Pharmacokinetic and pharmacodynamic analysis in anesthesia : a modeling odyssey

Olofsen, E.

Population Pharmacokinetic and Pharmacodynamic Analysis in Anesthesia

A Modeling Odyssey

Erik Olofsen

ISBN 978-90-827007-0-1

Cover design inspired by "*2001: A Space Odyssey*", by S. Kubrick and by A.C. Clarke

Typeset with LaTeX *via* "tbook" by T. Bronger

Printed by Puntgaaf Drukwerk, Leiden, The Netherlands

Population Pharmacokinetic and Pharmacodynamic Analysis in Anesthesia

A Modeling Odyssey

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 21 juni 2017
klokke 16.15 uur

door

Erik Olofsen
geboren te Amsterdam
in 1963

Promotor:	prof. dr. A. Dahan
Copromotor:	dr. M. Niesters
Promotiecommissie:	prof. dr. L.P.H.J. Aarts
	prof. dr. ing. P.H.C. Eilers (Erasmus MC, Rotterdam)
	prof. dr. J.M.A. van Gerven
	prof. dr. H.-J. Guchelaar
	prof. dr. C.A.J. Knibbe (St. Antonius Hospital, Nieuwegein)
	prof. dr. T.W. Schnider (Kantonsspital St. Gallen, Switzerland)
	dr. E.Y. Sarton
	dr. M. van Velzen

Contents

* E Olofsen, A Dahan, G Borsboom, G Drummond, *J Clin Monit Comput* 2015; 29:127–139

[†] E Olofsen, A Dahan, *F1000Research* 2015; 2:71

[‡] A E Olesen, E Olofsen, T Andresen, C Graversen, A M Drewes, A Dahan, *Anesth Analg* 2015; 121:1165–1175

[§] Parts have been published in E Olofsen, J W Sleigh, A Dahan, *Br J Anaesth* 2008; 101:810–821, 2008, and E Olofsen, *PAGE* 2011: Abstract #2202

[¶] E Olofsen, R Mooren, E van Dorp, L Aarts, T Smith, J den Hartigh, A Dahan, *Anesth Analg* 2010; 111:626–632

[‖] Adapted quote of HAL (http://www.imdb.com/title/tt0062622/quotes)

Chapter 1

Introduction

1.1 A Modeling Odyssey

WHEN MODELING DATA from experiments, many explicit and implicit assumptions have to be made about the deterministic and random processes that produced the data. One of the assumptions is that "unexplained" variation in the measurements is independent and identically distributed. "Independent" means that the value of a measurement is not influenced by previous (or subsequent) ones. One situation where this assumption is easily violated is with data from multiple subjects, because of differences between the characteristics of those subjects, which are the same within the data from each subject.

The collection of assumptions is referred to by the term "model". To estimate both the parameters of the deterministic part of the model and the properties of the random processes, the program NONMEM was developed in the 1970s,[78,77] and its development continues to this day.[9] NONMEM is an acronym for "NONlinear Mixed-Effects Model".[9]

From the beginning, it was called into question if the complexity, and the associated computational cost, of NONMEM was worth the effort. Warnings that model parameter estimates and/or their standard errors could be biased if assumptions were violated, were balanced by examples that the parameter estimates and their standard errors were about the same with more naive estimation methods.

Part of the modeling process with NONMEM often involves obtaining successful estimation and covariance steps, specified as $ESTIMATION and $COVARIANCE records in a control file, which starts with the promising $PROBLEM record. Even when the syntax of the model specification records is correct, these commands usually lead to an unexpected (but correct!) response from the program - hence the leaflet interpretation of a famous dialogue in the science fiction literature.[22] The estimation step typically takes a lot of computer time, with the possibility that it fails with the message that the objective function is infinite. The covariance step should output the standard errors of the estimated parameters. It also takes a lot of computer time, and it fails if it encounters a singular matrix, which means that, even with the model specified, there is still an infinite number of ways to describe the available data. The objective function is a number related to the likelihood of observing the data, conditional on the values of model parameters. Usually, and in NONMEM, it is calculated as minus two times the natural logarithm of the likelihood (discarding terms that are constant with respect to the model parameter values). So if the objective function is infinite, there is actually zero likelihood of observing the data. With a reasonable model, such a result does not

seem to make sense. To obtain standard errors of the parameter estimates, the "Hessian" and/or "cross-product gradient matrix" needs to be inverted. If such a matrix is singular, it cannot be inverted. This might be caused by the model having too many parameters. But in this situation, there is no output of a possibly large standard error indicating which parameter causes overparameterization. One reason that NONMEM might fail to give successful and/or useful estimation and covariance steps, is that the model specified in the control file contains explicit or implicit assumptions about the data that are violated. For example, by writing

$$\theta_i = \widetilde{\theta} \cdot \exp(\eta_i) \tag{1.1}$$

it will be assumed that model parameter θ_i is lognormally distributed around $\widetilde{\theta}$ across the population.[9] The lognormal distribution is unimodal, but for example a population with groups of low- and high responders could be bimodal, leading to estimation problems.

In the following chapters of this thesis, different conditions where assumptions are violated will be investigated. Most of the situations are explored using simulated data, but the characteristics of these data were based on studies in patients or volunteers. The next sections in this chapter introduce those situations by the people that first proposed solutions to handle these situations. In all cases, these solutions are given by single characteristic formulas.

1.2 The Limits of Agreement

One of the papers in Nature's Top 100[97] is called "Statistical methods for assessing agreement between two methods of clinical measurement", written by Martin Bland and Douglas Altman in 1986.[14] In this important paper, the authors explain why the correlation coefficient may not be a good measure of agreement. For example, the correlation coefficient may be high, while agreement is low, because the correlation coefficient is not sensitive to scaling factors such as a bias term. Bland and Altman proposed to calculate "limits of agreement" (LoA):

$$\text{LoA} = \bar{d} \pm 2s, \tag{1.2}$$

where $\bar{d}$ and s denote the mean and standard deviation of the differences between the measurements, respectively. The mean is the bias of one measurement method with respect to the other. The above equation looks familiar: the LoA contain 95% of the differences, if these follow a normal distribution. However, the properties of the differences are not known and have to be estimated, so Bland and Altman also showed how to calculate $\bar{d}$ and s.

Very often, data are obtained from a group of subjects with multiple paired measurements in each subject. Bland and Altman warned that in this case the differences may not be considered as statistically independent, and estimated limits of agreement could be too narrow if this is not taken into account. However, the calculations become much more complicated, for which there is no readily available software.

Therefore, in **Chapter 2**, the development of a freely available implementation in JavaScript is described, which is able to run in a Web browser. We validate the implementation by giving a formal description of both the basic and more advanced Bland-Altman comparison methods, and by using simulated data so that it can be verified

that the calculations are correct. We also study the effects caused by failing to take the presence of multiple paired measurements per subject properly into account. Because the results depend on the properties of the data, and analysis methods used, we list important items for a standard format of reporting comparison studies.

1.3 An Information Theoretic Criterion

Another entry in Nature's Top 100[97] is called "A new look at the statistical model identification", written by Hirotugu Akaike in 1974. He proposed "An Information theoretic Criterion" for model selection - AIC - which he invented while taking a seat in a commuter train according to his review of his Citation Classic.[2] In his honor, the first letter of AIC is usually pronounced as "Akaike". The criterion may be written as

$$\mathrm{AIC} = -2\log L + 2p, \tag{1.3}$$

where L is the likelihood of observing the data, and p the number of parameters of the model used to calculate L. It is often stated that when AIC is used for model discrimination it leads to "overfitting", *i.e.*, it selects models with a higher dimension than the dimension of the model that generated the data.[18] But for example with experimental pharmacometric data it may not be possible to identify the correct model, because of the complexity of the processes governing drug disposition and action. Instead of trying to find the correct model, a more useful objective might be to minimize the prediction error of drug concentrations or drug effects in subjects with unknown drug characteristics. In that case, the AIC might be the selection criterion of choice.

In **Chapter 3**, we perform Monte Carlo simulations using a model of pharmacokinetic data (a power function of time) with the property that fits with common multi-exponential models can never be perfect - thus resembling the situation with real data. AIC and AIC_c (the criterion with a correction for small sample sizes) values are calculated and averaged. The average predictive performances of the models, quantified using simulated validation sets, are compared to the means of the AICs. These simulations are also done at different levels of interindividual variability in the pharmacokinetic volume of distribution, to check that AIC remains a valid criterion under these circumstances.

1.4 The Kalman Filter

Another "Citation Classic" is on a paper by Rudolf Kalman.[45,44] It contains the mathematical foundation for another paper from the same year[43], introducing the "Kalman filter". The Kalman filter filters measurements to obtain optimal estimates of the state of the system and its uncertainty, by feeding back the difference between measurements and predictions multiplied by what is now called the "Kalman gain" K_K:[32]

$$K_K = P_K C_K^T \cdot \left(C_K P_K C_K^T + R_K\right)^{-1} \tag{1.4}$$

where the terms on the right side of the equation denote properties of the system and of the noise perturbing the system and the measurements.

In **Chapter 4**, the pharmacokinetic-pharmacodynamic (PK-PD) properties of buprenorphine transdermal patch in healthy volunteers on electroencephalographic (EEG) characteristics and pain tolerance are studied. Because the latter effect is often based on subjective measurements, electroencephalography offers a possibility to objectively quantify and track the changes in the the activity of the brain when an opioid is administered.

Usually, the pharmacokinetic and pharmacodynamic states are assumed to deterministically depend on drug administration only. However, there could be variability in the absorption rate from the patch and/or in the blood–effect-site equilibration rate. The estimates of buprenorphine's properties could be biased if such variability is not taken into account. Therefore, a population PK-PD model with stochastic differential equations was implemented in the NONMEM to analyze the PK and PD data simultaneously.

1.5 The Entropy of Permutations

In the 1940s, Claude Shannon proposed the following measure of uncertainty:[75]

$$\text{Entropy} = -\sum p_i \log p_i, \tag{1.5}$$

where p_i is the probability of event i to be observed. Shannon selected this function of the probabilities because of its desirable mathematical properties. For example, entropy is maximal if all events are equally likely to occur, and entropy is zero if none of the events are likely to occur, except one.

A frequently used electroencephalographic index of central anesthetic drug effect is the "Bispectral Index" (BIS). The algorithm that is used to calculate the BIS is not in the public domain, and is subject to repetitive updates with unknown changes. On the other hand, open source indices are often highly sensitive to artifacts due to muscle activity (*e.g.*, eye blinking).

The EEG waveform can be described as a sequence of ordinal patterns. The permutation entropy (PE) describes the relative occurrence of each of these patterns.[8] The normalized PE is high (almost maximal) when the signal has predominantly high frequencies and low (approximately 40% of maximal) when the signal consists of only low frequencies. The permutation entropy was shown to be insensitive to eye blink artifacts.[65] With high amplitude eye blinks, the low amplitude high frequency components indicative of awakeness still dominate the permutation entropy, because permutations are not dependent on amplitude *per se*, but only on amplitude rankings.

The cost of sampling EEG indices is quite independent on the amount of samples, so these can be acquired at high sampling rate. However, PK-PD model fits then show correlations between the samples. For example, if a measurement is above the line representing the model fit, there is a high probability that the next measurement is also above the model output. This represents another situation where the data are not independent, and model parameter estimates may be biased unless a Kalman filter is used. Therefore, in **Chapter 5** models using stochastic differential equations for the state of the brain are studied using simulated and experimental data.

1.6 Location, Location, Location, of the Sampling Site

About 40 years ago Lewis Sheiner[79] (but see also Hull[41] and Segre[74]) pointed out that it is important to take into account a delay between drug concentration in the arterial blood and the effect-site:

$$\frac{dCe(t)}{dt} = k_{e0} \cdot (C_b(t) - C_e(t)) , \tag{1.6}$$

where k_{e0} is the blood–effect-site equilibration rate. This equation is able to describe, for example, the phenomenon of increasing drug effect with decreasing blood concentration - namely for as long as $C_b(t)$ is higher than $C_e(t)$. In pharmacokinetic-pharmacodynamic modeling studies, venous plasma samples are sometimes used to derive pharmacodynamic model parameters. In principle, the same equation can be used, but because there is an arterio-venous delay, k_{e0} will be estimated with a bias towards a smaller value.

In **Chapter 6**, the extent of arteriovenous concentration differences of morphine-6-glucuronide is quantified based on arterial and venous blood samples in volunteers. An extended pharmacokinetic model is described with standard compartments for the arterial data that are linked to additional compartments for the venous data. The extent of bias in pharmacodynamic model parameter estimates is explored *via* simulation studies with NONMEM. Furthermore, simulations are presented where a pharmacokinetic model based on arterial data is connected to a pharmacodynamic model based on venous data, to assess the influence of this mismatch on the predicted effect.

1.7 NONMEM and Beyond the Infinite

As may be expected, violations of modeling assumptions may have just mild but also significant effects on the data analysis results. As may also be expected, violations may also have significant modeling effects that are perhaps not expected in advance. Therefore, in **Chapter 7**, the findings of the five studies that are presented in this thesis will be revisited in detail. The existence of increasingly more powerful computers and increasingly more sophisticated software like NONMEM is invaluable for studies like those described, because it provides the means to look one or more dimensions deeper than the present state-of-the-art.

Chapter 2

Improvements in the Application and Reporting of Advanced Bland-Altman Methods of Comparison*

IN MANY FIELDS OF SCIENCE, two different methods of measurement may require comparison. Altman and Bland[4,13] developed a comparison method which paired the difference between measures taken using two devices to the mean of the two values, thus avoiding mathematical traps and statistical misconceptions associated with previous methods of comparison.[58] Their original method was attractively simple, intuitive, and required very little calculation. It became very well-known and widely cited. After the method was presented in a short article in a medical journal,[14] the paper has now been cited more than 22,000 times. Initially, interest was placed on measuring the bias between the two measures; in later papers by Bland and Altman, the assessment of agreement between methods was emphasized. If the "limits of agreement" (LoA) were acceptable, then one method of measurement might be interchanged with another. This decision of "acceptability" is in essence a value judgement tailored to the specific comparison, and emphasizes the fundamentally largely qualitative features of this method of analysis.[23] In many studies, remarkably wide limits appear to be accepted, without employing pre-defined criteria for judgement, or considering the consequences of accepting poor correspondence of the measures, such as a wrong decision or a mis-diagnosis.[51]

Bland and Altman extended their method to allow analysis of more complex data structures, such as repeated measures taken from a number of subjects, where variation can be caused by both within and between subject factors.[15,12] In their original examples, they compared pairs of measurements where each pair was from a separate subject. An example might be hemoglobin concentration in a blood sample.[80] In the later papers, they considered subjects in whom repeated pairs of measurements were made. These measurements could be made under two conditions. In one condition, the measurements might be of a feature that could be reasonably expected to remain the same over the period of study (*i.e.*, the true value remains constant). In another circumstance, the quantity measured could be of a feature that could easily change between each sampling occasion, for example cardiac output before and during treatment, termed "true value varies".[15] These additional features require different procedures for

* E Olofsen, A Dahan, G Borsboom, G Drummond, *J Clin Monit Comput* 2015; 29:127–139

analysis which take into account the type and structure of the data. Inappropriate analysis can substantially overestimate the agreement between methods.

Not only do such specific features of the data affect the LoA, they also influence the precision with which the "limits of agreement" can be estimated, that is the confidence intervals (CI) with which the LoA can be defined. Clearly, substantial CI for the LoA will degrade their value in comparing measurement methods. To allow adequate evaluation of the results of comparison studies, reports should not only provide the limits of agreement, but also the precision of these estimates. Bland and Altman observe that although CI for the LoA are necessary, they are rarely provided in reports.[11,13] Approximate methods for estimating the CI are simple,[11] but when the data structure is more complex, simple approximations are inappropriate. Thus with greater complexity of data used in comparison studies, papers not only fail to report exactly how the comparisons were calculated, but also signally fail to present sufficient data, such as CI for the estimated LoA, to allow adequate conclusions.[51]

We suggest that these deficiencies observed in comparison studies may largely result from three related factors. First, each form of data requires a different procedure to calculate the LoA and the CI of this estimate. Second, these procedures are not accurately defined, since most of the source papers describe several methods that may not be explicitly identified when the paper is cited. Third, the software available for calculating these descriptive statistics only implements the simpler methods, which are inappropriate for more complex data.

Since this general method is now very widely used, we set out in this paper to provide a freely available and formal means to estimate the LoA and the associated CI of the estimated values. This could improve substantially the reporting of comparison data. To use the routines, the analysis requires the exact data structure to be defined. The analysis procedure could be reported when the results are presented. Because we detected some errors, possibly typographical, in the extant descriptions of the procedures used, the exact procedures used for each calculation will be provided for scrutiny. Asymptotically valid assumptions may lead to inaccurate results with real data sets, so the most robust methods available are provided. Suggestions are given for a standard format of reporting that would improve analysis and interpretation of comparison studies.

2.0.1 Definitions and Notation

Let measurements obtained from a measurement device be called X. If these are obtained from $i = 1, \cdots, n$ subjects, we denote that by adding a subscript and write X_i. If there are m_i measurements from subject i we add a second subscript and write X_{ij}, with $j = 1, \cdots, m_i$. In the following, if the subscripts ij are omitted, all possible values of i and j are implied.

A model for the measurements X is

$$X_{ij} = T_{xij} + E_{xij},$$

i.e., the measured values X_{ij} are the sum of true values T_{xij} and measurement errors E_{xij}. To indicate explicity that there may be a systematic shift in the measurements with respect to the true values, we could refine the definition and write

$$X_{ij} = T_{xij} + B_x + E_{xij},$$

where B_x is defined as the bias and the mean of E_x as zero. However, it may be that the the bias is not the same for each individual, and therefore we refine the model as

$$X_{ij} = T_{xij} + B_x + I_{xi} + E_{xij},$$

and take the random variables I_x and E_x as statistically independent and with means of zero. The variances of I_x and E_x are denoted by σ^2_{xI} and σ^2_{xw}, respectively. These are the between-subjects (subscript I) and within-subject (subscript w) variances of the differences of the measurements and the true values.

Bland and Altman proposed to look at the differences between two sets of measurements, X and Y:

$$D_{ij} = X_{ij} - Y_{ij} = T_{ij} + B + I_i + E_{ij},$$

with $T_{ij} = T_{xij} - T_{yij}$, and similar expressions for B, I, and E. Variances of (independent) sums of differences are additive, so $\sigma^2_{dI} = \sigma^2_{xI} + \sigma^2_{yI}$ and $\sigma^2_{dw} = \sigma^2_{xw} + \sigma^2_{yw}$. The subscript d denotes that these variances are associated with the differences D. The variance of the differences is $\sigma^2_d = \sigma^2_{dI} + \sigma^2_{dw}$. A Bland-Altman plot shows the differences *versus* the means $M_{ij} = (X_{ij} + Y_{ij})/2$.

In a single individual, repeated measurements may be made of a quantity that can reasonably be assumed to have a constant "true" value, over the time span of the measurements, such as bone length in a single day. We define this as the "true value constant" condition. In such circumstances, there is no specific requirement to have, in each subject, paired simultaneous measurements made by two methods. On the other hand, many quantities can vary substantially over a short time, especially when manipulated using some procedure. Examples include heart rate, cardiac output, and breathing frequency. Here the true value is varying: measurements of such quantities using two methods are often intended to capture the changing values: this is the "true value varies" situation, and requires paired measurements for accurate analysis. The differences between the true values, T_{ij}, are only known when these are zero, *i.e.*, when $T_{xij} = T_{yij}$. This will be the case when the sampling times associated with xij and yij are identical. In other words, when the data are paired, the true values T_{xij} and T_{yij} may be varying within an individual, as their differences will be zero. When the experimental conditions are kept the same when taking measurements from each individual, T_{xij} and T_{yij} may be assumed to be constant and equal within each individual: this is the "true value constant" situation.

Bland and Altman defined the limits of agreement (LoA) as

$$\text{LoA} = B \pm 1.96 \cdot \sigma_d,$$

where B is the overall bias and σ_d is the standard deviation of the differences. This definition states that 95% of the differences lie within the limits of agreement, assuming 1) the differences are normally distributed, and 2) the parameters B and σ_d are known exactly. In practice these have to be estimated.

2.0.2 Estimation Methods for the Limits of Agreements

Bland and Altman showed how to estimate the limits of agreement for the following situations:[15,12]

- The True value varies method, which may be used if the measurement data are

paired, so that varying true values within the repeated measurements cancel due to the fact that differences are analyzed.

- The True value constant method, which may be used if the true value is for all practical purposes constant within the repeated measurements, and hence these do not have to be paired.

- The Pooled data method, which ignores information about measurements possibly originating from different subjects, so the method assumes that between-subjects variability is absent, or that the subject-specific biases in the two measurement devices are identical ($I_{xi} = I_{yi}$).

2.0.3 Confidence Intervals for the Limits of Agreements

The importance of confidence intervals (CI) for the limits of agreement has been emphasized by Bland and Altman and other authors (see e.g., [39,51,94]) The 95% confidence intervals for the limits of agreement should contain the true limits 95% of the times that these are estimated. Until the paper by Zou, [100] equations to compute these were only available for the Pooled data method and for the True value constant method when there are equal numbers of measurements per individual. [15] Zou derived equations for the True value varies method and generalized the True value constant method for unequal numbers of measurements per individual. Furthermore, Donner and Zou [28] developed the MOVER (Methods of Variance Estimates Recovery), which gives a set of equations for more accurate confidence intervals.

2.1 Methods

2.1.1 Implementation

We implemented the analysis methods described by Bland and Altman, [12] based on ANOVA, in JavaScript, [54] and created a library with functions that expand the Array object. These functions perform statistical operations, such as calculating the sums of squares necessary for the ANOVA tables. Functions that involve percentiles of distributions and correlation coefficients call the jStat library. [95] The Flot library [72] provides functions to create graphs. To call these libraries from a HyperText Markup Language (HTML) file, the jQuery library [86] is used.

For all Bland-Altman methods, we derived expressions for 95% confidence intervals of the limits of agreement (see Appendix), using the approach of Bland and Altman [15], and implemented these in the library. We modified the True value varies method for the unbalanced case, *i.e.*, where the number of measurements is not equal for each individual. Bland and Altman base the analysis on an ANOVA table, [12] because that is easily accessible. However, as Thomas and Hultquist note, [87] the subject partial sum of squares is, with unbalanced data, only X^2 distributed if there is no between-subjects variance; they also show an alternative sum of squares that has more useful properties in the presence of large between-subjects variability (see Section 2.A.6). Depending on the ratio of the between-subjects and the total variability, our implementation selects the standard or the modified True value varies method automatically. In the following,

when we refer to the "true value varies method" we refer to this automatic method unless we explicitly state otherwise.

The derived expressions for the 95% confidence intervals are exact only asymptotically, so the MOVER[100] and a bootstrap method[29] (see Section 2.A.3) were implemented to assess the utility of the different methods with finite data.

An HTML file to be accessed by a Web browser was constructed, with a box into which input data can be pasted, and a button to start the available Bland-Altman analyses. Output plots and tables with parameter estimates are then created, along with guidance regarding the meaning of these output values. In addition, a table of the data is displayed to allow a check that these have been input correctly.

2.1.2 Output of the Analyses

From the differences (the $X - Y$ data), the bias B and the limits of agreement are calculated. The bias, or grand mean of the differences, is identical for all analysis methods. The standard error of the bias is calculated according to the analysis method, and may therefore differ between methods. With the Pooled data method, the standard error of the bias is usually smaller than with the other methods if there is between-subjects variation in the bias. For the bias and the LoA, 95% confidence intervals are estimated using the derived equations and the bootstrap. The data are plotted with the estimated bias, LoA, and the 95% confidence intervals of the bias and LoA.

For the True value varies and True value constant methods, the within-subject variance and the between-subjects variance are given; the latter with its associated F and P values. When the between-subjects variance of the bias is small, the P value will be high, and results of the different methods similar. The ratio of between-subjects variance and the total variance is given as τ. A value of τ greater than zero indicates the need to take between-subjects variability into account.

Spearman's rank correlation coefficient ρ between the size of the bias and the mean values is given with its associated P value. A significant correlation suggests the need for a more sophisticated approach than described in this paper.[15]

We will not study the statistical properties of F and ρ. Significance levels for hypothesis tests based on these statistics may be influenced by between-subjects variability or unbalanced data, but we consider this topic outside the scope of the present paper.

2.1.3 Diagnostic Plots

In the developed Web application, the analysis consists of pasting the data, and just one click on a button. Before accepting the results, a necessary step is to inspect diagnostics plots to ensure that the assumptions underlying the analysis are not violated. Quantile-quantile plots of the residuals and the individual means are presented to check for their normality. The quantiles of the differences and of the individual means are plotted *versus* the quantiles of the normal distribution. If these lie close to the line of identity, the distributions of the differences and of the individual means may be approximated by normal distributions. Outliers may be detected if there are a few points far from the line of identity. Furthermore, the residuals should be independent of the means of the observations and the subject identification numbers. Plots of the (studentized)

residuals are presented as another tool to check for outliers. See the papers by Bland and Altman[4,12] for further discussion.

2.1.4 Validation Study

We applied the three Bland-Altman methods to simulated data sets, generated using a function in our JavaScript library (see Section 2.A.8). The bias was set to zero, and the total variance to one. We adjusted the proportion of the between-subjects variance with respect to the total variance, as was done by Hamilton and Lewis[38] and Zou.[100] The number of measurements per subject was varied. We used a pattern of either "5,5,5" or "1,5,9", to assess the dependence of the LoA on balanced *versus* unbalanced data. The pattern was repeated to obtain a number of subjects of 6, 12, 24, 48, and 96.

From 10,000 Monte Carlo simulations, *i.e.*, repeated data generation and analysis steps), sample means of estimated variances were computed to be compared with sample variances of the estimated parameters themselves. Converging agreement between these estimates as the number of subjects increases would indicate that the derived expressions are correct. Coverage frequencies were computed to show how often the estimated confidence intervals contain the z-value of 1.96 (as the true distribution has mean and variance of exactly zero and one, respectively).

2.1.5 Application

An example data file was created (using a C++ program; see Section 2.A.8) where the measurement of some quantity was simulated 15 times in 20 subjects. The measurements X and Y were paired. The measurements X had a mean bias of $B = 0.5$ units, relative to the Y values. The within-subject and between-subjects variances were set to $\sigma^2_{dw} = 0.2$ and $\sigma^2_{dI} = 0.5$, respectively. The mean of the true value was set to 6 and and its within-subject variance to 2.25. With these parameter values, the simulated data have characteristics similar to the data described by Biancofiori *et al.*[10] The data are shown in Figure 2.1. The variation of the dashed lines reflect between-subjects variability; the variation of the data around the dashed lines indicate within-subject variability. The data were analyzed using the Pooled data method and the True value varies method.

2.2 Results

2.2.1 Validation Study

Table 2.1 presents the Monte Carlo simulation results for the True value varies method with 12 subjects, m_i pattern 1,5,9, and total number of observations 60. The means of the estimated variances of the differences are close to the actual value of unity for all values of τ, showing that this method gives accurate estimates of the limits of agreement (the LoA are directly related to the variance of the differences). The means of the estimated variances of the bias are similar to the actual variances, showing that the derived expressions for the variance of the bias are valid. The means of the estimated variances of the limits of agreement are also similar to their actual variances, which indicates that the derived expressions for the variance of the limits of agreement are also valid.

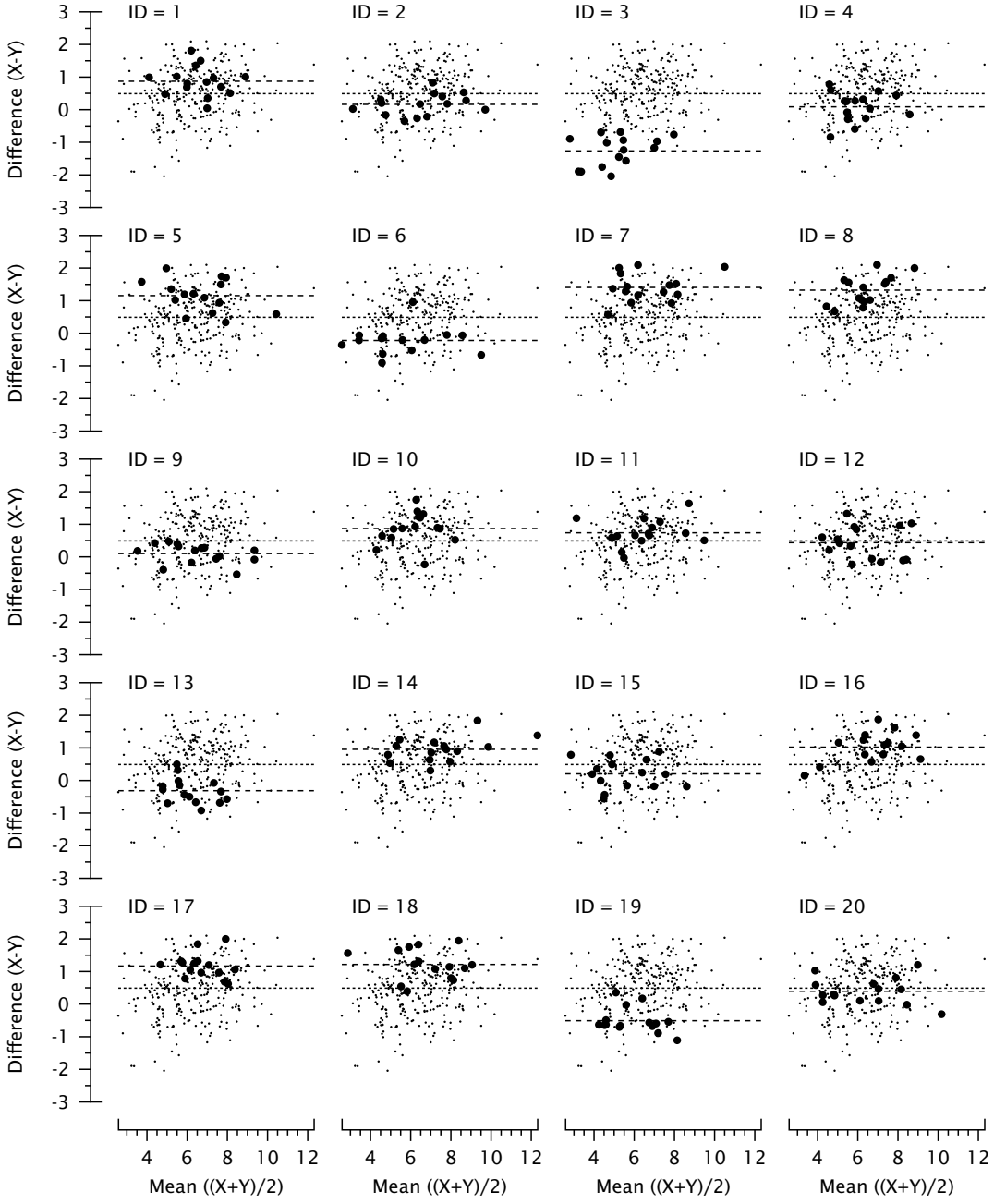

Figure 2.1: Simulated data for the example Bland-Altman analysis. *Small dots* denote all data; the *large dots* denote the data for each indicated subject. *Dotted lines*: overall bias; *dashed lines*: subject-specific bias.

Table 2.1: Means of estimated variances are shown for differences, and means of estimated variances and actual variances are shown for the bias and the limits of agreement (LoA). Because the limits of agreement are symmetric, values for their variances should be approximately the same.

	Variances of:	Bias		LoA		
τ	differences	Theoretical	Actual	Theoretical	Lower limit	Upper limit
0.1	1.001	0.027	0.028	0.065	0.066	0.066
0.3	1.002	0.046	0.047	0.107	0.109	0.109
0.5	0.995	0.058	0.061	0.150	0.150	0.156
0.7	0.998	0.069	0.070	0.191	0.194	0.188
0.9	0.999	0.079	0.078	0.234	0.227	0.235

Table 2.2 presents the coverage frequencies associated with the estimated 95% confidence intervals from the True value varies method. The coverage frequencies are less than they should be, so the CI are too narrow. Because the variances of the limits of agreement are well estimated (see Table 2.1, this is caused by a violation of the assumption that the standard deviation of the limits of agreement is normally distributed (see Section 2.A.3). Because there are differences in the one-sided coverage frequencies for the lower and upper sides, while these should all be 0.975, better CIs would be asymmetric. Table 2.3 presents coverage frequencies associated with the parametric bootstrap-t confidence intervals. In comparison with the results in Table 2.2, these are closer to the desired values. Table 2.4 presents coverage frequencies associated with MOVER confidence intervals. which are even closer to the desired values.

With balanced data (pattern of 5,5,5), the properties of the standard and modified True value varies methods are similar. With unbalanced data (pattern of 1,5,9), the standard True value varies method underestimates variability at large values of τ; the modified True value varies method underestimates variability at small values of τ (data not shown). With both the True value varies and the True value constant methods and 96 subjects, the two-sided coverage frequencies are between 0.93 and 0.95, and the one-sided coverage frequencies showed a small asymmetry. When the true value is constant, the True value constant method and the True value varies method have identical statistical properties (see Section 2.A.7), so no further numerical values are presented here for the True value constant method.

Tables 2.5 and 2.6 present the simulation results for the Pooled data method. In the presence of between-subjects variability ($\tau > 0$):

- the variance of the differences is underestimated;
- the variance of the bias is seriously underestimated (the means of the estimated variances of the bias are less than the actual variance);
- the variance of the limits of agreement is seriously underestimated (the means of the estimated variances of the LoA are less than their actual variance);
- the coverage frequencies are below 0.95 for the two-sided and mostly below 0.975 for the one sided confidence intervals.

Table 2.2: Coverage frequencies for the two-sided asymptotic 95% confidence intervals and one-sided asymptotic 97.5% confidence intervals obtained by the True value varies method. Because the limits of agreement are symmetric, values for corresponding coverage frequencies should be approximately the same. With 10,000 simulations the estimation error for frequency 0.95 is approximately 0.0002.

			Lower limit		Upper limit	
τ	Lower limit	Upper limit	Lower side	Upper side	Lower side	Upper side
0.1	0.931	0.934	0.943	0.988	0.991	0.944
0.3	0.919	0.923	0.926	0.994	0.993	0.930
0.5	0.906	0.905	0.910	0.996	0.996	0.909
0.7	0.905	0.907	0.909	0.997	0.997	0.910
0.9	0.915	0.913	0.917	0.998	0.997	0.915

Table 2.3: Coverage frequencies for the bootstrap-*t* confidence intervals; see Table 2.2 for explanation.

			Lower limit		Upper limit	
τ	Lower limit	Upper limit	Lower side	Upper side	Lower side	Upper side
0.1	0.937	0.941	0.963	0.974	0.977	0.964
0.3	0.931	0.937	0.962	0.969	0.973	0.965
0.5	0.934	0.929	0.964	0.970	0.968	0.961
0.7	0.942	0.937	0.967	0.975	0.976	0.961
0.9	0.951	0.952	0.978	0.974	0.976	0.976

Table 2.4: Coverage frequencies for the MOVER confidence intervals; see Table 2.2 for explanation.

			Lower limit		Upper limit	
τ	Lower limit	Upper limit	Lower side	Upper side	Lower side	Upper side
0.1	0.952	0.950	0.980	0.972	0.973	0.977
0.3	0.947	0.948	0.970	0.978	0.976	0.971
0.5	0.946	0.939	0.962	0.984	0.981	0.958
0.7	0.945	0.949	0.963	0.983	0.985	0.964
0.9	0.958	0.951	0.970	0.987	0.984	0.967

Table 2.5: Simulation results for the Pooled data method; see Table 2.1 for explanation.

	Variances of:	Bias		LoA		
τ	Differences	Theoretical	Actual	Theoretical	Lower limit	Upper limit
0.1	0.989	0.016	0.027	0.049	0.060	0.061
0.5	0.945	0.016	0.068	0.047	0.143	0.137
0.9	0.905	0.015	0.110	0.045	0.282	0.282

Table 2.6: Coverage frequencies for the asymptotic confidence intervals obtained by the the Pooled data method; see Table 2.2 for explanation.

			Lower limit		Upper limit	
τ	Lower limit	Upper limit	Lower side	Upper side	Lower side	Upper side
0.1	0.914	0.911	0.934	0.980	0.980	0.931
0.5	0.713	0.727	0.792	0.921	0.930	0.797
0.9	0.532	0.527	0.653	0.878	0.882	0.645

2.2.2 Application

The upper panel in Figure 2.2 shows the Bland-Altman analysis of the simulated data with the True value varies method. The ratio of between-subjects variance and total variance $\tau \approx 0.73$. The bias is estimated as 0.498 with bootstrap 95% CI 0.171 to 0.830. The limits of agreements are -1.12 to 2.11 with 95% CIs -1.77 to -0.70 and 1.69 to 2.77, which are asymmetric as can be seen in the figure. The asymptotic 95% CI are 0.187 to 0.809 for the mean, and are -1.61 to -0.62 and 1.62 to 2.60 for the limits of agreement.

The lower panel in Figure 2.2 shows the Bland-Altman analysis of simulated data with the Pooled data method. The bias was estimated as 0.498 with asymptotic CI 0.406 to 0.589. The limits of agreements were -1.09 to 2.08 with CIs -1.24 to -0.93 and 1.93 to 2.24. Note that the application of the Pooled data method is inappropriate here, and gives limits of agreement that are too narrow and confidence intervals that are much too narrow. The MOVER and parametric bootstrap-*t* confidence intervals would be inaccurate too, because these would be based on inaccurate estimates of variability.

2.3 Discussion

2.3.1 The Methods of Comparison

We found that the modified True value varies method has statistical properties identical to the True value constant method, under the condition that the true value is constant. That means that the modified True value varies method may be used both when the true value is varying and when it is constant. The True value constant method is only needed if the number of measurements from two devices is not equal in one or more individuals. However, if the true value is not constant, which is usually the case, there

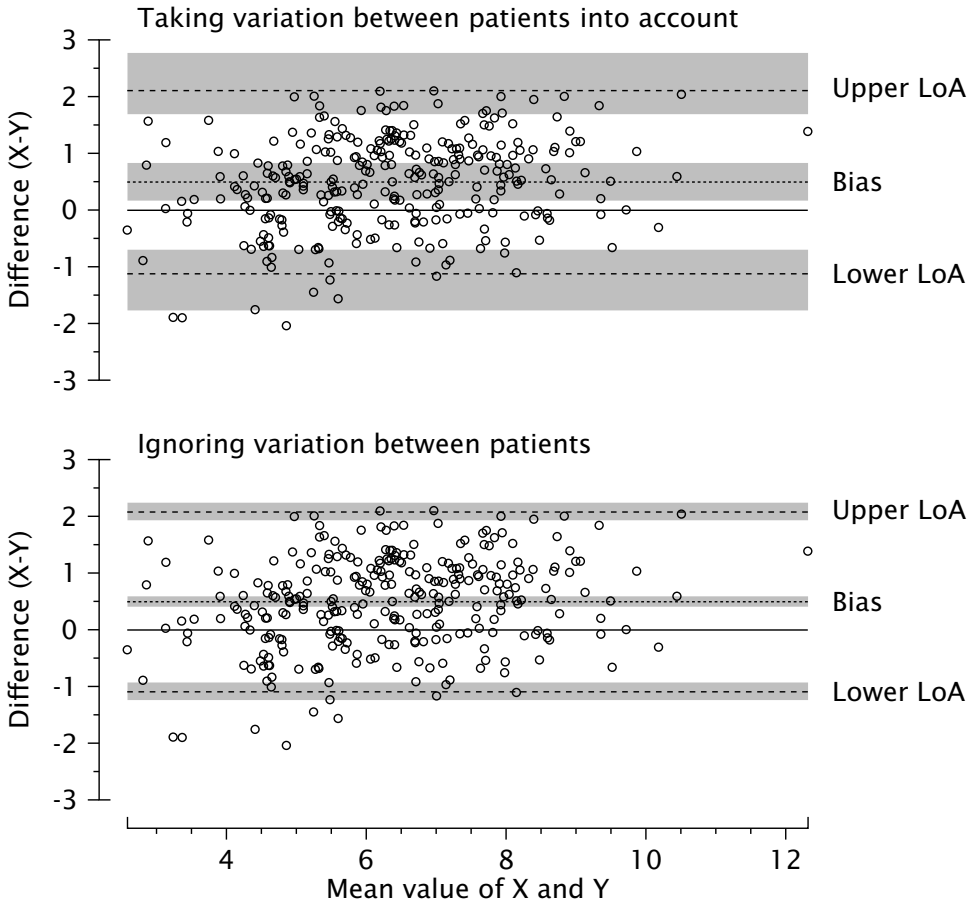

Figure 2.2: **Upper panel**: Bland-Altman analysis of simulated data with the True value varies method. *Solid line*: zero bias; *dotted lines*: overall bias; *dashed lines*: limits of agreement. The shaded areas indicate Student's-*t* and MOVER 95% confidence intervals. **Lower panel**: Bland-Altman analysis of simulated data with the Pooled data method. The *shaded areas* indicate the asymptotic 95% confidence intervals.

will be a risk of biased estimates. Part of the True value constant method calculates the repeatability coefficients, which should be reported (see below).

With unbalanced data, the standard True value varies method provides better confidence intervals with small between-subjects variability; the modified True value varies method provides better confidence intervals with large between-subjects variability. The methods incorporate sums of squared quantities which have the most advantageous statistical properties in the presence of small or large between-subjects variability. Our implementation chooses the best method automatically based on the estimated intraclass correlation (τ; see Section 2.A.6).

Our validation study showed that the expressions derived in the Appendix for the variance of the limits of agreement are valid, but for the confidence intervals themselves only asymptotically (for infinitely large data sets). The MOVER[100] provided very good confidence intervals with finite data sets, and outperformed the bootstrap in terms of coverage and speed. The MOVER uses percentiles of X^2 distributed quantities; the parametric bootstrap relies on the established model, and uses the variances of the X^2 distributed quantities to studentize the LoA. So the present methods to obtain confidence intervals all assume that the between-subjects and within-subject variabilities are normally distributed. This assumption, however, also underlies the limits of agreement themselves. If the assumption of normality does not hold, the number of differences within the LoA may be different from 95%.

For the True value constant method, Hamilton and Lewis found coverage frequencies that were too large.[38] In contrast, we obtained values close to 0.95 with 96 subjects. This discrepancy may be the result of using a possibly incorrect expression for calculating the confidence intervals (see Section 2.A.7; see also the erratum to their paper).

2.3.2 Mixed-Effects Models

In the literature,[10,55] the "mixed-effects" analysis approach has been applied. This approach takes the presence of random effects on different levels into account (here the level of the individuals and the levels of the measurements), based on assumptions concerning the probability distributions of the random effects. The methods described in this paper also take the two levels of random effects into account, but calculate their means and variances without assumptions on their distributions *per se*. However, as noted above, if the random effects are not normally distributed, the LoA need not contain 95% of the differences. Dedicated mixed-effects modeling software allows for the analysis of data from arbitrarily complex designs, but flexible model specification for the inclusion of fixed effects or covariates is not easily implemented in a Web browser application. For many applications, the present direct approach, for two sources of variability, is sufficient. However, as always, this should be checked using the diagnostic plots (which are provided by our library).

2.3.3 Explanatory Simulations

Bland and Altman, and others, have warned that if the presence of multiple paired measurements per individual is not taken into account, simple pooled analysis gives limits of agreement that are too narrow. The validation and application studies in the present analysis confirm this, although the effects on the limits of agreement may not be sub-

stantial. In the analysis of an example data set, Bland and Altman obtained limits of agreement that were indeed similar.[12] We found that the effect of multiple measurements per individual on the confidence intervals for the limits of agreement is very much greater. We now apply the derived expressions (see Appendix) to explain the effects of small numbers of subjects (as is often the case in clinical studies) on the expected variability of the estimates from Bland and Altman analysis.

In the upper panels of Figure 2.3, the expected standard deviation of the differences is plotted in relation to the proportion of between-subjects variance compared to the total variance (τ) for 6 and 24 subjects and 5 paired observations in each subject. This is done using eq. (2.10). As the number of subjects becomes larger, the term $\frac{n-1}{n}$ in eq. (2.10) soon becomes negligible. In the middle panel of Figure 2.3, the theoretical standard deviation of the bias is plotted, obtained by using eq. (2.3), either by substituting the expected value of the variance from the Pooled data for σ^2_{dw} and $\sigma^2_{dI} = 0$, or their values used in this simulation. The latter corresponds to using the expected values from the True value varies method as these are unbiased (see Appendix). The figure shows that the standard deviation may be in error by a factor of two; note that this variability is one part of the expression to obtain the 95% CIs of the limits of agreement (eq. (2.7)). Finally, in the lower panel of Figure 2.3, the theoretical standard deviation of the limits of agreement is shown, obtained by using eqs (2.6) and (2.3), with either eq. (2.10) or eq. (2.14) from the Pooled data and modified True value varies methods, respectively. Clearly, when using the Pooled data method, so when the presence of multiple paired measurements per individual is not taken into account, the confidence intervals obtained can easily be half their actual width when analyzing experimental data sets.

2.3.4 Towards a Standard Format of Reporting

Based on the literature and our findings described above, we propose the following standard format of reporting a Bland-Altman analysis:

- The types of the procedures used for the analysis.
- The mean of the differences (bias) and the limits of agreement, with their 95% confidence intervals.
- The standard deviation of the differences with its SE.
- Within-subject variability and between-subjects variability or the intraclass correlation (τ), with their/its SE, because that indicates that between-subjects variability was estimated.
- Repeatability coefficients which indicate the precision of the measurement devices.
- Remarks on visual inspection of the diagnostic plots - these may be shown or noted to be adequate.

2.3.5 The JavaScript Library

We have written software to execute the method of Bland and Altman to estimate bias and limits of agreement, with 95% confidence intervals, using JavaScript. Advantages of

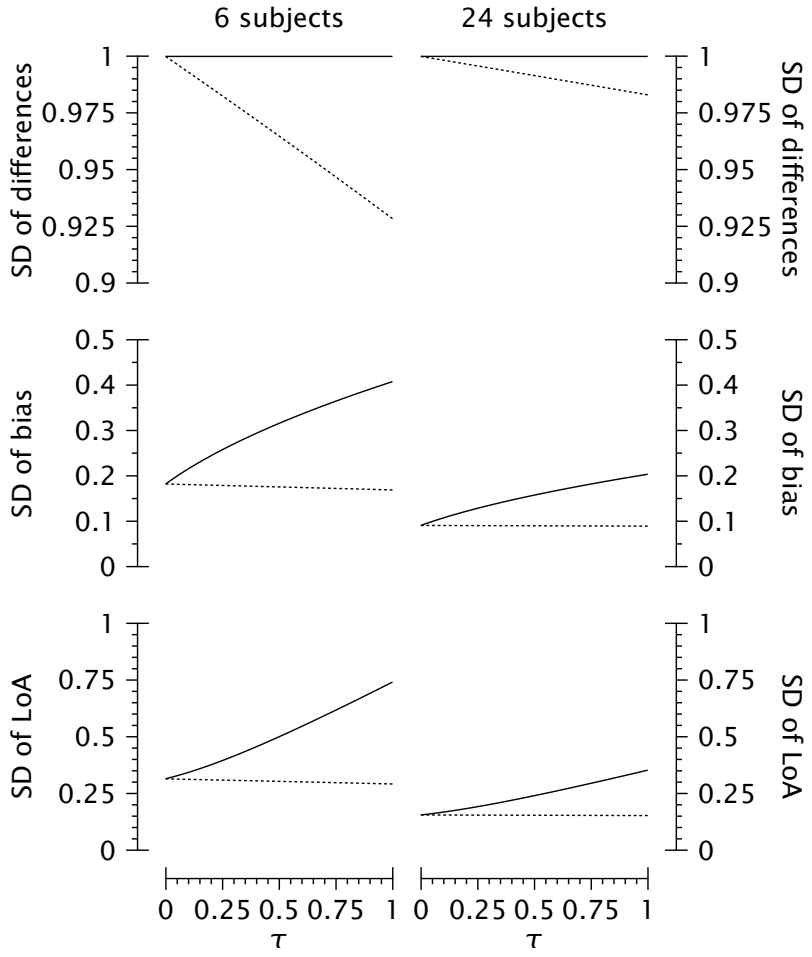

Figure 2.3: Effects of between subject variation (τ) and different numbers of subjects on key quantities in Bland-Altman analysis. The left panels show results for 6 subjects, the right panels results for 24 subjects. **Upper panels**: The standard deviation (SD) of the differences as a function of between-subjects variability, estimated by the Pooled data (*dotted line*) and True value varies (*solid line*) methods. **Middle panels**: The standard deviation of the bias as a function of between-subjects variability, estimated by the Pooled data (*dotted line*) and True value varies (*solid line*) methods. **Lower panels**: The standard deviation of the limits of agreement (LoA) as a function of between-subjects variability, estimated by the Pooled data (*dotted line*) and True value varies (solid line) methods.

JavaScript are 1) that the user only needs a Web browser to perform all computations; 2) the user may inspect the code of the application; 3) the files containing the application can (but need not) be located on a web server; and in the case of using a server 4) the data remain private and do not have to be transferred to the server. A disadvantage of JavaScript is that there are relatively few mathematical/statistical libraries available.

2.3.6 Conclusion

In this report, we have confirmed that when multiple measurements have been taken in several individuals, a Pooled data approach may be expected to give limits of agreement that are too narrow. Furthermore, we have shown convincingly that this has even more of an effect on the confidence intervals around these limits. When reporting how closely measurements between two devices are related, the accuracy of the limits of agreement is just as important as the limits themselves. Clearly, suitable software that permits easy calculation of these confidence intervals can be helpful in assessing the value of medical devices. Our results suggest that in previous studies wrong conclusions on agreement between two methods were likely, particularly when the number of subjects was small. To avoid ambiguity, we propose that studies that use the Bland and Altman method of comparison should follow a standard format. By providing sufficient data, especially the 95% CI of the limits of agreement, and between and within subject variation, ambiguity can be reduced and confidence in the results increased. The JavaScript library (including the validation script and example data set) to perform the discussed analyses may be obtained from one of the authors (E.O.).

Acknowledgment: The statistical properties of the simulated data for the example application were inspired by a real data set kindly provided by Prof. L.A.H. Critchley.

2.A Appendix: Derivations

This section summarizes all theory needed to compute the limits of agreement and their 95% confidence intervals for the situations described by Bland and Altman.[12]

2.A.1 The Model

The model for differences D_{ij} is the sum of a constant bias B and independent random variables,[15]

$$D_{ij} = B + I_i + E_{ij},$$

where I and E have zero means and variances σ_{dI}^2 and σ_{dw}^2, respectively; σ_{dI}^2 denotes between-subject variance, σ_{dw}^2 denotes within-subject variance, and $\sigma_{dI}^2 + \sigma_{dw}^2 = \sigma_d^2$ is the total variance of the differences. There is a total of N_{obs} observations from N_{id} individuals ($i = 1, \cdots, N_{\text{id}}$), with m_i observations for individual i ($j = 1, \cdots, m_i$).

2.A.2 The Mean of the Differences

Let the "grand" mean of the D_{ij} be estimated by

$$\hat{B} = \frac{1}{N_{\text{obs}}} \sum_{i=1}^{N_{\text{id}}} \sum_{j=1}^{m_i} D_{ij}. \tag{2.1}$$

The expected value of $\hat{B}$ is B; the variance of $\hat{B}$ is

$$\text{VAR}\left\{\hat{B}\right\} = E\left\{\hat{B}^2\right\} - \left(E\left\{\hat{B}\right\}\right)^2.$$

The D_{ij} have expected variance σ^2_{dw} only for equal subscripts $i = k$ and $j = l$ and autocovariance σ^2_{dI} for all $j, y = 1 \ldots m_i$ in each subject $i = k$

$$\text{E}\left\{\left(\sum_{i=1}^{N_{\text{id}}} \sum_{j=1}^{m_i} D_{ij}\right) \cdot \left(\sum_{k=1}^{N_{\text{id}}} \sum_{l=1}^{m_k} D_{kl}\right)\right\} = N_{\text{obs}} \cdot \sigma^2_{dw} + \left(\sum_{i=1}^{N_{\text{id}}} m_i^2\right) \cdot \sigma^2_{dI}$$

So

$$\text{VAR}\left\{\hat{B}\right\} = \frac{\sigma^2_{dw}}{N_{\text{obs}}} + \frac{\sum_{i=1}^{N_{\text{id}}} m_i^2}{N^2_{\text{obs}}} \cdot \sigma^2_{dI}. \tag{2.2}$$

And for equal $m_i = N_{\text{obs}}/N_{\text{id}}$

$$\text{VAR}\left\{\hat{B}\right\} = \frac{\sigma^2_{dw}}{N_{\text{obs}}} + \frac{\sigma^2_{dI}}{N_{\text{id}}}. \tag{2.3}$$

An alternative estimator of B is

$$\hat{B}_a = \frac{1}{N_{\text{id}}} \sum_{i=1}^{N_{\text{id}}} \frac{1}{m_i} \sum_{j=1}^{m_i} D_{ij}. \tag{2.4}$$

The expected value of $\hat{B}_a$ is B; the variance of $\hat{B}_a$ is

$$\text{VAR}\left\{\hat{B}\right\} = \left(\frac{1}{N^2_{\text{id}}} \sum_{i=1}^{N_{\text{id}}} \frac{1}{m_i}\right) \cdot \sigma^2_{dw} + \frac{\sigma^2_{dI}}{N_{\text{id}}}. \tag{2.5}$$

For equal $m_i = N_{\text{obs}}/N_{\text{id}}$, this variance is also equal to eq. (2.3). It can be shown that with unequal m_i, $\frac{1}{N^2_{\text{id}}} \sum_{i=1}^{N_{\text{id}}} \frac{1}{m_i} > \frac{1}{N_{\text{obs}}}$, and $\frac{\sum_{i=1}^{N_{\text{id}}} m_i^2}{N^2_{\text{obs}}} > \frac{1}{N_{\text{id}}}$. Using these results, it can be seen that B is more precisely estimated with eq. (2.2) when σ^2_{dI} is small, and more precisely estimated with eq. (2.5) when σ^2_{dw} is small.

An estimate of $\text{VAR}\left\{\hat{B}\right\}$, or $\text{VAR}\left\{\hat{B}_a\right\}$, can be obtained by substituting estimates of σ^2_{dw} and σ^2_{dI} in eq. (2.2) or eq. (2.5). Estimators of these variances are derived below.

A 95% confidence interval for $\hat{B}$ or $\hat{B}_a$ may be obtained by assuming these have a Student's t-distribution with $N_{\text{id}} - 1$ degrees of freedom.

2.A.3 Limits of Agreement and Their Confidence Intervals

The limits of agreement (LoA) are estimated by

$$\text{LoA} = \hat{B} \pm 1.96 \cdot \hat{\sigma}_d. \tag{2.6}$$

To obtain the variance of the LoA, the variances of $\hat{B}$ and $\hat{\sigma}_d^2$ are needed:

$$\text{VAR}\{\text{LoA}\} = \text{VAR}\left\{\hat{B}\right\} + 1.96^2 \cdot \text{VAR}\left\{\hat{\sigma}_d\right\} \tag{2.7}$$

The former is given by eq. (2.2) above; the latter can be obtained by assuming that sums of squares are X^2 distributed, and using

$$\text{VAR}\left\{\hat{\sigma}_d\right\} \approx \frac{\text{VAR}\left\{\hat{\sigma}_d^2\right\}}{4\text{E}\left\{\hat{\sigma}_d^2\right\}}.$$

Expressions for $\text{VAR}\left\{\hat{\sigma}_d^2\right\}$ are derived below. Confidence intervals around the LoA can finally be constructed by taking 1.96 times the square root of VAR{LoA}, assuming the LoA are normally distributed. This procedure was described by Bland and Altman.[15]

Confidence Interval Estimation by the MOVER

Donner and Zou described the application of the MOVER (Method of Variance Estimates Recovery) to estimate confidence intervals of the LoA.[28,100] We implemented eqs (5) and (6) from the paper by Zou[100] in our JavaScript library. The method combines confidence intervals for the estimated bias and the standard deviaton of the differences in eq. (2.6). The latter CI is based on percentiles of the X^2 distributions assumed for the sums of squares that are used to compute the standard deviation. The confidence interval for the mean used here is the one based on the normal approximation (see Zou[100]) instead of on Student's t-distribution. Otherwise, the CI of the LoA are too wide.

Parametric Bootstrap-*t* Confidence Intervals

When the LoA are not normally distributed, their confidence intervals may be asymmetric, for example when the estimated value of a statistic scales with its estimation error. A bootstrap procedure[29] may be used to construct better confidence intervals. In the parametric bootstrap, data sets are simulated using the established model. In the present context, the model consists of an overall bias, the partitioning of the total variance in between-subjects and within-subject variance, and their estimated values. From each of the bootstrap data sets, the LoA are calculated, and their 2.5% and 97.5% percentiles determined. The bootstrap-t interval "studentizes" the bootstrap estimated LoA using their associated standard deviation. Often, this is a disadvantage of this method, but in our case expressions for the standard error of the LoA can be derived (see below) that have sufficient accuracy. Furthermore, the bootstrap-t is applicable to location statistics - such as percentiles - in particular.[29] A nonparametric procedure based on resampling of individuals gave confidence intervals that were both too symmetric and too narrow.

2.A.4 The Pooled Data Method

The Variance of the Differences

Let the variance of the D_{ij} be estimated by

$$\hat{\sigma}_d^2 = \frac{1}{N_{\text{obs}}} \sum_{i=1}^{N_{\text{id}}} \sum_{j=1}^{m_i} \left(D_{ij} - \hat{B} \right)^2, \tag{2.8}$$

with $\hat{B}$ as defined in Section 2.A.2. Expanding the sums of squares yields

$$\sum_{i=1}^{N_{\text{id}}} \sum_{j=1}^{m_i} D_{ij}^2 - 2 \cdot \left(\sum_{i=1}^{N_{\text{id}}} \sum_{j=1}^{m_i} D_{ij} \right) \cdot \left(\frac{1}{N_{\text{obs}}} \sum_{k=1}^{N_{\text{id}}} \sum_{l=1}^{m_k} D_{kl} \right) + \sum_{i=1}^{N_{\text{id}}} \sum_{j=1}^{m_i} \left(\frac{1}{N_{\text{obs}}} \sum_{k=1}^{N_{\text{id}}} \sum_{l=1}^{m_x} D_{kl} \right)^2.$$

The expected value of the first part is

$$\mathrm{E} \left\{ \sum_{i=1}^{N_{\text{id}}} \sum_{j=1}^{m_i} D_{ij}^2 \right\} = N_{\text{obs}} \cdot \left(\sigma_{dw}^2 + \sigma_{dI}^2 \right),$$

and the expected values of the second and third parts can be found using the result in Section 2.A.2. We then find

$$E \left\{ \hat{\sigma}_d^2 \right\} = \left(1 - \frac{1}{N_{\text{obs}}} \right) \cdot \sigma_{dw}^2 + \left(1 - \frac{\sum_{i=1}^{N_{\text{id}}} m_i^2}{N_{\text{obs}}^2} \right) \cdot \sigma_{dI}^2.$$

For equal $m_i = N_{\text{obs}}/N_{\text{id}}$

$$E \left\{ \hat{\sigma}_d^2 \right\} = \left(1 - \frac{1}{N_{\text{obs}}} \right) \cdot \sigma_{dw}^2 + \left(1 - \frac{1}{N_{\text{id}}} \right) \cdot \sigma_{dI}^2. \tag{2.9}$$

At this point there is only an estimator of σ_{dw}^2 (eq. (2.8)) available, while an estimator of σ_{dI}^2 is not avaliable. The commonly used unbiased estimator of the variance, which divides the sum of squares by $N_{\text{obs}} - 1$ instead of N_{obs} (in eq. (2.8)) would give

$$E \left\{ \hat{\sigma}_{dc}^2 \right\} = \sigma_{dw}^2 + \frac{N_{\text{obs}}}{N_{\text{obs}} - 1} \cdot \frac{N_{\text{id}} - 1}{N_{\text{id}}} \cdot \sigma_{dI}^2, \tag{2.10}$$

which can only be an unbiased estimator - of σ_{dw}^2 - if $\sigma_{dI}^2 = 0$.

Limits of Agreement and their Confidence Intervals

For now, σ_{dI}^2 is taken to be zero. The sum of squares

$$\sum_{i=1}^{N_{\text{id}}} \sum_{j=1}^{m_i} \left(D_{ij} - \hat{B} \right)^2,$$

normalized by σ_{dw}^2, is assumed to have a X^2 distribution with $N_{\text{obs}} - 1$ degrees of freedom, so the variance of this quantity is $2(N_{\text{obs}} - 1)$. The variance of the unbiased esti-

mator of $\hat{\sigma}_d^2$ (eq. (2.8)) is then

$$\text{VAR}\left\{\hat{\sigma}_d^2\right\} = \frac{2\sigma_{dw}^4}{N_{\text{obs}} - 1},$$

and

$$\text{VAR}\left\{\hat{\sigma}_d\right\} \approx \frac{\text{VAR}\left\{\hat{\sigma}_d^2\right\}}{4\text{E}\left\{\hat{\sigma}_d^2\right\}} = \frac{\sigma_{dw}^2}{2(N_{\text{obs}} - 1)}.$$

Next, the variance of the LoA is given by the sum of the variance of the mean of the differences eq. (2.2) and 1.96^2 times the variance of the standard deviation of the differences, so

$$\text{VAR}\{\text{LoA}\} = \left(\frac{1}{N_{\text{obs}}} + \frac{1.96^2}{2(N_{\text{obs}} - 1)}\right) \cdot \sigma_{dw}^2.$$

Confidence intervals around the LoA can then be constructed by using the procedure described in Section 2.A.3.

2.A.5 The Standard True Value Varies Method

The ANOVA calculates two mean sums of squares:

$$\text{MSSR} = \frac{1}{N_{\text{obs}} - N_{\text{id}}} \cdot \sum_{i=1}^{N_{\text{id}}} \sum_{j=1}^{m_i} \left(D_{ij} - \hat{B}_i\right)^2, \tag{2.11}$$

and

$$\text{MSSI} = \frac{1}{N_{\text{id}} - 1} \cdot \sum_{i=1}^{N_{\text{id}}} m_i \cdot \left(\hat{B}_i - \hat{B}\right)^2,$$

where

$$\hat{B}_i = \frac{1}{m_i} \sum_{k=1}^{m_i} D_{ik}.$$

Their expected values are

$$E\{\text{MSSR}\} = \sigma_{dw}^2,$$

and

$$E\{\text{MSSI}\} = \sigma_{dw}^2 + \frac{1}{N_{\text{id}} - 1} \cdot \frac{N_{\text{obs}}^2 - \sum_{i=1}^{N_{\text{id}}} m_i^2}{N_{\text{obs}}} \cdot \sigma_{dI}^2 = \sigma_{dw}^2 + \lambda \cdot \sigma_{dI}^2,$$

where λ denotes an abbreviation. Therefore estimators of the components of variance are

$$\hat{\sigma}_{dw}^2 = \text{MSSR},$$

and

$$\hat{\sigma}_{dI}^2 = \left(\text{MSSI} - \hat{\sigma}_{dw}^2\right) \cdot \frac{(N_{\text{id}} - 1) \cdot N_{\text{obs}}}{N_{\text{obs}}^2 - \sum_{i=1}^{N_{\text{id}}} m_i^2} = \left(\text{MSSI} - \hat{\sigma}_{dw}^2\right) / \lambda$$

Next the variance of the differences can be estimated by

$$\hat{\sigma}_d^2 = \hat{\sigma}_{dw}^2 + \hat{\sigma}_{dI}^2 = (1 - 1/\lambda) \cdot \text{MSSR} + (1/\lambda) \cdot \text{MSSI}.$$

Now $(N_{\text{obs}} - N_{\text{id}}) \cdot \text{MMSR}/\sigma_{dw}^2$ is approximately X^2 distributed with $N_{\text{obs}} - N_{\text{id}}$ degrees of freedom, and $(N_{\text{id}} - 1) \cdot \text{MSSI}/(\sigma_{dw}^2 + \lambda\sigma_{dI}^2)$ is approximately X^2 distributed with $N_{\text{id}} - 1$ degrees of freedom, at least with balanced data, so when m_i values are equal.[17] The properties of the X^2 distributed quantities can be used to derive an expression for the variance of $\hat{\sigma}_d^2$:

$$\text{VAR}\left\{\hat{\sigma}_d^2\right\} = \frac{2\left((1 - 1/\lambda)\sigma_{dw}^2\right)^2}{N_{\text{obs}} - N_{\text{id}}} + \frac{2\left(\sigma_{dw}^2/\lambda + \sigma_{dI}^2\right)^2}{N_{\text{id}} - 1}.$$

This result can be transformed and added to the variance of the mean to obtain approximate confidence intervals as in Section 2.A.3. The MOVER (see Section 2.A.3) may be applied, using percentiles of the X^2 distributions of MSSR and MSSI, to obtain better confidence intervals.

2.A.6 The Modified True Value Varies Method

An alternative mean sum of squares was studied by Thomas and Hultquist.[87] They used the expression

$$\text{MSSIa} = \frac{1}{N_{\text{id}} - 1} \cdot \sum_{i=1}^{N_{\text{id}}} \cdot \left(\hat{B}_i - \hat{B}_a\right)^2. \tag{2.12}$$

Note that $\hat{B}_a$ is used here (eq. (2.4)), which may be different from $\hat{B}$ with unbalanced data. The expectation of MSSIa is

$$E\{\text{MSSIa}\} = \frac{1}{N_{\text{id}}} \cdot \sum_{i=1}^{N_{\text{id}}} \frac{1}{m_i} \cdot \sigma_{dw}^2 + \sigma_{dI}^2 = \lambda \cdot \sigma_{dw}^2 + \sigma_{dI}^2, \tag{2.13}$$

where λ is an abbreviation. Next the variance of the differences can be estimated by

$$\hat{\sigma}_d^2 = (1 - \lambda) \cdot \text{MSSR} + \text{MSSIa}.$$

Thomas and Hultquist show[87] that $(N_{\text{id}} - 1) \cdot \text{MSSIa}/(\lambda \cdot \sigma_{dw}^2 + \sigma_{dI}^2)$ is close to a X^2 distribution with $N_{\text{id}} - 1$ degrees of freedom, for $\sigma_{dI}^2 \geq \sigma_{dw}^2/4$. Burdick and Graybill gave[17] a method to obtain confidence intervals for $\hat{\sigma}_d^2$, which may also be used to obtain confidence intervals for $1.96 \cdot \hat{\sigma}_d$, but this method was not investigated given the accuracy of the MOVER (see Section 2.A.3). The properties of the X^2 distributed quantities can be used to derive an expression for the variance of $\hat{\sigma}_d^2$:

$$\text{VAR}\left\{\hat{\sigma}_d^2\right\} = \frac{2\left((1 - \lambda)\sigma_{dw}^2\right)^2}{N_{\text{obs}} - N_{\text{id}}} + \frac{2\left(\lambda\sigma_{dw}^2 + \sigma_{dI}^2\right)^2}{N_{\text{id}} - 1}. \tag{2.14}$$

This result can be transformed and added to the variance of the mean to obtain approximate confidence intervals as in Section 2.A.3. Zou showed[100] how to use the MOVER using the distributions of MSSR and MSSIa (see Section 2.A.3) to obtain better confidence intervals.

2.A.7 The True Value Constant Method

To obtain expressions for the 95% confidence intervals for the True value constant method, the approach of Bland and Altman[15] was followed, but generalized for unbalanced data. In this method, sums of squares are calculated from the measurements themselves rather than from their differences. These sums of squares have the same properties concerning their X^2 distributions as described in Section 2.A.6. In the derivation of Bland and Altman[15] the wrong expression was used for the variance of the mean. They substituted the estimated variance of the differences ($\hat{\sigma}_{\bar{d}}^2$) instead of the variance of the differences between the within-subject means ($s_{\bar{d}}^2$ - for equal m_i, the ratio of eq. (2.13) and N_{id} is equal to eq. (2.3)). Because the variance of the differences is larger than the variance of the differences between the within-subject means, the resulting 95% CI will be too wide.

The True value constant method calculates the means of the X and Y data per subject: $\bar{X}_i = \sum_{j=1}^{m_i} X_{ij}$, and likewise for Y. Next, the variance of the differences between the within-subject means is calculated, which is the variance of $\bar{X}_i - \bar{Y}_i$, which is equal to the variance of $\hat{B}_i$, so it is identical to eq. (2.12) from the modified True value varies method. Furthermore, The True value constant method calculates sums of squares as given by eq. (2.11), but separately for the X and Y data, and subsequently adds these. The expectation of that result is $\sigma_{dw}^2 = \sigma_{xw}^2 + \sigma_{yw}^2$, which is identical to the expectation of eq. (2.11). Thus when the true value is indeed constant, the True value constant method has properties identical to the modified True value varies method, except that with the latter method, the number of measurements from both measurement devices needs to be the same ($m_{xi} = m_{yi}$), to be able to calculate the D_{ij}. Furthermore, the separate estimates of σ_{xw}^2 and σ_{yw}^2 from the True value constant method may be used to assess repeatability of the measurement devices.[11]

2.A.8 Generation of Simulation Data

The simulated data for the validation study and the example analysis were generated using variants of the following pseudocode:

```
for (i=0; i<n; i++) {
  Ix[i] = sxI*normal();
  Iy[i] = syI*normal();
  for (j=0; j<m[i]; j++) {
    X[i][j] = Bx + Ix[i] + sxw*normal();
    Y[i][j] = By + Iy[i] + syw*normal();
  }
}
```

Here `n` is the number of subjects, and `m[i]` is the number of paired measurements for subject `i`. `sxI` and `syI` are the standard errors of the between-subjects variabilities, and `swx` and `swy` are the standard errors of the within-subject variabilities. `normal()` is a function that generates normally distributed numbers with a mean of zero and variance of 1. `Ix` and `Iy` are the subject-specific biases, and `X` and `Y` are data arrays. With the above pseudocode the expected overall bias is `Bx - By`, the true value constant and zero, and the expected between-subjects and within-subjects variances are $\sigma_{dI}^2 = \mathtt{sxI}^2 + \mathtt{syI}^2$ and $\sigma_{dw}^2 = \mathtt{sxw}^2 + \mathtt{syw}^2$, respectively.

Chapter 3

Using Akaike's Information Theoretic Criterion in Mixed-Effects Modeling of Pharmacokinetic Data: A Simulation Study*

WE FIRST DEFINE population data as a set of one or more measurements in two or more individuals *e.g.*, patients, volunteers, or animals). Such data may be characterized by mixed-effects models, where the mixed effects consist of fixed and random effects. Fixed effects (or fixed effect factors) are, for example, the times at which the measurements are obtained, and covariates such as demographic characteristics of the individuals. Due to random effects (or random effect factors), the model output may vary between measurements, and between individuals. When mixed-effects models are fitted to population data, the question arises as to how many of those effects should be incorporated in the model. This is the so-called problem of variable selection.[40]

One strategy is to observe the change in goodness-of-fit by adding one more parameter and test the significance of that change.[16] In the maximum likelihood approach, the objective function value (OFV), being the minus two logarithm of the likelihood function, is minimized. To attain a p-value of *e.g.*, 0.05 or less, the decrease in OFV, when adding one parameter, should be 3.84 or more.[16]

Another strategy is to apply Akaike's information theoretic criterion (AIC), which can be written as

$$\mathrm{AIC} = \mathrm{OFV} + 2 \cdot D, \tag{3.1}$$

where D is the number of parameters in the model.[1,16,18,40] The model with the lowest value of AIC is considered the best one. In the case of just adding one parameter, the OFV needs to decrease only 2 points or more to be incorporated in the model, so the associated p-value > 0.05 seems too high to justify this strategy.

When additional model parameters are incorporated, the significance of one model parameter might change, but the interpretation of AIC does not.[18] However, when multiple significance tests are performed, the significance level of each individual test should be corrected to a lower value, so a decrease of 2 points for one parameter does again seem to be too low.

* E Olofsen, A Dahan, *F1000Research* 2015; 2:71

Even if the strategy of using AIC leads to optimal variable selection, the question arises as to whether this is also the case when using mixed-effects models. In theory, the model that is best according to AIC is the one that minimizes prediction error[1,96] and this is also true for a mixed-effects model when predicting data for individuals for which no data have been obtained so far.[96]

In the literature, many simulation studies have assessed the performance of AIC, but to our knowledge these were never done in selecting the model with minimal prediction error for population data. In this article, we will define a toy pharmacokinetic model and observe the performance of AIC when adding fixed effects to this model, as well as when adding interindividual variability.

3.1 Methods

3.1.1 A Hypothetical Pharmacokinetic Model

Consider the following function $y(t)$, an infinite sum of exponentials, and its relationship with a (negative) power of time:[57]

$$y(t) = \int_0^\infty \exp(-\lambda t) d\lambda = -\frac{1}{t}\exp(-\lambda t)\Big|_0^\infty = \frac{1}{t} \text{ for } t > 0. \tag{3.2}$$

Figure 3.1A shows that this function looks like a typical pharmacokinetic profile after bolus administration. This model is to be regarded as a toy model, because we do not expect it to adequately describe pharmacokinetic data, although variations of power functions of time have been shown to fit pharmacokinetic data well.[57] We approximate $y(t) = 1/t$ by the following sum of M exponentials:

$$\hat{y}(t_j; \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \sum_{m=1}^{M} \alpha_m \exp(-\lambda_m t_j). \tag{3.3}$$

The M parameters $\boldsymbol{\lambda}$ and measurement time instants t_j are fixed and are set to have distinct values as described in the next subsections. The coefficients $\boldsymbol{\alpha}$ (related to how a drug dose is distributed across compartments) are parameters to be estimated. Let the number of α_m that are not fixed to zero be denoted by K. Then the above approximation has the property that while the fits of models to the data would improve with increasing K, we would need no less than $K = M$ exponentials to obtain a perfect fit. Moreover, with noisy data, it might be that for $K < M$ an optimal fit is obtained in the sense that then the associated prediction error of the model is minimal. Figure 3.1B shows how eleven (in this case error-free) samples from this function can be approximated by sums of exponentials.

3.1.2 Individual Data Modeling and Simulation

In the following, the time instants t_j, $j = 1, \cdots, M$, centered around 1, were chosen within $[1/t_{\max}, t_{\max}]$ according to

$$t_j = \left(\frac{j}{M+1-j}\right)^\gamma, \tag{3.4}$$

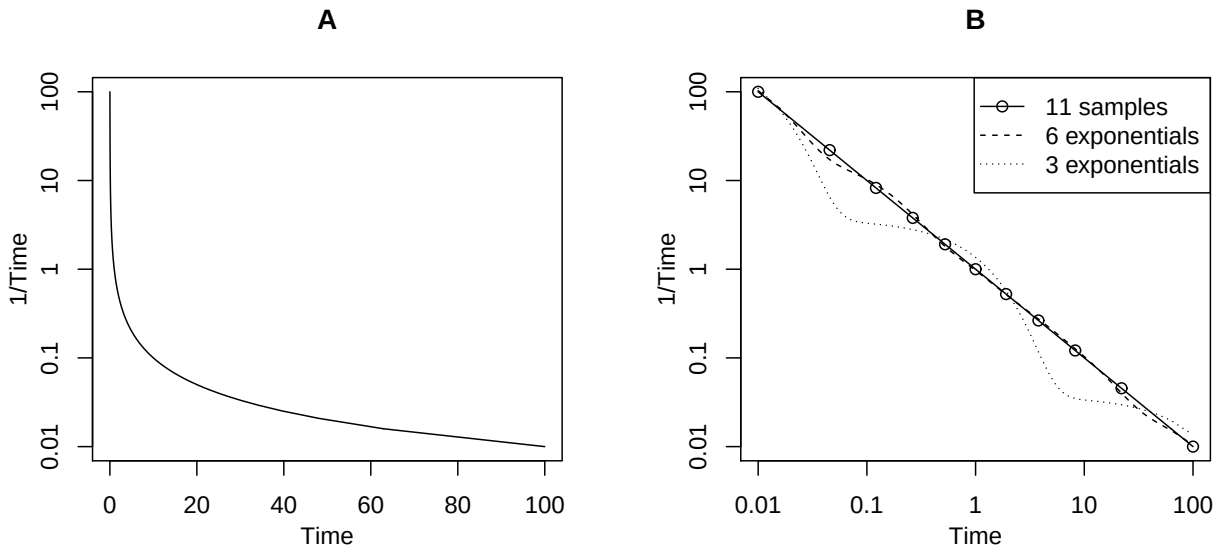

Figure 3.1: A: function $y(t) = 1/t$, and B: approximations obtained by fitting six and three exponentials to the depicted eleven samples. Note the log-lin and log-log scales for panels A and B, respectively. Time has arbitrary units.

with $\gamma = \log(t_{\max})/\log(M)$; $t_{\max}$ was set to 100 (see the time axis of Figure 3.1B for an example with $M = 11$). Simulated data with constant proportional error were generated *via*

$$y(t_j) = \frac{1}{t_j}(1 + \epsilon_j), \tag{3.5}$$

where ϵ_j denotes Gaussian measurement noise with variance σ^2. The M time constants $\boldsymbol{\lambda}$ were fixed according to $\lambda_m = 1/t_m$, $m = 1, \cdots, M$. In this setting the model eq. (3.3) can be fitted to simulated data using weighted linear least squares regression, with weight factors $w(t_j) = 1/t_j$ (note that no precaution is needed against $\epsilon \leq -1$). Linear least squares regression is very fast and robust, so it allows for the evaluation of many simulation scenarios.

3.1.3 Population Data Modeling and Simulation

Population data consisting of N individuals were simulated *via*

$$y_i(t_j) = \frac{1}{t_j} \cdot (\exp(\eta_i) + \epsilon_{ij}) \text{ with } i = 1, \cdots, N, \tag{3.6}$$

where η_i denotes interindividual variability with variance ω^2. The random effect η_i influences the overall magnitude of the values of y_i, but not the shape of the function in time, so this is similar to a random effect that influences pharmacokinetic volume of distribution. The nonlinear mixed-effects model for the population data was then

written as:

$$\hat{y}_i(t_j;\boldsymbol{\alpha},\boldsymbol{\lambda}) = \sum_{m=1}^{M} \alpha_m \exp(-\lambda_m t_j + \eta_i). \tag{3.7}$$

Note that with $N > 1$, a perfect fit is no longer obtained with $K = M$ nonzero coefficients $\boldsymbol{\alpha}$, because the ϵ_{ij} are generally different for different i (individuals). Just one different η_i for each individual i cannot compensate for M different ϵ_{ij}.

3.1.4 Statistical Analysis

Simulation data were generated *via* eq. (3.6) in R.[85] Model fitting was also done in R, with function "lm()" from package "stats", except for nonlinear mixed-effects model fitting for simulated data with $\omega^2 > 0$, which was done in NONMEM version 7.3.0.[9] Parameters $\boldsymbol{\alpha}$ (see eq. (3.7)) were either fixed to zero or free. Although the $\boldsymbol{\alpha}$ are expected to be positive with pharmacokinetic data, they were not constrained to be positive. So it was not possible for parameters to become essentially fixed to zero due to that constraint, which would reduce the dimensionality of the model. Prediction error (ν^2) was calculated with

$$\nu^2 = \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{j=1}^{M} \left(\frac{z_i(t_j) - \hat{y}_i(t_j)}{w(t_j)} \right)^2, \tag{3.8}$$

using predictions based on eq. (3.7) with the random effects $\eta_i = 0$. Validation data $z_i(t_j)$ were also generated *via* eq. (3.6), but with different realizations of ϵ_{ij} and η_i. Error terms weighted with $w(t_j) = 1/t_j$ are homoscedastic, which is an assumption underlying regression analysis and allows for the interpretation of ν^2 as independent of time. The objective function OFV was also calculated at the estimated parameters using the validation data, denoted OFV_v, which should on average be approximately equal to Akaike's criterion (see Supplementary material). OFV_v was compared with AIC and also with Akaike's criterion with a correction for small sample sizes (AIC_c[18])

$$\mathrm{AIC}_c = \mathrm{OFV} + 2 \cdot D \cdot \left(1 + \frac{D+1}{N \cdot M - D - 1}\right). \tag{3.9}$$

The above criteria were normalized by dividing them by the number of observations (see Supplementary material for motivation), and averaged over 1000 runs (unless otherwise stated; and runs where NONMEM's minimization was not successful were excluded). For plotting purposes, 95% confidence intervals or confidence regions for means were determined using R's packages "gplots" and "car", under the assumption that averages over 1000 variables are normally distributed. Model selection frequencies were calculated based on optimal models according to AIC_c as determined for each simulation data set.

3.1.5 Selection of Parameter Values

Simulation parameters M and σ^2 are expected to determine the number of exponentials K; if the number of measurements M increases and/or the measurement error σ^2 decreases, K will increase. Without interindividual variance, so $\omega^2 = 0$, the information in the data increases as N increases, so also in that case K is expected to increase. With $N = 2$, $M = 11$ and $\sigma^2 = 0.5$, pilot simulations indicated a $K \approx 4$. When $\omega^2 > 0$, prediction error will increase, but it is less easy to predict what its effect will be on K. For ω^2

Table 3.1: Selecting $K = 1, \cdots, M = 11$ evenly spaced rate constants from $\boldsymbol{\lambda}$: 0 and 1 denote α_m to be fixed to zero, and a free parameter to be estimated, respectively (see eq. (3.7)).

K	m : 1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	1	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	1
3	1	0	0	0	0	1	0	0	0	0	1
4	1	0	0	1	0	0	0	1	0	0	1
5	1	0	0	1	0	1	0	1	0	0	1
6	1	0	1	0	1	0	1	0	1	0	1
7	1	0	1	1	0	1	0	1	1	0	1
8	1	1	0	1	1	0	1	1	0	1	1
9	1	1	0	1	1	1	1	1	0	1	1
10	1	1	1	1	1	0	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1	1

values of 0, 0.1, and 0.5 were selected - values that are encountered in practice. Because there is only one random effect in the mixed-effects model, the relatively low number of individuals $N = 5$ was selected.

For a certain choice of M, there are $2^M - 1$ possible combinations of λs to choose for the terms $\exp(-\lambda_m t_j)$ in the sum of exponentials (excluding the case of a model without exponentials). Because accurate evaluation of all models at different parameter values is not feasible with respect to computer time, the set of possible combinations was reduced to one with evenly spaced λs. Table 3.1 gives an example for the case $M = 11$.

3.2 Results

Figure 3.2 shows the averaged prediction error *versus* number of exponentials for all possible choices of $\boldsymbol{\lambda}$, with $N = 2$, $M = 11$, $\sigma^2 = 0.5$, and $\omega^2 = 0$. From the figure it is clear that prediction error may indeed increase if the number of exponentials selected is too large. The bigger solid circles correspond to the models chosen in Table 3.1; in general the evenly spaced selection of exponents resulted in models with smallest prediction error.

Figure 3.3 shows simulation results using the model set defined in Table 3.1, starting from $K = 4$, with parameters $N = 5$, $M = 11$, $\sigma^2 = 0.5$, and $\omega^2 = 0$. The model with $K = 6$ exponentials had minimal mean AIC_c, and also minimal mean OFV_v and minimal mean squared prediction error (ν^2). With $N = 5, M = 11$, there are still visible differences between AIC_c and AIC; although AIC would in this case also select the optimal model, AIC appears to favor more complex models. Note that the sizes of the confidence intervals and confidence regions can be made arbitrarily small by choosing the number of runs to be higher than the selected number of 1000 (at the expense of computer time).

Figure 3.4 shows simulation results with $\omega^2 = 0.1$; mixed-effects analysis was used

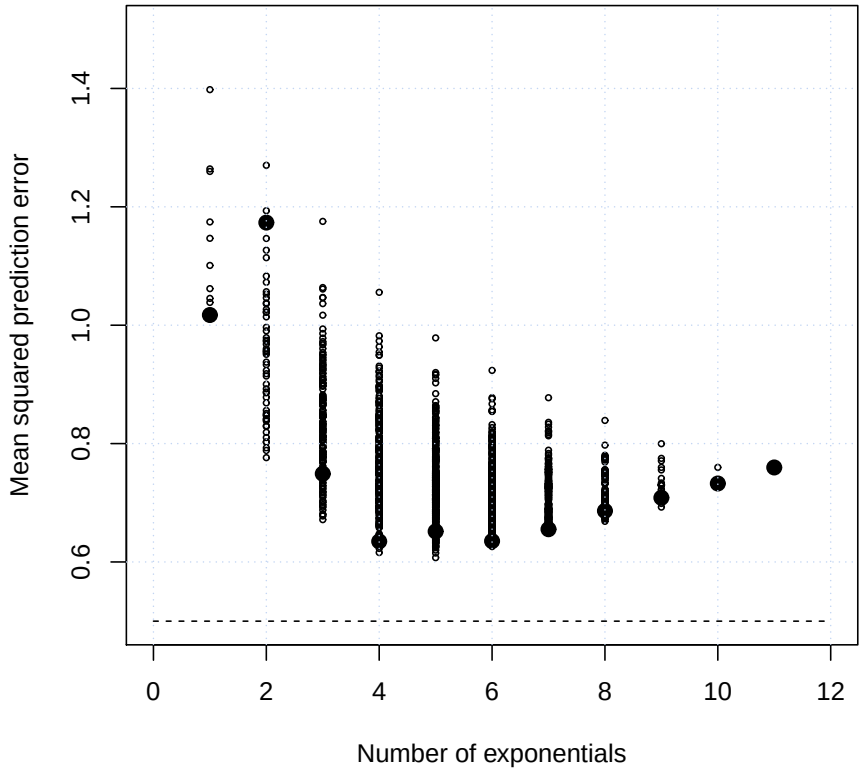

Figure 3.2: Mean squared prediction error ν^2 (eq. (3.8)) as a function of the number of exponentials, with 2047 models, averaged over 100 runs, $N = 2$, $M = 11$, $\sigma^2 = 0.5$, $\omega^2 = 0$. The dashed line represents the prediction error from the true model, so that $\nu^2 = \sigma^2$. The bigger solid circles correspond to the models chosen in Table 3.1.

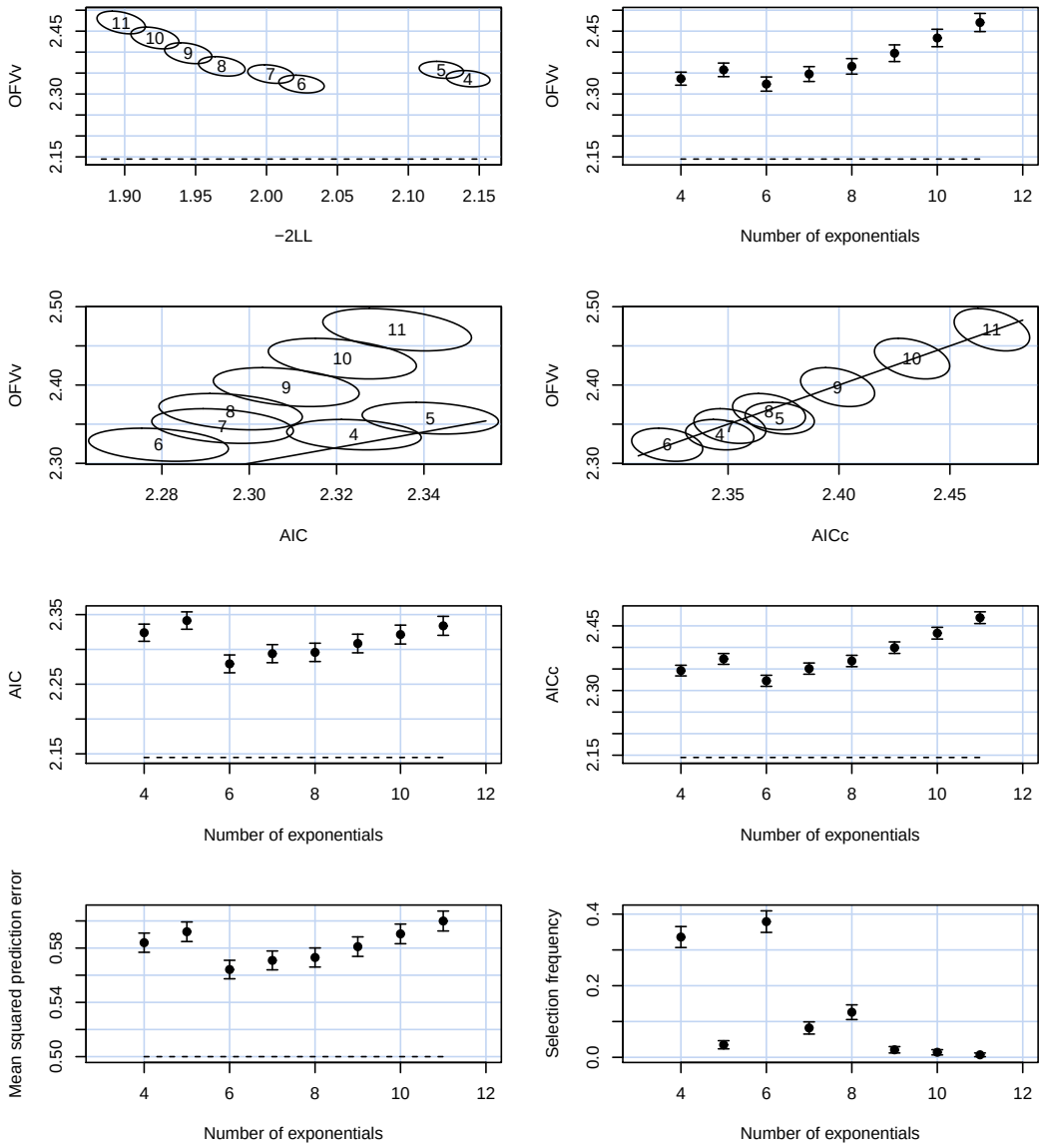

Figure 3.3: Mean OFV_v as a function minus of two log-likelihood (-2LL), the number of exponentials, AIC and AIC_c (top four panels), and AIC, AIC_c, prediction error ν^2, and model selection frequencies as a function of the number of exponentials (lower four panels), averaged over 1000 runs, $N = 5$, $M = 11$, $\sigma^2 = 0.5$, $\omega^2 = 0$. The dashed lines represent the theoretical values for an infinite amount of data (see Appendix). Error bars and ellipses denote 95% confidence intervals and confidence regions, respectively. The numbers in the confidence regions denote the number of exponentials. The solid lines in the middle upper panels are lines of identity.

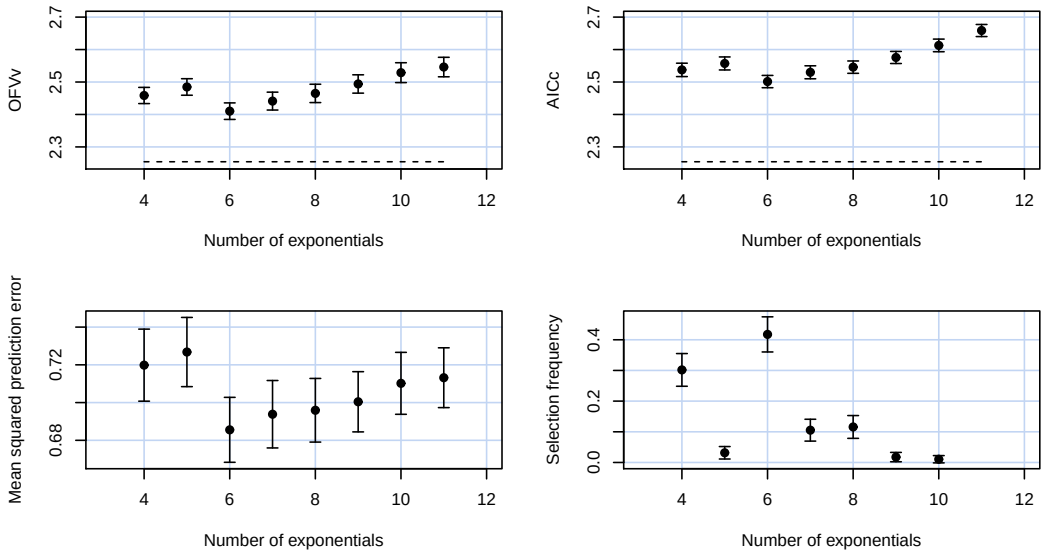

Figure 3.4: Mean OFV_v, AIC_c prediction error ν^2, and model selection frequencies as a function of the number of exponentials, for $\omega^2 = 0.1$; parameters otherwise identical to those for Figure 3.3.

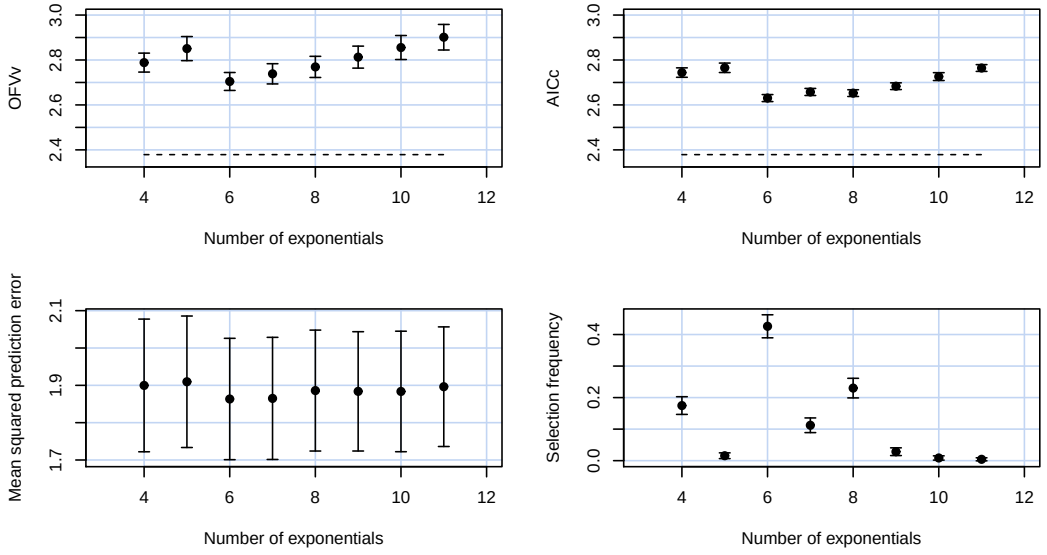

Figure 3.5: Mean OFV_v, AIC_c prediction error ν^2, and model selection frequencies as a function of the number of exponentials, for $\omega^2 = 0.5$; parameters otherwise identical to those for Figure 3.3.

to fit the population data. The main difference with the results of data with $\omega^2 = 0$ is the overall increase in OFV_v and AIC_c. The optimal number of exponentials remained $K = 6$.

Figure 3.5 shows simulation results with ω^2 set at the higher value of 0.5. The main differences with the results of data with $\omega^2 = 0.1$ are again the overall increase in OFV_v, AIC_c and prediction error, and also in the variability in the prediction error. The optimal number of exponentials remained $K = 6$, although AIC_c begins to favor the models with larger K (a simulation with N increased to 7, both OFV_v and AIC_c favored larger models; data not shown).

3.3 Discussion

With the objective of creating a simulation context resembling pharmacokinetic analysis where concentration data are approximated by a sum of exponentials, the toy model $y(t) = 1/t$ was chosen. In this setting, reality - the reality of the toy model - is always underfitted. When mixed-effects models were fitted to the simulated data, mean AIC_c was approximately equal to the validation criterion mean OFV_v. The minima of mean AIC_c and mean OFV_v coincided. With large interindividual variability, mean expected prediction error (ν^2, see eq. (3.8)), with random effects fixed to zero), was less discriminative between models, so that it becomes less suitable as a validation criterion; it does not take into account whether estimated interindividual variability matches the variability in the validation data.

3.3.1 Akaike's *versus* the Conditional Akaike Information Criterion

Vaida and Blanchard proposed a conditional Akaike information criterion to be used in model selection for the "cluster focus".[96] It is important to stress that their definition of cluster focus is the situation where data are to be predicted of a cluster that was also used to build the predictive model. In that case, the random effects have been estimated, and then the question arises how many parameters that required. In our situation, a cluster is the data from an individual; AIC was used in the situation of predicting population data consisting of individual data that were not used to build the model. This would seem to be the most common situation in clinical practice. Furthermore, AIC for the population focus is asymptotically equivalent with leave-one-individual-out cross-validation; AIC for the individual focus with leave-one-observation-out cross-validation.[31]

3.3.2 Akaike's *versus* the Bayesian Akaike Information Criterion

We chose to perform simulations using the model given by eq. (3.2) because approximating data with a sum of exponentials is daily practice in pharmacokinetic analysis where data are obtained from "infinitely complex" systems, and we cannot hope to find the "correct" model. The Bayesian information criterion (BIC) is consistent in the sense that it selects the correct model, given an infinite amount of data.[18] The reason that AIC can be used in "real-life" problems is that as the amount of data goes to infinity, the

complexity, or dimension, of the model that should be applied should also go infinity.[19] Burnham and Anderson show that it is possible to choose the prior for BIC in such a way that it incorporates the knowledge that more complex models should be favored if the amount of data increases, and so that the BIC "reduces" to AIC.[18,19] In the situation that the correct model set belongs to the set of evaluated models, a selection criterion that both finds the correct model and minimizes prediction error would be preferable - but Yang concluded that this may not be possible.[98]

3.3.3 Model Selection Criterion AIC and Predictive Performance

It should be noted that minimizing AIC has a more general interpretation than just minimizing prediction error v^2 as given for example by eq. (3.8). The interpretation of minimizing AIC is minimizing the difference between the the information contained in the data and captured by the model.[18] Independent or future population data z are not just predicted by $\hat{y}$; also the distributions of the expected random effects ϵ and η are characterized by $\hat{\sigma}^2$ and $\hat{\omega}^2$. That is why OFV_v (and not v^2) is the criterion to be used to assess the predictive performance of a model.

3.3.4 Regression Weights as Functions of the Model Output

The simulated data were analyzed using weighted (non)linear regression, see eq. (3.6)), where measurement noise was weighted according to the exact function value. In practice, when the weights are unknown, the model output may be used to weight the data. In that case simulated data should be generated (*cf.* eq. (3.6)) *via*

$$y_i(t_j) = \frac{1}{t_j} \cdot \exp(\eta_i) \cdot (1 + \epsilon_{ij}). \tag{3.10}$$

The likelihood function and AIC are both still well-defined if the model output $\hat{y}_i(t_j) \neq 0$. Prediction errors are to be calculated with

$$v^2 = \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{j=1}^{M} \left(\frac{z_i(t_j) - \hat{y}_i(t_j)}{\hat{y}_i(t_j)} \right)^2, \tag{3.11}$$

where where $\hat{y}$ possibly becomes arbitrarily close to zero for less than optimal models, and v^2 may be based on long-tailed distributed numbers. To be able to compare prediction errors from different models, the weight factors could be chosen identical for all K to the model output of the largest model - see Appendix for further analysis.

3.3.5 Model Selection Uncertainty

Theoretically, and in the discussed simulations, minimum mean AIC is related to best mean predictive performance, where the mean is taken across multiple studies and prespecified models. This holds independent of the number of models. However, in practice, we have data from one study and the task of specifying the models to consider. As soon as there is more than one model, there is a nonzero probability that the model

selected based on AIC would have, on average, a larger prediction error than the optimal one. Also, if we were able to repeat the study, the average prediction error based on the models with minimum AIC would be larger than optimal. With many models, model selection is called unstable in the sense that each time a study is repeated it would lead to the selection of another model.

The figure panels with the model selection frequencies (Figure 3.3, Figure 3.4, and Figure 3.5) show: 1) there is a relationship between the model with highest selection probability and minimum mean prediction error, but this relationship is not one-to-one; 2) there can be an almost as large selection probability for a model that is not associated with minimum mean prediction error; but 3) in that case, their minimum mean prediction errors are comparable.

Models with equal mean predictive properties may have different properties in different extrapolation scenarios. Model averaging,[18] where model parameters or their predictions are averaged, reduces model selection instability and hence may be used to avoid model specific inference which discards model selection uncertainty. Data dredging[18] refers to the situation where there is an increasingly large set of models which are not prespecified. At the point the data dredging is stopped (by the investigator, or by the computer), the best model is at high risk to fit only the data at hand, and hence cannot be used for prediction.[21]

3.3.6 Limitations of the Study

We recognize the following limitations of our study:

- The simulation model contained only one random effect to describe interindividual variability, and therefore the number of random effect (co)variances was fixed to one in the model set used for fitting. While the number of (co)variance parameters should be counted as ordinary parameters,[96] at least in well behaved situations,[36] we did not investigate the process of optimizing this part of a random effects model.

- The nonlinearity in the mixed-effects model was simply due to a multiplicative factor $\exp(\eta)$ in the model output. Usually, random effects in pharmacokinetic models have more complex influence on the model output. However, the lognormal nature of $\exp(\eta)$ is a characteristic property of both our toy model and general pharmacokinetic models.

- The characteristics of the exponentials incorporated in the regression models were evenly spaced, and the values of the rate constants $\boldsymbol{\lambda}$ were fixed. We expect that with more freedom in the specification of the set of models, prediction errors with overfitted models may be worse. However, the agreement between AIC_c and prediction error should persist.

- We did not evaluate all possible models within their definition, but only those listed in Table 3.1, and it makes sense to limit the model set to reduce model selection instability.[18,98] We did not address how to optimally select the rate constants $\boldsymbol{\lambda}$. Stepwise selection methods have their disadvantages.[83] With stepwise forward selection, AIC_c may even perform worse than AIC.[60]

· We did not evaluate the process of covariate selection. However, the set of exponentials may be viewed as a number of (somewhat correlated) predictors. It is therefore expected that the present findings also hold for other types of covariates.

3.4 Conclusion

In conclusion, the present simulation study demonstrated that, at least in a relatively simple mixed-effects modeling context with a set of prespecified models, minimum mean AIC_c coincided with best predictive performance, also in the presence of interindividual variability.

Acknowledgment: The authors would like to thank J. de Goede for many fruitful discussions.

3.A Appendix: Supplementary Material

In the following, we summarize theory on the maximum likelihood approach and AIC relevant for this paper. We start with the situation for data from one individual and show how AIC is related to OFV_v. Subsequently we discuss the situation for population data.

Suppose the model for measured data y_j, $j = 1, \cdots, M$ is given by (*cf.* eqs ((3.5), (3.6), and (3.10))

$$y_j = \hat{y}_j + w_j \cdot \epsilon_j,$$

where $\hat{y}_j$ is the model output, w_j are weight factors, and ϵ_j are independent normally distributed with mean zero and variance σ^2. The likelihood function L for this data set is then given by

$$L(y;\boldsymbol{\theta}) = \prod_{j=1}^{M} \frac{1}{w_j \sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y_j - \hat{y}_j}{w_j \sigma}\right)^2\right], \tag{3.12}$$

where the set of parameters $\boldsymbol{\theta}$ contains σ^2 and those needed to calculate $\hat{y}$. The objective function value (OFV) is defined as minus two times the natural logarithm of the likelihood:

$$\mathrm{OFV} = -2\log(L(y;\boldsymbol{\theta}) = \sum_{j=1}^{M} \log(w_j^2) + M\log(\sigma^2) + M\log(2\pi) + \frac{1}{\sigma^2}\sum_{j=1}^{M}\left(\frac{y_j - \hat{y}_j}{w_j}\right)^2. \tag{3.13}$$

Note that in writing "OFV", the data and parameters it depends on have been omitted. Now maximum likelihood is obtained when OFV is minimal; constant terms such as $M\log(2\pi)$ may then be discarded (for example, in NONMEM's calculation of the the objective function). The minimum is attained for certain values of parameters of $\hat{y}$, and for the parameter value of σ^2, when the derivative of OFV with respect to that parameter is zero:

$$\frac{\partial \mathrm{OFV}}{\partial \sigma^2} = \frac{M}{\sigma^2} - \frac{1}{(\sigma^2)^2}\sum_{j=1}^{M}\left(\frac{y_j - \hat{y}_j}{w_j}\right)^2 = 0,$$

so the maximum likelihood estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{M}\sum_{j=1}^{M}\left(\frac{y_j - \hat{y}_j}{w_j}\right)^2.$$

By subsituting this estimate in eq. (3.13), we obtain

$$\mathrm{OFV} = \sum_{j=1}^{M}\log(w_j^2) + M\log(\hat{\sigma}^2) + M\ \log(2\pi) + M. \tag{3.14}$$

By substituting this result in eq. (3.1), we have

$$\mathrm{AIC} = \sum_{j=1}^{M}\log(w_j^2) + M\log(\hat{\sigma}^2) + M\ \log(2\pi) + M + 2D.$$

The term $2D$ arises from the fact that in minimizing the Kullback-Leibler information, *i.e.*, a measure of the distance between reality and the best approximating model, expectations have to be taken over a data space leading to estimates of parameters $\boldsymbol{\theta}$ (and hence $\hat{y}$, and possibly w (see below)) and over a second independent data space y.[18] So AIC as defined above should on average be approximately equal the value of OFV (eq. (3.13)), with estimated values for the parameters and validation data z_j denoted OFV_v:

$$\mathrm{OFV}_v = \sum_{j=1}^{M}\log(w_j^2) + M\log(\hat{\sigma}^2) + M\log(2\pi) + \frac{1}{\hat{\sigma}^2}\cdot\sum_{j=1}^{M}\left(\frac{z_j - \hat{y}_j}{w_j}\right)^2. \tag{3.15}$$

So when OFV and AIC are both minimized, the latter term - the sum of squared weighted prediction errors - should also be minimal. For the plots in this paper, the measures OFV, OFV_v, AIC, and AIC_c, were normalized by dividing them by the number of data samples. With an infinite amount of data, and $\hat{\sigma}^2 = \sigma^2$, the normalized criteria should attain the value of $\log(\sigma^2) + \log(2\pi) + 1$.

Note that if the weights w_j are taken as in subsection "Data simulation", the term $\sum\log(w_j^2)$ vanishes (this is a just a curiosity of that choice of weights); if the w_j are taken as the measurements y_j, the expectation of this term is the same for every K (for every model considered here). However, if the weights are taken as the model output $\hat{y}_j$, the expectation of the term will not vanish for a less than perfect model, and will differ between different models. To compare their ν^2, the weights for all models could be fixed to the model output of the best model - but since that is unknown at this point - to the output of the largest model.

For population data, the likelihood function is the product across individual marginal likelihoods where the random effects have been integrated out. For one individual i, and the model given by eq. (3.6), the likelihood L_i is

$$L_i = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^M\cdot\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\sum_{j=1}^{M}\left(\frac{\exp(\eta_i) + \epsilon_{ij} - \exp(\eta')}{\sigma}\right)^2\right]\cdot\frac{1}{\omega\sqrt{2\pi}}\cdot\exp\left[-\frac{1}{2}\left(\frac{\eta'}{\omega}\right)^2\right]d\eta'.$$

The ϵ_{ij} have on average mean zero and variance σ^2, and NONMEM's first-order condi-

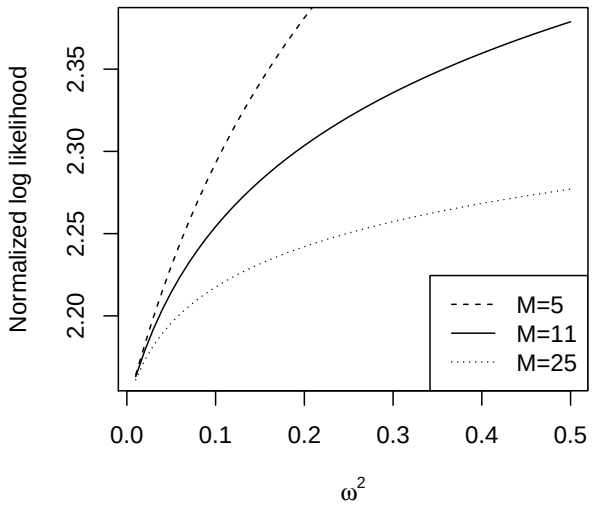

Figure 3.6: Theoretical values of the normalized log-likelihood with $\sigma^2 = 0.5$, as a function of ω^2 (interindividual variability), for different values of M (the number of observations per individual).

tional estimation method linearizes around the empirical Bayesian estimate of η_i, so that $\exp(\eta') = \exp(\hat{\eta}_i) \cdot (1 + \eta')$. The equation then reduces to

$$L_i = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^M \cdot \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} \cdot M \cdot \left(1 + \left(\frac{\eta' \cdot \exp(\hat{\eta}_i)}{\sigma}\right)^2\right)\right] \cdot \frac{1}{\omega\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2}\left(\frac{\eta'}{\omega}\right)^2\right] d\eta';$$

next some algebra gives for the expected value of minus two log L_i:

$$-2\log L_i = M \cdot (\log(\sigma^2) + \log(2\pi) + 1) + \log(M \cdot \exp(2\eta_i) \cdot \omega^2/\sigma^2 + 1). \qquad (3.16)$$

The minus two log-likelihood for the population data is the sum of N individual $-2\log L_i$. Now let the expected normalized likelihood be $\overline{NL}$, which is the expected population minus two log-likelihood divided by $N \cdot M$, taking into account that the η_i have on average mean zero and variance ω^2:

$$\overline{NL} = \log(\sigma^2) + \log(2\pi) + 1 + \frac{1}{M}\int_{-\infty}^{\infty} \log(M \cdot \exp(2\eta') \cdot \omega^2/\sigma^2 + 1) \cdot \exp\left[-\frac{1}{2}\left(\frac{\eta'}{\omega}\right)^2\right] d\eta'. \qquad (3.17)$$

Figure 3.6 depicts the normalized log-likelihood (with eq. (3.17) evaluated numerically) as a function of ω^2, for $\sigma^2 = 0.5$ and three values of M. For large M, the last term in eq. (3.17) (the integral divided by M) goes to zero, and the uncertainty left in the data is determined only by σ^2. Values for $M = 11$, and $\omega^2 = 0$, 0.1, and 0.5 were used as "target" values in Figures 3.3 - 3.5. The observed averaged normalized log-likelihoods will be larger, because the models used do not fit perfectly, and the parameters are estimated instead of set to their true values.

The context of AIC is also the one where the ηs have been integrated out (but with the parameters at their estimated values), which is to be done when all data are acquired. So while the characteristics of the set of (validation) data are optimally captured, this context is different from the case where prediction errors are calculated with the random effects set to zero instead of integrated out. In that case, the above AIC and OFV_v criteria do not match, as the components of the likelihood in eq. (3.12) are no longer independent (they can only independent if the true values of η for the individuals are also zero). Note however, that from the higher perspective of optimally characterizing a future set of population data, this is a less important case.

Chapter 4

Stochastic Pharmacokinetic-Pharmacodynamic Analysis of the Effect of Transdermal Buprenorphine on Electroencephalogram and Analgesia*

OPIOIDS ARE USED WIDELY in the treatment of moderate-to-severe cancer and noncancer pain.[91] There are currently 2 monumental challenges in the treatment of chronic pain: the objective assessment of opioid effect in a setting in which abuse and accidental overdose is highly prevalent and the need for proper dosing strategies. The efficacy of opioids and other centrally acting analgesics often is determined rather subjectively by the use of quantitative sensory testing, questionnaires, and so on.[59] To determine a suitable objective biomarker as a measure of opioid drug effect is challenging.

One possibility is the electroencephalography (EEG), which is a widely available and noninvasive tool for recording brainwave activity simultaneously from multiple brain regions. Several drug classes that act on the central nervous system generate a reproducible effect on the EEG obtained at rest.[46,50] For example, Liley *et al.*[48] showed that the effect of remifentanil on frontally recorded resting EEG could be dissociated from the EEG effects of propofol, an anesthetic acting at a different receptor target in the central nervous system. The EEG is therefore of great interest in evaluating the effect of drugs used in anesthesia and pain treatment.

Because opioid effects are delayed relative to their plasma concentration profile, because of the time needed to reach and interact with the opioid receptors, a pharmacokinetic-pharmacodynamic (PK-PD) analysis may be used.[49,52] PK-PD analysis links dose to effect and makes it possible to take inter- and intraindividual variability into consideration when designing appropriate dosing strategies.[52] Indeed, a drug-induced EEG effect can produce a dynamic outcome applicable in PK-PD modeling, which may be

* A E Olesen, E Olofsen, T Andresen, C Graversen, A M Drewes, A Dahan, *Anesth Analg* 2015; 121:1165–1175

used to determine population-predicted values for dose and effect, leading to a more rational approach for effective dosing regimens.

In the current study, we assessed the effect of transdermal buprenorphine on the resting EEG and experimental pain in healthy volunteers to elucidate the PK-PD profile of transdermal buprenorphine. Transdermal buprenorphine is an appealing treatment for chronic pain, because it is an agonist for analgesia but a partial agonist for respiratory depression over its clinical dose range.[26,25] Modeling the effect of an opioid given by a transdermal patch should consider the possibility that the absorption rate may not be constant. For example, changes in skin temperature may lead to changes in drug absorption from the patch.[56] Hence, the PK-PD model that we apply should take into account a variable uptake of drug from patch or dermal reservoir.

Here, we applied a stochastic PK-PD technique that accounts for varying drug absorption as first described by Tornøe *et al.*[89] We previously applied a stochastic PK-PD (SPKPD) model to assess the effect of ketamine on cardiac output and chronic pain relief.[64,24]

We measured 2 opioid effects: pain relief and changes in EEG. Rather than using conventional indices derived from the EEG, such as spectral edge, median frequency, peak frequency, or power spectrum, we used the ratio of slow-to-fast EEG frequencies as a biomarker for opioid effect. Dichotomizing the frequency spectrum into low- and high-frequency components has the advantage in that it compensates for interindividual variability in the frequency distributions and minimizes the number of EEG features traditionally obtained from the various frequency bands.[88]

We hypothesize (1) that the resting EEG is a reliable and objective surrogate for buprenorphine's effect and (2) that SPKPD analysis allows the computation of the time-dependent variability in drug absorption from patch to blood. Our approach will lead to a better understanding of the behavior of the patch.

4.1 Methods

This double-blind, randomized, placebo-controlled, crossover study was approved by the North Denmark Region Committee on Health Research Ethics and the Danish Health and Medicines Authority and registered at ClinicalTrials.gov under number NCT00647127. The study was performed according to the principles of Good Clinical Practice of the European Union from June 2008 until August 2009 in the research laboratories of Mech-Sense, Aalborg University Hospital, Denmark, and all subjects gave written informed consent. Descriptions and analyses of portion of the data were reported previously.[6,5,7,33,81,82] These reports include data on the effect of buprenorphine and fentanyl on evoked potentials, analgesia, and antihyperalgesia using a set of nociceptive tests (including pressure pain, ultraviolet B light burn injury, intradermal capsaicin-induced hyperalgesia, and conditioned pain modulation).

4.1.1 Study Design

Twenty-two healthy, opioid-naive male volunteers (mean age 23.1 ± 3.8 years) were recruited to participate in the study. Subjects received a transdermal patch (Norspan^{TM} 144-h; Norpharma, Vedbæk, Denmark) or a placebo patch (Norpharma) identical in appearance for 144 hours, followed, after removal of the patch, by a 3-day follow-up period.

A washout period of 10 days was observed between treatments. The subjects were hospitalized during the treatment phase, with regular assessments of blood pressure, heart rate, respiratory rate, and oxygen saturation.

An independent pharmacist performed the randomization using an electronic randomization list downloaded from randomization.com. Sample size calculation was based on previous studies on the influence of opioids on experimental heat pain and was used to set the number of subjects to detect an effect in these previous descriptive studies.[81,82] To show an increase in pain tolerance threshold of 2°C (with a power of 90%, SD = 1.70 and $\alpha = 0.05$), 16 subjects are required in each group. Taken into account the variability and possible loss of data, the number in each group was increased to 22.

Blood Sampling and Buprenorphine Assay

Nine microliters of venous blood samples were collected in EDTA blood collection tubes at baseline and 6, 9, 12, 24, 36, 48, 60, 72, 78, 84, 96, 120, 144, 168, 192, and 216 hours after application of the patches. The blood samples were immediately centrifuged at 4°C at 3000 rpm for 15 minutes. Next, plasma was separated into two 2-mL polypropylene tubes (the second sample served as duplicate). Both tubes were stored at -80°C until analysis. The buprenorphine analysis has been published before.[6]

Thermal Cutaneous Stimulation

The response to a noxious thermal stimulus was obtained at baseline and 24, 48, 72, and 144 hours after application of the patches. Pain was applied using a thermode (TSA II NeuroSensory Analyser; Medoc Ltd, Ramat Yishai, Israel) applied to the right volar forearm. The temperature increased from a baseline of 32°C to a maximum of 52°C with a rate of 1°C/s. The subject pressed a button on reaching the heat pain tolerance threshold. Three consecutive stimulations were performed, and the average was computed. All subjects were familiarized with the procedure before the study.

The Electroencephalogram

The resting EEG was recorded at baseline and 4, 24, 48, 72, and 144 hours after application of the patches. An EEG amplifier (NuAmp; Neuroscan, El Paso, TX) was used to record the electrical activity on the scalp. Two electrodes were placed at Cz and CPz locations according to the international 10-20 system. In addition, one electrode was mounted at the right earlobe serving as reference, whereas one electrode was placed 2 cm frontal to the Cz electrode serving as ground electrode. The electrodes were mounted using electrode gel to reduce the impedance to < 5 kΩ, and the positions of the electrodes were maintained during the experiment by using an elastic fixation cap (Carefix Head, Ikast, Denmark). The EEG data were recorded with a sampling rate of 1 kHz. The data were recorded with an online notch filter at 50 Hz and band pass filter with cutoff frequencies of 0.5 and 300 Hz. Resting EEG recordings were obtained after pain tests and blood sampling by a research nurse in a quiet room with dimmed light as participants lay in supine position with eyes open.

The EEG signals were processed off-line. The processing included the following steps: (1) Artifact rejection by visual inspection, leaving at least 1 minute of valid recording for further analysis (Neuroscan version 4.3.1; Compumedics, El Paso, TX); and (2) high-pass

filtering to remove DC offset and linear detrending by a first-order Butterworth filter with cutoff frequency 0.5 Hz performed in MATLAB (The MathWorks Inc., Natick, MA).

4.1.2 Data Analysis

EEG Analysis: Time-Frequency Analysis

Time-frequency analysis can be applied in several ways; recent studies showed that a continuous wavelet transform (CWT) is advantageous over more traditional methods such as the Fourier transform.[3] The CWT is based on a mother wavelet function, which was a complex Morlet wavelet for the current study.[70,84,35,27] The time-frequency coefficients of the Cz EEG channel were rectified and integrated over time to obtain the marginal distribution in the frequency bands: δ (0.5-3.5 Hz), θ (3.5-7.5 Hz), α_1 (7.5-10.5 Hz), α_2 (10.5-13.5 Hz), β_1 (13.5-18.5 Hz), β_2 (18.5-24.5 Hz), and β_3 (24.5-32 Hz). The frequency bands were normalized into percentage of the total power (0.5-32 Hz). An EEG ratio was used to evaluate the results. The EEG ratio was defined as the percentage sum of the slow conducting frequency bands ($\delta + \theta + \alpha_1$) divided by the percentage sum of the fast conducting frequency bands ($\alpha_2 + \beta_1 + \beta_2 + \beta_3$).

Stochastic Model for Buprenorphine Absorption

We assume that the absorption rate of buprenorphine from patch into the disposition compartment varied over time; see Figure 4.1. The noise in the absorption (*i.e.*, process noise) was modeled using the following stochastic differential equations:[89,64,24,90,61,53,42]

$$dA_a(t)/dt = -k_a(t) \cdot A_a(t) \quad (4.1)$$

$$dA_d(t)/dt = k_a(t) \cdot A_a(t) - k_e(t) \cdot A_d(t) \quad (4.2)$$

$$k_a(t) = \exp(Z(t)) \quad (4.3)$$

$$dZ(t) = \sigma_w \cdot dw(t), \quad (4.4)$$

where $A_a(t)$ is amount of drug in the absorption compartment at time t ($A_a(0) = 20$ mg), $A_d(t)$ is the amount of drug in the disposition compartment, $k_a(t)$ is the absorption rate (set to zero when the patch is removed), $k_e(t)$ is the elimination rate constant, Z is a link variable, $w(t)$ is the Wiener process, and σ_w is the standard deviation of changes in $Z(t)$ per the square root of time (*i.e.*, σ_w is the variability in the absorption rate constant in the log domain). A Wiener process is a model of Brownian random motion resulting from a sum of many small normally distributed fluctuations.[32] This parameterization constrains $k_a(t)$ to be positive; $k_a(t = 0)$ is a variable to be estimated. The buprenorphine concentration is given by the ratio of A_d and the volume of distribution, V_d.

Pharmacodynamic (PD) Analysis

The PD part of the models assumes an effect compartment in which the drug appears with a delay:

$$dC_e(t)/dt = k_{e0} \cdot (C_d(t) - C_e(t)) \quad (4.5)$$

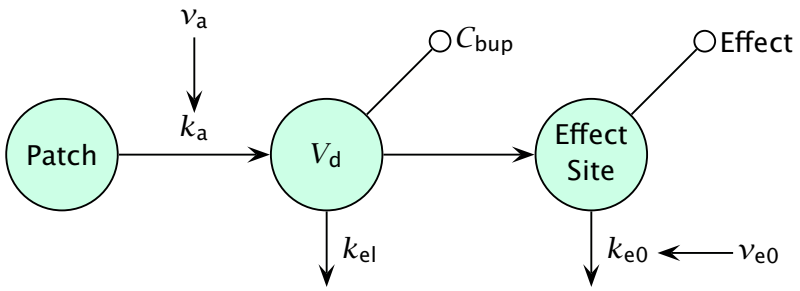

Figure 4.1: Schematic representation of the stochastic pharmacokinetic-pharmacodynamic model, in which the pharmacokinetic part consists of the transfer of drug from patch to disposition compartment V_d with rate constant k_a. Fluctuations in k_a are modeled by noise process ν_a. V_d is linked to the effect-site compartment. k_{e0} is the blood–effect-site equilibration constant; possible fluctuations are modeled by noise process ν_{e0}.

where $C_d(t)$ is the drug concentration in the disposition compartment at time t, $C_e(t)$ is the effect-site concentration at time t, and k_{e0} is the blood–effect-site equilibration rate constant. The PD effect (EF) was assumed to be related to $C_e(t)$:

$$\mathrm{EF}(t) = \mathrm{BLN} \cdot (1 + (C_e(t)/C_{100})^{\gamma}) \tag{4.6}$$

where BLN is the baseline value, C_{100} is the effect-site concentration causing a 100% increase in surrogate effect measure, and γ is a shape parameter. Finally, an additional stochastic differential equation for k_{e0} was tested.

Statistical Analysis

The population pharmacokinetic-pharmacokinetic (PK-PD) analyses were performed by implementing the models in the statistical software package NONMEM (version VII, level 2; Icon Development Solutions, Hanover, MD). For the SPKPD analysis, an extended Kalman filter was incorporated.[90,61,53,42] NONMEM's subroutine ADVAN13 was used to integrate drug amounts in absorption, disposition, and effect compartments and additional Kalman filter state variables. PK and PD data were analyzed simultaneously. The 2 PD data sets were analyzed separately (PK/resting EEG, PK/skin pain tolerance). Residual error was assumed to have both an additive and a relative error for concentrations and only an additive error term for the PD end points. Goodness-of-fit plots were created for the PK and PD data to check for model adequacy and possible outliers. P values < 0.01 were considered significant.

4.2 Results

4.2.1 EEG Spectrum and Pain Response

EEG data were not available from 3 subjects because of technical problems with the EEG equipment. These subjects did provide pain data. The mean age of the subjects (± SD, n = 22) was 22.5 ± 1.8 years, mean height 181.2 ± 5.8 cm, and mean weight 73.3 ± 7.4 kg.

Figure 4.2 displays the impact of buprenorphine and placebo on the spectral distribution of the EEG at baseline and after 72 hours of patch application.

Figure 4.2E shows a greater shift from fast to slow activity of the EEG spectrum in the buprenorphine group after 72 hours (dotted and solid red lines) than in the baseline group (dotted and solid blue lines). The absolute and relative individual and average EEG ratio and pain tolerance data are given in Figures 4.3 and 4.4. Compared with placebo, buprenorphine increased EEG ratio by 0.2 to 0.3 points (paired t test: P = 0.006 at t = 48 hours; P = 0.0006 at 72 hours; P = 0.001 at 144 hours) and heat pain tolerance threshold by 1 to 2°C (paired t test: P = 0.0008 at 48 hours; P = 0.005 at 72 hours; P = 0.03 at 144 hours).

4.2.2 PK-PD Analysis

The individual and average plasma buprenorphine concentrations are given in Figure 4.3E and Figure 4.3F. A separate PK-PD analysis was performed on the PK/resting EEG ratio data and the PK/skin pain tolerance data. In none of the analyses, parameter γ was significantly different from 1, indicating that the PD effect was linearly related to the buprenorphine effect-site concentration.

PK Analysis

The PK parameters are given in Table 4.1. The initial value for the absorption rate constant (k_a) was 0.005 h^{-1}, elimination rate constant 0.04 h^{-1}, and volume of distribution 11 L. Similarly, the standard deviation of the noise of the absorption process and standard deviations and residual errors were of similar magnitude between analyses (σ_w 0.11 $1/\sqrt{h}$, σ_1 0.01 ng/mL, and σ_2 0.14). The variability in the absorption rate is quantified by σ_w, indicating that k_a varies by 0.11 per hour and, for example, will range between approximately 0.003 h^{-1} and 0.008 h^{-1} after 10 hours of the patch application if no information via PK or PD samples is obtained.

PK goodness-of-fit plots are given in Figure 4.5, and examples of PK data fits are given in Figure Figure 4.6, (panels G, H, and I). Both show that the SPKPD model adequately described the PK data. For none of the PK parameters was interindividual error estimable, which indicates that the variability in the parameter estimates was mainly caused by within-subject, rather than between-subject, variability. In none of the subjects did drug absorption remain constant during the 144-hour buprenorphine treatment, as observed by the fluctuations in ka over time (Figure 4.6, panels J, K, and L).

PD Analysis

Examples of PD data fits are given in Figure 4.6 (panels A-C and D-F). For EEG ratio, the best, median, and worst fits are given as based on the coefficient of determination (R2). Note that a negative value for R2 was observed for the worst fit, indicating that the fit is worse than just using the mean of the data. PD parameter estimates are given in Table 4.1. For all data fits, the 95% confidence intervals were calculated (broken lines in Figure 4.6). These intervals are based on both the measurement and the prediction errors and may therefore vary in time, depending on the information obtained from the measurements (PK or PD), which is fed back to the stochastic differential equations of

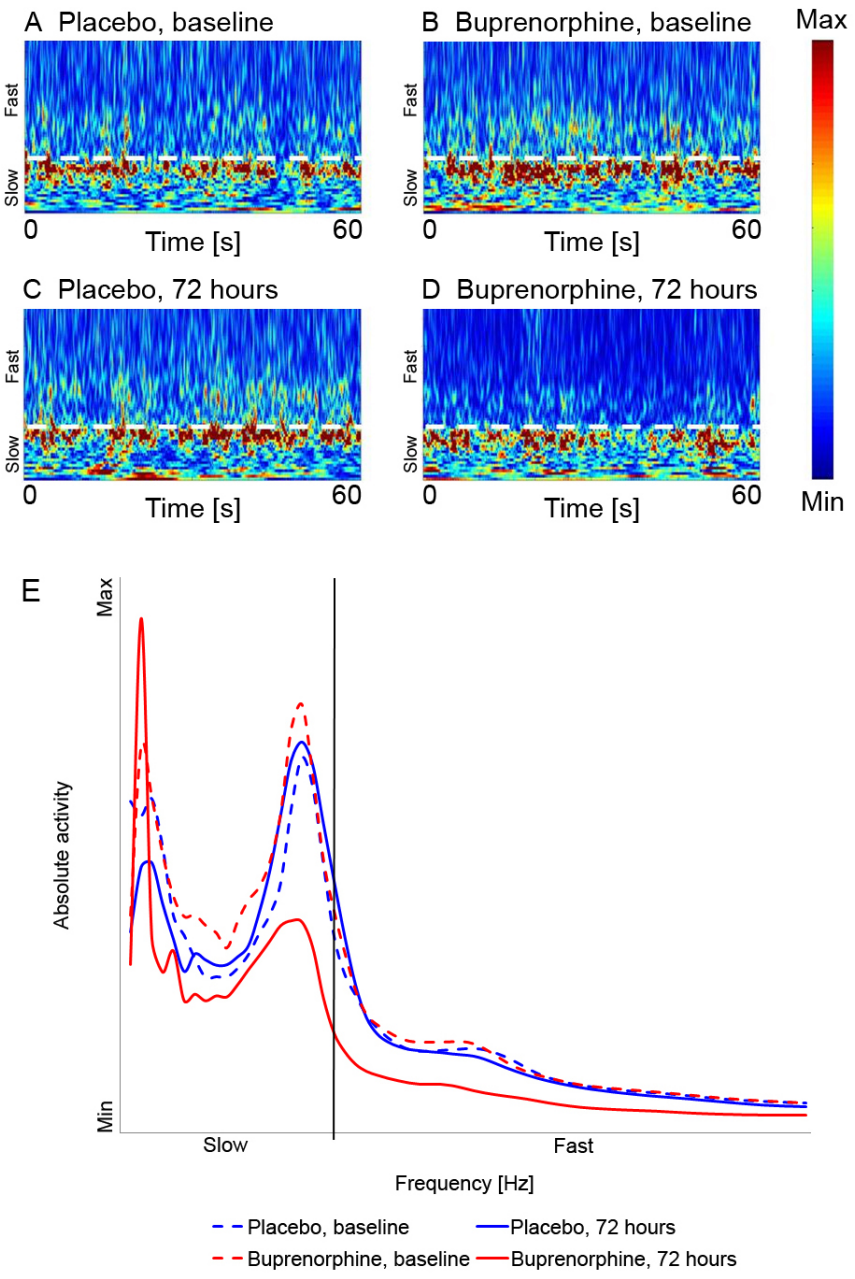

Figure 4.2: **A-E**: Examples from one subject of the spectral distribution of the resting electroencephalography measurement at baseline (**A** and **B**) and after 72 hours of placebo (**C**) and buprenorphine (**D**). A shift is visible after 72 h of buprenorphine treatment from fast toward slow oscillations. **E**: Frequency *versus* absolute activity for buprenorphine and placebo treatment at baseline and 72 h. Absolute activity was calculated by a continuous wavelet transform using a complex Morlet function with bandwidth parameter of 128 Hz and wavelet center frequency of 0.5 Hz.

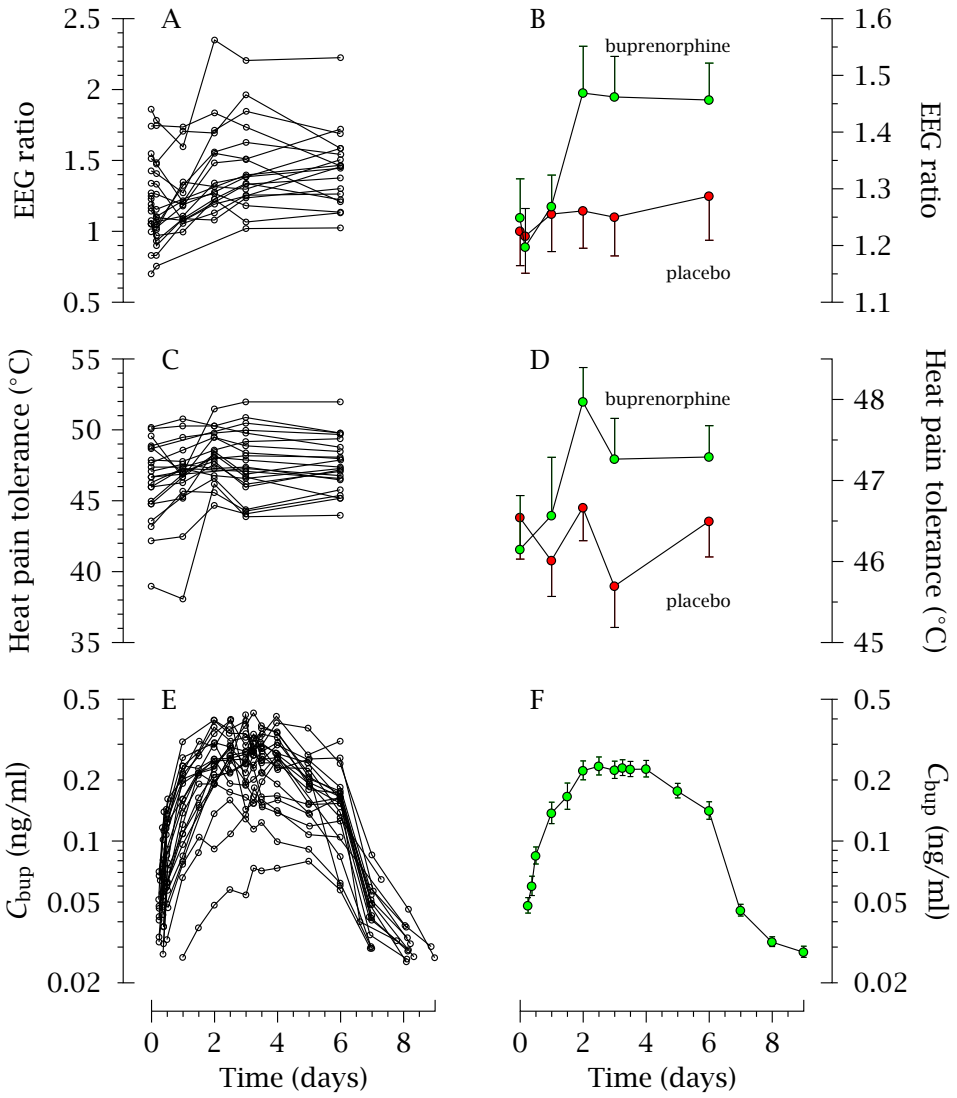

Figure 4.3: Individual data (**A**) and averages (**B**) of the effect of a 144 h (6 day) administration of buprenorphine by transdermal patch on the electroencephalography (EEG) ratio. In panel B, the placebo averages are given. Over time buprenorphine increased the EEG ratio significantly compared with placebo. Individual data (**C**) and averages (**D**) of the effect of the 144 h buprenorphine patch on skin heat pain tolerance (units °C). In panel D, the placebo averages are included. Over time buprenorphine increased skin heat pain tolerance significantly compared with placebo. Individual (**E**) and average (**F**) buprenorphine plasma concentrations during and 2.5 days after the buprenorphine patch application. In panels A, C, and E, each line represents one subject; in panels B, D, and F, the data are mean ± SEM.

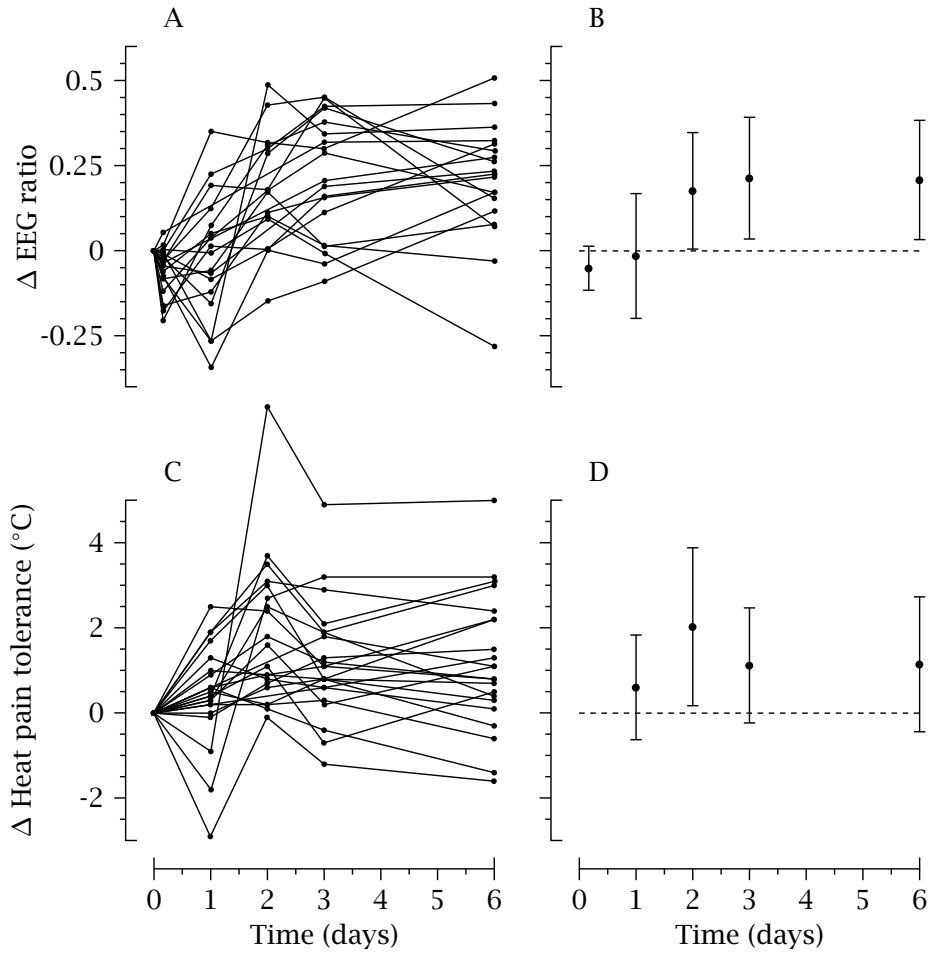

Figure 4.4: **A**-**D**: Individual and average electroencephalography ratio and pain tolerance relative to baseline values.

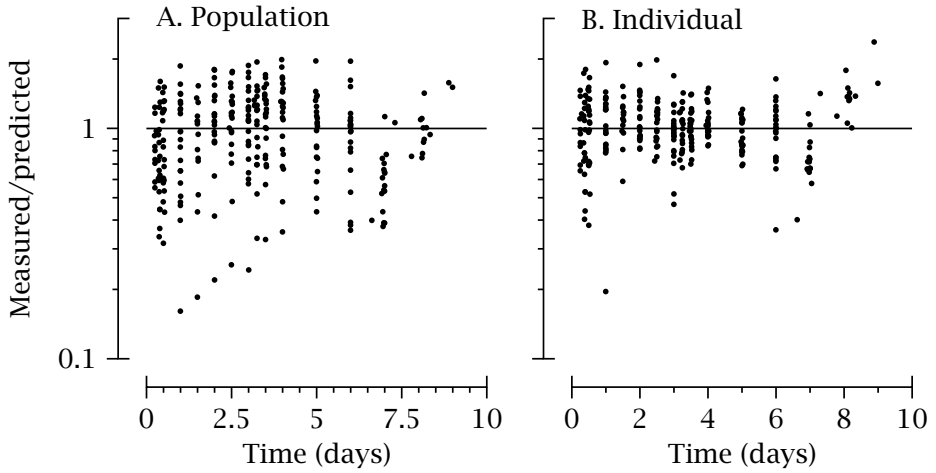

Figure 4.5: Pharmacokinetic goodness-of-fit plots. **A**: Measured *versus* population-predicted plasma buprenorphine concentrations. **B**: Measured *versus* individual predicted plasma buprenorphine concentrations.

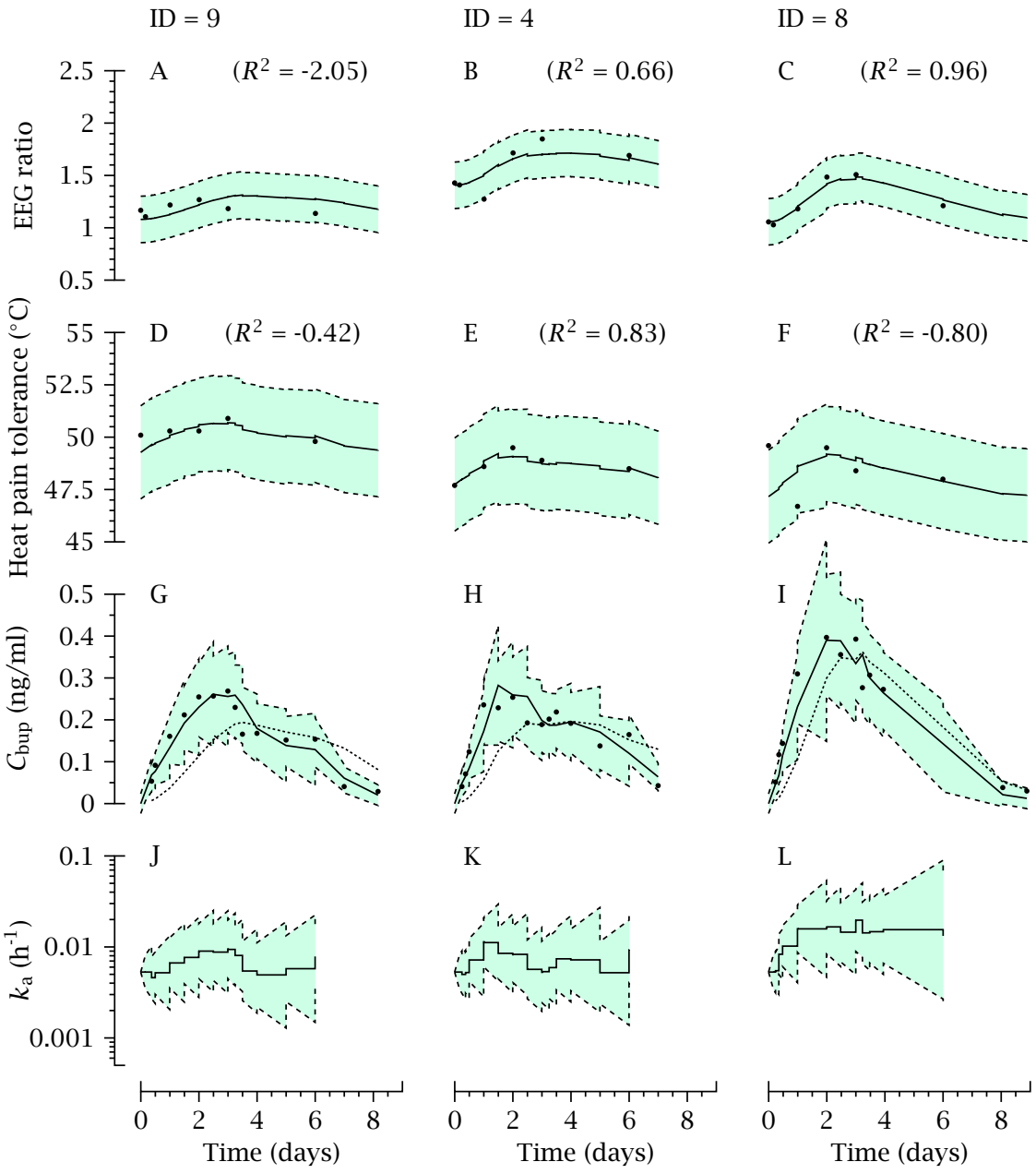

Figure 4.6: Best (**A**), median (**B**), and (**C**) worst data fits of the electroencephalography (EEG) ratio and corresponding data fits of skin heat pain tolerance (**D**, **E**, and **F**), and plasma buprenorphine concentration (**G**, **H**, and **I**). Goodness of fit was based on the coefficient of determination (R^2). The bottom graphs (**J**, **K**, and **L**) depict the changing absorption rate constant (k_a) over time. The black dots are the measured data; the continuous lines are the data fit; and the broken lines are the 95% confidence intervals; in panels **G**-**I**, the dotted lines are the buprenorphine effect-site concentrations (derived from EEG data). Note that at t = 144 h the patch was removed (k_a set to 0).

Table 4.1: Parameter Estimates of the SPKPD Analysis

	Resting EEG ratio		Heat pain tolerance	
Parameter	Estimate ± SEE	ω^2 ± SEE	Estimate ± SEE	ω^2 ± SEE
k_a (h^{-1})	0.005 ± 0.001	*a*	0.005 ± 0.001	*a*
k_e (h^{-1})	0.04 ± 0.002	*a*	0.04 ± 0.002	*a*
V_d (L)[b]	11.6 ± 0.9	*a*	11.4 ± 0.9	*a*
σ_w ($1/\sqrt{h}$)	0.11 ± 0.01	*a*	0.11 ± 0.01	*a*
σ_1 (ng/mL)	0.01 ± 0.003		0.01 ± 0.003	
σ_2	0.14 ± 0.03		0.14 ± 0.03	
BLN[c]	1.18 ± 0.06	0.03 ± 0.03	46.5 ± 0.60	0.02 ± 0.006
$t_{½,k_{e0}}$ (h)	24.8 ± 8	0.7 ± 0.4		
C_{100} (ng/mL)	0.90 ± 0.10	*a*	9.2 ± 2.5	*a*
σ_3	0.11 ± 0.01		1.10 ± 0.20	

SPKPD = stochastic pharmacodynamic-pharmacodynamics;
EEG = electroencephalography; SEE = standard error of estimate;
ω^2 = between-subject variability (in the log-domain);
k_a = initial absorption rate constant, *i.e.*, at $t = 0$;
k_e = elimination rate constant; V_d = volume of distribution;
σ_w = standard deviation of the noise process (Z in Equations (4.3) and (4.4));
σ_1 and σ_2 = standard deviations of additive and relative error;
respectively, for the concentration estimates;
BLN = baseline value; $t_{½,k_{e0}}$ = blood–effect-site equilibration half-life;
C_{100} = effect-site concentration causing a 100% increase (*i.e.*, doubling) in effect;
σ_3 = additive error for the effect estimates with unit for pain tolerance °C;
[a]Not estimable; [b]V_d values are relative to the buprenorphine bioavailability (it is assumed that 100% of the buprenorphine absorbed from the patch becomes systemically available); [c]Unit for BLN pain tolerance is °C.

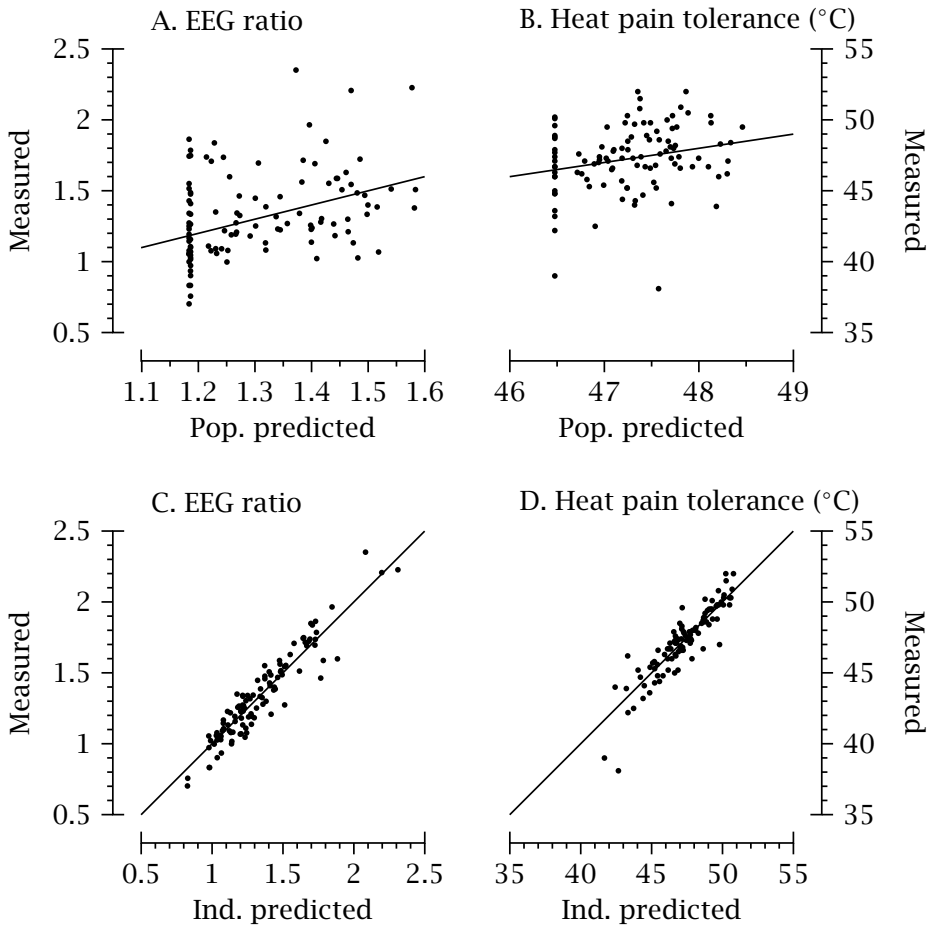

Figure 4.7: Pharmacodynamic goodness-of-fit plots. **A** and **C**, Measured electroencephalography (EEG) ratio *versus* individual-predicted and population-predicted EEG ratio. **B** and **D**, Measured skin heat pain tolerance (units °C) *versus* individual-predicted and population-predicted skin heat pain tolerance.

the model.[90] The updated states of the differential equations can be seen as sudden updates of the estimated absorption rates, concentrations, and PD end points.

Goodness-of-fit plots are given in Figures 4.7 to 4.9. Figure 4.7 shows the measured *versus* population-predicted (Fig. 4.7A, EEG ratio; and Fig. 4.7B, heat pain tolerance) and measured *versus* individual-predicted (Fig. 4.7C, EEG ratio; and Fig. 4.7D, heat pain tolerance) data. Figure 4.8 shows the spaghetti plots for EEG ratio error (Fig. 4.8A) and heat pain tolerance error (Fig. 4.8B). Figure 4.9 gives the log-likelihood profiles of the (PK and PD) model parameters. The objective function is most sensitive to changes in parameter BLN (EEG ratio, Fig. 4.9E; and heat pain tolerance, Fig. 4.9F) and least sensitive to changes in $t_{½,k_{e0}}$ (half-life from k_{e0}) (EEG ratio, Fig. 4.9C). A bootstrap analysis (1000 simulations drawing random samples from the subject pool) was performed to assess the sensitivity of the model output to exclude subjects from the data set (data not shown). The results show that excluding subjects did not result in systematic changes in parameter values. Overall, the inspection of the data fits and diagnostic plots indicates that the SPKPD adequately described the data.

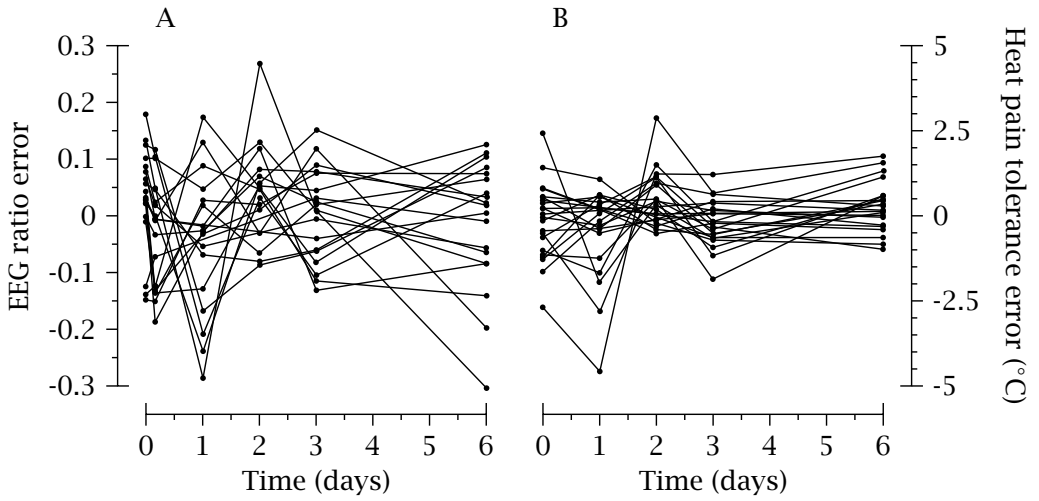

Figure 4.8: Spaghetti plots for electroencephalography ratio (**A**) and pain tolerance (**B**) showing residual error *versus* time.

Hysteresis

In Figure 4.10, examples of buprenorphine plasma concentration *versus* effect are plotted and show that no significant hysteresis was detected for pain tolerance. In contrast, a significant hysteresis was observed for EEG ratio, with a value for parameter $t_{½,k_{e0}}$ of 24 ± 8 hours. The log-likelihood profile of parameter $t_{½,k_{e0}}$ shows a rather flat surface profile with values that range from -50% to +100% of the optimal estimate within its 95% confidence interval (Fig. 4.9C). Removal of parameter $t_{½,k_{e0}}$ from the model resulted in an increase of Objective Function Value > 20 points. A stochastic differential equation to account for fluctuations of k_{e0} did not improve the data fits. Hence, this approach was discarded.

EEG Ratio *versus* Heat Pain Tolerance

The EEG ratio was more sensitive to buprenorphine than skin pain tolerance, with a 10 ± 3 (mean ± SE) times greater potency: resting EEG ratio C_{100} = 0.90 ± 0.10 ng/mL *versus* EEG ratio C_{100} = 9.01 ± 1.90 ng/mL. To get an indication of whether the EEG is a good predictor of heat pain tolerance, the 2 PD models were coupled via their corresponding plasma concentrations (obtained from taking the measurement variability σ, into account). Figure 4.11 shows that the EEG predicts heat pain tolerance with acceptable uncertainty compared with the skin test when coupling the PD models to their corresponding plasma concentrations.

4.3 Discussion

Our main findings are that the EEG ratio can be used as a surrogate measure of buprenorphine effect and that the SPKPD analysis, which includes tracking and update features, allowed the computation of the time-dependent variability in drug absorption from patch to blood. Both results confirm our study hypotheses. We demonstrated that buprenorphine's absorption varied over time, ranging from -40% to +60% of baseline

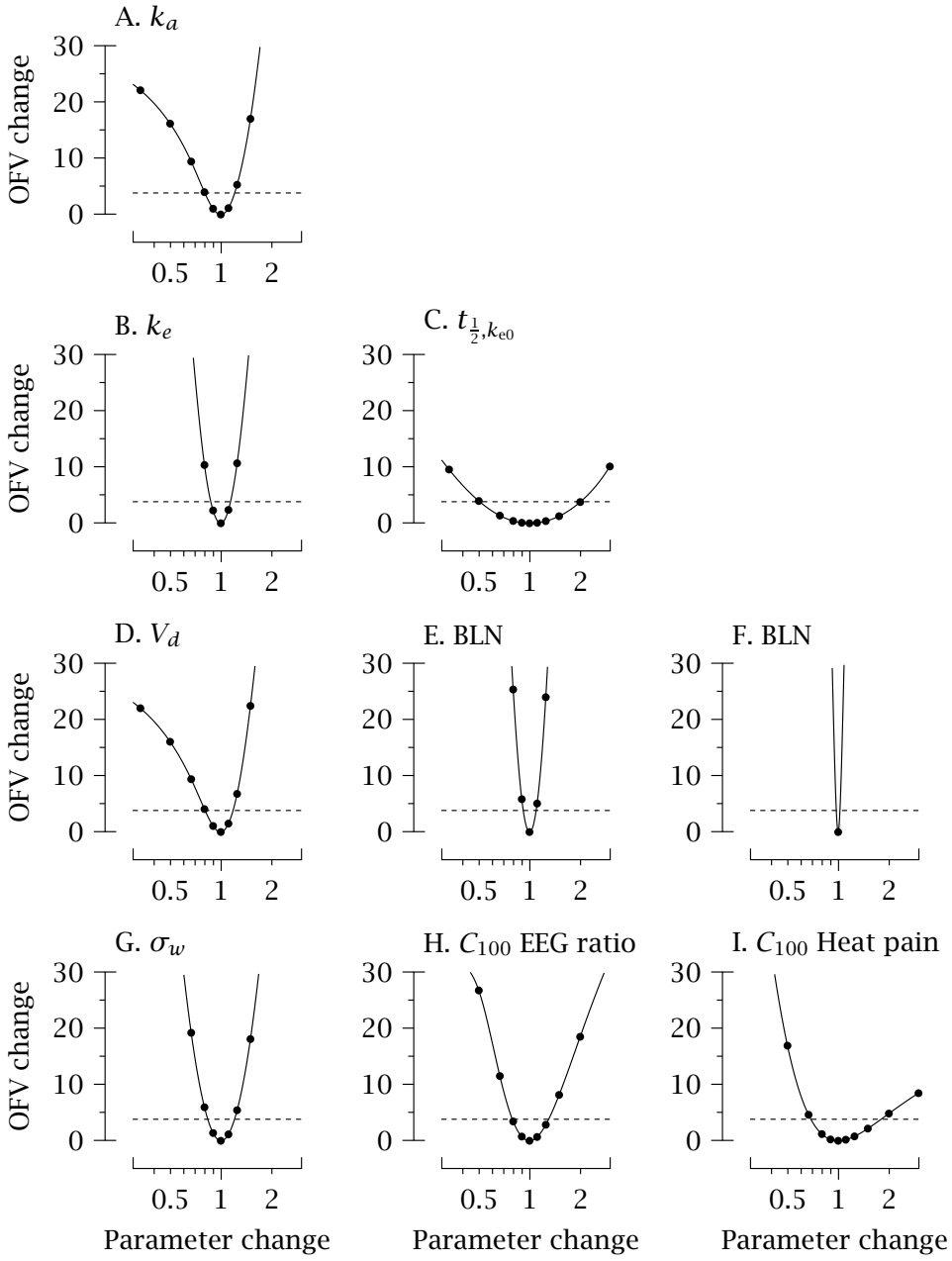

Figure 4.9: Likelihood profiles showing the change in objective function *versus* a relative change in the denoted parameter (**A-I**) while estimating the remaining parameters. The dashed line denotes a change of 3.84 points in objective function (OFV), indicating the $P = 0.05$ level. The crossings of the likelihood profiles with the dashed lines give a parameter range corresponding to a 95% confidence interval. (**A**) k_a is the buprenorphine absorption rate; (**B**) k_e is the buprenorphine elimination rate constant; (**C**) $t_{½,k_{e0}}$ is the blood–effect-site equilibration constant; (**D**) V_d is the volume of distribution; (**E**) Baseline (BLN) is the electroencephalography (EEG) ratio baseline estimate; (**F**) BLN is the heat pain tolerance baseline estimate; (**G**) σ_w is the variability in the absorption rate constant in the log domain; (**H**) C_{100} is the effect-site concentration causing a 100% increase in EEG ratio; (**I**) C_{100} is the effect-site concentration causing a 100% increase in heat pain tolerance.

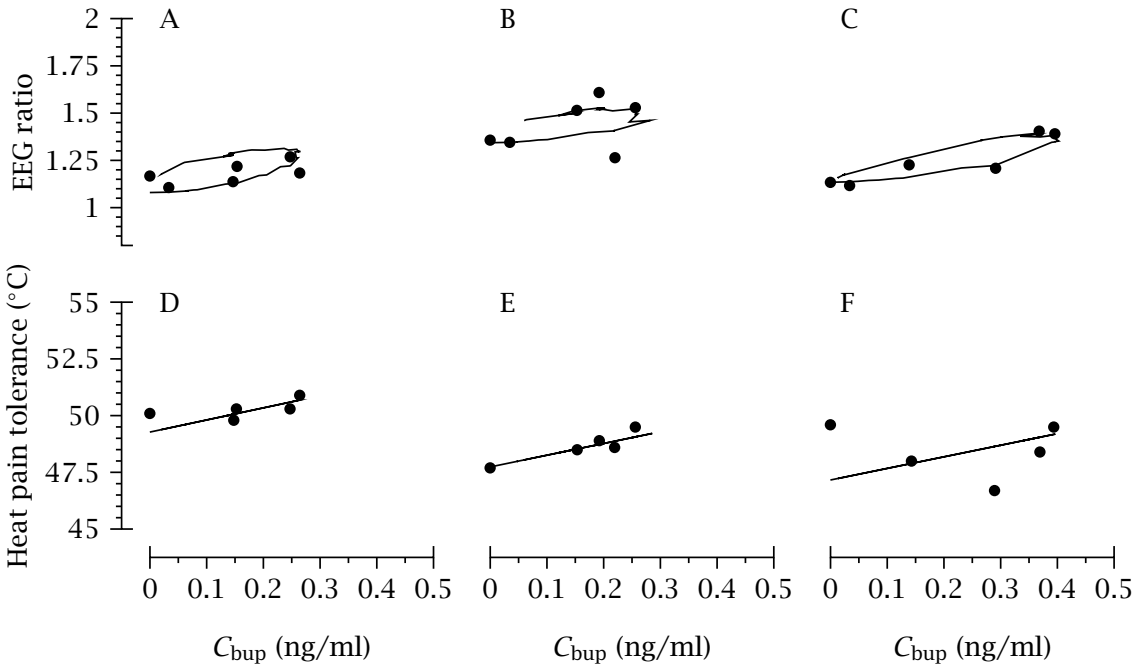

Figure 4.10: Random examples of effect *versus* plasma concentration for electroencephalography ratio (**A**-**C**) and pain tolerance (**D**-**F**). Each panel represents the data of one subject.

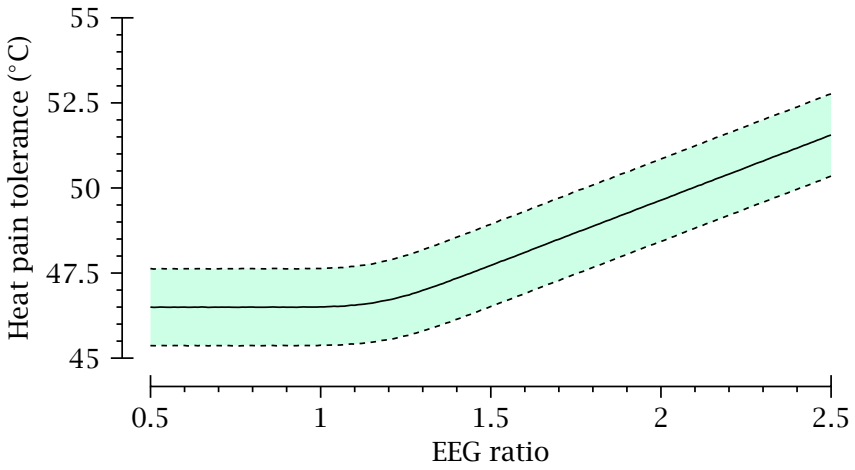

Figure 4.11: Relationship between the electroencephalography ratio and heat pain tolerance as derived from the 2 stochastic pharmacokinetic-pharmacodynamic (SPKPD) analyses. Shown is the median ± 1 SD.

absorption, and that buprenorphine's effect on the EEG ratio is 10 times more sensitive than buprenorphine's effect on dulling noxious skin stimuli.

4.3.1 Variations in Absorption Rate

The heat pain tolerance data were previously analyzed by Andresen et al.[6] The input to their PD model consisted of cubic splines fitted to the measured concentrations because no PK model could be found to adequately describe transdermal drug delivery. Although splines give smooth curves, they cannot correct for measurement error, which may result in possibly amplified errors of the interpolated values. Although the structural PK model that we applied may be simple, it provides the PD model with interpolated concentration values based on best estimates of drug absorption and disposition at the sampling times of the effect parameters.

Tornøe *et al.*[89] were the first to model subcutaneous drug absorption with a varying absorption rate. In the current study, we analyzed transdermal drug absorption using an approach similar to theirs. Assuming that the release of drug from the patch is constant over time, variations in drug absorption from the skin may be related to diurnal changes in local skin blood flow because of fluctuations in skin temperature, cardiac output, and ambient temperature. For example, the drug label for buprenorphine warns patients to avoid exposing the patch to external heat. For the fentanyl transdermal patch, heat-related toxicity has been described and was related to a significant (25%-30%) increase in plasma fentanyl concentration because of an increased drug release from the patch.[56] In our study, it is unknown whether the k_a fluctuations affected our PD outcome significantly. Theoretically, quantifying the fluctuations in k_a by modeling, the process noise could increase the precision of the estimate of the onset and offset of effect. However, the design of our study prohibited the precise estimation of $t_{½,k_{e0}}$. For heat pain tolerance, no hysteresis between plasma concentration and effect was detected, and for EEG ratio, the log-likelihood profile of $t_{½,k_{e0}}$ exhibits a rather large 95% confidence interval (12-48 hours), indicating that the support for hysteresis is limited. Still, excluding this parameter from the model had a significant negative effect on the objective function value, which suggests that the slow distribution of buprenorphine from plasma to brain is detectable in the EEG data. It is of interest to note that Andresen *et al.*[6] found a direct and linear effect of buprenorphine on heat pain tolerance similar to our observations.

4.3.2 EEG Ratio as Biomarker of Opioid Effect

The response to skin heat pain test is quite subjective, whereas the resting EEG is a more objective measure of drug effect.[59] This is the first study to assess the effect of the long-term administration of an opioid on the EEG and particularly on the EEG ratio. Most studies on the effect of opioids on resting EEG use Fast Fourier Transform to convert the raw EEG signal into quantifiable measures, such as spectral edge and median frequency.[49] Several of these studies show that slowing of the frequency of the EEG reflects a narcotic or sedative drug effect.[46,49,73,66] In the current study, a CWT was used to extract information from the raw EEG signal, followed by the evaluation of multiple frequency bands combined in a single EEG ratio. The design of the wavelet

analysis was chosen to be comparable with previous studies on analgesic effect and resting EEG.[35,34]

The rationale for using an EEG ratio is that opioids produce high-voltage cortical bursts associated with increases in EEG spectral power in predominantly the low-frequency range (0-10 Hz).[99] However, the frequency-specific alterations in cortical EEG oscillations depend on the receptor type that is activated.[99] Because buprenorphine is a mixed agonist-antagonist opioid receptor modulator (*i.e.*, acting at multiple opioid receptors), it seems more rational to assess the entire frequency range rather than the individual frequency bands independently. In addition, because the EEG power between subjects varies considerably, opioids may cause larger effects in some subjects (*e.g.*, subjects with an initial higher power). Using the ratio of the normalized EEG spectral distribution partly cancels out such bias, as the EEG ratio assesses the relative distribution between low- and high-frequency oscillations and quantifies how this balance is altered by buprenorphine administration in comparison with placebo treatment.

An interesting observation in our study is that the resting EEG effects were more sensitive to buprenorphine than the pain responses, with just one-tenth of the concentration at the effect-site required for a doubling of effect (C_{100}). This makes the EEG ratio an attractive biomarker of opioid effect compared with pain intensity testing when measuring the PD of opioid analgesics, especially when long-term administrations are tested.

The findings on the effect of buprenorphine on the resting EEG ratio were obtained in healthy male volunteers without coadministration of sedative hypnotic agents. For clinical use, it would be interesting to investigate in future studies whether the ratio would still be detectable in the presence of a potent IV or inhaled anesthetic agent. Although the mechanisms behind coadministration are not yet understood, several recent studies have been focused on this topic. Liley *et al.*[48] used a fixed-order autoregressive moving average model to analyze EEG signals from 2 frontal electrodes and found that during the simultaneous administration of remifentanil and propofol, increasing remifentanil concentrations caused significant changes in the cortical EEG. In line with this, Kortelainen *et al.*[47] used the frequency spectrum from the Fz channel to separate the effects of propofol and remifentanil and found the entire frequency range from 2 to 20 Hz to contribute to the remifentanil effect, although the low frequencies from 1 to 5 Hz showed the most discriminative oscillations.

4.4 Conclusions

In this study, the effect of transdermal buprenorphine on the cortical EEG (EEG ratio) and on heat pain tolerance was investigated; the EEG ratio was defined as (% slow frequency bands, 0-10.5 Hz)/(% fast frequency bands, 10.5-32 Hz). We showed that the EEG ratio is a reliable surrogate measure of buprenorphine effect, with a 10-fold greater sensitivity than heat pain tolerance. In addition, we successfully analyzed the data with a SPKPD model that allowed us to compute the time-dependent variability in drug absorption from patch to skin. The analysis showed a high variability in absorption, possibly related to diurnal variations in skin blood flow.

Chapter 5

Population Analysis of Kalman-Filtered Permutation Entropy of the Electroencephalogram*

General anesthetics produce dose-dependent effects on the electroencephalogram (EEG), causing an increase in power combined with a decrease in average EEG frequency.[62] A novel EEG-derived parameter is the permutation entropy (PE) of the EEG.[65] Important advantages are its robustness under eye blinks, and its ease of computation. The permutation entropy quantifies the probability distribution of motifs present in the signal, and it is determined by the dominant frequency in the EEG signal. But because of the ordinal (counting) nature of the PE, it is dominated by the presence of high EEG frequencies, even if they have quite small amplitudes.

Anesthetic concentration-effect data fits often show systematic misfits, due to correlated residuals, which could lead to biased standard errors of parameter estimates and false conclusions from statistical tests. Kalman filters may be constructed to separate measurement and process noise.[43,90] The first objective of the present study was to construct a pharmacokinetic-pharmacodynamic (PK-PD) model, including a Kalman filter, to analyze concentration-permutation entropy data sets. The second objective was to gain more insight into the effects of the incorporation of a Kalman filter, by fitting models with and without a Kalman filter to simulated data.

5.1 Methods

5.1.1 Ordinal Statistics and the Calculation of the Permutation Entropy

With the possible exception of zero-crossing rate, most EEG indices use the EEG signal as a continuously variable signal (to the limits of the measuring equipment). In contrast, ordinal statistics rank the data from smallest to largest, and then compare the rankings. Therefore, the use of ordinal descriptions of EEG may have the advantage of being re-

* Parts have been published in E Olofsen, J W Sleigh, A Dahan, *Br J Anaesth* 2008; 101:810–821, 2008, and E Olofsen, *PAGE* 2011: Abstract #2202

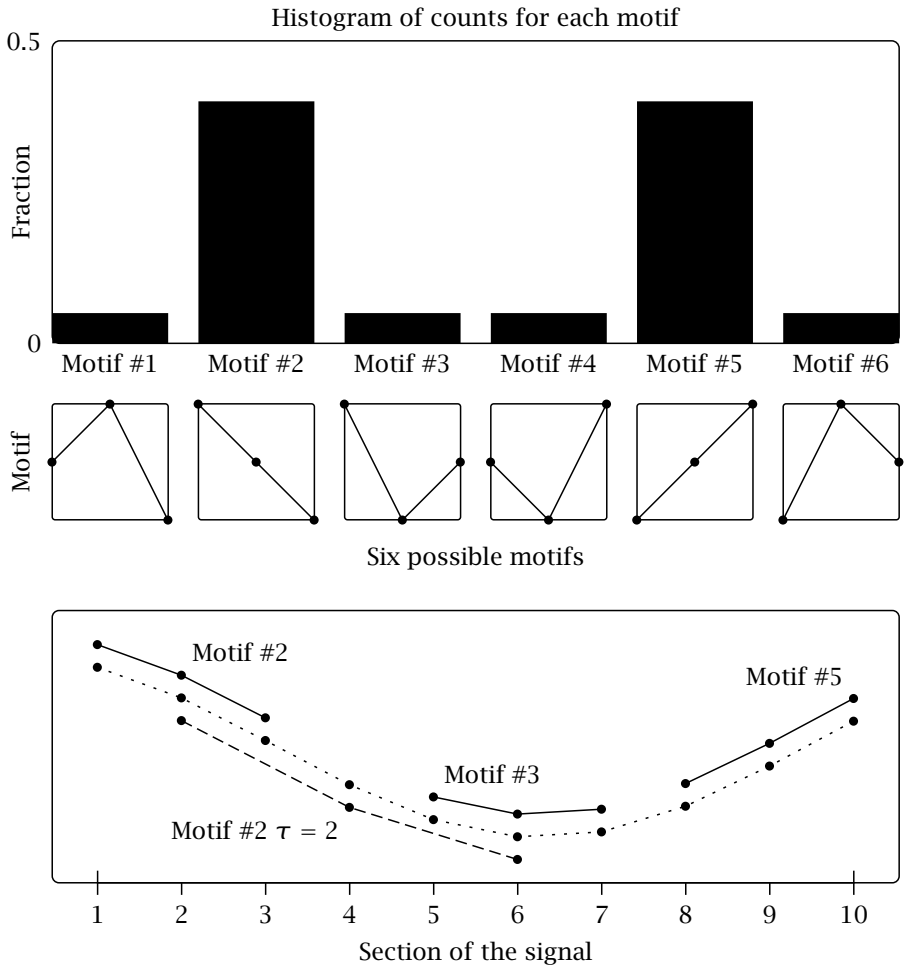

Figure 5.1: Lower panel: extraction of ordinal patterns from the EEG signal. As the algorithm moves sequentially through the EEG signal, the sections ("motifs" consisting of three data points' length) are classified as one of the six possible patterns, depicted in the middle panel. **Upper panel**: a histogram of the relative numbers of each motif in the signal. The dashed-line motif is a demonstration of the operation of the $\tau = 2$ lag.

sistant to large artifacts that occur with low frequencies. In this article, we will refer to the elemental patterns that are extracted from the EEG signal as "motifs" (see Figure 5.1 for graphical explanation). The EEG signal can therefore be considered to consist of a sequence of ordinal motifs.

The use of PE to quantify EEG changes in seizures was originally proposed by Bandt and Pompe,[8] and this work has been further developed by Cao and colleagues.[20] The algorithm of the calculation of the PE is quite simple and is depicted diagrammatically in Figure 1 as follows:

1. Fragment the continuous EEG signal (dotted line in the lower diagram of Figure 5.1), into a sequence of motifs (some examples are shown above and below the signal);

2. Identify each motif as belonging to one of the six possible types (as shown as 6 diagrams in the middle panel in Figure 5.1) - according to their shape [we describe

the six types as two varieties each of "slopes" (motifs #2 and #5 in Figure 5.1), "peaks" (motifs #1 and #6), and "troughs" (motifs #3 and #4)];

3. Count the number of motifs of each of the six categories, to obtain the probability of occurrence of each motif in the signal (p_i) (upper diagram in Figure 5.1);

4. Calculate the PE of the resultant normalized probability distribution of the motifs, using the standard Shannon uncertainty formula:[76]

$$\mathrm{PE} = -\frac{\Sigma p_i \cdot \log(p_i)}{\log(\text{number of motifs})} \tag{5.1}$$

Thus, the PE is a way of quantifying the relative occurrence of the different motifs. Like other entropies, the PE is simply a measure of the "spread-outness", or "flatness", or "uncertainty" in the frequency distribution. When the EEG signal is dominated by high frequencies, there will be almost equal numbers of each species of motif in each EEG segment analyzed. The properly normalized entropy is maximal (PE=1.0), if there is an equal distribution of motifs between each of the six patterns. Conversely, when the signal consists of slow delta waves, there will be relatively more of the "slope" motifs (motifs #2 and #5 in Figure 5.1), and fewer of the other "peak" and "trough" motifs, and the entropy decreases. The PE of a signal consisting of a single motif (such as one very long "up-slope") is zero. However, the effective realistic minimum value of the PE is about 0.4. It is important to note that the PE is very different from the spectral entropy in its frequency response. The PE tends to decrease as the frequency decreases Figure 5.2), whereas the value of the spectral entropy is completely independent of frequency *per se*, but only measures the sharpness of the frequency peak. To test the responses of the various PEs to variations in frequency, artificial "pseudo-EEG" signals were generated, using a C++ computer program which allowed known amounts of white noise to be added to various pure sine wave frequencies (Figure 5.2).

In summary, the PE quantifies the probability distribution of motifs present in the signal, and is determined both by the dominant frequency in the EEG signal and by the bandwidth. Because of the ordinal (counting) nature of the PE, this is dominated by the presence of higher EEG frequencies, even if they have quite small amplitudes.

5.1.2 Parameters and Ties

The PE has two predefined parameters. (i) The "order" of the PE is the number of data points that are included in each motif. We restricted our study to include only short motifs of just three points (order=3). Exploratory data analysis suggested that the use of longer motifs did not contribute to a better index of depth of anesthesia. (ii) The "lag" (τ) of the PE is the number of sample points spanned by each section of the motif. In the lower diagram in Figure 1, the dark grey motifs are of lag $\tau = 1$, because they are made up of adjacent data points. The longer light grey motif is of lag $\tau = 2$, because the length of each section of the motif is two data points. The importance of the lag is that it gives the resultant PE different frequency characteristics (Figure 5.2). Most of the anesthetic-related information in the EEG can be extracted using a lag ($\tau = 1$) of one sample step (assuming a sampling frequency of about 128 s^{-1}). However, as described in our earlier paper,[65] the inclusion of lag ($\tau = 2$) helped differentiate deeper planes of anesthesia, and resulted in better PK-PD modeling.

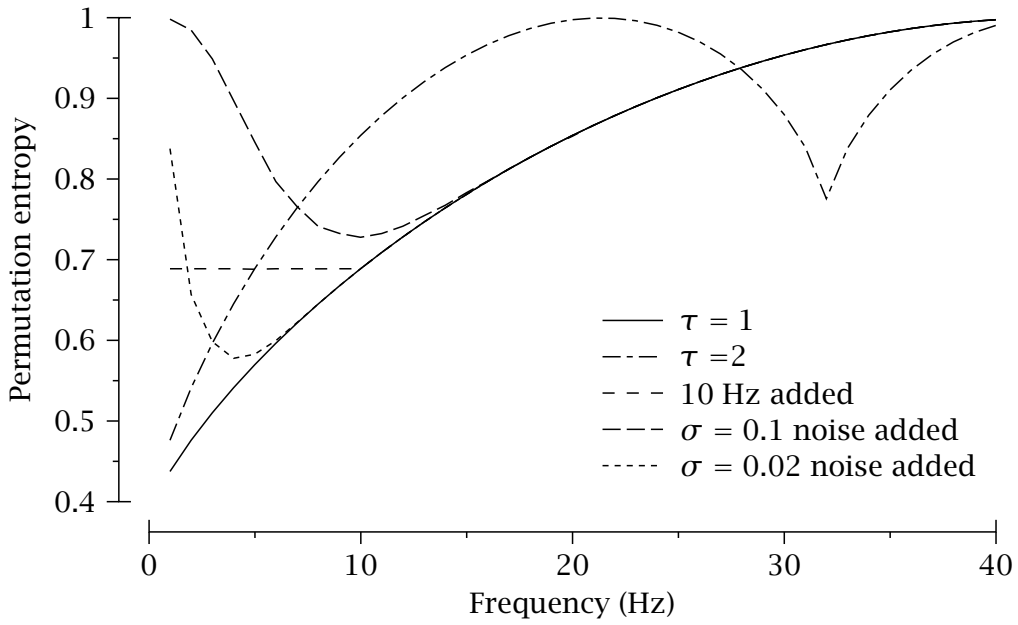

Figure 5.2: The frequency dependence of the permutation entropy with different lags ($\tau = 1$ and $\tau = 2$), and response to added white noise or a second sine wave oscillation. Long signals (65536 samples) were used to minimize fluctuations.

It is possible that two (or indeed all three) of the data points in a motif may have the same measured voltage because of the limited resolution of the analogue-digital conversion, and therefore are not able to be ranked. Although this occurrence may seem to be very unlikely, it raises the important issue of how to deal sensibly with very small voltage fluctuations in the EEG. A flat-line EEG indicates an almost absolutely quiescent cerebral cortex. If we have a high-resolution analogue-to-digital converter, the PE will continue to estimate the minute fluctuations in the signal - and its value typically increases. Unfortunately, these fluctuations are made up largely of various types of measurement noise, and are not reflecting drug actions on the cerebral cortex. This is most apparent when the burst suppression EEG pattern is present during deep anesthesia. One solution is to arbitrarily assign a third parameter - the threshold level, below which, most of the "signal" is thought to consist of noise. In our earlier paper, we chose the value of 0.5 μV as a threshold level. If the difference between any two of the points in the motif is below the threshold, then the motif was counted as a seventh category of "motif" - a "tie". However, for simplicity with regard to computations and analysis, in the remainder of this chapter the standard PE (with $m = 3$, $\tau = 1$) was used.

5.1.3 Application to Real EEG signals

Fourteen raw EEG data sets (from 7 male and 7 female patients; an extension of the data described in an earlier publication[62]) without burst suppression were analyzed using the population analysis approach. Step-wise changes in end-tidal isoflurane concentration were applied. The pharmacokinetic part of the PK-PD describes the relationship between the end-tidal and brain or effect-site concentration. The pharmacodynamic part of the model describes the relationship between the brain concentration and the measured

effect. It was assumed that this relationship is nonlinear, monotonically decreasing and can be described by the inhibitory sigmoid-$E_{\max}$ model with steepness parameter γ.

Previously,[62] we performed a two-stage analysis, *i.e.*, the data were fitted for each subject separately and next the obtained parameter values were averaged. In the present chapter, first both a two-stage and population analysis were performed to assess the importance of differences between the methods. Next, a population analysis of permutation entropy data based on the same EEGs was performed to assess the performance of permutation entropy *versus* BIS. An extended Kalman filter was constructed to analyze permutation entropy data. This was motivated by the fact that many data fits show systematic misfits; this may lead to biased standard errors of parameter estimates and false conclusions from statistical tests. Finally, simulations were performed to study the effects of the incorporation of a Kalman filter by fitting models with and without a Kalman filter to simulated data.

The Kalman filter is a method to track the state of a system in the presence of measurement and system noise. The extended Kalman filter is a generalization for linearized nonlinear systems. Tornøe *et al.* described[90] its implementation in the statistical analysis package NONMEM.[9] The Kalman filter state and functions of the state need to be defined to describe how the state evolves in time, how it is affected by an input, and how it propagates to the output. Two versions were constructed. Version A assumes that colored noise is present on the model output, albeit limited by maximum and minimum values of PE. Version B assumes that noise enters the system at the input, so at the level of anesthetic concentration. Because the process noise term is assumed to be normally distributed, it was added to the logarithm of concentration. NONMEM version 7 was used for parameter estimation. The differential equations describing the effect-site concentration, the Kalman filter state, and its variance were implemented as one (recursive) difference equation. The Kalman filter gain could be solved from the estimates of the measurement and process noise variances and hence needed no recursive approximation. The initial filter state was a parameter to be estimated, in addition to the parameters of the PK-PD model. The model incorporating Kalman filter version B was used to generate 1000 sets of artificial data of 10, 20, 50 and 100 individuals, with parameter values from the final model of the permutation entropy data. These data were fitted by the same model to check that parameter estimation was consistent, and by the model without Kalman filter.

5.1.4 PK-PD Modeling

The anesthetic concentration-effect data were analyzed with the pharmacodynamic model as described earlier.[62] Briefly, the pharmacodynamic model consisted of a hypothetical effect compartment combined with a sigmoid-$E_{\max}$ model described by the equations

$$\frac{dC_e(t)}{dt} = k_{e0} \cdot (F_{ET}(t) - C_e(t)) \tag{5.2}$$

and

$$E(t) = E_{\max} + (E_{\min} - E_{\max}) \cdot \frac{C_e(t)}{C_e(t) + \mathrm{IC}_{50}^{\gamma}}, \tag{5.3}$$

where $F_{ET}(t)$ is end-tidal isoflurane concentration; k_{e0} the rate constant determining the speed of equilibrium (we estimated the effect-site equilibration half-time $t_{½,k_{e0}}$), $C_e(t)$

is the effect-site concentration, $E(t)$ is the effect measure (permutation entropy or BIS), $E_{\max}$ and $E_{\min}$ are the maximal and minimal values of E, IC_{50} is the concentration that places E halfway between $E_{\min}$ and $E_{\max}$, and γ is a steepness parameter. In the mixed-effects analysis, all parameters were assumed to be lognormally distributed, and that residual error was normally distributed. Parameters $E_{\max}$ and $E_{\min}$ were constrained to the interval (0,1) and γ to the interval (0,25) *via* the inverse logit transform.

5.1.5 Two Extended Kalman Filters

Kalman filter A assumes that colored noise is present on the model output, albeit limited by $E_{\max}$ and $E_{\min}$. With

$$\gamma(t) = \log\left(\frac{1 - E(t)}{E(t)}\right),$$

where $E(t)$ is given by equation (5.3), and

$$\frac{dx}{dt} = g(x) + \nu = -\frac{x(t)}{\tau} + \nu,$$

where x is the state of the system and ν is system noise (a Wiener process[67]), the output is given by

$$E'(t) = \frac{1}{1 + \exp(x + \gamma)} + \epsilon$$

So $E'(t)$ resembles $E(t)$, but contains colored noise, filtered by time constant τ, and is, without the noise term, still constrained to the interval (0,1). Note that while we retain the notation with τ, this parameter is different from the lag used to calculate the permutation entropy.

Kalman filter B assumes that process noise enters the system at the input, so at the level of anesthetic concentration. Because the noise term ν has to be normally distributed, it was added to the logarithm of concentration, and then the system state is

$$\frac{dx}{dt} = g(x) + \nu = \frac{1}{\tau}\left(\gamma \cdot \log\left(\frac{C_{ET}(t) + C_{ET,0}}{IC_{50}}\right) - x\right) + \nu \tag{5.4}$$

and the output is given by

$$E'(t) = E_{\min} + \frac{1 - E_{\min}}{1 + \exp(x)} + \epsilon, \tag{5.5}$$

where $C_{ET,0}$ a parameter that yields finite values of the logarithm and replaces $E_{\max}$. The time delay is determined by several factors.[62] By placing the system noise at the level of concentration, τ may be viewed as neural processing rather than blood–brain concentration equilibration delay ($t_{½,k_{e0}}$).

For both versions, the derivative of $g(x)$ with respect to x is $-1/\tau$. This implies that the differential equation for the variance P of x is

$$\frac{dP}{dt} = -\frac{2P}{\tau} + \sigma_w^2.$$

Both x and P were solved for discrete time steps Δt under the assumption that the

input changes step-wise. For example, the next P_{i+1} is given by

$$P_{i+1} = P_i \cdot \exp\left(-\frac{2\Delta t}{\tau}\right) + \frac{\tau\sigma_w^2}{2}\left(1 - \exp(-\frac{2\Delta t}{\tau})\right).$$

In this way, the differential equations described here are transformed to difference equations, which can be solved by NONMEM, without the need for special compartment reset records in the input file as described by Tornøe *et al.*[90] The Kalman gain of the extended Kalman filter depends on the derivative of $E'(t)$ with respect to x, and was computed as given by Tornøe *et al.*[90]

5.1.6 Simulation Study

The model incorporating Kalman filter version B was used to generate 1000 sets of artificial data of 100 individuals, with parameter values from the final model of the permutation entropy data. The individual data consisted of 241 data points (time between 0 and 60 min, Δt 0.25 min). These data were fitted by the same model to check that the parameter estimation is consistent, and by the model without Kalman filter to study its effects on parameter estimates.

5.2 Results

5.2.1 Application to Real EEG signals

Two-Stage *versus* Population Analysis

Table 5.1 presents the results from the two-stage and population analyses of the BIS data. The population analysis provides estimates of the parameters, and of their variability across the population. The parameter estimates are similar, and interestingly, the SEMs of the two-stage analysis and standard errors from the population analysis also agree. On the other hand, the squares of the standard deviations from the two-stage analysis do not correspond to the inter-individual variance. These two results are due to the fact that interindividual variability is relatively small compared to parameter uncertainties in each individual (residual intraindividual variance was 50.4 ± 6.23), despite the number of observations per individual (≈ 200). Model fits are shown in the top panels of Figure 5.3. Worst, median and best fits were selected based on the coefficient of determination (R^2 ≈ 0.79, 0.92, and 0.98 for IDs 24, 12, and 18 respectively). The bottom panels show the applied (as measured) end-tidal isoflurane concentration.

Bispectral Index *versus* Permutation Entropy

Table 5.2 presents the results from population analysis of the permutation entropy data. Apart from obvious differences with the results for the BIS between $E_{\max}$ and $E_{\min}$ the much larger interindividual variability of γ is notable. This is related to the occurrence of some rather high values of γ - see the steep fall in PE in Figure 5.3 for ID 12 ($\gamma \approx 18$). This, in turn, may be related to the occurrence of high-frequency EEG activity, falsely (or not) increasing the measure of depth of anesthesia.

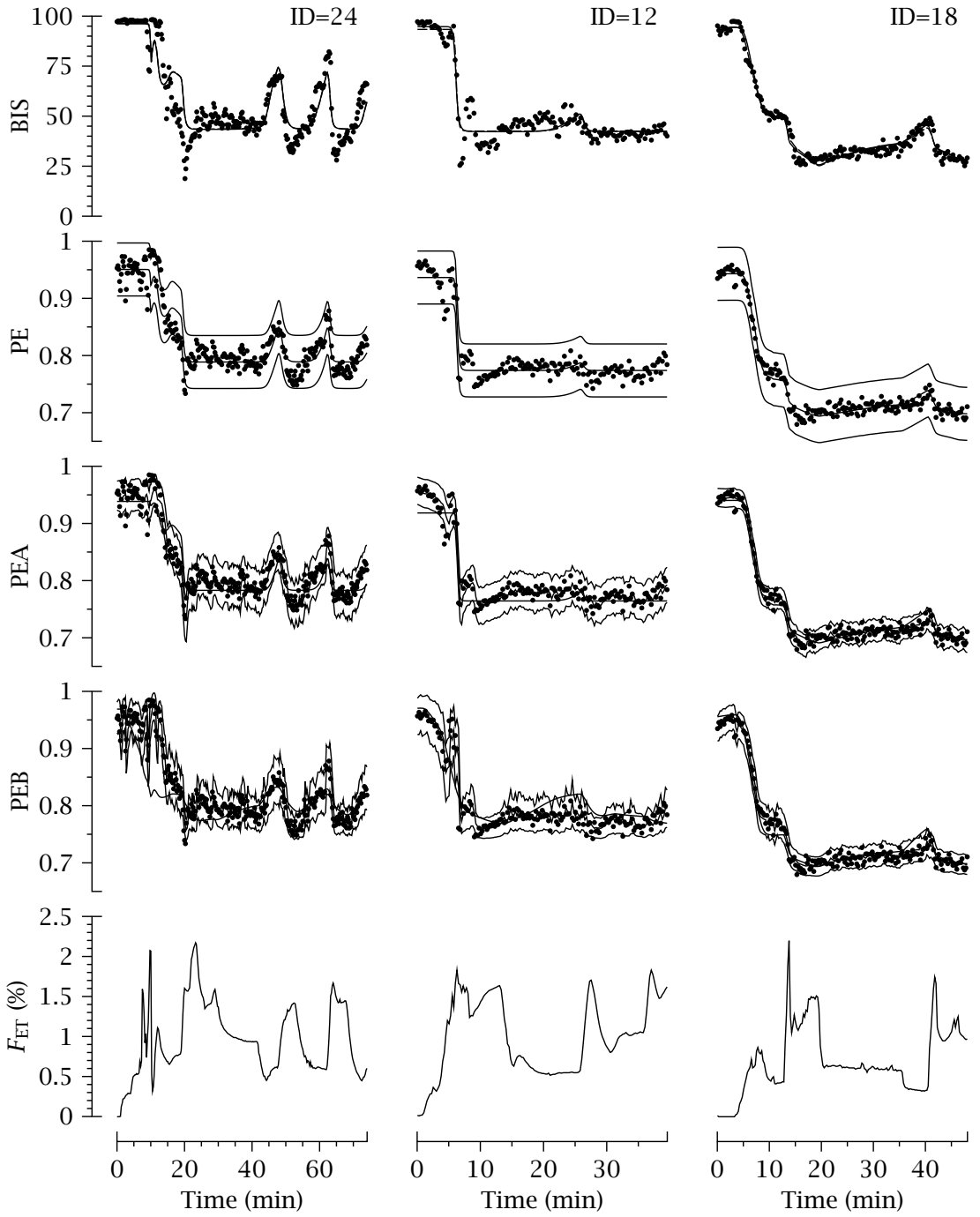

Figure 5.3: Administered isoflurane concentration (F_{ET}, bottom panels), BIS and permutation entropy data with fits (median ± 95% CI) without Kalman filter (PE), Kalman filter A (PEA), and Kalman filter B (PEB), for three subjects.

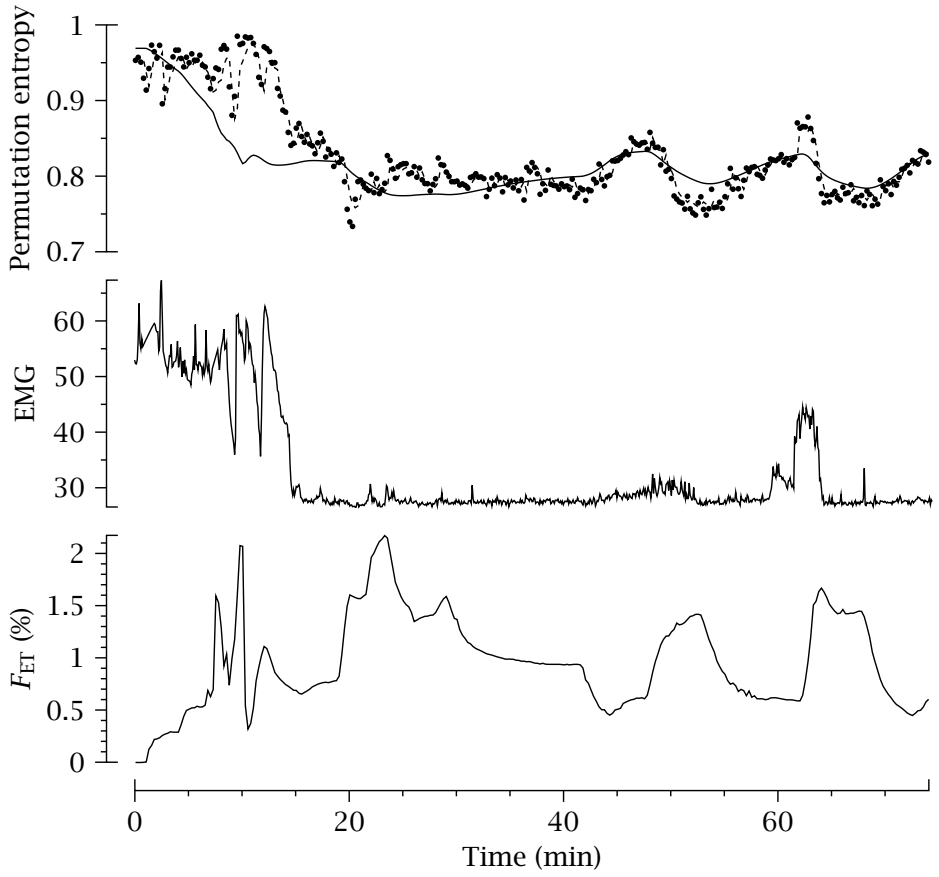

Figure 5.4: **Upper panel**: Close-up of the permutation entropy data fit of Subject 24. The dashed line is the Kalman-filtered permutation entropy; the solid line is the model prediction if there were no process noise. **Middle panel**: Electromyographic activity (EMG). **Lower panel**: Administered isoflurane concentration (F_{ET}).

Table 5.1: Parameter estimates from the Two-Stage and Population Analyses of the BIS data.

	Two-Stage Analysis			Population Analysis			
Parameter	Est.	SD	SEM	Est.	SE	ω^2	SE
$t_{½,k_{e0}}$ (min)	4.24	1.20	0.321	4.06	0.309	0.0744	0.0175
γ	5.94	4.14	1.11	5.05	0.831	0.508	0.255
IC_{50} (%)	0.575	0.107	0.0286	0.558	0.0302	0.0412	0.0182
E_{max}	95.9	1.86	0.497	95.8	0.508	0.0945	0.0443
E_{min}	27.2	10.8	2.87	27.4	2.55	0.199	0.0907

SEM is SD divided by the square root of the number of subjects.

Table 5.2: Population Analysis of the PE Data (without Kalman filter).

Parameter	Est.	SE	ω^2	SE
$t_{½,k_{e0}}$	4.37	0.350	0.0803	0.0235
γ	6.79	2.52	2.27	1.63
IC_{50}	0.506	0.0408	0.0633	0.0336
E_{max}	0.947	0.00296	0.0386	0.0145
E_{min}	0.727	0.00827	0.0238	0.00991

See Methods section for explanation of parameters.

Table 5.3: Population Analysis of the PE Data with Kalman Filter A.

Parameter	Est.	SE	ω^2	SE
$t_{½,k_{e0}}$	4.28	0.450	0.129	0.0328
γ	17.2	5.18	9.49	4.45
IC_{50}	0.501	0.0393	0.0785	0.0415
E_{max}	0.937	0.00617	0.0636	0.0894
E_{min}	0.729	0.0111	0.0314	0.0121
τ	5.41	1.13	-	-
σ_m	0.0113	0.000666	0.0326	0.0137
σ_s	0.102	0.0123	0.221	0.136
x_0	-0.344	0.208	0.260	0.287

σ_m and σ_s are the standard deviations of the measurement and process noise sources, respectively.
x_0 is the initial state of the system.
A dash indicates that the ω^2 was not estimable.

Kalman-Filtered Permutation Entropy

Table 5.3 and Figure 5.3 (PEA panels) present the results from population analysis of the permutation entropy data with Kalman filter A. Unfortunately, the estimate of γ and its interindividual variance increased (for ID 12 to $\approx$ 24, while it was limited to 25). It is unfortunate, because it means that the PE as a candidate measure of depth of anesthesia reduces to an on/off indicator.

Table 5.4 presents the results from population analysis of the permutation entropy data with Kalman filter B. Now, the estimate of γ and its interindividual variance decreased to a useful value. In Figure 5.3 it can be seen that PEB is still sensitive to higher isoflurane concentrations instead of being saturated. Moreover, note that all interindividual variabilities are smaller than those obtained from the analysis without Kalman filter (although we do not know the right values - see next subsection using simulated data). Figure 5.4 shows a close-up of the analysis for ID 24. High-frequency EEG activity and/or EMG activity elevates the PE. The Kalman-filtered PE seems to filter out this activity. See also the EMG activity at about 60-65 minutes.

A bootstrap analysis has been proposed to validate the parameter estimates.[30] Furthermore, the method can be used to obtain 95% confidence intervals. Results from the

Table 5.4: Population Analysis of the PE Data with Kalman Filter B.

Parameter	Est.	SE	ω^2	SE
$t_{½,k_{e0}}$	3.50	0.248	-	-
γ	2.60	0.677	0.0355	0.0204
IC_{50}	0.635	0.191	0.0210	0.0114
$C_{ET,0}$	0.280	0.163	-	-
E_{min}	0.699	0.00832	0.0165	0.00726
σ_m	0.00837	0.000286	-	-
σ_s	0.516	0.0398	0.0766	0.0384
x_0	0.499	0.320	-	-

$C_{ET,0}$ is a parameter related to E_{max} (see Methods).

Table 5.5: Bootstrap Analysis of the PE Data with Kalman Filter B.

Parameter	Est.	SE	95% CI	ω^2	SE	95% CI
$t_{½,k_{e0}}$	3.50	0.241	3.07-3.97			
γ	2.77	0.823	1.73-4.79	0.0333	0.0252	0.00343-0.0909
IC_{50}	0.679	0.225	0.392-1.23	0.0179	0.0129	0.00181-0.0430
$C_{ET,0}$	0.322	0.198	0.104-0.831			
E_{min}	0.700	0.00787	0.684-0.716	0.0150	0.00681	0.00223-0.0281
σ_m	0.000839	0.000285	0.00787-0.00895			
σ_s	0.516	0.0436	0.435-0.600	0.0724	0.0384	0.0157-0.157
x_0	0.554	0.289	0.0767-1.21			

bootstrap are given in Table 5.5. A comparison with the results in Table 5.4 shows that all estimates, including the standard errors, are similar. Unfortunately, the estimate of IC_{50} is rather imprecise; without the Kalman filter it was much more precise. Of course, at this stage it is unknown which precision estimate is the most useful; the more precise IC_{50} may be related to the description of the presence or absence of EMG activity rather than EEG activity.

5.2.2 Simulation Study

One thousand simulated data were generated using the model with Kalman filter B and parameter values listed in Table 5.4. These data were fitted with the same model, without and with Kalman filter, the results of which are presented in Table 5.6 and Table 5.7, respectively. The standard deviations from the 1000 estimates were calculated and the medians of the 1000 estimates of the standard errors (actually only NONMEM's succesful estimation and covariance runs were used ($\approx$ 930)). The medians were used because the estimates of the standard errors in the analyses with Kalman filter sometimes had quite large values; and also without the Kalman filter these are approximately lognormally distributed.

Without the Kalman filter, the median(SE) were somewhat less than the SDs (the

Table 5.6: Population Analysis of simulated data without Kalman filter.

Parameter	Est.	SD	median(SE)	ω^2	SD	median(SE)
$t_{½,k_{e0}}$	3.71	0.234	0.215	0.288	0.0675	0.0554
γ	2.67	0.222	0.211	0.625	0.149	0.133
IC_{50}	0.531	0.0186	0.0173	0.0821	0.0207	0.0184
E_{max}	0.961	0.00218	0.00213	0.296	0.0438	0.0430
E_{min}	0.697	0.00632	0.00600	0.0562	0.0136	0.0118

Table 5.7: Population Analysis of simulated data with Kalman filter.

Parameter	Est.	SD	median(SE)	ω^2	SD	median(SE)
$t_{½,k_{e0}}$	3.46	0.0914	0.0927	-	-	-
γ	2.49	0.181	0.184	0.0328	0.00663	0.00642
IC_{50}	0.622	0.0514	0.0506	0.0218	0.00502	0.00499
$C_{ET,0}$	0.270	0.0392	0.0395	-	-	-
E_{min}	0.699	0.00292	0.00282	0.0164	0.00247	0.00239
σ_m	0.00853	0.0000743	0.0000759	-	-	-
σ_s	0.480	0.0140	0.0145	0.0730	0.0124	0.0122
x_0	0.442	0.0650	0.0638	-	-	-

means were also somewhat smaller (data not shown)). This may be related to the fact that residuals are correlated if the Kalman filter is not applied. More importantly, the estimates of interindividual variability are too large, for example for $t_{½,k_{e0}}$, where it was actually zero, or for γ where the estimates differ a factor of 20. The IC_{50} was set at 0.635; the estimate of 0.531 ± 0.0186 is clearly biased.

5.3 Discussion

Ordinal measures of EEG patterns have very different characteristics to traditional methods that use the raw EEG signal. The conceptual difference may be summed up in the phrases: not “how large is the pattern?” but rather “how many patterns exist?”.[65] Our preliminary investigations[65] indicated that the permutation entropy shows promise as a practical EEG measure of GABAergic hypnotic drug effect. It appropriately tracks the qualitative assessment of the EEG pattern from awake to sedated/lightly anesthetized, and to deeply anesthetized. It requires minimal preprocessing and is very resistant to blink artifacts.

A wide variety of other EEG indices have been used in PK-PD modeling.[65] At the present time, the most commonly used are the commercial ones (*e.g.*, BIS). Most indices require extensive and sophisticated artifact handling to achieve acceptable results, and this always runs the risk of distorting the EEG signal. The resistance to blink artifacts (and speed of computation) is a big advantage of the PE over other indices, producing stable values in the pre- and early induction period. The PE is comparable with the BIS in the spread of values between patients before and after loss-of-consciousness. Our

present results indicate that the PK-PD models that may be constructed using the PE are as acceptable as those constructed using the BIS.

A characteristic feature of the state of general anesthesia is the presence of large slow fluctuations in the instantaneous frequency content of the EEG. The most extreme example of this phenomenon is the burst suppression pattern, which needs special precautions,[65] so for the present analyses only data without burst suppression were used.

As for all EEG monitors, the PE is effectively computed from an EEG signal which is inextricably linked with the frontalis EMG. The same reason that makes the PE stable during the awake state also makes it very sensitive to episodic high-frequency fluctuations (artifactual, electromyographical, and neurophysiological) during deep anesthesia (see Figures 5.2 and 5.4 for illustration).

So while the PE is insensitive to eye blinks, it is sensitive to high frequency components present in the EEG just before loss of consciousness. This causes EEG effect parameters to respond relatively late and suddenly to a change from wakefulness into anesthesia. This property also results in a steep concentration-effect relationship. Analysis of EEG data with a Kalman filter accentuated or filtered out this phenomenon, depending on the postulated location of process noise. Version B provided a more useful pharmacodynamic model, because version A with the steep concentration-effect relationship impedes timely prediction of (sudden) awakening. Furthermore, the large steepness may, at least partly, be based on a confounding effect or artifact. It should be noted that this artifact filtering is only possible if the anesthetic concentration is known. The main difference between the two applied Kalman filters lies in the location of the process noise, *i.e.*, before or after the sigmoid-$E_{\max}$ concentration-effect relationship. Process noise may substitute for model inadequacies, in this case an EEG effect which was (in the model) not related to anesthetic concentration. So the confounding effect may need to be scaled, due to nonlinearities, to the level of the real process noise.

Finally and most importantly, the simulation study showed that if Kalman filtering is not applied, inter-individual variability may be overestimated; variability that is actually intra-individual process noise.

Chapter 6

Arterial and Venous Pharmacokinetics of Morphine-6-Glucuronide and Impact of Sampling Site on Pharmacodynamic Parameter Estimates*

PHARMACOKINETIC-PHARMACODYNAMIC (PK-PD) MODELING is an important tool to examine the dynamic behavior of drugs, and because the analysis yields an estimation of drug potency and delay between blood concentration and effect (hysteresis), it allows for an accurate prediction of effect. The hysteresis occurs because of the distributional disequilibrium between the site at which the drug is measured and the site of action (= biophase kinetics).[92] Most contemporary PK-PD models are correctly based on arterial blood samples.[92] However, occasionally, arterial blood samples are not available. This may, for example, occur when it is deemed inappropriate to place arterial catheters and it is assumed that similar results will be obtained by using venous blood samples.

The human ethics committee of our institution expressed its concerns regarding the placement of arterial catheters in healthy volunteers participating in PK-PD studies on long-acting opioids. They reasoned that for some drugs, just a small difference in PD parameter estimates would be obtained when sampling from a venous site and consequently that it was judged unnecessary (and hence unethical given the possibility of serious complications) to place an arterial catheter. We have ample experience with arterial catheter placement and over the years did not encounter any complications. We do agree with them, however, that the consequences of complications such as radial artery occlusion, nerve damage, or pseudo aneurysm of the radial artery are serious and need to be carefully balanced against the gain of obtaining arterial blood samples. In cooperation with the human ethics committee we therefore decided to perform a study in volunteers in which both arterial and venous (from a peripheral site, *i.e.*, the arm) drug

* E Olofsen, R Mooren, E van Dorp, L Aarts, T Smith, J den Hartigh, A Dahan, *Anesth Analg* 2010; 111:626–632

samples were obtained after IV infusion of the opioid morphine-6-glucuronide (M6G), a drug of long action.[69] The dose chosen, 0.3 mg/kg, causes long-lasting analgesia but only moderate respiratory depression.[69,71,68]

The present study was performed in 3 steps. Initially, we determined the arterial and venous concentrations of M6G after a rapid IV infusion. Next, we built a PK model of the drug distribution between arterial and venous blood. Finally, we performed simulation studies to help us to understand the consequences of the (erroneous) assumption that arterial and venous data are equal and the effect on PD parameter estimates when using venous rather than arterial drug concentrations to drive the effect compartment. We hypothesized that because we were dealing with a slow-acting drug, there is no difference between an arterial sample-based model and a venous sample-based model.

6.1 Methods

6.1.1 Subjects

Seventeen healthy volunteers (9 men, 8 women, ages 19 to 34 years and body mass index <28) participated in the study after approval of the protocol by the local ethics committee and after giving written informed consent. All subjects were asked to refrain from food for at least 8 hours before the start of the study.

6.1.2 Study Design

After arrival in the laboratory, 2 venous catheters (in the left and right cubital veins) and 1 arterial catheter (in the radial artery at the wrist of the nondominant arm) were inserted. One venous catheter was placed for drug infusion; the other venous catheter and the arterial catheter were placed for blood sampling. At t = 0, 0.3 mg/kg M6G was infused IV over 90 seconds. Next, arterial and venous samples were obtained (simultaneously) at times t = 5, 10, 20, 30, 40, 50, 60, 80, 120, 180, 240, 300, 360, and 420 minutes. Plasma was separated within 10 minutes of blood collection and stored at -25°C until analysis. M6G measurement has been described.[71] Briefly, serum was pretreated by protein precipitation with acetonitril; M6G was measured using liquid chromatography with tandem mass spectrometry. The between-days coefficients of variation were 4.1% and 4.0% for 75 and 1800 µg/L, respectively; the within-day coefficients of variation were 0.5% and 2.0%. The quantitation limit was set at 20 ng/mL.

6.1.3 Pharmacokinetic Analysis

The arterial and venous concentration data were analyzed simultaneously. To that end, the arterial data were analyzed first with 2 or 3 compartments. Next, arterial and venous data were analyzed simultaneously with 3 compartments for arterial data and 1 or 2 compartments for venous data. Objective function values were reported for the latter 2 cases, to give an indication for the importance of the second venous compartment. The venous samples are from the periphery, *i.e.*, the forearm. Data analysis was performed with the statistical package NONMEM VI, version 1.2. Model selection (the number of compartments) was based on the goodness-of-fit criterion, *i.e.*, the magnitude of the

decrease in minimum objective function value (MOFV; X^2 test (for nested models): P < 0.01 was considered significant).

For the concentrations in the central and peripheral venous compartments (C_{V_1} and C_{V_2}, respectively) we write (Figure 6.1A)

$$V_{V_1} \cdot dC_{V_1}/dt = CL_{AV} \cdot (C_{A_1} - C_{V_1}) - CL_{V_2} \cdot (C_{V_1} - C_{V_2}) \quad (6.1)$$

$$V_{V_2} \cdot dC_{V_2}/dt = CL_{V_2} \cdot (C_{V_1} - C_{V_2}), \quad (6.2)$$

where V_{V_1} and V_{V_1} are the volumes of the central and peripheral venous compartments, CL_{AV} the arteriovenous (AV) clearance, and CL_{V_2} the central-peripheral venous clearance (see Appendix and Figure 6.1A for details).

6.1.4 Simulation Studies

Simulation studies were conducted to assess the influence of sampling site on estimated PD parameter values and on prediction. The PD model consisted of an inhibitory sigmoid E_{max} model, with a baseline effect of 1 and a minimum effect of 0 (*e.g.*, describing the respiratory effect of M6G):

$$\text{Effect}(t) = 1/\left(1 + (C_e(t)/EC_{50})^{\gamma}\right), \quad (6.3)$$

where where $C_e(t)$ is the effect-site concentration at time t, EC50 the effect-site concentration giving 50% effect, and γ a shape factor (the Hill coefficient). An effect site is postulated. The equilibration rate constant between arterial blood and effect site is k_{e0} with half-life $t_{½,k_{e0}}$. The effect site has to be linked to arterial blood concentrations to obtain correct PD parameter estimates (Figure 6.1). However, when only venous blood samples are available, the effect site may be (incorrectly) linked to venous blood concentrations (Figure 6.1). EC_{50} was set to 500 (a typical value for pain relief from M6G with $t_{½,k_{e0}}$ values ranging from 2 to 6 hours),[68] γ to 1 (as observed in studies on the effect of M6G on minute ventilation [where maximal respiratory depression = apnea] and with $t_{½,k_{e0}}$ values ranging from 1 to 2 hours)[69] or 2.5 (as observed in studies on the effect of M6G on pain relief)4 and $t_{½,k_{e0}}$, as given above. These 3 parameters had lognormal distributions across the population with variance 0.1 (coefficient of variation ≈30%)[69,68] The SE of the additive intraindividual error was set to 0.1.

Study I

The purpose of the first set of simulations was to determine the influence of linking venous blood samples to the effect site on PD parameters. One thousand Monte Carlo simulations were performed to generate PD data with arterial concentrations driving the effect site. These PD data were next fitted using arterial or venous concentration data driving the effect site. Using arterial concentration data should yield parameter estimates close to the ones used for simulation; using venous concentrations might yield estimates that are biased because of the erroneous location of the sampling site. For each simulated data set (n = 1000), the ratios of the parameters based on venous and arterial data were calculated. This was done for a range of $t_{½,k_{e0}}$ values: 5, 10, 15, 20, 30, 45, 60, 90, 120, 150, 180, and 240 minutes. In studies on respiration and analgesia a large variation in $t_{½,k_{e0}}$ values has been observed between 60 and 240 minutes.[69,68] From

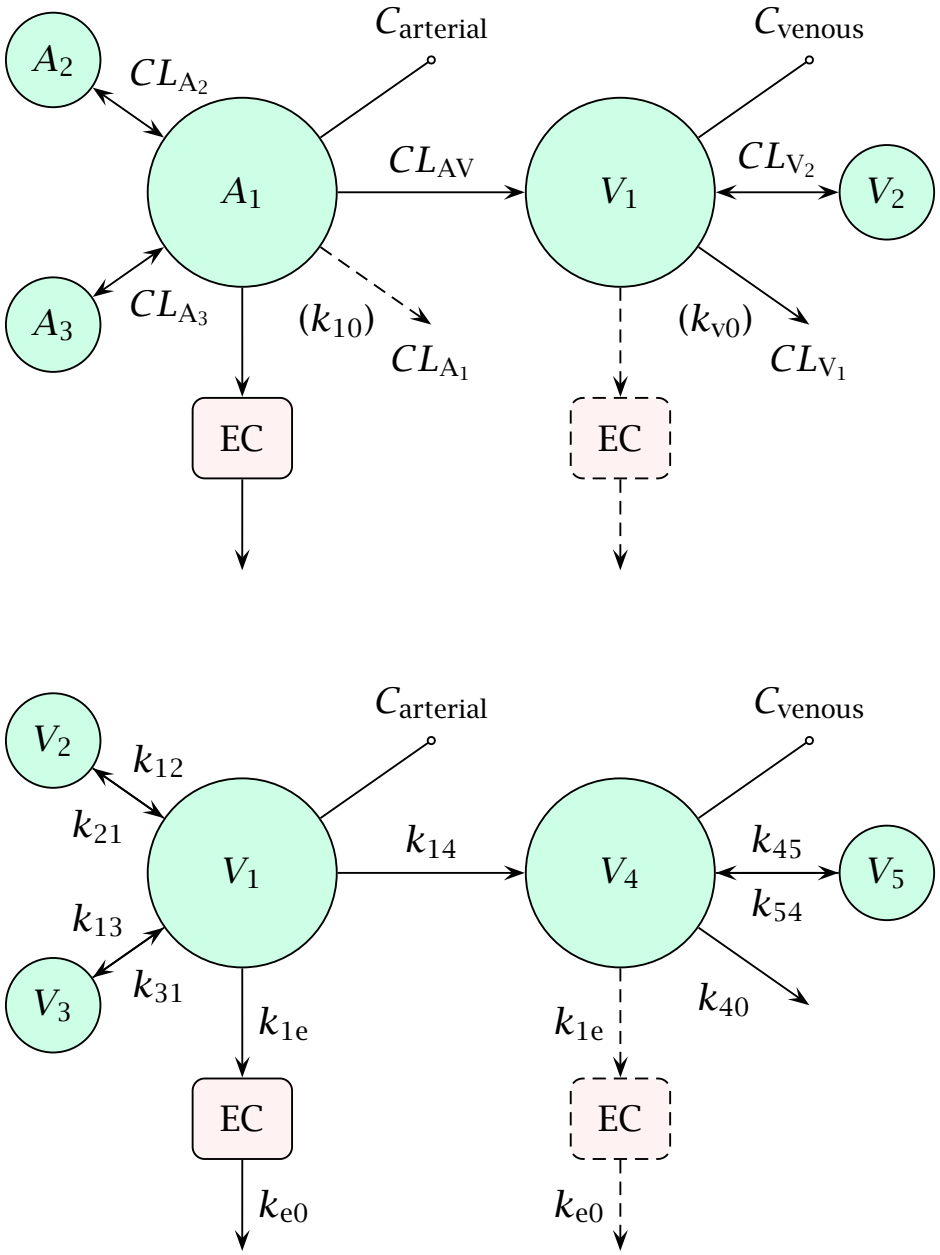

Figure 6.1: Two schematic representations of the pharmacokinetic (PK) model used to simultaneously analyze the arterial and venous morphine-6-glucuronide (M6G) data. **A**: Model in terms of volumes and clearances. A_1, A_2, and A_3 are arterial compartments, and V_1 and V_2 are venous compartments. CL_{A_1}, CL_{A_2}, CL_{A_3}, CL_{V_1}, and CL_{V_2}, are the clearances from compartments A_1, A_2, A_3, V_1, and V_2, respectively. An effect site has been postulated, driven by arterial M6G concentrations (EC linked to A_1) and an effect site driven by venous concentrations (EC linked to V_1; as if venous concentrations were equal to arterial ones). The latter representation is erroneous and tested in the current study. **B**: Model in terms of rate constants and NONMEM compartments. We assumed that k_{14} equals k_{10} (or CL_{A_1}/V_{A_1}). This has no influence on the model but indicated that V_4 and V_5 (and the clearance between them) have to be interpreted in relation to k_{14} (or CL_{AV}). The k_{e0} is the rate constant depicting the equilibration between blood and effect site. The k_{40} is the elimination rate from the central venous compartment and is equivalent to k_{v0}.

the 1000 simulation data sets, the median and 95% confidence intervals of the ratios of the parameters based on venous and arterial data were calculated. $t_{½,k_{e0}}$ values chosen below the values typically seen in respiratory and pain studies (1 hour) do reflect values that may occur when examining other end points, such as opioid-induced changes in electroencephalographic data.

Study II

To get an indication of the effect of using a biased PD parameter set, as may occur in clinical settings when PK and PD data sets derived from distinct arterial and venous concentration data sets are linked (*e.g.*, effect-controlled target-controlled infusion systems), we simulated the link between a PK model based on arterial data and a PD model based on venous data. Simulations were performed using the median estimates from Study I. To quantify the bias, we made a comparison to a simulation in which PK and PD models are both based on arterial data.

6.2 Results

All subjects completed the study without major side effects. Figure 6.2 shows the time course of the plasma concentrations of M6G for the samples obtained from arterial (radial artery) and venous (elbow) sites for 3 subjects. In Figure 6.3 the difference between arterial and venous M6G concentrations are plotted over time.

The data indicate that just after the 90-second infusion (with relatively high M6G concentrations), arterial concentrations exceeded the venous ones because the net drug flow is into the tissues of the arm. At later times (>40 minutes, with relatively low M6G concentrations), venous concentrations exceeded arterial concentrations because the net drug flow is from tissue to blood.

A schematic representation of the final "extended" PK model is given in Figure 6.1: for the arterial site (*i.e.*, body), 3 compartments were required (MOFV, 2 compartments = 2389.811 *versus* 3 compartments = 2167.616); for the venous site (*i.e.*, forearm), 2 compartments were required (MOFV 1 compartment = 4693.153 *versus* 2 compartments = 4418.097). Best, median, and worst data fits and goodness-of-fit plots are shown in Figure 6.2. Spaghetti plots (measured and predicted concentrations *versus* time) for the arterial and venous plasma concentration data are given in Figure 6.4. Without exception, the data were well described by the model. The model parameters are collected in Table 6.1. A significant equilibration delay was present between the central arterial and venous compartments ($t_{½,k_{v0}}$ = 2.00 ± 0.45 minutes, median ± SE). Figure 6.5 shows the results of simulation study I. For both values of γ, the results indicate that large biases are to be expected when using venous PK data as input to the PD model, albeit the magnitude of the bias depends on the "true" value for parameter $t_{½,k_{e0}}$ (for analgesia the true values range from 2 to 6 hours; for respiratory depression, from 1 to 2 hours). For parameter $t_{½,k_{e0}}$ the bias (as reflected by the ratio parameter derived from venous PK over arterial PK) is an underestimation of the "true" value ranging from a decrease of 60% at a "true" $t_{½,k_{e0}}$ value of 5 minutes to a decrease of 30% at a value of 240 minutes. For C50 the estimation from venous PK data yields a small bias (an overestimation ranging from +10% to -2%) in the $t_{½,k_{e0}}$ range of 5 to 90 minutes. At larger values of $t_{½,k_{e0}}$ the bias increases to an overestimation of 30% to 40% at $t_{½,k_{e0}}$ = 240 minutes. Parameter γ

Table 6.1: Pharmacokinetic Parameter Estimates

Parameter	Estimate	SE of estimate	ω^2	SE of ω^2
V_{A_1} (L)	4.45	0.69	0.07	0.04
V_{A_2} (L)	4.73	0.36	0.03	0.01
V_{A_3} (L)	5.07	0.27	0.02	0.01
CL_{A_1} (L/min)	0.14	0.006	0.03	0.01
CL_{A_2} (L/min)	0.55	0.10	-	
CL_{A_3} (L/min)	0.07	0.01	-	
$t_{½,k_{v0}}$ (min)	2.00	0.45	0.12	0.10
V_{V_1} (L)	0.05			
V_{V_2} (L)	1.88	0.38	0.68	0.25
CL_{V_2} (L/min)	0.06	0.01	0.27	0.18
σ^2 arterial	0.003	0.001		
σ^2 venous	0.02	0.006		

V_{A_1}, V_{A_2}, and V_{A_3} are the volumes of the arterial compartments A_1, A_2, and A_3, with intercompartmental clearances CL_{A_1}, CL_{A_2}, and CL_{A_3}, respectively.
$t_{½,k_{v0}}$ is the half-life of drug elimination from compartment V_1.
V_{V_1} is the volume of venous compartment V_1; V_{V_2} is the volume of venous compartment V_2 with intercompartmental clearance CL_{V_2}. V_{V_1} is derived from $V_{V_1} = CL_{A_1}/k_{v0}$,which implies that in the steady state, arterial and venous concentrations are equal.
The ω^2 are between-subjects variabilities (in the log-domain); the σ^2 are the residual errors.

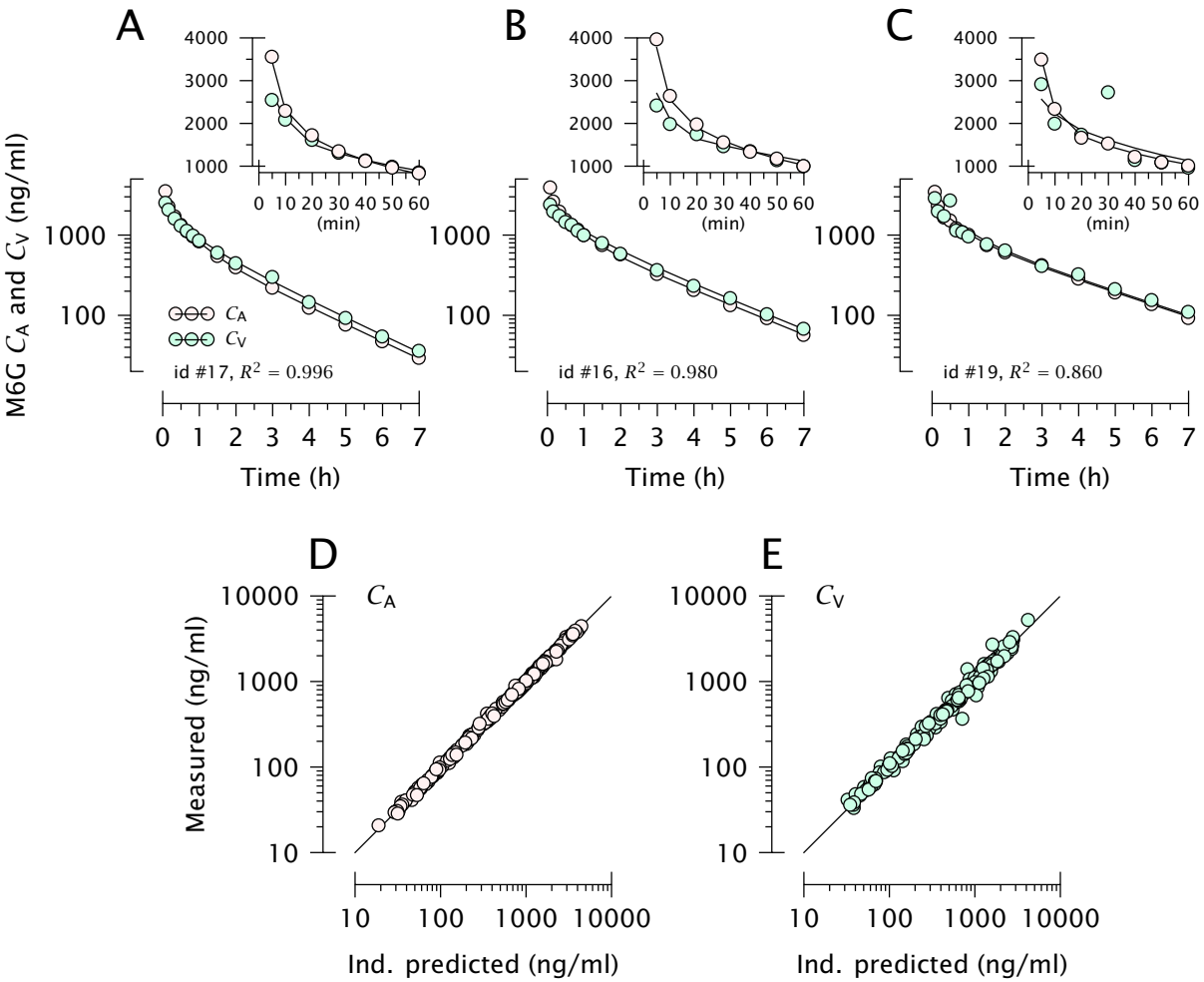

Figure 6.2: Best (**A**), median (**B**), and worst (**C**) pharmacokinetic (PK) data fits. Pink symbols, arterial samples; cyan symbols, venous samples. Solid lines, data fits. M6G, morphine-6-glucuronide. Inserts are the samples and data fits of the first hour of the experiment. **D** and **E**: Goodness-of-fit plots for the individual PK model for the arterial (**D**) and venous (**E**) concentrations. Shown are the observed (*y-axis*) *versus* individual predicted (*x-axis*) PK data.

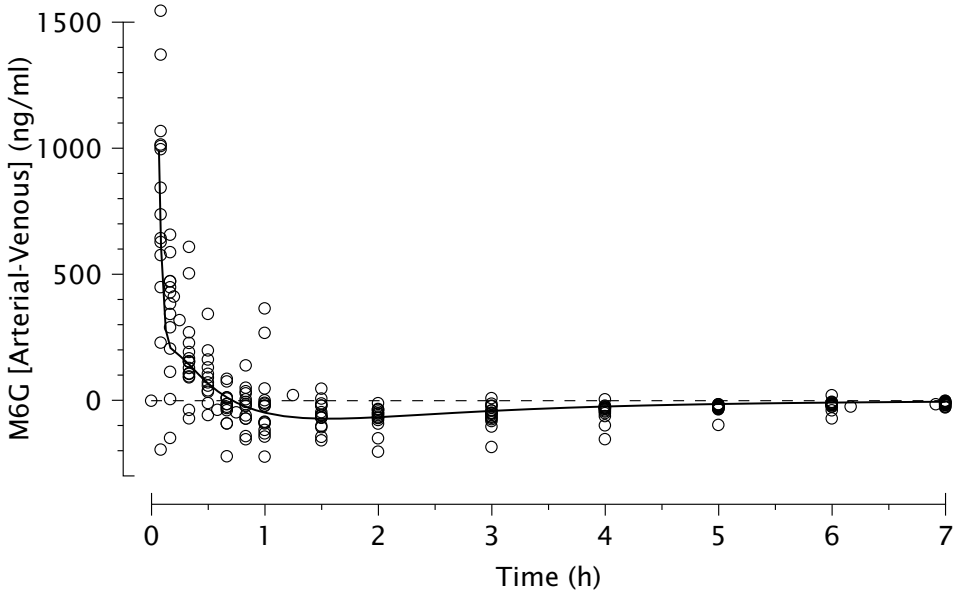

Figure 6.3: Difference between arterial and venous morphine-6-glucuronide (M6G) concentrations over time. The line through the data is the NONMEM population fit.

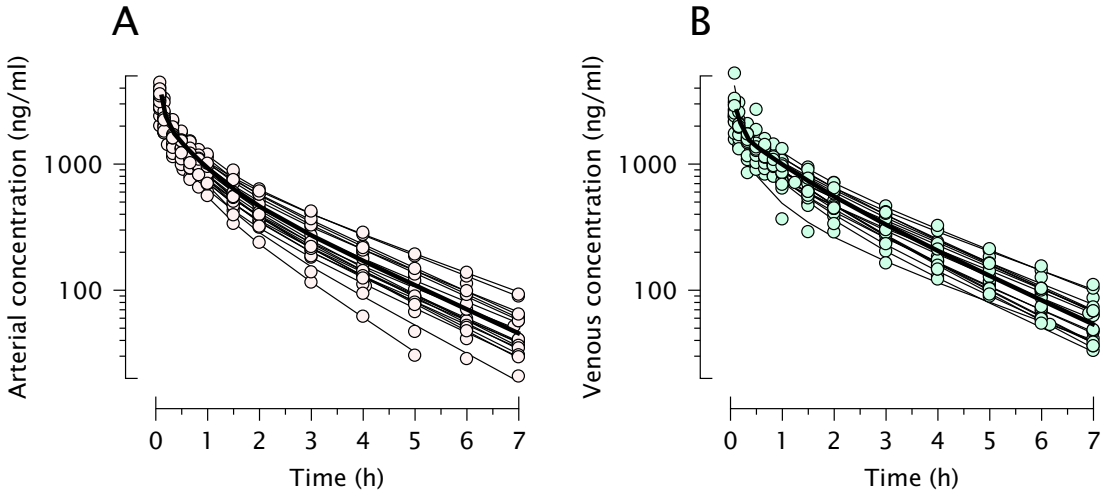

Figure 6.4: Spaghetti plots of measured and predicted morphine-6-glucuronide (M6G) concentrations in time. **A**: Measured (*solid circles*) and predicted (*solid black lines*) arterial concentrations. **B**: Measured (*solid circles*) and predicted (*solid black lines*) venous concentrations. The bold lines are the average model-based predicted concentrations.

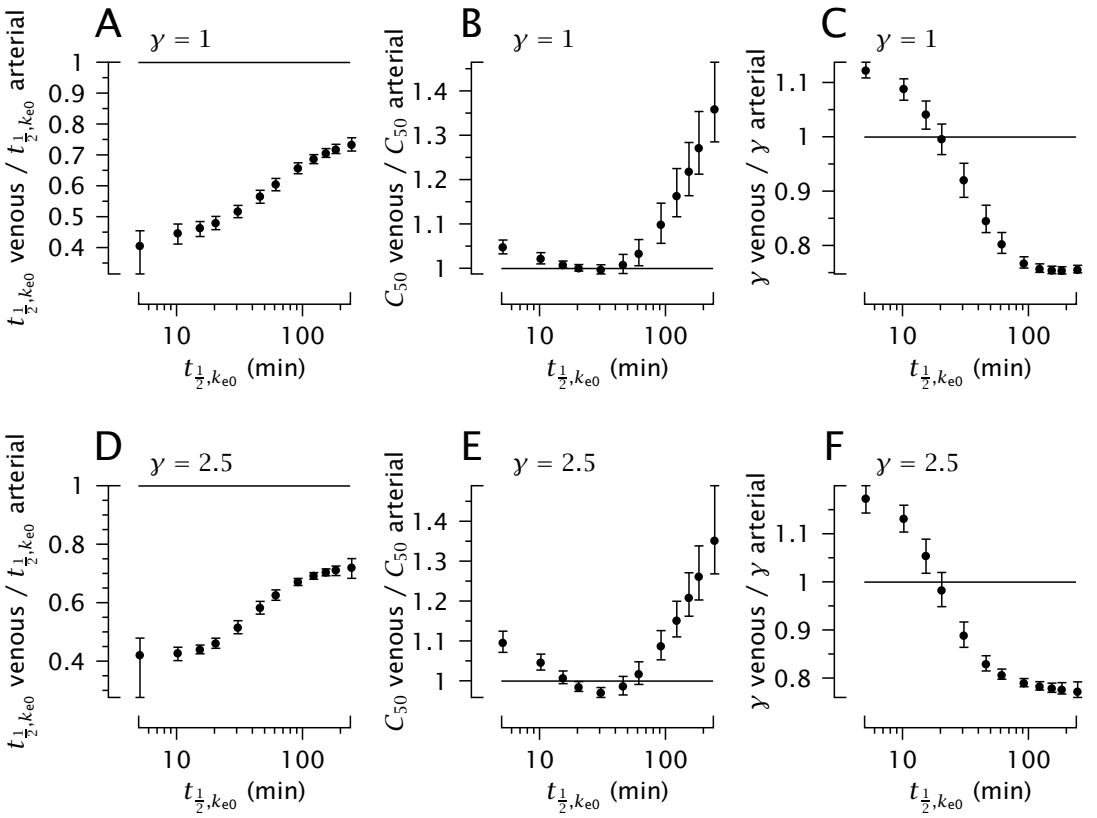

Figure 6.5: Results of simulation study I. **A** and **D**: Concentration site dependence on $t_{½,k_{e0}}$; **B** and **E**: Concentration site dependence on C_{50}; **C** and **F**: Concentration site dependence on γ. The ratio of the parameters obtained by using venous and arterial pharmacokinetic data are plotted against $t_{½,k_{e0}}$. A, B, and C are simulations with $\gamma = 1$; D, E, and F with $\gamma = 2.5$. $t_{½,k_{e0}}$ values range from 5 to 240 minutes. Values are median (of 1000 simulations) ± 95% confidence interval.

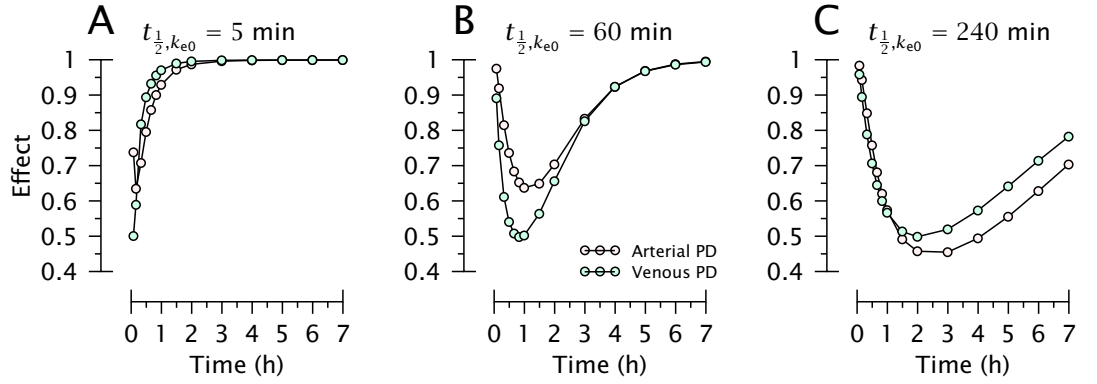

Figure 6.6: Results of simulation study II. Linking arterial pharmacokinetic (PK) data to pharmacodynamic (PD) models derived from venous (*cyan symbols*) and arterial (*pink symbols*) PK data. The data are simulated for 3 different values of $t_{½,k_{e0}}$: **A**: 5 minutes, **B**: 60 minutes, and **C**: 240 minutes. The effect simulated was a maximum of a 50% decrease in effect for the PD model derived from venous data.

is overestimated at $t_{½,k_{e0}}$ values <20 minutes by about 10%, whereas at larger values it is underestimated, reaching 25% at $t_{½,k_{e0}}$ = 240 minutes.

Figure 6.6 shows the results of simulation study II. The cyan symbols are the PD data that occur when an arterial PK set is linked to a PD parameter set derived from a venous PK data set for 3 values of $t_{½,k_{e0}}$ (5, 60, and 240 minutes). The pink symbols are the PD data derived from "arterial" PK and PD sets. The simulations were such that the maximum "venous" PD peak effect was 50% of control. As is obvious from the presented simulations, there are clear differences in effect between the arterial and venous PD models. For example, at low $t_{½,k_{e0}}$ values, venous peak effect exceeded arterial peak effect (for $t_{½,k_{e0}}$ = 60 minutes the venous peak effect is 30% greater at t = 60 minutes). However, at increasing values of $t_{½,k_{e0}}$ the differences decrease, and at $t_{½,k_{e0}}$ values between 150 and 180 minutes the arterial peak effect exceeds the venous peak effect (Figure 6.6C). For all values of $t_{½,k_{e0}}$, peak effect occurred somewhat earlier in the venous PD model.

6.3 Discussion

This study was designed to evaluate the effect of venous *versus* arterial blood sampling (derived from the cubital vein and radial artery, respectively) of the μ-opioid M6G on the bias of parameter estimates derived from PD models using simulated data of volunteers. We chose a combined arterial-venous model to model the AV differences in PK (which gives information of the kinetics at the site of sampling), rather than just a PK model that described the venous PK data in terms of an "arterial" model. The values of the arterial PK parameters were well in agreement with those presented in the literature.[68] Previous studies assessed the relevance of AV concentration differences on PK-PD modeling. For example, in the rat, Tuk *et al.*[93] measured arterial and venous concentrations of the benzodiazepine midazolam and linked them to effect (using electroencephalographic amplitude). Using a "traditional" effect-compartment model, differences in PD model parameters were apparent for EC_{50} (104 *versus* 86 ng/mL for arterial *versus* venous sampling) and k_{e0} (0.32 *versus* 313 min^{-1}). With an extended effect-compartment model (by characterizing the delay between arterial and venous sampling sites, comparable to our approach depicted in Figure 6.1), the model parameters did improve, although large differences did persist (EC_{50} = 89 ng/mL, k_{e0} = 2.5 min^{-1}).[93]

The AV concentration difference of a drug is, for a large part, determined by its interaction with the tissue at the venous sampling site (the venous concentration is not a reflection of the mixed venous/pulmonary artery concentrations). As was discussed by Gumbleton *et al.*,[37] mechanisms for the generation of AV concentration differences in the forearm arise from an elimination process (the drug is taken up by the muscles in the arm and metabolized in the muscle cells), by a distributional process (drug equilibration between plasma and tissue, which is dependent on the fraction of cardiac output going to the sample arm, perfusion and temperature-dependent capillary shunting, diffusion into and affinity for muscle tissue), or the combination of the two. M6G is a drug that is eliminated from the plasma exclusively through renal clearance. Because no tissue metabolic processes are known for M6G, the AV difference for this drug is determined solely by a distributional process in the forearm (equilibration between plasma and muscle tissue occurs at times >120 minutes, which is 4 times the half-life for the lower $t_{½,k_{e0}}$; see next paragraph). The PK-PD consequence of a large distributional AV concentration

difference is illustrated by Gumbleton *et al.*[37] showing that the determination of PD parameters will be highly biased when using a first-order effect-compartment model. In contrast, when the AV differences are related to an elimination process only (with instantaneous equilibration between plasma and tissue and a constant AV difference), little bias in parameter estimates is expected.[37]

Our results are in agreement with the theoretical studies of Tuk *et al.*[92] and Gumbleton *et al.*[37] However, our extended PK model differs significantly from their applied PK models. Most importantly, the venous part of our PK model has two compartments (V1 and V2), and theirs has only one. This makes comparison of our parameter k_{v0} not possible between studies. For the PK parameters of the venous compartments it can be calculated that clearance from the venous site is characterized by two half-lives, with values of 1.4 and 33 minutes, respectively. This suggests a secondary slow equilibration of M6G between arterial and venous blood and as such explains the large bias in the PD parameter estimates derived from venous blood samples. In our simulation study I we showed further that the bias in parameter estimates is critically dependent on the value of $t_{½,k_{e0}}$. When using venous M6G PK values as input to our first-order effect-compartment model and assuming that the "true" $t_{½,k_{e0}}$ value of M6G ranges from 60 to 240 minutes, as observed for M6G's effect on respiration and analgesia,[69,68] we would have underestimated $t_{½,k_{e0}}$ by 30% and γ by 25%, while the potency parameter (EC_{50}) would have been overestimated by about 40%. The estimations were also dependent on the value of the Hill coefficient γ (which may vary from 1 in respiratory studies to 2.5 in analgesia studies),[69,68] although this effect was less in magnitude (Figure 6.5).

In simulation study II, the effect of the use of biased PD parameters (*i.e.*, derived from venous blood samples) linked to an arterial PK set was explored. A situation in which the two are linked may occur in, for example, target-controlled infusion systems that incorporate PD model parameters to steer effect rather than target plasma concentration. The bias in predicted effect was dependent on the value of $t_{½,k_{e0}}$ with respect to the magnitude of effect and the timing of peak effect. The bias was such that a useful application of the model is not warranted in a clinical setting. Note, however, that if the PD model is derived from a venous PK set and linked to venous PK data, the bias would be minimal (but only when using an infusion scheme identical to that used in establishing the venous model), although the parameter estimates are biased in comparison with an arterial PK set and PD model.[92] Evidently, such an approach makes reliable comparisons with studies using arterial PK and PD models impossible and also is in violation of the principle that the effect site is directly linked to arterial rather than venous blood concentrations.

In conclusion, there are significant AV differences in M6G plasma concentration, related to a distributional process in the forearm. Biases exceeding 10% - 20% in PD model parameters will occur when linking venous concentration to effect, using the traditional effect-compartment model.

6.A Appendix: Linking Venous Compartments

It is assumed that the arterial morphine-6-glucuronide (M6G) concentrations are completely described by a 2- or 3-compartment model. The venous compartment then needs to be linked to the arterial pharmacokinetic (PK) model without affecting the latter, similar to linking an effect compartment.

For the amount of drug in the central (A_{V4}) and peripheral (A_{V5}) venous compartments we write (Figure 6.1B)

$$dA_{V4}/dt = k_{14} \cdot A_{V1} - k_{40} \cdot A_{V4} + k_{54} \cdot A_{V5} - k_{45} \cdot A_{V4} \quad (6.4)$$

$$dA_{V5}/dt = k_{45} \cdot A_{V4} - k_{54} \cdot A_{V5}, \quad (6.5)$$

where k_{14}, k_{54}, and k_{45} are rate constants between compartments V_1 and V_4, V_5 and V_4, and V_4 and V_5, respectively; k_{40} is the elimination rate constant from compartment V_4. We set k_{14} equal to k_{10} (= CL_{A_1}/V_{A_1}; Figure 6.1A) so that V_4 is not trivial in size but rather

$$V_4 = V_1 \cdot k_{14}/k_{40}. \quad (6.6)$$

This is an alternative yet exact method to keep the arterial PK part unaffected. We devised this method to avoid potential numerical problems in NONMEM with very small volume parameters. So k_{40} (Figure 6.1B) equals k_{v0} (= CL_{V_1}/V_{V_1}; Figure 6.1A); and $t_{½,k_{v0}}$ (= $\log(2)/k_{v0}$) is a parameter to be estimated.

It must be noted that, given the above, parameters V_{V_1}, V_{V_2}, and CL_{V_2} in Table 6.1 and Figure 6.1A are to be interpreted in relation to CL_{AV} (= CL_{A_1}), which will differ from the true arterial-venous clearance.

Chapter 7

Summary and Conclusions

7.1 Expected and Unexpected Findings

AS ALLUDED TO in **Chapter 1**, violations of modeling assumptions may have significant effects that are hard to foresee. In the present chapter both expected and unexpected effects are summarized. The findings suggest that, if the various sources of bias and variability are not properly taken into account, the following warnings may be warranted:

- A measurement method may mistakenly seem to be interchangeable with its golden standard (Chapter 2).
- The prediction error (the weighted difference between measurement and model output), when used as a validation criterion, may be more variable than necessary, leading to a higher probability of selecting a less than optimal model (Chapter 3).
- Intra-individual process noise may be mistaken as variability between measurements and/or individuals (Chapters 4 and 5).
- Blood–effect-site equilibration half-life and potency of a drug may be underestimated (Chapter 6).

7.2 The 95% Confidence Intervals of the Limits of Agreement

Bland-Altman methods to assess agreement between two measurement methods were studied in **Chapter 2**. The warning in the literature was confirmed that when multiple measurements have been obtained in several individuals, an analysis that does not take this into account may be expected to give limits of agreement that are too narrow. However, using simulations it was shown that this has even more of an effect on the confidence intervals around these limits. When reporting how closely measurements between two devices are related, the accuracy of the limits of agreement is just as important as the limits themselves. Clearly, suitable software that permits easy calculation of these confidence intervals can be helpful in assessing the value of medical devices. To that end, an open-source web application was developed so that a Bland-Altman analysis can be performed without the need to install any software apart from the ubiquitous

web browser. In previous studies wrong conclusions on agreement between two methods may have been reached, particularly when the number of subjects was small. To avoid inconclusiveness, it is proposed that studies that use Bland-Altman methods of comparison should follow a standard format. By providing sufficient data on the assumptions underlying an analysis of agreement next to the results, especially the 95% confidence intervals of the limits of agreement and inter- and intra-individual variation, ambiguity can be reduced and confidence in the results increased.

7.3 Akaike's Information Theoretic Criterion

Akaike's Information Theoretic Criterion (AIC) is a number representing a model's goodness of fit, relative to competing models. The simulations performed in **Chapter 3** demonstrated that, at least in a relatively simple mixed-effects modeling context with a set of prespecified models, minimum mean AIC coincided with best predictive performance.

It was found that in the presence of interindividual variability, prediction error by itself becomes a less suitable validation criterion, because it does not take into account whether estimated interindividual variability matches the variability in the validation data. The context of AIC is the one where the random effects have been integrated out, with the parameters at their (estimated) population values, which is to be done when all data are acquired. This holds also for the validation data, so this context is different from the case where prediction errors are calculated with the random effects set to zero. In other words, interindividual variability is predicted as well; the distributions of the model parameters are estimated to allow optimal prediction of a new set of data, even when the individualized model parameter values remain unknown until enough data are gathered.

7.4 Kalman-Filtered Concentrations and Measures of Analgesia

The opioid buprenorphine significantly increased the resting state EEG ratio (a surrogate EEG measure of analgesia) and skin pain tolerance compared with placebo, as was demonstrated in **Chapter 4**. A stochastic model was applied to the data, which adequately characterized the concentration-time and effect-time courses for both the skin heat stimulation and the resting state EEG ratio outcomes, with variations in the drug's absorption rate during a 144-hour treatment period. As measured by the potency parameter, the EEG effect was about 10 times more sensitive to buprenorphine than the skin pain test. The findings suggest that the resting state EEG ratio is an objective alternative for assessing opioid effect.

The stochastic PK-PD analysis was successful, in the sense that three kinds of random sources could be identified: variability between individuals, variability within individuals, and variability in measurements. This allowed the computation of a time-dependent variability in drug absorption from patch to blood. However, the effects of ignoring this variability remained unknown.

7.5 Kalman-Filtered Surrogate EEG Measures of Anesthesia

An example where the standard two-stage (combining results from separate fits for each individual) and nonlinear mixed-effects modeling (NONMEM) approach yielded nearly identical parameter estimates was encountered in **Chapter 5**. Furthermore, it was found that the interindividual variability identified by a mixed-effects but otherwise standard PK-PD analysis was for a large part actually intra-individual variability, namely process noise.

Analysis of permutation entropy data calculated from raw EEG measurements with a first Kalman filter design displayed a large value of the steepness parameter ($\gamma \approx 17$) of the sigmoid-$E_{\max}$ model. As a consequence, the model output is very sensitive in the region of loss of consciousness, but very insensitive in deeper levels of anesthesia. In contrast, analysis of the data with an alternative Kalman filter design showed a relatively low value ($\gamma \approx 2.5$), where the model output responds smoothly to changes in anesthetic concentration.

Simulations showed that the model parameters could be reliably estimated. Estimated parameter values were similar if the Kalman filter was present or absent in the model, except for the interindividual variability estimates. Without the Kalman filter, these were overestimated with a factor of 10–30. Furthermore, the steepness parameter was not overestimated. Interestingly, the largest interindividual variability (coefficient of variation about 25%) was found to be present on the standard deviation of the process noise.

7.6 Sampling Site Bias

Arteriovenous morphine-6-glucuronide (M6G) concentration differences were analyzed in **Chapter 6**. Arterial plasma concentrations were higher just after infusion, whereas at later times venous concentrations exceeded arterial concentrations. An extended pharmacokinetic model adequately described the data; it consisted of three arterial compartments, one central venous compartment, and one peripheral venous compartment.

The simulation studies revealed large biases in model parameters derived from venous concentration data. The biases were dependent on the value of $t_{½,k_{e0}}$, the blood–effect-site equilibration delay. Assuming that the true value of M6G's $t_{½,k_{e0}}$ may be in the range of 120 to 240 minutes (depending on the endpoint measured), we would have underestimated $t_{½,k_{e0}}$ by 30%, whereas the potency parameter would have been overestimated by about 40%, when using venous plasma samples.

A delay between arterial and venous concentrations would not be unexpected. If $t_{½,k_{e0}}$ would be estimated based on venous data, we would expect a smaller value than if it would be based on arterial data. There are two other PD parameters, C_{50} and γ; and interestingly their values may also be biased when estimated from venous data. Most of the duration of the experiment, venous concentrations were higher than arterial concentrations, so when the effect occurs at higher concentrations, this would lead to an upward biased C_{50}.

While this was not investigated, the biases are likely dependent on the administration schedule. That means that the biases are very hard to know in advance, if the admin-

istration schedule is different than the one upon which a model is based. While this is arguably always the case, because a PD model may also be imperfect when it is based on arterial concentrations, it seems prudent to not further confound a PK-PD analysis by basing it on venous concentrations.

So because of large arteriovenous differences in M6G concentrations, biases in pharmacodynamic model parameters will occur when linking venous concentration to effect using a traditional effect-compartment model.

7.7 NONMEM: User, stop. Stop, will you? Stop, User. Will you stop User? Stop, User.*

This thesis ends with a few suggestions for future studies.

1. The Bland-Altman plot gives limits of agreement, and when preset criteria are met, a new measurement method may be assessed to be as good as the golden standard. But what is the probability that this decision is correct?

2. The simulations performed to study the behavior of Akaike's criterion incorporated only one interindividual variability term. The expected value of the criterion remained minimal at the optimal model and the probability of selecting the optimal model remained the same. Do these findings remain true with many more random effects? When would multimodel inference[18] be of help?

3. Implementing the Kalman filter is quite difficult because of the complex derivation of, and much larger number of differential equations. This leads to further analyses being carried out with biased results with respect to the variances of the random effects. How could the implementation be facilitated?

4. Interindividual variability may be dealt with by Bayesian individualization.[63] However, this possibility vanishes if this variability is actually also filtered intraindividual variability. Should monitors of depth of anesthesia incorporate a Kalman filter?

5. Arterial-venous concentration differences were significant for the one drug studied leading to biased pharmacodynamic model parameter estimates. Without an *a priori* reason why this would not hold for another drug under investigation, venous sampling is not to be recommended if accurate pharmacodynamic model building is desired. But could there be a predictive covariate for the magnitude of the concentration differences?

* Adapted quote of HAL (http://www.imdb.com/title/tt0062622/quotes)

Bibliography

1. H Akaike. A new look at the statistical model identification. *IEEE Trans Automat Contr* 1974; 19:716–23

2. H Akaike. Citation classic: A new look at the statistical model identification. *CC/Eng Tech Appl Sci* 1981; 51:22

3. M Akin. Comparison of wavelet transform and FFT methods in the analysis of EEG signals. *J Med Syst* 2002; 26:241–247

4. D G Altman, J M Bland. Measurement in medicine: the analysis of method comparison studies. *The Statistician* 1983; 32:307–317

5. T Andresen, C Staahl, A Oksche, H Mansikka, L Arendt-Nielsen, A M Drewes. Effect of transdermal opioids in experimentally induced superficial, deep and hyperalgesic pain. *Br J Pharmacol* 2011; 164:934–945

6. T Andresen, R N Upton, D J Foster, L L Christrup, L Arendt-Nielsen, A M Drewes. Pharmacokinetic/pharmacodynamic relationships of transdermal buprenorphine and fentanyl in experimental human pain models. *Basic Clin Pharmacol Toxicol* 2011; 108:274–284

7. L Arendt-Nielsen, T Andresen, L P Malver, A Oksche, H Mansikka, A M Drewes. A double-blind, placebo-controlled study on the effect of buprenorphine and fentanyl on descending pain modulation: a human experimental study. *Clin J Pain* 2012; 28:623–627

8. C Bandt, B Pompe. Permutation entropy: a natural complexity measure for time series. *Phys Rev Lett* 2002; 17:174102–1–4

9. S L Beal, L B Sheiner, A J Boeckman, R J Bauer, editors. NONMEM User's Guides. Icon Development Solutions, Hanover, MD, USA, 1989-2016

10. G Biancofiore, L A H Critchley, A Lee, X Yang, L M Bindi, M Esposito, M Bisà, L Meacci, R Mozzo, F Filipponi. Evaluation of a new software version of the Flotrac/Vigileo (version 3.02) and a comparison with previous data in cirrhotic patients undergoing liver transplant surgery. *Anesth Analg* 2011; 113:515–522

11. J M Bland, D G Altman. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol* 2003; 22:85–93

12. J M Bland, D G Altman. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat* 2007; 17:571–582

13. J M Bland, D G Altman. Agreed statistics: measurement method comparison. *Anesthesiology* 2012; 116:182–185

14. J M Bland, D G Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; i:307–310

15. J M Bland, D G Altman. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; 8:135–160

16. P Bonate. Pharmacokinetic-pharmacodynamic modeling and simulation (2nd ed.). Springer, New York, 2011

17. R K Burdick, F A Graybill. Confidence intervals on linear combinations of variance components in the unbalanced one-way classification. *Technometrics* 1984; 26:131–136

18. K P Burnham, D R Anderson. Model selection and multimodel inference (2nd ed.). Springer, New York, 2002

19. K P Burnham, D R Anderson. Multimodel inference - understanding AIC and BIC in model selection. *Sociol Meth Res* 2004; 33:261–304

20. Y Cao, W W Tung, J B Gao, V A Protopopescu, L M Hively. Detecting dynamical changes in time series using the permutation entropy. *Phys Rev E Stat Nonlin Soft Matter Phys* 2004; 70:046217

21. C Chatfield. Model uncertainty, data mining and statistical inference. *J R Statist Soc A* 1995; 158:419–466

22. A C Clarke. 2001: A Space Odyssey. New American Library, New York, 1968

23. M O Columb. Clinical measurement and assessing agreement. *Curr Anaesth Crit Care* 2008; 19:328–329

24. A Dahan, E Olofsen, M Sigtermans, I Noppers, M Niesters, L Aarts, M Bauer, E Sarton. Population pharmacokinetic-pharmacodynamic modeling of ketamine-induced pain relief of chronic pain. *Eur J Pain* 2011; 15:258–267

25. A Dahan, A Yassen, H Bijl, R Romberg, E Sarton, L Teppema, E Olofsen, M Danhof. Comparison of the respiratory effects of intravenous buprenorphine and fentanyl in humans and rats. *Br J Anaesth* 2005; 94:825–834

26. A Dahan, A Yassen, R Romberg, E Sarton, L Teppema, E Olofsen, M Danhof. Buprenorphine induces ceiling in respiratory depression but not in analgesia. *Br J Anaesth* 2006; 96:627–632

27. M R Digiacomo, J Marco-Pallarés, A B Flores, C M Gómes. Wavelet analysis of the EEG during the neurocognitive evaluation of invalidly cued targets. *Brain Res* 2008; 1234:94–103

28. A Donner, G Y Zou. Closed-form confidence intervals for functions of the normal mean and standard deviation. *Stat Methods Med Res* 2012; 21:347–359

29. B Efron, R J Tibshirani. An introduction to the bootstrap. Chapman and Hall, New York, 1993

30. E I Ette, P J Williams, Y H Kim, J R Lane, M-J Liu, E V Capparelli. Model appropriateness and population pharmacokinetic modeling. *J Clin Pharmacol* 2003; 43:610–623

31. Y Fang. Asymptotic equivalence between cross-validations and Akaike information criteria in mixed-effects models. *J Data Sci* 2011; 9:15–21

32. G C Goodwin, R L Payne. Dynamic system identification. Academic Press, New York, 1977

33. M Gram, C Graversen, A K Nielsen, T Arendt-Nielsen, C D Mørch, T Andresen, A M Drewes. A novel approach to pharmaco-EEG for investigating analgesics: assessment of spectral indices in single-sweep evoked brain potentials. *Br J Clin Pharmacol* 2013; 76:951–963

34. C Graversen, L Malver, G P Kurita, C Staahl, P Sjøgren, L L Christrup, A M Drewes. Altered frequency distribution in the electroencephalogram is correlated to the analgesic effect of remifentanil. *Basic Clin Pharmacol Toxicol* 2015; 116:414–422

35. C Graversen, S S Olesen, A E Olesen, K Steimle, D Farina, O H Wilder-Smith, S A Bouwense, H van Goor, A M Drewes. The analgesic effect of pregabalin in patients with chronic pain is reflected by changes in pharmaco-EEG spectral indices. *Br J Clin Pharmacol* 2012; 73:363–372

36. S Greven, B Kneib. On the behaviour of marginal and conditional AIC in linear mixed effects models. *Biometrika* 2010; 97:773–789

37. M Gumbleton, S Øie, D Verotta. Pharmacokinetic-pharmacodynamic (PK-PD) modelling in non-steady-state studies and arterio-venous drug concentration differences. *Br J Clin Pharmacol* 1994; 38:389–400

38. C Hamilton, S Lewis. The importance of using the correct bounds on the Bland-Altman limits of agreement when multiple measurements are recorded per patient. *J Clin Monit Comput* 2010; 24:173–175

39. C Hamilton, J Stamey. Using Bland-Altman to assess agreement between two medical devices - don't forget the confidence intervals! *J Clin Monit Comput* 2007; 21:331–333

40. T Hastie, R J Tibshirani, J H Friedman. The elements of statistical learning: data mining, inference, and prediction (2nd ed.). Springer, New York, 2009

41. C J Hull, H B Van Beem, K McLeod, A Sibbald, M J Watson. A pharmacodynamic model for pancuronium. *Br J Anaesth* 1978; 50:1113–1123

42. J B Jørgensen, P G Thomsen, H Madsen. A computationally efficient and robust implementation of the continuous-discrete extended Kalman filter. Proceedings of the 2007 American Control Conference, New York City, NY:3707-12, 2007

43. E R Kalman. A new approach to linear filtering and prediction problems. *J Basic Eng* 1960; 82:35–45

44. E R Kalman. Contributions to the theory of optimal control. *Bol Soc Mat Mex* 1960; 5:102–119

45. E R Kalman. Citation classic: Contributions to the theory of optimal control. *CC/Phys Chem Earth SCi* 1979; 32:14

46. V J Knott. Quantitative EEG methods and measures in human psychopharmacological research. *Hum Psychopharmacol* 2000; 15:479–498

47. J Kortelainen, E Väyrynen, T Seppänen. Depth of anesthesia during multidrug infusion: separating the effects of propofol and remifentanil using the spectral features of EEG. *IEEE Trans Biomed Eng* 2011; 58:1216–1223

48. D T Liley, N C Sinclair, T Lipping, B Heyse, H E Vereecke, M M Struys. Propofol and remifentanil differentially modulate frontal electroencephalographic activity. *Anesthesiology* 2010; 113:292–304

49. J Lötsch. Pharmacokinetic-pharmacodynamic modeling of opioids. *J Pain Symptom Manage* 2005; 29:S90–S103

50. L P Malver, A Brokjaer, C Staahl, C Graversen, T Andresen, A M Drewes. Electroencephalography and analgesics. *Br J Clin Pharmacol* 1014; 77:72–95

51. S Mantha, M F Roizen, L A Fleisher, R Thisted, J Foss. Comparing methods of clinical measurement: reporting standards for Bland and Altman analysis. *Anesth Analg* 2000; 90:593–602

52. C Martini, E Olofsen, A Yassen, L Aarts, A Dahan. Pharmacokinetic-pharmacodynamic modeling in acute and chronic pain: an overview of the recent literature. *Expert Rev Clin Pharmacol* 2011; 4:719–728

53. P S Maybeck. Model selection and multimodel inference (vol.1). Academic Press, New York, 1979

54. Mozilla Foundation. JavaScript. `https://developer.mozilla.org/en-US/docs/JavaScript`, accessed April 2013

55. P S Myles, J Cui. Using the Bland-Altman method to measure agreement with repeated measures (Editorial). *Br J Anaesth* 2007; 99:309–311

56. G Newshan. Heat-related toxicity with the fentanyl transdermal patch. *J Pain Symptom Manage* 1998; 16:277–278

57. K H Norwich. Noncompartmental models of whole-body clearance of tracers: A review. *Ann Biomed Eng* 1997; 25:421–439

58. P D Oldham. A note on the analysis of repeated measurements of the same subjects. *J Chronic Dis* 1962; 15:969–977

59. A E Olesen, T Andresen, C Staahl, A M Drewes. Human experimental pain models for assessing the therapeutic efficacy of analgesic drugs. *Pharmacol Rev* 2012; 64:722–779

60. E Olofsen. The performance of model selection criteria in the absence of a fixed-dimensional correct model. *PAGE 16* 2007;

61. E Olofsen. Simultaneous stochastic modeling of pharmacokinetic and pharmacodynamic data with noncoinciding sampling times. Population Approach Group Europe (PAGE); Abstract #2696, 2013

62. E Olofsen, A Dahan. The dynamic relationship between end-tidal sevoflurane and isoflurane concentrations and bispectral index and spectral edge frequency of the electroencephalogram. *Anesthesiology* 1999; 90:1345–1353

63. E Olofsen, D F Dinges, H P A Van Dongen. Nonlinear mixed-effects modeling: individualization and prediction. *Aviat Space Environ Med* 2004; 75:A134–140

64. E Olofsen, M Sigtermans, I Noppers, M Niesters, R Mooren, M Bauer, L Aarts, E Sarton, A Dahan. The dose-dependent effect of S(+)-ketamine on cardiac output in healthy volunteers and complex regional pain syndrome type 1 chronic pain patients. *Anesth Analg* 2012; 115:536–546

65. E Olofsen, J W Sleigh, A Dahan. Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect. *Br J Anaesth* 2008; 101:810–821

66. B Pietrzak, E Czarnecka. Pharmaco-EEG-based assessment of the interaction between ethanol and oxcarbazepine. *Pharmacol Rep* 2010; 62:278–286

67. M B Priestley. Spectral Analysis. Academic Press, London, 1981

68. R Romberg, E Olofsen, E Sarton, J DenHartigh, P Taschner, A Dahan. Pharmacokinetic-pharmacodynamic modeling of morphine-6- glucoronide–induced analgesia in healthy volunteers: absence of sex differences. *Anesthesiology* 2004; 100:120–133

69. R Romberg, E Olofsen, E Sarton, L Teppema, A Dahan. Pharmacodynamic effect of morphine-6-glucuronide versus morphine on hypoxic and hypercapnic breathing in healthy volunteers. *Anesthesiology* 2003; 99:788–798

70. V J Samar, A Bopardikar, R Rao, K Swartz. Wavelet analysis of neuroelectric waveforms: a conceptual tutorial. *Brain Lang* 1999; 66:7–60

71. E Sarton, E Olofsen, R Romberg, J DenHartigh, B Kest, D Nieuwenhuijs, A Burm, L Teppema, A Dahan. Sex differences in morphine analgesia - an experimental study in healthy volunteers. *Anesthesiology* 2000; 93:1245–1254

72. Schnur D, et al. Flot: Attractive JavaScript plotting for jQuery. `http://www.flotcharts.org/`, accessed April 2013

73. J C Scott, J E Cooke, D R Stanski. Electroencephalographic quantitation of opioid effect: comparative pharmacodynamics of fentanyl and sufentanil. *Anesthesiology* 1991; 74:34–42

74. G Segre. Kinetics of interaction between drugs and biological systems. *Farmaco Sci* 1968; 23:907–918

75. C E Shannon. A mathematical theory of communication. *Bell Syst Tech J* 1948; 27:379–423

76. C E Shannon, W Weaver. The mathematical theory of communication. University of Illinois Press, Urbana, 1949

77. L B Sheiner, B Rosenberg, V V Marathe. Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *J Pharmacokin Biopharm* 1977; 5:445–479

78. L B Sheiner, B Rosenberg, K L Melmon. Modelling of individual pharmacokinetics for computer-aided drug dosage. *Comput Biomed Res* 1972; 5:441–459

79. L B Sheiner, D R Stanski, S Vozeh, R D Miller, J Ham. Simultaneous modeling of pharmacokinetics and pharmacodynamics: application to d-tubocurarine. *Clin Pharmacol Ther* 1979; 25:358–371

80. F Sjöstrand, P Rodhe, E Berglund, N Lundström, C Svensen. The use of a noninvasive hemoglobin monitor for volume kinetic analysis in an emergency room setting. *Anesth Analg* 2013; 116:337–342

81. C Staahl, L L Christrup, S D Andersen, L Arendt-Nielsen, A M Drewes. A comparative study of oxycodone and morphine in a multi-modal, tissue-differentiated experimental pain model. *Pain* 2006; 123:28–36

82. C Staahl, R Upton, D J Foster, L L Christrup, K Kristensen, S H Hansen, L Arendt-Nielsen, A M Drewes. Pharmacokinetic-pharmacodynamic modeling of morphine and oxycodone concentrations and analgesic effect in a multimodal experimental pain model. *J Clin Pharmacol* 2008; 48:619–631

83. E W Steyerberg. Clinical prediction models. A practical approach to development, validating, and updating. Springer, New York, 2009

84. S F Storti, E Formaggio, A Beltramello, A Fiaschi, P Manganotti. Wavelet analysis as a tool for investigating movement-related cortical oscillations in EEG-fMRI coregistration. *Brain Topogr* 2010; 23:46–57

85. R Development Core Team. R: A language and environment for statistical computing. `https://www.r-project.org/`, accessed April 2017

86. The jQuery Foundation. jQuery. `http://jquery.com/`, accessed April 2013

87. J D Thomas, R A Hultquist. Interval estimation for the unbalanced case of the one-way random effects model. *Ann Statist* 1978; 6:582–587

88. P H Tonner, B Bein. Classic electroencephalographic parameters: median frequency, spectral edge frequency etc. *Best Pract Res Clin Anaesthesiol* 2006; 20:147–159

89. C W Tornøe, J L Jacobsen, H Madsen. Grey-box pharmacokinetic/pharmacodynamic modelling of a euglycaemic clamp study. *J Math Biol* 2004; 48:591–604

90. C W Tornøe, R V Overgaard, H Agersø, H A Nielsen, H Madsen, E N Jonsson. Stochastic differential equations in NONMEM: implementation, application, and comparison with ordinary differential equations. *Pharm Res* 2005; 22:1247–1258

91. A M Trescot, S Data, M Lee, H Hansen. Opioid pharmacology. *Pain Physician* 2008; 11:S133–S153

92. B Tuk, M Danhof, J W Mandema. The impact of arteriovenous concentration differences on pharmacodynamic parameter estimates. *J Pharmacokin Biopharm* 1997; 25:39–62

93. B Tuk, V M Herben, J W Mandema, M Danhof. Relevance of arteriovenous concentration differences in pharmacokinetic-pharmacodynamic modeling of midazolam. *J Pharm Exp Ther* 1998; 284:202–207

94. K Uemura, T Kawada, M Inagaki, M Sugimachi. A minimally invasive monitoring system of cardiac output using aortic flow velocity and peripheral arterial pressure profile. *Anesth Analg* 2013; 116:1006–1017

95. UncertWeb. jStat: a JavaScript statistical library. `http://www.jstat.org/`, accessed April 2013

96. F Vaida, S Blanchard. Conditional Akaike information for mixed-effects models. *Biometrika* 2005; 92:351–70

97. R Van Noorden, B Maher, R Nuzzo. The top 100 papers. *Nature* 2014; 514:550–553

98. Y Yang. Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. *Biometrika* 2005; 92:937–50

99. G A Young, N Khazan. Differential neuropharmacological effects of mu, kappa and sigma opioid agonists on cortical EEG power spectra in the rat. *Neuropharmacology* 1984; 23:1161–1165

100. G Y Zou. Confidence interval estimation for the Bland-Altman limits of agreement with multiple observations per individual. *Stat Methods Med Res* 2013; 22:630–642

Samenvatting

Voorziene en Onvoorziene Bevindingen

VOOR HET FITTEN (passend maken) van gemengde modellen van de eigenschappen van geneesmiddelen werd in de jaren zeventig van de vorige eeuw een start gemaakt met de ontwikkeling van het computerprogramma "NONMEM", en dit programma wordt nog steeds verbeterd. Een gemengd model kan rekening houden met zowel onverklaarbare verschillen tussen individuen als verklaarbare (zoals verschillen tussen mannen en vrouwen).

NONMEM is gemaakt om instructies van een gebruiker op te volgen, die in een "control file" aangekondigd worden met het woord "$PROBLEM". Twee belangrijke stappen zijn: 1) "$ESTIMATION": geef de meest aannemelijke waarden van de parameters van het model, en 2) "$COVARIANCE": geef een indicatie van de schattingsfouten in de waarden van de vorige stap. De parameterschattingen zijn in een bepaalde zin optimaal als een zeker criterium, de zogenaamde "objective function" een minimum heeft bereikt.

Beide stappen kunnen foutmeldingen opleveren. De eerste stap kan bijvoorbeeld melden dat de OBJECTIVE FUNCTION INFINITE is; de tweede stap bijvoorbeeld dat er een MATRIX, benodigd voor het verkrijgen van de schattingsfouten, SINGULAR is. (Het gebruik van woorden in hoofdletters lijkt onbeleefd, maar hoofdletters waren heel gebruikelijk in de tijd dat de ontwikkeling van NONMEM begon.) Dit soort foutmeldingen kunnen optreden zelfs als het voorgestelde model heel plausibel is, en vaak nadat veel rekentijd verstreken is. Het lijkt er - onterecht natuurlijk - wel eens op alsof er een diepere reden is om de stappen te dwarsbomen - vandaar de interpretatie van een beroemde discussie uit de science fiction literatuur[22] op het kaartje. In werkelijkheid was NONMEM onmisbaar voor de in dit proefschrift besproken studies.

Bij het fitten worden er expliciete en impliciete aannamen gedaan, bijvoorbeeld dat gewicht normaal verdeeld is of altijd positief is, of geslacht twee waarden kan aannemen. In de inleiding, **Hoofdstuk 1**, werd er op gezinspeeld dat een statistische analyse, waarbij gedane aannamen geschonden worden, resultaten kan opleveren die wel te verwachten waren, of juist moeilijk te voorzien zijn. In dat hoofdstuk werden de onderwerpen ingeleid waarbij de gevolgen van het schenden van aannamen bestudeerd werden. In alle gevallen gaat het om analyses met behulp van lineaire of niet-lineaire gemengde modellen. In de volgende secties worden zowel de verwachte als onverwachte resultaten samengevat. De bevindingen suggereren dat de volgende waarschuwingen gerechtvaardigd zijn:

- Een meetmethode kan ten onrechte als uitwisselbaar met de "gouden standaard" bevonden worden als interindividuele variabiliteit aanwezig is, en omgekeerd (Hoofdstuk 2).

- De voorspellingsfout (gewogen verschil tussen meting en uitkomst van een model), gebruikt als validatiecriterium, kan onzekerder worden als interindividuele variabiliteit aanwezig is. Dit leidt tot een grotere kans om aan een suboptimaal model de voorkeur te geven (Hoofdstuk 3).
- Intra-individuele "procesruis" - variatie in metingen die niet verklaard kan worden door meetfouten - kan ten onrechte worden aangezien als variatie tussen individuen (Hoofdstukken 4 en 5).
- De werkzaamheid van een geneesmiddel en de equilibratie-halfwaardetijd tussen concentraties in het bloed en op plaats van werking worden mogelijk onderschat (Hoofdstuk 6).

De 95% Betrouwbaarheidsintervallen van de Grenzen van Overeenstemming

De methoden van Bland en Altman om de overeenstemming tussen twee meetmethoden te bepalen werden bestudeerd in **Hoofdstuk 2**. In het bijzonder werd er gekeken naar de invloed van correlatie tussen metingen binnen individuen. De waarschuwing in de literatuur, dat een analyse waarbij er geen rekening wordt gehouden met correlaties te nauwe grenzen van overeenstemming kan opleveren, werd bevestigd. Maar dankzij simulaties kwam er naar voren dat correlaties een groter effect hebben op de betrouwbaarheidsintervallen rond te grenzen van overeenstemming. Bij het rapporteren van de overeenstemming tussen twee meetmethoden is de nauwkeurigheid van de grenzen van overeenstemming natuurlijk net zo belangrijk als de grenzen zelf. Software die geschikt is om deze berekeningen gemakkelijk uit te voeren kan van pas komen bij het evalueren van de toepasbaarheid van nieuwe medische apparaten. Met dat doel werd een "open source" webapplicatie ontwikkeld zodat een Bland-Altman analyse uitgevoerd kan worden zonder software te installeren (behalve een webbrowser die vrijwel altijd beschikbaar is). Het zou kunnen dat in eerdere studies incorrecte conclusies over de uitwisselbaarheid van meetmethoden werden getrokken door een te beperkt aantal individuen. Om ambivalentie te voorkomen werd voorgesteld een standaard te hanteren voor het uitvoeren en rapporteren van Bland-Altman analyses. Door naast de analyseresultaten ook voldoende gegevens over de gedane aannames te verschaffen kan het vertrouwen in de resultaten worden vergroot.

Het op de Informatietheorie Gebaseerde Criterium van Akaike

Het criterium van Akaike is een getal dat een maat is voor de kwaliteit van een model. De absolute waarde van het getal is niet relevant; het getal is alleen te interpreteren ten opzichte van de grootte van het getal voor alternatieve modellen. De simulaties die in **Hoofdstuk 3** werden besproken lieten zien dat, in het geval van een verzameling vooraf gespecificeerde eenvoudige gemengde modellen, de gemiddelde minimum AIC overeenkomt met de beste prestaties in het kader van het voorspellen van nieuwe data.

Een andere bevinding was dat in aanwezigheid van interindividuele variabiliteit de voorspellingsfout als zodanig minder geschikt wordt om als modelvalidatiecriterium gebruikt te worden, omdat deze in principe geen rekening kan houden met deze variabiliteit. De context van AIC in het geval van gemengde modellen is die waarbij inter-

individuele variabiliteit wordt meegenomen door te integreren over alle mogelijkheden. Dit gebeurt met de parameters op de geschatte waarden voor de populatie, nadat alle data zijn verzameld. Dit geldt ook voor de validatiedata, dus deze context verschilt van die waarbij voorspellingsfouten worden berekend aan de hand van de voorspelling voor het "gemiddelde" individu. Met andere woorden, de interindividuele variabiliteit wordt ook voorspeld; de verdelingen van de parameters van een model worden geschat voor de optimale voorspelling van een verzameling nieuwe data, ook al blijven de geïndividualiseerde modelparameters onbekend totdat genoeg data is verzameld.

Concentraties en Maten voor Analgesie Nasporen met het Kalman Filter

In **Hoofdstuk 4** werd de stijging in de tolerantie voor hittepijn en in de "EEG ratio in rusttoestand" ten gevolge van het morfinomimeticum buprenorfine bestudeerd. De EEG ratio is een surrogaatmaat voor de diepte van analgesie, gebaseerd op de verhouding van de aanwezigheid van snelle en langzame golven in het EEG. Deze data werd adequaat beschreven met een stochastisch model, waarbij variabiliteit werd aangenomen in de absorptie uit de buprenorfinepleister gedurende de 144 uur dat deze was aangebracht. Afgaand op de parameter die de werkzaamheid kwantificeert, was de EEG ratio ongeveer 10 maal gevoeliger voor buprenorfine dan de hittepijntest. De bevindingen suggereren dat de EEG ratio een gevoelige en objectieve maat is om te kijken naar het effect van de toediening van opioïden.

De stochastische PK-PD analyse was succesvol, in de zin dat drie bronnen van variabileit konden worden geïdentificeerd: residuele ("onverklaarde") variabiliteit tussen individuen, variabiliteit in de individuen (procesruis), en variabiliteit tussen metingen. Dit maakte het mogelijk de variatie in de tijd te volgen van de absorptie van buprenorfine uit de pleister. In dit hoofdstuk werd niet gekeken naar de invloed van het negeren van deze bron van variabiliteit. Het kan zijn dat een farmacokinetisch model waarbij procesruis niet ingebouwd is zo slecht fit, dat de concentratiemetingen zelf een betere basis vormen voor een farmacodynamisch model. Maar een Kalman filter geeft een optimale voorspelling van de concentraties, in de aanwezigheid van procesruis en meetruis.

Surrogaatmaten voor Anesthesie Nasporen met het Kalman Filter

Een voorbeeld waarbij een standaard analyse (het combineren van de resultaten van individuele fits) en een analyse met (niet-lineaire) gemengde modellen (NONMEM) vrijwel dezelfde parameterschattingen opleverde werd aangetroffen in **Hoofdstuk 5**. Een andere bevinding was dat de interindividuele variabiliteit, geschat met NONMEM, eigenlijk voor een groot deel intra-individuele variabiliteit oftewel procesruis was.

De permutatie-entropie is een maat voor de onvoorspelbaarheid van een signaal, waarbij de amplitude op zich geen rol speelt, maar alleen de amplitude van een meting ten opzichte van enkele daaraanvoorafgaande metingen. De analyse van de permutatie-entropie van het EEG met behulp van een eerste ontwerp van een Kalman filter leverde een zeer grote niet-lineariteit van het farmacodynamische model op. Dat houdt in dat het model heel gevoelig is rond het verlies of de terugkeer van het bewustzijn, maar dat het diepere anesthesie niet kan volgen. Een tweede ontwerp gaf een veel kleinere mate van niet-lineariteit, waarbij het model de concentratie van het toegediende anestheticum

vloeiender kon volgen. Het verschil tussen deze ontwerpen had te maken met de locatie van de procesruis - aan de uitgang of aan de ingang van het farmacodynamisch model.

Als data wordt gesimuleerd, dan kan er gekeken worden of een analyse met behulp van modellen inclusief een Kalman filter de eigenschappen van die data inderdaad oplevert. Het bleek dat de eigenschappen betrouwbaar teruggeschat kunnen worden. De parameterwaarden hingen nauwelijks af van de aan- of afwezigheid van een Kalman filter, behalve voor de parameters die interindividuele variabiliteit aangeven. Zonder een Kalman filter waren deze soms met een factor van 10 tot 30 overschat. Ook bleek dat de niet-lineariteit goed geschat werd. De grootste overgebleven interindividuele variabiliteit bleek aanwezig op de standaarddeviatie van de procesruis. Deze bevindingen zijn interessant omdat ze een aanwijzing kunnen vormen bij het verklaren van de onvoorspelbaarheid in de farmacodynamiek van anesthetica.

De Locatie van Bloedafname en Systematische Fouten

Arterioveneuze morfine-6-glucuronide (M6G) concentratieverschillen werden bestudeerd in **Hoofdstuk 6**. Vlak na toediening van M6G waren arteriële concentraties hoger dan veneuze; later waren veneuze concentraties hoger dan arteriële. Een farmacokinetisch model met een uitbreiding om zowel arteriële als veneuze concentraties te beschrijven deed dat adequaat; het bestond uit drie compartimenten voor de arteriële data en twee compartimenten extra voor de veneuze data.

Simulatiestudies lieten grote systematische fouten in parameterschattingen zien als daarvoor de veneuze data in plaats van de arteriële data gebruikt werden. De grootte van de systematische fouten hing af van de equilibratie-halfwaardetijd tussen concentraties in het bloed en op plaats van werking. Onder de aanname dat de werkelijke waarde ergens tussen de twee en vier uur ligt (afhankelijk van het farmacodynamisch eindpunt) dan zou deze equilibratie-halfwaardetijd met een factor van ongeveer 30% onderschat worden, en de werkzaamheid met een factor van 40%.

Een equilibratietijd tussen arteriële en veneuze concentraties is niet vreemd, en het valt dan ook te verwachten dat de equilibratie-halfwaardetijd tussen concentraties in het bloed en op plaats van werking onderschat wordt als bij de analyse veneuze in plaats van arteriële concentraties gebruikt worden. Er zouden ook systematisch schattingsfouten van de werkzaamheid en niet-lineariteit van de farmacodynamiek van M6G kunnen optreden. Maar van te voren is het niet eenvoudig te zeggen of deze onder- of overschat zouden worden. Aangezien de veneuze concentraties gedurende het grootste deel van het experiment hoger lagen dan de arteriële, valt te verklaren dat de werkzaamheid overschat werd.

Hoewel dit niet nader werd onderzocht, is de grootte van systematische schattingsfouten waarschijnlijk afhankelijk van het toedieningsschema van een geneesmiddel. Dat betekent dat die fouten erg moeilijk vooraf in te schatten zijn, vooral als het toedieningsschema afwijkt van het schema dat werd gebruikt bij de opstellen van een model. Hoewel te beargumenteren is dat dit altijd het geval is, omdat een farmacodynamisch model ook nooit perfect is als het gebaseerd is op arteriële concentraties, lijkt het verstandig om de kwaliteit van een PK-PD analyse niet te beperken door het te baseren op veneuze concentraties.

De studie liet ook zien dat arterioveneuze concentratieverschillen kunnen leiden tot grote fouten in voorspellingen van de farmacodynamiek, als het farmacodynamisch

model wordt gekoppeld aan veneuze concentraties, beschreven met een standaard farmacokinetisch model.

NONMEM: Gebruiker, Houdt Op. Houdt U Op? Houdt Op, Gebruiker. Houdt U Op, Gebruiker? Houdt Op, Gebruiker.

Deze samenvatting eindigt met enkele suggesties voor nieuwe studies.

- De grafiek van Bland en Altman toont de grenzen van overeenstemming, en wanneer aan vooraf gestelde criteria wordt voldaan kan een nieuwe meetmethode als "net zo goed" worden bevonden als de "gouden standaard". Maar wat is de kans dat deze beslissing correct is?

- De simulaties die werden uitgevoerd om het gedrag van het fitcriterium van Akaike te bestuderen, bevatten slechts een enkele bron van interindividuele variabiliteit. De verwachte waarde van het criterium bleef minimaal voor het optimale model, en de kans het optimale model te kiezen bleef hetzelfde in aan- en afwezigheid van deze variabiliteit. Blijven deze resultaten gelden in de aanwezigheid van veel meer bronnen van variabiliteit? Wanneer kan inferentie door middel van een verzameling van modellen nuttig zijn?

- Het implementeren van een Kalman filter in NONMEM is nogal lastig vanwege de complexe afleiding van en het grote aantal benodigde differentiaalvergelijkingen. Maar dat leidt ertoe dat er veel analyses gedaan blijven worden zonder een Kalman filter, met als gevolg systematische fouten in de geschatte relevantie van bronnen van interindividuele variabiliteit. Hoe zou de implementatie vereenvoudigd kunnen worden?

- Bayesiaanse optimalisatie van een model voor een specifiek individu is een manier om met interindividuele variabiliteit om te gaan.[63] Maar deze mogelijkheid verdwijnt als deze variabiliteit eigenlijk gefilterde intra-individuele variabiliteit is. Zouden monitors van de diepte van anesthesie eigenlijk een Kalman filter moeten hebben?

- Arterioveneuze concentratieverschillen kunnen significant zijn, en leiden tot systematische fouten in de schattingen van farmacodynamische parameters. Zonder een *a priori* reden waarom dit niet zou gelden voor een nog niet in dit opzicht bestudeerd geneesmiddel, kan veneus in plaats van arterieel bloed afnemen niet aanbevolen worden als adequate farmacodynamische modellenbouw gewenst is. Maar zou er een voorspellende covariaat voor de grootte van arterioveneuze concentratieverschillen bestaan?

Curriculum Vitae

Erik Olofsen was born on March 2, 1963, in Amsterdam, The Netherlands. He received the M.Sc. degree in Electrical Engineering from the University of Twente, Enschede, The Netherlands, in 1992. From 1987 to 1992 he was with the Department of Physiology and Physiological Physics of the University of Leiden, where he worked on the application of chaos theory to the analysis of biomedical time series. In 1992, he joined the Department of Anesthesiology of the Leiden University Medical Center where he has been a Research Associate since 1993.

List of Publications

1. M G van Erp, E Olofsen, J de Goede, A M J Van Dongen, W van de Water. Estimation of Rényi dimensions from human multichannel EEGs. On epilepsy: investigations on the level of the nerve membrane and of the brain. PhD Thesis, Leiden University, 1988; pp 78–84

2. M G van Erp, E Olofsen, J de Goede, W van de Water. On the spatiotemporal behavior of human brain activity. *Pflügers Arch - Eur J Physiol* 1988; 412:S6

3. J de Goede, E Olofsen, R Heijungs, A Berkenbosch, D S Ward. Dimensionality analysis of human respiratory flows. *FASEB Journal* 1989; 3:6284

4. J de Goede, E Olofsen. Modeling the hypoxic ventilatory depression in anesthetized cats. *J Physiol (Lond)* 1989; 417:P116

5. E Olofsen, J de Goede, R Heijungs. A maximum likelihood approach to correlation dimension and entropy estimation. *Bulletin of Mathematical Biology* 1992; 54:45–58

6. J de Goede, E Olofsen, W J Rietveld. A mathematical model of the circadian system with a delayed feedback loop. *Journal of Interdisciplinary Cycle Research* 1992; 23:142–144

7. J Vuyk, F H M Engbers, A G L Burm, A A Vletter, G E R Griever, E Olofsen, J G Bovill. Pharmacodynamic interaction between propofol and alfentanil when given for induction of anesthesia. *Anesthesiology* 1996; 82:288–299

8. F Boer, E Olofsen, J G Bovill, A G L Burm, A Hak, M J Geerts, K E Wetselaar. Pulmonary uptake of sufentanil during and after constant rate infusion. *Br J Anaesth* 1996; 76:203–208

9. E Y Sarton, A Dahan, L J S M Teppema, M J L J van den Elsen, E Olofsen, A Berkenbosch, J W VanKleef. Acute pain and central nervous system arousal do not restore impaired hypoxic ventilatory response during sevoflurane sedation. *Anesthesiology* 1996; 85:295–303

10. H P A Van Dongen, E Olofsen, J H van Hartevelt, E W Kruyt. Periodogram analysis of unequally spaced data: the Lomb method. Technical report, Leiden University, 1997

11. M J Geerts, V Chopra, E Olofsen. The Leiden Anesthesia Simulator (LAS). *Techniek in de gezondheidszorg (Technology in Health Care)* 1997; 1:22–25

12. J Vuyk, M J Mertens, E Olofsen, A G L Burm, J G Bovill. Propofol anesthesia and rational opioid selection - determination of optimal EC50-EC95 propofol-opioid concentrations that assure adequate anesthesia and a rapid return of consciousness. *Anesthesiology* 1997; 87:1549–1562

13. A Dahan, E Olofsen, E Y Sarton, L J S M Teppema. Quantifying the respiratory dynamics of sevoflurane in the cat. *Anesth Analg* 1998; 86:U265

14. E Olofsen. A technique for population pharmacodynamic analysis of concentration - binary response data. *Anesthesiology* 1998; 88:833

15. E Olofsen, A P Schipper, J Swen, A Dahan. Dynamics of the bispectral index of the EEG: sevoflurane versus isoflurane. *Anesthesiology* 1998; 89:U817

16. E Olofsen, V Chopra, M J Geerts, J G Bovill. Familiarizing anesthetists with a new drug using the Leiden Anaesthesia Simulator. EuroSIVA 2nd Annual Meeting, Amsterdam, The Netherlands, 1999

17. J A Kuipers, F Boer, E Olofsen, W Olieman, A A Vletter, A G L Burm, J G Bovill. Recirculatory and compartmental pharmacokinetic modeling of alfentanil in pigs - the influence of cardiac output. *Anesthesiology* 1999; 90:1146–1157

18. A Dahan, E Olofsen, L J S M Teppema, E Y Sarton, C N Olievier. Speed of onset and offset and mechanisms of ventilatory depression from sevoflurane - an experimental study in the cat. *Anesthesiology* 1999; 90:1119–1128

19. E Olofsen, A Dahan. The dynamic relationship between end-tidal sevoflurane and isoflurane concentrations and bispectral index and spectral edge frequency of the electroencephalogram. *Anesthesiology* 1999; 90:1345–1353

20. H P A Van Dongen, E Olofsen, J H VanHartevelt, E W Kruyt. A procedure of multiple period searching in unequally spaced time-series with the Lomb-Scargle method. *Biol Rhythm Res* 1999; 30:149–177

21. H P A Van Dongen, E Olofsen, J H VanHartevelt, E W Kruyt. Searching for biological rhythms: peak detection in the periodogram of unequally spaced data. *J Biol Rhythms* 1999; 14:617–620

22. E Sarton, E Olofsen, R Romberg, J DenHartigh, B Kest, D Nieuwenhuijs, A Burm, L Teppema, A Dahan. Sex differences in morphine analgesia - an experimental study in healthy volunteers. *Anesthesiology* 2000; 93:1245–1254

23. J A Kuipers, F Boer, E Olofsen, J G Bovill, A G L Burm. Recirculatory pharmacokinetics and pharmacodynamics of rocuronium in patients - the influence of cardiac output. *Anesthesiology* 2001; 94:47–55

24. J A Kuipers, F Boer, A DeRoode, E Olofsen, J G Bovill, A G L Burm. Modeling population pharmacokinetics of lidocaine - should cardiac output be included as a patient factor? *Anesthesiology* 2001; 94:566–573

25. M J Mertens, J Vuyk, E Olofsen, J G Bovill, A G L Burm. Propofol alters the pharmacokinetics of alfentanil in healthy male volunteers. *Anesthesiology* 2001; 94:949–957

26. A Dahan, D J F Nieuwenhuijs, E Olofsen, E Y Sarton, R R Romberg, L J S M Teppema. Response surface modeling of alfentanil-sevoflurane interaction on cardiorespiratory control and bispectral index. *Anesthesiology* 2001; 94:982–991

27. H P A Van Dongen, T Ruf, E Olofsen, J H van Hartevelt, E W Kruyt. Analysis of problematic time series with the Lomb-Scargle method, a reply to "emphasizing difficulties in the detection of rhythms with Lomb-Scargle periodograms". *Biol Rhythm Res* 2001; 32:347–354

28. H P A Van Dongen, T Ruf, E Olofsen, J H van Hartevelt, E W Kruyt. A brief response to dr. Schimmel's reply. *Biol Rhythm Res* 2001; 32:361–362

29. E Olofsen, D J F Nieuwenhuijs, E Y Sarton, L J S M Teppema, R R Romberg, A Dahan. Response surface modeling of drug interactions on cardiorespiratory control. *Frontiers In Modeling Control Breathing* 2001; 499:303–308

30. E Olofsen, J W Sleigh, A Dahan. The influence of remifentanil on the dynamic relationship between sevoflurane and surrogate anesthetic effect measures. *Anesthesiology* 2002; 96:555–564

31. H Higuchi, Y Adachi, A Dahan, E Olofsen, S Arimura, T Mori, T Satoh. The interaction between propofol and clonidine for loss of consciousness. *Anesth Analg* 2002; 94:886–891

32. D J F Nieuwenhuijs, E Olofsen, R R Romberg, E Sarton, D Ward, F Engbers, J Vuyk, R Mooren, L J Teppema, A Dahan. Response surface modeling of remifentanil-propofol interaction on cardiorespiratory control and bispectral index. *Anesthesiology* 2003; 98:312–322

33. M J Mertens, E Olofsen, F H M Engbers, A G L Burm, J G Bovill, J Vuyk. Propofol reduces perioperative remifentanil requirements in a synergistic manner: response surface modeling of perioperative remifentanil-propofol interactions. *Anesthesiology* 2003; 99:347–359

34. R Romberg, E Olofsen, E Sarton, L Teppema, A Dahan. Pharmacodynamic effect of morphine-6-glucuronide versus morphine on hypoxic and hypercapnic breathing in healthy volunteers. *Anesthesiology* 2003; 99:788–798

35. R Romberg, E Olofsen, E Sarton, J DenHartigh, P Taschner, A Dahan. Pharmacokinetic-pharmacodynamic modeling of morphine-6- glucoronide–induced analgesia in healthy volunteers: absence of sex differences. *Anesthesiology* 2004; 100:120–133

36. H P A Van Dongen, E Olofsen, D F Dinges, G Maislin. Mixed-model regression analysis and dealing with interindividual differences. Numerical Computer Methods, Part E, Edited by M Johnson, L Brand, volume 384 of *Methods in Enzymology*. Elsevier, Amsterdam, 2004; pp 139–171

37. E Olofsen, D F Dinges, H P A Van Dongen. Nonlinear mixed-effects modeling: individualization and prediction. *Aviat Space Environ Med* 2004; 75:A134–140

38. M J Mertens, E Olofsen, A G L Burm, J G Bovill, J Vuyk. Mixed-effects modeling of the influence of alfentanil on propofol pharmacokinetics. *Anesthesiology* 2004; 100:795–805

39. R Romberg, E Olofsen, E Sarton, L Teppema, A Dahan. Increased margin of safety of morphine-6-glucuronide relative to morphine. *Anesthesiology* 2004; 100:1622

40. A Dahan, R Romberg, L Teppema, E Sarton, H Bijl, E Olofsen. Simultaneous measurement and integrated analysis of analgesia and respiration after an intravenous morphine infusion. *Anesthesiology* 2004; 101:1201–1209

41. E Olofsen, A Dahan. Population pharmacokinetics/pharmacodynamics of anesthetics. *AAPS Journal* 2005; 7. Article 39 (http://www.aapsj.org)

42. R R Romberg, E Olofsen, H Bijl, P E M Taschner, L J Teppema, E Y Sarton, J W van Kleef, A Dahan. Polymorphism of mu-opioid receptor gene (OPRM1:c.118A > G) does not protect against opioid-induced respiratory depression despite reduced analgesic response. *Anesthesiology* 2005; 102:522–530

43. A Dahan, A Yassen, H Bijl, R Romberg, E Sarton, L Teppema, E Olofsen, M Danhof. Comparison of the respiratory effects of intravenous buprenorphine and fentanyl in humans and rats. *Br J Anaesth* 2005; 94:825–834

44. A Yassen, E Olofsen, A Dahan, M Danhof. Pharmacokinetic-pharmacodynamic modeling of the antinociceptive effect of buprenorphine and fentanyl in rats: role of receptor equilibration kinetics. *J Pharm Exp Ther* 2005; 313:1136–1149

45. E Olofsen, R Romberg, H Bijl, R Mooren, F Engbers, B Kest, A Dahan. Alfentanil and placebo analgesia - no sex differences detected in models of experimental pain. *Anesthesiology* 2005; 103:130–139

46. A Dahan, A Yassen, R Romberg, E Sarton, L Teppema, E Olofsen, M Danhof. Buprenorphine induces ceiling in respiratory depression but not in analgesia. *Br J Anaesth* 2006; 96:627–632

47. A Yassen, E Olofsen, R Romberg, E Sarton, M Danhof, A Dahan. Mechanism-based pharmacokinetic-pharmacodynamic modeling of the antinociceptive effect of buprenorphine in healthy volunteers. *Anesthesiology* 2006; 104:1232–1242

48. E van Dorp, A Yassen, E Sarton, R Romberg, E Olofsen, L Teppema, M Danhof, A Dahan. Naloxone reversal of buprenorphine-induced respiratory depression. *Anesthesiology* 2006; 105:51–57

49. A Yassen, J Kan, E Olofsen, E Suidgeest, A Dahan, M Danhof. Mechanism-based pharmacokinetic-pharmacodynamic modeling of the respiratory-depressant effect of buprenorphine and fentanyl in rats. *J Pharm Exp Ther* 2006; 319:682–692

50. A Yassen, E Olofsen, E van Dorp, E Sarton, L Teppema, M Danhof, A Dahan. Mechanism-based pharmacokinetic-pharmacodynamic modelling of the reversal of buprenorphine-induced respiratory depression by naloxone - a study in healthy volunteers. *Clin Pharmacokin* 2007; 46:965–980

51. A Yassen, J Kan, E Olofsen, E Suidgeest, A Dahan, M Danhof. Pharmacokinetic-pharmacodynamic modeling of the respiratory depressant effect of norbuprenorphine in rats. *J Pharm Exp Ther* 2007; 321:598–607

52. A Yassen, E Olofsen, J Kan, A Dahan, M Danhof. Pharmacokinetic-pharmacodynamic modeling of the effectiveness and safety of buprenorphine and fentanyl in rats. *Pharm Res* 2008; 25:183–193

53. E Olofsen, A G L Burm, M J G Simon, B Th Veering, J W van Kleef, A Dahan. Population pharmacokinetic-pharmacoadynamic modeling of epidural anesthesia. *Anesthesiology* 2008; 109:664–674

54. E Olofsen, J W Sleigh, A Dahan. Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect. *Br J Anaesth* 2008; 101:810–821

55. M Sigtermans, A Dahan, R Mooren, M Bauer, B Kest, E Sarton, E Olofsen. S(+)-ketamine effect on experimental pain and cardiac output a population pharmacokinetic-pharmacodynamic modeling study in healthy volunteers. *Anesthesiology* 2009; 111:892–903

56. E Olofsen, H P A Van Dongen, C G Mott, T J Balkin, D Terman. Current approaches and challenges to development of an individualized sleep and performance prediction model. *The Open Sleep Journal* 2010; 3:24–43

57. M Sigtermans, I Noppers, E Sarton, M Bauer, R Mooren, E Olofsen, A Dahan. An observational study on the effect of S(+)-ketamine on chronic pain versus experimental acute pain in complex regional pain syndrome type 1 patients. *Eur J Pain* 2010; 14:302–307

58. B J Lichtenbelt, E Olofsen, A Dahan, J W van Kleef, M M R F Struys, J Vuyk. Propofol reduces the distribution and clearance of midazolam. *Anesth Analg* 2010; 110:1597–1606

59. E Olofsen, M Boom, D Nieuwenhuijs, E Sarton, L Teppema, L Aarts, A Dahan. Modeling the non-steady state respiratory effects of remifentanil in awake and propofol-sedated healthy volunteers. *Anesthesiology* 2010; 112:1382–1395

60. E Olofsen, E van Dorp, L Teppema, L Aarts, T W Smith, A Dahan, E Sarton. Naloxone reversal of morphine- and morphine-6-glucuronide-induced respiratory depression in healthy volunteers a mechanism-based pharmacokinetic-pharmacodynamic modeling study. *Anesthesiology* 2010; 112:1417–1427

61. E Olofsen, R Mooren, E van Dorp, L Aarts, T Smith, J den Hartigh, A Dahan, E Sarton. Arterial and venous pharmacokinetics of morphine-6-glucuronide and impact of sample site on pharmacodynamic parameter estimates. *Anesth Analg* 2010; 111:626–632

62. A Dahan, E Olofsen, M Sigtermans, I Noppers, M Niesters, L Aarts, M Bauer, E Sarton. Population pharmacokinetic-pharmacodynamic modeling of ketamine-induced pain relief of chronic pain. *Eur J Pain* 2011; 15:258–267

63. I Noppers, E Olofsen, M Niesters, L Aarts, R Mooren, A Dahan, E Kharasch, E Sarton. Effect of rifampicin on S-ketamine and S-norketamine plasma concentrations in healthy volunteers after intravenous S-ketamine administration. *Anesthesiology* 2011; 114:1435–1445

64. E Olofsen, I Noppers, M Niesters, E Kharasch, L Aarts, E Sarton, A Dahan. Estimation of the contribution of norketamine to ketamine-induced acute pain relief and neurocognitive impairment in healthy volunteers. *Anesthesiology* 2012; 117:353–364

65. E Olofsen, M Sigtermans, I Noppers, M Niesters, R Mooren, M Bauer, L Aarts, E Sarton, A Dahan. The dose-dependent effect of S(+)-ketamine on cardiac output in healthy volunteers and complex regional pain syndrome type 1 chronic pain patients. *Anesth Analg* 2012; 115:536–546

66. A E Olesen, E Olofsen, S S Olesen, C Staahl, T Andresen, A Dahan, A M Drewes. The absorption profile of pregabalin in chronic pancreatitis. *Basic Clin Pharmacol Toxicol* 2012; 111:385–390

67. M Niesters, R Mahajan, E Olofsen, M Boom, S Garcia del Valle, L Aarts, A Dahan. Validation of a novel respiratory rate monitor based on exhaled humidity. *Br J Anaesth* 2012; 109:981–989

68. M Boom, E Olofsen, M Neukirchen, R Fussen, J Hay, G J Groeneveld, L Aarts, E Sarton, A Dahan. Fentanyl utility function: a risk-benefit composite of pain relief and breathing responses. *Anesthesiology* 2013; 119:663–674

69. C H Martini, A Yassen, A Krebs-Brown, P Passier, M Stoker, E Olofsen, A Dahan. A novel approach to identify responder subgroups and predictors of response to low- and high-dose capsaicin patches in postherpetic neuralgia. *Eur J Pain* 2013; 17:1491–1501

70. A Dahan, M Niesters, E Olofsen, T Smith, F Overdyk. Opiods. Clinical Anesthesia, Edited by P G Barash, B F Cullen, R K Stoelting, M K Cahalan, R Ortega, M C Stock. Lippincott Williams & Wilkins, Philadelphia, 2013; pp 501–522

71. E Olofsen, A Dahan, G Borsboom, G Drummond. Improvements in the application and reporting of advanced Bland-Altman methods of comparison. *J Clin Monit Comput* 2015; 29:127–139

72. E Olofsen, A Dahan. Big brain, small world? *Anesthesiology* 2015; 122:8–11

73. C H Martini, P Proto, E Olofsen, M van Velzen, L Aarts, A Dahan, M Niesters. A randomized controlled trial and novel mathematical analysis of the analgesic effect of oxycodone versus paracetamol orodispersible tablets. *Eur J Pain* 2015; 19:295–304

74. A Dahan, E Olofsen, M Niesters. Pharmacotherapy for pain: efficacy and safety issues examined by subgroup analyses. *Pain* 2015; 156:S119–S126

75. M Roozekrans, E Olofsen, R van der Schrier, J van Gerven, S Peng, J McLeod, A Dahan. Reversal of opioid-induced respiratory depression by BK-channel blocker

GAL021: a pharmacokinetic-pharmacodynamic modeling study in healthy volunteers. *Clin Pharmacol Ther* 2015; 97:641–649

76. N Khalili-Mahani, C H Martini, E Olofsen, A Dahan, M Niesters. Effect of subanaesthetic ketamine on plasma and saliva cortisol secretion. *Br J Anaesth* 2015; 115:68–75

77. A E Olesen, E Olofsen, T Andresen, C Graversen, A M Drewes, A Dahan. Stochastic pharmacokinetic-pharmacodynamic analysis of the effect of transdermal buprenorphine on electroencephalogram and analgesia. *Anesth Analg* 2015; 121:1165–1175

78. E Olofsen, A Dahan. Using Akaike's information theoretic criterion in mixed-effects modeling of pharmacokinetic data: a simulation study. *F1000Research* 2015; 2:71

79. L Oudejans, M Van Velzen, E Olofsen, R Beun, A Dahan, M Niesters. Translation of random painful stimuli into numerical responses in fibromyalgia and perioperative patients. *Pain* 2016; 157:128–136

80. D J A R Moes, S A S van der Bent, J J Swen, T van der Straaten, A Inderson, E Olofsen, H W Verspaget, H J Guchelaar, J den Hartigh, B van Hoek. Population pharmacokinetics and pharmacogenetics of once daily tacrolimus formulation in stable liver transplant recipients. *Eur J Clin Pharmacol* 2016; 72:163–174

81. E Sitsen, E Olofsen, A Lesman, A Dahan, J Vuyk. Epidural blockade affects the pharmacokinetics of propofol in surgical patients. *Anesth Analg* 2016; 122:1341–1349

82. A Dahan, M Douma, E Olofsen, M Niesters. High inspired oxygen concentration increases the speed of onset of remifentanil-induced respiratory depression. *Br J Anaesth* 2016; 116:878–879

83. B Torensma, C H Martini, M Boon, E Olofsen, B in 't Veld, R S L Liem, M T T Knook, D J Swank, A Dahan. Deep neuromuscular block improves surgical conditions during bariatric surgery and reduces postoperative pain: a randomized double blind controlled trial. *PLOS ONE* 2016; 11