2005]. The EM algorithm consists of the expectation 'E' step followed by the maximization 'M' step, which are iterated until convergence is reached. In context of NLME modeling, the available EM algorithms, both deterministic and stochastic, are usually based on a similar M–step, while the main difference is the E–step. Following Walker [1996]; Mentre and Gomeni [1995] we will describe a deterministic EM algorithm. Furthermore, the description of the M–step by Walker [1996]; Mentre and Gomeni [1995] is complemented by the details how fixed effects which are not associated with random effects can be estimated.

Finally, the presented deterministic EM algorithm is evaluated in a simulation & estimation study using a classical compartmental PK model and compared to the SAEM algorithm implemented in the R–package *saemix* (Comets et al. [2011]). For the underlying situation, this comparison demonstrates that the deterministic estimation approach performs sufficiently well. The details can be found in the Section 3.3.9.

Population PBPK modeling of levofloxacin

In Chapter 5, using the deterministic EM algorithm, we demonstrate how mechanistical PBPK modeling and the lumping methodology by Pilari and Huisinga [2010] can be combined to analyze population PK data and illustrate this for the drug levofloxacin.

To this aim, a PBPK model as described by Pilari and Huisinga [2010] and the tissue-distribution model introduced by Rodgers et al. [2005b] is used. The model parameters for the human physiology and physicochemical properties of levofloxacin are extracted from different literature sources. In a next step, we developed a new approach to include known physiological-based sources of BSV into the PBPK model by individualizing the physiological information within the PBPK model based on patient's covariates. The essentials of this new approach was published in Huisinga et al. [2012]. The new approach—a linear scaling of the organ size with lean body weight—most appropriately reproduces the observed BSV in organ size compared to other existing approaches. Additionally, in Huisinga et al. [2012] we demonstrated how the BSV in apparent PK parameters of classical compartmental models is expected to be affected by patient's body size. The successful quantification of such covariate effects based on population PK analysis is often very challenging because the effect size is relatively weak compared to the variability which is usually present in the data. On top of this individualization, due the importance of the cardiac output parameter for the drug distribution, a literature–based uncertainty is used to refine the individual cardiac output parameter in the NLME estimation.

In a last step, to examine the link between tissue concentration predicted by PBPK modeling and measurements obtained by microdialysis, the details of the experimental setting of the microdialysis method are reproduced and considered in the model. Therefore, it has to be considered that in the microdialysis experiment the drug concentration is collected over predefined time periods and not measured at specific time points as plasma PKs. Moreover, the experimentally determined correction factor and, importantly the associated uncertainty are incorporated into this assessment. On the other hand, because the microdialysis measurement are taken from the ISF—belonging to the vascular space of the tissue—the corresponding concentration within the PBPK model has to be determined.

The simulations of the microdialysis measurements shows that the expected uncertainty and variability is rather high. However, the measured microdialysis tissue data lies within the predicted range. Concludingly, within the framework of this simulation, the developed PBPK model adequately predicted the average microdialysis tissue concentrations as well as the associated variability for both adipose and muscle tissue. For the first time, such a comparison of human tissue PK data obtained by microdialysis and PBPK predictions is presented.

In summary, a new approach how PBPK models can be used to analyze population PK data, integrate sources of known variability and to quantify the unexplained between patient variability is presented. Additionally, the new approach is used to estimate missing physicochemical properties and to adapt the literature–based physiological parameters. Finally, clinical microdialysis measurements and the population PBPK modeling approach were compared. The developed population PBPK model is able to adequately predict adipose and muscle ISF tissue concentrations, emphasizing the consistency between both approaches.

Mechanistic NLME modeling in Psycholinguistics

Obviously, the difficulty to combine mechanistic and NLME modeling is not only confined to the field of pharmacology. Furthermore, mechanistic modeling approaches are not limited to motivate only the structural, but also the stochastic part of a model. For example, in the field of psycholinguistic Magnuson et al. [2003] conducted an experiment to study the recognition competition of phonologically similar words. The study participants listened to vocal utterances and were exposed to visual stimuli at the same time while their eye movements were tracked over time. The experiments were repeated N times and the results were represented in the form that at time t participant 'X' fixated item '1' in K_1, item '2' in K_2 and item '3' in K_3 out of the N experiments. Mirman et al. [2008] based their analysis of this study on the aggregated data $K_j/N,\ 1 \leq j \leq 3$ and used linear mixed effects modeling assuming a normally distributed WSV model. Thus, they neglected the fact that $K_j/N \in [0, 1]$ which differs from the support of the normal distribution $(-\infty, \infty)$, but even more, they neglected the discrete stochastic nature of the experiment and as a consequence accepted a loss of information.

As a structural model Mirman et al. [2008] used an empirically motivated polynomial function with degree 3 and analyzed each of the three items separately. This approach has a number of drawbacks, (i) the parameters of polynomial function with degree 3 do not have a direct link to the process of interest or the experiment, (ii) the dependency, i.e. the competition, between the measured frequency of the three items K_1, K_2 and K_3 has been neglected and (iii) through the aggregation, the discrete nature of the experiment and for this reason the number of experiment executions N has been neglected.

Therefor, in Chapter 6 in a first step, we developed a structural model which takes into account the competition between the phonologically similar words. The structural model describes the probability time-course to look at each of the three items. Because the observations in this experiment are discrete items, which were fixated, we describe a discrete WSV model motivated by the experimental setting and included this into the NLME analysis of the eye movement data. The conjugated Dirichlet-multinomial distribution is chosen as WSV model, where the multinomial distribution takes into account the multivariate, discrete stochastic nature of the observations, while the Dirichlet distribution allows for flexible modeling of the shape of the WSV distribution, in particular w.r.t. the order of magnitude of the variance and the skewness of the distribution.

With the results of the analysis we were able show that the mentioned drawbacks of the Mirman et al. [2008] analysis could be solved. Using conjugated priors in NLME modeling offers potential for further investigation, for example how conjugated priors of the normal distribution could be incorporated in NLME modeling and EM context.

In summary, in this thesis the mathematical framework of NLME modeling based on deterministic approximations is systemically presented and an effective NLME estimation algorithm is described. Thereon based, we introduced a new approach showing how mechanism-driven NLME modeling can be used in the analysis of repeated measurement studies for two real world problems from quantitative pharmacology and psycholinguistics.

Part I.

The Mathematical Theory of Nonlinear Mixed Effects Modeling

1.1. Introduction

In a repeated measurement study, the quantity of interest is measured for the same subject under varying experimental conditions. The objective is to investigate the relationship between the varying conditions—the independent variables x—and the measurement—the dependent variables y, and to describe this process with a mathematical model f. This design arises in many life sciences with the aim to quantitatively describe a process not only for a single subject, but for the entire population of interest. To this end, the repeated measurement design is applied to a number of subjects $1, \dots, N$ assuming them to be a random sample of the population of interest. This results in the sample $\boldsymbol{y_1}, \dots, \boldsymbol{y_N}$, with $\boldsymbol{y_i} \in \mathbb{R}^{n_i}$ containing the $n_i \geq 1$ repeated measurements for each subject

$$\boldsymbol{y_i} = (\boldsymbol{y_{i1}}, \dots, \boldsymbol{y_{in_i}})^T,\ 1 \leq i \leq N,$$

with a total number of measurements $N_n := \sum_{i=1}^{N} n_i$.

An established assumption is that the mathematical model f, relating the dependent to the independent variable, is identical across the population. Furthermore, it is assumed that $\boldsymbol{y_1}, \dots, \boldsymbol{y_N}$ are realizations of a random sample $\boldsymbol{Y_1}, \dots, \boldsymbol{Y_N} \sim F_\beta^{\boldsymbol{Y}}$, where $F_\beta^{\boldsymbol{Y}}$ denotes a distribution function which usually depends on some unknown parameter $\beta \in \Upsilon^\beta$, where Υ^β denotes the parameter space of β. For example, a widely applied statistical model for repeated measurement data is given by

$$\boldsymbol{Y_{ij}} = f(x_{ij}; \boldsymbol{\Theta_i}) + \boldsymbol{\epsilon_{ij}},\ \boldsymbol{\epsilon_{ij}} \sim \mathcal{N}(0, \sigma^2),\ \boldsymbol{\Theta_i} \sim F^{\boldsymbol{\Theta}}, \tag{1.1}$$

for $1 \leq j \leq n_i$ and $1 \leq i \leq N$, with $f : \mathbb{R}^{k_{ij}} \times \mathbb{R}^p \to \mathbb{R}$ and $\sigma > 0$, where $k_{ij} \geq 0$ denotes the dimension of the independent variables $x_{ij} \in \mathbb{R}^{k_{ij}}$, $p > 0$ denotes the dimension of $\boldsymbol{\Theta_i}$ and $F^{\boldsymbol{\Theta}}$ denotes the distribution function of $\boldsymbol{\Theta_1}, \dots, \boldsymbol{\Theta_N}$.

Based on these assumption, the so–called population analysis aims to specify the model f, denoted as structural model, the distribution $F^{\boldsymbol{\Theta}}$, denoted as the BSV model, and the so–called WSV model describing the link between the model predictions and the measurements. A widely used and powerful approach to determine these three parts is given by the NLME approach. A NLME model contains fixed and random effects, where fixed effects are those parameters which are assumed to be identical in the population, while the random effects describe the variability within parameters in the study population. This variability between subjects is represented by the random variables $\boldsymbol{\Theta_1}, \dots, \boldsymbol{\Theta_N}$ and their distribution $F^{\boldsymbol{\Theta}}$. The mix of fixed and random effects led to the terminology 'mixed effects modeling'. The term 'nonlinear' indicates that the structural model $f(x_i; \theta_i)$ is a nonlinear function of $\theta_i \in \mathbb{R}^p$.

The WSV model presented in Equation (1.1) is denoted as additive error model, in the following section the most widely applied parametric WSV and BSV models are briefly presented, motivated and discussed. Previously, a brief introduction to pharmacokinetics is given; the PK of a drug is often characterized using NLME modeling and we will use this field for illustration purposes in the first part of this thesis.

In the Chapters 2 and 3, the utility of a NLME analysis is presented and a mathematical methodology and framework to perform them are described. The determination of parameter estimates in the NLME context is based on the maximum likelihood principle; a detailed introduction to the maximum likelihood principle can for example be found in Wasserman

[2004]; Boos and Stefanski [2013] or Wakefield [2013]. The NLME approach will be illustrated based on data from a PK study of the drug levofloxacin in Section 3.3. The example demonstrates that the NLME approach is a very powerful tool to analyse population data from a repeated measurement study. In addition, the example is used to introduce useful modeling tools to diagnose the model's ability to describe the underlying data, and to test the robustness of the model. These tools are then used in the application in Part II.

1.1.1. Pharmacokinetics

Pharmacokinetics (PK) belongs to the field of pharmacology and means *movement of the drug*. Heuristically, it can be described as the science of what the body does to a drug, whereas pharmacodynamics (PD), another field of pharmacology, is the science of what the drug does to the body. The PK of a drug are broadly composed of the following 4 processes:

- Absorption: the process from administration of a drug until it enters the circulating blood,
- Distribution: the process how the drug is distributed in the body,
- Metabolism: how the drug is transformed into other, pharmacological active or non–active, molecules,
- Excretion: how the drug is removed from the body.

These four processes are the basis of the finally observed drug concentration–time course in the body. An comprehensive overview over PK and PD can be found in Rowland and Tozer [2011].

Repeated measurement studies and the corresponding population analysis are a necessary tool to characterize the PK of a drug within a population. The dependent variable is usually the drug concentration in blood plasma or serum taken at different time points after the administration of the drug. The independent variables are the sampling time points, the dosing information and the subject's covariates which might influence the PKs of a drug, such as ethnicity, age, etc.

In PK, in general, the function $f(x_i; \theta_i)$, which describes the concentration–time course of the drug, is a nonlinear function with respect to the parameter θ_i.

1.2. Parametric Within and Between Subject Variability Models

The WSV represents the variability which can arise between measurements within each subject. It usually comprises rather diverse sources. In pharmacokinetics, the main sources for this deviation are, see Bonate [2011, p. 243],

- Uncertainty in the determination of the drug concentration in the plasma or tissue fluid sample due to the bio–analytical method,
- Model misspecification and incorrect assumptions, e.g. error–free documentation of administration times.

In general, the error of the bio–analytical method in the determination of the drug concentrations is proportional to the drug concentration, EMEA [2011, sec. 4]. According to the guideline (EMEA [2011, sec. 4]), the magnitude of this variation has to be quantified based on repeating the concentration determination several times for one of the samples. If the estimated WSV is larger than the variability of the analytical method, the difference can be interpreted as a model misspecification or incorrect assumptions. Naturally, the objective of mathematical modeling is to minimize the magnitude of this remaining part of WSV.

An important implicit assumption in many modeling approaches is the correct determination and documentation of the values of the independent variables. If these values are not error-free, so–called *errors in variables* models can be applied, see [Schennach, 2004].

For example, in pharmacokinetics it sometimes occurs, that the sampling or dosing times are not correctly documented, e.g. see Karlsson et al. [1998]. The sampling time points in a clinical PK study are usually fixed *a priori* to have an as informative as possible sampling regime. However, the exact compliance of this scheme—within the clinical routine—is often not feasible and deviations are most likely not always documented. In general, this can be expected to have negligible impact. However, in a time span where rapid changes in the concentration occur, an incorrect documentation of the sampling time points will have larger impact on the WSV. In general, such rapid changes in concentration appear during and shortly after the dosing, or absorption phase. Naturally, in these phases higher concentrations arise, see Figure 1.1.

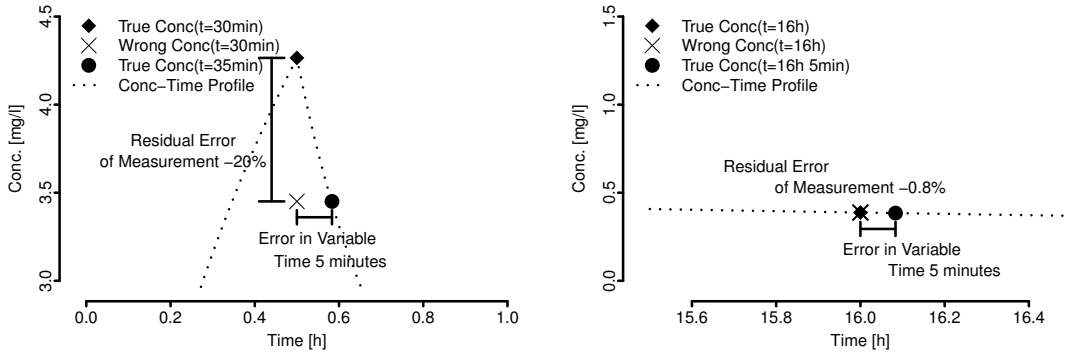

Figure 1.1.: *Errors in Variables: Both plots show the concentration–time profiles of a two compartment model, with a* 30 *minutes infusion. In the left plot a sample is taken at* $t = 35$ *min,* 5 *minutes after the planned time without the documentation of this deviation. The resulting residual variation between true value and measurement is* -12%*. In the right plot, about* 16 *hours later, the sample is measured* 5 *minutes after the planned time again. The deviation here is only* -0.5%*.*

Even so errors in variables might frequently occur, the impact on the correct determination of the model parameters is expected to be small. Thus, in general, one is either not able or not interested to successfully distinguish between the errors in variables and the WSV. For example, Karlsson et al. [1998] investigated the impact of errors in the sampling times. They analyzed the substance moxondine and observed, that "a random shift of ± 10 min according to a uniform distribution did not lead to any appreciable change in the parameter estimates apart from an increase in the residual error magnitude". As a consequence, the uncertainty related to the determination of the independent variables is allocated to the

WSV.

Assume for subject i the observations $\boldsymbol{y_{i1}}, \ldots, \boldsymbol{y_{in_i}}$ are realizations of a random vector

$$\boldsymbol{Y_i} = (\boldsymbol{Y_{i1}}, \ldots, \boldsymbol{Y_{in_i}})^T = f(x_i; \theta_i) + \boldsymbol{\epsilon_i},\ \boldsymbol{\epsilon_i} \sim \mathcal{N}_{n_i}(0, I_{n_i}\sigma_i^2), \tag{1.2}$$

with $\boldsymbol{\epsilon_i} = (\boldsymbol{\epsilon_{i1}}, \ldots, \boldsymbol{\epsilon_{in_i}})^T$, $f(x_i; \theta_i) := (f(x_{i1}; \theta_i), \ldots, f(x_{in_i}; \theta_i))^T$, $I_{n_i} = \operatorname{diag}(1, \ldots, 1) \in \mathbb{R}^{n_i \times n_i}$ and $\theta_i \in \mathbb{R}^p$. As a consequence $\boldsymbol{Y_i}$ is multivariate normally distributed with entries $\boldsymbol{Y_{ij_1}}$ and $\boldsymbol{Y_{ij_2}}$, $j_1 \neq j_2$, being independent of each other. The model presented in Equation (1.2) is called additive error model and reflects a constant, concentration independent variance around the model predictions.

A reasonable solution to consider the behavior illustrated in Figure 1.1 and the proportionality of the bio–analytical error is to assume the WSV to be concentration dependent, i.e. the magnitude of WSV increases with the model predictions. A common strategy to model a concentration dependent WSV is to weight the residual term with a weighting function $w : \mathbb{R}^{k_{ij}} \times \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}$, where ζ could be a further parameter to influence the weighting. Thus, the variance-covariance of $\boldsymbol{Y_{ij}}$ is given by

$$\boldsymbol{Y_{ij}} \sim \mathcal{N}_{n_i}\left(f(x_{ij}; \theta_i),\, w(x_{ij}; \theta_i, \zeta)^2\, \sigma_i^2\right). \tag{1.3}$$

The variability in $\boldsymbol{Y_{ij}}$ is often quantified by the coefficient of variation (CV), defined as

$$\mathrm{CV\%}\left(\boldsymbol{Y_{ij}}\right) = 100 \cdot \frac{\sqrt{\mathrm{Var}_{\beta_i}\left[\boldsymbol{Y_{ij}}\right]}}{\mathbf{E}_{\beta_i}\left[\boldsymbol{Y_{ij}}\right]}, \tag{1.4}$$

where $\mathbf{E}_{\beta_i}\left[\boldsymbol{Y_{ij}}\right]$ and $\mathrm{Var}_{\beta_i}\left[\boldsymbol{Y_{ij}}\right]$ denote the expected value and variance–covariance of $\boldsymbol{Y_i}$ defined as

$$\mathbf{E}_{\beta}\left[\boldsymbol{Y_{ij}}\right] = \int y\, p^{\boldsymbol{Y_{ij}}}(y; \beta)\, dy, \tag{1.5}$$

$$\mathrm{Var}_{\beta}\left[\boldsymbol{Y_{ij}}\right] = \int \left(y - \mathbf{E}_{\beta}\left[\boldsymbol{Y_{ij}}\right]\right)^2 p^{\boldsymbol{Y_{ij}}}(y; \beta)\, dy, \tag{1.6}$$

Accordingly, the CV of both introduced error models is given by:

- Additive error model:

$$\boldsymbol{Y_{ij}} \sim \mathcal{N}\left(f(x_{ij}; \theta_i),\, \sigma_i^2\right),$$
$$\mathrm{CV\%}\left(\boldsymbol{Y_{ij}}\right) = 100 \cdot \frac{\sigma_i^2}{f(x_{ij}; \theta_i)}.$$

- Weighted error model:

$$\boldsymbol{Y_{ij}} \sim \mathcal{N}\left(f(x_{ij}; \theta_i),\, w^2(x_{ij}; \theta_i, \zeta)\, \sigma_i^2\right),$$
$$\mathrm{CV\%}\left(\boldsymbol{Y_{ij}}\right) = 100 \cdot \frac{w(x_{ij}; \theta_i, \zeta)\, \sigma_i^2}{f(x_{ij}; \theta_i)}.$$

A popular choice of the weighting function is $w(x_{ij};\theta_i,\zeta) = f(x_{ij};\theta_i)$. Using f as weighting relates the magnitude of WSV for a measurement $\boldsymbol{y_{ij}}$ to its predicted value $f(x_{ij};\theta_i)$.

The resulting error model is called proportional error model

$$\boldsymbol{Y_{ij}} = f(x_{ij};\theta_i)\,(1+\boldsymbol{\epsilon_{ij}}) \text{ with } \mathrm{CV\%}\,(\boldsymbol{Y_{ij}}) = 100\cdot\sigma_i^2. \tag{1.7}$$

In a classical regression context, compared to the additive error model, the corresponding maximum likelihood estimation results in a so–called penalized least squares problem

$$(\hat{\theta}_i,\hat{\sigma}_i^2)(\boldsymbol{Y_i}) = \underset{\theta_i\in\mathbb{R}^p,\sigma_i^2\in\mathbb{R}_{>0}}{\arg\min} \sum_{j=1}^{n_i}\left(\log\left(f^2(x_{ij};\theta_i)\sigma_i^2\right) + \frac{(\boldsymbol{Y_{ij}} - f(x_{ij};\theta_i))^2}{f^2(x_{ij};\theta_i)\sigma_i^2}\right), \tag{1.8}$$

where the estimation of θ_i and σ_i^2 are not independent of each other.

In PK, the random variable $\boldsymbol{Y_i}$ describes the measured drug concentrations of subject i. Naturally, concentrations are non–negative. For the additive, as well as the proportional error model, the probability that negative concentrations can be measured is positive $P(\boldsymbol{Y_{ij}} < 0) > 0$ which leads to a theoretical inconsistency. An approach to avoid this is it to apply the log transformation on the data $\boldsymbol{y_i}$ and the model f, and to assume an additive error in the log transformed space

$$\boldsymbol{Y_i} = \log f(x_i;\theta_i) + \boldsymbol{\epsilon_i},\ \boldsymbol{\epsilon_i} \sim \mathcal{N}_{n_i}\left(0,\, I_{n_i}\sigma_i^2\right). \tag{1.9}$$

In regression analysis log–transformations of data and model are a widely used if a heteroscedastic error is assumed, e.g. see Beal and Sheiner [1988].

The additive error model in Equation (1.9) is equivalent to the so called exponential error model

$$\boldsymbol{Y_i} = f(x_i;\theta_i)\,e^{\boldsymbol{\epsilon_i}} \text{ with } \mathrm{CV\%}\,(\boldsymbol{Y_{ij}}) = 100\cdot\sqrt{e^{\sigma_i^2}-1}. \tag{1.10}$$

The exponential error model satisfies the assumption that only positive measurements are possible $P(\boldsymbol{Y_i} \le 0) = 0$ and attains a constant CV%, implying a concentration dependent WSV. Furthermore, for a small WSV σ^2 it is $\sqrt{e^{\sigma^2}-1} \approx \sigma^2$ and the difference between exponential and proportional error model is negligible.

In general, a parametric distribution is assumed for the random effects characterizing the BSV. If the underlying data is sufficient and no distributional assumption regarding the random effects can be made, so–called non–parametric methods can be applied. However, in practice these approaches are rarely used, particularly in life sciences the assumption of normally or log–normally distributed parameter within a population are well accepted, see Karlsson et al. [1998]; Limpert et al. [2001].

The assumption that $\boldsymbol{\Theta_i}$ is multivariate normally distributed implies the possibility to model the single entries of $\boldsymbol{\Theta_i}$ as normally or log–normally distributed. For example, if the k-th entry of $\boldsymbol{\Theta_i}$ is log–normally distributed, this entry can be replaced by $\exp\boldsymbol{\Theta'_{i(k)}}$, where $\boldsymbol{\Theta'_{i(k)}}$ is normally distributed. The marginal distribution of the original $\boldsymbol{\Theta_i}$ and $\boldsymbol{\Theta'_i}$, which is identical to $\boldsymbol{\Theta_i}$ except that the k-th entry is replaced as described, are identical. However,

in general, the exact correlation structure is not retained under the exp–transformation. Although, properties like independence, positive or negative correlations are preserved under strict monotone increasing transformations, e.g. such as log– or exp–transformations, see Nelsen [2007].

In summary, in the following it is assumed that the error is additive. This generalization allows to consider an additive error for untransformed data and model, as well as for log–transformed data and model (exponential error model). Furthermore, a normal distribution is chosen to characterize the BSV ($F^{\boldsymbol{\Theta}}$) .

2. NLME methodology using deterministic approximations of the likelihood function

2.1. Mixed Effects Models

Based on the considerations as presented in Section 1.1, we define a NLME model for the additive error model, with normally distributed error and normally distributed population parameters as

$$\boldsymbol{Y_i} = f(x_i; \boldsymbol{\Theta_i}, \vartheta) + \boldsymbol{\epsilon_i} \text{ with } \boldsymbol{\epsilon_i} \sim \mathcal{N}_{n_i}\left(0,\, I_{n_i}\sigma^2\right) \text{ and } \boldsymbol{\Theta_i} \sim \mathcal{N}_p\left(\theta,\, \Omega\right), \tag{2.1}$$

for $1 \leq i \leq N$, with positive definite covariance matrix $\Omega \in \mathbb{R}^{p\times p}_{>0}$, $\boldsymbol{\epsilon_i}$ and $\boldsymbol{\Theta_i}$ being independent from each other, $\theta \in \mathbb{R}^p$ is usually denoted as *typical* value and $\vartheta \in \mathbb{R}^q$.

Another common parametrization of Equation (2.1) is given by

$$\boldsymbol{Y_i} = \tilde{f}(x_i; \boldsymbol{\eta}, \tilde{\vartheta}) + \boldsymbol{\epsilon_i} \text{ with } \boldsymbol{\epsilon_i} \sim \mathcal{N}_{n_i}\left(0,\, I_{n_i}\sigma^2\right) \text{ and } \boldsymbol{\eta} \sim \mathcal{N}_p\left(0,\, \Omega\right), \tag{2.2}$$

with $\tilde{\vartheta} = (\vartheta, \theta)$, $\boldsymbol{\eta} = \boldsymbol{\Theta_i} - \theta$ and $\tilde{f}(x_i; \boldsymbol{\eta}, \tilde{\vartheta}) = f(x_i; \theta + \boldsymbol{\eta}, \vartheta)$.

The parameter vectors θ and ϑ are constant within the population; therefore, they are called fixed effects. The random variables $\boldsymbol{\Theta_i}$ or $\boldsymbol{\eta}$ describe the variability between subjects and are called random effects. The distinction between parameters that are constant across the population ('pure' fixed effects ϑ and fixed effects associated with random effects θ) and parameters that vary across the population represents a crucial advantage of the NLME approach. For example, in general, in a study population PK parameters are typical examples of random effects $\boldsymbol{\Theta_i}$ and the typical values of these parameters would be represented by fixed effects associated with random effects (θ). However, in many examples BSV cannot be identified in all PK parameters; these parameters would be considered as fixed effects not associated with random effects (ϑ).

The distribution of $\boldsymbol{Y_i}$ is defined by the distributions of $\boldsymbol{\Theta_i}$ and $\boldsymbol{\epsilon_i}$. The likelihood function to derive the ML estimates is given by

$$L_N\left(\vartheta, \theta, \Omega, \sigma^2; \boldsymbol{y_1}, \ldots, \boldsymbol{y_N}\right) := \prod_{i=1}^{N} p^{\boldsymbol{Y_i}}\left(\boldsymbol{y_i}; \vartheta, \theta, \Omega, \sigma^2\right) \tag{2.3}$$

has to be maximized.

To determine the analytical expression of L_N, the analytical expression of $p^{\boldsymbol{Y_i}}$ is necessary. However, because f is nonlinear, in general, the distribution of $f\left(x_i; \boldsymbol{\Theta_i}, \vartheta\right)$ and thereby the distribution of $\boldsymbol{Y_i}$ cannot be expressed in an analytically closed form.

Though, using the law of total probability for continuous random variables, $p^{\boldsymbol{Y_i}}$ can be written as

$$p^{\boldsymbol{Y_i}}\left(\boldsymbol{y_i}; \beta\right) = \int_{\mathbb{R}^p} p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}\left(\boldsymbol{y_i}, \theta_i; \vartheta, \sigma^2\right) p^{\boldsymbol{\Theta_i}}\left(\theta_i; \theta, \Omega\right)\, d\theta_i, \tag{2.4}$$

with $\beta = (\vartheta, \theta, \Omega, \sigma^2)$ and where $p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}$ denotes the probability density function (p.d.f.) of the conditional random variable $(\boldsymbol{Y_i}|\boldsymbol{\Theta_i} = \theta_i)$.

The random variables $\boldsymbol{Y_{ij_1}}$ and $\boldsymbol{Y_{ij_2}}$, $j_1 \neq j_2$, describe two different observations at time t_{ij_1} and t_{ij_2} within the same individual, i.e. both depend on $\boldsymbol{\Theta_i}$ and are not independent. However, following the assumption, the conditional random variables $Y_{ij_1}|\boldsymbol{\Theta_i}$ and $Y_{ij_2}|\boldsymbol{\Theta_i}$ are independent and the conditional distribution of $\boldsymbol{Y_i}|\boldsymbol{\Theta_i} = \theta_i$ is given by

$$(\boldsymbol{Y_i}|\boldsymbol{\Theta_i} = \theta_i) \sim \mathcal{N}_{n_i}\left(f\left(x_i; \theta_i, \vartheta\right), I_{n_i}\sigma^2\right), \tag{2.5}$$

with p.d.f.

$$p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}\left(\boldsymbol{y_i}, \theta_i; \vartheta, \sigma^2\right) = \frac{1}{\sqrt{(2\pi\sigma^2)^{n_i}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{n_i}(\boldsymbol{y_{ij}} - f(x_{ij}; \theta_i, \vartheta))^2\right). \tag{2.6}$$

Furthermore, the p.d.f. $p^{\boldsymbol{\Theta_i}}$ is given by

$$p^{\boldsymbol{\Theta_i}}(\theta_i; \theta, \Omega) = \frac{1}{\sqrt{(2\pi)^p|\Omega|}} \exp\left(-\frac{1}{2}(\theta_i - \theta)^T\Omega^{-1}(\theta_i - \theta)\right). \tag{2.7}$$

In the following, the abbreviations $l_i(\theta_i) := p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}(\boldsymbol{y_i}; \theta_i\vartheta, \sigma^2)$ representing the WSV and $h(\theta_i) := p^{\boldsymbol{\Theta_i}}(\theta_i; \theta, \Omega)$ represents the BSV are used.

Thus, Equation (2.4) can be written as

$$p^{\boldsymbol{Y_i}}(\boldsymbol{y_i}; \beta) = \int l_i(\theta_i)h(\theta_i)\, d\theta_i. \tag{2.8}$$

In general, no analytical solution of the integral in Equation (2.8) exists if f is a nonlinear function. As a consequence, the likelihood function L_N has no closed-form expression on the basis of which the ML estimates can be defined. Concludingly, using a deterministic approach to NLME modeling the unknown probability distribution of $\boldsymbol{Y_i}$ has to be approximated. To this end, a commonly used approach is to approximate the integrand in Equation (2.8) such that the integral can be solved and expressed analytically.

In the next subsection the most commonly used approximation—namely, the Laplacian, the first–order conditional estimation (FOCE) and the first-order (FO) method—are described. The Laplacian and FOCE approximation consist of two steps: first, the mode $\hat{\theta}_i$ of $l_i \cdot h$ given β and $\boldsymbol{y_i}$ is determined, secondly, using the so–called Laplace method $p^{\boldsymbol{Y_i}}$ is approximated based on $\hat{\theta}_i$. The FO used the typical population value θ instead of the mode for the Laplace method. Because the estimation of $\hat{\theta}_i$ in the Laplacian and FOCE method itself is a nonlinear regression problem, only the FO method provide a close-form expression of the approximated likelihood function which can be optimized, e.g. using numerical methods, to determine the ML estimates. The determination of ML estimates based on the Laplacian and FOCE method is more complex and an alternating process is necessary: first, $\hat{\theta}_i$ is determined for all subjects $1 \leq i \leq N$ and secondly, the approximated likelihood function given the estimates $\hat{\theta}_1, \ldots, \hat{\theta}_N$ is optimized with respect to β, see Beal and Sheiner [1998, p. 3].

An alternative iterative approach is based on the EM algorithm, e.g. see Dempster et al. [1977]. The advantage of the EM algorithm consists of the fact that no closed-form solution

or approximation of the likelihood function is necessary. Given the parameters $\beta^{(k)}$ the algorithm iteratively determines new parameter values $\beta^{(k+1)}$ such that

$$L_N(\beta^{(k+1)}; \boldsymbol{y_1}, \ldots, \boldsymbol{y_N}) \geq L_N(\beta^{(k)}; \boldsymbol{y_1}, \ldots, \boldsymbol{y_N}). \tag{2.9}$$

An iteration of EM algorithm consists of two steps, the E-step and the M-step which are iterated until a pre-defined convergence criterion is reached. Several EM algorithms in the context of NLME modeling exists, which usually have a similar M-step, while the E-step differs. The most popular EM algorithms in the context of NLME modeling use stochastic approximation in the E-step (Kuhn and Lavielle [2005]; Walker [1996]). These algorithms share the advantage that the approximation accuracy can be controlled with the number of iterations in the stochastic approximation step. I.e. increasing the number of iterations, which comes with an increase of computational cost, improves accuracy. In contrast to this, the accuracy of deterministic approximations mainly depends on the complexity of the NLME model, in particular the complexity of f and the choice of the random effects model, and the available data. As a consequence, for complex models and/or sparse data situations stochastic methods seems to be superior in terms of stability and parameter bias compared to deterministic methods, e.g. see Liu and Wang [2016]; Bauer et al. [2007]; Gibiansky et al. [2012]; Plan et al. [2012]. Nevertheless, for rather simpler NLME models and dense data Liu and Wang [2016]; Bauer et al. [2007]; Gibiansky et al. [2012]; Plan et al. [2012] consistently conclude that deterministic methods seem to perform sufficiently well. Mechanistic PK modeling is usually applied in the analysis of rather small studies with dense sampling for all subjects, e.g. preclinical studies or clinical phase I studies dedicated to investigate PK and/or PK interactions with comedication or food. Concludingly, if the NLME model set up is not too complex, combining mechanistic modeling with deterministic NLME methods is expected to be sufficient, while typically going with lower computational effort compared to a stochastic method.

Such a deterministic EM algorithm in the context of NLME modeling, where the E-step is derived using deterministic approximations, is described in Section 3.1. As the basis for this, the most commonly used deterministic approximations of the likelihood function are described in the next section.

2.2. Deterministic Approximations of the Likelihood Function

2.2.1. Introduction

In this section the Laplacian, the FOCE and the FO approximations of likelihood function defined in Equation (2.4) are presented. These approximations are described in several publications; a good overview can be found in Wang [2007]. For simplification, Wang [2007] only considered an univariate random effect $\boldsymbol{\Theta_i}$, while in the following this is generalized to the multivariate case. Moreover, in the following the differences between the three approximation methods are illustrated.

All three methods are based on the Laplace method for the approximation of integrals of the form

$$I(m, g) := \int e^{mg(\theta_i)}\, d\theta_i,$$

for a three times continuous differentiable function g with $I(m, g) < \infty$ for all $m > 0$. The approximation of the integral is based on the approximation of g by its second order Taylor series approximation $\tilde{g}$ around $\hat{\theta}_i \in \mathbb{R}^p$

$$g(\theta_i) = \underbrace{g(\hat{\theta}_i) + \nabla_{\theta_i} g(\hat{\theta}_i)^T(\theta_i - \hat{\theta}_i) + \frac{1}{2}(\theta_i - \hat{\theta}_i)^T \nabla^2_{\theta_i} g(\hat{\theta}_i)(\theta_i - \hat{\theta}_i)}_{=:\tilde{g}(\theta_i)} + R_3(\theta_i), \tag{2.10}$$

with a rest term $R_3(\theta_i) \to 0$ for $\theta_i \to \hat{\theta}_i$. The integral over the second order polynomial $\tilde{g}$ can be solved analytically by relating $\tilde{g}$ to the multivariate normal p.d.f. and using the fact that the integral over the p.d.f. is always identical to 1

$$I(m, \tilde{g}) = e^{mg(\hat{\theta}_i)} \sqrt{\frac{(2\pi)^p}{m|-\nabla^2 g(\hat{\theta}_i)|}} e^{-\frac{1}{2} m \nabla g(\hat{\theta}_i)^T \left(\nabla^2 g(\hat{\theta}_i)\right)^{-1} \nabla g(\hat{\theta}_i)} \tag{2.11}$$

If $\hat{\theta}_i = \arg\max_{\theta_i \in \mathbb{R}^p} g(\theta_i)$, the approximation error of the approximation in Equation (2.11) vanishes with increasing $m \to \infty$, see De Bruijn [1958, chap. 4] or Wong [2001, chap. 2].

Using the Laplace method for Equation (2.8) the term mg is given by $g(\theta_i) := \log l_i(\theta_i) + \log h(\theta_i)$ with m is fixed to 1. I.e. given the observations, the model and the parameter values, only the choice of $\tilde{g}$ and therewith the choice of $\hat{\theta}_i$ influences the approximation accuracy.

Approximating g by its second order Taylor approximation at $\hat{\theta}_i \in \mathbb{R}^p$ and applying the Laplace method to Equation (2.8) leads to

$$-2 \log p^{\boldsymbol{Y_i}}(\boldsymbol{y_i}; \beta) \approx -p \log 2\pi - 2g(\hat{\theta}_i) + \log \left| -\nabla^2 g(\hat{\theta}_i) \right| \tag{2.12}$$

$$+ \nabla g(\hat{\theta}_i)^T \left(\nabla^2 g(\hat{\theta}_i) \right)^{-1} \nabla g(\hat{\theta}_i). \tag{2.13}$$

Vonesh [1996] stated that the associated approximation accuracy is improving with increasing n_i. To determine $\tilde{g}$, the first and second derivatives of $\log l_i$ and $\log h$ have to be derived

$$\nabla \log l_i(\theta_i) = \frac{\nabla l_i(\theta_i)}{l_i(\theta_i)} = \frac{1}{\sigma^2} \sum_{j=1}^{n_i} \nabla_{\theta_i} f(x_{ij}; \theta_i, \vartheta) \, (\boldsymbol{y_{ij}} - f(x_{ij}; \theta_i, \vartheta)), \tag{2.14}$$

$$\nabla \log h(\theta_i) = \frac{\nabla h(\theta_i)}{h(\theta_i)} = \Omega^{-1}(\theta - \theta_i), \tag{2.15}$$

$$\begin{aligned} \nabla^2 \log l_i(\theta_i) &= \frac{\nabla^2 l_i(\theta_i)}{l_i(\theta_i)} - \frac{\nabla l_i(\theta_i) \nabla l_i(\theta_i)^T}{l_i^2(\theta_i)} \\ &= \frac{1}{\sigma^2} \sum_{j=1}^{n_i} \nabla^2_{\theta_i} f(x_{ij}; \theta_i, \vartheta) \, (\boldsymbol{y_{ij}} - f(x_{ij}; \theta_i, \vartheta)) \\ &\quad - \frac{1}{\sigma^2} \sum_{j=1}^{n_i} \nabla_{\theta_i} f(x_{ij}; \theta_i, \vartheta) \nabla_{\theta_i} f(x_{ij}; \theta_i, \vartheta)^T, \end{aligned} \tag{2.16}$$

$$\nabla^2 \log h(\theta_{i0}) = -\Omega^{-1}. \tag{2.17}$$

Considering that the p.d.f. of $\boldsymbol{Y_i}$ does not depend on θ_i, i.e. $\nabla^2_{\theta_i} \log p^{\boldsymbol{Y_i}}(\boldsymbol{y_i}; \beta) = 0$, the Hessian of g can be written as

$$\nabla^2_{\theta_i} g(\theta_i) = \nabla^2_{\theta_i} \left(\log l_i(\theta_i) + \log h(\theta_i) - \log p^{\boldsymbol{Y_i}}(\boldsymbol{y_i}; \beta) \right). \tag{2.18}$$

On the other hand, according to Bayes' theorem it is

$$\log p^{\boldsymbol{\Theta_i}|\boldsymbol{Y_i}=\boldsymbol{y_i}}(\theta_i; \beta, \boldsymbol{y_i}) = \log \underbrace{p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}(\boldsymbol{y_i}; \theta_i, \vartheta, \sigma^2)}_{=l_i(\theta_i)} + \log \underbrace{p^{\boldsymbol{\Theta_i}}(\theta_i; \theta, \Omega)}_{=h(\theta_i)} - \log p^{\boldsymbol{Y_i}}(\boldsymbol{y_i}; \beta). \tag{2.19}$$

Combining Equations (2.18) and (2.19) implies that

$$\nabla^2 g(\theta_i) = \nabla^2_{\theta_i} \log p^{\boldsymbol{\Theta_i}|\boldsymbol{Y_i}=\boldsymbol{y_i}}(\theta_i; \boldsymbol{y_i}, \beta) =: -\left(C^{\boldsymbol{\Theta_i}|\boldsymbol{Y_i}=\boldsymbol{y_i}}(\theta_i) \right)^{-1}.$$

For a random vector $\boldsymbol{X} \sim \mathcal{N}_r(\mu, \Sigma)$ the Hessian of its log p.d.f. is the negative, inverse of the variance–covariance matrix, i.e.

$$\nabla^2_{\boldsymbol{x}} \log p^{\boldsymbol{X}}(\boldsymbol{x}; \mu, \Sigma) = -\Sigma^{-1}.$$

Hence, if the random variable $(\boldsymbol{\Theta_i}|\boldsymbol{Y_i} = \boldsymbol{y_i})$ is normally distributed, the respective variance–covariance matrix is given by $C^{\boldsymbol{\Theta_i}|\boldsymbol{Y_i}=\boldsymbol{y_i}}(\theta_i)$. In the following, an abbreviated notation is used for the conditional variable $\boldsymbol{Z_i} := (\boldsymbol{\Theta_i}|\boldsymbol{Y_i} = \boldsymbol{y_i})$.

In fact, the distribution of $\boldsymbol{Z_i}$ depends on the distribution of $\boldsymbol{Y_i}$, which is not known. Therefore, it can not be derived analytically. However, from Bayes' theorem we can conclude that an approximation of $p^{\boldsymbol{Y_i}}$ is equivalent to an approximation of $p^{\boldsymbol{Z_i}}$. The latter is presented in the next section and is used to determine the E-step of the deterministic EM algorithm.

The Laplacian, FOCE and FO method which are discussed in the following sections are based on the above Laplace approximation and only differ in the choice of $\hat{\theta}_i$.

2.2.2. Laplacian Method

In the context of NLME models, for the Laplacian approximation of $-2 \log p^{\boldsymbol{Y_i}}$, the Laplace method (see Equation (2.11)) is applied for $\hat{\theta}_i$ as the point where g has its maximum. I.e. given the data $\boldsymbol{y_i}$ and the parameters β, θ_i is defined as

$$\begin{aligned} \hat{\theta}_i(\boldsymbol{y_i}; \beta) &= \underset{\theta_i \in \mathbb{R}^p}{\arg\max} \ \log l_i(\theta_i) + \log h(\theta_i) \\ &= \underset{\theta_i \in \mathbb{R}^p}{\arg\min} \ \frac{1}{\sigma^2} \sum_{j=1}^{n_i} (\boldsymbol{y_{ij}} - f(x_{ij}; \theta_i, \vartheta))^2 + (\theta - \theta_i)^T \Omega^{-1} (\theta - \theta_i). \end{aligned} \tag{2.20}$$

Under suitable regularity conditions, it follows that $\nabla g(\hat{\theta}_i(\boldsymbol{y_i}; \beta)) = 0$ and $\nabla^2 g(\hat{\theta}_i(\boldsymbol{y_i}; \beta))$ is negative definite. The estimate $\hat{\theta}_i(\boldsymbol{y_i}; \beta)$ for subject i is called empirical Bayes estimate (EBE), maximum a-posterior estimate or post-hoc estimate. In NLME modeling, the EBE represents the individual parameter estimates.

Using Equations (2.11, 2.14, 2.15, 2.16) and (2.17), the Laplacian approximation is given by

$$\begin{aligned}-2\log p^{\boldsymbol{Y_i}}(\boldsymbol{y_i};\beta) \approx\ & n_i \log 2\pi + n_i \log \sigma^2 + \frac{1}{\sigma^2}\sum_{j=1}^{n_i}\left(\boldsymbol{y_{ij}} - f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta)\right)^2 \\ & + \log|\Omega| + \left(\theta - \hat{\theta}_i(\boldsymbol{y_i};\beta)\right)^T \Omega^{-1}\left(\theta - \hat{\theta}_i(\boldsymbol{y_i};\beta)\right) \\ & + \log\left|\frac{1}{\sigma^2}\sum_{j=1}^{n_i}\nabla_{\theta_i} f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta)\,\nabla_{\theta_i} f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta)^T\right. \\ & \left. - \frac{1}{\sigma^2}\sum_{j=1}^{n_i}\nabla^2_{\theta_i} f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta)\,(\boldsymbol{y_{ij}} - f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta)) + \Omega^{-1}\right|. \end{aligned} \tag{2.21}$$

In order to apply the Laplacian approximation, in a first step $\hat{\theta}_i(\boldsymbol{y_i};\beta)$ has to be determined. In general, the parameter $\hat{\theta}_i(\boldsymbol{y_i};\beta)$ cannot be derived analytically and has to be determined using numerical optimization methods. As a consequence, the Laplacian approximation does not provide a closed-form expression of $\log p^{\boldsymbol{Y_i}}$ as function of $\boldsymbol{y_i}$ and β.

The following consideration are of use for the deterministic EM algorithm which is described in Section 3.1. The Laplacian approximation in Equation (2.21) is equivalent to

$$\begin{aligned}-\log p^{\boldsymbol{Y_i}}(\boldsymbol{y_i};\beta) \approx\ & -\log p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}(\boldsymbol{y_i};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta,\sigma^2) - \log p^{\boldsymbol{\Theta_i}}(\hat{\theta}_i(\boldsymbol{y_i};\beta);\theta,\Omega) \\ & + \log p^{\boldsymbol{Z_i^{LA}}}(\hat{\theta}_i(\boldsymbol{y_i};\beta);\beta,\boldsymbol{y_i}),\end{aligned}$$

with $\hat{\theta}_i(\boldsymbol{y_i};\beta)$ defined as above and $\boldsymbol{Z_i^{LA}}$ denoting the random variable with distribution defined by the p.d.f.:

$$p^{\boldsymbol{Z_i^{LA}}}(\theta_i;\beta,\boldsymbol{y_i}) = \frac{1}{\sqrt{(2\pi)^p\left|C^{\boldsymbol{Z_i^{LA}}}\right|}}\exp\left(-\frac{1}{2}(\theta_i - \hat{\theta}_i(\boldsymbol{y_i};\beta))^T\left(C^{\boldsymbol{Z_i^{LA}}}\right)^{-1}(\theta_i - \hat{\theta}_i(\boldsymbol{y_i};\beta))\right),$$

where $C^{\boldsymbol{Z_i^{LA}}} := \left(-\nabla^2 g(\hat{\theta}_i(\boldsymbol{y_i};\beta))\right)^{-1}$. Hence, the random variable $\boldsymbol{Z_i^{LA}}$ is normally distributed

$$\boldsymbol{Z_i^{LA}} \sim \mathcal{N}_p\left(\hat{\theta}_i(\boldsymbol{y_i};\beta),\, C^{\boldsymbol{Z_i^{LA}}}\right). \tag{2.22}$$

It is worth noting, that both the p.d.f.s of $\boldsymbol{Z_i}$ and $\boldsymbol{Z_i^{LA}}$ have $\hat{\theta}_i(\boldsymbol{y_i};\beta)$ as their mode. Furthermore, based on Bayes' theorem it is

$$\begin{aligned}\log p^{\boldsymbol{Y_i}}(\boldsymbol{y_i};\beta) = & \log p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}(\boldsymbol{y_i};\theta_i,\sigma^2) + \log p^{\boldsymbol{\Theta_i}}(\theta_i;\theta,\Omega) \\ & - \log p^{\boldsymbol{Z_i}}(\theta_i;\beta,\boldsymbol{y_i}).\end{aligned}$$

This implies that approximating the distribution of $\boldsymbol{Z_i}$ by the distribution of $\boldsymbol{Z_i^{LA}}$ as defined in Equation (2.22) and applying Bayes' theorem leads to the Laplacian approximation of $p^{\boldsymbol{Y_i}}$ for $\theta_i = \hat{\theta}_i(\boldsymbol{y_i};\beta)$.

2.2.3. First–Order Conditional Estimation Method

A further simplification of the Laplacian approximation is given by the FOCE approximation. Reconsidering the definition of observed and expected Fisher information (see for example Wakefield [2013, p. 37ff]), the term $-\nabla^2 \log l_i(\theta_i)$ represents an estimate of the observed Fisher information of $(\boldsymbol{Y_i}|\boldsymbol{\Theta_i} = \theta_i)$ given the data $\boldsymbol{y_i}$. The idea behind the FOCE method is to approximate the estimate of the observed Fisher by an estimate of the (expected) Fisher information $-\mathbf{E}_\beta\left[\nabla^2_{\theta_i} \log p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}((\boldsymbol{Y_i}|\boldsymbol{\Theta_i} = \theta_i); \theta_i, \vartheta, \sigma^2)\right]$. As a consequence, the often numerical challenging computation of $\nabla^2 l_i$ is avoided, because

$$\begin{aligned} -\mathbf{E}_\beta\left[\nabla^2_{\theta_i} \log p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}(\boldsymbol{Y_i}|\boldsymbol{\Theta_i} = \theta_i; \theta_i, \vartheta, \sigma^2)\right] &= \frac{1}{\sigma^2}\sum_{j=1}^{n_i} \nabla_{\theta_i} f(x_{ij}; \theta_i, \vartheta)\, \nabla_{\theta_i} f(x_{ij}; \theta_i, \vartheta)^T \\ &\quad - \frac{1}{\sigma^2}\sum_{j=1}^{n_i} \nabla^2_{\theta_i} f(x_{ij}; \theta_i, \vartheta)\, \underbrace{(\mathbf{E}_\beta\left[\boldsymbol{Y_{ij}}|\boldsymbol{\Theta_i} = \theta_i\right] - f(x_{ij}; \theta_i, \vartheta))}_{=0}, \end{aligned} \tag{2.23}$$

see Equation (2.16). Using $-\nabla^2 \log l_i(\theta_i) \approx -\mathbf{E}_\beta\left[\nabla^2_{\theta_i} \log p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}(\boldsymbol{Y_i}|\boldsymbol{\Theta_i} = \theta_i; \theta_i, \vartheta, \sigma^2)\right]$ leads to the following simplification

$$\nabla^2 g(\theta_i) \approx -\frac{1}{\sigma^2}\sum_{j=1}^{n_i} \nabla_{\theta_i} f(x_{ij}; \theta_i, \vartheta)\, \nabla_{\theta_i} f(x_{ij}; \theta_i, \vartheta)^T - \Omega^{-1} =: -\left(C^{\boldsymbol{Z_i^{FOCE}}}\right)^{-1}. \tag{2.24}$$

Finally, for $\theta_i = \hat{\theta}_i(\boldsymbol{y_i}; \beta)$ (see Equation (2.20)) the FOCE approximation of $\log p^{\boldsymbol{Y_i}}$ is given by

$$\begin{aligned} -2\log p^{\boldsymbol{Y_i}}(\boldsymbol{y_i}; \beta) \approx\ & n_i \log 2\pi + n_i \log \sigma^2 + \frac{1}{\sigma^2}\sum_{j=1}^{n_i} (\boldsymbol{y_{ij}} - f(x_{ij}; \hat{\theta}_i(\boldsymbol{y_i}; \beta), \vartheta))^2 \\ &+ \log|\Omega| + (\theta - \hat{\theta}_i(\boldsymbol{y_i}; \beta))^T \Omega^{-1} (\theta - \hat{\theta}_i(\boldsymbol{y_i}; \beta)) \\ &+ \log\left|\frac{1}{\sigma^2}\sum_{j=1}^{n_i} \nabla_{\theta_i} f(x_{ij}; \hat{\theta}_i(\boldsymbol{y_i}; \beta), \vartheta)\, \nabla_{\theta_i} f(x_{ij}; \hat{\theta}_i(\boldsymbol{y_i}; \beta), \vartheta)^T + \Omega^{-1}\right|. \end{aligned} \tag{2.25}$$

Analogously to the Laplacian method, approximating the distribution $\boldsymbol{Z_i}$ by the distribution of

$$\boldsymbol{Z_i^{FOCE}} \sim \mathcal{N}_p\left(\hat{\theta}_i(\boldsymbol{y_i}; \beta),\, C^{\boldsymbol{Z_i^{FOCE}}}\right) \tag{2.26}$$

and applying Bayes' theorem leads to the FOCE approximation in $\theta_i = \hat{\theta}_i(\boldsymbol{y_i}; \beta)$.

According to Wang [2007], the FOCE approximation is analogous to approximating the distribution of $\boldsymbol{Y_i}$ by the distribution of $\tilde{\boldsymbol{Y}}_{\boldsymbol{i}}$, with

$$\tilde{\boldsymbol{Y}}_{\boldsymbol{i}} = f(x_i; \hat{\theta}_i(\boldsymbol{y_i}; \beta), \vartheta) + \mathcal{J}_{\theta_i} f(x_i; \hat{\theta}_i(\boldsymbol{y_i}; \beta), \vartheta)\, (\boldsymbol{\Theta_i} - \hat{\theta}_i(\boldsymbol{y_i}; \beta)) + \boldsymbol{\epsilon_i}, \tag{2.27}$$

with $\boldsymbol{\Theta_i} \sim \mathcal{N}_p(\theta, \Omega)$, $\boldsymbol{\epsilon_i} \sim \mathcal{N}_{n_i}\left(0, I_{n_i}\sigma^2\right)$ and $\mathcal{J}_{\theta_i} f(x_i; \hat{\theta}_i(\boldsymbol{y_i}; \beta), \vartheta)$ denoting the Jacobian matrix of f w.r.t. θ_i evaluated at $\hat{\theta}_i(\boldsymbol{y_i}; \beta)$. This implies, that $\tilde{\boldsymbol{Y}}$ is normally distributed

with p.d.f. defined in Equation (2.25) and with

$$\begin{aligned}\mathbf{E}_\beta\left[\tilde{\boldsymbol{Y}}\right] &= f(x_i;\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta) + \mathcal{J}_{\theta_i} f(x_i;\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta)\,(\theta - \hat{\theta}_i(\boldsymbol{y_i};\beta)),\\ \mathrm{Var}_\beta\left[\tilde{\boldsymbol{Y}}\right] &= \mathcal{J}_{\theta_i} f(x_i;\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta)\,\Omega\,\mathcal{J}_{\theta_i} f(x_i;\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta) + I_{n_i}\sigma^2.\end{aligned}$$

A heuristic argument to derive Equation (2.27) is to replace f by its first order approximation at $\theta_i = \hat{\theta}_i(\boldsymbol{y_i};\beta)$

$$f(x_{ij};\theta_i,\vartheta) \approx f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta) + \nabla_{\theta_i} f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta)\,(\boldsymbol{\Theta_i} - \hat{\theta}_i(\boldsymbol{y_i};\beta)). \tag{2.28}$$

Analogously, to the previous Section 2.2.2, the FOCE approximation does not have a closed-form expression, if $\hat{\theta}_i(\boldsymbol{y_i};\beta)$ cannot be expressed in closed form. From a theoretical point of view, the approximation accuracy of the FOCE method for g is expected to be worse compared to the Laplacian method at least around $\hat{\theta}_i(\boldsymbol{y_i};\beta)$ because an additional approximation step is used.

2.2.4. First–Order Method

In the context of PK data analysis, Beal and Sheiner [1982] introduced the FO method to derive ML estimates of a NLME model. The FO approximation consists of two simplifications compared to the Laplacian approximation in Section 2.2.2. The first simplification is to approximate g by its Taylor approximation around the expectation of $\boldsymbol{\Theta_i}$, i.e. $\hat{\theta}_i := \mathbf{E}_\beta[\boldsymbol{\Theta_i}] = \theta$ (compared to using the mode $\hat{\theta}_i(\boldsymbol{y_i};\beta)$). In general, θ is not the mode of g and therefore it is $\nabla g(\theta) \neq 0$. The second simplification—analogously to the FOCE approximation—consists in approximation an estimate of the observed by an estimate of the expected Fisher information of $(\boldsymbol{Y_i}|\boldsymbol{\Theta_i} = \theta_i)$. As a result, the FO approximation of $\log p^{\boldsymbol{Y_i}}$ is given by

$$\begin{aligned}-2\log p^{\boldsymbol{Y_i}}(\boldsymbol{y_i};\beta) \approx\ & n_i\log 2\pi + n_i\log\sigma^2 + \frac{1}{\sigma^2}\sum_{j=1}^{n_i}(\boldsymbol{y_{ij}} - f(x_{ij};\theta,\vartheta))^2\\ &+ \log|\Omega| + \log\left|\frac{1}{\sigma^2}\sum_{j=1}^{n_i}\nabla_{\theta_i} f(x_{ij};\theta,\vartheta)\,\nabla_{\theta_i} f(x_{ij};\theta,\vartheta)^T + \Omega^{-1}\right|\\ &+ \left(\frac{1}{\sigma^2}\sum_{j=1}^{n_i}\nabla_{\theta_i} f(x_{ij};\theta,\vartheta)\,(\boldsymbol{y_{ij}} - f(x_{ij};\theta,\vartheta))\right)^T\\ &\cdot\left(-\frac{1}{\sigma^2}\sum_{j=1}^{n_i}\nabla_{\theta_i} f(x_{ij};\theta,\vartheta)\,\nabla_{\theta_i} f(x_{ij};\theta,\vartheta)^T - \Omega^{-1}\right)^{-1}\\ &\cdot\left(\frac{1}{\sigma^2}\sum_{j=1}^{n_i}\nabla_{\theta_i} f(x_{ij};\theta,\vartheta)\,(\boldsymbol{y_{ij}} - f(x_{ij};\theta,\vartheta))\right).\end{aligned} \tag{2.29}$$

Using the definitions of l_i, h and g, Equation (2.29) can be written as

$$\begin{aligned}\log p^{\boldsymbol{Y_i}}(\boldsymbol{y_i};\beta) \approx \log l_i(\theta) + \log h(\theta) \\ \underbrace{+\frac{p}{2}\log 2\pi - \frac{1}{2}\log|H_i(\theta)| - \frac{1}{2}\nabla g(\theta)^T\left(-H_i(\theta)^{-1}\right)\nabla g(\theta))}_{-w(\theta)},\end{aligned} \tag{2.30}$$

with $H_i(\theta) = (\nabla l_i(\theta)\cdot\nabla l_i(\theta)^T)/l_i^2(\theta) + \Omega^{-1}$ denoting the approximation of $\nabla^2 g(\theta)$. Furthermore, the term w can be expressed as

$$\begin{aligned}w(\theta) &= -\frac{p}{2}\log 2\pi - \frac{1}{2}\log|H_i(\theta)^{-1}| - \frac{1}{2}\nabla g(\theta)^T H_i(\theta)^{-1}\nabla g(\theta) \\ &= -\frac{p}{2}\log 2\pi - \frac{1}{2}\log|H_i(\theta)^{-1}| \\ &\quad - \frac{1}{2}\left(\theta - (\theta - H_i(\theta)^{-1}\nabla g(\theta))\right)^T H_i(\theta)\left(\theta - (\theta - H_i(\theta)^{-1}\nabla g(\theta))\right).\end{aligned}$$

Thus, w is identical to the following p.d.f. evaluated at $\theta_i = \theta$

$$\begin{aligned}\log p^{\boldsymbol{Z_i^{FO}}}(\theta_i;\beta,\boldsymbol{y_i}) := -\frac{p}{2}\log 2\pi - \frac{1}{2}\log|H_i(\theta)^{-1}| \\ - \frac{1}{2}\left(\theta_i - (\theta - H_i(\theta)^{-1}\nabla g(\theta))\right)^T H_i(\theta)\left(\theta_i - (\theta - H_i(\theta)^{-1}\nabla g(\theta)^T)\right),\end{aligned}$$

of the respective random variable

$$\boldsymbol{Z_i^{FO}} \sim \mathcal{N}_p\left(\theta - C^{\boldsymbol{Z_i^{FO}}}\nabla g(\theta),\, C^{\boldsymbol{Z_i^{FO}}}\right),\ \text{with } C^{\boldsymbol{Z_i^{FO}}} = H_i(\theta)^{-1}. \tag{2.31}$$

Hence, Equation (2.30) can be written as

$$\log p^{\boldsymbol{Y_i}}(\boldsymbol{y_i};\beta) \approx \log l_i(\theta) + \log h(\theta) + \log p^{\boldsymbol{Z_i^{FO}}}(\theta;\beta,\boldsymbol{y_i}).$$

In conclusion, approximating the distribution of $\boldsymbol{Z_i}$ by the distribution of $\boldsymbol{Z_i^{FO}}$ and applying Bayes' theorem leads to the FO approximation for $\theta_i = \theta$.

Analogously to the FOCE method, the approximated p.d.f. of $\boldsymbol{Y_i}$ can be considered as the density of the random variable

$$\tilde{\boldsymbol{Y_i}} = f(x_i;\theta,\vartheta) + \mathcal{J}_{\theta_i} f(x_i;\theta,\vartheta)\,(\boldsymbol{\Theta_i} - \theta) + \boldsymbol{\epsilon_i}, \tag{2.32}$$

with $\boldsymbol{\Theta_i} \sim \mathcal{N}_p(\theta,\Omega)$, $\boldsymbol{\epsilon_i} \sim \mathcal{N}_{n_i}\left(0,\, I_{n_i}\sigma^2\right)$. This implies that

$$\tilde{\boldsymbol{Y_i}} \sim \mathcal{N}_{n_i}\left(f(x_i;\theta,\vartheta),\, \mathcal{J}_{\theta_i} f(x_i;\theta,\vartheta)\,\Omega\,\mathcal{J}_{\theta_i} f(x_i;\theta,\vartheta) + I_{n_i}\sigma^2\right).$$

Again, a heuristic argument to derive Equation (2.32) is to linearize $f(x_i;\theta_i,\vartheta)$ around $\theta_i = \theta$

$$f(x_{ij};\theta_i,\vartheta) \approx f(x_{ij};\theta,\vartheta) + \nabla_{\theta_i} f(x_{ij};\theta,\vartheta)\,(\boldsymbol{\Theta_i} - \theta).$$

For the FO method, no intermediate step to derive $\hat{\theta}_i$ (see Equation (2.11)) is needed. Therefore, the likelihood function has a closed-form expression.

In the next example the three approximation approaches are illustrated based on the data and model introduced in Example 2.2.5.

2.2.5. Example: Distribution of the Empirical Bayes Estimates

In the previous section, Laplacian, FOCE and FO method were derived. Furthermore, a new interpretation of these methods was introduced. Basically, each method can be interpreted as an approximation of $\boldsymbol{Z_i} = (\boldsymbol{\Theta_i}|\boldsymbol{Y_i} = \boldsymbol{y_i})$ by a multivariate normal distribution. The methods only differ in the choice of the corresponding expectation and variance-covariance, see Table 2.1.

Table 2.1.: *Approximation of the distribution of $\boldsymbol{Z_i}$: All three Laplace based methods are equivalent in assuming a normal distribution. In the second column, the point in which the Taylor series approximation of g is developed, denoted as $\hat{\theta}_i$, is present.*

Method	$\hat{\theta}_i$	$\mathbf{E}_\beta[\boldsymbol{Z_i}]$	$\mathrm{Var}_\beta[\boldsymbol{Z_i}]$
Laplacian	$\hat{\theta}_i(\boldsymbol{y_i};\beta)$	$\hat{\theta}_i(\boldsymbol{y_i};\beta)$	$\left(\frac{\nabla_{\theta_i} l_i(\hat{\theta}_i(\boldsymbol{y_i};\beta))\,\nabla_{\theta_i} l_i(\hat{\theta}_i(\boldsymbol{y_i};\beta))^T}{l_i^2(\hat{\theta}_i(\boldsymbol{y_i};\beta))} + -\frac{\nabla^2_{\theta_i} l_i(\hat{\theta}_i(\boldsymbol{y_i};\beta))}{l_i(\hat{\theta}_i(\boldsymbol{y_i};\beta))} + \Omega^{-1}\right)^{-1}$
FOCE	$\hat{\theta}_i(\boldsymbol{y_i};\beta)$	$\hat{\theta}_i(\boldsymbol{y_i};\beta)$	$\left(\frac{\nabla_{\theta_i} l_i(\hat{\theta}_i(\boldsymbol{y_i};\beta))\,\nabla_{\theta_i} l_i(\hat{\theta}_i(\boldsymbol{y_i};\beta))^T}{l_i^2(\hat{\theta}_i(\boldsymbol{y_i};\beta))} + \Omega^{-1}\right)^{-1}$
FO	θ	$\theta - \left(\frac{\nabla_{\theta_i} l_i(\theta)\,\nabla_{\theta_i} l_i(\theta)^T}{l_i^2(\theta)} + \Omega^{-1}\right)^{-1} \nabla g(\theta)$	$\left(\frac{\nabla_{\theta_i} l_i(\theta)\,\nabla_{\theta_i} l_i(\theta)^T}{l_i^2(\theta)} + \Omega^{-1}\right)^{-1}$

For illustration of the different methods the levofloxacin dataset and NLME model which will be introduced later in the Example 3.3 is considered. The NLME model consists of four random effects parameters Cl, V_1, Q and V_2 and the values of the population parameters θ, Ω and σ^2 are given by the ML estimates which will be presented later. Note, in this model no fixed effects not associated with a random effect ϑ were included.

A stochastic Markov-Chain Monte-Carlo (MCMC) approach was used to create a sampling distribution of $\boldsymbol{Z_i}$ for a given set of parameter $\beta = (\theta, \Omega, \sigma^2)$. The MCMC and the Laplace–based approaches are used to approximate the distribution of $\boldsymbol{Z_i}$. The results are compared w.r.t. expectation and variance of $\boldsymbol{Z_i}$, where the stochastic approximation was used as benchmark. Simulations and analysis throughout the whole thesis were performed in R Core Team [2012].

An introduction to Metropolis–Hastings algorithms and their properties can be found in Chib and Greenberg [1995]. To draw a sample $\theta_i^{(r)}$ from the sampling distribution of $\boldsymbol{Z_i}$ with p.d.f.

$$p^{\boldsymbol{Z_i}}(\theta_i;\beta,\boldsymbol{y_i}) = \frac{\overbrace{p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}(\theta_i;\theta,\sigma^2,\boldsymbol{y_i})p^{\boldsymbol{\Theta_i}}(\theta_i;\theta,\Omega)}^{=:\exp g(\theta_i)}}{p^{\boldsymbol{Y_i}}(\boldsymbol{y_i};\beta)} \propto \exp g(\theta_i),$$

which is known up to a constant, a candidate from a distribution where a sampling technique is established is drawn. The latter distribution is denoted as proposal distribution. Based on a certain criterion the candidate is rejected or accepted. Repeating the procedure leads to a Markov chain which converges to the aimed–at distribution.

As proposal distribution the multivariate normal distribution with p.d.f. $p(\theta_i; \theta_i^{(r-1)}, C^{\boldsymbol{Z_i^{LA}}})$ was chosen, centered around the last accepted candidate $\theta_i^{(r-1)}$ and with variance–covariance $C^{\boldsymbol{Z_i^{LA}}}$, as defined in Section 2.2.2. Because the proposal distribution depends on the last accepted candidate, the actual and the last step are not independent, therefore this algorithm is a random walk Metropolis–Hastings algorithm. Furthermore, the proposal distribution is symmetric in the sense that $p(\theta_i^{(r)}; \theta_i^{(r-1)}, C^{\boldsymbol{Z_i^{LA}}}) = p(\theta_i^{(r-1)}; \theta_i^{(r)}, C^{\boldsymbol{Z_i^{LA}}})$. The convergence of this algorithm and the corresponding rate was investigated by Mengersen and Tweddie [1996]. The Metropolis-Hastings algorithm described in Algorithm 1 was used to create a sample $\theta_i^{(1)}, \ldots, \theta_i^{(R)}$ of size R. Gelman et al. [1994] stated that for the practical implementation of this type of random walk Metropolis–Hastings algorithm, with a symmetric proposal density and corresponding constant variance-covariance term, an acceptance rate between 15 % to 40 % is preferred. To optimize the computational efficiency and to improve the acceptance rate of the algorithm, the choice of the proposal distribution can be further improved, e.g. by taking into consideration the random walk for the variance of the proposal distribution as well. The NLME software packages Monolix, the R–package *saemix* as well as NONMEM (for the SAEM method) apply such a strategy. However, as we did not aim for time–optimized implementation, we used the symmetric proposal distribution as described above.

The algorithm was applied for each of the 24 subjects in the levofloxacin dataset and iterated $R = 20\,000$ times. The recommendation regarding the acceptance rate was adequately matched with a rate between 36 % to 38 % for all subjects. For one subject with good agreement between stochastic and deterministic approximation, the results are illustrated in Figure 2.1. For all parameters the normal densities and expectations values resulting from the Laplacian and FOCE method are in good agreement with the empirical density (with empirical density is meant, that the empirical frequency distribution is scaled in a way that the histogram has an area of one). The bivariate dependence structure is very well matched by the Laplacian and FOCE approximations. Overall, the Laplacian and FOCE methods are visually not distinguishable and provide good approximations. In contrast to that, the FO method shows poor agreement with the empirical density.

A subject showing a worse performance in approximating $\boldsymbol{Z_i}$ by the Laplace–based methods is shown in Figure 2.2. The shape of the empirical distribution for Q and V_2 is left–skewed. As a consequence, the expectations values are over–estimated for Laplacian and FOCE method. In addition, a significant difference between Laplacian and FOCE approximation can be observed, with a better approximation accuracy of the Laplacian method. Due to the left–skewed marginal distributions obviously also the bivariate dependence structures are not well matched by the Laplacian and FOCE approximations. Overall, the Laplacian and FOCE methods differ only slightly from each other and provide good approximations. Again, the FO method shows poorer agreement with the empirical density compared to Laplacian and FOCE method.

Based on the law of large numbers (LLN) mean and sample variance are used to estimate $\mathbf{E}_\beta[\boldsymbol{Z_i}]$ and variance–covariance matrix $\mathrm{Var}_\beta[\boldsymbol{Z_i}]$. The results are compared with the

Algorithm 1 Random Walk Metropolis–Hastings Algorithm to Create the Sampling Distribution of $\boldsymbol{Z_i}$.

Define number of iterations size R, $\beta := (\theta, \Omega, \sigma^2)$ and $g(\theta_i) := \log p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}(\boldsymbol{y_i}; \theta_i, \sigma^2) + \log p^{\boldsymbol{\Theta_i}}(\theta_i; \theta, \Omega)$.
Estimate $\hat{\theta}_i(\boldsymbol{y_i}; \beta) := \arg\min\ -g(\theta_i)$ and calculate $C_{\boldsymbol{Z_i^{LA}}} := \left(-\nabla^2 g(\hat{\theta}_i(\boldsymbol{y_i}; \beta))\right)^{-1}$.
Set $\theta_i^{(0)} = \hat{\theta}_i(\boldsymbol{y_i}; \beta)$.
Set $r = 0$.
Set $n_{\text{accept}} = 0$.
repeat
  Set $r = r + 1$.
  Sample a candidate $\widehat{\theta_i}$ from $\mathcal{N}_p\left(\theta_i^{(r-1)}, C^{\boldsymbol{Z_i^{LA}}}\right)$.
  Estimate acceptance ratio $\log a = g(\widehat{\theta_i}) - g(\theta_i^{(r-1)})$.
  Sample a realization u from the uniform distribution $U(0, 1)$.
  if $a > u$ **then** ▷ Acceptance
    Set $\theta_i^{(r)} = \widehat{\theta_i}$.
    Set $n_{\text{accept}} = n_{\text{accept}} + 1$.
  else ▷ Rejection
    Set $\theta_i^{(r)} = \theta_i^{(r-1)}$.
  end if
until $r = R$
Define acceptance rate n_{accept}/R.

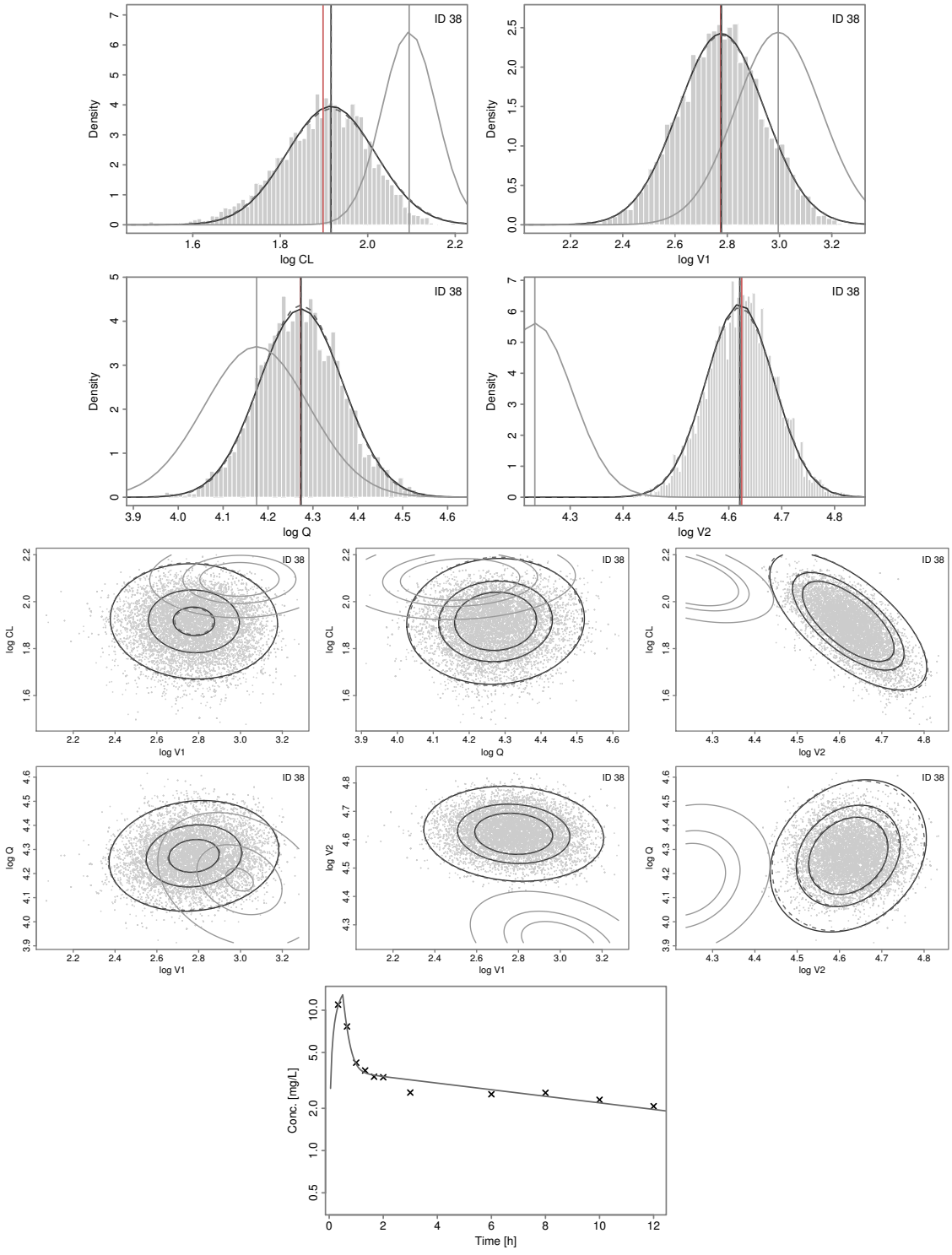

Figure 2.1.: *Example: Adequate approximation quality. Results of stochastic and Laplacian (black solid line), FOCE (grey dashed line) and FO (grey solid line) approximations of the distribution of* $\boldsymbol{Z_i}$ *exemplified for one subject of the levofloxacin study. Each histogram shows empirical density and the associated mean (red vertical line), approximated marginal densities and respective expectation values. The bivariate scatter plots shows the empirical bivariate distribution with the respective approximated densities (contour lines). Lower plot shows the observations (black crosses) for the respective subject and the individual predictions (grey solid line) based on the final model results.*

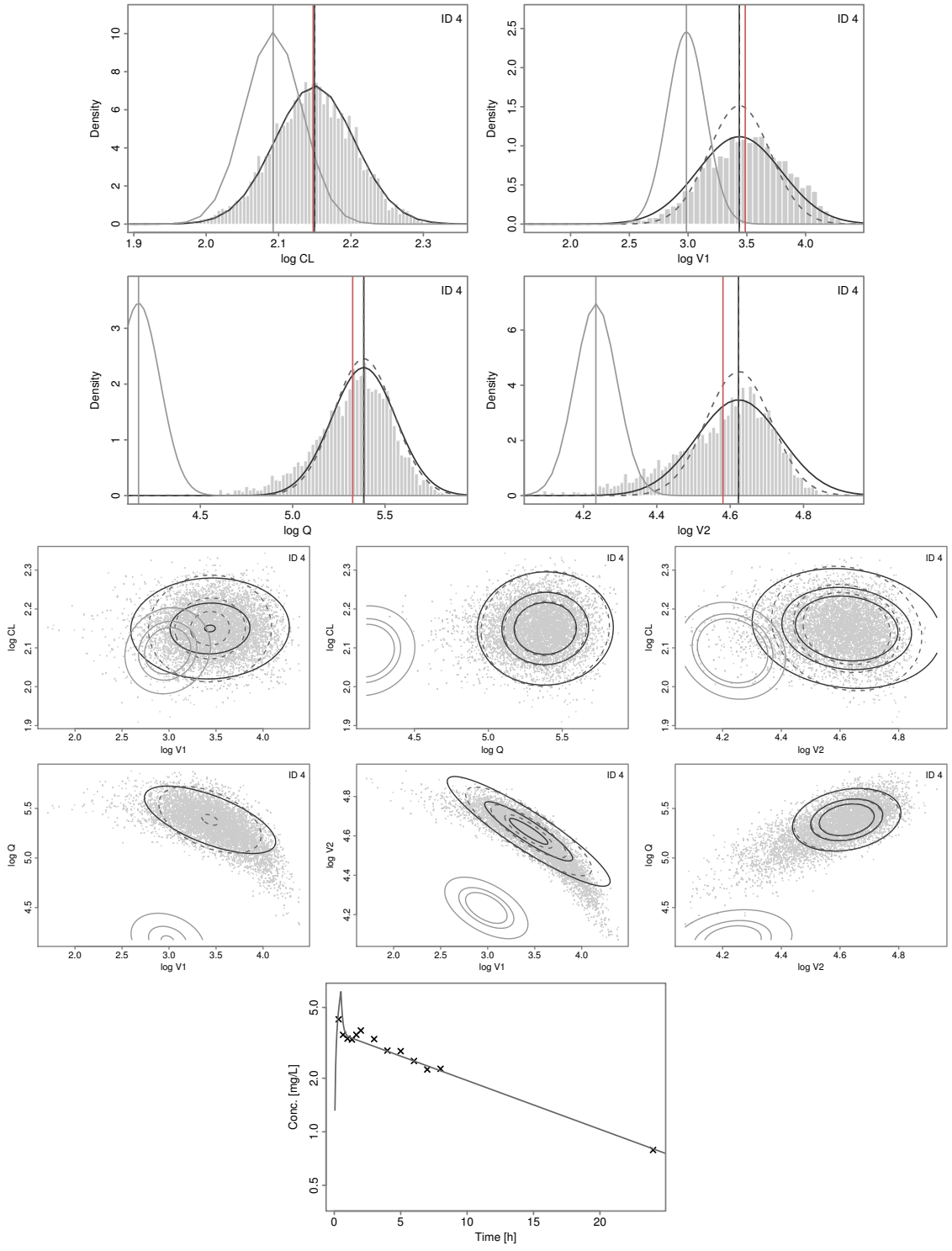

Figure 2.2.: *Example: Inaccuracy of deterministic approximation. Results of stochastic and Laplacian (black solid line), FOCE (grey dashed line) and FO (grey solid line) approximations of the distribution of* $\boldsymbol{Z_i}$ *exemplified for one subject of the levofloxacin study. Each histogram shows empirical density and the associated mean(red vertical line), approximated marginal densities and respective expectation values. The bivariate scatter plots shows the empirical bivariate distribution with the respective approximated bivariate densities (contour lines). Lower plot shows the observations (black crosses) for the respective subject and the individual predictions (grey solid line) based on the final model results.*

Laplacian, FOCE and FO approximation in terms of relative bias:

$$\text{relBias}(\%) := \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{\beta}_i - \beta_0}{\beta_0}\right), \tag{2.33}$$

where $\hat{\beta}_i$ represents the approximated parameter value of a component of $\mathbf{E}_\beta[\boldsymbol{Z_i}]$ or $\text{Var}_\beta[\boldsymbol{Z_i}]$, and β_0 the respective reference value determined using stochastic approximation (SA).

The results are listed in in Table 2.2. For all methods the relative bias is smallest for the expectation value and the relative bias for the diagonals is smaller compared to the off–diagonals of $\text{Var}_\beta[\boldsymbol{Z_i}]$. For the underlying situation, Laplacian and FOCE method provide very adequate approximation of $\mathbf{E}_\beta[\boldsymbol{Z_i}]$. In general, all methods tend to underestimate the entries of $\text{Var}_\beta[\boldsymbol{Z_i}]$. Overall, for all parameter the range of bias in the underlying subjects is similar for Laplacian and FOCE method, while for most parameters the FO method performs markedly worse.

Table 2.2.: *Mean and range (in brackets) of relative bias of the Laplace–based estimators for* $\mu^{\boldsymbol{Z_i}} = \mathbf{E}_\beta[\boldsymbol{Z_i}]$ *and* $C^{\boldsymbol{Z_i}} = \text{Var}_\beta[\boldsymbol{Z_i}]$.

	Laplacian	FOCE	FO
$\mu^{\boldsymbol{Z_i}}_{(1)}$ ($\log$ Cl)	0.3 % (0.02 to 1.0)	0.3 % (0.02 to 1.00)	2.2 % (−14.0 to 50.0)
$\mu^{\boldsymbol{Z_i}}_{(2)}$ ($\log \text{V}_1$)	−0.07 % (−5.6 to 1.5)	−0.07 % (−5.6 to 1.5)	1.7 % (−18.9 to 28.6)
$\mu^{\boldsymbol{Z_i}}_{(3)}$ ($\log$ Q)	−0.05 % (−0.8 to 1.1)	−0.05 % (−0.8 to 1.1)	2.2 % (−21.8 to 29.6)
$\mu^{\boldsymbol{Z_i}}_{(4)}$ ($\log \text{V}_2$)	−0.09 % (−0.3 to 0.9)	−0.09 % (−0.3 to 0.9)	0.9 % (−10.1 to 23.8)
$C^{\boldsymbol{Z_i}}_{(1,1)}$	−3.1 % (−12.7 to 4.4)	−2.3 % (−14.9 to 9.3)	−3.5 % (−77.0 to 112.9)
$C^{\boldsymbol{Z_i}}_{(2,2)}$	−5.3 % (−23.5 to 5.8)	−15.6 % (−46.1 to 7.6)	−1.7 % (−79.5 to 114.9)
$C^{\boldsymbol{Z_i}}_{(3,3)}$	−2.9 % (−31.3 to 15.8)	−13.6 % (−42.0 to 9.1)	−12.6 % (−80.1 to 91.0)
$C^{\boldsymbol{Z_i}}_{(4,4)}$	−7.0 % (−57.9 to 3.4)	−9.3 % (−59.4 to 6.9)	−6.9 % (−83.6 to 102.4)
$C^{\boldsymbol{Z_i}}_{(1,2)}$	−15.6 % (−392.6 to 106.3)	−30.0 % (−478.0 to 162.8)	−50.6 % (−305.3 to 376.6)
$C^{\boldsymbol{Z_i}}_{(1,3)}$	−4.9 % (−82.5 to 60.4)	−4.7 % (−86.4 to 75.6)	9.1 % (−85.0 to 138.3)
$C^{\boldsymbol{Z_i}}_{(1,4)}$	−17.4 % (−448.0 to 55.8)	−13.8 % (−487.6 to 50.1)	61.6 % (−378.3 to 1371.9)
$C^{\boldsymbol{Z_i}}_{(2,3)}$	−10.9 % (−306.8 to 110.0)	−65.8 % (−1102.1 to 87.9)	21.4 % (−431.5 to 1338.9)
$C^{\boldsymbol{Z_i}}_{(2,4)}$	−16.1 % (−154.0 to 4.3)	−31.2 % (−209.0 to 7.5)	−361.3 % (−88.0 to 4507.4)
$C^{\boldsymbol{Z_i}}_{(3,4)}$	−13.9 % (−188.0 to 123.3)	−6.4 % (−169.4 to 480.8)	−242.3 % (−2415.1 to 7859.8)

The results are in agreement with the findings by Bonate [2011, p. 257], that in general the approximation accuracy is increasing in the following order

$$\text{Laplacian} \geq \text{FOCE} \gg \text{FO}.$$

In the next section the EM algorithm for NLME models is introduced. Within this algorithm it is necessary to determine $\mathbf{E}_\beta[\boldsymbol{Z_i}]$ and $\text{Var}_\beta[\boldsymbol{Z_i}]$ for all $1 \leq i \leq N$.

3. Numerical determination of the Maximum Likelihood Estimates

ML estimates are parameter values that maximize the likelihood function. In general, for NLME models there is no closed-form expression of the likelihood function available and approximations have to be made. Thus, the maximum likelihood estimates are often defined as the maximizer of some approximated likelihood function.

The Laplacian, FOCE and FO approximation for the p.d.f. $p^{\boldsymbol{Y_i}}$ given the parameter β were introduced in the previous section. As already mentioned, in contrast to the FO method, the Laplacian and the FOCE methods depend on the mode $\hat{\theta}_i(\boldsymbol{y_i};\beta)$ of $p^{\boldsymbol{Z_i}}$ (where $\boldsymbol{Z_i} = (\boldsymbol{\Theta_i}|\boldsymbol{Y_i} = \boldsymbol{y_i})$), which itself depends on β. Therefore, the Laplacian and FOCE approximation also do not provide a closed-form which could be used to define ML estimates, see Beal and Sheiner [1998]. For the additive error model—the one which is considered here—the WSV $\mathrm{Var}_\beta\left[\boldsymbol{Y_{ij}}|\boldsymbol{\Theta_i} = \theta_i\right] = \sigma^2$ does not depend on the realization of $\boldsymbol{\Theta_i}$. Hence, Vonesh and Chinchilli [1997, p. 345] suggest to apply an iterative approach to derive the ML estimates, switching back and forth between estimating $\hat{\theta}_i(\boldsymbol{y_i};\beta)$ and optimizing $p^{\boldsymbol{Y_i}}$ w.r.t. to β until some predefined convergence criterion is met. The algorithm is described in detail on page 33 (see Algorithm 2). One disadvantage of this relatively simple approach is the absence of any mathematical proof that the iteration improves the likelihood function value and even more, convergences to the ML estimates. Under some regularity conditions, Vonesh [1996] showed that with increasing number of subjects N and increasing numbers of observations n_i for all subjects, the maximizer of the approximated likelihood function using the Laplacian method β^*_{LA} converges to the maximizer of the likelihood function (the ML estimate β^*):

$$(\beta^*_{\mathrm{LA}} - \beta) = O\left(\max\left\{\frac{1}{\sqrt{N}}, \frac{1}{\min\{n_1,\ldots,n_N\}}\right\}\right)$$

where O denotes the Landau symbol.

An alternative, theoretically well–founded iterative approach is presented in the next section.

3.1. The deterministic EM–Algorithm

In the following chapter an EM algorithm in the context of NLME modeling is presented. It is based on a deterministic (Laplace) approximation of the integral over the random effects involved in the likelihood function. To distinguish this approach from Monte Carlo and stochastic approximations of the integral (Walker [1996]; Kuhn and Lavielle [2005]), we denote the variant presented here as *deterministic EM*.

Algorithm 2 Derivation of the ML estimates based on the Laplacian approximation

Define $g(\theta_i) := \log p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}(\boldsymbol{y_i}; \theta_i, \vartheta, \sigma^2) + \log p^{\boldsymbol{\Theta_i}}(\theta_i; \theta, \Omega)$.
Define $\text{OFV}\,(\beta; \theta_1, \ldots, \theta_N, \boldsymbol{y}) := -\sum_{i=1}^{N} \left(g(\theta_i) + \log\left|-\nabla^2_{\theta_i} g(\theta_i)\right|\right)$.
(For $\theta_i = \hat{\theta}_i(\boldsymbol{y_i}; \beta)$ for all $1 \leq i \leq N$, OFV corresponds to the Laplacian approximation of $\log p^{\boldsymbol{Y}}$ up to a constant.)
Set $r = 0$.
Choose initial values: $\beta^{(0)} = (\theta^{(0)}, \Omega^{(0)}, \sigma^{2\,(0)})$.
repeat
Set $r = r + 1$.
(a) For all $i = 1, \ldots, N$: Estimate $\hat{\theta}_i(\boldsymbol{y_i}; \beta^{(r-1)}) := \underset{\theta_i}{\arg\min}\ -g(\theta_i)$.
(b) Estimate $\beta^{(r)} := \underset{\beta}{\arg\min}\ \text{OFV}\left(\beta; \hat{\theta}_1(\boldsymbol{y_1}; \beta^{(r-1)}), \ldots, \hat{\theta}_N(\boldsymbol{y_N}; \beta^{(r-1)}), \boldsymbol{y}\right)$.
until Convergence
Results: $\hat{\beta}_{\text{itLA}} = \beta^{(r)}$.

Dempster et al. [1977] demonstrated that in case the likelihood function L does not have a close form, but can be expressed as

$$L(\beta; \boldsymbol{y}) = p^{\boldsymbol{Y}}(\boldsymbol{y}; \beta) = \int p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y}, \boldsymbol{\theta}; \beta)\, d\boldsymbol{\theta},$$

where $\boldsymbol{\Theta} = (\boldsymbol{\Theta_1}, \ldots, \boldsymbol{\Theta_N})^T$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)^T$ and with joint p.d.f. $p^{\boldsymbol{Y},\boldsymbol{\Theta}}$ of $(\boldsymbol{Y}, \boldsymbol{\Theta})$, the iteration of the

E–Step: Given the parameter $\beta \in \Upsilon^\beta$, derive the expectation

$$\begin{aligned} Q(\beta', \beta) &:= \mathbf{E}_\beta\left[\log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y}, \boldsymbol{Z}; \beta')\right] \\ &= \int \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y}, \boldsymbol{\theta}; \beta')\, p^{\boldsymbol{\Theta}|\boldsymbol{Y}=\boldsymbol{y}}(\boldsymbol{\theta}; \beta)\, d\boldsymbol{\theta}, \end{aligned} \tag{3.1}$$

for $\beta' \in \Upsilon^\beta$, where Υ^β denotes the parameter space of β and $\boldsymbol{Z} := (\boldsymbol{\Theta}|\boldsymbol{Y} = \boldsymbol{y})$,

M–Step: Determine

$$\tilde{\beta} = \underset{\beta' \in \Upsilon^\beta}{\arg\max}\ Q(\beta', \beta). \tag{3.2}$$

results in a mapping $\beta \to \tilde{\beta}$ with the following crucial property

$$L(\tilde{\beta}; \boldsymbol{y}) \geq L(\beta; \boldsymbol{y}).$$

Dempster et al. [1977] denoted the tuple $(\boldsymbol{Y}, \boldsymbol{\Theta})$ as complete data, where $\boldsymbol{Y}$ represents the incomplete data and is observed, i.e. $\boldsymbol{Y} = \boldsymbol{y}$, while $\boldsymbol{\Theta}$ is not observed directly and estimated based on $\boldsymbol{y}$.

Wu [1983] showed that under some regulatory conditions the sequence of EM iterations $\beta^{(0)}, \ldots, \beta^{(k)}$ converges to a local maximum of the likelihood function L. I.e., if $L(\cdot; \boldsymbol{y})$ is

uni–modal, the sequence converges to the global maximum of L, namely the ML estimate. Because the uni–modality is difficult to show, different initial values $\beta^{(0)}$ should be tested to increase the chance of finding the global maximum of L and to check the robustness of the algorithm, as suggested in most optimization routines, e.g. see Gilli and Schumann [2010].

Among others, Mentre and Gomeni [1995]; Walker [1996]; Kuhn and Lavielle [2005] proposed an EM algorithms in the context of NLME modeling.

Using that $p^{\boldsymbol{Y_i},\boldsymbol{\Theta_i}}(\boldsymbol{y_i},\theta_i;\beta) = p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}(\boldsymbol{y};\theta_i,\vartheta,\sigma^2)\, p^{\boldsymbol{\Theta_i}}(\theta_i;\theta,\Omega)$, Equation (3.1) can be written as

$$Q(\beta',\beta) = \sum_{i=1}^{N} \int \left(\log p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}(\boldsymbol{y_i};\theta_i,\vartheta',\sigma^{2\prime}) + \log p^{\boldsymbol{\Theta_i}}(\theta_i;\theta',\Omega')\right)\; p^{\boldsymbol{Z_i}}(\theta_i;\theta,\Omega)\; d\theta_i, \quad (3.3)$$

with $\boldsymbol{Z_i} = (\boldsymbol{\Theta_i}|\boldsymbol{Y_i} = \boldsymbol{y_i})$.

In the context of NLME modeling, Walker [1996] showed that the E–step can be expressed as

E–Step: Given the data $\boldsymbol{y_i}$ and a parameter β, derive expectations $\mathbf{E}_\beta[\boldsymbol{Z_i}]$ and variances $\mathrm{Var}_\beta[\boldsymbol{Z_i}]$, for all subjects $i = 1,\ldots,N$,

If the NLME model does not contain fixed effect parameters not associated with a random effect ϑ, the M–Step can be solved analytically. Accordingly, different EM–like algorithms, such as proposed by Mentre and Gomeni [1995]; Walker [1996]; Kuhn and Lavielle [2005] mainly differ in the derivation of the E–step. To this end, Mentre and Gomeni [1995] use analytical approximations, while Walker [1996]; Kuhn and Lavielle [2005] use stochastic approximations. Walker [1996] uses the importance sampling method, while Kuhn and Lavielle [2005] used an MCMC approach. Both stochastic approaches have the advantage to be more precise if large numbers of iterations in the stochastic approximation step are used.

The M–Step

Under suitable regularity conditions, the following equation has to be fulfilled for $\tilde{\beta}$ as defined in Equation (3.2)

$$\nabla_{\beta'} Q(\beta',\beta)\big|_{\beta'=\tilde{\beta}} = 0,$$

with $\tilde{\beta} \in \Upsilon^\beta$. Again, under suitable regularity conditions, according to Leibniz integral rule (e.g. see Flanders [1973]), the integration and differentiation can be changed which leads to

$$\sum_{i=1}^{N} \mathbf{E}_\beta\left[\nabla_{\beta'} \log p^{\boldsymbol{Y_i},\boldsymbol{\Theta_i}}(\boldsymbol{y_i},\boldsymbol{Z_i};\beta')\right]\Big|_{\beta'=\tilde{\beta}} = 0. \quad (3.4)$$

Using Bayes' theorem (as in Equation (3.3)) the gradients of $\log p^{\boldsymbol{Y_i},\boldsymbol{\Theta_i}}$ with respect to θ', Ω', ϑ' and $\sigma^{2\prime}$ are given by:

1. Fixed effects associated with a random effect θ:

$$\begin{aligned}\nabla_{\theta'} \log p^{\boldsymbol{Y_i},\boldsymbol{\Theta_i}}(\boldsymbol{y_i},\boldsymbol{Z_i};\beta') &= \nabla_{\theta'} \log p^{\boldsymbol{\Theta_i}}(\boldsymbol{Z_i};\theta',\Omega') \\ &= \Omega'^{-1}(\theta' - \boldsymbol{Z_i}),\end{aligned}$$

2. Between–subject variability Ω:

$$\begin{aligned}\nabla_{\Omega'} \log p^{\boldsymbol{Y_i},\boldsymbol{\Theta_i}}(\boldsymbol{y_i}, \boldsymbol{Z_i}; \beta') &= \nabla_{\Omega'} \log p^{\boldsymbol{\Theta_i}}(\boldsymbol{Z_i}; \theta', \Omega') \\ &= \frac{1}{2}\Omega'^{-1}\left(-\Omega' + (\theta' - \boldsymbol{Z_i})(\theta' - \boldsymbol{Z_i})^T\right)\Omega'^{-1},\end{aligned}$$

3. Fixed effects not associated with a random effect ϑ:

$$\begin{aligned}&\nabla_{\vartheta'} \log p^{\boldsymbol{Y_i},\boldsymbol{\Theta_i}}(\boldsymbol{y_i}, \boldsymbol{Z_i}; \beta') = \nabla_{\vartheta'} \log p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta}(\boldsymbol{y_i}; \boldsymbol{Z_i}, \vartheta', \sigma^{2'}) \\ &= -\nabla_{\vartheta'} \frac{1}{2\sigma^{2'}} \sum_{j=1}^{n_i} \left(\boldsymbol{y_{ij}} - f(x_{ij}; \boldsymbol{Z_i}, \vartheta')\right)^2 \\ &\approx -\nabla_{\vartheta'} \frac{1}{2\sigma^{2'}} \sum_{j=1}^{n_i} \left(\boldsymbol{y_{ij}} - f(x_{ij}; \boldsymbol{Z_i}, \vartheta^*) - \nabla_\vartheta f(x_{ij}; \boldsymbol{Z_i}, \vartheta^*)^T(\vartheta' - \vartheta^*)\right)^2 \\ &= \frac{1}{\sigma^{2'}} \sum_{j=1}^{n_i} \nabla_\vartheta f(x_{ij}; \boldsymbol{Z_i}, \vartheta^*) \left(\boldsymbol{y_{ij}} - f(x_{ij}; \boldsymbol{Z_i}, \vartheta^*) - \nabla_\vartheta f(x_{ij}; \boldsymbol{Z_i}, \vartheta^*)^T(\vartheta' - \vartheta^*)\right),\end{aligned}$$

   where f was approximated by its first order linearization w.r.t. ϑ' at ϑ^*. This linearization is necessary in order to have a closed analytical expression for the M step,

4. Within–subject variability σ^2:

$$\begin{aligned}\partial_{\sigma^{2'}} \log p^{\boldsymbol{Y_i},\boldsymbol{\Theta_i}}(\boldsymbol{y_i}, \boldsymbol{Z_i}; \beta') &= \partial_{\sigma^{2'}} \log p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}}(\boldsymbol{y_i}; \boldsymbol{Z_i}, \vartheta', \sigma^{2'}) \\ &= -\frac{n_i}{2\sigma^{2'}} + \frac{1}{2\sigma^{4'}} \sum_{j=1}^{n_i} \left(\boldsymbol{y_{ij}} - f(x_{ij}; \boldsymbol{Z_i}, \vartheta')\right)^2.\end{aligned}$$

Hence, given β and the distribution of $\boldsymbol{Z_i}$ (note that distribution of $\boldsymbol{Z_i}$ depends on $\boldsymbol{y_i}$ and β), a solution of Equation (3.4) for the single components of $\tilde{\beta}$ is given by:

1. Fixed effects associated with a random effect:

$$\tilde{\theta} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{E}_\beta \left[\boldsymbol{Z_i}\right],$$

2. Between–subject variability:

$$\tilde{\Omega} = \frac{1}{N} \sum_{i=1}^{N} \left(\mathrm{Var}_\beta \left[\boldsymbol{Z_i}\right] + \left(\tilde{\theta} - \mathbf{E}_\beta \left[\boldsymbol{Z_i}\right]\right)\left(\tilde{\theta} - \mathbf{E}_\beta \left[\boldsymbol{Z_i}\right]\right)^T\right),$$

   where $\mathrm{Var}_\beta \left[\boldsymbol{Z_i}\right] = \mathbf{E}_\beta \left[\boldsymbol{Z_i}\ \boldsymbol{Z_i}^T\right] - \mathbf{E}_\beta \left[\boldsymbol{Z_i}\right] \mathbf{E}_\beta \left[\boldsymbol{Z_i}\right]^T$ was used.

3. Fixed effects not associated with a random effect ($\vartheta^* = \vartheta$):

$$\begin{aligned}\tilde{\vartheta} = &\left(\sum_{i=1}^{N} \sum_{j=1}^{n_i} \left(\boldsymbol{y_{ij}} - \mathbf{E}_\beta \left[f(x_{ij}; \boldsymbol{Z_i}, \vartheta)\right] + \nabla_\vartheta \mathbf{E}_\beta \left[f(x_{ij}; \boldsymbol{Z_i}, \vartheta)^T\right] \vartheta\right)\right) \\ &\cdot \left(\sum_{i=1}^{N} \sum_{j=1}^{n_i} \nabla_\vartheta \mathbf{E}_\beta \left[f(x_{ij}; \boldsymbol{Z_i}, \vartheta)\right] \nabla_\vartheta \mathbf{E}_\beta \left[f(x_{ij}; \boldsymbol{Z_i}, \vartheta)^T\right]\right)^{-1},\end{aligned}$$

4. Within–subject variability:

$$\tilde{\sigma}^2 = \frac{1}{N_n} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \mathbf{E}_\beta \left[\left(\boldsymbol{y_{ij}} - f(x_{ij}; \boldsymbol{Z_i}, \vartheta') \right)^2 \right].$$

As mentioned in Section 2.1, θ and ϑ are both fixed effects and constant within the population with the difference that θ is associated with a random effect, while ϑ is not. Furthermore, a NLME model can also be defined based on the alternative parameterization of fixed effects as described in Equation (2.2). However, from the above description it becomes obvious that in the context of the EM algorithm the determinations of θ and ϑ differ, which finally could lead to different results using the alternative parametrization.

In the next paragraph the E–step is presented to derive approximations for the unknown expectations and variances in the terms above.

The E–Step

Because of the nonlinearity of f, in general, the expectation $\mathbf{E}_\beta [\boldsymbol{Z_i}]$ cannot be derived analytically and stochastic (e.g. see Walker [1996]; Kuhn and Lavielle [2005]) or deterministic approximations have to be used. In this work, we focus on deterministic approximations. In context of EM modeling, Mentre and Gomeni [1995] introduced such an algorithm. The E–step described in the following coincides to the one described by Mentre and Gomeni [1995], however, it will be derived using the new argumentation of the FOCE method derived in Section 2.2.3. It should be noted, that the presented M and E–step are independent of each other, as such, the FOCE based E–step could be replaced by a Laplacian, FO or stochastic based E–step. The FOCE method was chosen because a reduced computational effort is expect compared to the Laplacian approximation and a better approximation accuracy is expected compared to the FO method. Furthermore, the use of the FOCE method instead of the Laplacian corresponds to using the Fisher instead of the observed Fisher information to approximate variances of estimators and is therefore more robust to changes in the underlying data, e.g. against outliers. However, as Walker [1996] stated: "The accuracy of such approximations is a problem which is hard to evaluate".

In Section 2.2.3 it was shown that from the FOCE approximation of the likelihood function it follows that $\mathbf{E}_\beta [\boldsymbol{Z_i}] \approx \hat{\theta}_i(\boldsymbol{y_i}; \beta)$ (with $\hat{\theta}_i(\boldsymbol{y_i}; \beta)$ defined in Equation (2.20)) and $\mathrm{Var}_\beta [\boldsymbol{Z_i}] \approx C^{\boldsymbol{Z_i^{FOCE}}}$ (with $C^{\boldsymbol{Z_i^{FOCE}}}$ defined in Equation (2.24)). For updating the WSV in the M–Step, the following expectation has to be derived

$$\mathbf{E}_\beta \left[\left(\boldsymbol{y_{ij}} - f(x_{ij}; \boldsymbol{Z_i}, \vartheta') \right)^2 \right] = \boldsymbol{y_{ij}^2} - 2\boldsymbol{y_{ij}} \mathbf{E}_\beta \left[f(x_{ij}; \boldsymbol{Z_i}, \vartheta') \right] + \mathbf{E}_\beta \left[f(x_{ij}; \boldsymbol{Z_i}, \vartheta')^2 \right]. \quad (3.5)$$

In general, the distribution of $f(x_{ij}; \boldsymbol{Z_i}, \vartheta')$ is not known due to the nonlinearity of f. However, for f being continuously differentiable and $\nabla_{\theta_i} f(x_{ij}; \theta_i; \vartheta') \neq 0$, $\forall \theta_i \in \mathbb{R}^p$, the distribution can be approximated using the Delta method (e.g. see Lehmann and Romano [2005, Theorem 11.2.14 and 12.4.1])

$$f(x_{ij}; \boldsymbol{Z_i}, \vartheta') \overset{approx}{\sim} \mathcal{N} \Big(f(x_{ij}; \hat{\theta}_i(\boldsymbol{y_i}; \beta), \vartheta'), \\ \nabla_{\theta_i} f(x_{ij}; \hat{\theta}_i(\boldsymbol{y_i}; \beta), \vartheta')^T \, C^{\boldsymbol{Z_i^{FOCE}}} \, \nabla_{\theta_i} f(x_{ij}; \hat{\theta}_i(\boldsymbol{y_i}; \beta), \vartheta') \Big).$$

Hence, Equation (3.5) can be approximated by

$$\mathbf{E}_\beta\left[\left(\boldsymbol{y_{ij}} - f(x_{ij};\boldsymbol{Z_i},\vartheta')\right)^2\right] \approx (\boldsymbol{y_{ij}} - f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta'))^2 + \nabla_{\theta_i} f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta')^T C^{\boldsymbol{Z_i^{FOCE}}} \nabla_{\theta_i} f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta').$$

The same considerations applies for updating ϑ.

Finally, a deterministic version of the EM for NLME models is given by

E–Step: Determine EBEs given $\beta = (\theta, \Omega, \vartheta, \sigma^2)$ for all subjects $i = 1, \ldots, N$:

$$\hat{\theta}_i(\boldsymbol{y_i};\beta) = \underset{\theta_i \in \mathbb{R}^p}{\arg\max} \log p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}(\boldsymbol{y_i};\theta_i,\vartheta,\sigma^2) + \log p^{\boldsymbol{\Theta_i}}(\theta_i;\theta,\Omega),$$

and the inverse of the variance–covariance of $\boldsymbol{Z_i}$ approximated by $C^{\boldsymbol{Z_i^{FOCE}}}$, see Equation (2.24).

M–Step: 1. The updated fixed effect associated with a random effect is defined as the mean of the EBEs

$$\tilde{\theta} = \frac{1}{N}\sum_{i=1}^{N} \hat{\theta}_i(\boldsymbol{y_i};\beta).$$

2. The updated BSV variance–covariance term is defined as the sum of the variance of the EBEs and the mean of the variances of the EBEs estimators

$$\tilde{\Omega} = \frac{1}{N}\sum_{i=1}^{N}\left\{(\tilde{\theta} - \hat{\theta}_i(\boldsymbol{y_i};\beta))(\tilde{\theta} - \hat{\theta}_i(\boldsymbol{y_i};\beta))^T + C^{\boldsymbol{Z_i^{FOCE}}}\right\}.$$

3. The updated fixed effect not associated with a random effect is given by

$$\tilde{\vartheta} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left\{\boldsymbol{y_{ij}} - f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta) + \nabla_\vartheta f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta)^T\vartheta)\right\}}{\sum_{i=1}^{N}\sum_{j=1}^{n_i}\nabla_\vartheta f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta)\,\nabla_\vartheta f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta)^T}.$$

4. The updated WSV variance term is given by

$$\tilde{\sigma^2} = \frac{1}{N_n}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left\{\left(\boldsymbol{y_{ij}} - f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\tilde{\vartheta})\right)^2 + \left(\nabla_{\theta_i} f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\tilde{\vartheta})^T\right) C^{\boldsymbol{Z_i^{FOCE}}} \left(\nabla_{\theta_i} f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\tilde{\vartheta})\right)\right\},$$

which corresponds to a composition of the average residuals and variance in model predictions.

The iterative algorithm on page 33 (Algorithm 2) for the determination of the ML estimates based on the Laplacian and FOCE approximation consists of two numerical optimization steps, while in the EM algorithm presented above only one numerical optimization in the E–step is needed. Furthermore, the EM approach theoretically guarantees that the overall likelihood will—at least—not decrease after each iteration, e.g. see Wu [1983]. The principle of the algorithm is illustrated in Figure 3.1. Finally, for the practical application of the presented EM algorithm a stopping criterion of the algorithm as depicted in Figure 3.1 has to be defined.

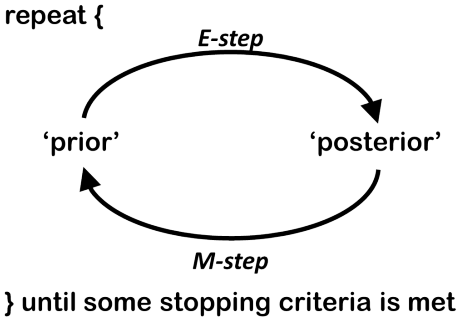

Figure 3.1.: *Schematic work process of the EM algorithm. In the r-th step of the algorithm, the E–step will derive 'posterior' distributions of $(\boldsymbol{\Theta_i}|\boldsymbol{Y_i} = \boldsymbol{y_i})$ for all individuals given 'prior' information $\beta^{(r-1)}$. Based on these, the M–step updates the 'prior' $\beta^{(r-1)} \rightarrow \beta^{(r)}$, which is the input for the $(r+1)$-th step.*

For example, one of the following stopping criteria could be used

1. The final estimates are defined as those values, where changes in the parameter values become sufficiently small between two EM iterations:

$$\text{For some predefined } \delta > 0 : \left|\beta^{(r)} - \beta^{(r-1)}\right| \leq \delta,$$

   where for $x \in \mathbb{R}^p$, the notation $|x| \leq \delta$ means, that for every component of x it is $|x_k| \leq \delta$, for $k = 1, \ldots, p$.

2. The final estimates are defined as those values, when changes in a predefined value, e.g. the approximated likelihood function value $\tilde{L}$, become sufficiently small between two EM iterations:

$$\text{For some predefined } \delta > 0 : \left|\tilde{L}(\beta^{r}) - \tilde{L}(\beta^{r-1})\right| \leq \delta.$$

3. The performance of the algorithm is monitored and on a subjective basis it is decide when an adequate convergence status is reached, for example this could be based on only one specific parameter with largest interest.

Independent of the used stopping criterion, the results should be reviewed using diagnostics tools like the ones which were introduced in the Examples 3.3.6, 3.3.5, such as goodness–of–fits (GOFs) plots and visual predictive checks (VPCs).

3.2. Precision of EM estimates

The precision of an estimate, quantified by the estimated standard error (SE), is usually defined using the Cramér-Rao bound as an approximation to the variance of the corresponding estimator, for more details see for example Newey and McFadden [1994] or Lehmann and Romano [2005, sec. 12.4.1]. In general, for a NLME model the derivation of the Fisher and also observed Fisher matrix (see [Efron and Hinkley, 1978]) is rather complex because no closed-form expression of the likelihood function exists.

So far, it has not be shown how to derive the expected Fisher information for NLME estimates. Walker [1996] and Guo and Thompson [1994] derived the observed Fisher matrix using stochastic integration. Based on their work, we will describe a method to determine an estimate of the observed Fisher information $I_{\boldsymbol{Y}}$ in the context of NLME modeling

$$\widehat{I}_{\boldsymbol{Y}}(\beta) = -\nabla^2_\beta \log p^{\boldsymbol{Y}}(\boldsymbol{y};\beta). \tag{3.6}$$

We will extend the derivation by Walker [1996] and Guo and Thompson [1994] and will also consider fixed effects not associated with random effects. Parts of this formula cannot be expressed as closed-form and have to be approximated using either stochastic or analytical approximations.

An equivalent representation of Equation (3.6) is given by the so–called Louis' formula (missing information principle), see Louis [1982] and the Appendix 8, page 126 for details,

$$\begin{aligned}\widehat{I}_{\boldsymbol{Y}}(\beta) = \mathbf{E}_\beta\left[-\nabla^2_\beta \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta)\right] \\ - \mathrm{Var}_\beta\left[\nabla_\beta \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta)\right],\end{aligned} \tag{3.7}$$

with $\boldsymbol{Z} := (\boldsymbol{\Theta}|\boldsymbol{Y} = \boldsymbol{y})$. The first derivatives of $\log p^{\boldsymbol{Y},\boldsymbol{\Theta}}$ have already been derived in Section 3.1, the second derivatives are relatively easy to obtain and an approximation of the distribution of $\boldsymbol{Z_i}$ was derived in the E–step (as demonstrated in Section 3.1). Because of the independence of $(\boldsymbol{Y_i}, \boldsymbol{\Theta_i})$ and $(\boldsymbol{Y_j}, \boldsymbol{\Theta_j})$, for $i \neq j$, we have

$$\begin{aligned}p^{\boldsymbol{Z}}(\boldsymbol{\theta};\boldsymbol{y},\beta) = p^{\boldsymbol{\Theta}|\boldsymbol{Y}=\boldsymbol{y}}(\boldsymbol{\theta};\boldsymbol{y},\beta) = \frac{p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{\theta};\beta)}{p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)} \\ = \frac{\prod_{i=1}^N p^{\boldsymbol{Y_i},\boldsymbol{\Theta_i}}(\boldsymbol{y_i},\theta_i;\beta)}{\prod_{i=1}^N p^{\boldsymbol{Y_i}}(\boldsymbol{y_i};\beta)} = \prod_{i=1}^N p^{\boldsymbol{\Theta_i}|\boldsymbol{Y_i}=\boldsymbol{y_i}}(\theta_i;\boldsymbol{y_i},\beta) = \prod_{i=1}^N p^{\boldsymbol{Z_i}}(\theta_i;\boldsymbol{y_i},\beta),\end{aligned}$$

for $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)^T$, $\theta_i \in \mathbb{R}^p$ for $1 \leq i \leq N$.

Thus, given the final estimates of the EM algorithm $\hat{\beta} = (\hat{\theta}, \hat{\Omega}, \hat{\vartheta}, \hat{\sigma}^2)$, the first derivatives of $\log p^{\boldsymbol{Y},\boldsymbol{\Theta}}$ are given or can be approximated by (see page 34)

- $a = \nabla_\theta \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta) = \Omega^{-1}\sum_{i=1}^N(\theta - \boldsymbol{Z_i})$,
- $b = \nabla_\Omega \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta) = \frac{1}{2}\Omega^{-1}\sum_{i=1}^N\left(-\Omega + (\theta - \boldsymbol{Z_i})(\theta - \boldsymbol{Z_i})^T\right)\Omega^{-1}$,

- $c = \nabla_\vartheta \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta) \approx \frac{1}{\sigma^2} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \nabla_\vartheta f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\hat{\vartheta})$
  $\cdot \Big(\boldsymbol{y_{ij}} - f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\hat{\vartheta}) - \nabla_\vartheta f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\hat{\vartheta})^T(\vartheta - \hat{\vartheta})\Big)$, where $\hat{\theta}_i(\boldsymbol{y_i};\beta)$ was defined in Equation (2.20),
- $d = \nabla_{\sigma^2} \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta) = -\frac{N_n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{N} \sum_{j=1}^{n_i} (\boldsymbol{y_{ij}} - f(x_{ij};\boldsymbol{Z_i},\vartheta))^2$.

The Hessian given by Walker [1996] is extended by the fixed effects not associated with random effects ϑ

$$-\nabla^2_\beta \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta) = \begin{pmatrix} A & AB & 0 & 0 \\ AB^T & B & 0 & 0 \\ 0 & 0 & C & 0 \\ 0 & 0 & 0 & D \end{pmatrix},$$

with

- $A = \frac{\partial^2}{\partial\theta^2} \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta) = N\Omega^{-1}$,
- $(AB)_{qr} = \left(\frac{\partial^2}{\partial\Omega\partial\theta} \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta)\right)_{qr} = \mathbb{1}_q^T \Omega^{-1} \left(\frac{\partial}{\partial\Omega_r}\Omega\right) \Omega^{-1} \sum_{i=1}^{N} (\boldsymbol{Z_i} - \theta)$, for $1 \le q \le p$, $1 \le r \le p^2$, where $\mathbb{1}_q$ denotes the q-th column of the $p \times p$ identity matrix, p denotes the number of random effects and Ω_r the r-th entry of Ω in the sense that $\Omega_1 = \Omega_{(1,1)}$, $\Omega_2 = \Omega_{(1,2)}$, $\ldots$,$\Omega_{p+1} = \Omega_{2,1}$,$\ldots$,$\Omega_{p^2} = \Omega_{(p,p)}$.
- $B_{qr} = \frac{1}{2} \operatorname{trace}\left(\Omega^{-1} \left(\frac{\partial}{\partial\Omega_q}\Omega\right) \sum_{i=1}^{N} \left[2\left(\boldsymbol{Z_i} - \theta\right)\left(\boldsymbol{Z_i} - \theta\right) - \Omega\right] \Omega^{-1} \left(\frac{\partial}{\partial\Omega_r}\Omega\right)\right)$ for $1 \le q, r \le p^2$,
- $C = -\frac{\partial^2}{\partial\vartheta^2} \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta) = -\frac{1}{\sigma^{2\prime}} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \nabla_\vartheta f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\hat{\vartheta}) \nabla_\vartheta f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta),\hat{\vartheta})^T$,
- $D = -\frac{N_n}{2\sigma^4} + \frac{1}{2\sigma^6} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \left(\boldsymbol{y_{ij}} - f(x_{ij};\boldsymbol{Z_i},\hat{\vartheta})\right)^2$.

Because A, c and C are independent of $\boldsymbol{Z_i}$, the respective expectation values are constant (with variance 0). All remaining terms depend on $\boldsymbol{Z_i}$. As a matter of fact the terms b, d, AB, B and D are nonlinear functions of $\boldsymbol{Z_i}$, such that the corresponding expectation and variance in Equation (3.7) cannot be derived analytically and further approximations have to be used.

To avoid further linearization and because this step is only performed once at the end of the algorithm, we recommend to use a stochastic integration method as proposed by Guo and Thompson [1994] to derive the distributions of a, b, d, A, AB, C and D. For simplification, we further recommend to use the approximated distribution of $\boldsymbol{Z_i}$ derived in Section 2.2.3, although MCMC methods can be used to derive the distribution of $\boldsymbol{Z_i}$ more precisely.

3.3. Example: The EM–Algorithm —Classical Population Analysis of Levofloxacin Plasma Data

To exemplify NLME modeling and the deterministic EM algorithm, the levofloxacin dataset as described in the next Section 3.3.1 is used. A brief description of the drug levofloxacin is given later in Section 5.1. Prior to the NLME analysis, we briefly present two alternative and simple methods, namely the naive pooling and the two–stage approach, to analysis the data. The mathematical background of both approaches is described in the Appendix 9 (see page 128 and page 129). Due to their simplicity, both methods are occasionally used for the estimation of parameters in the PBPK context. For example, the naive pooling is used for parameter estimation in the PBPK context if the study population is assumed to be very homogenous and only few PK samples are available per individual, e.g. as in many preclinical studies (Tsamandouras et al. [2013]). On the other hand, if dense individual PK profiles are available, the data is often analysed on the individual level only—representing one part of the two–stage approach, e.g. see Levitt and Schnider [2005]. The performance of the naive pooling, two–stage and NLME approach will be compared based on the analysis of the levofloxacin study and a simulations & estimation study.

The optimization in the naive pooling, the two–stage and the E–step of the EM algorithm was performed using the optimization package *ucminf* (Nielsen and Mortensen [2012]).

3.3.1. Levofloxacin Dataset

The NLME approach will be illustrated based on the analysis of PK data of levofloxacin which was collected in four studies conducted by Prof. Markus Zeitlinger from the Medical University of Vienna, Austria (Zeitlinger et al. [2003, 2007]; Bellmann et al. [2004]). Levofloxacin was measured in plasma, muscle and adipose ISF tissue, the latter two were obtained by microdialysis. All studies were single dose studies with 12, 30 and 60 minutes intravenous (i.v.) infusion of 500 mg levofloxacin. Furthermore, several patients covariates were collected: age, sex, body height, body weight, albumin concentration in plasma and the hematocrit value.

For our purpose only those patients were included in the analysis for whom information about the patient's sex, body height and body weight was provided. The resulting dataset consisted of 24 patients, 19 male and 5 female with a total of 293 plasma (total drug) samples (9 to 14 samples per subject), 193 adipose tissue samples from 19 patients (6 to 12 samples per subject) and 181 muscle tissue samples from 18 patients (6 to 12 samples per subject). The median age was 62 years, with a range of 44 to 80 years.

3.3.2. The structural model

The raw data of the plasma PK data is shown in Figure 3.2. The visual assessment of the raw data strongly suggest that the pharmacokinetics of levofloxacin follow a bi–phasic kinetic with a fast distribution phase within the first two hours, followed by a slower elimination phase. Thus, we chose a two compartment model as structural model to describe the pharmacokinetics of levofloxacin. This is in line with previous investigation of the plasma

PK of levofloxacin, e.g. see Fish and Chow [1997].

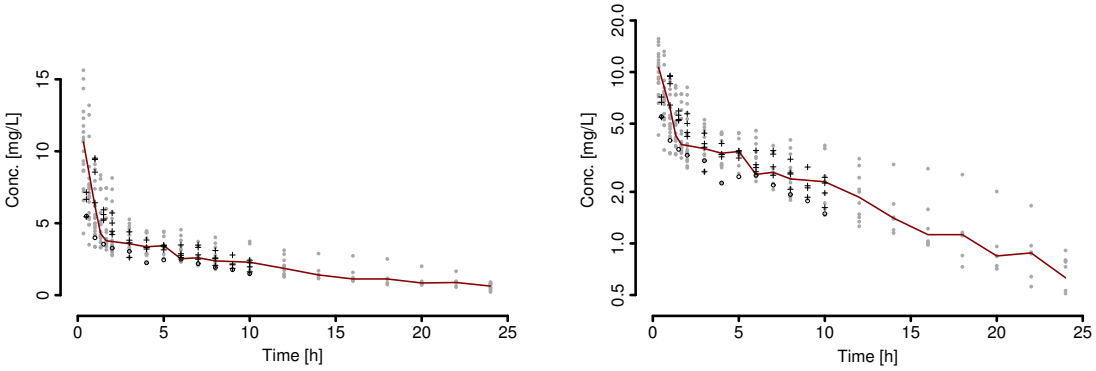

Figure 3.2.: *Plasma PK measurements of $N = 24$ subjects for linear (left) and log–scale (right): Black circles (1 subject), gray (19 subjects) and black crosses (4 subjects) show observations following 12, 30 and 60 minutes infusion, respectively. The red solid line shows median of measurements from the 19 subjects with 30 minutes infusion.*

A schematic illustration of a two compartment model is shown in Figure 3.3. The ODE system describing the rate of change of the amounts A_1 and A_2 in the central and peripheral compartment is given by

$$\begin{aligned} \frac{dA_1(t)}{dt} &= -k_{12}A_1(t) + k_{21}A_2(t) - k_{10}A_1(t) + r(t) \\ \frac{dA_2(t)}{dt} &= +k_{12}A_1(t) - k_{21}A_2(t) \end{aligned} \tag{3.8}$$

with exchange- and elimination rate constants $k_{12} = \mathrm{Q} \,/\, \mathrm{V}_1$, $k_{21} = \mathrm{Q} \,/\, \mathrm{V}_2$ and $k_{10} = \mathrm{Cl} \,/\, \mathrm{V}_1$, and infusion rate $r(t)$, given as $r(t) = \text{dose}/\mathrm{T}_{\text{inf}}$ for $t \leq T_{\text{inf}}$ and $r(t) = 0$ for $t > T_{\text{inf}}$, where $T_{\text{inf}} > 0$ denotes the infusion duration. The parameters Cl, V_1, V_2 and Q denote the clearance, the apparent volumes of distributions of the central and the peripheral compartment and the inter–compartmental clearance, respectively. The system of ODEs in Equation 3.8 can be solved analytically, e.g. see Dubois et al. [2011]. The drug concentration in the central compartment represents the plasma concentration. Thus, the structural model is given by $f(x_i; \theta_i) := C_1(t) = A_1(t) / \mathrm{V}_1$.

Because, PK parameters Cl, V_1, Q and V_2 and can only take positive values, a log–transformation is applied and the following parametrization for f is used

$$\theta_i = (\log \mathrm{Cl}_i, \log \mathrm{V}_{1i}, \log \mathrm{Q}_i, \log \mathrm{V}_{2i})^T,$$

with V_1, V_2 in unit L and Q, Cl in unit L/h. The independent variable x_i consists of the individual dose administered, the infusion duration and the sampling time points $t_{i1}, \ldots, t_{in_i}$.

3.3.3. The Within–Subject Variability Model

The WSV is assumed to be proportional to the predicted concentrations. Following the considerations described in Section 1.2 the log–transformed measurements $\log(\boldsymbol{y_i})$ are assumed

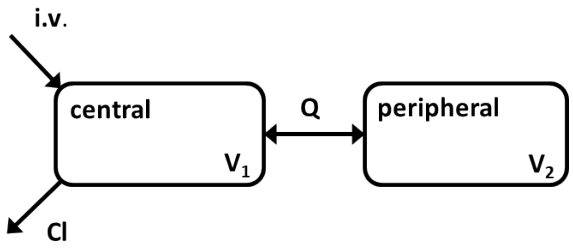

Figure 3.3.: *Schematic representation of a two compartment model: central compartment represents the dosing and observation compartment—e.g. the plasma—from where the drug is eliminated, the peripheral compartment represents sub-part of the body with different kinetics as the central compartment.*

to be realizations of the random variable $\boldsymbol{Y_i}$ defined by

$$\boldsymbol{Y_i} = \log f(x_i; \theta_i) + \boldsymbol{\epsilon_i},\ \boldsymbol{\epsilon_i} \sim \mathcal{N}_{n_i}\left(0,\ I_{n_i}\sigma^2\right),$$

with $\sigma^2 > 0$.

3.3.4. The Between–Subject Variability Model

The BSV is characterized by

$$\boldsymbol{\Theta_i} \sim \mathcal{N}_4\left(\theta, \Omega\right),$$

with $\theta = (\log \mathrm{Cl}, \log \mathrm{V}_1, \log \mathrm{Q}, \log \mathrm{V}_2)^T \in \mathbb{R}^4$ and $\Omega \in \mathbb{R}^{4\times 4}_{>0}$. Due to the log–transformation, the parameters Cl, V_1, Q and V_2 are log–normally distributed which is a widely used assumption, see Karlsson et al. [1998].

3.3.5. The Naive Pooling Approach

Under the assumption of an additive error for the log–transformed data, the objective function of the naive pooling analysis is given by

$$\mathrm{OFV}_{\mathrm{NP}}(\theta; \boldsymbol{y_1}, \ldots, \boldsymbol{y_N}) = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left(\log \boldsymbol{y_{ij}} - \log f(x_{ij}; \theta)\right)^2,$$

where OFV denotes the objective funtion value (OFV). The estimates are defined as

$$\hat{\theta}_{\mathrm{NP}} = \underset{\theta\in\mathbb{R}^p}{\arg\min}\ \mathrm{OFV}\left(\theta; \boldsymbol{y_1}, \ldots, \boldsymbol{y_N}\right) \text{ and } \hat{\sigma}^2_{\mathrm{NP}} = \frac{\mathrm{OFV}\left(\hat{\theta}_{\mathrm{NP}}; \boldsymbol{y_1}, \ldots, \boldsymbol{y_N}\right)}{N_n}.$$

For a successful optimization the choice of initial values $\theta_i^{(0)}$ is important and usually different initial values should be tested, see Bonate [2011, p. 114]. For levofloxacin we could extract values for clearance (8.64 L/h to 13.56 L/h) and total volume of distribution ($= \mathrm{V}_1 + \mathrm{V}_2$)

(89 L to 112 L) from Fish and Chow [1997]. For the regression analysis the respective log-values of 20 L, 80 L, 10 L/h and 30 L/h are chosen as initial values for $\log V_1$, $\log V_2$, $\log Cl$ and $\log Q$, respectively. Thereby, the choice for $\log V_1$, $\log V_2$ and $\log Q$ is partly arbitrary.

The results of the naive pooling approach are displayed in Table 3.1. Per definition, the naive pooling approach does not provide information about BSV. As a consequence, the WSV estimate covers both sources of variability–within and between subject variability.

Table 3.1.: *Results of the naive pooling approach, WSV expressed as CV.*

	Cl	V_1	Q	V_2
$\exp\hat{\theta}$	8.3 L/h	17.5 L	82.8 L/h	72.3 L
WSV	29.0 %			

The GOF plots are shown in Figure 3.4 and indicate that the naive pooling approach adequately describes the pooled data.

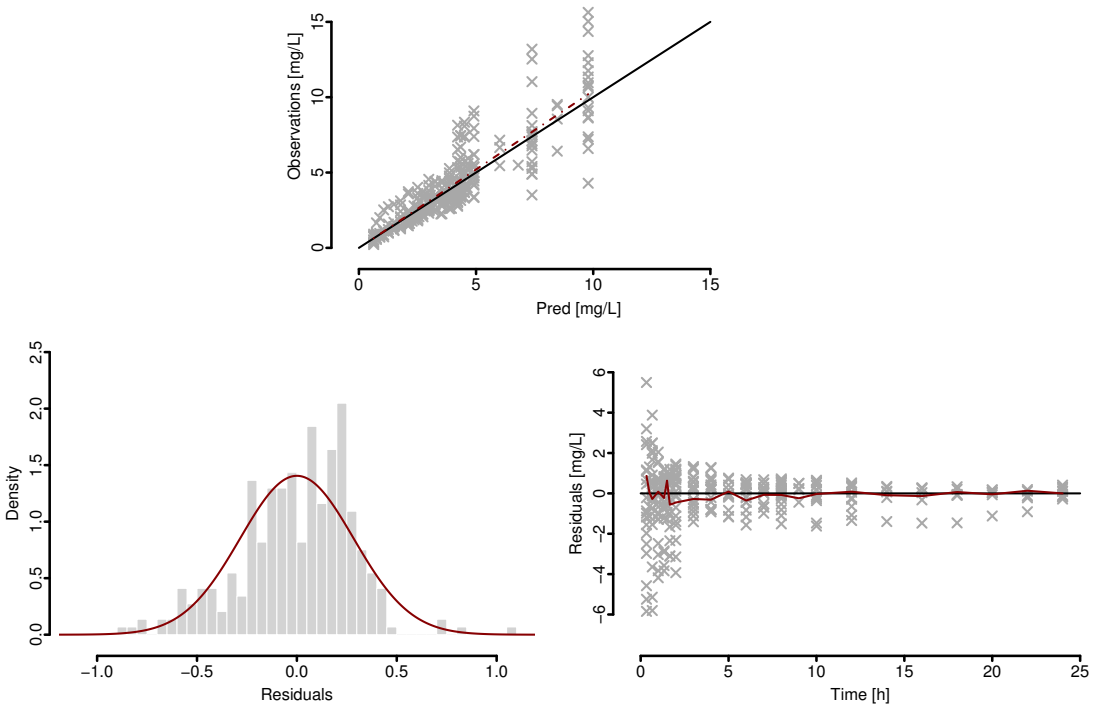

Figure 3.4.: *Goodness of fit plots of the naive pooling analysis: Upper plot shows observations against the predictions (Pred). In the upper graphs, solid black line represents the identity line and red dashed line represent a smoothing spline. Lower left plot shows the empirical density of the residuals, the red line indicates the estimated density of the WSV. Lower right plots shows the residuals over time, red line shows the median of the residuals for each time point.*

3.3.6. The Two–Stage Approach

In the first step the individual ML estimate $\hat{\theta}_i(\boldsymbol{y_i})$ is determined for all subjects $1 \leq i \leq N$

$$\hat{\theta}_i(\boldsymbol{y_i}) = \underset{\theta_i \in \mathbb{R}^p}{\arg\min} \underbrace{\sum_{j=1}^{n_i} (\log \boldsymbol{y_{ij}} - \log f(x_{ij}; \theta_i))^2}_{=:\mathrm{OFV}_{\mathrm{TS}}(\theta_i; \boldsymbol{y_i})} .$$

The same initial values as for the naive pooling approach (Section 3.3.5) were used. The individual residual variance is given by $\hat{\sigma}_i(\boldsymbol{y_i}, \hat{\theta}_i(\boldsymbol{y_i})) = \mathrm{OFV}_{\mathrm{TS}}(\hat{\theta}_i(\boldsymbol{y_i}); \boldsymbol{y_i})/n_i$.

The results of the second stage are stated in Table 3.2. The sample means of the parameter estimates are in a reasonable magnitude of order and in line with literature values. The variance of the BSV is defined as the diagonals of $\hat{\Omega}$. The WSV was estimated to be 8.0 % (CV).

Table 3.2.: *Results of the two–stage analysis: Sample mean and range of individual parameters, BSV and WSV expressed as CV (e.g.* $\mathrm{CV\%\,(Cl)} = 100 \cdot \sqrt{e^{\omega^2_{\log \mathrm{Cl}}} - 1}$*).*

	Cl	V_1	Q	V_2
$\exp \hat{\theta}$	7.9 L/h	19.8 L	61.5 L/h	69.0 L
Range of $\exp \hat{\theta}_i$	3.9–11.5 L/h	7.2–60.2 L	20.6–473.8 L/h	27.2–122.6 L
BSV	30.6 %	64.8 %	71.0 %	42.8 %
WSV	8.0 %			
Range	3.6–16.5 %			

The small WSV can also be observed in the GOF plots, see Figure 3.5. Observations and individual predictions are in good agreement, furthermore the model does not tend to a systematical under– or over–prediction. The residuals are symmetrically distributed around 0, relatively constant over the observation period. Please note, that only few observations are available for $t \geq 12$ h—approximately 6 per time point.

In conclusion, the two–stage approach is able to describe the individual observed profiles very adequately. However, the range of the individual estimates is noticeably broad, particularly in Q with minimal and maximal deviation from the population mean of −66.5 % and 670 %, respectively.

3.3.7. The NLME Approach

As initial values for the deterministic EM algorithm for the fixed effects the values were chosen according to the description in Section 3.3.5, for the BSV for all parameters a value of CV% 30 and for the WSV a value of CV% 10 was chosen. As convergence criterion of the EM algorithm three significant digits in the fixed effects were chosen; only the fixed effects were considered to save computational time and because they were considered to be the most relevant parameters.

Figure 3.6 illustrates how the EM algorithm works in case the initial values are determined from the two–stage analysis: In general, the two–approach over-estimates the BSV and

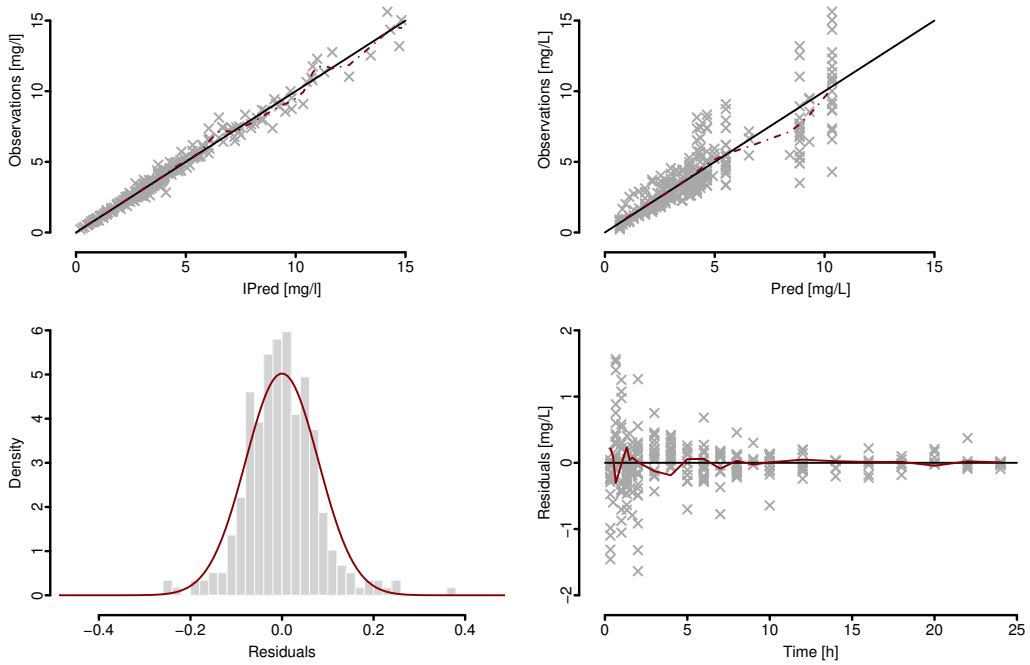

Figure 3.5.: *Goodness of fit plots of the two–stage analysis: Upper left plot shows observations against the individual predictions (IPred), upper right plot shows observations against the predictions based on the population mean $\hat{\theta}$ (Pred). In the upper graphs, solid black line represents the identity line and red dashed line represent a smoothing spline. For both plots apply: the closer the black and the red lines are, the better. Lower left plot shows the empirical density of the residuals, the red line indicates the estimated density of the WSV. Lower right plots shows the residuals over time, red line shows the median of the residuals for each time point.*

under-estimates the WSV, the EM algorithm adjusts WSV and BSV estimates until an equilibrium between both sources of variability is reached.

The results of the EM algorithm are displayed in Table 3.3. In addition to the parameter estimates, the so–called η– and ϵ–shrinkage in percentage for the variance estimates are provided.

Shrinkage

The shrinkage estimate of a variance term—η for a BSV, and ϵ for WSV—indicates whether the information in the underlying data is sufficient to determine individual estimates for that parameter. The shrinkage estimates are defined as (Savic and Karlsson [2009])

$$\eta\text{–shr}(\hat{\Omega}_{\text{k,k}}) := \left(1 - \frac{\widehat{\text{Sd}}\left((\theta_1(\boldsymbol{y_1};\beta),\ldots,\theta_N(\boldsymbol{y_N};\beta))_k\right)}{\sqrt{\hat{\Omega}_{k,k}}}\right),$$
$$\epsilon - \text{shr}(\hat{\sigma}^2) := \left(1 - \frac{\widehat{\text{Sd}}\left(\hat{\epsilon}_{11},\ldots,\hat{\epsilon}_{Nn_N}\right)}{\sqrt{\hat{\sigma}^2}}\right), \tag{3.9}$$

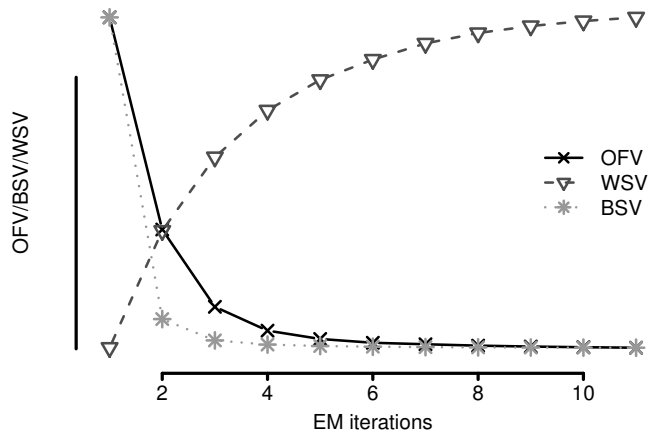

Figure 3.6.: *Exemplified performance of the EM algorithm for levofloxacin if parameters determined from the two–stage approach are used as initial values: The three lines show how the OFV (crosses and solid line), the product of the BSVs terms (asterix and dotted line) and the WSV (triangles and dashed line) evolve over the iterations.*

with $(\theta_1(\boldsymbol{y_1};\beta),\ldots,\theta_N(\boldsymbol{y_N};\beta))_k$ denoting the k-th entry of each EBE, $\hat{\Omega}_{k,k}$ denoting the k-th diagonal entry of $\hat{\Omega}$, $\hat{\epsilon}_{ij} := \boldsymbol{y_{ij}} - f(x_{ij};\hat{\theta}_i(\boldsymbol{y_i};\beta))$ denoting the residuals of the individual model predictions and observations and $\widehat{\text{Sd}}$ denotes the sample standard deviation. Please note, the above definition only holds true for an additive error model. For more details see Savic and Karlsson [2009]. The above definition is commonly used in applied pharmacometric analysis, however in statistics, typically the variance is used instead of the standard deviation, see Lavielle and Ribba [2016] or Xu et al. [2012]. This is motivated by the consideration that the estimate of the population variance (Ω) is composed of the (explained) variance of the EBEs $\theta_1(\boldsymbol{y_1};\beta),\ldots,\theta_N(\boldsymbol{y_N};\beta)$ and of the (unexplained) variance $\text{Var}_\beta[\boldsymbol{Z_i}]$, where the later is assumed to be orthogonal (uncorrelated).

The η–shrinkage is an indicator whether conclusion based on the EBEs are reliably, such as "the distribution of the individual parameters and/or the correlation structure of the random effects" (Lavielle and Ribba [2016]). From the definition of the shrinkage it is obvious, that an optimal value would be close to 0 and that theoretically, if the BSV is estimated using the presented EM algorithm, negative values are not possible. According to Savic and Karlsson [2009], shrinkage values in the magnitude up to 30 % are acceptable. Large shrinkage values are typically observed if only uninformative observations (w.r.t. to the parameter) for the majority of subjects are available and the associated EBEs shrink towards the population value.

For the underlying model the shrinkage estimates are all below 30 %. This indicates that the information contained in the data is sufficient to reliably determine EBEs for all random effects parameters.

The GOF plots indicate an adequate description of the data, see Figure 3.7. The comparison of the empirical distribution of the residuals and the respective density of the WSV reveals a slight over-prediction of the WSV, as already indicated by the ϵ-shrinkage value.

Table 3.3.: *Results of the classical population PK analysis using the deterministic EM algorithm: Typical population parameters and range of EBEs are presented on the original scale (e.g.* $\mathrm{Cl} = e^{\log \mathrm{Cl}}$*), BSV and WSV are expressed as CV (e.g.* $\mathrm{CV\%}(\mathrm{Cl}) = 100 \cdot \sqrt{e^{\omega^2_{\log \mathrm{Cl}}} - 1}$*).*

	Cl	V_1	Q	V_2
Typical value $\exp \hat{\theta}$	8.1 L/h	19.9 L	65.0 L/h	69.2 L
Range of $\exp \hat{\theta}_i$	4.2–11.4 L/h	10.9–39.3 L	25.2–214.7 L/h	31.6–110.7 L
BSV	27.3 %	46.3 %	38.5 %	54.3 %
η-shrinkage	0.6 %	8.1 %	0.6 %	2.2 %
WSV	9.9 %			
ϵ-shrinkage	17.8 %			
OFV	−286.5			

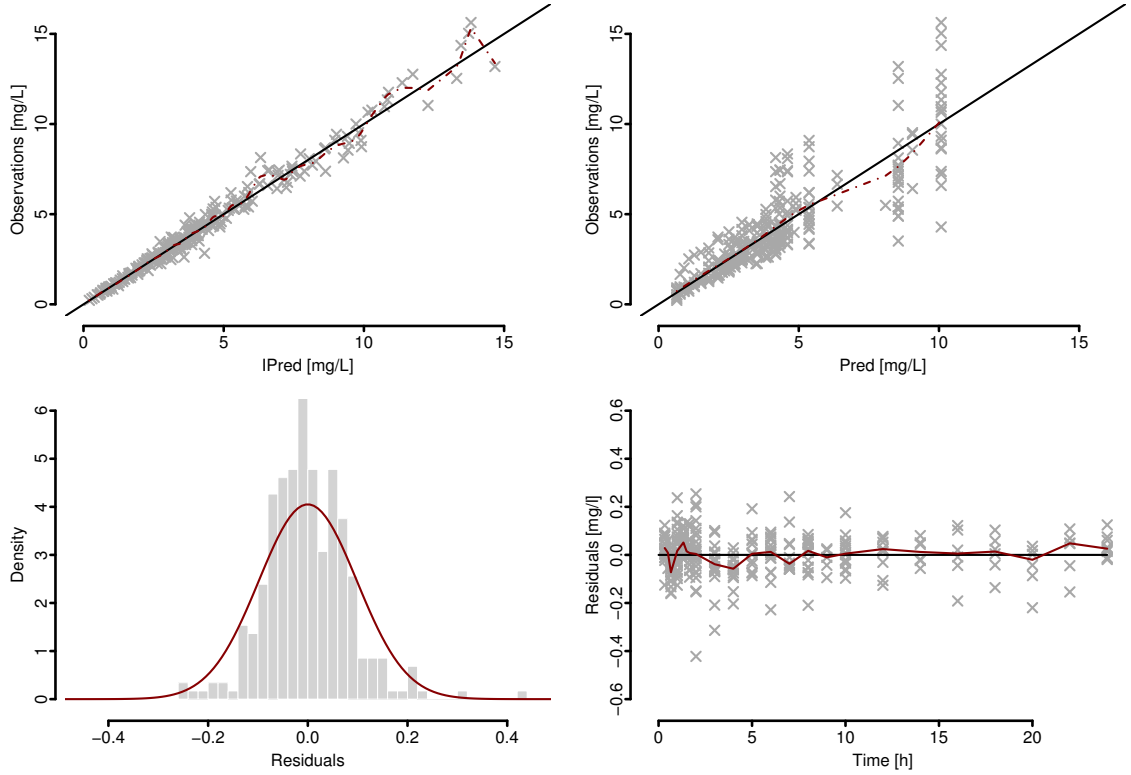

Figure 3.7.: *Goodness of fit plots of NLME analysis: Upper left plot shows observations against the individual predictions (IPred), upper right plot shows observations against the predictions based on the population mean* $\hat{\theta}$ *(Pred). In the upper graphs, solid line represents the identity line and red dashed line represent a smoothing spline. Lower left plot shows the empirical density of the residuals (with empirical density is meant, that the empirical frequency distribution is scaled in a way that the histogram has an area of one), the red line indicates the estimated density of the WSV. Lower right plots shows the residuals over time, red line shows the median of the residuals for each time point.*

A further diagnostic tool, to examine whether the estimated variability terms are in the correct order of magnitude is given by the so–called VPC.

The VPC

The purpose of the VPC is to visually check how well the model performs, in particular how well the overall variability is captured by the model, e.g. see Karlsson et al. [1998]. This diagnostic tool is created in the following procedure: with the final model estimates, PK observations are simulated n_{sim} times with the underlying study design. For the original study, as well as for every simulated study all measurements associated with one time point or predefined time interval, so–called *bins*, are binned together. For each bin, quantiles (e.g. $10^{th}, 50^{th}$ and 90^{th}) of all measurements within the bin are estimated. Thus, for each time point or interval of the original study and each of the pre-defined quantiles one estimate is available, while for the simulated studies n_{sim} estimates of the quantiles are generated. The estimates of the simulated studies are summarized using median and 5^{th} and 95^{th}. The resulting statistics are visually compared with the quantiles of the original study. A more detailed description is presented in the Algorithm 3 on page 49.

In the best case, the observed quantiles lie within the $5^{th} - 95^{th}$ interval of the simulated quantiles indicating an adequate choice of the structural and random effects model as well as adequate parameter estimates.

Algorithm 3 Create a VPC

Define number of virtual studies to be simulated n_{sim}.
Define n binning points/intervals $\{(t_{1,l}, t_{1,u}), \ldots, (t_{n,l}, t_{n,u})\}$ (l=lower, u=upper).
Based on subjects EBEs, individual predictions are determined for planned (plan) and actual (act) dosing. For the VPC the dose–corrected observations $\boldsymbol{y_i}(\text{plan}) = \boldsymbol{y_i}(\text{act}) \cdot f(\text{plan})/f(\text{act})$ are used.
For the original study, estimate quantiles of all (dose–corrected) observations $\boldsymbol{y}(t)$ with $t_{j,l} \leq t < t_{j,u}$, where t are the associated measurements times, allocate the time $t_j := \text{median}(t; t_{j,l} \leq t < t_{j,u})$ to the estimated quantile, for each bin $1 \leq j \leq n$.
Set $r = 0$
repeat
  Set $r = r + 1$.
  Sample parameters from the estimated parameter distributions for each individual.
  Simulate predictions for each individual based on the sampled parameter.
  Sample residuals and generate simulated measurements.
  For the resulting r^{th} simulated study, estimate quantiles of simulated $y(t)$ with $t_{j,l} \leq t < t_{j,u}$, where t are the associated measurements times, allocate the time $t_j := \text{median}(t; t_{j,l} \leq t < t_{j,u})$ to the estimated quantile, for $1 \leq j \leq n$.
until $r = n_{sim}$
for $1 \leq j \leq n$ **do**
  Estimate prediction intervals (e.g. $5^{th} - 95^{th}$) for all quantiles for each bin.
end for

For some design parameters, e.g. absolute amount of drug administered, observations and predictions can be normalized w.r.t. the respective design variable and normalized VPC can

be created as done for underlying levofloxacin population PK analysis.

The bins should be defined, such that the prediction curve does not change significantly within each bin. In addition, each interval should contain a sufficient number of measurements—from different subjects—so that quantiles can be estimated robustly, e.g. it is not reasonable to estimate the 5^{th} and 95^{th} quantiles of less than 20 measurements. In this case either smaller quantile ranges should be chosen, e.g. 10^{th} and 90^{th}, or the binning interval has to be adjusted if possible.

The VPC for the levofloxacin analysis is shown in Figure 3.8. For all time points median, 10^{th} and 90^{th} quantile of the observations lie in the respective prediction interval of the simulated quantiles. This indicates a good agreement of the estimated typical profile and the median observations, as well as for the estimated and observed variability. Howsoever, for $t \geq 12$ h the prediction interval increases due to the low number of samples within each bin (on average 6), which makes it difficult to assess whether the variability in the later phase is correctly captured or slightly overestimated by the model.

In summary, the GOF plots and the VPC show that the model adequately describes the data. Furthermore, the estimates are in good agreement with the results obtain by Fish and Chow [1997], who reported a total volume of distribution between 89 L to 112 L and a clearance between 8.6 L/h to 13.56 L/h, whereas in the current analysis values of 89.1 L and 8.1 L/h were estimated.

3.3.8. Comparison of Naive Pooling, Two–Stage and NLME (using the deterministic EM algorithm) approach

In this section the naive pooling, two–stage and NLME approach are compared with each other on the basis of the levofloxacin study. Therefore, a simulation & estimation study was conducted using the final parameters estimates of the classical levofloxacin population PK analysis presented in the previous section. In total, $n_{\text{sim}} = 100$ studies were simulated. For each simulated study, 24 (number of subjects) parameter vectors were sampled from the distribution estimated in Section 3.3.7. Based on the sampled parameter values, levofloxacin concentration for the given sampling time points were simulated. These simulated concentrations were transformed to simulated observations by adding a randomly sampled residual error. Each trial was analysed using naive pooling, two–stage and NLME approach. For each method, relative bias (relBias) and relative root mean squarred error (RSME) was estimated based on the results of the n_{sim} analyses

$$\begin{aligned} \text{relBias} &:= \frac{1}{n_{\text{sim}}} \sum_{l=1}^{n_{\text{sim}}} \left(\frac{\hat{\beta}_l - \beta_0}{\beta_0} \right), \\ \text{relRMSE} &:= \sqrt{\frac{1}{n_{\text{sim}}} \sum_{l=1}^{n_{\text{sim}}} \left(\frac{\hat{\beta}_l - \beta_0}{\beta_0} \right)^2}, \end{aligned} \tag{3.10}$$

where for the l-th study, $\hat{\beta}_l$ represents the estimated parameter value and β_0 is the respective known input parameter. The results are displayed in Table 3.4

For the underlying situation, a small study population with dense and balanced sampling, the deterministic EM provides the less biased estimates compared to the naive pooling and

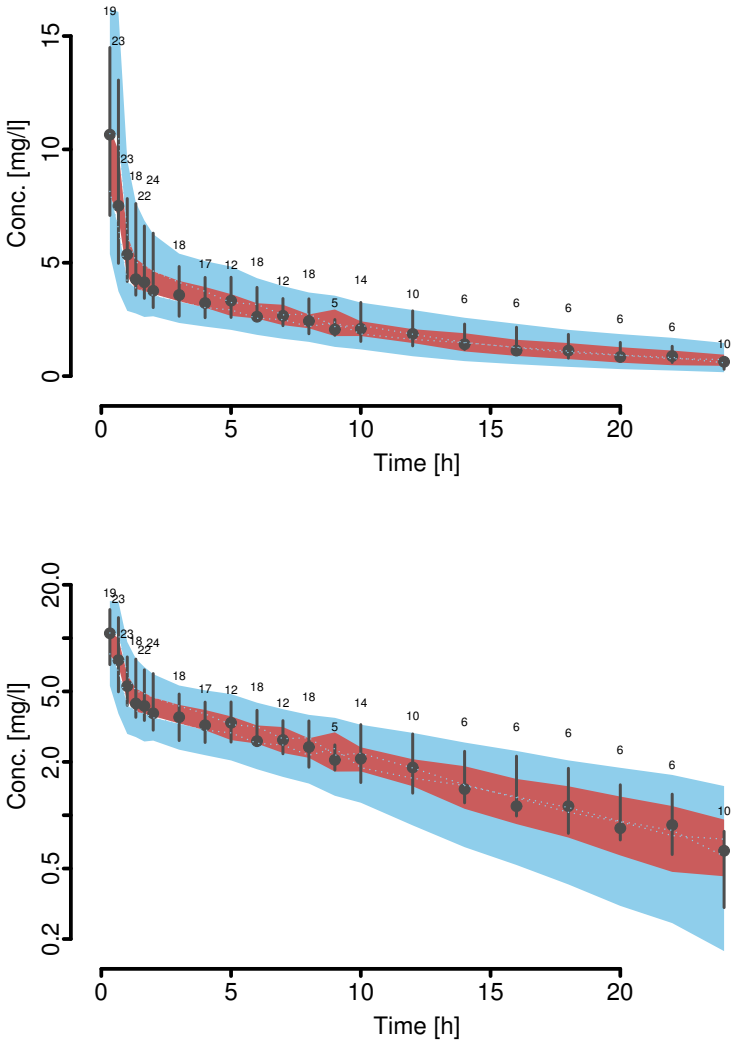

Figure 3.8.: *VPC for the classical population PK analysis for linear (top) and log–scale (bottom): The black bars and dots show the* $10^{th}, 90^{th}$ *and the* 50^{th} *(the median) quantiles of the observations for each time point. The numbers above each bar specifies the underlying number of observations at this time point. The blue shaded area shows the* 95^{th} *error band around the* 5^{th} *and the* 95^{th} *quantiles of the simulated measurements, the red area shows the* 95^{th} *error band of the median of the simulated measurements. The dotted blue line shows the upper and lower range of the prediction interval of the* 10^{th} *and* 90^{th} *quantiles.*

two–stage approach. In particular, for the random effect terms the deterministic EM algorithm outperform the two–stage approach, while the naive pooling approach per definition does not provide estimates for BSV. As expected, in average the two–stage approach overestimates the BSV and underestimates the WSV, with a relative large variability represented by the relative root–mean–square error. This is in line with previous findings by Sheiner

Table 3.4.: *Comparison of naive pooling, two–stage and NLME (deterministic EM) approach based on relative bias and relative RMSE (in brackets) of the results of the simulation & estimation study.*

	Naive Pooling	Two–Stage	deterministic EM
typical value Cl	5.5 % (8.9)	−2.7 % (9.2)	−1.3 % (5.9)
typical value V_1	6.1 % (14.2)	−8.2 % (14.0)	−0.8 % (9.8)
typical value Q	12.3 % (19.9)	−2.0 % (13.6)	1.5 % (11.9)
typical value V_2	3.7 % (9.2)	0.6 % (8.1)	0.8 % (7.5)
BSV Cl	na	166.2 % (851.6)	−3.1 % (15.7)
BSV V_1	na	54.9 % (85.0)	−8.6 % (22.4)
BSV Q	na	29.1 % (57.6)	−3.1 % (19.6)
BSV V_2	na	7.8 % (28.4)	−4.1 % (18.3)
WSV	na	−18.2 % (18.6)	−0.6 % (4.8)

and Beal [1980] that the "two–stage approach produce good estimates of mean kinetics, but biased and imprecise estimates of inter-individual variability".

A main disadvantage of the naive pooling approach is the missing ability to differentiate between different sources of variability. Moreover, assuming that PK varies between patients, "the method violates the assumption of independence in the residuals, as it is likely that observations within an individual are correlated" (Bonate [2005]). This validity of the assumption is illustrated for the underlying analyses in Figure 3.9. For each subject the mean of the residuals was estimated

$$\bar{\epsilon}_i(\theta_i) := \frac{1}{n_i}\sum_{j=1}^{n_i}(f(x_{ij};\theta_i) - \boldsymbol{y_{ij}}),$$

for $\theta_i \in \{\hat{\theta}_{\mathrm{NP}}, \hat{\theta}_i(\boldsymbol{y_i}), \hat{\theta}_i(\boldsymbol{y_i};\beta)\}$ and for $1 \leq i \leq N$, where $\hat{\theta}_{\mathrm{NP}}$ denotes the estimate resulting from the naive pooling approach and $\hat{\theta}_i(\boldsymbol{y_i})$ and $\hat{\theta}_i(\boldsymbol{y_i};\beta)$ are the individual estimates from two–stage and NLME analysis, respectively. For the two–stage and even more so for the NLME approach the average residuals are close to zero, fitting the assumptions that $\mathbf{E}[\boldsymbol{\epsilon_i}] = 0$. For the naive pooling approach it seems natural, that for some subjects the associated PK profiles are either entirely over– or under–predicted.

For the underlying example, the above comparison demonstrates the main disadvantages of naive pooling and two–stage and the strengths of the NLME approaches: the deterministic EM produces nearly unbiased estimates with a slight tendency to underestimate the BSV, the two–stage approach generates biased estimates of the random effects parameters (in general, overestimates BSV and underestimates the WSV), while using the naive pooling approach no BSV can be determined.

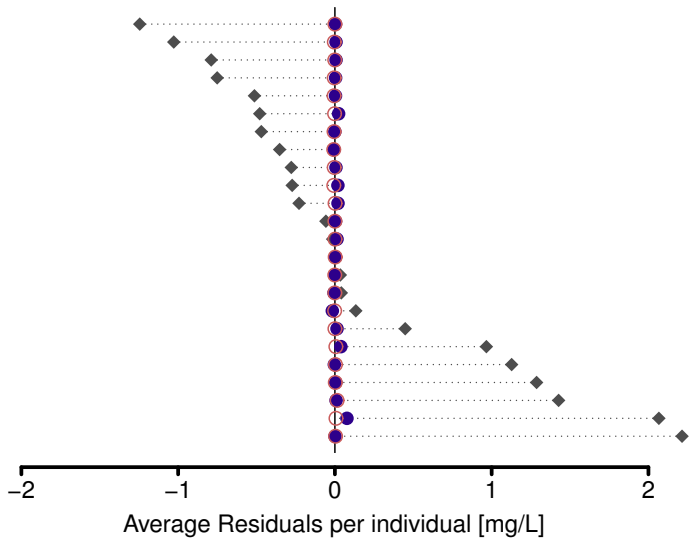

Figure 3.9.: *Comparison of the average individual residuals of naive pooling (grey diamonds), two–stage (blue dots) and NLME approach (red circles) for each of the* 24 *subjects.*

3.3.9. Evaluation of the deterministic EM algorithm

In this section the efficiency of the deterministic EM algorithm described in Section 3.1 is examined and compared to the SAEM algorithm by Kuhn and Lavielle [2005], implemented in R (package *saemix* Comets et al. [2011]). In a first step, the analysis presented in Section 3.3 is repeated using the SAEM algorithm. The parameter estimates are compared with the results presented in Section 3.3 in Table 3.5.

Table 3.5.: *Comparison of the results of the classical population PK analysis of levofloxacin using the deterministic EM (detEM) and SAEM algorithm: Typical values are presented on original scale (e.g.* $\mathrm{Cl} = e^{\log \mathrm{Cl}}$*), BSV and WSV are expressed as CV (e.g.* $\mathrm{CV\%}\,(\mathrm{Cl}) = 100 \cdot \sqrt{e^{\omega^2_{\log \mathrm{Cl}}} - 1}$*).*

	Cl	V_1	Q	V_2
typical value SAEM	8.0 L/h	19.6 L	63.8 L/h	69.4 L
typical value detEM	8.1 L/h	19.9 L	65.0 L/h	69.2 L
relative differences	1.4 %	1.8 %	1.9 %	−0.3 %
BSV SAEM	28.5 %	51.2 %	61.6 %	40.3 %
BSV detEM	27.3 %	46.3 %	54.3 %	38.5 %
relative differences	−4.2 %	−9.6 %	−11.8 %	−4.3 %
WSV SAEM	9.2 %			
WSV detEM	9.9 %			
relative differences	7.3 %			

The parameter estimates of both algorithms are in good agreement with less than 2 % deviation for the fixed effects estimates and deviation of approximately −4 % to 12 % for the

random effects. The computational time was approximately 22 s and 95 s for the deterministic EM and the SAEM algorithm, respectively, while we aimed for as similar conditions as possible for both algorithms.

To confirm these results in a broader setting, a simulation & estimation study was conducted as described in the previous Section 3.3.8 for two scenarios:

(S1) random effects on all PK parameters, i.e. $\theta \in \mathbb{R}^4$ and $\Omega \in \mathbb{R}^{4\times 4}_{>0}$,

(S2) random effects only on Cl and V_2, i.e. $\theta \in \mathbb{R}^2$, $\Omega \in \mathbb{R}^{2\times 2}_{>0}$ and $\vartheta \in \mathbb{R}^2$.

For scenario *(S2)* the same multivariate distribution for Cl and V_2 as in *(S1)* was used. Each simulated study was analysed using the deterministic EM and the SAEM algorithm with identical initial values. For each algorithm, relative bias and relative RSME was estimated based on the results of the n_{sim} analyses, see Equation (3.10).

The results are displayed in Tables 3.6 and 3.7. For the underlying problem—dense sampling in a small study population and relatively simple model structure (in terms of structural and random effects model)—both EM algorithms perform sufficiently well in reproducing in average nearly unbiased estimates with acceptable root–mean–square error. The SAEM algorithm performs slightly better, especially for the random effects terms compared to the deterministic EM algorithm. Similar observations were made by Johansson et al. [2014], where a broader setting of situations was investigated and in average the random effects were underestimated by the deterministic methods (FOCE, Laplacian).

Table 3.6.: *Comparison of the deterministic EM and SAEM algorithm based on relative bias and relative RMSE (in brackets) of the results of the simulation & estimation study S1 with random effects on all four parameters.*

	SAEM	deterministic EM
typical value Cl	−0.3 % (5.5)	1.3 % (5.9)
typical value V_1	2.6 % (10.0)	−0.8 % (9.8)
typical value Q	−1.6 % (11.7)	1.5 % (11.9)
typical value V_2	0.4 % (7.5)	0.8 % (7.5)
BSV Cl	−2.4 % (15.8)	−3.1 % (15.7)
BSV V_1	−1.6 % (21.0)	−8.6 % (22.4)
BSV Q	0.2 % (20.3)	−3.1 % (29.6)
BSV V_2	−2.9 % (19.2)	−4.1 % (18.3)
WSV	−0.02 % (6.3)	−0.6 % (4.8)

Because the ML estimator $\hat{\beta}$ is asymptotically unbiased, the mean squarred error (MSE) defined as

$$\text{MSE}(\hat{\beta}) := \mathbf{E}_\beta\left[(\hat{\beta}-\beta)^2\right] = \text{Var}_\beta\left[\hat{\beta}\right] + \text{Bias}^2(\hat{\beta})$$

converges towards $\text{Var}_\beta\left[\hat{\beta}\right]$ for $n_{\text{sim}} \to \infty$, e.g. see Lehmann and Romano [2005, Section 12.4.1]. Using the *saemix* package, estimates for the variance of $\hat{\beta}$ were determined. Hence, as

Table 3.7.: *Comparison of the deterministic EM and SAEM algorithm based on relative bias and relative RMSE (in brackets) of the results of the simulation & estimation study S2 with random effects only on* Cl *and* V_2.

	SAEM	deterministic EM
typical value Cl	0.2 % (6.0)	1.0 % (6.1)
typical value V_1	0.1 % (4.0)	−1.7 % (4.1)
typical value Q	0.3 % (2.8)	1.0 % (3.0)
typical value V_2	−0.6 % (8.4)	−0.7 % (8.2)
BSV Cl	−5.1 % (17.6)	−5.9 % (17.6)
BSV V_2	−4.8 % (15.9)	−6.3 % (16.0)
WSV	0.01 % (4.2)	0.1 % (4.2)

a consistency check, the relative RSME and the relative standard error (RSE) obtained from the *saemix* package were compared for the untransformed parameter estimates in Table 3.8. The values are in a good agreement considering the relatively small size of the simulation & estimation study of $n_{\text{sim}} = 100$. Concludingly, for the deterministic EM algorithm the relative RSME shown in Table 3.5 could be used to derive the corresponding RSE of the estimates.

Table 3.8.: *Comparison of the relative RSME (relRMSE) resulting from the simulation & estimation studies S1 and S2 and RSE for the SAEM algorithm both based on the untransformed parameter, e.g. for a typical value/fixed effects parameter* (θ) $\log$ Cl *instead of* Cl*, for a BSV variance parameter* $\Omega_{1,1}$ *instead of the respective CV representation and for the WSV* σ *instead of the respective CV representation.*

	relRMSE (S1)	RSE (S1)	relRMSE (S2)	RSE (S2)
θ_1 ($\log$ Cl)	2.6 %	2.8 %	2.9 %	2.6 %
θ_2 ($\log V_1$)	3.2 %	3.7 %	1.3 %	2.7 %
θ_3 ($\log$ Q)	2.8 %	2.9 %	0.7 %	0.9 %
θ_4 ($\log V_2$)	1.8 %	1.9 %	2.0 %	1.5 %
$\Omega_{1,1}$ (BSV Cl)	30.4 %	30.8 %	31.6 %	32.1 %
$\Omega_{2,2}$ (BSV V_1)	38.0 %	35.3 %	–	–
$\Omega_{3,3}$ (BSV Q)	35.1 %	31.7 %	–	–
$\Omega_{4,4}$ (BSV V_2)	35.0 %	30.3 %	28.3 %	31.6 %
σ (WSV)	6.3 %	5.0 %	4.2 %	4.5 %

Obviously, based on this investigation no general conclusion can be drawn. However, the comparison shows that the deterministic EM algorithm performs sufficiently well for the underlying situation. We are aware that the levofloxacin data is a rather ideal example with dense data following single dose PK profile following a relatively simple two–compartment kinetics. Moreover, it should be noted, that not only the methodology between the algorithms differs but also input arguments related to the implementation, e.g. the convergence

criterion or the number of iterations in the stochastic approximation step for the SAEM, which obviously also affects the outcome of such an investigation.

4. Discussion

In the first Part of this thesis a comprehensive overview over the NLME methodology, including the derivation of ML estimates and a deterministic EM algorithm in this context was presented. The NLME approach is a mixed model approach, where fixed and random effects can be incorporated. A major challenge in estimating the model parameter of a NLME model, which is based on the ML method, is that there does not exist an analytical solution of the likelihood function. As a consequence, the commonly used Laplacian, FOCE and FO methods are based on approximations of the likelihood function to derive the ML estimates. These approximations and the respective mathematical background were revisited in Section 2.2.

In this context we introduced new motivations of these approximations. Instead of approximating the likelihood function, the distribution of $\boldsymbol{Z_i} = (\boldsymbol{\Theta_i}|\boldsymbol{Y_i} = \boldsymbol{y_i})$ is approximated by the normal distribution of the random variable $\boldsymbol{Z_i^{LA}}, \boldsymbol{Z_i^{FOCE}}$ or $\boldsymbol{Z_i^{FO}}$. The distribution of $\boldsymbol{Z_i^{LA}}$ and $\boldsymbol{Z_i^{FOCE}}$ are centered around the same mode as $\boldsymbol{Z_i}$, while $\boldsymbol{Z_i^{FO}}$ is centered around the typical population value θ. Thus, $\boldsymbol{Z_i^{LA}}$ and $\boldsymbol{Z_i^{FOCE}}$ only differ in the variance–covariance matrix: the variance–covariance of $\boldsymbol{Z_i^{LA}}$ is given by the inverse of an estimate of the observed Fisher matrix of $\boldsymbol{Z_i}$, while the variance–covariance of $\boldsymbol{Z_i^{FOCE}}$ is given by the inverse of an estimate of the expected Fisher matrix of $\boldsymbol{Z_i}$.

We illustrated the accuracy of approximating the expectation and variance–covariance of $\boldsymbol{Z_i}$ by $\boldsymbol{Z_i^{LA}}, \boldsymbol{Z_i^{FOCE}}$ and $\boldsymbol{Z_i^{FO}}$ based on the analysis of the levofloxacin plasma PK data in Section 2.2.5, whereas the reference distribution of $\boldsymbol{Z_i}$ was derived using a MCMC approach. For this example, the comparison revealed that the Laplacian and the FOCE approximation method provide comparable and adequate approximations for both expectation and variance–covariance, clearly outperforming the FO approximation method, which is in line with earlier findings Bauer et al. [2007]. Of course this example does not replace a rigorous investigation and no general conclusion can be drawn.

However, because for the Laplacian and the FOCE method the mode of $\boldsymbol{Z_i}$ has to be determined using numerical optimization, these methods do not provide a closed-form expression of the approximated likelihood function. As a consequence, the ML estimates can only be determined using an iterative algorithm.

Based on the new interpretation of the FOCE approximation, i.e. approximating the distribution of $\boldsymbol{Z_i}$ by the normal distribution of $\boldsymbol{Z_i^{FOCE}}$, we described a deterministic EM algorithm for NLME models in Section 3 following the algorithms presented by Mentre and Gomeni [1995]; Walker [1996]. Compared to Mentre and Gomeni [1995]; Walker [1996], we complemented the description by details to include and derive the ML estimates for fixed effects not associated with a random effects in the NLME model.

In Example 3.3 we used the deterministic EM algorithm to analysis the population PK of levofloxacin and conducted a simulation & estimation study. Therein, we illustrate how each step of the EM algorithm equilibrates the magnitude of BSV versus the magnitude

of WSV. The result of a simulation & estimation study outlines that the NLME approach outperforms the two–stage and the naive pooling approach in providing adequate estimates for the BSV as well as the WSV for the underlying situation. Furthermore, the simulation & estimation study was used to evaluate the performance of the deterministic EM algorithm compared to the SAEM algorithm implemented in the R–package *saemix* (Comets et al. [2011]). For the underlying situation, this comparison demonstrates that the deterministic estimation approach performs sufficiently well. The details can be found in Section 3.3.9.

In a next step, the performance of the deterministic EM algorithm should be verified on a wider range of applications. For example, different types of problems, with different degrees of complexity regarding the structural and/or stochastic sub-model, and/or the data, e.g. the number of subjects and the numbers of measurements per subject, could elaborate the strengths, weaknesses and limitations of the deterministic EM algorithm, e.g. compared to stochastic EM algorithms.

Furthermore, the presented EM algorithm could be extended to other commonly used types of WSV models, e.g. the proportional or combined error model. In these error models, the WSV depends on $\boldsymbol{\Theta_i}$; e.g. in the proportional error model

$$\boldsymbol{Y_{ij}} = f(x_{ij}; \boldsymbol{\Theta_i}, \vartheta) \cdot (1 + \boldsymbol{\epsilon_{ij}}), \text{ with } \boldsymbol{\epsilon_{ij}} \sim \mathcal{N}\left(0, \sigma^2\right),$$

the estimation of the EBEs $\hat{\theta}_i$ is no longer independent of σ^2. As a consequence, neither the Laplacian based algorithm (see page 33), nor the presented deterministic EM algorithm can be applied, because the iteration steps are no longer independent of each other. According to Beal and Sheiner [1998] one option is to replace the random variable $\boldsymbol{\Theta_i}$ by the respective expectation value θ in the update step for the WSV. As a consequence, the WSV is proportional to the typical population prediction and not to the individual prediction. Beal et al. [1989-2013] offer an interaction option to avoid this problem, but the details of this algorithm are not accessible.

As an alternative, in case of a proportional error and reasonable small WSV, we propose to apply a log transformation of observations and model, combined with an additive error which is expected to approximate the proportional error model very well.

In the next Part II, the presented deterministic EM algorithm is applied to two real world applications.

Part II.

Combining Nonlinear Mixed Effects– and Mechanistic Modeling in Pharmacometrics and Psychometrics

5. Quantitative Pharmacology: Integrating Mechanisms of Drug Distribution and Variability into the Analysis of Population Pharmacokinetic Data

The following chapter is planned to be published; it originated from a cooperation with Prof. Charlotte Kloft, Institute of Clinical Pharmacy, FU Berlin, and her former PhD student André Schäftlein.

5.1. Introduction

In this section the EM algorithm described in the first part of this thesis is applied to the analysis of population pharmacokinetic data. The drug which is investigated is levofloxacin. Levofloxacin is a broad spectrum antibiotic against Gram-positive and Gram-negative bacteria and was approved by the Food and Drug Administration (FDA) in 1996. It is primarily used in treatments of infections of the respiratory and urinary tract, skin and soft tissues and is administered orally or as i.v. infusion, see Fish and Chow [1997]; Neckel et al. [2002]. After its approval it became a blockbuster and was ranked as one of the most frequently prescribed drugs in the US.

To ensure an as safe and as efficacious as possible treatment the pharmacology of a drug is studied in many preclinical and clinical studies. For antibiotics, the emergence of resistance puts both safety and efficacy at risk. The extensive use of levofloxacin increases this risk and underlines the high relevance for a correct understanding of its pharmacology and to preserve the drug's antibacterial activity in future.

The efficacy of an antibiotic is defined as its ability to inhibit bacterial growth or even kill bacteria; a comprehensive review describing the principles and ideas behind, is given by Drusano [2004]. The ability to inhibit bacterial growth is first tested in *in vitro* studies. A standard experiment consists of determination of the minimum inhibitory concentration (MIC) w.r.t. a certain bacterial strain. The MIC is defined as lowest concentration which inhibits visual bacterial growth in a standardized experimental setting. This experiment can be conducted very easily and provides information about (i) whether the antibiotic works against this type of bacteria, and (ii) what drug concentrations at the site of infection should be reached.

As the readout of those experiments are done visually, they are subjective. Furthermore, the efficacy is only quantified in a binary signal: whether the number of colony forming units (CFUs)—an estimate of the bacteria count—after 18 to 24 hours of incubation with the antibiotic $\leq$ number of CFUs at start of the experiment. More informative *in vitro* studies are based on the determination of time–kill curves, which quantify the time–course of bacterial growth or killing depending on the drug concentration.

Additionally, preclinical experiments are conducted in animals, where the conditions are assumed to be more comparable to the human physiology and disease as in *in vitro* studies. An important objective of such preclinical experiments is to identify a surrogate parameter derived from plasma PK profiles which correlates with the efficacy of the drug. Those parameters are called pharmacokinetic–pharmacodynamic (PK/PD) indices. In general, a class of standard indices is tested, e.g.

- $C_{\max}/MIC$, where $C_{\max}$ denotes the peak drug concentration,
- AUC_{0-24}/MIC, where AUC_{0-24} denotes the area under the concentration–time curve (AUC) over 24 hours after start of drug treatment,
- $T_{\geq MIC}$, the duration of time where $C(t) \geq MIC$,

and the one which shows the highest correlation with the predefined endpoint, e.g. number of CFUs at infection site after a certain time, is chosen. For example, for levofloxacin AUC_{0-24}/MIC is the most predictive index for the majority of bacteria strains, see Andes and Craig [2002]; Ambrose et al. [2003]; EUCAST [2007]. Brunner et al. [2005] stated that "in most cases, only the concentration of unbound antibiotic in the ISF at the infection site promotes the antibacterial effect". As a consequence, indices be based on these concentrations are expected to be more appropriate. However, in humans the ISF concentration are very difficult to assess appropriately. Hence, PK/PD indices from plasma might vary substantially in relation to the target site of infection.

In general, these preclinical results are successfully translated into clinics. Heuristically, one could argue that antibiotics target bacteria, which behave similar in different host species. I.e. as long as the target is reached, the mechanism of action and the efficacy should be similar among species. Andes and Craig [2002] draw the conclusion that "animal models have been very useful for determining ... (2) time–course of antimicrobial activity *in vivo*, (3) PK/PD indices correlating with efficacy and (4) magnitudes of the PK/PD index required for efficacy".

Recently, it is getting more and more important to not only describe the efficacy of an antibiotic, but to also quantify the emergence of antibiotic resistance as a function of the drug concentration; this quantification is performed in an analogous procedure as described above, e.g. see Drusano et al. [2006]. Based on these considerations and because in humans only the clinical response—i.e. curing or progress of infections—can be measured, the PK/PD relationship established in preclinical studies are assumed to be similar in humans.

In summary, the plasma PK/PD relationship of antibiotics are usually well understood and the major risk is that too low antibiotic drug concentrations lead to a poor antibacterial activity and likely to resistance, e.g. see Drusano [2004]. However, in general it is not known whether drug concentrations in plasma and within the tissues—the drug target—are similar. Thus, the objective of this work was to learn more about the target concentrations and the link to plasma kinetics. In order to do this, the so–called PBPK modeling approach was used. This method uses prior knowledge about species physiology, e.g. tissue volumes and blood flows, and physicochemical properties of the drug to predict the PK in the whole body. An introduction to PBPK models is given in the next Section 5.2.

Another important aspect is to characterize the BSV of the PK in the patient population to assess whether all patients are treated effectively. Thus, a further objective was to derive a mechanistically motivated approach to integrate BSV into the PBPK model. This is very helpful to identify factors which explain differences between patients or sub–populations, and to use those factors to individualize the dosing if necessary.

The advantage of using PBPK models compared to the often used empirically motivated compartmental models, as used in the Example 3.3, are numerous. The classical approach is only based on the available data; i.e. the data should be reliable and contain all relevant information needed for. Furthermore, classical PK parameters are only apparent parameters and in this sense they are difficult to interpret in a physiologically meaningful manner.

The PBPK modeling approach is based on prior knowledge, i.e. it depends on different data sources, e.g. species– or drug–related. This allows to easily incorporate physiological differences between patient populations, and to make predictions for unstudied patient populations. For example, an obese study population will certainly have an increased fraction of adipose tissue compared to a normal population, which could easily be accounted for in a PBPK model. This is a valuable property, especially, if a sub–population should be investigated, which is critical to be investigated in a clinical study, e.g. due to ethical consideration (e.g. pregnant women, elderly patients or neonates).

One disadvantage of the PBPK approach is that it relies on a large number of model parameters. A large proportion of these parameters, usually the species- and non–drug related parameters, are literature–based and statements about the uncertainty or variability are in general not provided. The drug specific parameters are often not known and have to been derived from *in vitro* or pre–clinical *in vivo* experiments. This makes it sometimes difficult to develop a PBPK model and to judge the reliability of the results. If plasma PK data are available, a common procedure is to estimate the missing, or adapt the existing, parameter values based on these plasma data, e.g. see Tsamandouras et al. [2013]. However, there exist only few examples (e.g. see Ploeger et al. [2001]; Krauss et al. [2013, 2015]; Tsamandouras et al. [2015]) where such approaches have been used in a population context when considering random variability in parameters between individuals.

Another issue of the PBPK approach is that the details of drug–specific processes, e.g. the binding and distribution within tissue or tissue sub–spaces, are not known. Within humans these processes are difficult or impossible to study with experiments. As a consequence, fundamental preclinical experiments are conducted in animals and one assumes that the studied processes can be extrapolated across species. As soon as clinical data is available, one can examine whether the PBPK model, and the underlying assumptions are consistent with the data. However, we were in the exceptional situation to have access to human tissue PK measurements of levofloxacin in the ISF, a common site of infections, obtained by so–called microdialysis. These data gave us the unique opportunity to investigate and verify the mechanistic tissue binding and distribution model of Rodgers et al. [2005b] for humans.

In summary, the objectives of this chapter are:

1. To build a PBPK model for levofloxacin incorporating prior knowledge about the mechanism of drug disposition.

2. To integrate BSV using prior knowledge about the patient population.

3. To establish an approach how population PK plasma data and NLME modeling can be used to refine the PBPK model and estimate missing parameters and the variability between patients.

4. To compare the tissue ISF predictions with the corresponding PK measurements obtained by the microdialysis technique.

A schematic presentation of the resulting new mechanism–driven modeling approach is shown in Figure 5.1. The illustrated process will explained in detail based on the analysis of the levofloxacin plasma PK data in the following sections.

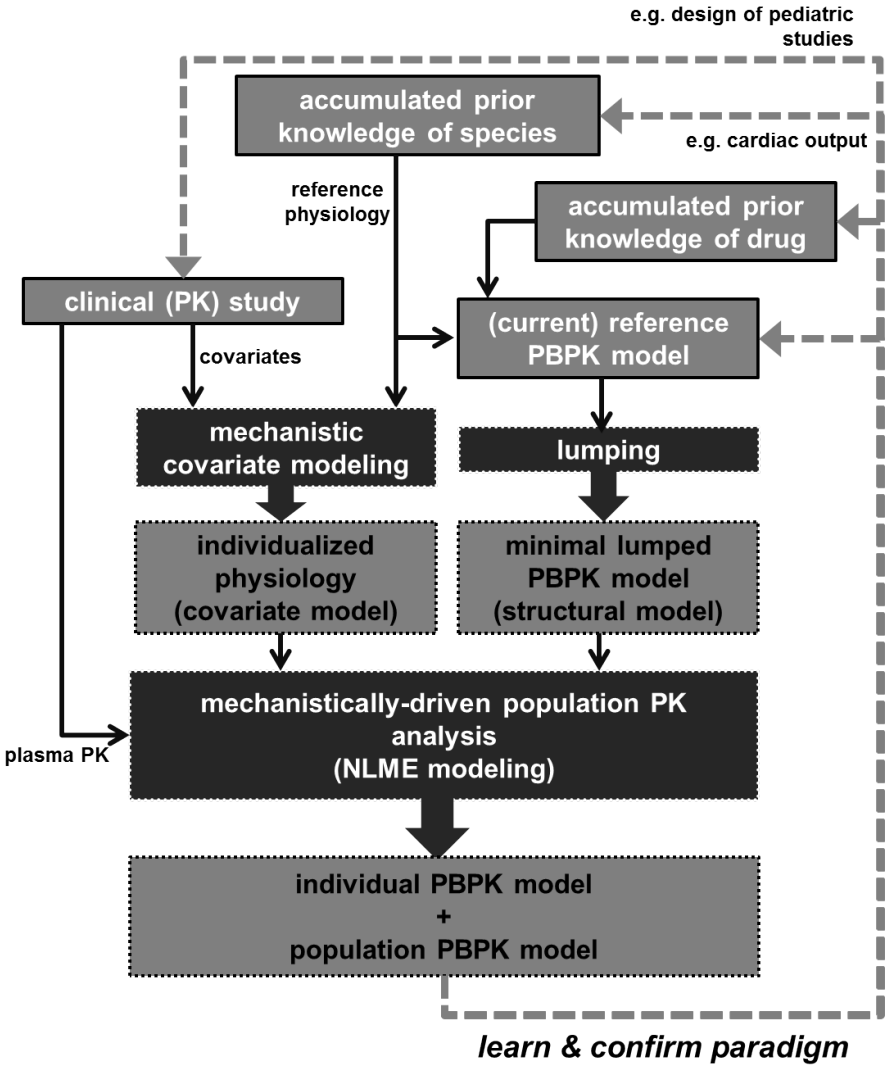

Figure 5.1.: *Schematic illustration of the mechanism–driven analysis of population PK data. Grey boxes & white printing depict the entire input, black boxes & white printing depict the processing steps and grey boxes & black printing depict the results of each processing step.*

The following chapter starts with the development of the PBPK model for levofloxacin in

humans. The PBPK model is individualized using individual patient information according to our work described in Huisinga et al. [2012]. Based on this model, it is shown how the population plasma PK data can be analyzed with a mechanistic modeling approach using the deterministic EM algorithm in combination with the lumping approach by Pilari and Huisinga [2010]. The results were compared to the results of the classical population PK analysis presented in Section 3.3. Finally, PBPK model–based ISF microdialysis concentrations were predicted and compared with microdialysis measurements obtained in muscle and adipose ISF.

5.2. Mechanistic Modeling in Pharmacokinetics: Physiologically Based Pharmacokinetics

In the introduction to PK in Section 1.1.1, the four main processes of PK were introduced: absorption, distribution, metabolism and excretion. In the underlying levofloxacin studies, levofloxacin was administered as an i.v. infusion. Thus, we assumed that there is no absorption process involved and the drug is immediately available in the blood. In Section 3.3 the PK of levofloxacin was characterized by a two compartment model. Metabolism and excretion were summarized in the clearance parameter. PBPK models are based on the same mathematical framework as classical compartmental models and are described by an ODE system. Compared to a classical compartmental model, which in general consists of one to four compartments, in PBPK modeling each tissue and organ is described separately and represented as a separate compartment in the model. These compartments are connected via regional blood flows. Each tissue is further divided into usually three sub–components: the vascular, the interstitial and the cellular space. Based on the drug class and the physiology of the tissue the approaches by Rodgers et al. [2005a,b]; Rodgers and Rowland [2006, 2007] allow to predict the so–called tissue–to-blood partition coefficient. Using these partition–coefficients the drug concentration within each tissue can be predicted. For more information about PBPK modeling, the tutorial by Jones and Rowland-Yeo [2013] offer a comprehensive introduction with further references about the history, development and practical applications of this wide topic.

5.2.1. Whole Body Physiologically Based Pharmacokinetic Modeling

To model the PK of levofloxacin, we used a 13 compartment PBPK model comprising the most important tissues and organs: adipose (*adi*), artery (*art*), bone (*bon*), brain (*bra*), gut (*gut*), heart (*hea*), kidney (*kid*), liver (*liv*), lung (*lun*), muscle (*mus*), skin (*ski*), spleen (*spl*) and vein (*ven*). A schematic presentation of the whole body PBPK model is shown in

Figure 5.2; the corresponding ODE system is given by the following equations

$$
\begin{aligned}
V_{ven}\frac{d}{dt}C_{ven}(t) &= \mathrm{co}\cdot(C_{in,\,ven}(t)-C_{ven}(t))+r(t) \qquad (5.1)\\
V_{lun}\frac{d}{dt}C_{lun}(t) &= f_{lun}^{\mathrm{co}}\cdot\mathrm{co}\cdot\left(C_{ven}(t)-\frac{C_{lun}(t)}{K_{lun:blo}}\right)\\
V_{art}\frac{d}{dt}C_{art}(t) &= \mathrm{co}\cdot\left(\frac{C_{lun}(t)}{K_{lun:blo}}-C_{art}(t)\right)\\
V_{tis}\frac{d}{dt}C_{tis}(t) &= f_{tis}^{\mathrm{co}}\cdot\mathrm{co}\cdot\left(C_{art}(t)-\frac{C_{tis}(t)}{K_{tis:blo}}\right)\\
V_{kid}\frac{d}{dt}C_{kid}(t) &= f_{kid}^{\mathrm{co}}\cdot\mathrm{co}\cdot\left(C_{art}(t)-\frac{C_{kid}(t)}{K_{kid:blo}}\right)-\mathrm{Cl}_{kid,\,int}\cdot C_{kid}(t)\\
V_{liv}\frac{d}{dt}C_{liv}(t) &= f_{liv}^{\mathrm{co}}\cdot\mathrm{co}\cdot\left(C_{in,\,liv}-\frac{C_{liv}(t)}{K_{liv:blo}}\right)-\mathrm{Cl}_{liv,\,int}\cdot C_{liv}(t),
\end{aligned}
$$

with $tis \in \{hea, bra, spl, gut, ski, bon, adi, mus\}$ and

$$
\begin{aligned}
C_{in,\,ven} &= \sum_{tis\in\{hea,bra,spl,gut,ski,bon,adi,mus,liv\}} f_{tis}^{\mathrm{co}}\cdot\frac{C_{tis}(t)}{K_{tis:blo}},\\
C_{in,\,liv} &= \frac{1}{f_{liv}^{\mathrm{co}}}\sum_{tis\in\{art,spl,gut\}} f_{tis}^{\mathrm{co}}\cdot\frac{C_{tis}(t)}{K_{tis:blo}},
\end{aligned}
$$

with the following notations: V_{tis} denotes the physiological volume of the tissue, co the cardiac output, f_{tis}^{co} the fractional regional tissue blood flow expressed as fraction of cardiac output, $K_{tis:blo}$ the tissue–to–blood partition coefficient, $r(t)$ the dose infusion rate and $\mathrm{Cl}_{tis,\,int}$ the intrinsic tissue clearance. According to Rodgers et al. [2005b] the total drug concentration C_{tis} can be interpreted as the "concentration of drug in a tissue outside the blood perfusing it". Obviously, the initial conditions of the above system are given by $C_{ven}(0) = \ldots = C_{liv} = 0$, under the assumption that system is drug free before the first administration.

Only $K_{tis:blo}$ and $\mathrm{Cl}_{tis,\,int}$ are drug specific parameters. The remaining parameters are species specific and are obtained from literature. The sources of the physiological parameters are described in Table 5.1. Because several of the tissue–specific values were not available for humans, these values were fixed to the corresponding animal values.

In the next section, we describe how $K_{tis:blo}$ is determined based on physicochemical properties of levofloxacin.

Tissue Distribution

Levofloxacin belongs to the so–called type I zwitterions (Rodgers et al. [2005b]), i.e. the molecule has one negatively charged functional group, one positively charged functional group and at least one basic ionization constant $pKa \geq 7$ (Rodgers and Rowland [2006]). The tissue distribution is modeled according to Rodgers et al. [2005b], which is based on the assumption that for each tissue the drug exchange between the three subspaces is fast

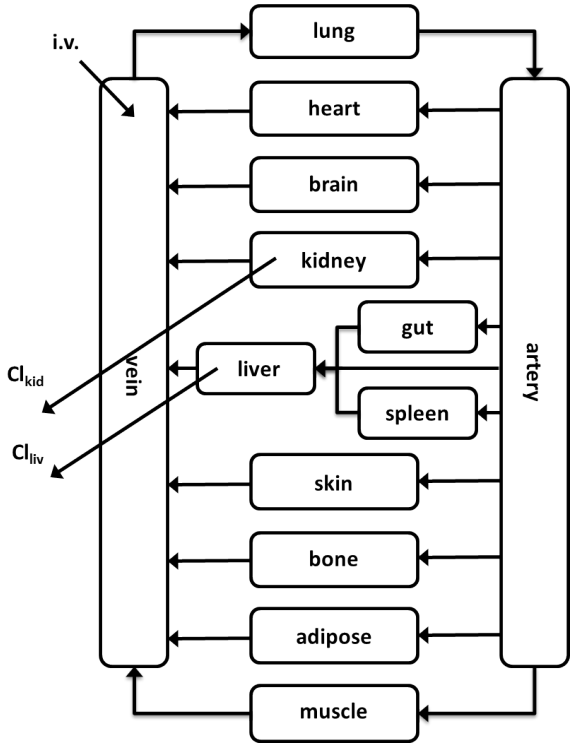

Figure 5.2.: *13 compartment PBPK model comprising the most important organs. The drug is administered i.v., distributed via the blood into the tissues and eliminated via the liver (hepatic elimination) and the kidneys (renal elimination).*

compared to the regional blood flow (perfusion rate limited tissue model, Jones and Rowland-Yeo [2013]). I.e. every concentration change in the vascular space of a tissue will immediately lead to a concentration change in the interstitial and cellular space and vice versa. In the extracellular space—the vascular (*vas*) and interstitial space—levofloxacin binds to the protein albumin and in the cellular space (*cell*) levofloxacin interacts with neutral lipids, neutral phospholipids and has a predominant affinity to acidic phospholipids, see Figure 5.3.

Assuming that the unbound and unionized concentrations in the three subspaces are identical at steady state, Rodgers et al. [2005b] show that the tissue–to–unbound plasma (*up*) partition coefficient can be approximated by

$$\begin{aligned} K^{ss}_{tis:up} &= \frac{C^{ss}_{tis}}{C^{ss}_{up}} \\ &= f^{V_{tis}}_{isf} + f^{V_{tis}}_{vas} + f^{V_{tis}}_{cell} \cdot I_{cell} \\ &\quad + f^{V_{tis}}_{nl} \cdot I_{nl} \cdot K^{tis}_{nl:w} + f^{V_{tis}}_{npl} \cdot I_{npl} \cdot K^{tis}_{npl:w} \\ &\quad + f^{V_{tis}}_{apl} \cdot I_{apl} \cdot Ka_{apl}, \end{aligned} \tag{5.2}$$

with the following notations: $f^{V_{tis}}_{ss/vas/isf/cell/nl/npl/apl}$ denotes the volume of the sub–space as fraction of the total tissue volume, I describes the ionization based on the Henderson–Hasselbalch equation (Rowland and Tozer [2011]) depending on the ionization constants of

Table 5.1.: *Physiological parameter values of the reference PBPK model. For the tissues in brackets reference values for the human were available, all other parameters were either completely available for human, or only rat parameters existed.*

reference values	species	literature source
anthropometric covariates	human	ICRP [2002, chap. 2.3.1.]
organ mass & density	human	ICRP [2002, chap. 2.3.1.], Brown et al. [1997]
blood flow rates	human	ICRP [2002, chap. 2.3.1.]
fractions of interstitial space	rat, human (*adi*)	Kawai et al. [1994]; Engelhardt et al. [1971]
fractions of neutral lipids	human	Poulin and Theil [2009]
fraction of neutral phospholipids	human	Poulin and Theil [2009]
fraction of intra–cellular acidic phospholipids	rat, human (*mus*, *pla*, *ery*)	Poulin and Theil [2009]
fraction of total tissue water	rat, human (*adi*, *mus*, *pla*, *ery*)	Poulin and Theil [2009]; Engelhardt et al. [1971]
interstitial–plasma albumin concentration ratio	rat, human (*adi*, *mus*)	Poulin and Theil [2009]; Ellmerer et al. [2002]
interstitial–plasma lipoprotein concentration ratio	rat	Poulin and Theil [2009]

levofloxacin and the *pH* value in the tissue subspace, *nl* denotes the neutral lipids concentrations, $K_{nl:w}$ denotes the neutral lipids–to–water partition coefficient, which is assumed to be identical for all tissues except adipose, *npl* denotes the neutral phospholipids concentrations and $K_{npl:w}$ denotes the neutral phospholipids–to–water partition coefficient, which is assumed to be identical for all tissues except adipose, *apl* denotes the acidic phospholipids concentrations and Ka_{apl} the association constant to acidic phospholipids. The partition coefficients $K_{nl:w}$ and $K_{npl:w}$ can be derived from the octanol–water partition coefficient $P_{o:w}$, which in general is determined *in vitro* and is a measure for the lipophilicity of a molecule.

According to Rodgers and Rowland [2006], the binding to the *apl* dominates the tissue distribution for type I zwitterions, thus Ka_{apl} represents an important parameter. Unfortunately, these parameters are rarely investigated *in vitro*. However, Rodgers et al. [2005b] considered that human blood cells—mainly erythrocytes (*ery*)—contain acidic phospholipids

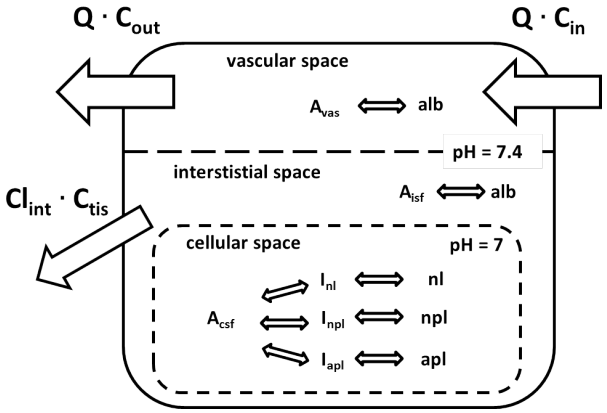

Figure 5.3.: *Tissue distribution of a type I zwitterion according to Rodgers et al. [2005b] adjusted for levofloxacin. In the vascular (vas) and interstitial space (isf) levofloxacin binds to albumin (alb). After ionization in the cellular space (denoted by I), levofloxacin binds to neutral lipids (nl), neutral phospholipids (npl) and acidic phospholipids (apl).*

and do not consist of vascular or interstitial space fluids ($f_{isf}^{V_{ery}} = f_{vas}^{V_{ery}} = 0$). Thus, rearranging Equation (5.2) for $tis = ery$ leads to

$$Ka_{apl} = \frac{K_{ery:up} - f_{cell}^{V_{ery}} I_{cell} - f_{nl}^{V_{ery}} I_{nl} K_{nl:w}^{ery} - f_{npl}^{V_{ery}} I_{npl} K_{npl:w}^{ery}}{f_{apl}^{V_{ery}} I_{apl}}. \tag{5.3}$$

The only unknown parameter value in this equation is $K_{ery:up}$. Considering that blood mainly consists of plasma and erythrocytes, and neglecting the other blood cells components, the drug amount in the erythrocytes can be expressed as the difference of drug amount in blood and in plasma $A_{ery} = A_{blo} - A_{pla}$, where A denotes the drug amount $A = V \cdot C$. Furthermore, for our patients the individual erythrocytes fraction of the total blood, named hematocrit hct, was measured and one has $V_{ery} = \text{hct} \cdot V_{blo}$ and $V_{pla} = (1 - \text{hct}) \cdot V_{blo}$. I.e. the erythrocytes concentration can be expressed by the weighted blood and plasma concentrations

$$C_{ery} = \frac{1}{\text{hct}} C_{blo} - \frac{1 - \text{hct}}{\text{hct}} C_{pla}$$

and the erythrocytes–to–unbound plasma concentration ratio is given by

$$K_{ery:up} := \frac{C_{ery}}{C_{up}} = \frac{K_{blo:pla} - (1 - \text{hct})}{f_{up} \cdot \text{hct}}, \tag{5.4}$$

where $K_{blo:pla} := C_{blo}/C_{pla}$ denotes the blood–to–plasma ratio which can be determined *in vitro*, and f_{up} denotes the unbound fraction of drug in plasma which is not bound to proteins.

In summary, the physicochemical properties of levofloxacin needed for the tissue distribution model introduced above are

- the ionization constant of the positively and the negatively charged functional groups, denoted by pKa_1 and pKa_2, to derive the ionized and unionized fractions of the drug within the different tissue subspaces,
- the octanol–water partition coefficient $P_{o:w}$, usually reported as $\log P_{o:w}$, to derive the binding to the neutral lipids and neutral phospholipids,
- the blood–to–plasma ratio $K_{blo:pla}$ as a surrogate for the affinity to the acidic phospholipids—this parameter is species dependent, compared to the previous properties.

In the next step the patient's individual protein binding is derived.

Protein Binding

The fraction of unbound drug in plasma is defined as the ratio of the unbound and the total concentration $f_{up} := C_{up}/C_{pla}$, where the index u denotes the unbound fraction. The dominating protein binding of levofloxacin in plasma and interstitial space is to the protein albumin (*alb*) (see Fish and Chow [1997]; Sun and He [2008])

$$\text{levofloxacin} + \text{albumin} \leftrightarrows \text{levofloxacin:albumin}.$$

Based on the levofloxacin–albumin affinity constant Ka_{alb} determined by Sun and He [2008] and assuming a linear protein binding the fraction unbound in plasma can be approximated by

$$f_{up} = \frac{1}{1 + Ka_{alb}\, C_{alb}}, \tag{5.5}$$

with albumin plasma concentration C_{alb} (Rowland and Tozer [2011, chap. 4]).

In summary, based on the physicochemical properties of the drug $K_{tis:up}$ is determined. Using f_{up} and $K_{blo:pla}$ the tissue–to–unbound plasma partition coefficient $K_{tis:up}$ defined in Equation (5.2) can easily be transformed into the tissue–to–blood and tissue–to–plasma partition coefficient by

$$K_{tis:blo} = f_{up}\,\frac{K_{tis:up}}{K_{blo:pla}} \text{ and } K_{tis:pla} = f_{up}\,K_{tis:up}. \tag{5.6}$$

The last process to finally predict the tissue concentrations is the drug elimination from the tissue, see Figure 5.3 which is described in the next section.

Clearance

The present work did not consider an *a priori* approach to predict the clearance based on *in vitro* data as introduced for the drug distribution. Rather, the total plasma clearance was estimated based on the plasma PK data. The plasma clearance was transformed into an intrinsic tissue clearance for the integration into the PBPK model (5.1). To this end, it was assumed that levofloxacin is mainly cleared by filtration and re–absorption in the kidneys ($\sim 80\%$) and only to a small amount metabolized in the liver ($\leq 5\%$) (Fish and Chow [1997]). Thus, we assumed that total clearance is composed of $f_{kid}^{\text{Cl}} = 80\%$ renal

(kidney) clearance and $f_{liv}^{\mathrm{Cl}} = 5\%$ hepatic (liver) clearance. As the remaining 15% could not be allocated to a specific organ, we assumed $f_{ven}^{\mathrm{Cl}} = 15\%$ to be cleared form the vein.

To transform the estimated total plasma clearance into an intrinsic organ clearance, we consider that under steady state conditions, that is, $d/dt\, C_{tis} = 0$ in model (5.1) one has

$$\begin{aligned} 0 &= Q_{tis} \cdot C_{in} - (Q_{tis} + K_{tis:blo} \cdot \mathrm{Cl}_{tis,\,int}) \cdot (C_{tis}/K_{tis:blo}), \\ \Leftrightarrow \underbrace{\frac{C_{in} - C_{tis}/K_{tis:blo}}{C_{in}}}_{=:E_{tis}} &= \frac{K_{tis:blo} \cdot \mathrm{Cl}_{tis,\,int}}{Q_{tis} + K_{tis:blo} \cdot \mathrm{Cl}_{tis,int}} \end{aligned} \tag{5.7}$$

where C_{in} is the inflowing concentration into the tissue according to Equation (5.1) and the term $C_{tis}/K_{tis:blo}$ represents the out–flowing tissue concentration. The so–called tissue extraction ratio E_{tis} is defined as the fraction of drug removed from the tissue during passage through the tissue. For organ blood clearance we have $\mathrm{Cl}_{tis,\,blo} = f_{tis}^{\mathrm{Cl}} \cdot \mathrm{Cl}_{blo} = \mathrm{Q}_{tis} \cdot E_{tis}$, with total blood clearance $\mathrm{Cl}_{blo} = \mathrm{Cl}_{pla}\,/K_{blo:pla}$ and Q_{tis} the tissue blood flow.

Thus, the intrinsic clearance as function of the blood clearance is given by

$$\mathrm{Cl}_{tis,\,int} = \frac{f_{tis}^{\mathrm{Cl}}\ \mathrm{Cl}_{blo}}{K_{tis:blo}\,(1 - E_{tis})} = \frac{f_{tis}^{\mathrm{Cl}}\ \mathrm{Cl}_{blo}}{\hat{K}_{blo:tis}}, \tag{5.8}$$

with $\hat{K}_{blo:tis} := K_{tis:blo}\,(1 - E_{tis})$ for $tis \in \{kid, liv\}$. This model is called the well–stirred tissue model and is widely used in mechanistic PK modeling; a detailed motivation and derivation of this model can be found in Rowland and Tozer [2011, chap. 2 and appendix].

Now, all processes relevant for the tissue–distribution are described. Assuming the concentration ratios to be constant $K_{tis:blo} = K_{tis:blo}^{ss}$ over time, blood and tissue concentration–time courses can predicted.

In the next section these processes are individualized based on the physiological information of the patients.

5.2.2. Mechanistic Modeling of Physiological Inter–Individual Variability

Based on the individually measured hematocrit and plasma albumin concentration and Equations (5.4) and (5.5) the erythrocytes–to–plasma partition coefficient and the protein binding are individualized, respectively.

For patients without measured hematocrit value (18 out of 24), the missing covariate was replaced by the literature reference stratified on sex—43% for male and 38% for female patients (ICRP [2002, chap. 2.3.1.]). For patients without measured plasma albumin concentration (8 out of 24), the missing covariate was replaced by the median of the available albumin concentrations stratified on sex—40 g/L for male and 46.7 g/L for female patients. This resulted in a bound fraction $(1 - f_{up})$ of $20.6 \pm 4.2\,\%$ (mean $\pm$ standard deviation), which is in good agreement with the results obtained by Neckel et al. [2002] for levofloxacin $(21.9 \pm 3.2\,\%)$.

Obviously, also the body composition differs between individuals, e.g. the tissue volumes and blood flow rates. There are approaches published to account for these differences in PBPK modeling, e.g. see Price et al. [2003]; Willmann et al. [2007]. However, these approaches either only poorly reflect experimentally observed physiological BSV, or they only

achieve this by using randomly sampled physiological parameters according to the experimental observed variability. The latter means that unexplained variability is integrated in the model. In contrast, our focus was to link variations in the anatomical and physiological parameters to covariates, which we denote as mechanistically derived BSV. As a consequence, we developed our own approach (Huisinga et al. [2012]). The idea was to collect published information about the relationship between anthropometric properties, like body weight or height, and physiological parameters used in PBPK modeling.

The finding was that there seems to be a lack of literature which investigates the individual body composition to such a detail we would need. In general, several publications focus on only one tissue and often statements about the BSV are missing. A relatively rich source is given by an autopsy study conducted by de la Grandmaison et al. [2001]. They provide the mean values and BSVs for the organ weights of heart, lung, liver, spleen, pancreas, kidney and thyroid gland. Furthermore, they observed a correlation between organ weight and body height and body–mass–index (BMI). The basis of this conclusion was a linear regression where only the coefficient of determinations were provided. Unfortunately, the authors did not provide the details of the analysis upon a request, nor the original data to further investigate those dependencies. However, their results were very useful because they showed that (i) there is link between anthropometry and organ weights and (ii) they provided statements about the BSV of the organ weights.

For each individual a PBPK parameter database x_i is created containing the individualized anatomical and physiological parameters based on the patients age, sex and covariates like body weight (BW), body height (BH), BMI ($BMI[\mathrm{kg/m^2}] = BW[kg]/(BH[\mathrm{m}])^2$), and body–surface–area (BSA) $BSA[\mathrm{m^2}] = \sqrt{BH[\mathrm{m}] \cdot BW[\mathrm{kg}]/(36[\mathrm{kg/m^3}])}$ using the following approach:

1. Definition of reference female ref_{female} and male subject ref_{male} based on literature data, see Table 5.1); e.g. the reference subject of a Caucasian woman (≥ 18 years) has a body weight of $BW^{ref,female} = 60\,\mathrm{kg}$, a body height of $BH^{ref,female} = 1.63\,\mathrm{m}$ and for example a blood volume of $V_{blo}^{ref,female} = 3.9\,\mathrm{L}$. Because all patients in the underlying study were adult Caucasians, we only stratified for sex.

2. Because it is known that cardiac output is correlated with age (Luisada et al. [1980]), the reference population is further stratified by age classes 18–39, 40–59, 60–69, 70–79, 80–89 years. In Luisada et al. [1980] the respective means and variability for cardiac output are provided.

3. Based on the individual covariates the adipose organ weight W is approximated by $W_{adi}[\mathrm{kg}] = BW[\mathrm{kg}] - LBW[\mathrm{kg}]$, with lean–body–weight (LBW). If LBW was not measured—as in the underlying study—the following approximations by Janmahasatian et al. [2005] were used

$$LBW[\mathrm{kg}] = \frac{9270 \cdot BW[\mathrm{kg}]}{8780 + 244[\mathrm{m^2/kg}] \cdot BMI[\mathrm{kg/m^2}]}\,, \text{ for female,}$$
$$LBW[\mathrm{kg}] = \frac{9270 \cdot BW[\mathrm{kg}]}{6680 + 216[\mathrm{m^2/kg}] \cdot BMI[\mathrm{kg/m^2}]}\,, \text{ for male.} \qquad (5.9)$$

   The organ weight can be transformed into a volume by using the organ's density $\mathrm{V}_{adi}[\mathrm{L}] = d_{adi}[\mathrm{L/kg}] \cdot W_{adi}[\mathrm{kg}]$, with $d_{adi} = 0.92\mathrm{L/kg}$ (see Brown et al. [1997]).

4. Based on ICRP [2002, chap. 3.3.1.] it was assumed that the brain volume is constant $V_{bra} = V_{bra}^{ref}$ within the reference population, and that the organ density of $d_{bra} = 1$ (see Brown et al. [1997]).

5. For the skin it was considered that the volume is equal to the skin surface area (BSA) times the skin thickness Δ_{ski}. According to Price et al. [2003] it was assumed that the thickness is constant over the body and within a reference population. Thus, one has $V_{ski} = BSA \cdot \Delta_{ski}^{ref}$ and an organ density of $d_{ski} = 1$ (see ICRP [2002]).

6. The remaining individual body weight was defined as $BW_{rem} := BW - W_{adi} - W_{bra} - W_{ski}$. The remaining body weight of the reference female/male subject was defined as $BW_{rem}^{ref} := BW^{ref,female/male} - W_{adi}^{ref,female/male} - W_{bra}^{ref,female/male} - W_{ski}^{ref,female/male}$, respectively. For the remaining tissues we assumed the proportion of the tissue volumes w.r.t. remaining body weight BW_{rem} to be identical to these proportions of the reference subject: $W_{tis}/BW_{rem} = W_{tis}^{ref,female/male}/BW_{rem}^{ref,female/male}$ for $tis \in \{art, gut, hea, kid, liv, lun, bon, mus, spl, ven\}$. For example, the amount of muscle tissue would be 47 % and 54 % of BW_{rem} for all female and male patients, respectively.

7. According to the description above individual values for the hematocrit hct and the fraction unbound f_{up} are included in the indvidual PBPK parameter database x_i as well.

The approach was qualified based on the distribution properties provided by de la Grandmaison et al. [2001]. In comparison to the existing approaches, the herein proposed *LBW scaling approach* predicts more realistic variations of the organ weights as the approaches by Willmann et al. [2007]—based on body height scaling—and by Price et al. [2003]—based on regression using different covariates. Our approach was further supported by the data published by Nyengaard and Bendtsen [1992]. The dataset contained the individual kidney weight and the covariates sex, body height and body weight of 18 adult males and 18 adult females. In Figure 5.4 the correlation with the LBW is shown; the correlation was higher compared to the other anthropometric variables body height, body weight or BMI.

Although there are more sources of physiological BSV, e.g. in the tissue composition (see White et al. [1991]), only sources which could be related to measurable covariates were included in this approach.

5.2.3. Lumping of PBPK Models

In this section we introduce the so–called lumping approach. The idea of lumping is to lower the dimension of a PBPK model and to establish a link to the lower dimensional classical compartmental models while maintaining the mechanistic basis of a PBPK model. As a consequence, the computational effort to solve the associated ODE system numerically is reduced and parameter estimation is faster and more stable.

Pilari and Huisinga [2010] derived a method where they distinguish between two kinds of lumped model, the first—denominated as minimal lumped model—is defined as the model with as few compartments as possible, where the ability to describe the blood and/or plasma concentration is maintained; the second—denominated as mechanistical lumped model—is defined as the model with as few compartments as possible, where for all tissues the tissue

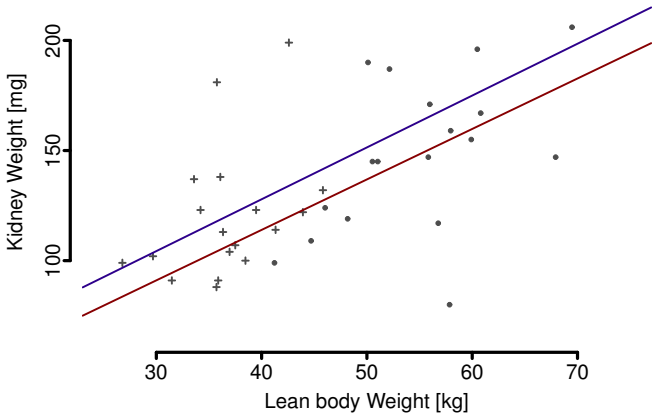

Figure 5.4.: *Kidney weight vs. lean body weight: Crosses represent the female subjects and dots represent the male subjects. Solid lines represent the linear regression of kidney weight as function of lean body weight showing an correlation (Pearson's correlation coefficient) of* 0.38 *for female (red) and* 0.49 *for male (blue) subjects, respectively.*

concentrations can be derived from the lumped model. Both approaches exploit "the fact that drug concentrations in different compartments of the whole–body PBPK model are often strongly kinetically dependent on each other" (Pilari and Huisinga [2010]).

Until now, there is no mathematical criterion established how the lumping and the number of lumped compartments k has to be chosen for a minimal or mechanistical lumped model. In a lumped model the central compartment L_1 contains the blood tissues, while the remaining tissues are assigned to one of the k compartments

$$\begin{aligned} L_1 &= \{art, ven, \ldots, tis_{1,n_1}\},\ n_1 \geq 0, \\ L_2 &= \{tis_{2,1}, \ldots, tis_{2,n_2}\},\ n_2 > 0, \\ &\vdots \\ L_k &= \{tis_{k,1}, \ldots, tis_{k,n_k}\},\ n_k > 0, \end{aligned}$$

and $1 \leq k \leq 13$. Using the approach by Pilari and Huisinga [2010], tissues are lumped together, if the normalized concentration profiles $C_{tis}/\hat{K}_{tis:blo}$ are similar. Thereby, the numbers of minimal or mechanistically lumped compartments, as well as lumping itself is defined based on visual inspection of the normalized profiles. Note, the composition of the lumped compartments is not unique and is subjective to some extend. However, Pilari and Huisinga [2010] obtained that in general the minimal lumped model consists of one to three compartments—as in classical compartment modeling—and the mechanistical lumped model in general consists of four compartments; thereby the central compartment usually comprises well perfused tissues like lung, kidney and heart, while the peripheral compartment comprises the poorly perfused tissues like adipose, muscle and bone.

If $k > 1$, the structure of the lumped model is defined in the way that the central compartment is directly connected to all peripheral compartments, while the peripheral compartments are not connected among each other. The elimination is not restricted to a certain

lumped compartment. Although, usually it is considered from the central compartment using the blood clearance—as usually done in classical compartment modeling.

For the perfusion–rate limited model, the corresponding ODE system is given by

$$\begin{aligned}
\mathrm{V}_{L_1} \frac{d}{dt} C_{L_1}(t) &= \sum_{j=2}^{k} \mathrm{Q}_{L_j} \left(C_{L_j}(t) - C_{L_1}(t) \right) - \mathrm{Cl}_{L_1} \, C_{L_1}(t) + r(t) \\
\mathrm{V}_{L_2} \frac{d}{dt} C_{L_2}(t) &= \mathrm{Q}_{L_2} \, (C_{L_1}(t) - C_{L_2}(t)) - \mathrm{Cl}_{L_2} \, C_{L_2}(t) \\
&\vdots \\
\mathrm{V}_{L_k} \frac{d}{dt} C_{L_k}(t) &= \mathrm{Q}_{L_k} \, (C_{L_1}(t) - C_{L_k}(t)) - \mathrm{Cl}_{L_k} \, C_{L_k}(t),
\end{aligned}$$

with $C_{L_1}(0) = \ldots = C_{L_k}(0) = 0$. The apparent PK parameters are given by

- the lumped volumes of distribution

$$V_{L_j} := \sum_{tis \in L_j} V_{tis} \, \hat{K}_{tis:blo}, \text{ for } 1 \leq j \leq k, \tag{5.10}$$

  where $\hat{K}_{tis:blo} = (1 - E_{tis}) \cdot K_{tis:blo}$ denotes the tissue–to–blood partition coefficient corrected by the amount of drug removed from the tissue, presented by the extraction ratio E_{tis},

- for $k > 1$ the inter-compartmental clearance(s)

$$\mathrm{Q}_{L_j} := \mathrm{co} \cdot \sum_{tis \in L_j} f_{tis}^{\mathrm{co}}, \text{ for } 2 \leq j \leq k, \tag{5.11}$$

- and the clearances

$$\mathrm{Cl}_{L_j} = \sum_{tis \in L_j} \mathrm{Cl}_{tis,\, blo}, \text{ for } 1 \leq j \leq k. \tag{5.12}$$

The concentrations in the lumped compartments are defined as the weighted sum of the tissue concentrations, see Pilari and Huisinga [2010]:

$$C_{L_j}(t) = \sum_{tis \in L_j} \frac{V_{tis}}{V_{L_j}} C_{tis}(t), \text{ for } 1 \leq j \leq k.$$

Furthermore, Pilari and Huisinga [2010] show that the lumped concentrations can be transformed back into the original tissue concentrations by correcting with the tissue–to–blood partition coefficient

$$C_{tis}(t) \approx \frac{\hat{K}_{tis:blo}}{K_{L_j}} \cdot C_{L_j}(t), \text{ with } tis \in L_j, \tag{5.13}$$

with $K_{L_j} := \sum_{tis \in L_j} \frac{V_{tis}}{V_{L_j}} \cdot \hat{K}_{tis:blo}$.

To illustrate the lumping approach at this point, the parameters of the PBPK model which are derived and presented in the following sections are used for the simulations.

Based on the visual inspection of the partition–coefficient normalized concentrations time profiles shown in Figure 5.5, we choose $k = 2$ compartments for the minimal lumped model of levofloxacin and the lumping as depicted in the schematic representation in Figure 5.6. The normalized tissue concentrations advise that the tissues assigned to the central compartment are more or less identical, while in the peripheral compartment the normalized muscle concentrations differ from the other peripheral tissues. Based on this, we infer that the chosen lumping model might be appropriate as a minimal lumped model to successfully describe the venous plasma concentration, but it will not be acceptable as a mechanistical lumped model where all tissues concentrations should be reproducible. As shown in Figure 5.7, the chosen minimal lumped model adequately describes the vein concentration of levofloxacin.

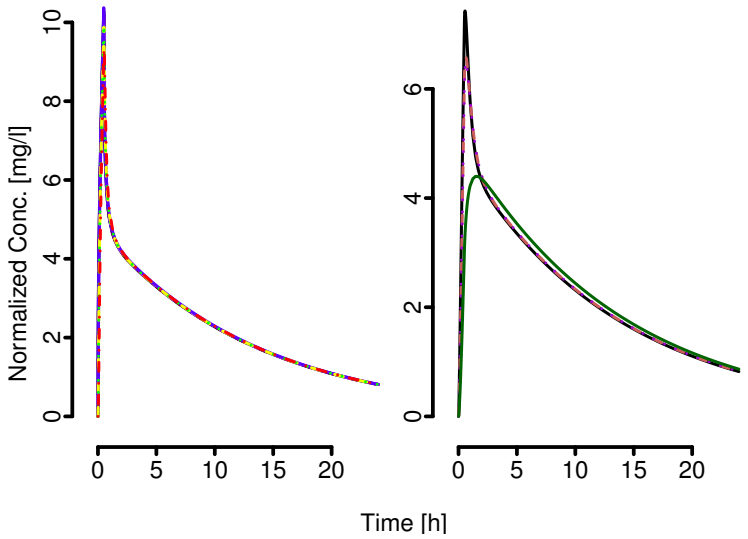

Figure 5.5.: *Tissue partition–coefficient normalized tissue concentrations of the full PBPK model. Left: the tissues assigned to the central compartment. Right: the tissues assigned to the peripheral compartment. In the left plot no distinction between the single tissues is possible, while in the right plot the normalized profiles of the skin, bone, and adipose are similar, but the profile of the muscle (green) significantly deviates from the others.*

The normalized profiles indicate that a mechanistical lumped model is given by splitting the peripheral compartment into one compartment comprising the skin tissue, one comprising the bone and the adipose tissue, and one comprising the muscle tissue. The results of such a model are shown in Figure 5.8. The figures illustrate that the lumping methodology is able to reproduce the same information provided by a 13 compartment model with only four compartments, considering that the kinetics between tissues might be very similar. Additionally, it also gives a hint about the maximal identifiable compartments in a classical analysis; as example, based on plasma data two compartments at most would be identifiable for levofloxacin.

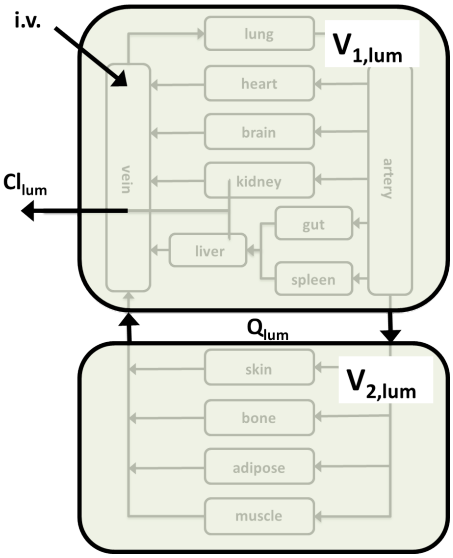

Figure 5.6.: *Schematic presentation of the minimal lumped two compartment model. The central compartment comprises the well perfused tissue artery, brain, gut, heart, kidney, liver, lung, spleen and vein and the peripheral compartment comprises the poorly perfused tissues adipose, bone, muscle and skin.*

5.3. Results

5.3.1. Population Analysis of Levofloxacin Plasma Data Based on a Minimal Lumped PBPK Model

In this section, the developed PBPK model, the mechanistically integrated BSV and the lumping method are combined with the EM algorithm to estimate the missing drug–related parameters and refine important physiological parameters. As mentioned above, the physicochemical parameters of levofloxacin which are needed for PBPK modeling are the ionization constants pKa_1 and pKa_2, the octanol–water partition coefficient $\log P_{o:w}$ and the blood–to–plasma ratio $K_{blo:pla}$. Additionally, the total blood clearance Cl_{blo} is needed. A literature review revealed several sources for the pKa and the $\log P_{o:w}$ values, see Table 5.2. There was an inter–publication variation, complicating the choice of those values. Besides, we had no findings about the human blood–to–plasma ratio for levofloxacin. Because of its importance it was decided to estimate this parameter based on the plasma data. The idea to use minimal lumped models to estimate missing parameters is not new, e.g. see Cao and Jusko [2012]. However, until now this has not been put into a population context.

For other substances an intra- and inter–subject variability of $K_{blo:pla}$ was reported, e.g. see Høiseth et al. [2009] and Schwilke et al. [2009]. Furthermore, $K_{blo:pla}$ is an important parameter to derive the cellular distribution, see Equations (5.3) and (5.4), where BSV is very likely to occur. For example, White et al. [1991] compiled data about the composition of the adipose tissues, and observed that the percentage of tissue water varies around CV% 45, the percentage of lipids around CV% 18 and the percentage of proteins around CV% 80 in adults. Using $K_{blo:pla}$ as a surrogate for these processes it is reasonable to expect BSV on

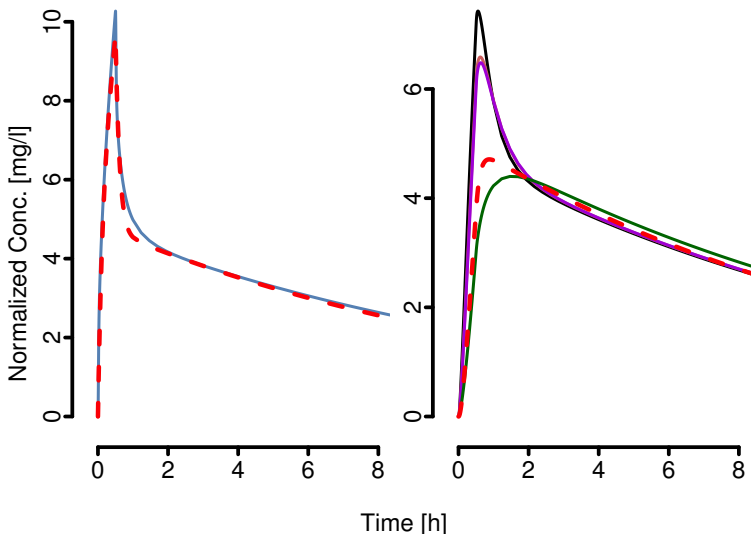

Figure 5.7.: *Tissue partition–coefficient normalized tissue concentrations of the full PBPK model compared with the minimal lumped concentrations (dashed red lines) for the first 8 hours. For the central compartment (left) only the vein (blue) and the lumped concentration are displayed, indicating a good agreement between both. The right plot confirms the observation in Figure 5.5 that additional compartments are necessary to reproduce the muscle tissue successfully. As already indicated by the distribution half life, the plot confirms that there are also deviations between the skin (black) and the bone (rose) and adipose (violet). The right plot nicely illustrates how the lumped concentrations form a mixture of the normalized tissue concentrations.*

$K_{blo:pla}$.

To find the most appropriate combination of the physicochemical parameters, different values for pKa and $\log P_{o:w}$ were tested in combination with the estimation of $K_{blo:pla}$ and the associated BSV.

For the integrated lumped PBPK and NLME analysis of the plasma PK data the following approach was developed. Via the lumping the structural model is defined; as described in Section 5.2.3 for levofloxacin the minimal lumped model is given by a two compartment model. As in our situation, usually not for all parameter of the PBPK model values are available. However, the model has to be fully parameterized to define the structural model via the lumping. Therefore, for unknown parameter values different values should be tested and it should be examined whether the resulting minimal lumped models differ in the number of compartments and the composition of the lumped compartments. Furthermore, subsequently it should be tested if the structural model used for the parameter estimation is identical to the minimal lumped model resulting using the final parameter estimates.

The structural model is parametrized by $\log \mathrm{Cl}$, $\log K_{blo:pla}$ and a correction factor Δ_{lco} to adjust the prior information of the cardiac output co. The apparent PK parameters $\mathrm{V}_1(x_i, \Delta_{\mathrm{lco}}, \log \mathrm{Cl}, \log K_{blo:pla})$, $\mathrm{V}_2(x_i, \Delta_{\mathrm{lco}}, \log \mathrm{Cl}, \log K_{blo:pla})$ and $\mathrm{Q}(x_i, \Delta_{\mathrm{lco}})$ depend on the individualized physiological data x_i which was defined in the previous Section 5.2.2 and which is needed to parameterize a PBPK model x_i, see for example Table 5.1. Therefore, in a first step for each individual the independent variables x_i is generated based on a

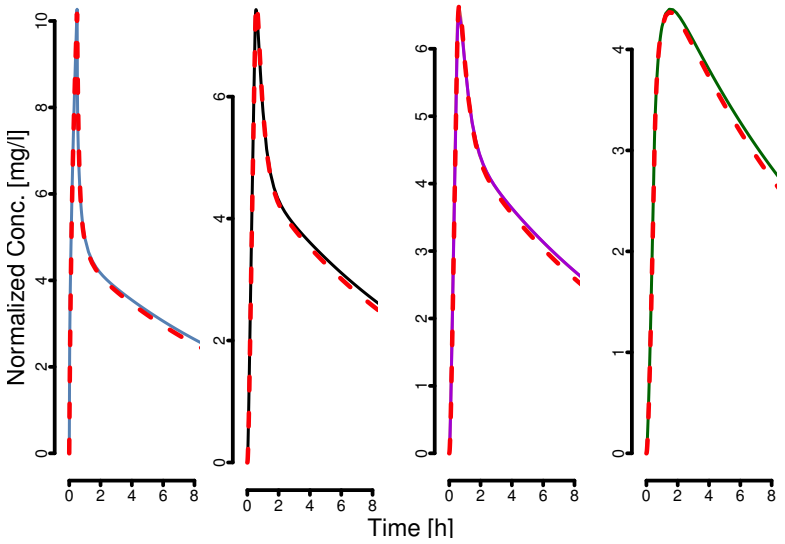

Figure 5.8.: *Normalized concentration of the full PBPK model compared with the mechanistical lumped concentrations (dashed red lines) for the first 8 hours. For the central compartment (left) only the vein (blue) and the lumped concentration are displayed. The second plot shows the normalized skin tissue concentration (black). The third plot shows the normalized bone (rose) and adipose (violet) concentrations. The last plot shows the normalized muscle (green) concentrations. The predictions of the lumped model are in excellent agreement with the normalized tissue concentrations for all tissues.*

comprehensive physiology database and the individual covariates according to the description in Section 5.2.2. Given x_i and values for $\log \mathrm{Cl}$, $\log K_{blo:pla}$ and Δ_{lco} the apparent parameters are derived using the following sequential approach:

(1) Calculate cardiac output $\mathrm{co}(x_i, \Delta_{\mathrm{lco}}) = \mathrm{co}^{\mathrm{prior}}(x_i)\, e^{\Delta_{\mathrm{lco}}}$ and update tissue blood flows accordingly $\mathrm{Q}_{tis}(x_i, \Delta_{\mathrm{lco}}) = f_{tis}^{\mathrm{co}} \cdot \mathrm{co}(x_i, \Delta_{\mathrm{lco}})$ for all tissues.

(2) Translate plasma clearance Cl into the blood clearance

$$\mathrm{Cl}_{blo}(\log \mathrm{Cl}, \log K_{blo:pla}) = \exp(\log \mathrm{Cl} - \log K_{blo:pla}).$$

(3) Using (1) and (2) the extraction ratio of the eliminating organs is determined

$$E_{tis}(x_i, \Delta_{\mathrm{lco}}, \log \mathrm{Cl}, \log K_{blo:pla}) = \frac{f_{tis}^{\mathrm{Cl}}\ \mathrm{Cl}_{blo}(\log \mathrm{Cl}, \log K_{blo:pla})}{Q_{tis}(x_i, \Delta_{\mathrm{lco}}}.$$

(4) Calculate partition–coefficients $K_{ery:up}(x_i, \log K_{blo:pla})$ and $K_{tis:pla}(x_i, \Delta_{\mathrm{lco}}, \log \mathrm{Cl}, \log K_{blo:pla})$ for all tissues according to Equations (5.4) and (5.6).

(5) Calculate lumped PK parameters:

$$\mathrm{V}_1(x_i, \Delta_{\mathrm{lco}}, \log \mathrm{Cl}, \log K_{blo:pla}) = \sum_{tis \in cen} V_{tis}(x_i)\, \hat{K}_{tis:pla}(x_i, \Delta_{\mathrm{lco}}, \log \mathrm{Cl}, \log K_{blo:pla}),$$

$$\mathrm{V}_2(x_i, \Delta_{\mathrm{lco}}, \log \mathrm{Cl}, \log K_{blo:pla}) = \sum_{tis \in per} V_{tis}(x_i)\, \hat{K}_{tis:pla}(x_i, \Delta_{\mathrm{lco}}, \log \mathrm{Cl}, \log K_{blo:pla}),$$

$$\mathrm{Q}(x_i, \Delta_{\mathrm{lco}}) = \sum_{tis \in per} Q_{tis}(x_i, \Delta_{\mathrm{lco}})$$

where $\hat{K}_{tis:pla} = (1 - E_{tis})\, K_{tis:pla}$ and the central compartment *cen* comprises the artery, vein, lung, heart, kidney, spleen, gut, liver, brain and the peripheral compartment *per* comprises the skin, bone, adipose and muscle tissue, using the individual scaled tissue volumes.

Thus, the apparent parameters are dynamic functions of the parameters $\log \mathrm{Cl}$, $\log K_{blo:pla}$ and Δ_{lco}. Since the approach requires several intermediate computations (like the determination of the $K_{tis:pla}(x_i, \Delta_{\mathrm{lco}}, \log \mathrm{Cl}, \log K_{blo:pla})$ relying on detailed physiological data), the implementation in standard population PK software tools on simple script structures, like NONMEM, is very cumbersome. Particularly, in case more complex PBPK and/or tissue–distribution model would be used.

We included random effects on the clearance, the blood–to–plasma ratio and the cardiac output. Thus, the random effects model is given by

$$\boldsymbol{\Theta_i} \sim \mathcal{N}_3\left(\theta, \Omega\right),$$

with $\theta \in \mathbb{R}^3$ and $\Omega \in \mathbb{R}^{3\times 3}_{>0}$, where the off–diagonals $\Omega_{1,3} = \Omega_{3,1} = \Omega_{2,3} = \Omega_{3,2} = 0$ and $\Omega_{3,3} = \omega^2_{\mathrm{co}}$, where ω^2_{co} represents variance of the BSV of the cardiac output which was fixed to a literature value (Luisada et al. [1980]).

As in Section 3.3, the log–transformed measurements ($\log \boldsymbol{y_i}$) are assumed to be realizations of $\boldsymbol{Y_i}$ given by

$$\boldsymbol{Y_i} = \log f(x_i; \boldsymbol{\Theta_i}) + \boldsymbol{\epsilon_i},\ \boldsymbol{\epsilon_i} \sim \mathcal{N}_{n_i}\left(0, I_{n_i}\sigma^2\right),$$

with $\sigma > 0$ and as convergence criterion of the deterministic EM algorithm three significant digits in the fixed effects were chosen.

Table 5.2 shows the minimum of the OFV after a successful completion of the EM algorithm for several combinations of the physicochemical parameter values. For most of the combinations of $\log P_{o:w}$ and *pKa* no differences in the ability to describe the underlying plasma data—represented by the OFV—are observable. The parameter estimate of the clearance is independent of the drug distribution and thus independent of the physicochemical properties. Also the estimate of Δ_{lco} was not affected by the choice of *pKa* and $\log P_{o:w}$. However, the estimate of $K_{blo:pla}$ varied for the different combinations, indicating a correlation between these parameters. In conclusion, this means one cannot distinguish between the contribution of each physicochemical parameter to the tissue distribution process based on plasma data.

Table 5.2.: *OFV value of the minimal lumped two compartment model based on different combinations of physicochemical properties found for levofloxacin in literature. The OFV indicates that the physicochemical parameters have an impact, but only if outside a certain range, e.g.* $-0.62 \leq \log P_{o:w} \leq 0.63$.

$\log P_{o:w}$	pKa (5.7, 7.9) [1987]	(5.5, 8.0) [2011]	(6.05, 8.22) [2007]
0.63 [2012]	−227.4	−227.5	−225.2
−0.62 [2000]	−228.0	−228.0	−226.1
1.49 [2005]	−183.3	−175.2	648.1
−0.42 [1998]	−228.0	−228.0	−226.1

Thus, as the final model we chose a combination with the lowest OFV: $pKa = (5.7, 7.9)$ (Kitzes-Cohen [1987]) and $\log P_{o:w} = -0.62$ (Tamai et al. [2000]). The results of this model are displayed in Table 5.3. If the results should be translated back into the full PBPK model, the procedure described above can simply be executed backwards.

Table 5.3.: *Results of the minimal lumped two compartment model: The typical values for the population parameters are transformed to the original scale (e.g.* $\mathrm{Cl} = e^{\log \mathrm{Cl}}$*), BSV and WSV are expressed as CV (e.g.* $\mathrm{CV\%}(\mathrm{Cl}) = 100 \cdot \sqrt{e^{\omega^2_{\mathrm{Cl}}} - 1}$*).*

	Cl	$K_{blo:pla}$	Δ_{lco}
Typical value	8.2 L/h	0.98	−0.2 log(L/min)
BSV [CV%]	26.3 %	6.3 %	fixed to 23 %
η–shrinkage	2.0 %	3.1 %	−4.8 %
WSV [CV%]	12.3 %		
ϵ–shrinkage	14.2 %		
OFV	−228.0		

The GOF plots in Figure 5.9 shows that the models adequately describes the data. Also the VPC in Figure 5.10 shows good agreement between simulations and measurements w.r.t. the typical profile as well as the observed variability. For all time points median, 10^{th} and 90^{th} quantile of the observations are included in the respective prediction interval of the simulated quantiles. This indicates a good agreement of the estimated typical profile and the median observations, as well as for the estimated and observed variability. Howsoever, for $t \geq 12$ h the prediction interval increases due to the low number of samples within each bin (in average 6), which makes it difficult to assess whether the variability in the later phase is correctly captured or slightly overestimated by the model.

Except for Δ_{lco}, the shrinkage values are in an acceptable order of magnitude indicating that the individual data is sufficient to determine EBEs for Cl and $K_{blo:pla}$. The negative η–shrinkage for Δ_{lco} indicates an under–prediction of the corresponding BSV. In this case the BSV was fixed to a literature prior, and increasing the BSV would theoretically result in

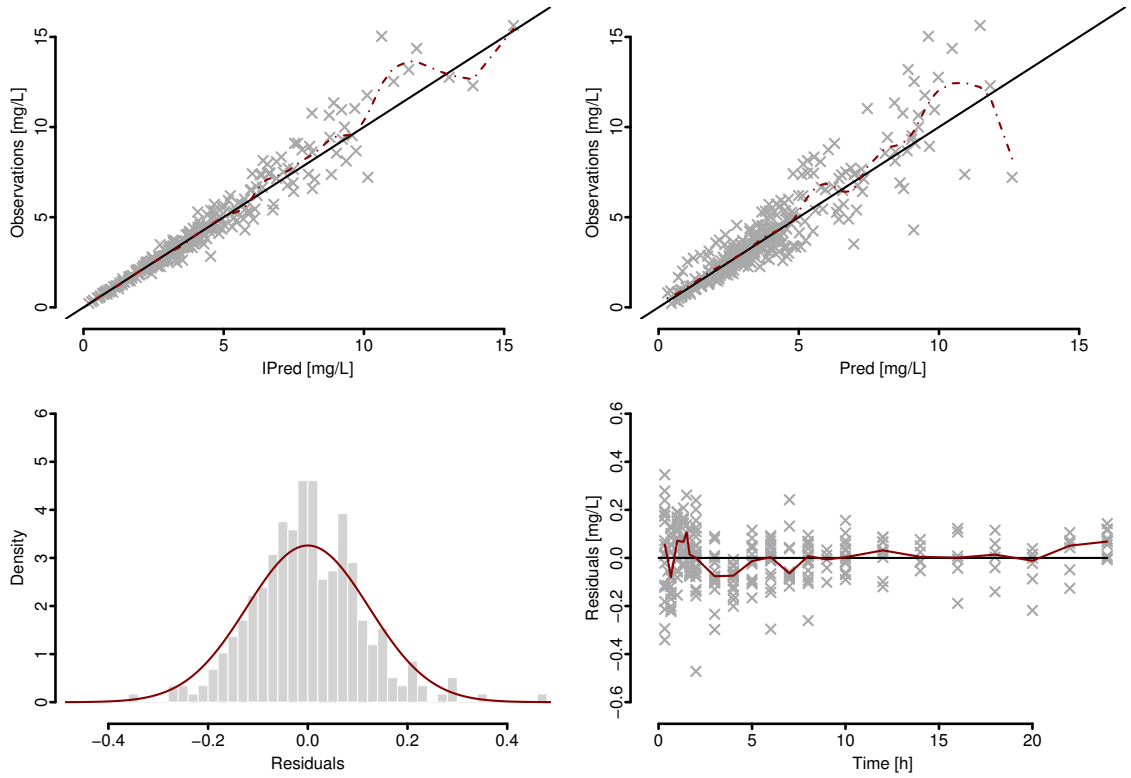

Figure 5.9.: *Upper left plot shows observations against the individual predictions (IPred), upper right plot shows observations against the predictions based on the population mean $\hat{\theta}$ (Pred). In the upper graphs, solid line represents the identity line and red dashed line represent a smoothing spline. Lower left plot shows the empirical distribution of the residuals, the red line indicates the estimated density of the WSV. Lower right plots shows the residuals over time, red line shows the median of the residuals for each time point.*

unphysiological values of the cardiac output. The negative η–shrinkage could indicate that BSV in tissue blood flow is not appropriately described by BSV in cardiac output. However, because our analysis was only based on 24 patients we thought it is worth trusting the well studied literature value and accepted the negative η–shrinkage.

The typical value of $\Delta_{\text{lco}} = -0.2$ corresponds to a correction factor of the cardiac output of by $-18\,\%$ on average.

As mentioned above, it was not possible to find information about the human $K_{blo:pla}$ for levofloxacin. However, Rodgers and Rowland [2006] reported the blood–to–plasma ratio for the drug ofloxacin to be 0.92. It is reasonable to assume that both compounds have similar physicochemical properties, because levofloxacin is the enantiomer of ofloxacin, i.e. the dominating pharmacologically active moiety of ofloxacin is given by levofloxacin.

Instead of estimating $K_{blo:pla}$, the lumped partition–coefficients K_{L_1} and K_{L_2} as described in Equation (5.13) could be estimated. Using the lumped partition–coefficients would avoid the determination of the partition–coefficients for all tissues in each iteration of the NLME estimation algorithm. As a consequence, the approach could be implemented

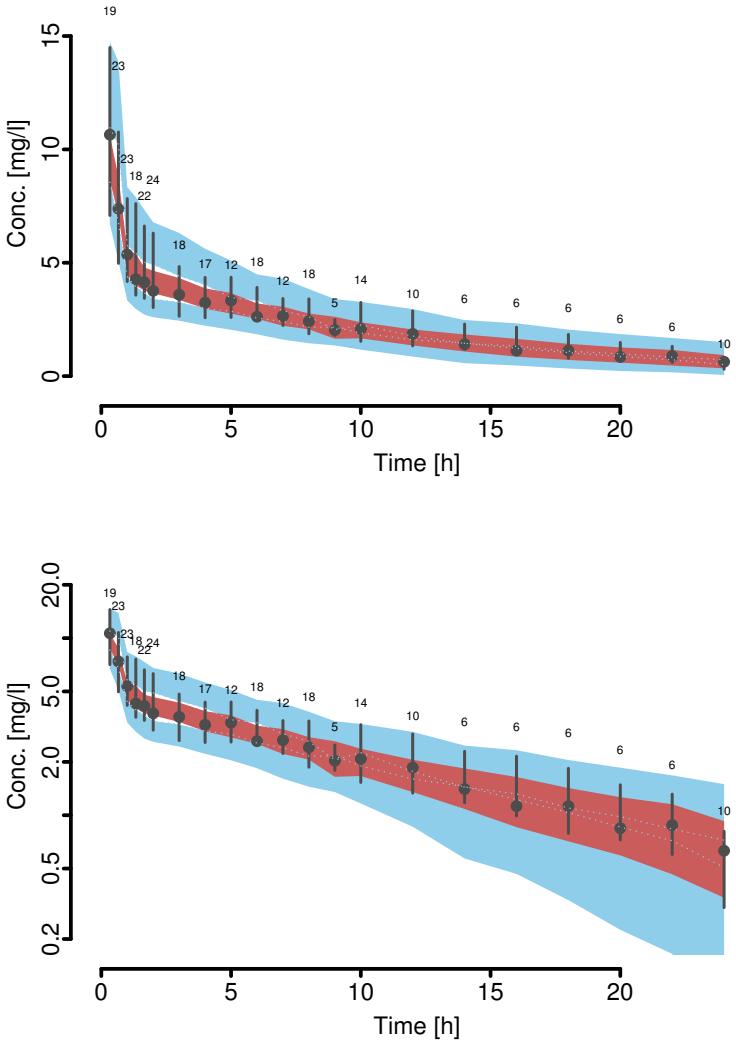

Figure 5.10.: *VPC for the lumped PBPK population analysis for linear (top) and log–scale (bottom): The black bars and dots show the* $10^{th}, 90^{th}$ *and the* 50^{th} *(the median) quantiles of the observations for each time point. The numbers above each bar specifies the underlying number of observations at this time point. The blue shaded area shows the* 95^{th} *error band around the* 5^{th} *and the* 95^{th} *quantiles of the simulated measurements, the red line shows the* 95^{th} *error band of the median of the simulated measurements. The dotted blue line shows the upper and lower range of the prediction interval of the* 10^{th} *and* 90^{th} *quantiles.*

almost as easily as a classical population PK model in standard NLME software tools. Obviously, as a consequence the distinction to the classical population PK approach diminishes. However, using our approach has the crucial advantage that the mechanistically–derived covariate relationships are considered and consequently correlations between the apparent parameters are explicitly induced. As the case maybe, complex interactions and correla-

tions between covariates and the associated impact on the physiology will be captured. As Rowland [2014] stated "ignoring this issue can lead to unrealistic prediction intervals, and conclusion".

After all, using blood–to–plasma ratio instead of the lumped partition coefficients comes with extremely useful advantages from a mechanistic modeling point of view:

- The estimated parameters can be used for the parameterization of the whole body PBPK model and all advantages of PBPK modeling are accessible; in Section 5.3.3 an example of this advantage is given.
- Beyond the impact of the mechanistically–derived covariates on the physiology, a natural correlation between the apparent PK parameters is induced based on the assumption that the distribution into the cellular subspace is similar in all tissues.

Lastly, using two lumped partition–coefficient parameters results in the estimation of one additional fixed effects parameter and eventually one additional random effects parameter compared to the estimation of $K_{blo:pla}$ and the associated BSV.

5.3.2. Comparison of Classical and Lumped Two Compartment Model

In this section the result of the analysis based on the minimal lumped PK model, described in the previous Section 5.3.1, is compared with the results of the classical compartmental analysis presented in Section 3.3. Both approaches provide an adequate description of the data as shown in the VPCs in Figures 3.8 and 5.10 and in the GOF plots shown in Figures 3.7 and 5.9. There is a slight tendency that the classical model better predicts the individual observations as the lumped model, while population predictions seem to be more appropriate in the lumped model.

In the classical compartmental model four model parameters and the corresponding BSV were estimated. In the lumped model approach two parameters and the corresponding BSV, and one parameter—the correction factor—with fixed BSV were estimated. Because the structure of both models was identical, the empirical model had more degrees of freedom and it is obvious that a better description of the data can be expected. This can be observed in the 'IPred vs. Observations' and the 'Residuals vs. Time' plots as well as in the smaller WSV (CV% 10 vs. CV% 12, see Table 5.4). In their work Neckel et al. [2002] describe the bio–analytical method used in the underlying studies and quantified the within– and between day accuracy for the plasma assays to lie between 78.2 % to 106.9 %—i.e. they observed mean deviations of −26.2 % to 6.9 % of the reference concentrations. In conclusion, this means that both WSV estimates are in reasonable order of magnitude.

In Table 5.4 the parameter estimates are compared. To that end, the results of the new approach were transformed into the classical PK parameters. This was done by randomly drawing 1000 virtual patients with replacement from the underlying study population, and $K_{blo:pla}$ and Δ_{lco} from the estimated distribution. Using the lumping methodology the apparent parameters were derived. From the resulting empirical distribution the median (identical to $\exp\theta$ for the log-normal distribution) and CV, assuming a log-normal distribution of the parameters, were estimated. The typical values are similar while the BSV is

larger than in the classical compartmental approach. The lumped model integrated individual information from the patients, the 'Pred vs. Measurements' plots indicate that the unexplained variability was slightly decreased in the lumped model. However, we think, the objective of modeling is not to reduce the unexplained variability as much as possible, but rather identify the 'true' unexplained variability, even if this would lead to a worsening of the model according to the objective function value. In this sense, we align ourselves with Agoram [2014] questioning whether the same objective criteria on mechanistic model should be applied as on purely data–driven models.

Table 5.4.: *Comparison of the population PK parameters of the classical and the minimal lumped modeling approach. Except for the clearance, the latter are based on drawing* 1000 $\log K_{blo:pla}$, Δ_{lco} *from the corresponding population distribution, sampling random patients from the study population and estimating the typical values and WSV for the apparent PK parameters* V_1, V_2 *and* Q *assuming a log–normal distribution.*

	Cl	V_1	Q	V_2
$\exp\hat{\theta}$ classical	8.1 L/h	19.9 L	65.1 L/h	69.2 L
$\exp\hat{\theta}$ lumped	8.1 L/h	17.8 L	56.8 L/h	73.8 L
BSV classical	27.3 %	46.4 %	38.5 %	53.8 %
BSV minimal lumped	26.3 %	23.4 %	15.1 %	30.0 %
WSV classical	9.9 %			
WSV minimal lumped	12.3 %			
OFV classical	−286.5			
OFV minimal lumped	−228.0			

In a last step the individual PK parameter estimates of the two approaches were compared. The scatter–plot in Figure 5.11 shows the high correlation between both models. The plot emphasizes the similarity between both models. The range of the individual apparent PK parameter derived from the lumped model (11.0 L to 24.2 L, 38.7 L/h to 68.9 L/h and 38.1 L to 108.8 L in V_1, Q and V_2) is narrower as the of the classical compartmental analysis (see Table 3.3, 10.9 L to 39.3 L, 25.2 L/h to 214.7 L/h and 31.6 L to 110.7 L in V_1, Q and V_2). However, none of the models is the 'true' model and to decide which model is more appropriate is very subjective.

5.3.3. Population Prediction of ISF Data based on a lumped PBPK model

Based on the previous section, a fully parametrized PBPK model is available. In Section 5.2.1 the methodology to predict tissue distribution based on tissue composition and physicochemical properties of the drug was presented. Various assumptions enter these predictions. Moreover, anatomical and physiological information are most likely affected with uncertainty, which is difficult to assess. In this section we want to investigate the performance of the tissue distribution model and make use of the unique situation to have levofloxacin drug concentration measurements in the ISF of the adipose and muscle tissue.

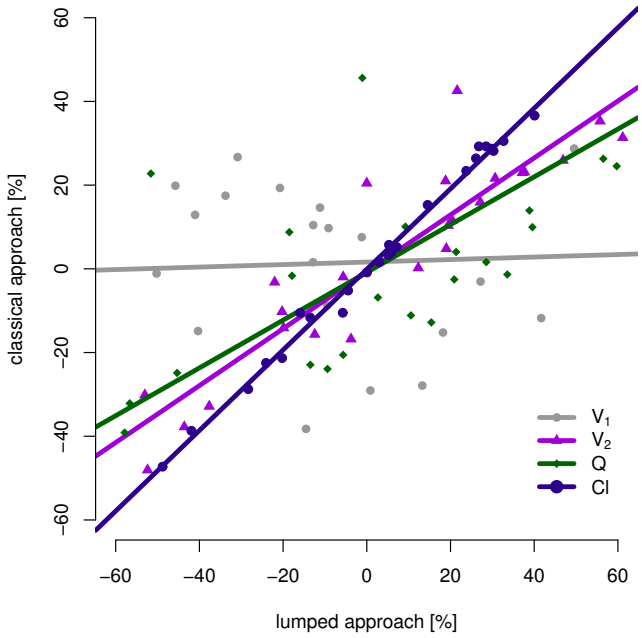

Figure 5.11.: *Comparing the classical (y-axis) and the lumped (x-axis) model w.r.t. the individual EBE of the apparent PK parameters* V_1, V_2, Cl *and* Q *expressed in fold change of the respective typical parameter estimate in Table 5.4. All parameter, except* V_1*, are strongly correlated.*

The ISF measurements were obtained by the so–called retro- and microdialysis method. It is important to know, that this technique—similar to the PBPK approach—underlies several assumptions and only indirectly measures the tissue concentrations. In the next section the microdialysis concept is introduced. Afterwards, the corresponding concentrations of the PBPK model are derived and finally the predictions of the PBPK and the microdialysis approach are compared.

Retro- and Microdialysis

We briefly explain the ideas, principles and limitations of the retro- and microdialysis technique; further information of microdialysis can be found in the comprehensive review Plock and Kloft [2005].

Microdialysis is a sampling technique to approximate the concentration time profile of a drug within the extracellular tissue space. Alternatively, the tissue concentrations can either be determined by tissue biopsy or imaging technology in combination with radio–labeled drug molecules. However, for the biopsy repeated samples over time are difficult to obtain within the same subject, while imaging methods do not allow a direct measurement of the concentration and allocation to a specific subspace.

In microdialysis experiments a pump, a probe and a vial are used. The probe is inserted into the ISF and perfused by a perfusat (*perf*) which is infused by the pump. The membrane of the probe is permeable so that unbound drug molecules can cross the membrane and diffuse

from the ISF into the perfusat and vice versa. The pump produces a constant flow rate of the perfusat through the probe. After passing the probe, the perfusat is collected in a vial over some time interval $[t_1, t_2]$, e.g. 30 minutes. Finally, the dialysate concentration in the vial $C_{\mu D}(t_2, t_1)$ is determined. As a result, the measured concentration $C_{\mu D}(t_2, t_1)$ is linked to the average concentration $C_{dial}(t)$ in the dialysate in this time interval

$$C_{\mu D}(t_2, t_1) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} C_{dial}(t)\, dt,\ t_2 > t_1.$$

The process is illustrated in Figure 5.12. Depending on the drug, the species and the insertion occasion, the experimental parameters, e.g. the membrane permeability, the flow rate of the perfusat and the collection interval have to be calibrated thoroughly.

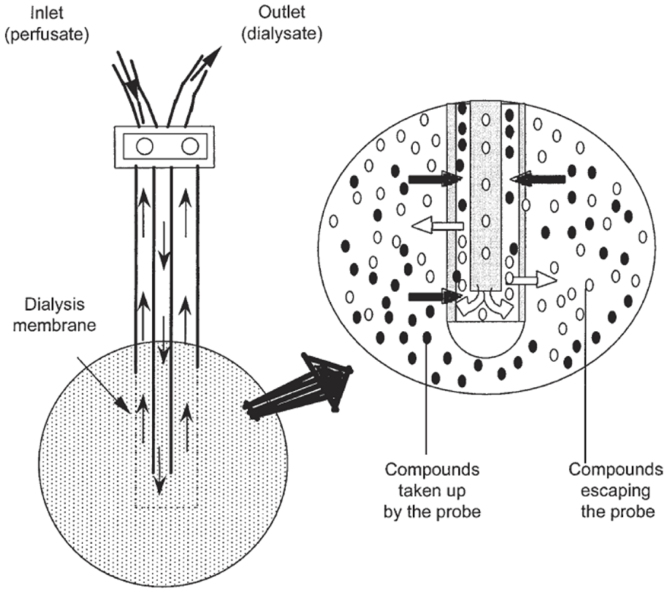

Figure 5.12.: *Cartoon of the microdialysis technique taken from Barbe et al. [2001]. The perfusat perfuses the probe, the drug molecules cross the probes membrane and diffuse from the ISF into the probe and vice versa. The dialysate is collected in a vial which is not displayed here.*

The generic idea is, that the concentration in the dialysate C_{dial} only represents a fraction—denoted as relative recovery RR—of the concentration in the ISF C_{isf}

$$C_{dial} = RR \cdot C_{isf} \text{ and } C_{\mu D}(t_2, t_1) = RR \cdot \underbrace{\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} C_{isf}(t)\ dt}_{=:C_{isf}(t_2, t_1)},\ 0 < RR \leq 1. \tag{5.14}$$

Thus, to predict $C_{isf}(t_2, t_1)$ based on $C_{\mu D}(t_2, t_1)$ the fraction RR has to be determined. Assuming the drug exchange between tissue ISF and perfusat is similar in both directions, the same experiment can performed with $C_{isf} = 0$ and $C_{perf} > 0$, where C_{perf} denotes the

known inflowing concentration from the pump into the probe, which is usually set to 0 in a microdialysis experiment. Hence, the relative drug loss is given by

$$RR := \frac{C_{perf} - C_{dial}}{C_{perf}}. \tag{5.15}$$

This experiment is called retrodialysis and usually forgoes the actual microdialysis measurements before the drug is administered and $C_{isf} = 0$.

Obviously, the role of the relative recovery is very important. The uncertainty affected with this estimate will be carried over to predictions of the tissue concentrations $C_{isf}(t_2, t_1)$; in particular, for compounds with low relative recoveries values, the uncertainty in the ISF predictions will be amplified. As example, imagine the following situation:

- In the retrodialysis experiment a perfusat concentration of $C_{perf} = 1$ mg/L was chosen, in the vial a concentration of $C_{\mu D}(t_2, t_1) = 0.8$ mg/L was determined. The precision of the analytical method is assumed to be proportional to the measured concentration and with reasonable coefficient of variation of CV% 10. I.e. the relative recovery is given by $RR = 0.2$ with 95% confidence interval of [0.04, 0.35], assuming that the adjustment of C_{perf} was error–free.
- In the microdialysis setting a concentration of $C_{\mu D}(t_2, t_1) = 1$ mg/L was determined; the same proportional error of CV% 10 is assumed.
- The predicted average ISF concentration in the collection interval is given by $C_{isf}(t_2, t_1) = 1/0.2 = 5$ mg/L. Incorporating the uncertainty in RR and in $C_{\mu D}(t_2, t_1)$ the 95% prediction interval is given by [2.6, 22] mg/L.

In conclusion, for small values of the relative recovery (RR) already small variations will have large impact on the prediction accuracy. On the other hand, the same considerations with $C_{\mu D,\, rd} = 0.2$ mg/L, i.e. a large relative recovery, would lead to a relatively narrow 95 % prediction interval of [1, 1.5] mg/L.

The example illustrates how the error propagation influences the final predictions. The accuracy of the microdialysis based predictions decreases with decreasing relative recovery values. Usually, the retrodialysis experiment is only performed once, while the microdialysis experiment is repeated several times to describe the full concentration time profile. Thus, a potential error in the retrodialysis measurement influences all ISF predictions. To successfully quantify the WSV, repeated measurements of the retrodialysis as well as the microdialysis would be necessary.

Obviously, the determination of RR is flow dependent, for a slow flow rate the RR will increase and vice versa. To further mitigate the risks of misspecification of RR, the robustness of the RR w.r.t. different flow rates of the perfusat should be tested. However, there are still unsolved issues, e.g. how the insertion of the probe impacts the surrounding tissue and its kinetics or the recovery, for example Clough et al. [2002] observed that (*i*) "local vasodilation, ..., had a significant impact on dialysis recovery from the tissue space" and (*ii*) "manipulation of local blood flow influenced the recovery".

In summary, microdialysis is a helpful tool to predict *in vivo* tissue kinetics, however it is very difficult to quantitatively interpret the results to quantify the accuracy of the predictions.

Modeling of Tissue Interstitial Fluid Concentrations

In this section the concentrations in the PBPK model which corresponds to the tissue ISF concentrations and the assigned microdialysis measurements is derived. A fundamental principle in the PK field is the assumption that after a certain time of a permanent infusion the unbound drug concentrations reach distribution equilibrium in the body. How fast the tissue-plasma equilibrium is reached, depends on the tissue distribution rate $k_{tis} = Q_{tis}/(V_{tis} \cdot K_{tis:up})$, with half life $\log 2/k_{tis}$. Based on the physicochemical properties estimated in the previous Section 5.3.1, tissue distribution half lifes of a typical patient are $\sim$ 14 minutes for adipose and $\sim$ 40 minutes for muscle tissue, compared to $\leq$ 2 minutes for well-perfused tissues like lung or kidney.

Under steady state conditions, interstitial concentration $C_{tis,isf}(t)$ can either be related to the unbound arterial or to the concentration leaving the tissue into the vein

$$C^{ss}_{tis,\,isf} = C^{ss}_{up} = \frac{C^{ss}_{up}}{C^{ss}_{tis}} \cdot C^{ss}_{tis} = \frac{C^{ss}_{tis}}{K_{tis:up}}. \tag{5.16}$$

From Section 5.2.3 and from Figure 5.13, one can infer that for some tissues $C_{tis}(t)/K_{tis:up} \neq C_{up}(t)$ during the distribution phase, i.e. in non–steady state conditions. Because the distribution into the tissue spaces is usually slower compared to the distribution within blood, this has to be considered in the prediction of the unbound tissue ISF concentrations and we assumed that the latter expression in Equation (5.16) corresponds to the ISF concentrations $C_{tis,\,isf}(t) = C_{tis}(t)/K_{tis:up}$.

Instead of simulating the tissue PK profiles with the full PBPK model, again the lumping technique was used. Because we aim for an appropriate description of adipose and muscle tissue, both tissues were separated from the peripheral compartment in the minimal lumped model and single compartments were assigned, see Figure 5.14.

Comparison

The microdialysis measurements $C_{\mu D}(t_2, t_1)$ represents the average drug concentration in the collection time interval. In a classical compartmental population analysis Tunblad et al. [2004] showed that the description in Equation (5.14) is more appropriate as using the so–called mid point method, i.e. $C_{isf}((t_2 - t_1)/2) := C_{\mu D}(t_2, t_1)$ is defined as the concentration value at the middle of the time interval, to link $C_{\mu D}(t_2, t_1)$ and C_{isf} for brain microdialysis data.

Above we have shown that WSV has huge impact on the prediction range, thus for an appropriate comparison WSV has to be considered in the PBPK simulations. Because the output of the model are ISF concentrations, they have to be corrected with the RR to compare them with microdialysis measurements. Thus, in the simulations below the uncertainty assigned to retrodialysis samples, as well as the uncertainty assigned to the microdialysis is incorporated.

In addition to the bio–analytical uncertainty, the WSV also comprises uncertainty related to model misspecification and error in the dependent variables (see Section 1.2) which has to be considered. Because the WSV has not been estimated, we assumed the WSV to be CV% 12, as derived in the analysis of plasma data. Because in general microdialysis data is

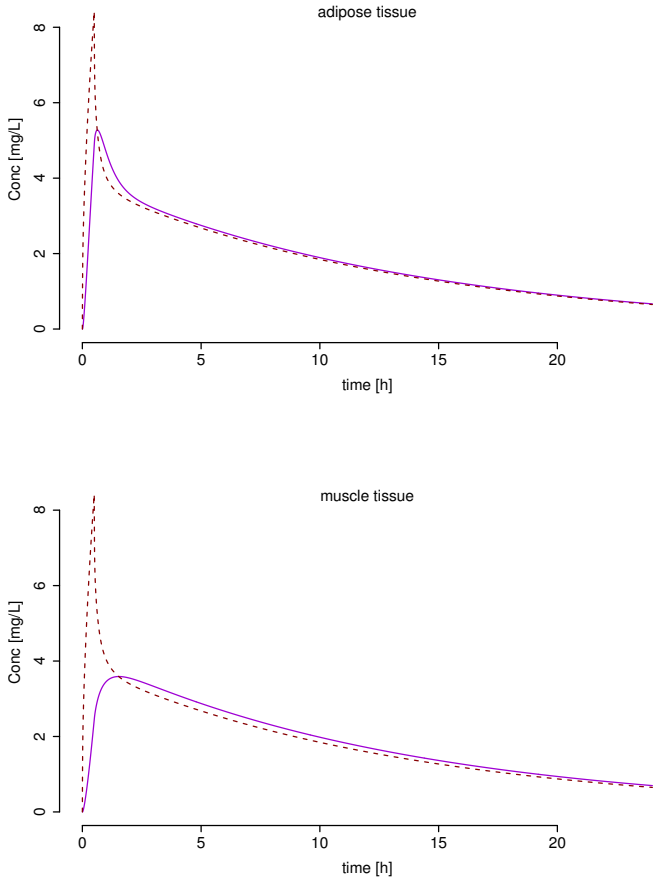

Figure 5.13.: *Non–steady state ISF tissue concentrations according to Equation* (5.16) *for adipose (top) and muscle (bottom) after half an hour infusion of levofloxacin: solid line shows* $C_{\text{tis}}/K_{tis:up}$ *and dashed line shows* C_{up}.

more error–prone as plasma data this is a reasonable and conservative approximation of the lowest possible WSV one could expect in retro- and microdialysis samples. This is in line with findings of Tunblad et al. [2004], who investigated microdialysis in vein and brain and estimated a similar magnitude for the respective proportional error.

Finally, a comparison via a VPC was performed. Therefore, microdialysis measurements were simulated according to the underlying study using the lumped PBPK model as described above, incorporating BSV as estimated in Section 5.3.1 and WSV. These simulations were compared to microdialysis measurements. In summary, the simulations were generated using the following approach:

- For each patient draw a random parameter vector $(\log \text{Cl}, \log K_{blo:pla}, \Delta_{\text{lco}})^T$ of the BSV random effects model

$$\boldsymbol{\Theta_i} \sim \mathcal{N}_3 \left(\theta, \Omega\right),$$

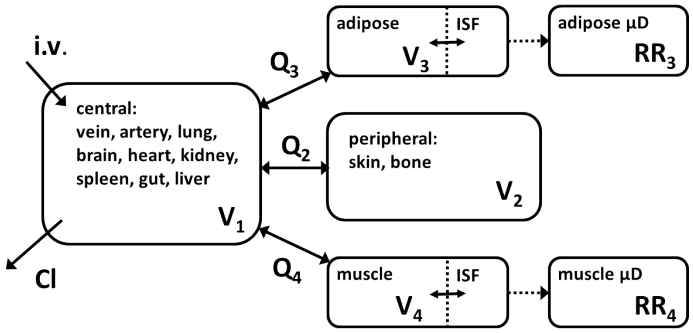

Figure 5.14.: *Lumped 4 compartment model where adipose and muscle concentrations are described by their own compartments. The corresponding interstitial spaces form a sub-compartment in each compartment, with instantaneous drug exchange. From the ISF the fraction RR of drug diffuses into the microdialysis probe.*

with θ and Ω fixed to the results obtained in the analysis of the plasma data in Section 5.3.1.

- For each patient draw one residual $\boldsymbol{\epsilon}_{rd} \sim \mathcal{N}\left(0, \sigma^2\right)$ for the adipose, and one for the muscle retrodialysis measurement—assuming an additive error on the log–scale, with WSV σ of CV% 12. Assemble the residual, the patient's perfusat and retrodialysis sample to a simulated relative recovery according to Equation (5.15)

$$RR_{tis} = \frac{C_{tis,\,perf} - C_{tis,\,\mu D} \cdot e^{\boldsymbol{\epsilon}_{rd}}}{C_{tis,\,perf}}.$$

- The microdialysis concentration is simulated based on RR_{tis} and Equation (5.14). This simulation already contains WSV due to the previous step. Of course, also the detection of the concentration in the vial is error affected. Thus, a residual error $\boldsymbol{\epsilon}_{md} \sim \mathcal{N}\left(0, \sigma^2\right)$ with identical WSV σ^2 of CV% 12 is added:

$$C_{tis,\,\mu D}(t_2, t_1) = \left(\frac{1}{t_2 - t_1} \cdot \int_{t_1}^{t_2} y_{RR_{tis}} \cdot C_{tis,\,isf}(t)\, dt \right) \cdot e^{\boldsymbol{\epsilon}_{md}}.$$

- Please note, in this assessment the microdialysis measurements are not used in an estimation context, but only compared to simulations based on the parameter estimates obtained from using the plasma PK data and the above considerations.

The resulting simulations are compared with the observations using the VPC approach; the comparison is shown in Figure 5.15. The width of the area of the prediction interval indicate the huge impact due to the uncertainty in the determination of the relative recovery. Moreover, the width also reflects the fact that only 5 to 13 measurements are used to derive the quantiles, i.e. the 10^{th} and 90^{th} quantiles will nearly represent the complete range of simulations and measurements for some time points. For the adipose tissue, the simulated typical behavior as well as its variability show good agreement with the corresponding

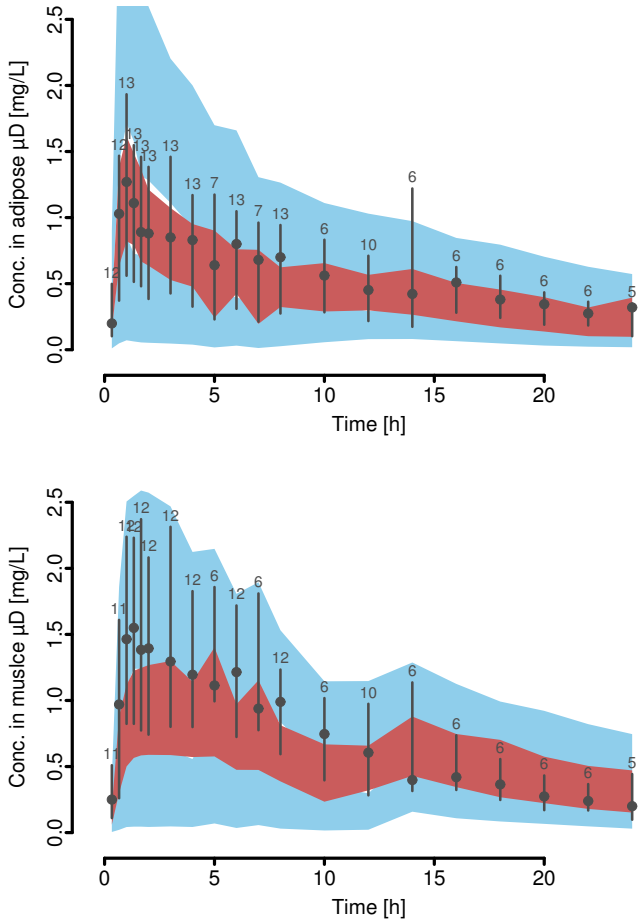

Figure 5.15.: *Quantiles of the microdialysis observations and the model simulations for µD measurements in adipose and muscle tissue. The black bars and dots show the* $10^{th}, 90^{th}$ *and the* 50^{th} *(median) quantiles of the measurements and for each time point. The blue shaded area shows the* 95^{th} *error band around the* 10^{th} *and the* 90^{th} *quantiles of the simulated measurements, the red line shows the* 95^{th} *error band of the median of the simulated measurements, respectively. From the width of the prediction interval the impact of the WSV can be obtained. For the adipose tissue the model predictions and µD data are in adequate agreement. The fast increase and decrease in the beginning, and the peak at around one hour after dosing is also predicted by the model. For the muscle µD there is a slight disagreement between model predictions and measurements, the model predicts a slower absorption in the tissue as observable from the data. The up- and down behavior of the curve between five and seven hours results from the fact that the collection interval of the microdialysis probe varies between one and two hours. Please note, for all subjects with µD measurements the infusion duration was* 30 *minutes and no dosing–specific adjustment of the data as for the plasma data had to be applied.*

microdialysis measurements. This indicates that both approaches predict also similar ISF concentrations.

For the muscle tissue the curves of the typical profiles deviates from the microdialysis measurements. The distribution is predicted to be slower as observed in the data, i.e. the micro-

dialysis measurement in the distribution phase are under–predicted, while in the elimination phase they are over-predicted by the model. This behavior would be transferred to the ISF level one–to–one. However, even if simulations and measurements of the muscle tissue only moderately agree with each other, the typical AUC of the first 8 hours AUC_0^8 =21.6 mg h/ml predicted by the PBPK model is similar to the estimated AUC_0^8 =22.1 mg h/ml obtained by Zeitlinger et al. [2003] in a non–compartmental analysis.

The data was collected in different studies with different sampling schemes: in one study comprising 6 patients, only samples after 12 hours were taken, while in the other studies only samples $\leq$ 12 hours were taken. Thus, for non of the patients a full 24 hours profile was available.

Overall, both approaches are in adequate agreement with slight deviations for the muscle tissue. However, having in mind that the modeling results were obtained without fitting but only predicting the microdialysis data, and the weaknesses of both approaches—most notably the respective assumptions—the results are promising and suggest that the underlying tissue distribution model is a useful tool to predict ISF tissue concentrations.

5.4. Discussion and Conclusion

In this chapter we successfully combined NLME and mechanistic PBPK modeling to analyze population PK data. In detail, we worked out the following new aspects:

(1) Introduced mechanistically motivated BSV into the PBPK model.

(2) Presented a mechanistically–driven population PK analysis in a NLME context.

(3) Derived an interpretation of microdialysis measurements in the context of PBPK modeling and tested the consistency of the *a priori* prediction of the tissue partition coefficients of Rodgers et al. [2005b] and the microdialysis tissue data.

The mechanistically motivated BSV was based on literature information about the quantitative relationship between anthropometric patient variables and physiological parameters. We demonstrated how such known relationships can easily be incorporated into a PBPK model. However, in Huisinga et al. [2012] we have shown that the physiological differences between patients due to changes in anthropometry in general only causes minor variability in the PK compared to other, unknown sources of BSV, e.g. variability in the partition–coefficients.

In the next step, we showed how the theoretically derived lumping approach can be used in the analysis of population PK data and, depending on the objectives of the analysis, is an equal competitor to the classical compartmental analysis. Based on the mechanism–driven analysis of the population plasma PK data, we were able to adjust the physiological information—namely cardiac output, and estimate missing parameters and the magnitude of unexplained BSV which we assigned to the blood–to–plasma ratio, a surrogate for the cellular distribution of levofloxacin. By doing so, a BSV in the tissue partition–coefficients is introduced. The presented approach constitutes a compromise between a purely empirical driven model building—which is per definition only data dependent—and a purely mechanistic approach—which depends on *in vitro* data and literature information.

In a next step it would be interesting to integrate a Baysian approach into the NLME approach for parameters were prior information are available, e.g. similar as used by Krauss et al. [2013, 2015], e.g. cardiac output (in the presented approach the *prior* distribution for cardiac output was fixed to values obtained from literature and only the individual, posterior mode (EBE) was determined).

The third point reveals an important aspects. The scope of Rodgers et al. [2005b] approach was to *a priori* predict the partition coefficient between tissue and unbound plasma; their approach does not aim to predict concentrations within some subspaces of the tissue. Moreover, a typical PBPK model is not developed in such detail that every tissue subspace can be assessed. We examined how the Rodgers et al. [2005b] approach can be applied to predict the concentration in the ISF subspace of the tissue. Assuming steady–state conditions and based on a so–called blood–flow limited tissue distribution, the ISF PBPK predictions were defined as unbound tissue concentration leaving into the vein. These concentration on the one hand reflect vascular space kinetics but also consider the tissue–specific distribution. The *in vivo* microdialysis and the *in silico* population PBPK approach predicted similar tissue kinetics and showed only slight deviation for the muscle tissue, see Figure 5.10. However, compared to a purely data–driven model, for a mechanistically–driven model different model diagnostic standards should be considered, see Agoram [2014]. Concludingly, although the true kinetics are not known, the good agreement between the two independent approaches is promising.

Ball et al. [2014] used similar ideas and verified their animal PBPK model with brain microdialysis data obtained in rats. The model was then used for extrapolation to human. In these extrapolations they also included BSV using the approach by Jamei et al. [2013]. However, neither a population approach to analyze the preclinical animal data was used, nor the uncertainty associated with the microdialysis was incorporated.

There could be several reasons for the deviations observed for the muscle tissue. For example, we have shown that the microdialysis based predictions strongly depend on the retrodialysis experiment and the respective experimental parameters. In further studies, the associated uncertainty should be investigated and quantified by repeating the retrodialysis experiment. Additionally, with a repetition of the microdialysis experiment within the same subjects after a wash–out phase, the consistency of the concentration–time profiles could be investigated. At the same time, the plasma kinetics should be measured to confirm that no intra–subject variability occurred.

On the other hand, the deviations in the muscle tissue could also be caused by model misspecification in the PBPK model: the tissue distribution model simplifies the tissue to basically one well–stirred compartment and only predicts the average tissue concentration. Moreover, in the PBPK model we assumed that composition and perfusion of the adipose and muscle tissue is identical all over the body. Concludingly, we expect identical kinetics in a tissue independent from the exact location, for example muscle tissue in the legs and arms. In further studies this hypothesis could be challenged by conducting a microdialysis experiment with simultaneously measuring the drug concentrations in the same tissue (e.g. muscle) at different locations (e.g. right/left arm and right/left leg).

The approximation of the ISF predictions in the PBPK model (see Equation (5.16)) is based on the same assumptions underlying the tissue distribution model by Rodgers

et al. [2005b]: the tissue distribution only depends on passive diffusion and for a continuous drug infusion, steady state can be reached, i.e. unbound drug levels in all body fluids are identical. Based on several microdialysis studies, Liu and Derendorf [2003] conclude that this "does not always hold true because many studies have shown the discrepancy between free drug levels in plasma vs. tissue". However, for levofloxacin we showed that incorporating the uncertainty in the microdialysis based ISF predictions, the unbound tissue predictions based on the modeling and microdialysis approach are in good agreement. Concludingly, we cannot support their conclusion.

The uncertainty, in particular in the assumptions of the tissue distribution model, is one of the major weaknesses of the presented mechanistic population PK approach. A future aspect to gain further confidence, would be to apply and test the new approach for other drugs. A verification of the results could consists of determining the blood–to–plasma ratio *ex vivo* and compare it to the estimated value.

Compared to the microdialysis approach, the presented modeling method to predict the target concentrations is only based on plasma data which is easier to assess. Furthermore, once a model is established and validated, not only the tissue kinetics can be predicted, but also patients or sub–populations characteristics can be taken into account, e.g. obese patients or disease specific modifications of the physiology. However, if a classical compartmental analysis is preferred, the approach by Pilari and Huisinga [2010] should be considered to decide whether the concentrations in the central or the peripheral compartment are more appropriate predictors of the target concentrations driving the efficacy, depending on the tissue and on the drug properties.

In the introduction of this chapter, the objectives were motivated by emphasizing the important link between PK and efficacy of antibiotics. We showed that the derived PBPK for levofloxacin adequately describe the plasma as well as the tissue ISF concentrations. In a next step, the relationship between plasma and tissue PK derived indices, e.g. the duration of time over MIC could be investigated. If necessary, correction factors could be derived to minimize the potential differences. Furthermore, as population PK models are already used for the treatment individualization and optimization, e.g. see Vinks [2002, 2014], population PBPK models could be utilized for the same purpose with even more precise individualization and with the advantage to provide predictions for the drug concentrations at the target.

The underlying chapter was based on levofloxacin. Nonetheless, this chapter does not only cover results of the levofloxacin analysis, but also introduces new principles and ideas to combine the different methodologies—which can also be used and applied to other drugs, presumed the mechanism of distribution is known. The input of the presented mechanism–driven modeling approach is the accumulated prior knowledge about the drug, species and/or study population, and the data collected in a clinical study. Based on this—at the time available—information a PBPK model is defined. In a next step, the reference physiology is individualized based on the individual covariates of the study participants. In order to reduce the complexity of the PBPK model and to identify the parameters which can be estimated based on the plasma PK data the lumping–approach is applied. Finally, the individual physiology, the minimal lumped PBPK model and the NLME approach is integrated into the analysis of the plasma PK data. Based on the results of this analysis a population

PBPK model can be created. Furthermore, the determined EBEs can be used to generate individual PBPK models for each subject of the study. The new information generated by this analysis can be fed back to the accumulated know–how of the drug, species and/or study population. Furthermore, the population PBPK model can be used for simulations, extrapolations and planning of further clinical trials, e.g. a study in a pediatric population. This represents a further contribution to the learn & confirm paradigm across different stages of drug development. A schematic presentation of the new approach was shown in Section 5.1 in Figure 5.1.

6. Quantitative Psycholinguistics: Modeling the Competition in the Time Course of Spoken Word Recognition

The following chapter is planned to be published and originated from on a cooperation with Prof. Reinhold Kliegl, Institute of Cognitive Psychology, University of Potsdam. The chapter focuses on the modeling aspects and only briefly the psycholinguistic background is introduced.

6.1. Introduction

The last chapter, as well as the illustrating examples in part I dealt with repeated measurement studies in pharmacokinetics. In this chapter, we switch to a different discipline—the field of psycholinguistics. Psycholinguistic research investigates the development and comprehension of language using quantitative assessments of language–related behavior to decide between theoretical propositions. In one very prominent research paradigm subjects' direction of gaze at a computer monitor is measured at a very fine time scale (e.g., every ten milliseconds) while they listen to a sentence describing the content of the display. This experimental approach is used "to exploit the systematic relationship between eye movements and speech processing" (Huettig et al. [2011]) and is called the visual word paradigm (VWP).

The VWP was first introduced by Cooper [1974] and reached new prominence twenty years later Tanenhaus et al. [1995]. It is well established that there exists a relationship between eye movements and the time–course of mental comprehension of spoken language (Rayner [1998, 2009]). Therefore, analysis of the eye movements afford inferences about how comprehension unfolds over time. For example, Cooper [1974] experiments revealed that the participants' attention (indicated by the direction of gaze) was attracted to those objects that where mentioned by the speaker, e.g. the likelihood of looking at the picture of a dog increased if a dog was mentioned by the speaker. Consequently, it is possible to check whether a systematic manipulation of part of the speech input, often quite subtle, leads to an effect on when and for how long an object is looked at. For example, we can ask whether phonologically or semantically similar words lead to different time courses in the direction of gaze, how the time course of language comprehension differs between languages and how it changes as children develop their language ability. There is research on how language comprehension changes across the adult life span, on how diseases such as stroke or Alzheimer affect language–related processes, and on many other questions. An overview of topics addressed with the VWP can be found in the review by Huettig et al. [2011].

In the present work, data reported in a study conducted by Magnuson et al. [2003] were reanalyzed. The goal of this study was to investigate how geometric objects described by

phonologically similar words compete for attracting the gaze. In a first phase, participants were trained to associate simple, artificial words, e.g. 'pibo', with geometric figures. Each word had three so–called neighbors or distractors: an onset–matching neighbor called 'cohort', which differed only in the final vowel, e.g. 'pibu', an onset–mismatching neighbor called 'rhyme', which differed only in its initial consonant, e.g. 'dibo', and one dissimilar item, which differed in the first and last phonemes, e.g. 'dibu'. After participants had learned the associations, the experiment started and the geometric object (e.g. associated with 'pibo') was displayed with two arbitrary unrelated geometric objects and one of the three distractor objects. Thus, the experiments tests whether, e.g., a cohort or a rhyme distractor object leads to a stronger distraction from the target item than an unrelated distractor object. The experiment was repeated with different words and under varying distractor conditions. For the analyses, the measurements were aggregated for each participant by experimental condition and time bin. For example, for a given participant, the rhyme neighbor was used as a distractor in 16 experiments. At time point 2 s the participant fixated the target in 12 out of the 16 experiments. Thus, the result was provided as relative frequency of $0.75 = 12/16$ target fixations at time 2 s.

From a modeling perspective this study and the recorded measurements differ in some interesting features from the population PK setting used in the previous chapter. Eye movements were tracked with a high temporal resolution; the number of samples and sampling times were identical for all participants. Moreover, the experiment was repeated with different words (i.e., across experiments). Hence, one not only has repeated measurements over time, but the experiments were also repeated under identical and varying experimental conditions. Magnuson et al. [2003] reported the relative frequency of the target, distractor and unrelated item fixations over time under different experimental conditions. Another difference to PK data is that there are almost no missing data, but the eye movement itself is a relatively noisy process.

As far as the current practice in the analyses of these kind of eye movement data is concerned, Mirman et al. [2008] identified three gaps: (*i*) appropriate analysis of the time courses, (*ii*) characterization of individual effects—that is the description and quantification of the BSV distribution, and (*iii*) interpretation of individual differences—meaning identifying factors that explain inter–individual differences. Mirman et al. [2008] suggest to use linear mixed effects modeling to address these gaps. As a structural model they used a polynomial function describing the average eye movement time–course and exemplified their approach based on the dataset by Magnuson et al. [2003]. In their analysis they only considered the measurements within time intervals they had identified to be theoretical importance. For example, in this time interval the relative frequency of target fixation increases over time. Mirman et al. [2008] modeled this increase with a polynomial function with degree 3, which obviously converges to $\pm\infty$, a value which is outside of the range of possible measurements $[0, 1]$. Therefore, their model is only applicable in this narrow time window and has no predictive value beyond. In addition, the model parameters are difficult to interpret: E.g. what does it mean that the typical value of the coefficient of the quadratic order term is -0.2264 for the cohort distractor vs. -0.1476 for the rhyme distractor?

Another drawback of the approach by Mirman et al. [2008] is that target and distractor fixation were analyzed independently of each other. Obviously, there is a competition between both and the time courses are strongly correlated. Indeed, from a theoretical perspective, one could argue that the quantification of the competition between target and distractor is the most interesting question of this experiment. With independent analyses one arrives only at an indirect rather than a direct assessment of the most relevant piece of information.

Here we propose an alternative, more comprehensive approach which we denote as the 'competition model'. For our analyses, we used the same dataset as Mirman et al. [2008], but propose a structural model that takes into account the competition between target and distractor and, in addition, uses a parameterization that allows easier interpretation of the estimated parameter values. To this end, the analysis incorporates all data from the target, distractor and unrelated item fixations simultaneously. As a consequence the information is utilized more effectively. The analysis was performed within the NLME framework using the EM algorithm described in Section 3.1.

Moreover, we explicitly take into account that the originally measured data are discrete, i.e., at time point t the participant fixated k out of n times the target. Naturally, the tuple (k, n) contains more information than the aggregated version the fixation proportion k/n. Unfortunately, the original data in the form (k, n) was not reported in the dataset and we had to retrieve k and n according to the information contained in the publication by Magnuson et al. [2003]. To take into account the stochastic nature of the experiment, we assumed a discrete stochastic distribution describing the WSV and combined this with a NLME approach.

In summary, we complement the gaps identified by Mirman et al. [2008] in the analysis of eye–movement time courses with the following issues:

(i) The structural model should
   - be capable to describe and predict the full time course of measurements,
   - be able to improve the interpretation of the time course, compared to e.g. only visual inspection of the raw data, or simple descriptive statistical analysis,

(ii) The extraction of information contained in the data should be optimized, e.g. by incorporating the competition between the different competitor items and by taking also into account the absolute number of experiments that were used to estimate the relative frequencies of target, distractor and unrelated item fixations. Thus, the modeling uses much more of the information contained in the data delivered by this experiment than previous analyses.

In the next section the details of the study by Magnuson et al. [2003] are briefly described. Subsequently, the structural model to describe the time–course of the probability to look at the target, distractor and the unrelated item is introduced. This structural model is first applied in a naive pooling analysis. This analysis revealed first insights in the model and the data. Based on these results, a NLME analysis is conducted to quantify the distractor effects as well as the BSV and WSV. In a last step the discrete data is analyzed using an

modified NLME approach, where the so–called Dirichlet–multinomial distribution is used as WSV distribution.

6.2. Material and Methods

The study was conducted by Magnuson et al. [2003] and comprised $N = 16$ participants. The experiments were performed on two consecutive days. Each day started with a training session, followed by the experimental trials. The objective of the training was to have participants learn associations between geometric figures and short novel words. The geometric figures were randomly created before the start of the experiment, ensuring that figures are comparable in memorability. Each word, e.g. 'pibo', had three distractors with an associated geometric figure each: a (cohort) neighbor, which differed in the final vowel, e.g. 'pibu', a (rhyme) neighbor, which differed in its initial consonant, e.g. 'dibo', and a dissimilar item, which differed in the first and last phonemes, e.g. 'dibu'. The frequency of the items were controlled during the training and separated into low and high frequency occurrence.

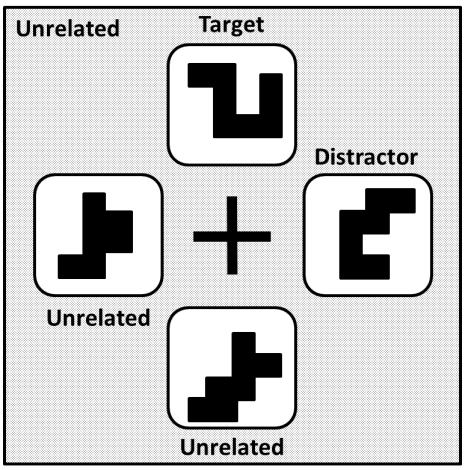

Figure 6.1.: *A cartoon of the arrangement displayed to the participants (of course, without the notations). The participant heard a word and four different item were shown: one of the items (the target) was associated with the spoken word, one or none of the figures belonged to the phonological distractor of the word, and two or three items were phonologically unrelated to the spoken word.*

In total 16 sets of words were trained. During the actual test each item appeared 6 times, once with a 'cohort', once with a 'rhyme' and four times with 'none' distractor (e.g. including the dissimilar item), resulting in a total of 96 trials, 64 without a distractor, 16 with a 'cohort' distractor and 16 with a 'rhyme' distractor. The experiments were conducted according to the description by Magnuson et al. [2003]

(1) A central fixation cross appeared on the screen. The participant clicked on the cross to begin the trial.

(2) After 500 ms, four shapes appeared, see Figure 6.1. After 750 ms, the participant heard the instruction 'Look at the cross' through headphones. The participant fixated

the cross, clicked on it with the mouse, and continued to fixate the cross until they heard the next instruction (as they had been trained).

(3) 500 ms after clicking on the cross, the instruction was given, e.g. 'Click on the pibu'. 150 ms after the participant clicked on an item, all four items disappeared.

The description above reflects the experimental protocol. Importantly, however, compliance was not controlled during the test. It is worth mentioning, that in the analysis and dataset of Magnuson et al. [2003], only successfully performed trials were included, i.e. these where the correct object was clicked in the last step. Moreover, only the relative frequencies were reported, i.e. the dataset contained the following type of record: at time 2 seconds the target fixation proportion was 0.875, instead of providing the original results, e.g. 14 out of 16 correct accomplished experiments.

For the discrete modeling approach, we aimed at recreating the original measurements by assuming that for a 'cohort' and a 'rhyme' distractor at most 16 trials and for the 'none' distractor most 64 trials were performed. Based on each relative frequency value k/n, we tested which values for the original k and n were most appropriate under these considerations. Because the relative frequency values were rounded to the 3^{rd} digit, the recreated original values might be inaccurate. This might marginally impact the results of the following analysis, however the conclusions for the new approach we present here are not affected by this.

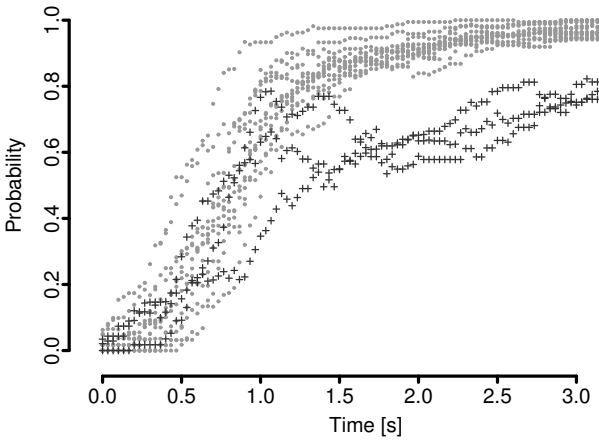

Figure 6.2.: *Raw data of Magnuson et al. [2003]. The plot shows the target fixation proportions of the trials without a distractor (condition 3). The black crosses depict the measurement of the three subjects which deviate from the rest of the population. These subjects also deviate under the other experimental conditions. Hence, they were excluded from the analysis.*

For simplification, the analysis was only based on the measurements obtained on the second day. On the second day, the participants were already used to the experimental procedure and consequently eye movements were less noisy.

As exemplified in Figure 6.2 for the experimental condition where no distractor was present, we identified three participants with deviating eye movement profiles. As the same

deviation is present for the other experimental conditions, for simplification these three subjects were excluded from the analysis. Moreover, for the underlying analysis we ignored that there were differences in the training frequency of the different words—as this was balanced between the distractor conditions we expect this to be negligible.

In summary, the dataset we used for our analysis contained 13 participants and we focused on the impact of the distractor—whether a cohort distractor (condition 1), a rhyme distractor (condition 2) or none distractor was present (condition 3).

Later in this section, we will compare two so–called nested models with each other. To decide which model more appropriately describes the data, the so–called likelihood ratio test (LRT) will be applied which is a widely used statistical test for this purpose.

The likelihood ratio test

The LRT is a statistical test to decide whether one model—the alternative model M_1—statistical significantly better describes the underlying data as the null model M_0. The LRT can be used, if the two models M_1 and M_0 are nested, i.e. the alternative and the null model are identical, except that in the null–model only a subset of parameters of the alternative model are estimated, and the remaining parameters are fixed to pre-chosen values. The statistical hypothesis can be formulated as:

$$H_0 : \ \beta = \beta_0 \text{ versus } H_1 : \ \beta = \beta_1,$$

with $\beta_0 \in \Upsilon^0 = \{b_1\} \times \cdots \{b_k\} \times \Upsilon^{\beta_{(k+1)}} \times \cdots \times \Upsilon^{\beta_{(p)}}$, $\beta_1 \in \Upsilon^1 = \Upsilon^{\beta_{(1)}} \times \cdots \times \Upsilon^{\beta_{(p)}}$, with $k > 0$ and $(b_1, \ldots, b_k)^T \in \Upsilon^{\beta_{(1)}} \times \cdots \times \Upsilon^{\beta_{(k)}}$, where $\Upsilon^{\beta_{(l)}}$ denotes the parameter space of the l-th component of β, for $1 \leq l \leq p$. Under the assumption of some regularity conditions, e.g. that b_l is an interior point of $\Upsilon^{\beta_{(l)}}$, for all $1 \leq l \leq k$, the respective test statistic is defined as the ratio of the maximum likelihood functions

$$\Lambda = -2 \log \left(\frac{\max\limits_{\beta \in \Upsilon^0} L(\beta; y)}{\max\limits_{\beta \in \Upsilon^1} L(\beta; y)} \right). \tag{6.1}$$

Wilks [1938] showed that in limit (regarding the number of subjects) the distribution of Λ converges to the χ^2_k distribution with k degrees of freedom. Although, in practice the number of subjects is always limited, and often rather small, the Wilks theorem is used to approximate the distribution of Λ by the χ^2 distribution. I.e., if the type–one error is chosen to be $\alpha = 5\%$, and the two models M_0 and M_1 are differing in 1 parameter, e.g. $\beta_1 = (\beta_{1\,(1)}, \ldots, \beta_{1\,(p)})^T$ and $\beta_0 = (\beta_{0\,(1)}, \ldots, \beta_{0\,(p-1)}, 0)^T$, the null hypothesis $H_0 : \ \beta = \beta_0$ will be rejected if $\Lambda > c$, where c is the quantile function of the χ^2_1–distribution, evaluated at $1 - \alpha$.

The LRT is a frequently used tool in model development, e.g. see Bonate [2011]. The corresponding mathematical details, e.g. the proof of Wilks theorem can be found in Lehmann and Romano [2005].

6.3. Results

6.3.1. The Structural Model: Integrating the Competition between Target, Distractor and Unrelated Items

The objective of the analysis was to describe the time course of the participant's target recognition and to what extent a distractor influences this process.

We describe this process by describing the underlying probability to fixate the target $p_{\text{tar}}(t) : \mathbb{R}_{\geq 0} \to [0,1]$, or the distractor $p_{\text{dis}}(t) : \mathbb{R}_{\geq 0} \to [0,1]$ or the unrelated items $p_{\text{unrel}}(t) : \mathbb{R}_{\geq 0} \to [0,1]$ over time. Because, the three items compete for the participants' attention, the associated probabilities are strongly correlated. Yet, in the past, this competition was not considered in the analysis of such data; the three curves were analyzed as independent time series. I.e. the nature of the underlying process was neglected and the additional information was lost and ignored. For example, the observation that a participant's fixation switches between the target and the distractor, contains different information than the observation that the fixations move randomly between all displayed items. Thus, the interaction between the items should be captured by the model. Our model captures this information by assuming that the probability to look at one item changes over time and is constantly exchanged between the three items, while maintaining the constraint that the participant had to look somewhere $p_{\text{tar}}(t) + p_{\text{dis}}(t) + p_{\text{unrel}}(t) = 1$ for $t \geq 0$.

Another implicit consideration is that the probability time course is continuous. This ensures a correlation between two consecutive time points, i.e. the random variables describing the occasion of eye fixation at time t_k and t_{k-1} are not independent of each other.

The following following ODE system was developed to fulfill the properties mentioned above and to describe the general trend observable in the data. The system comprises one ODE for each probability p_{tar}, p_{dis} and p_{unrel}, describing the interaction between the three probabilities, while maintaining the above constraints. A schematic view of the model is shown in Figure 6.3 and the ODE is given by

$$\begin{aligned}
\frac{d}{dt}p_{\text{tar}}(t) &= k_{\text{dis}\to\text{tar}}(t)\cdot p_{\text{dis}}(t) + k_{\text{unrel}\to\text{tar}}(t)\cdot p_{\text{unrel}}(t) \\
&\quad - (k_{\text{tar}\to\text{dis}}(t) + k_{\text{tar}\to\text{unrel}}(t))\cdot p_{\text{tar}}(t) && \text{(Target)} \\
\frac{d}{dt}p_{\text{dis}}(t) &= k_{\text{tar}\to\text{dis}}(t)\cdot p_{\text{tar}}(t) + k_{\text{unrel}\to\text{dis}}(t)\cdot p_{\text{unrel}}(t) \\
&\quad - (k_{\text{dis}\to\text{tar}}(t) + k_{\text{dis}\to\text{unrel}}(t))\cdot p_{\text{dis}}(t) && \text{(Distractor)} \\
p_{\text{unrel}}(t) &= 1 - p_{\text{tar}}(t) - p_{\text{dis}}(t) && \text{(Unrelated)}
\end{aligned}$$

with $k_{v\to w}(t) = K_{base} + K_{v\to w}\,\delta(t)$, for $v, w \in \{\text{tar}, \text{dis}, \text{unrel}\}$, $v \neq w$, where $\delta(t) = \frac{t^h}{t_{1/2}^h + t^h}$ denote a time delay for the on-set of the reaction. The incorporation of a time delay was necessary, because we observe a delayed response after starting the experiment in the measurements, see e.g. Figure 6.2. This delay can either be interpreted as a reaction process, or it represents an artifact of the poorly defined and controlled starting conditions of the experiments; e.g. that the eye movements were already recorded nor the experiment started and/or ensuring that the participants fixated the cross in the middle. The Hill coefficient h denotes the steepness of the increase in $k_{v\to w}$; for large values of h, the time delay term can be interpreted as a continuous implementation of a lag time t_{lag}, i.e. $k_{v\to w}(t) = 0$, for

$t < t_{\text{lag}}$, and $k_{v\to w}(t) = K_{v\to w}$, for $t \geq t_{\text{lag}}$. The half time parameter $t_{1/2}$ denotes the time where the rate parameters $k_{v\to w}(t)$ reaches half of its maximum value $K_{v\to w}$.

The initial state of the ODE system in Equation (6.2) is given by the observed values for $p_{\text{tar}}(0)$, $p_{\text{dis}}(0)$ and $p_{\text{unrel}}(0)$. These values were read out of the dataset. If the participants followed the protocol the initial states are $p_{\text{tar}}(0) = p_{\text{dis}}(0) = 0$ and $p_{\text{unrel}}(0) = 1$.

The baseline rate K_{base} was incorporated to arrange a baseline flow between all compartments, thus it controls the equilibrium state for $t \to \infty$ of the system. In Figure 6.4 the behavior is illustrated for small and large value of K_{base}. Introducing the baseline rate K_{base} was necessary to ensure that a minimum probability to look at one of the three items retains, i.e. $\lim_{t\to\infty} p_l(t) > 0$, $l \in \{\text{tar}, \text{dis}, \text{unrel}\}$.

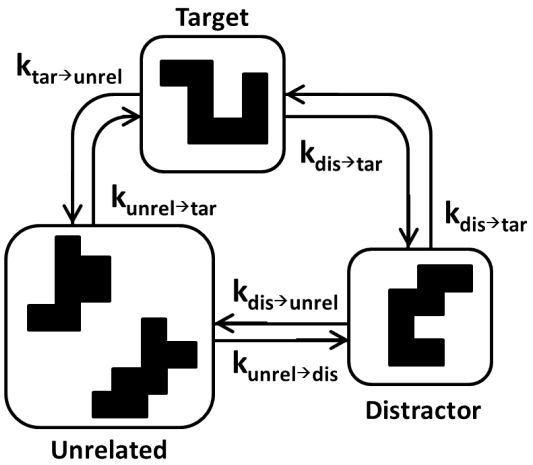

Figure 6.3.: *Schematic representation of the model: three compartments describe the probability to look at the target, the distractor and the unrelated item. The three compartments are connected with rates describing the probability exchange between the compartments.*

In summary, the model parameterization can be interpreted in the following way. The parameters K_{base}, h and $t_{1/2}$ affect all rate parameters $k_{v\to w}$ to correct for possible delays in the execution of the experiment and to avoid that $p_{\text{tar}}, p_{\text{dis}}, p_{\text{unrel}} \to 0$; in this sense these parameters are nuisance parameters. The more interesting parameters are the rates $K_{v\to w}$. Comparing these values leads to an interpretation of the underlying process. E.g. in general, the target is recognized at one point and it is expected that $K_{\text{dis}\to\text{tar}} > K_{\text{tar}\to\text{dis}}$ and $K_{\text{tar}\to\text{unrel}}, K_{\text{dis}\to\text{unrel}} \approx 0$. For a cohort distractor, where the distractors first consonant is identical to first consonant of the target, and the second consonant differs, it is expected that $K_{\text{unrel}\to\text{tar}} \approx K_{\text{unrel}\to\text{dis}}$ and $K_{\text{dis}\to\text{tar}} \gg K_{\text{tar}\to\text{dis}}$ while for the rhyme distractor, where the first consonant differs and the second consonant is identical, it is expected that $K_{\text{unrel}\to\text{tar}} > K_{\text{unrel}\to\text{dis}}$.

In the following section this model is denoted as the 'competition model'.

6.3.2. Naive Pooling Analysis of the Stratified Dataset

Before the NLME analysis was conducted, the naive pooling approach as described in the Appendix 9 (page 128) was applied for each experimental condition separately ('cohort',

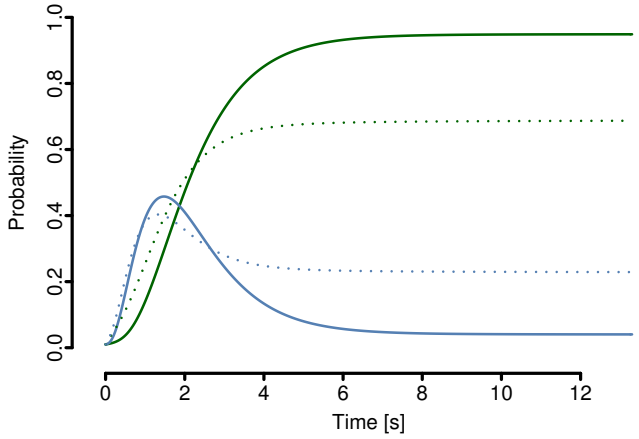

Figure 6.4.: *Illustration of the impact of K_{base}: green lines show the probability of target fixation, blue line shows the probability of distractor fixation, therein the solid lines represents a small value for K_{base} and the dashed line a larger value for K_{base}.*

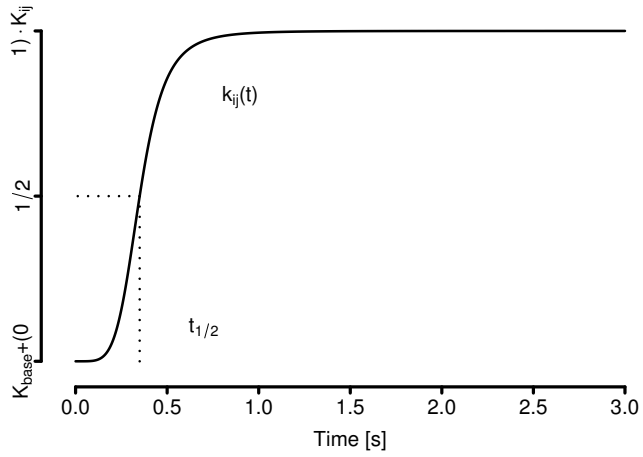

Figure 6.5.: *Curve showing the time delay of the parameters. The larger the Hill coefficient is, the steeper the curve. The vertical and horizontal dotted line indicate the time $t_{1/2}$ where half of the parameter value is reached.*

'rhyme' or 'none' distractor). The goal was to examine the ability to describe the data by this structural model, to test which parameters can be identified and to have robust initial estimates for NLME analysis.

The measurements depict frequencies of target, distractor and unrelated item fixations, and the competition model predicts the associated probabilities, i.e. measurements as well as predictions lie in $[0, 1]$. The WSV level is described by an additive error model, to avoid that $P(\boldsymbol{Y} > 1) > 0$ or $P(\boldsymbol{Y} < 0) > 0$ the data as well as the model f are logit-transformed (logit : $(0, 1) \to \mathbb{R}$ with $\operatorname{logit} p = \log(p) - \log(1 - p)$). Thus, it is assumed that the logit-transformed observations $\boldsymbol{y}_{\mathbf{cohort}}, \boldsymbol{y}_{\mathbf{rhyme}}, \boldsymbol{y}_{\mathbf{nonel}}$ (each containing the mea-

surements of all subjects under the respective experimental condition), are realizations of $\boldsymbol{Y}_{\mathbf{cohort}}, \boldsymbol{Y}_{\mathbf{rhyme}}, \boldsymbol{Y}_{\mathbf{none}}$ with distribution defined by

$$\boldsymbol{Y_l} = \operatorname{logit} f(x;\theta_l) + \boldsymbol{\epsilon_l},\ \boldsymbol{\epsilon_l} \sim \mathcal{N}_{n_l}\left(0, I_{n_l}\sigma_l^2\right),$$

for $\sigma_l > 0$, and distractor $l \in \{\text{cohort}, \text{rhyme}, \text{none}\}$ and $x = (t_1, \ldots, t_n)^T$ containing the measurement time points which were identical for all three distractor. The initial conditions of the system were fixed according to the experimental conditions; i.e. the participants fixated the unrelated item at the beginning of the experiment: $p_{\text{dis}}(0) = p_{\text{tar}}(0) = 0$ and $p_{\text{unrel}}(0) = 1$, because the logit function is not defined for 0 and 1, a threshold $c \in (0,1)$ is chosen and observed and predicted values $\leq c$ were replaced by c and values $\geq 1 - c$ were replaced by $1 - c$.

The naive pooling analysis revealed that not all model parameters could be identified successfully. To detect the parameters which are not identifiable, the following procedure was performed. In a first attempt, for the three experimental conditions, all model parameters were estimated. However, the respective Hessian $-\nabla^2_\theta \log p^{\boldsymbol{Y_l}}(\operatorname{logit} \boldsymbol{y_l}; \hat{\theta}_l)$ evaluated at the estimates $\hat{\theta}_l$ was not positive definite for all $l \in \{\text{cohort}, \text{rhyme}, \text{none}\}$. According to Gill and King [2004] this "can only be caused by multicollinearity or by including more explanatory variables (the parameters) than measurements". Because obviously the latter is not the case, the model parameters are highly correlated and not all of them can be identified successfully. Thus, one parameter at a time was excluded and the identifiability according to the above criterion was tested. The results and the final parameterization are shown in Table 6.1. Thereby, the parameterization is not unique and different parameterizations provide comparatively adequate description of the data. However, the parameterization which is most plausible was chosen, i.e. it is assumed that the physiological recognition process goes via unrel $\rightarrow$ dis $\rightarrow$ tar or unrel $\rightarrow$ tar and the competition model is parameterized by $K_{base}, K_{\text{unrel}\rightarrow\text{tar}}, K_{\text{unrel}\rightarrow\text{dis}}, K_{\text{dis}\rightarrow\text{tar}}, h$ and $t_{1/2}$. Moreover, the results are in good agreement with the expectations that the distraction by the cohort distractor seems to be more pronounced in the first time period as for the rhyme distractor, indicated by a larger $K_{\text{unrel}\rightarrow\text{dis}}$ in comparison to $K_{\text{unrel}\rightarrow\text{tar}}$.

The resulting probability–time profiles for all three conditions is shown in Figure 6.6. The model adequately describes the time–trend for each condition.

The resulting parameter value are in line with the visual impression of the data: the probability to look at the target increases faster if no distractor is present ('none' distractor). For the 'none' distractor, the model also provides the most appropriate data description indicated by the smallest WSV. This is in line with the considerations made earlier because eye movements are per se a volatile process, i.e. also during fixation of an object the eyes might shortly sway to another object within the field of view. Because in the 'none' distractor case the eye fixation of the four displayed objects is only assigned to 'unrelated' or 'target' instead of 'unrelated', 'target' and 'distractor' the noise in the data is decreased.

6.3.3. Auto–Correlated Residuals

The competition model and the parameterization are now identified. Looking at the individual participant's data, which is the basis for the NLME analysis, we observe that the time–course of the relative fixation proportion is a discontinuous process, see for example Figure 6.7. This observation follows from the considerations that:

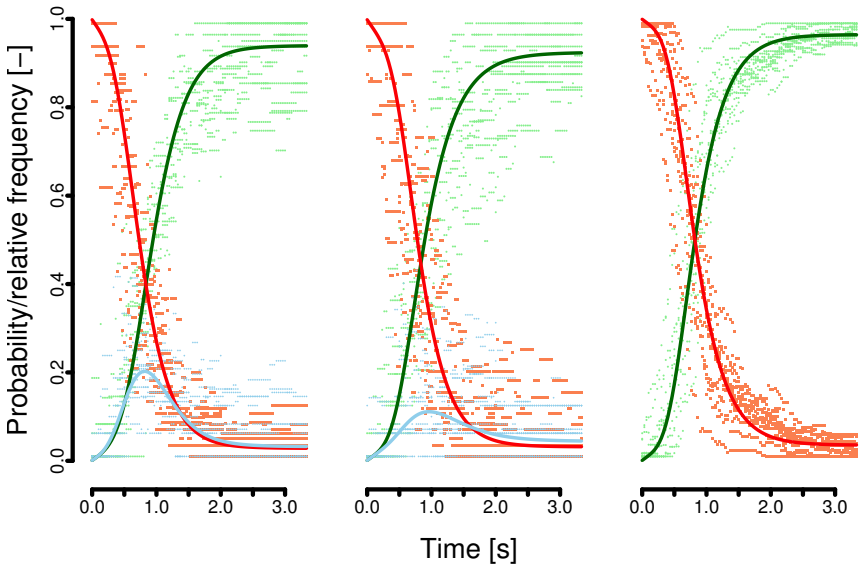

Figure 6.6.: *The plots shows the raw data of the* 13 *participants for each experimental condition: 'cohort', 'rhyme' or 'none' distractor (from left to right). The fixation proportion of the unrelated (red squares), target (green dots) and distractor item (blue diamond) and the results of the naive pooling analysis (corresponding colored lines) are displayed.*

Table 6.1.: *Results of the naive pooling approach. The parameter estimates are transformed to the original scale, e.g.* $K_{base} = e^{\log K_{base}}$. *In brackets the RSEs in [%] are displayed.*

	cohort	rhyme	none
K_{base} [1/s]	0.11 (7.4)	0.11 (8.0)	0.11 (5.9)
$K_{\text{unrel}\to\text{tar}}$ [1/s]	1.2 (19)	2.5 (11)	2.9 (3.3)
$K_{\text{unrel}\to\text{dis}}$ [1/s]	2.3 (9.5)	0.55 (16)	–
$K_{\text{dis}\to\text{tar}}$ [1/s]	5.0 (8.3)	2.5 (8.3)	–
$K_{\text{tar}\to\text{dis}}$ [1/s]	Not identifiable	Not identifiable	–
$K_{\text{tar}\to\text{unrel}}$ [1/s]	Not identifiable	Not identifiable	Not identifiable
$K_{\text{dis}\to\text{unrel}}$ [1/s]	Not identifiable	Not identifiable	–
h [–]	2.4 (6.7)	3.0 (9.9)	3.1 (4.5)
$t_{1/2}$ [s]	0.67 (9.3)	0.66 (9.2)	0.62 (3.0)
WSV σ^2 [–]	1.1	1.3	0.74

- the fixation of an object usually lasts longer as the very short sampling step, the measurement within an individual are often constant over several samplings implying high auto-correlations,
- the minimal change in the fixation proportion is always 1/#{trials}.

In conclusion, the description of such occasionally constant data with a continuous model generates auto-correlated residuals. For example, if the model over-predicts the measurement at time t_{ij}, there is increased likelihood that it will also over-predict the next measurement

at time $t_{i\,j+1}$. As a consequence, the assumption of independent residuals is violated in the underlying situation.

To exemplary investigate the impact of this violation on the parameter estimation the following approach was considered. First, an arbitrary subject and condition which seems to show auto–correlated measurements was picked (ID 14 and Condition cohort) and the model was fitted to the data (using least-squares on the logit-scale, see the first stage of the two–stage approach in the Appendix 10, page 129). As shown in Figure 6.7 the model very adequately describes the time trend of the measurements. The left plot in Figure 6.8 demonstrate the presence of auto-correlation.

In the next step the individual estimates were used to simulate a profile and create auto-correlated residuals. To create auto-correlated residuals, we incorporate a correlation matrix $P \in \mathbb{R}^{n\times n}$ into the error model

$$\begin{gathered}\boldsymbol{Y} = P \cdot (\text{logit}\, f + \boldsymbol{\epsilon}) \\ \text{with } P_{i,j} = 0 \text{ except } P_{1,1} = 1,\ P_{j,\cdot} = P_{j-1,\cdot} \cdot X_j \text{ and } P_{j,j} = 1 - X_j,\end{gathered}$$

where $P_{r,s}$ denotes the r^{th} row and s^{th} column of P, with $\boldsymbol{\epsilon} \sim \mathcal{N}_n\left(0, I_n\sigma^2\right)$ and $X_j \sim$ Binomial$(n = 1, \pi = 0.7)$, for $2 \leq j \leq n$. The probability $\pi = 0.7$ was chosen manually by visual inspection of the results. Note, the WSV matrix $P \cdot I_n\sigma^2$ is no longer diagonal.

For example, for $n = 3$ measurements and $X_2 = 1$ and $X_3 = 0$ as realizations of X_j the matrix P is given by

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Finally, the simulated auto–correlated measurements were fitted to the model—ignoring the known auto–correlation—and compared with the 'true' input profile. The results are shown in Figure 6.8. The right plots indicates that (i) the above approach is able to create realistic simulations and (ii) even with neglecting the auto–correlation in this analysis, the 'true' profile could be reproduced. However, the residuals of the simulated measurements are in average larger than the residuals in the real measurements. Due to the auto–correlation, the WSV seems to be slightly overestimated.

In conclusion, although we are aware that the measurements are auto–correlated, the above considerations and results should emphasize that the impact is negligible. Although there exists approaches to deal with such auto–correlated residuals, e.g. see Karlsson et al. [1995] or coarse–graining, i.e. an artificial thinning of the data, for simplification in the following the auto–correlation is ignored and the violation of the assumption of independent residuals is accepted.

6.3.4. The NLME Analysis

In this section a NLME analysis of the dataset is conducted with the objective to quantify the effect of the rhyme and cohort distractor and the BSV as well as the WSV. The visual inspection of the data and the results of the naive pooling approach suggest that there

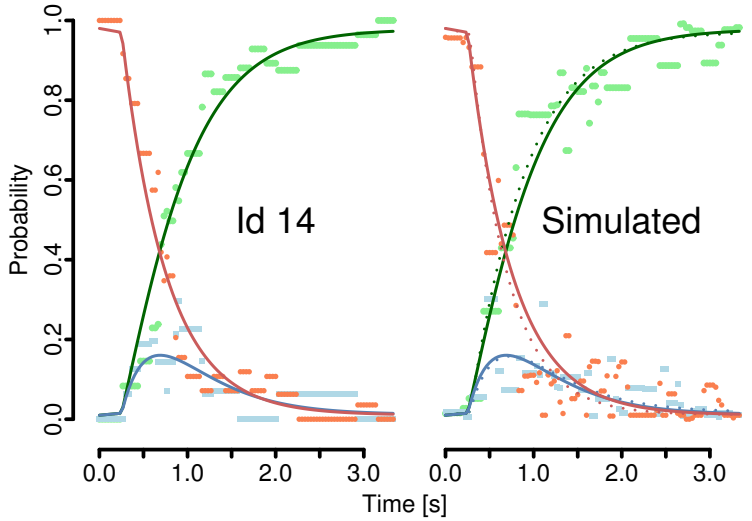

Figure 6.7.: *Autocorrelation: The left plot shows the data of subject* 14 *under condition* cohort. *The line indicates the individual least–squares fit. The right plot show the simulated auto–correlated measurements: The solid line shows the 'true' model which was used for simulation, and the dotted lines represents the fit to the simulated data ignoring the auto–correlation.*

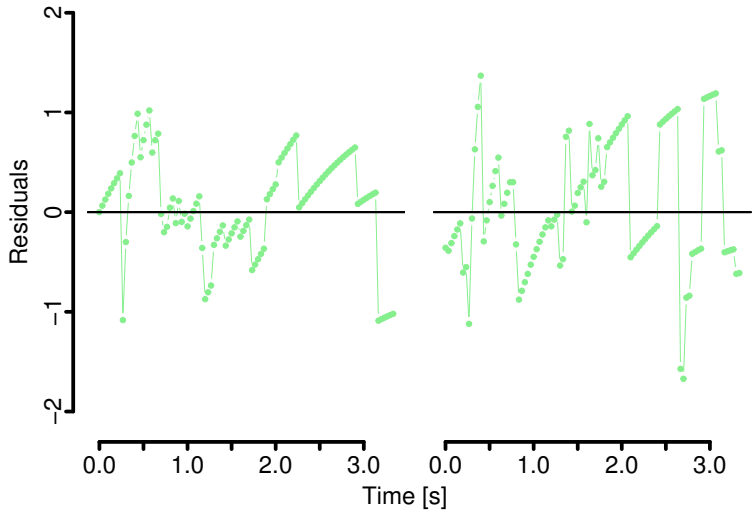

Figure 6.8.: *Residuals of the least–squares fit neglecting the auto–correlation for the target fixation of the observed data for subject* 14 *(left) and the simulated data (right).*

are differences between the different experimental conditions. Using the LRT introduced in Section 3.3 in combination with the NLME approach this question can be addressed quantitatively. To investigate whether the distractor has an effect, the dataset was analyzed under the null hypothesis H_0 'the distractor has no effect' and under the alternative hypothesis H_1 'the distractor has an effect, and the effects of cohort and rhyme differ'.

Both models were identical in the following aspects:

- Structural model $f(x_i, \theta_i)$ given by the competition model, i.e. a three compartment model, where x_i contains the sampling time points $t_1, \ldots, t_{101}$ and the different distractor conditions cond $\in \{$ cohort, rhyme, none$\}$, for $1 \leq i \leq 13$. The model is parameterized by $\log K_{base}$, $\log K_{31}$, $\log K_{32}$, $\log K_{21}$, $\log h$ and $\log t_{1/2}$. The log–transformation of the parameters was chosen to avoid non–positive parameter values.

- The random effects model is given by

$$\boldsymbol{\Theta_i} \sim \mathcal{N}_6 \left(\theta, \Omega\right),$$

  with $\theta \in \mathbb{R}^6$ and WSV $\Omega \in \mathbb{R}^{6\times 6}_{>0}$. The index i denotes the i^{th} participant, meaning the random effects are assumed on the participant's level.

- The logit–transformed measurements logit$(\boldsymbol{y_i})$ are assumed to be realizations of $\boldsymbol{Y_i}$ given by

$$\boldsymbol{Y_i} = \text{logit}\, f(x_i; \boldsymbol{\Theta_i}) + \boldsymbol{\epsilon_i},\ \boldsymbol{\epsilon_i} \sim \mathcal{N}_{n_i}\left(0, I_{n_i}\sigma^2\right),$$

  with WSV $\sigma > 0$.

- The ODE was solved by a numerical integration. In comparison to using analytical solutions this results in longer run times and larger inaccuracies in the different estimation steps—in particular for the numerical estimation of derivative terms. The convergence criterion used in the preceding EM applications (e.g. see Section 3.3) turned out to be too strict and no convergence could be achieved. However, monitoring the EM step revealed that the algorithm works fine. Thus, instead of setting a strict convergence criterion the algorithm was iterated 15 times. The potential reasons for the poorer performance of the EM algorithm for this application is discussed later in Section 6.4.

In the alternative model H_1 additional fixed effects ϑ were introduced. These additional fixed effects are correction factors Δ for the exchange rates, the Hill coefficient and the half time for the different distractors conditions:

$$k_{v\to w}(t, \text{cond}) = K_{base} + \delta(t) \cdot K_{v\to w} \cdot \Delta^{\text{cond}}_{v\to w}, \tag{6.2}$$
$$\text{for } (v \to w) \in \{(\text{unrel} \to \text{tar}), (\text{unrel} \to \text{dis}), (\text{dis} \to \text{tar})\},$$
$$k_{v\to w}(t, \text{cond}) = K_{base}, \tag{6.3}$$
$$\text{for } (v \to w) \in \{(\text{tar} \to \text{unrel}), (\text{dis} \to \text{unrel}), (\text{tar} \to \text{dis})\},$$
$$h(\text{cond}) = \theta_{\text{h}} \cdot \Delta^{\text{cond}}_{\text{h}} \text{ and } \text{t}_{1/2}(\text{cond}) = \theta_{\text{t}_{1/2}} \cdot \Delta^{\text{cond}}_{\text{t}_{1/2}}, \tag{6.4}$$

for cond $= \{$cohort, rhyme, none$\}$, h and $t_{1/2}$ defined in Equation (6.2) and we set the reference $\Delta^{\text{cond}}_{v\to w/h/t_{1/2}} = 1$ for cond = cohort. Note, for the unrelated condition the second compartment does not exists and it is $K_{32} = K_{21} = 0$.

Compared to H_0, in H_1 additional 8 fixed effects parameters were integrated, fixing these parameters to the value 1 the two models H_0 and H_1 are identical, thus H_0 and H_1 are

nested. The results of both models were compared with respect to the objective function (approximation of $-2\log\mathcal{L}$).

Based on the LRT, the null hypothesis can be rejected to a significance level of α, if the $(1-\alpha)$–quantile of the χ^2–distribution with 8 degrees of freedom is larger than $\Lambda = -2\log L(\boldsymbol{y}; H_0) - 2\log L(\boldsymbol{y}; H_1)$. For $\alpha = 1\%$ the corresponding χ^2 quantile is 26. As shown in Table 6.2, the additional parameters in H_1 lead to a reduction of 991 in the OFV. Thus, the null hypothesis 'the distractor has no effect' can be rejected. This finding is in agreement with the finding by Mirman et al. [2008] "that both cohort and rhyme competitor fixation were reliably different from unrelated fixation".

Table 6.2.: *The minimum objective function value and the WSV values of the null and the alternative model. Λ denotes minus two times the log likelihood ratio of the two models.*

	H_0	H_1
OFV	23 876	22 885
Λ	991	
σ^2 [–]	0.55	0.5

As shown in Table 6.3, incorporating the distractor effects in the analysis effectively reduces the BSV. According to the expectation, the BSV is reduced in general, except for h. This disagrees with the intuition. However, the relative difference is only 5% and we suspect that this is either due to the numerical error resulting from the EM algorithm or due to correlations within the parameter estimates.

Table 6.3.: *Reduction of the BSV: Differences in the BSV between the two models are expressed as relative change in the CV.*

	$\frac{\mathrm{CV\%}(H_0) - \mathrm{CV\%}(H_1)}{\mathrm{CV\%}(H_0)}$ [%]
K_{base}	29
$K_{\mathrm{unrel}\to\mathrm{tar}}$	51
$K_{\mathrm{unrel}\to\mathrm{dis}}$	53
$K_{\mathrm{dis}\to\mathrm{tar}}$	49
h	−5.7
$t_{1/2}$	26

In the next step we performed the model diagnostics introduced in the first part of this thesis I for H_1. All diagnostic plots reveal a robust data description by the model; due to the large number of different conditions and compartments it was decided to not included these plots here. Representatively, the VPC for the target fixation probability for the three experimental conditions is shown in Figure 6.9. The VPC outlines an adequate agreement between the simulated and observed quantiles. In particular, the typical trend of each condition is very well described by the model. However, the range of the 95 % prediction interval seems to be very large—this might be due to the small sample size of 13 participants or a hint for a over–prediction of the variability. The VPC for the other compartments look

comparable. However, the variability seems to be over–predicted. For example, this could result from an over–prediction of the WSV. As observed in the naive pooling analysis the magnitudes of the WSV differ between the three conditions; in a next an additional WSV terms for the different experimental conditions and/or the different compartment could be implemented.

Finally, we want to interpret the modeling result w.r.t. the initial questions how the distraction differs between a cohort and a rhyme distractor. The most interesting parameters are the fixed effects describing the typical probability exchange between the target, distractor and unrelated item fixation, these fixed effects are displayed in Table 6.4.

Table 6.4.: *The fixed effects of model H_1. Typical population parameters are presented on the original scale (e.g. $K_{base} = e^{\log K_{base}}$).*

	Cohort	Rhyme	None
$K_{\text{unrel}\to\text{tar}}$ [1/s]	0.76	1.7	1.8
$K_{\text{unrel}\to\text{dis}}$ [1/s]	1.3	0.56	–
$K_{\text{dis}\to\text{tar}}$ [1/s]	2.7	1.7	–
h [–]	15	3.4	27
$t_{1/2}$ [s]	0.33	0.47	0.33
K_{base} [1/s]		0.036	

The parameter estimates of the rate constant $K_{\text{dis}\to\text{tar}}$ reveal that the cohort distractor has a weaker attraction as the rhyme distractor. This finding and the estimates of $t_{1/2}$ are in agreement with the conclusion by Magnuson et al. [2003] "the rhyme effects became somewhat diminished, and they showed delayed onset compared with the cohorts". Moreover, the estimates of the rate constant $K_{\text{unrel}\to\text{tar}}$ emphasize the largest increase in target fixation is reached when no distractor is present.

In the next section we introduce a new modeling approach which is based on the same structural model but incorporates the stochastic nature of the experimental setting more appropriate.

6.3.5. The Dirichlet–Multinomial Approach

In this section a new approach which uses the original discrete data instead of the aggregated data is presented. Aggregating data, and the respective analysis leads to a possible loss of information in the analysis. E.g. imagine in the underlying study the distractor experiment was performed twice as often for a cohort as for a rhyme distractor. This fact would be neglected when using aggregated data, however, the results might differ especially for parameters which are shared in the model across the distractor conditions, e.g. the rates K_{base} or Kdt in the above model H_1. The following example illustrates the impact and thereby the weakness of analyzing aggregated data.

Assume that the probability of getting head by throwing a coin is $\boldsymbol{\pi}$ with a between coin variability $\text{Var}[\boldsymbol{\pi}]$. An experiment to estimate the expectation and the variance is per-

formed: N coins are chosen and each coin is thrown n_i times, for $1 \leq i \leq N$. The outcome is either head or tail, the number of heads are denoted by $k_1, \ldots, k_N$. Proceeding as above, the data would first be aggregated to the measurements $\boldsymbol{y} = (k_1/n_1, \ldots, k_n/n_N)^T$ and then analyzed; e.g. by estimating sample mean and variance. The results would be independent of the n_i's, meaning the results $1/5 = 0.2$ would have the same weight as $90/100 = 0.9$ which obviously depicts a strong simplification and loss of information. A more appropriate approach would also consider the size of each trial.

It is well known that the above experiment corresponds to a binomial distribution—a special case of the multinomial distribution with only two possible outcomes: head or tail. If one assumes that $\boldsymbol{\pi}$ is constant and there is no between coin variability, the distribution of $\boldsymbol{Y_i}$—the random variable describing the total number of heads in the i-th trial—is given by $\boldsymbol{Y_i} \sim \text{Binomial}(n_i, \boldsymbol{\pi})$, $1 \leq i \leq N$. However, if we expect a between coin variability we have to consider that $\boldsymbol{\pi}$ itself is a random variable taking values in $[0, 1]$. Thus, we have a mixed effects model where $p^{\boldsymbol{Y}|\boldsymbol{\pi}=p}$ is binomial distributed and $\boldsymbol{\pi}$ correspond to $\boldsymbol{\Theta_i}$, the random variable describing the BSV. Based on these assumptions, $\boldsymbol{Y_i}$ and $(\boldsymbol{Y_i}|\boldsymbol{\pi} = p)$ are discrete random variables and $\boldsymbol{\pi}$ is a continuous random variable. Furthermore, we assume that $(\boldsymbol{Y_i}|\boldsymbol{\pi} = p)$ and $(\boldsymbol{Y_j}|\boldsymbol{\pi} = p)$ are independent for $i \neq j$. Thus, the distribution of $\boldsymbol{Y_i}$ is given by the integral

$$p^{\boldsymbol{Y_i}}(\boldsymbol{k_i}; n_i) = \int_0^1 p^{\boldsymbol{Y_i}|\boldsymbol{\pi}=p}(\boldsymbol{k_i}; p, n_i) \cdot p^{\boldsymbol{\pi}}(x)\, dp,$$

for $1 \leq \boldsymbol{k_i} \leq n_i$, with $p^{\boldsymbol{Y_i}|\boldsymbol{\pi}=p}(\boldsymbol{k_i}; p, n_i) = \binom{n_i}{\boldsymbol{k_i}} \cdot p^{\boldsymbol{k_i}} \cdot (1-p)^{n_i - \boldsymbol{k_i}}$ denoting the binomial p.d.f.. There are several possibilities to chose the distribution of $\boldsymbol{\pi}$. A frequently used distribution in this situation is the beta distribution, with $\boldsymbol{\pi} \sim \mathcal{B}(\alpha_1, \alpha_2)$ and $\alpha_1, \alpha_2 > 0$. The beta distribution is a continuous distribution with support $[0, 1]$, and expectation and variance given by

$$\begin{aligned} \mathbf{E}[\boldsymbol{\pi}] &= \frac{\alpha_1}{\alpha_0} =: \bar{\pi}, \\ \text{Var}[\boldsymbol{\pi}] &= \bar{\pi} \cdot (1 - \bar{\pi}) \cdot \frac{1}{\alpha_0 + 1}, \end{aligned}$$

with $\alpha_0 = \alpha_1 + \alpha_2$, e.g. see Johnson [1995, chap. 25]. A key advantage of using the beta distribution combined with the binomial distribution is that the beta distribution is the so–called conjugate prior of the binomial distribution. I.e. compared to the situation in Section 2 the above integral has an analytical solution and the p.d.f. of $\boldsymbol{Y}_i$ is written as

$$p^{\boldsymbol{Y}_i}(\boldsymbol{k_i}; n_i, \alpha_1, \alpha_2) = \frac{\Gamma(n_i + 1)}{\Gamma(\boldsymbol{k_i} + 1) \cdot \Gamma(n_i - \boldsymbol{k_i} + 1)} \frac{\Gamma(\boldsymbol{k_i} + \alpha_1)}{\Gamma(n_i - \boldsymbol{k_i} + \alpha_2) \cdot \Gamma(n_i + \alpha_0)} \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1) \cdot \Gamma(\alpha_2)}$$

where Γ denotes the gamma function. The distribution of $\boldsymbol{Y}_i \sim \mathcal{BB}(n_i; \alpha_1, \alpha_2)$ is called beta–binomial distribution, and expectation and the variance of $\boldsymbol{Y}_i$ are given by

$$\begin{aligned} \mathbf{E}[\boldsymbol{Y_i}] &= n_i \cdot \bar{\pi}, \\ \text{Var}[\boldsymbol{Y_i}] &= \frac{\alpha_0 + n_i}{\alpha_0 + 1} \cdot n_i \cdot \bar{\pi} \cdot (1 - \bar{\pi}) = \frac{\alpha_0 + n_i}{\alpha_0 + 1} \cdot \text{Var}[(\boldsymbol{Y_i}|\boldsymbol{\pi} = \bar{\pi})]. \end{aligned}$$

Because $n_i > 0$ and $\alpha_0 > 0$, it is $\text{Var}[\boldsymbol{Y}_i] \geq \text{Var}[(\boldsymbol{Y}_i|\pi = \bar{\pi})]$.

In Table 6.5 the impact of using the aggregated vs. using the binomial or the beta–binomials model is exemplified. The example shows that the analysis of aggregated data neglects information, which most likely leads to an over-prediction of the inter–coin variability.

Table 6.5.: *Illustrative coin example: Comparing the analysis of aggregated and original scale data. The two columns Betabinomial show different results of the same experiment; aggregating the discrete data leads to identical relative frequencies, also the total number of experiments $\sum n_i$ is identical. In the first Betabinomial case, the largest trial $n_2 = 120$ shows the highest probability of getting head, while in second Betabinomial case the largest trial $n_1 = 125$ shows a much lower probability of getting head. In the Betabinomial models those differences in the trial size are considered while in the aggregated analysis each trial is weighted equally. The example illustrates the impact: in the first Betabinomial case the probability of getting head is 50% higher as in the second case. The inter–coin variability is more than doubled in the aggregated analysis.*

	Aggregated	Betabinomial	
	$\boldsymbol{k_i}/n_i$	$(\boldsymbol{k_i}, n_i)$	$(\boldsymbol{k_i}, n_i)$
y_1	0.2	$(1, 5)$	$(25, 125)$
y_2	0.6	$(72, 120)$	$(6, 10)$
y_3	0.5	$(24, 48)$	$(4, 8)$
y_4	0.25	$(1, 4)$	$(6, 24)$
y_5	0.2	$(3, 15)$	$(5, 25)$
$\sum n_i$	–	207	207
$\sum \boldsymbol{k_i}$	–	102	57
$\sum \boldsymbol{k_i} / \sum n_i$	–	0.53	0.24
$\mathbf{E}[\boldsymbol{\pi}]$	0.35	0.42	0.28
$\text{Var}[\boldsymbol{\pi}]$	0.04	0.017	0.0059

Going back to our study we have a similar situation. If no distractor is present the participant can fixate either the target or the unrelated items, this corresponds to head or tail. Under the rhyme or cohort distractor condition the participant can fixate either the target, the distractor or an unrelated item, thus one has three possible choices. The corresponding discrete distribution is given by the multinomial distribution. In this case the analog of the beta distribution is the so–called Dirichlet distribution. The Dirichlet distributions is the conjugate prior of the multinomial distribution. The multinomial distribution is a generalization of the binomial model, the Dirichlet distribution of the beta distribution and thus the beta–binomial distribution is a special case of the Dirichlet–multinomial $\mathcal{DM}$ distributions for $K = 2$. The introduced properties for the beta–binomial model apply as well to the Dirichlet–multinomial model.

The idea of our approach is now to use the original discrete data combined with the Dirichlet–multinomial model in the context of the NLME modeling approach. The WSV model $(\boldsymbol{Y_i}|\boldsymbol{\Theta_i} = \theta_i)$ of a NLME model—in the classical NLME situation given by a normal distribution—is replaced by the Dirichlet–Multinomial distribution (DirMult) distribution.

The underlying probability time course of $p(t)$ is described by the competition model developed in Section 6.3.1. Because we introduced the EM only for the situation that the WSV is described by a normal distribution, the E– and M–step have to be adapted to the new WSV model.

Adapting the EM-Algorithm

We briefly describe how the EM algorithm was adapted to be applicable in combination with a discrete WSV distribution.

In summary the parameter β comprises the fixed effects parameters θ, the associated BSV Ω and fixed effects not associated with a random effect ϑ. In the Dirichlet–multinomial model the Dirichlet parameter α_0 corresponds to the variance parameter σ^2 in the normal distribution model. In comparison to the usual M–step, the adapted update step of the WSV parameter α_0 cannot be derived in an analytically closed form. Thus, α_0 is considered as a fixed effect ϑ in the NLME context. The EM algorithm is performed analogously as introduced in Section 3.1, except for the following deviations:

- E-step: Analogously to the E–step presented in Section 3.1, we assume $\boldsymbol{Z_i} = (\boldsymbol{\Theta_i}|\boldsymbol{Y_i} = \boldsymbol{y_i})$ with

$$\boldsymbol{Z_i} \overset{approx.}{\sim} \mathcal{N}_p\left(\hat{\theta}_i(\boldsymbol{y_i};\beta),\, C^{\boldsymbol{Z_i^{LA}}}\right),$$
$$\text{with } \hat{\theta}_i(\boldsymbol{y_i};\beta) = \arg\max g(\theta_i) \text{ and } C^{\boldsymbol{Z_i^{LA}}} = \left(-\nabla^2 g(\hat{\theta}_i(\boldsymbol{y_i};\beta))\right)^{-1},$$

  with $g(\theta_i) = \log p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}(\boldsymbol{y_i};\theta_i,\vartheta) + \log p^{\boldsymbol{\Theta_i}}(\theta_i;\beta)$ and $p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}$ denotes the p.d.f. of the Dirichlet–multinomial distribution, and $p^{\boldsymbol{\Theta_i}}$ the p.d.f. of the multivariate normal distribution.

- M-step:
  (1) Update θ and Ω as introduced in Section 3.1.
  (2) Update ϑ (including the Dirichlet parameter $\alpha_0 > 0$):

$$\hat{\vartheta} = \arg\max_{\vartheta} \sum_{i=1}^{N} \log p^{\boldsymbol{Y_i}|\boldsymbol{\Theta_i}=\theta_i}(y_i;\hat{\theta}_i(\boldsymbol{y_i};\beta),\vartheta),$$

  using numerical optimization methods.

In the next section the DirMult adjusted EM algorithm is used to analysis the underlying discrete data in a NLME context. The results are compared with the associated results of the classical analysis of the aggregated data.

Results of the Mixed Effects Dirichlet–Multinomial Model

For the Dirichlet–multinomial approach we use an alternative random effects structure of a mixed effects model to address the main objective of the underlying study, namely the

quantification of the distractor effect. This alternative structure should illustrate how the same objective is addressed with different approaches using mixed effects modeling.

In the following, we assumed random effects for the experimental condition, but not for each participant; i.e. the individual participant effect is ignored here. The model we used is composed of the following parts:

- Structural model $f(x_i, \theta_i)$ given by the competition model with parameterization $k_{\text{unrel}\to\text{tar}}$, $k_{\text{unrel}\to\text{dis}}$ and $k_{\text{dis}\to\text{tar}}$ and the time delay $k_{x\to y}(t) = K_{base} + K_{x\to y} \cdot \delta(t)$. The remaining rates were set to the baseline $k_{\text{tar}\to\text{unrel}} = k_{\text{tar}\to\text{dis}} = k_{\text{dis}\to\text{unrel}} = K_{base}$.

- The random effects are $\log K_{31}$, $\log K_{21}$ and $\log K_{23}$ and the stochastic model is given by

$$\boldsymbol{\Theta_i} \sim \mathcal{N}_3\left(\theta, \Omega\right),$$

  with $\theta \in \mathbb{R}^3$ and WSV $\Omega \in \mathbb{R}^{3\times 3}_{>0}$. The baseline rate K_{base}, the Hill coefficient h and the half time $t_{1/2}$ were assumed to be condition independent and were thereby modeled as fixed effects not associated with a random effect.

- The WSV model is given by the Dirichlet–multinomial model

$$(\boldsymbol{Y_i}|\boldsymbol{\Theta_i} = \theta_i) \sim \mathcal{DM}(n_i; \alpha_i)$$

  with $\alpha_i = (\alpha_{i\,\text{tar}}, \alpha_{i\,\text{dis}}, \alpha_{i\,\text{unrel}})^T$, where $\alpha_{iv} = \alpha_0 \cdot p_v(t; \theta_i)$, $v \in \{\text{tar}, \text{dis}, \text{unrel}\}$, $\alpha_0 > 0$ and $1 \leq i \leq 3$ denoting the experimental condition cohort, rhyme or none distractor and Dirichlet parameter. For a specific time point t_j the vector $\boldsymbol{Y}_{ij} = (\boldsymbol{Y}_{ij,1}, \ldots, \boldsymbol{Y}_{ij,13})^T$ describes the result of repeating the experiment 13 times (number of participants) for the condition i.

- Due to the same reasons as stated in Section 6.3.4 we iterated the algorithm 15 times instead of setting a convergence criterion.

The results of the analysis are presented in Table 6.6. The conclusion based on the estimated parameter values is in line with the analysis of the aggregated data. The large value of α_0 indicates that $\text{Var}\left[\boldsymbol{Y_i}\right]$ is similar to the variability of the respective multinomial model without incorporating the Dirichlet level. The VPC in Figure 6.10 shows that the variability of the data varies over time; in the first phase (below one second) the model under–predicts this variability while in the latter phase the model over–predicts the variability. Unfortunately, the introduced Dirichlet–multinomial approach is not able to describe such time dependency of the variability appropriately. However, the median profiles are in good agreement. For illustration, in Figure 6.11 the observed and predicted (i.e. density) frequency of target fixation is presented for three arbitrary chosen time points. The plot emphasize the results obtained in the VPC, the medium profile is adequately described, while the variability in the early time point t_1 is under–predicted, in the time point t_2 slightly under–predicted and for the later time point t_3 the variability matches very well. Overall the variability in the data is more appropriately represented by the multinomial–Dirichlet NLME model as using a classical NLME approach on the aggregated data.

In the last step, we compared the results of the new approach with the analysis of the aggregated data using the classical NLME analysis with the same random effects structure as introduced above for the Dirichlet–multinomial approach. The parameter estimates of the classical analysis are presented in brackets in Table 6.6. The results of both approaches are very similar, emphasizing that the Dirichlet–multinomial model could successfully be applied and combined with the EM algorithm.

Table 6.6.: *Results of the NLME approach using DirMult model and the classical approach (in brackets). The transformed estimated are displayed, e.g.* $k_{base} = e^{\log k_{base}}$. *The rates were assumed to vary between the conditions (i.e. random effects); here the EBEs for each condition are presented.*

	Cohort	Rhyme	None
EBEs			
$K_{\text{unrel}\to\text{tar}}$ [1/s]	1.0 (1.1)	1.7 (2.0)	2.3 (2.4)
$K_{\text{unrel}\to\text{dis}}$ [1/s]	1.7 (1.8)	0.59 (0.5)	–
$K_{\text{dis}\to\text{tar}}$ [1/s]	3.3 (3.8)	2.0 (2.0)	–
Fixed effects			
K_{base} [1/s]	0.14 (0.08)		
h [–]	2.9 (3.5)		
$t_{1/2}$ [s]	0.54 (0.54)		
α_0 [–]	390 (–)		

6.4. Discussion and Conclusion

In this chapter we successfully applied the NLME methodology using the deterministic EM algorithm for data from a psycholinguistic experiment. We developed new aspects to take into account prior knowledge about the postulated cognitive mechanisms and the measurement of eye movements:

(i) Introduced a structural model which takes into account the competition for attraction between the target, the distractor and the unrelated item.

(ii) Introduced a new discrete WSV model which describes the underlying stochastic of the experiment more appropriately as using the aggregated data.

Based on the competition model, the model parameters can be interpreted more appropriately compared to the polynomial approach by Mirman et al. [2008]. Moreover, compared to the polynomial approach the competition model is able to describe the measurements over the full sampling period and to extrapolate beyond the measurement period. Using the competition model combined with the classical NLME analysis on the aggregated data as well as the Dirichlet–multinomial mixed effects model provided an adequate description of the data.

The analysis of the aggregated data was performed assuming a random effect on the participant's level. The integration of fixed effects to account for differences between the

experimental conditions led to a significant improvement of the model. For the Dirichlet–multinomial analysis we assumed that differences between participants can be neglected, and each participant only represents another repetition of the experiment. Here, the random effects were used to account for differences between the experimental conditions. This setting was chosen to demonstrate how the same questioning can be addressed with an alternative version of the mixed effects structure. Among others, leaving out the participant level results in absence of the auto–correlation phenomenon. Of course, if the differences between individuals are of interest the same mixed effects structure as assumed in Section 6.3.4 could be used in combination with the Dirichlet–multinomial distribution.

Howsoever, analysing the aggregated data with the 'classical' approach or the original data using the Dirichlet–multinomial approach successfully quantified the impact of the distractor item under the different experimental conditions with cohort, rhyme or none distractor present. The estimated parameter values of both approaches were in the same order of magnitude and provided similar conclusion for the study. This indicates that the adapted EM algorithm was successfully implemented for the Dirichlet–multinomial model. However, both, the classical and the Dirichlet–Multinomial approach were not able to adequately describe the time–changing variability.

As mentioned earlier, for the underlying data and the competition model the EM algorithm did not converge according to the convergence criterion we used in Section 5. We believe this is due to the accuracy of the numerical integration we used to solve the ODE system; in particular, this results in inferior accuracy of gradient and Hessian estimations which are important for the EM algorithm. Further investigation could show whether forcing a higher accuracy of the numerical ODE solver would solve this issue.

The idea to use discrete distribution in a mixed effects context is already widely applied for linear models. In this case, these discrete NLME models are called generalized NLME models, see Vonesh and Chinchilli [1997, chap. 8]. To the best of our knowledge, a non–linear mixed effects Dirichlet–multinomial model and the algorithm to derive the ML estimate have not been described yet. The successful implementation of the new methodology shows that different types of stochastic distribution can be combined with the EM algorithm and only small adaptions have to be conducted. Another new aspect which was introduced, is the use of a conjugate prior in combination with mixed effects modeling. The additional conjugate prior distribution—in our example the Dirichlet distribution—offers more flexibility to modify the shape of the discrete WSV distribution. In the underlying analysis we also tested the multinomial distribution without the Dirichlet prior, however, we observed that the estimation process was much more stable including the Dirichlet distribution which was an additional important argument to used the Dirichlet–multinomial model.

The possibilities for such a discrete mixed effects modeling approach are various: all repeated measurement experiments where the measurements are discrete or categorical and different subjects or conditions are involved. For example, in many clinical trials categorical scores are used to categorize the disease status and to observe whether this improves under treatment.

In a next step a simulation & estimation study should be conducted to investigate advantages and disadvantages of analyzing discrete data with a discrete WSV distribution vs. analyzing aggregated data with a continuous WSV distribution. Intuitively, especially in situations where number of experiments differs markedly, as illustrated in the coin–tossing example, the analysis based on a discrete WSV distribution should deliver more appropriate results as the analysis based on a continuous WSV distribution. Finally, we have to mention that the performance of the algorithm as well as the impact of the auto–correlated error has not been investigated systemically and appropriately for the underlying application. Thus, an important next step would be to follow up on this by conducting a simulation & estimation study.

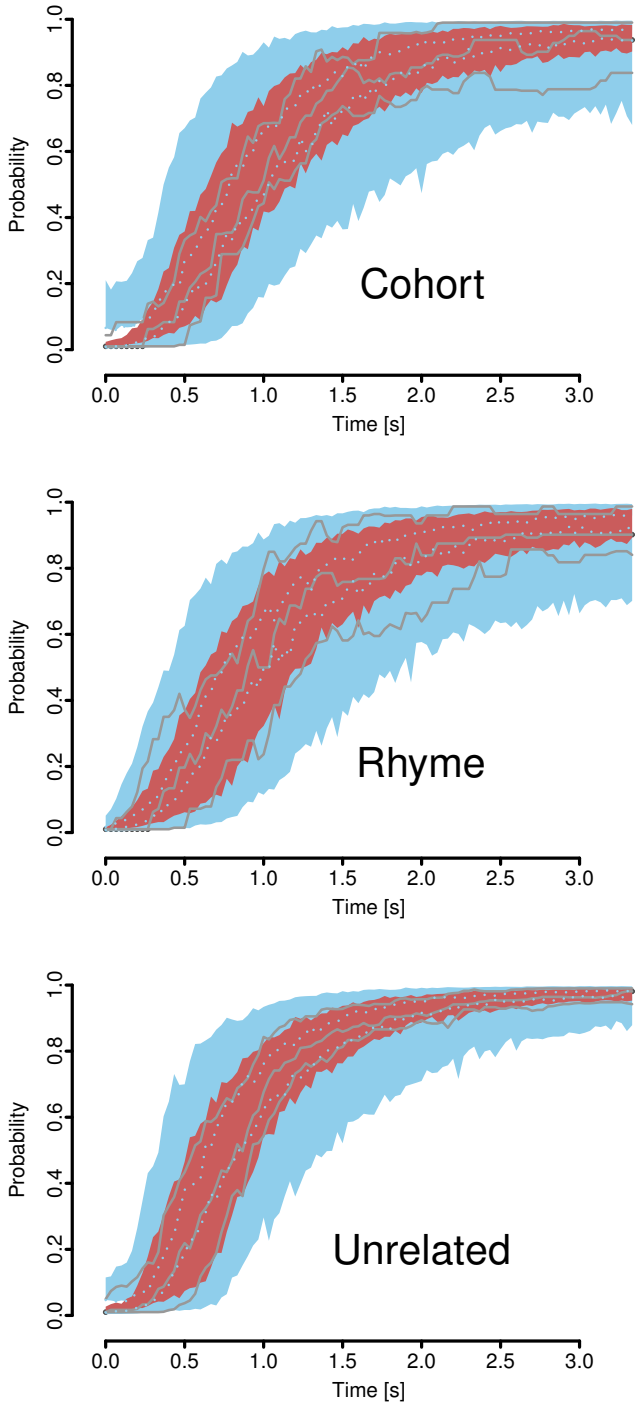

Figure 6.9.: *VPCs of the model H_1, only showing the results for the probability of the target p_{tar} for the three experimental conditions: cohort, rhyme and none distractor. The blue dotted lines indicate the boundaries of the prediction interval of the 10^{th} and 90^{th} quantiles, which overlap with the prediction interval of the median. The solid lines depict the 10^{th} and 90^{th} quantiles and the median of the measurements.*

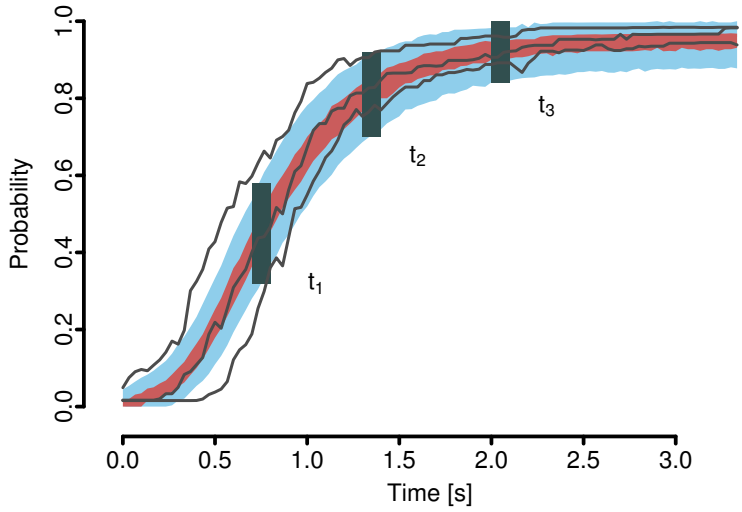

Figure 6.10.: *VPC of the Dirichlet-Multinomial model for the 'none' distractor condition a only for the target probability p_{tar}. The 'none' distractor condition was chosen because for this condition only two compartments are present and the Dirichlet–multinomial Model is similar to the beta–binomial model, the margins of this distribution are easier to illustrate. The $10^{th}, 50^{th}$ and 90^{th} quantiles are shown. The gray lines indicate the quantiles of the measurements, the shaded area of the model simulations. For three arbitrary chosen time points t_1, t_2 and t_3 cross–sections are created to illustrate the margins of the beta–binomial model on the original scale in Figure 6.11.*

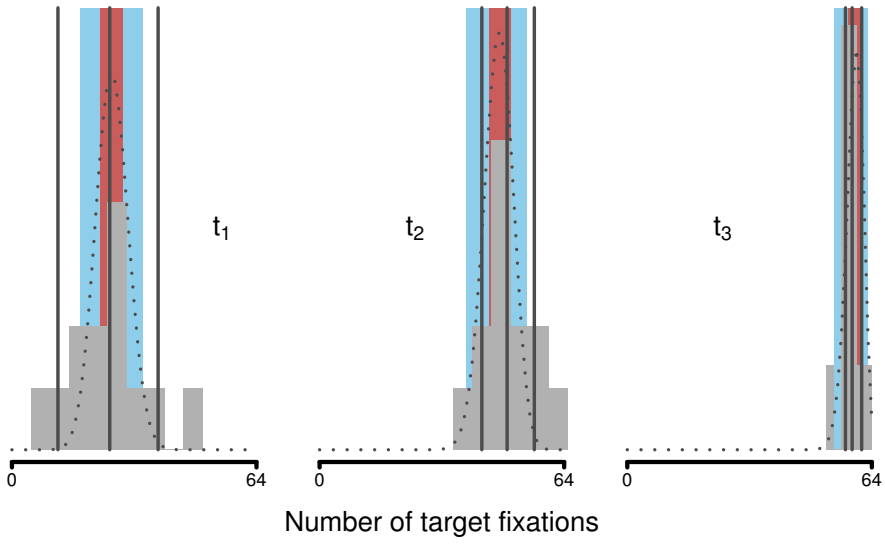

Figure 6.11.: *Cross–sections from Figure 6.10 at three different time points. Histograms represent the frequency of the measurements, vertical gray lines the $10^{th}, 50^{th}$ and 90^{th} quantiles of the measurements, the shaded areas the 95^{th} prediction intervals of the simulated quantiles.*

7. Conclusions

Finally, we want to briefly summarize each chapter including the main achievements of this thesis. In Part I we systematically derived and illustrated the Laplacian, FO and FOCE method to approximate the likelihood function of a (multivariate) NLME model. Therein, we presented a common interpretation of each approximation method which was used to outline the respective similarities and dissimilarities. Following the new interpretation, we described a deterministic EM algorithm based on the FOCE approximation to determine ML estimates of a NLME model.

In Part II, Chapter 5 we presented a new approach how mechanistic PBPK modeling can be integrated in the NLME framework to analyze population plasma PK data of the antibiotic levofloxacin using the deterministic EM algorithm. The resulting population PBPK model was used to predict ISF concentrations of levofloxacin which are the drug target—the site of infection. A comparison between the *in silico* predictions and *in vivo* microdialysis measurements in muscle and adipose tissue showed an adequate agreement. This indicates that the presented integrated approach—using plasma PK data and mechanism–driven NLME modeling to inform and optimize the PBPK model—led to an adequate *a priori* characterization of the tissue PKs of levofloxacin in humans.

The presented approach demonstrates how mechanistic PBPK models, which are usually developed in the early stage of drug development, can be used as basis for model building in the analysis of later stages, i.e. in clinical studies. As a consequence, the extensively collected and accumulated knowledge about species and drug are utilized and updated with specific volunteer or patient data. The NLME approach combined with mechanistic modeling reveals new insights for the mechanistic model, for example identification and quantification of variability in mechanistic processes. This represents a further contribution to the learn & confirm paradigm across different stages of drug development.

In Part II, Chapter 6 we introduced a new approach to analysis eye movement time–courses combining NLME and mechanistic modeling. Therein, the structural as well as the WSV model were mechanistically motivated reflecting the underlying experimental setting. The structural model takes into account the competition of the eyes' attraction by the different visual items, while the discrete Dirichlet–multinomial distribution was used as WSV model to account for the discrete nature of the measurements of the experiment. As a consequence, a simultaneous analysis of the entire data obtained from all experiments could be performed. Additionally, the mechanistic competition model reflects the underlying recognition process, facilitating an plausible interpretation of the experiment.

In summary, we successfully presented how NLME modeling and mechanistic modeling can be integrated to analyze repeated measurement data to a mechanism–driven model development based on two real world applications.

Appendix

8. Derivation of Louis' Formula

In the context of the EM algorithm a useful representation for an estimate of the observed Fisher information matrix $\widehat{I}_{\boldsymbol{Y}}(\beta) = -\nabla_\beta^2 \log p^{\boldsymbol{Y}}(\boldsymbol{Y};\beta)$ is given by the Louis' formula, see Louis [1982],

$$\begin{aligned}\widehat{I}_{\boldsymbol{Y}}(\beta) = \mathbf{E}_\beta\left[-\nabla_\beta^2 \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta)\right] \\ - \operatorname{Var}_\beta\left[\nabla_\beta \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta)\right].\end{aligned}$$

To avoid confusion due to different notation used by Louis [1982] and used by us, we considered it to be useful to include a brief derivation of the above formula. To this end, in a first step the first and second derivatives of $\log p^{\boldsymbol{Y}}$ have to be derived

$$\begin{aligned}\nabla_\beta \log p^{\boldsymbol{Y}}(\boldsymbol{y};\beta) &= \frac{\nabla_\beta p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)}{p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)}, \\ \nabla_\beta^2 \log p^{\boldsymbol{Y}}(\boldsymbol{y};\beta) &= \frac{\nabla_\beta^2 p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)}{p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)} + \left(\nabla_\beta \frac{1}{p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)}\right)\left(\nabla_\beta p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)\right)^T \\ &= \frac{\nabla_\beta^2 p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)}{p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)} + \left(-\frac{1}{(p^{\boldsymbol{Y}}(\boldsymbol{y};\beta))^2}\nabla_\beta p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)\right)\left(\nabla_\beta p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)\right)^T \\ &= \frac{\nabla_\beta^2 p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)}{p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)} - \left(\frac{1}{p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)}\nabla_\beta p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)\right)\left(\frac{1}{p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)}\nabla_\beta p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)\right)^T \\ &= \frac{\nabla_\beta^2 p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)}{p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)} - \left(\nabla_\beta \log p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)\right)\left(\nabla_\beta \log p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)\right)^T. \qquad (8.1)\end{aligned}$$

Using the same steps as above, $\log p^{\boldsymbol{Y},\boldsymbol{\Theta}}$ can be written as

$$\nabla_\beta^2 \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{\theta};\beta) = \frac{\nabla_\beta^2 p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{\theta};\beta)}{p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{\theta};\beta)} - \left(\nabla_\beta \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{\theta};\beta)\right)\left(\nabla_\beta \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{\theta};\beta)\right)^T. \qquad (8.2)$$

Furthermore, under sufficient regularity conditions, we have

$$\begin{aligned}
\frac{\nabla_\beta^2 p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)}{p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)} &= \frac{1}{p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)} \nabla_\beta^2 \int p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\theta;\beta)\, d\theta \\
&= \int \frac{\nabla_\beta^2 p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\theta;\beta)}{p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\theta;\beta)} \underbrace{\frac{p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\theta;\beta)}{p^{\boldsymbol{Y}}(\boldsymbol{y};\beta)}}_{=p^{\boldsymbol{\Theta}|\boldsymbol{Y}=\boldsymbol{y}}} d\theta \\
&= \mathbf{E}_\beta \left[\frac{\nabla_\beta^2 p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta)}{p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta)} \right] \\
&= \mathbf{E}_\beta \left[\nabla_\beta^2 \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta) \right] \\
&\quad + \mathbf{E}_\beta \left[\left(\nabla_\beta \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta) \right) \left(\nabla_\beta \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta) \right)^T \right], \qquad (8.3)
\end{aligned}$$

where $\boldsymbol{Z} := (\boldsymbol{\Theta}|\boldsymbol{Y} = \boldsymbol{y})$ and in the last step Equation (8.2) was used. An analogous calculation leads to

$$\nabla_\beta \log p^{\boldsymbol{Y}}(\boldsymbol{y};\beta) = \mathbf{E}_\beta \left[\nabla_\beta \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta) \right].$$

Thus, the second term in Equation (8.1) can be written as

$$\begin{aligned}
&\left(\nabla_\beta \log p^{\boldsymbol{Y}}(\boldsymbol{y};\beta) \right) \left(\nabla_\beta \log p^{\boldsymbol{Y}}(\boldsymbol{y};\beta) \right)^T = \\
&\quad \mathbf{E}_\beta \left[\nabla_\beta \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta) \right] \left(\mathbf{E}_\beta \left[\nabla_\beta \log p^{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y},\boldsymbol{Z};\beta) \right] \right)^T.
\end{aligned}$$

9. The Naive Pooling Approach

The naive pooling approach is a very simple approach, which can be used to estimate the average population parameters based on the observations $\boldsymbol{y_1}, \ldots, \boldsymbol{y_N}$, see Sheiner [1984]. The idea is to use the observations of all subjects pooled together and to estimate average population parameters only. Therefore, it is assumed that $\boldsymbol{y_1}, \ldots, \boldsymbol{y_N}$ are realizations of the random sample $\boldsymbol{Y} = (\boldsymbol{Y_1}, \ldots, \boldsymbol{Y_N})^T$ with

$$\boldsymbol{Y} = (\boldsymbol{Y_1}, \ldots, \boldsymbol{Y_N})^T = (f(x_1; \theta_{\text{NP}}), \ldots, f(x_N; \theta_{\text{NP}}))^T + \boldsymbol{\epsilon}, \tag{9.1}$$

with $\theta_{\text{NP}} \in \mathbb{R}^p$, $\boldsymbol{\epsilon} \sim \mathcal{N}_{N_n}\left(0,\, I_{N_n}\sigma^2_{\text{NP}}\right)$, $N_n = \sum_i^N n_i$, $I_{N_n} = \text{diag}\,(1, \ldots, 1) \in \mathbb{R}^{N_n \times N_n}$ and $\sigma^2 > 0$. This approach is useful, if the observations of the individuals are not informative enough to estimate reliable individual parameter values.

The ML estimator of θ_{NP} and σ^2_{NP} are given by

$$\begin{aligned}
\hat{\theta}_{\text{NP}}(\boldsymbol{Y}) &= \underset{\theta \in \mathbb{R}^p}{\arg\min} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \left(\boldsymbol{Y_{ij}} - f(x_{ij}; \theta)\right)^2, \\
\hat{\sigma^2}_{\text{NP}}(\boldsymbol{Y}; \theta) &= \frac{1}{N_n} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \left(\boldsymbol{Y_{ij}} - f(x_{ij}; \theta)\right)^2.
\end{aligned} \tag{9.2}$$

The inverse of the variance–covariance of the ML estimators can be approximated by the observed Fisher information, e.g. see Lehmann and Romano [2005, sec. 12.4.1] and Boos and Stefanski [2013, p. 66],

$$\begin{aligned}
I_{\boldsymbol{Y}}(\theta; \sigma^2) &= -\nabla^2_\theta \log p^{\boldsymbol{Y}}(\boldsymbol{Y}; \theta, \sigma^2) \\
&= \frac{1}{\sigma^2} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \nabla_\theta f(x_{ij}; \theta)\, \nabla_\theta f(x_{ij}; \theta)^T \\
&\qquad - \frac{1}{\sigma^2} \sum_{i=1}^{N} \sum_{j=1}^{n_i} (\boldsymbol{Y_{ij}} - f(x_{ij}; \theta))\, \nabla^2_\theta f(x_{ij}; \theta) \\
I_{\boldsymbol{Y}}(\sigma^2; \theta) &= -\frac{N_n}{2(\sigma^2)^2} + \frac{1}{(\sigma^2)^3} \sum_{i=1}^{N} \sum_{j=1}^{n_i} (\boldsymbol{Y_{ij}} - f(x_{ij}; \theta))^2 \\
&= \frac{N_n}{2(\sigma^2)^2},
\end{aligned} \tag{9.3}$$

Given a realization $\boldsymbol{y}$ of $\boldsymbol{Y}$, an estimate $\widehat{I_{\boldsymbol{Y}}}$ of the observed Fisher is given by evaluating Equation (9.3) at the ML estimates $\hat{\theta}_{\text{NP}}(\boldsymbol{y})$ and $\hat{\sigma^2}_{\text{TS}}(\boldsymbol{y}, \hat{\theta}_{\text{NP}}(\boldsymbol{y}))$ and by replacing the random variable $\boldsymbol{Y}$ by the observation $\boldsymbol{y}$.

10. The Two–Stage Approach

The two–stage approach is a simple approach to analyse repeated measurement studies. As the name already hints, the analysis is performed in two steps: first the individual parameter estimates are derived for each subject, and second, the individual estimates are described with some summary statistics. The principle of the two stage approach is among others described in Sheiner [1984], while details and the attention to detail might vary between authors.

10.1. First Stage: Determine Individual Parameter Estimates

Given the random sample $\boldsymbol{Y_1}, \ldots, \boldsymbol{Y_N}$ and under the assumption that

$$\boldsymbol{Y_i} = f(x_i; \theta_i) + \boldsymbol{\epsilon_i},\ \boldsymbol{\epsilon_i} \sim \mathcal{N}_{n_i}(0, I_{n_i}\sigma_i^2),$$

for $1 \leq i \leq N$, with $I_{n_i} = \operatorname{diag}(1, \ldots, 1) \in \mathbb{R}^{n_i \times n_i}$, $\sigma_i^2 > 0$ and $\theta_i \in \mathbb{R}^p$, the likelihood function for θ_i and σ_i^2 is given by

$$L_i : \mathbb{R}^p \times \mathbb{R}_{>0} \to [0, 1],\ L_i(\theta_i, \sigma_i^2) = p^{\boldsymbol{Y_i}}(\boldsymbol{Y_i}; \theta_i, \sigma_i^2).$$

The ML estimator of θ_i is given by

$$\begin{aligned} \hat{\theta}_i(\boldsymbol{Y_i}) &= \underset{\theta_i \in \mathbb{R}^p}{\arg\max}\ L_i(\theta_i, \sigma_i^2; \boldsymbol{Y_i}) \\ &= \underset{\theta_i \in \mathbb{R}^p}{\arg\min}\ -2 \log L_i(\theta_i, \sigma_i^2; \boldsymbol{Y_i}) \\ &= \underset{\theta_i \in \mathbb{R}^p}{\arg\min} \sum_{j=1}^{n_i} \left(\boldsymbol{Y_{ij}} - f(x_{ij}; \theta_i)\right)^2. \end{aligned} \tag{10.1}$$

Thus, the ML estimator of θ_i does not depend on the parameter σ_i^2. Given a realization $\boldsymbol{y_i}$ of $\boldsymbol{Y_i}$, the determination of the respective ML estimate $\hat{\theta}_i(\boldsymbol{y_i})$ corresponds to a classical nonlinear regression problem, e.g. see Deuflhard and Hohmann [2003, chap. 4], Bonate [2011, chap. 3]. Because f is nonlinear, the least squares minimization in Equation (10.1) can not be solved analytically and numerical minimization methods have to be applied, e.g. Gauss–Newton–type algorithms (e.g. see Deuflhard and Hohmann [2003, chap. 4.3], Bonate [2011, chap. 3]).

The ML estimator for σ_i^2 is given by

$$\hat{\sigma}_i^2(\boldsymbol{Y_i}; \theta_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} \left(\boldsymbol{Y_{ij}} - f(x_{ij}; \theta_i)\right)^2. \tag{10.2}$$

The inverse of the variance–covariance of the ML estimators for θ_i and σ_i^2 can be approximated by the observed Fisher information, e.g. see Lehmann and Romano [2005, sec. 12.4.1] and Boos and Stefanski [2013, p. 66],

$$\begin{aligned} I_{\boldsymbol{Y_i}}(\theta_i; \sigma_i^2) &= \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} \nabla_{\theta_i} f(x_{ij}; \theta_i) \, \nabla_{\theta_i} f(x_{ij}; \theta_i)^T, \\ &\qquad - \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (\boldsymbol{Y_{ij}} - f(x_{ij}; \theta_i)) \, \nabla_{\theta_i}^2 f(x_{ij}; \theta_i) \\ I_{\boldsymbol{Y_i}}(\sigma_i^2; \theta_i) &= -\frac{n_i}{2(\sigma_i^2)^2} + \frac{1}{(\sigma_i^2)^3} \sum_{j=1}^{n_i} (\boldsymbol{Y_{ij}} - f(x_{ij}; \theta_i))^2 \\ &= \frac{n_i}{2(\sigma_i^2)^2}. \end{aligned}$$

Besides, due to the properties of a ML estimator, for $n_i \to \infty$ we have $\hat{\theta}_i(\boldsymbol{Y_i}) \to \theta_i$ and $\hat{\sigma}_i^2(\boldsymbol{Y_i}) \to \sigma_i^2$, and in limit expected and observed Fisher information are equivalent, Wakefield [2013, p. 39]. Because $\mathbf{E}\left[\boldsymbol{Y_{ij}}\right] - f(x_{ij}; \theta_i) = 0$, for determining the expected Fisher information $\mathcal{I}_{\boldsymbol{Y_{ij}}}(\theta_i)$ the—as the case may be—challenging estimation of the hessian $\nabla_{\theta_i}^2 f(x_{ij}; \theta_i)$ is avoided.

In summary, in this section, the ML estimator for the individual parameters $\hat{\theta}_i(\boldsymbol{Y_i})$, the within subject variability $\hat{\sigma}_i^2(\boldsymbol{Y_i}; \theta_i)$ and the associated variance–covariance in terms of the observed Fisher information have been presented. Based on these results, the second step of the two stage approach is presented in the next section.

10.2. Second Stage

10.2.1. Within Subject Variability

The WSV quantifies the average magnitude of deviation between individual predictions and measurements. In general, it is assumed, that average deviation should be independent of the underlying subject. Thus, in a population context it is assumed that the WSV σ_i^2 are identical for all subjects

$$\sigma_1^2 = \ldots = \sigma_N^2 = \sigma^2.$$

For the additive error model this results in

$$\boldsymbol{Y_i} \sim \mathcal{N}_{n_i}\left(f(x_i; \theta_i), \, I_{n_i} \sigma^2\right), \tag{10.3}$$

for all $1 \leq i \leq N$.

Consequently, the likelihood introduced in the first stage is modified to

$$L_N : \mathbb{R}^p \times \ldots \times \mathbb{R}^p \times \mathbb{R}^{>0} \to [0, 1], \; (\theta_1, \ldots, \theta_N, \sigma^2) \to \prod_{i=1}^{N} p^{\boldsymbol{Y_i}}(\boldsymbol{Y_i}; \theta_i, \sigma^2). \tag{10.4}$$

The estimation of θ_i is independent of $\theta_j,\ j \neq i$ and σ^2. Hence, under the model (10.3) the ML estimator of θ_i is given by $\hat{\theta}_i(\boldsymbol{Y_i})$ (see Equation (10.1)).

The ML estimator of σ^2 is given by

$$\hat{\sigma}^2_{\mathrm{TS}}(\boldsymbol{Y}; \theta_1, \ldots, \theta_N) = \frac{1}{N_n} \sum_{i=1}^{N} \sum_{j=1}^{n_i} (\boldsymbol{Y_{ij}} - f(x_{ij}, \theta_i))^2, \tag{10.5}$$

with $N_n = \sum_{i=1}^{N} n_i$. If the number of observations is balanced within the population, i.e. $n_1 = \ldots = n_N$, it follows that $\hat{\sigma}^2_{\mathrm{TS}}(\boldsymbol{Y}; \theta_1, \ldots, \theta_N) = 1/N \sum_{i=1}^{N} \hat{\sigma}^2_i(\boldsymbol{Y_i}; \theta_i)$, where $\hat{\sigma}^2_i(\boldsymbol{Y_i}; \theta_i)$ are the results derived in the first stage.

Analogously to Section 10.1, the variance of $\hat{\sigma}^2$ can be approximated by the inverse of the observed Fisher information

$$\mathrm{Var}^{-1}\left[\hat{\sigma}^2_{\mathrm{TS}}(\boldsymbol{Y}; \theta_1, \ldots, \theta_N)\right] \approx \frac{N_n}{2(\hat{\sigma}^2_{\mathrm{TS}}(\boldsymbol{Y}; \theta_1, \ldots, \theta_N))^2}. \tag{10.6}$$

10.2.2. Between Subject Variability

In the two–stage approach the distribution $F^{\boldsymbol{\Theta}}$ characterizing the BSV is usually not specified explicitly but only descried using common statistical measures, like median, mean and standard deviation of the individual parameter, e.g. see FDA [1999, p. 4].

Sample mean and sample variance-covariance of $\hat{\theta}_1(\boldsymbol{Y_1}), \ldots, \hat{\theta}_N(\boldsymbol{Y_N})$ are given by

$$\begin{aligned} \hat{\theta}_{\mathrm{TS}} &:= \frac{1}{N} \sum_{i=1}^{N} \hat{\theta}_i(\boldsymbol{Y_i}), \\ \hat{\Omega}_{\mathrm{TS}} &:= \frac{1}{N} \sum_{i=1}^{N} (\hat{\theta}_i(\boldsymbol{Y_i}) - \hat{\theta}_{\mathrm{TS}})(\hat{\theta}_i(\boldsymbol{Y_i}) - \hat{\theta}_{\mathrm{TS}})^T. \end{aligned} \tag{10.7}$$

Analogously to the WSV, the magnitude of BSV is often presented as CV (see Equation (1.4))

$$\mathrm{CV\%}\left((\boldsymbol{\Theta_i})_k\right) = \frac{\sqrt{(\hat{\Omega}_{\mathrm{TS}})_{kk}}}{\left(\hat{\theta}_{\mathrm{TS}}\right)_k}, \tag{10.8}$$

for $1 \leq k \leq p$, where $(\boldsymbol{\Theta_i})_k$ denote the k-th entry of $\boldsymbol{\Theta_i}$, $(\hat{\theta}_{\mathrm{TS}})_k$ and $(\hat{\Omega}_{\mathrm{TS}})_{kk}$ represent an estimate of the expectation and variance of $(\boldsymbol{\Theta_i})_k$, respectively.

In summary, in the first step the individual parameters are determined. Based on these, in a second step, the WSV $\hat{\sigma}^2_{TS}$, the sample mean $\hat{\theta}_{\mathrm{TS}}$ representing the population average and the sample variance–covariance $\hat{\Omega}_{\mathrm{TS}}$ quantifying the BSV are determined. It is worth mentioning, that in the two–stage approach as described above each individual estimate is weighted identically.

References

B. Agoram. Perspective: Evaluating systems pharmacology models is different from evaluating standard pharmacokinetic–pharmacodynamic models. *CPT Pharmacometrics Systems Pharmacology*, 3, 2014.

P. G. Ambrose, S. M. Bhavnani, and R. C. J. Owens. Clinical pharmacodynamics of quinolones. *Infectious Disease Clinics of North America*, 17(3):529–543, 2003.

D. Andes and W. A. Craig. Animal model pharmacokinetics and pharmacodynamics: a critical review. *International Journal of Antimicrobial Agents*, 19:261–268, 2002.

K. Ball, F. Bouzom, J. M. Scherrmann, B. Walther, and X. Declèves. Comparing translational population-PBPK modelling of brain microdialysis with bottom–up prediction of brain–to–plasma distribution in rat and human. *Biopharmaceutics & Drug Disposition*, 2014.

P. Barbe, C. Darimont, P. Saint-Marc, and J. Galitzky. Measurements of white adipose tissue metabolism by microdialysis technique. *Methods in Molecular Biology*, 155:305–321, 2001.

R. J. Bauer, S. Guzy, and C. Ng. A survey of population analysis methods and software for complex pharmacokinetic and pharmacodynamic models with examples. *The AAPS Journal*, 9(1):60–83, 2007.

S. Beal, L. Sheiner, A. Boeckmann, and R. Bauer. *NONMEM 7.3.0 Users Guides*. Icon Development Solutions, Hanover, MD, 1989-2013. URL `ftp://nonmem.iconplc.com/Public/nonmem730/guides`.

S. B. Beal and L. B. Sheiner. Estimating population kinetics. *Critical Reviews in Biomedical Engineering*, 8:195–222, 1982.

S. L. Beal and L. B. Sheiner. Heteroscedastic nonlinear regression. *Technometrics*, 30(3): 327–338, 1988.

S. L. Beal and L. B. Sheiner. *NONMEM Users Guide - Part VII*, 1998.

R. Bellmann, G. Kuchling, P. Dehghanyar, M. Zeitlinger, E. Minar, B. X. Mayer, M. Müller, and C. Joukhadar. Tissue pharmacokinetics of levofloxacin in human soft tissue infections. *British Journal of Pharmacology*, 57(5):563–568, 2004.

P. Bonate. Recommended reading in population pharmacokinetic pharmacodynamics. *The AAPS Journal*, 7(2), 2005.

P. L. Bonate. *Pharmacokinetic-Pharmacodynamic Modeling and Simulation*. Springer, second edition, 2011.

D. D. Boos and L. A. Stefanski. *Essential Statistical Inference*. Springer Texts in Statistics. 2013.

R. P. Brown, M. D. Delp, S. L. Lindstedt, L. R. Rhomberg, and R. P. Beliles. Physiological parameter values for physiologically based pharmacokinetic models. *Toxicology and Industrial Health*, 13(4):407–484, 1997.

M. Brunner, H. Derendorf, and M. Müller. Microdialysis for in vivo pharmacokinetic/pharmacodynamic characterization of anti—infective drugs. *Current Opinion in Pharmacology*, 5:495–499, 2005.

Y. Cao and W. J. Jusko. Applications of minimal physiologically-based pharmacokinetic models. *Journal of Pharmacokinetics and Pharmacodynamics*, 39(6), 2012.

S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335, 1995.

G. F. Clough, P. Boutsiouki, M. K. Church, and C. C. Michel. Effects of blood flow on the in vivo recovery of a small diffusible molecule by microdialysis in human skin. *The Journal of Pharmacology and Experimental Therapeutics*, 302(2):681–686, 2002.

E. Comets, A. Lavenu, and M. Lavielle. SAEMIX, an R version of the SAEM algorithm. *20^{th} meeting of the Population Approach Group in Europe, Athens, Greence*, 2011. URL `http://www.page-meeting.org/default.asp?abstract=2173`. Abstr 2173.

R. M. Cooper. The control of eye fixation by the meaning of spoken language: A new methodology for the real–time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6:84–107, 1974.

N. G. De Bruijn. *Asymptotic Methods in Analysis*. Amsterdam: North-Holland, first edition, 1958.

G. de la Grandmaison, I. Clairand, and M. Durigon. Organ weight in 684 adult autopsies: new tables for a caucasoid population. *Forensic Science International*, 119(2):149–154, 2001.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

P. Deuflhard and A. Hohmann. *Numerical Analysis in Modern Scientific Computing: An Introduction*. Texts in Applied Mathematics 43. Springer, second edition, 2003.

G. L. Drusano. Antimicrobial pharmacodynamics: Critical interactions of 'bug and drug'. *Nature Reviews Microbiology*, 2:289–300, 2004.

G. L. Drusano, A. Louie, M. Deziel, and T. Gumbo. The crisis of resistance: Identifying drug exposures to suppress amplification of resistant mutant subpopulations. *Clinical Infectious Diseases*, 42:525–532, 2006.

A. Dubois, J. Betrand, and F. Mentre. Mathematical expressions of the pharmacokinetic and pharmacodynamic models implemented in the PFIM software, 2011. URL `http://www.pfim.biostat.fr/PFIM_PKPD_library.pdf`. Online; accessed 22-July-2014.

B. Efron and D. V. Hinkley. Assessing the accuracy of the maximum likelihood estimator: Observed vs expected Fisher information. *Biometrika*, 65(3):457–487, 1978.

M. Ellmerer, L. Schaupp, G. Brunner, G. Sendlhofer, A. Wutte, W. P., and T. Pieber. Measurement of interstitial albumin in human skeletal muscle and adipose tissue by open–flow microperfusion. *American Journal of Physiology: Endocrinology and Metabolism*, 278 (2):352–356, 2002.

EMEA. Guideline on bioanalytical method validation, 2011. URL `http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/08/WC500109686.pdf`. Online; accessed 11-March-2013.

A. Engelhardt, H. Liebermeister, T. H. Reuter, and K. Irmscher. Analyse der Wasserräume und des Lipidgehalts des menschlichen subcutanen Fettgewebes. *Zeitschrift fÃ¼r klinischer Chemie und klinischer Biochemie*, 9:356–360, 1971.

EUCAST. European committee on antimicrobial susceptibility testing: Levofloxacin—rationale for the EUCAST clinical breakpoints, version 1.5, 2007. URL `http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Rationale_documents/Levofloxacin_rationale_1.5.pdf`. Online; accessed 03-August-2014.

FDA. *Guidance for Industry–Population Pharmacokinetics*. Food and Drug Administration Center for Drug Evaluation and Research, 1999.

D. Fish and A. Chow. The clinical pharmacokinetics of levofloxacin. *Drug Disposition*, 32 (2):101–119, 1997.

H. Flanders. Differentiation under the integral sign. *The American Mathematical Monthly*, 80(6):615–627, 1973.

A. Gelman, G. O. Roberts, and W. R. Gilks. Efficient Metropolis jumping rules. *Baysian Statistics*, 5:599–607, 1994.

L. Gibiansky, E. Gibiansky, and R. Bauer. Comparison of nonmem 7.2 estimation methods and parallel processing efficiency on a target–mediated drug disposition model. *Journal of Pharmacokinetics and Pharmacodynamics*, 39(1), 2012.

J. Gill and G. King. What to do when your Hessian is not invertible. *Sociologial Methods & Research*, 33(1):54–86, 2004.

M. Gilli and E. Schumann. A note on 'good starting values' in numerical optimisation. *COMISEF Working Paper Series*, 44, 2010. URL `http://comisef.eu/files/wps044.pdf`. [Online; accessed 06-January-2014].

P. Girard and F. Mentre. A comparison of estimation methods in nonlinear mixed effects models using a blind analysis. 2005.

S. W. Guo and E. A. Thompson. Monte Carlo estimation of mixed models for large complex pedigrees. *Biometrics*, 50(2):417–432, 1994.

G. Høiseth, L. Morini, A. Polettini, A. S. Christophersen, L. Johnsen, R. Karinen, and J. Mørland. Serum/whole blood concentration ratio for ethylglucuronide and ethyl sulfate. *Journal of Analytical Toxicology*, 33:208–211, 2009.

F. Huettig, J. Rommers, and A. S. Meyer. Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychologica*, 137(2):151–171, 2011.

W. Huisinga, A. Solms, L. Fronton, and S. Pilari. Modeling interindividual variability in physiologically based pharmacokinetics and its link to mechanistic covariate modeling. *CPT: Pharmacometrics & Systems Pharmacology*, 1, 2012.

ICRP. Basic anatomical and physiological data for use in radiological protection reference values. Annals of the ICRP, ICRP Publication 89, 2002. URL `http://www.math.uni-bremen.de/riskom/pqra_ws_2004/handbuecher/ICRP_Basic_Anatomical/Annals_ICRP_32_1_277.pdf`. Online; accessed 01-August-2014.

M. Jamei, S. Marciniak, D. Edwards, K. Wragg, K. Feng, A. Barnett, , and A. Rostami-Hodjegan. The simcyp population based simulator: Architecture, implementation, and quality assurance. *In Silico Pharmacology*, 1(9), 2013.

S. Janmahasatian, S. B. Duffull, S. Ash, L. C. Ward, N. M. Byrne, and B. Green. Quantification of lean bodyweight. *Clinical Pharmacokinetics*, 44(10):1051–1065, 2005.

A. M. Johansson, S. Ueckert, E. L. Plan, A. C. Hooker, and M. O. Karlsson. Evaluation of bias, precision, robustness and runtime for estimation methods in NONMEM 7. *Journal of Pharmacokinetic and Pharmacodynamics*, 41(3), 2014.

Johnson. *Continuous Univariate Distributions—Volume 2*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., second edition, 1995.

H. M. Jones and K. Rowland-Yeo. Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development. *CPT: Pharmacometrics & Systems Pharmacology*, 2, 2013.

M. O. Karlsson, S. L. Beal, and L. B. Sheiner. Three new residual error models for population PK/PD analyses. *Journal of Pharmacokinetics and Biopharmaceutics*, 23(6):651–672, 1995.

M. O. Karlsson, E. N. Jonsson, C. G. Wiltse, and J. R. Wade. Assumption testing in population pharmacokinetic models: Illustrated with an analysis of moxonidine data from congestive heart failure patients. *Journal of Pharmacokinetics and Biopharmaceutics*, 26: 207–246, 1998.

R. Kawai, M. Lemaire, J. L. Steimer, A. Bruelisauer, W. Niederberger, and M. Rowland. Physiologically based pharmacokinetic study on a cyclosporin derivative, SDZ IMM 125. *Journal of Pharmacokinetics and Biopharmaceutics*, 22:327–365, 1994.

R. Kitzes-Cohen. Quinolones in cns infections. *Quinolone Bulletin*, 3, 1987.

M. O. Koeppe, R. Cristofoletti, E. F. Fernandes, S. Storpirtis, H. E. Junginger, S. Kopp, K. K. Midha, V. P. Shah, S. Stavchansky, J. B. Dressman, and D. M. Barends. Biowaiver monographs for immediate release solid oral dosage forms: levofloxacin. *Journal of Pharmaceutical Sciences*, 100(5):1628–1636, 2011.

M. Krauss, R. Burghaus, J. Lippert, M. Niemi, P. Neuvonen, A. Schuppert, S. Willmann, L. Kuepfer, and L. Görlitz. Using Bayesian-PBPK modeling for assessment of inter-individual variability and subgroup stratification. *In Silico Pharmacology*, 1, 2013.

M. Krauss, K. Tappe, A. Schuppert, L. Kuepfer, and L. Görlitz. Bayesian Population PBPK approach for a Physiologically realistic characterization of interindividual variability in clinically relevant populations. *PLOS One*, 10, 2015.

E. Kuhn and M. Lavielle. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 49:1020–1038, 2005.

M. Lavielle and B. Ribba. Enhanced Method for Diagnosing Pharmacometric Models: Random Sampling from Conditional Distributions. *Pharmaceutical Research*, 33(12), 2016.

E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer, third edition, 2005.

T. A. Leil. A bayesian perspective on estimation of variability and uncertainty in mechanism–based models. *CPT: Pharmacometrics & Systems Pharmacology*, 3, 2014.

D. G. Levitt and T. W. Schnider. Human physiologically based pharmacokinetic model for propofol. *BMC Anestesiology*, 5(4), 2005.

E. Limpert, W. A. Stahel, and M. Abbt. Log–normal distributions across the sciences—keys and clues. *BioScience*, 51(5):341–352, 2001.

P. Liu and H. Derendorf. Antimicrobial tissue concentrations. *Infectious Disease Clinics of North America*, 17:599–613, 2003.

X. Liu and Y. Wang. Comparing the performance of FOCE and different expectation–maximization methods in handling complex population physiologically–based pharmacokinetic models. *Journal of Pharmacokinetics and Pharmacodynamics*, 43:359–370, 2016.

T. A. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of Royal Statistical Society B*, 44(2):226–233, 1982.

A. A. Luisada, P. K. Bhat, and V. Knighten. Changes of cardiac output caused by aging: an impedance cardiographic study. *Angiology*, 31(2):75–81, 1980.

J. S. Magnuson, M. K. Tanenhaus, R. N. Aslin, and D. Dahan. The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, 132(2):202–227, 2003.

K. L. Mengersen and R. L. Tweddie. Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, 24(1):101–121, 1996.

F. Mentre and R. Gomeni. A two step iterative algorithm for estimation in nonlinear mixed effect models with an evaluation in population pharmacokinetics. *Journal of Biopharmaceutical Statistics*, 5:141–158, 1995.

J. Michot, C. Seral, F. Van Bambeke, M. Mingeot-Leclerq, and P. M. Tulkens. Influence of efflux transporters on the accumulation and efflux of four quinolones (ciprofloxacin, levofloxacin, garenoxacin, and moxifloxacin) in J774 macrophages. *Antimicrobial Agents and Chemotherapy*, 49(6):2429–2437, 2005.

D. Mirman, J. A. Dixon, and J. S. Magnuson. Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4):475–494, 2008.

M. Müller, A. de la Pena, and H. Derendorf. Minireview: Issues in pharmacokinetics and pharmacodynamics of anti–infective agents: Distribution in tissue. *Antimicrobial Agents and Chemotherapy*, 48(5):1441–1453, 2004.

U. Neckel, C. Joukhadar, M. Frossard, W. Jäger, M. Müller, and B. X. Mayer. Simultaneous determination of levofloxacin and ciprofloxacin in microdialysates and plasma by high-performance liquid chromatography. *Analytica Chimica Acta*, 463:199–206, 2002.

R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Lecture Notes Statistics, second edition, 2007.

W. K. Newey and D. McFadden. Chapter 36 large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245, 1994.

H. Nielsen and S. B. Mortensen. *ucminf: General-purpose unconstrained non-linear optimization*, 2012. URL `http://CRAN.R-project.org/package=ucminf`. R package version 1.1-3.

J. R. Nyengaard and T. F. Bendtsen. Glomerular number and size in relation to age, kidney weight, and body surface in normal man. *The Anatomical Record*, 232(3):194–201, 1992.

S. Pilari and W. Huisinga. Lumping of physiologically-based pharmacokinetic models and a mechanistic derivation of classical compartmental models. *Journal of Pharmacokinetics and Pharmacodynamics*, 37(4):365–405, 2010.

E. Plan, A. Maloney, F. Mentre, M. O. Karlsson, and J. Betrand. Performance comparison of various maximum likelihood nonlinear mixed–effects estimation methods for dose–response models. *The AAPS Journal*, 14(3):420–432, 2012.

N. Plock and C. Kloft. Microdialysis–theoretical background and recent implementation in applied life-sciences. *European Journal of Pharmaceutical Sciences*, 25:1–24, 2005.

B. Ploeger, T. Mensinga, A. Sips, W. Seinen, J. Meulenbelt, and J. DeJongh. The pharmacokinetics of glycyrrhizic acid evaluated by physiologically based pharmacokinetics modeling. *Drug Metabolism Reviews*, 33(2):125–147, 2001.

P. Poulin and P. Theil. Development of a novel method for predicting human volume of distribution at steady-state of basic drugs and comparative assessment with existing methods. *Journal of Pharmaceutical Sciences*, 98(12):4941–4961, 2009.

P. S. Price, R. B. Conolly, C. F. Chaisson, E. A. Gross, J. S. Young, E. T. Mathis, and D. R. Tedder. Modeling interindividual variation in physiological factors used in PBPK models of humans. *Critical Reviews in Toxicology*, 33(5):469–503, 2003.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372–422, 1998.

K. Rayner. Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8):1457–1506, 2009.

T. Rodgers and M. Rowland. Physiologically based pharmacokinetic modelling 2: Predicting the tissue distribution of acids, very weak bases, neutrals and zwitterions. *Journal of Pharmaceutical Scienes*, 95(6):1238–1257, 2006.

T. Rodgers and M. Rowland. Mechanistic approaches to volume of distribution predictions: understanding the processes. *Pharmaceutical Research*, 24(5):918–933, 2007.

T. Rodgers, D. Leahy, and M. Rowland. Tissue distribution of basic drugs: Accounting for enantiomeric, compound and regional differences amongst b-blocking drugs in rat. *Journal of Pharmaceutical Scienes*, 94(6):1259–1276, 2005a.

T. Rodgers, D. Leahy, and M. Rowland. Physiologically based pharmacokinetic modeling 1: Predicting the tissue distribution of moderate-to-strong bases. *Journal of Pharmaceutical Scienes*, 94(6):1237–1247, 2005b.

M. Rowland. PBPK: Where are we now and where might we be in the future? 2014.

M. Rowland and T. N. Tozer. *Clinical Pharmacokinetics and Pharmacodynamics: Concepts and Applications*. Lippincott Williams & Wilkins, fourth edition, 2011.

D. Sarisaltik and Z. Teksin. Bioavailability file: Levofloxacin. *Fabad Journal of Pharmaceutical Sciences*, 32:197–208, 2007.

R. M. Savic and M. O. Karlsson. Importance of shrinkage in empirical Bayes estimates for diagnostics: Problems and solutions. *American Association of Pharmaceutical Scientists Journal*, 11(3):558–569, 2009.

S. M. Schennach. Estimation of nonlinear models with measurement error. *Econometrica*, 72(1):33–75, 2004.

E. W. Schwilke, E. L. Karschner, R. H. Lowe, A. M. Gordon, J. L. Cadet, R. I. Herning, and M. A. Huestis. Intra- and intersubject whole blood/plasma cannabinoid ratios determined by 2–dimensional, electron impact GC-MS with cryofocusing. *Clinical Chemistry*, 55(6): 1188–1195, 2009.

V. K. Shahwal, B. K. Dubey, and M. Bhoumick. Preformulation study of levofloxacin. *International Journal of Advances in Pharmaceutics*, 1(1), 2012.

L. B. Sheiner. The population approach to pharmacokinetic data analysis: Rationale and standard data analysis methods. *Drug Metabolism Reviews*, 15:153–1715, 1984.

L. B. Sheiner and S. B. Beal. Evaluation of methods for estimating population pharmacokinetic parameters. i. Michaelis–Menten model: Routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics*, 8:553–571, 1980.

L. B. Sheiner and S. B. Beal. Evaluation of methods for estimating population pharmacokinetic parameters ii. biexponential model and experimental pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics*, 9:635–651, 1981.

L. B. Sheiner and S. B. Beal. Evaluation of methods for estimating population pharmacokinetic parameters. iii. monoexponential model: Routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics*, 11:303–319, 1983.

H. Sun and P. He. Characterization of interactions between fluoroquinolones and human serum albumin by CE–frontal analysis. *Chromatographia*, 68(11–12):969–975, 2008.

I. Tamai, J. Yamashita, Y. Kido, A. Ohnari, Y. Sai, Y. Shima, K. Naruhashi, S. Koizumi, and A. Tsuji. Limited distribution of new quinolone antibacterial agents into brain caused by multiple efflux transporters at the blood-brain barrier. *Journal of Pharmacology and Experimental Therapeutics*, 295(1):146–152, 2000.

M. K. Tanenhaus, M. J. Spivey-Knowlton, K. M. Eberhard, and J. Sedivy. Integration of visual and linguistic information spoken language comprehension. *Science*, 268:1632–1634, 1995.

N. Tsamandouras, A. Rostami-Hodjegan, and L. Aarons. Combining the "bottom-up" and "top-down" approaches in pharmacokinetic modelling: Fitting PBPK models to observed clinical data. *British Journal of Pharmacology*, 2013.

N. Tsamandouras, G. Dickinson, Y. Guo, S. Hall, A. Rostami-Hodjegan, A. Galetin, and L. Aarons. Development and application of a mechanistic pharmacokinetic model for simvastatin and its active metabolite simvastatin acid using an integrated population pbpk approach. *Pharmaceutical Research*, 2015.

K. Tunblad, M. Hammarlund-Udenaes, and E. N. Jonsson. An integrated model for the analysis of pharmacokinetic data from microdialysis experiments. *Pharmaceutical Research*, 21(9):1698–1707, 2004.

A. A. Vinks. The application of population pharmacokinetic modeling to individualized antibiotic therapy. *International Journal of Antimicrobial Agents*, 19(4):313–322, 2002.

A. A. Vinks. Population pharmacokinetic–pharmacodynamic modeling of anti–infective agents and its applications to individualized therapy. In A. A. Vinks, H. Derendorf, and J. W. Mouton, editors, *Fundamentals of Antimicrobial Pharmacokinetics and Pharmacodynamics*, pages 113–134. Springer New York, 2014.

E. F. Vonesh. A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika*, 83(2):447–452, 1996.

E. F. Vonesh and V. M. Chinchilli. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Marcel Dekker, first edition, 1997.

J. Wakefield. *Bayesian and Frequentist Regression Methods*. Springer Series in Statistics. 2013.

S. Walker. An EM algorithm for nonlinear random effects models. *Biometrics*, 52(3):934–944, 1996.

Y. Wang. Derivation of various NONMEM estimation methods. *Journal of Pharmacokinetics and Pharmacodynamics*, 34(5):575–93, 2007.

L. Wasserman. *All of Statistics*. Springer Texts in Statistics. Lecture Notes Statistics, 2004.

D. R. White, E. M. Widdowson, H. Q. Woodard, and J. W. Dickerson. The composition of body tissues (ii). fetus to young adult. *British Journal of Radiology*, 64:149–159, 1991.

S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 1938.

S. Willmann, K. Höhn, A. Edginton, M. Sevestre, J. Solodenko, W. Weiss, J. Lippert, and W. Schmitt. Development of a physiologically–based whole–body population model for assessing the influence of individual variability on the pharmacokinetics of drugs. *Journal of Pharmacokinetics and Pharmacodynamics*, 34(3):401–431, 2007.

R. Wong. *Asymptotic Approximations of Integrals*. Classics in Applied Mathematics. Academic Press, 2001.

C. F. J. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103, 1983.

S. Xu, M. Yuan, M. O. Karlsson, A. Dunne, and P. Nandy. Shrinkage in nonlinear mixed-effects population models: Quantification, influencing factors, and impact. *AAPS Journal*, 14(4):927–936, 2012.

M. Zeitlinger, P. Dehghanyar, B. X. Mayer, B. S. Schenk, U. Neckel, G. Heinz, A. Georgopoulos, M. Müller, and C. Joukhadar. Relevance of soft-tissue penetration by levofloxacin for target site bacterial killing in patients with sepsis. *Antimicrobial Agents and Chemotherapy*, 47(11):3548–3553, 2003.

M. Zeitlinger, F. Traunmüller, A. Abrahim, M. R. Müller, Z. Erdogan, A. Georgopoulos, M. Müller, and C. Joukhadar. A pilot study testing whether concentrations of levofloxacin in interstitial space fluid of soft tissues may serve as a surrogate for predicting its pharmacokinetics in lung. *Internation Journal of Antimicrobial Agents*, 29(1):44–50, 2007.

G. Zlotos, A. Bucker, M. Kinzig-Schippers, F. Sorgel, and U. Holzgrabe. Plasma protein binding of gyrase inhibitors. *Journal of Pharmaceutical Sciences*, 87(2):215–220, 1998.