UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy 83

Methodological Studies on Models and Methods for Mixed-Effects Categorical Data Analysis

MARIA C. KJELLSSON

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2008

ISSN 1651-6192
ISBN 978-91-554-7316-7
urn:nbn:se:uu:diva-9333

Dissertation presented at Uppsala University to be publicly examined in B42, BMC, Husargatan 3, Uppsala, Friday, November 21, 2008 at 13:15 for the degree of Doctor of Philosophy (Faculty of Pharmacy). The examination will be conducted in English.

Abstract

Kjellsson, M C. 2008. Methodological Studies on Models and Methods for Mixed-Effects Categorical Data Analysis. Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy* 83. 73 pp. Uppsala.
ISBN 978-91-554-7316-7.

Effects of drugs are in clinical trials often measured on categorical scales. These measurements are increasingly being analyzed using mixed-effects logistic regression. However, the experience with such analyzes is limited and only a few models are used.

The aim of this thesis was to investigate the performance and improve the use of models and methods for mixed-effects categorical data analysis. The Laplacian method was shown to produce biased parameter estimates if (i) the data variability is large or (ii) the distribution of the responses is skewed. Two solutions are suggested; the Gaussian quadrature method and the back-step method. Two assumptions made with the proportional odds model have also been investigated. The assumption with proportional odds for all categories was shown to be unsuitable for analysis of data arising from a ranking scale of effects with several underlying causes. An alternative model, the differential odds model, was developed and shown to be an improvement, in regard to statistical significance as well as predictive performance, over the proportional odds model for such data. The appropriateness of the likelihood ratio test was investigated for an analysis where dependence between observations is ignored, i.e. performing the analysis using the proportional odds model. The type I error was found to be affected; thus assessing the actual critical value is prudent in order to verify the statistical significance level. An alternative approach is to use a Markov model, in which dependence between observations is incorporated. In the case of polychotomous data such model may involve considerable complexity and thus, a strategy for the reduction of the time-consuming model building with the Markov model and sleep data is presented.

This thesis will hopefully contribute to a more confident use of models for categorical data analysis within the area of pharmacokinetic and pharmacodynamic modelling in the future.

Keywords: Pharmacodynamics, Categorical data, Markov model, Modelling, NONMEM, NLMIXED, Laplace, Gaussian quadrature, Back-Step Method, Proportional odds model, Differential odds model

Maria C. Kjellsson, Department of Pharmaceutical Biosciences, Division of Pharmacokinetics and Drug Therapy, Box 591, Uppsala University, SE-75124 Uppsala, Sweden

ISSN 1651-6192
ISBN 978-91-554-7316-7
urn:nbn:se:uu:diva-9333 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-9333)

Traue keiner Statistik, die du nicht selber gefälscht hast
The only statistics you can trust are those you falsified yourself

German proverb

Papers discussed

This thesis is based on the following papers, which will be referred to by their Roman numerals as assigned below.

I Siv Jönsson[*], **Maria C. Kjellsson**[*], Mats O. Karlsson.
Estimating bias in population parameters for some models for repeated measures ordinal data using NONMEM and NLMIXED.
J Pharmacokinet Pharmacodyn. 31: 299-320 (2004).

II **Maria C. Kjellsson**, Siv Jönsson, Mats O. Karlsson.
The Back-Step Method – Method for obtaining unbiased population parameter estimates for ordered categorical data.
AAPS J. 6: Article 19 (2004).

III **Maria C. Kjellsson**, Per-Henrik Zingmark, E. Niclas Jonsson, Mats O. Karlsson.
Comparison of proportional odds and differential odds models for mixed-effects analysis of categorical data.
J Pharmacokinet Pharmacodyn. Sept 23 (2008)

IV Hanna E. Silber, **Maria C. Kjellsson**, Mats O. Karlsson.
The impact of model misspecifications on the type I error rate for covariate inclusion; an investigation of continuous and categorical data.
Submitted

V **Maria C. Kjellsson**, Daniele Ouellet, Brian Corrigan, Mats O. Karlsson.
Modeling sleep data for a new drug in development using Markov mixed-effects models.
In manuscript

[*] Siv Jönsson and Maria C. Kjellsson have contributed equally to this work.

Contents

Abbreviations and symbols

AUC	Area under the plasma concentration time curve
BSM	Back-step method
C	Concentration
CL	Clearance
df	Degree of freedom
EBE	Empirical Bayes estimate
FO	First-order
FOCE	First-order conditional estimation
GQ	Gaussian quadrature
IIV	Inter-individual variability
IOV	Inter-occasion variability
IS	Initial sleeplessness
LPS	Latency to persistent sleep
LRT	Likelihood ratio test
MS	Multiple sclerosis
NAW	Number of awakenings
OFV	Objective function value
P	A parameter in a mixed-effects model
$\boldsymbol{P}$	A vector of several parameters, P
PD	Pharmacodynamic
PK	Pharmacokinetic
PSG	Polysomnography
PPC	Posterior predictive check
REE	Relative estimation error
SE	Standard error
SEF	Sleep efficiency
SHT	Simulation hypothesis test
TP	Transition probability
VPC	Visual predictive check
x_i	A covariate for the i^{th} individual
X_i	Vector with all covariates for the i^{th} individual
α	Intercept parameter
β	Differential effect parameter
ΔOFV	Difference in OFV

ε	Difference between predicted and observed value
η	Difference between population and individual parameter
ϕ	A parameter: θ or ω^2
Φ	Vector of parameters, ϕ
θ	Typical value, parameter in a mixed-effects model
κ	Difference between individual parameters for different occasions
π^2	Variance of κ's, parameter in a mixed-effects model
σ^2	Variance of ε's, parameter in a mixed-effects model
ω^2	Variance of η's, parameter in a mixed-effects model

Introduction

Data are traditionally defined as categorical or continuous, with the subclasses nominal and ordinal for categorical data and interval and ratio for continuous data[1]. Categorical data consist of variables with a finite number of values. The nominal variables are unordered, for example different cancer diagnosis, while ordinal variables have a natural order even though the exact spacing between the levels is unknown, for example pain relief. Binary variables, such as responder/non-responder, are a special case of categorical data with only two categories. Continuous data consist of variables with an infinite number of values where ratio data have an identifiable absolute zero point while interval data do not. Interval data can also be viewed as ordinal data where the spacing between the categories is known. If a continuous measurement is categorized, the variable is in between ordinal and interval data as the exact levels are known. These data are nevertheless analyzed as ordinal as the level of the categories in themselves usually do not reveal the spacing, that is level two is not necessary twice as high as level one.

An important aspect of clinical trials is to determine the pharmacokinetics (PK) of a drug, describing the time course of the systemic exposure following a given dose[2]. Even more essential is to investigate whether a drug has an effect and how this effect varies with the systemic exposure; that is the pharmacodynamics (PD) of the drug. Categorical data are commonly measured repeatedly within the same patient in clinical trials describing both disease symptoms as well as effects of the drug.

Traditionally, when analyzing data from a confirmatory clinical study the observations are analyzed by performing statistical testing using, for example t-test or Wilcoxon signed-rank test, comparing the measurements before and after drug administration at a certain time point. Another approach is to perform statistical testing on the measurements for the active and the placebo arm in the study. This type of analysis, even though commonly used, answers only a narrow question. Non-linear mixed-effects analysis, the method used in PK-PD modelling, is more appealing from a learning perspective[3] as it provides the opportunity to explore the time course of the effect using all measurements obtained at different times within the same individual.

Population PK-PD modelling

When model based analysis was first introduced within PK-PD analysis, the standard two stage approach was used to obtain summary measurements of all individuals, that is the population estimates. In the standard two stage approach each individual's PK and PD parameters are estimated separately, followed by assessment of the variability in the data by calculating statistical summary measurements based on the individual parameters[4-6]. A limitation with this method is the extensive sampling needed from each individual to be able to estimate the individual parameters and to obtain unbiased population parameters. Additionally, not always being able to characterise data from all subjects using the same structural models makes calculating the summary measurements complicated.

Another approach is mixed-effects modelling, where the data from all individuals are used and population parameters, that is mean and variance of the individual parameter distributions are estimated simultaneously. This approach allows the individuals to share information about the parameters, making estimation of parameters in sparse sampling designs possible. Model-based mixed-effects analysis has been used within the area of PK and PD since the 1980's[4].

Mixed-effects modelling

There are two types of parameters to be estimated in a mixed-effects model: the fixed effects and the variances of the random effects. The fixed effects are used to describe the overall trend in the data and these parameters are shared by all individuals in the population. The fixed effects in PK-PD models describe the basic PK and PD characteristics of a drug as well as the disease progression. An example of a PK characteristic is clearance (CL). Variability in the data originates from for example the variability between individuals, bioanalytical imprecision, sampling techniques and error in recording of dosing and sampling. To account for this at least two levels of variability are defined in the model: the variability between the individual prediction and observation (the residual error) and the variability between individuals (the inter-individual variability, IIV). These variabilities are implemented using random effects for which the variance is the estimated parameter. The IIV originates from differences between individuals, while the residual error originates from for example errors in dose or sampling times, bioanalytical error and model misspecification. Individual parameters can sometimes change between occasions, either randomly or because an underlying unknown physiological relationship is not accounted for in the model. To separate this intra-individual variability from the IIV, a third level of variability can be added to the statistical model, estimating the inter-occasion

variability (IOV)[7]. Part of the IIV, IOV and residual variability may be dependent on known observable factors of patient characteristics, also known as covariates and examples of such are concomitant medication or bioanalytical analysis method. The covariate model expresses the relationship between fixed effect parameters describing the overall trend in the data and different measured covariates. An example of a covariate model for renally cleared drugs is how creatinine clearance affects CL.

The individual parameter P_{ik} for the i^{th} individual at the k^{th} occasion can be described by:

$$P_{ik} = \theta \cdot e^{\eta_i + \kappa_{ik}},\ \eta_i \sim N\left(0, \omega^2\right) \text{ and } \kappa_i \sim N\left(0, \pi^2\right) \quad (1)$$

where θ is the fixed effect describing the mean (or typical) value of the parameter P in the population, η_i and κ_{ik} are the random effects with zero mean and variance, ω^2 and π^2, respectively. These random effects describe the IIV and IOV and these are implemented as variability in the fixed effects parameters. The random effects are normally distributed, but are usually included in the model to yield a lognormal distribution of the parameter P, as this is more frequently observed in biological systems. Any transformation of the normal distribution is possible, if this is biologically or physiologically relevant. If covariates are used to describe the individual parameter, this can be expressed as follows:

$$P_{ik} = \theta \cdot \left(1 + \theta_x \cdot \left(x_i - x_{mean}\right)\right) \cdot e^{\eta_i + \kappa_{ik}} \quad (2)$$

in which θ is the value of the parameter P for the typical individual; that is individuals with an observed covariate value, x_i equal to the mean covariate value in the population, x_{mean}, and θ_x is the slope of the change of parameter P.

If the model describing the trend in the data is $f(...)$, $\boldsymbol{P}_{ik}$ the vector of individual parameters described using fixed and random effects and X_i the vector of individual covariate values for all covariates included in the model, then the residual error can be implemented as shown in equation (3).

$$y_{ijk} = f\left(X_i, \boldsymbol{P}_{ik}\right) + \varepsilon_{ijk},\ \varepsilon_{ijk} \sim N\left(0, \sigma^2\right) \quad (3)$$

y_{ijk} is the observation and ε_{ij} is a normally distributed random variable with the variance σ^2, describing the residual error. The model describing the residual error may have many shapes, but the additive, as shown in equation (3), the proportional and a combination of the two are the most common

models. Other shapes of the residual error have been presented taking correlation between observations into account[8, 9]. Even though not the focus of this thesis the effect of ignoring some more complex shapes of the residual error was explored in paper III on a commonly used statistical test within PK-PD modelling.

Modelling categorical data

There are a number of options for how to analyze polychotomous data, all of which have both advantages and disadvantages. To illustrate, consider pain with levels no (0), mild (1), moderate (2) and severe pain (3). The data could be treated as continuous data, by transforming categories into numerical values, which facilitates the analysis, as there are many models available for continuous data. However, this approach is not feasible with few categories and seldom is this approach used for data with fewer than six categories. In addition, even though the observations only are allowed to take certain values, the predictions may be all values. Predicting a pain level of 1.25, should that be interpreted as mild or moderate pain? Another method is to dichotomise the data, which was a common approach in epidemiology about one to two decades ago[10-12]. In our example, the two categories would be pain or no pain, and these are analyzed using binary logistic regression. This gives advantages over the polychotomous data as the choice of model not is affected by the nature of the binary data; that is the same model can be used independent of if the data is ordinal or nominal. Additionally, the models for binary data are generally easier to explain to people not accustomed to models for categorical data. The information collected about the severity of the pain is nevertheless completely discarded with an increased risk of getting biased effect estimates[11] or not finding all relevant covariate relationships[12]. The last and preferred option is to use logistic regression keeping all categories as they were reported and this was the chosen method in this thesis. A disadvantage is however that the experience with this type of analysis is limited and that only a few models have been presented within mixed-effect modelling. Nevertheless, the latter is preferred as the information in the data is neither overestimated, as in the case of treating as continuous data, nor discarded as in the case of treating as dichotomous.

There are two major differences between models for continuous and categorical data analysis.

- The likelihood or probability of a certain score, rather than the values of the score itself, is used for the estimation of the parameters.
- Consequently, no residual error is defined in the model.

Link functions

Different link functions[13] are used to ensure that the probability of an event ranges from zero to one, while the estimated parameter ranges from –infinity to infinity. The logit or "log of odds" transformation is by far the most used transformation of categorical data, possibly for its mathematical tractability. The cumulative distribution of the responses is assumed to be logistically distributed and it is implemented as described in equation (4).

$$\begin{aligned} & logit\left(p\right) = \ln\frac{p}{1-p} = f\left(\boldsymbol{P}, X\right) \\ \Leftrightarrow\ & p = \frac{e^{f(\boldsymbol{P},X)}}{1+e^{f(\boldsymbol{P},X)}} \end{aligned} \tag{4}$$

where $f(\boldsymbol{P},X)$ is the function with parameters $\boldsymbol{P}$ describing how the probability varies with baseline conditions and covariates, X.

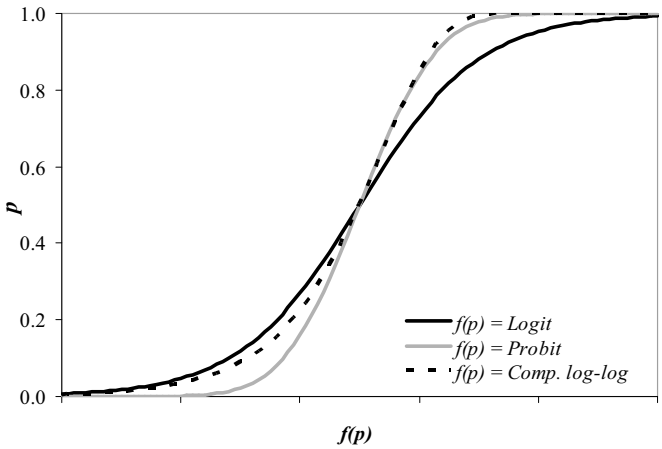

Figure 1. The cumulative probability distributions obtained with different link functions used for transforming categorical data; the logit, the probit and the complementary log-log. The values of $f(p)$ were chosen such that the curves for each transformation overlap at p=0.5. This was done to emphasize the differences of the distributions.

Other transformations are also possible, for example the probit transformation. This transformation will assume that the cumulative distribution of the observations is the inverse of the cumulative normal distribution, see equation (5).

$$probit\left(p\right)=\Phi^{-1}\left(p\right)=f\left(\boldsymbol{P},X\right)$$
$$\Leftrightarrow p=\Phi\left(f\left(\boldsymbol{P},X\right)\right) \quad (5)$$

Another quite common transformation is the complementary log-log, which assumes that the cumulative distribution is the Gumbel distribution, as given below:

$$complementary\ log\text{-}log\left(p\right)=\ln\left(\ln\left(1-p\right)\right)=f\left(\boldsymbol{P},X\right)$$
$$\Leftrightarrow p=1-e^{-e^{f\left(\boldsymbol{P},X\right)}} \quad (6)$$

The different transformations will give rise to slightly different cumulative probability distributions, schematically depicted in Figure 1. Notable is that both the logit and the probit transformations are symmetric while the complementary log-log is not. The largest differences are also seen at the tails of the distributions.

The proportional odds model

Lewis Sheiner presented in 1994[14] the proportional odds model for mixed-effects analysis of PK-PD categorical data using pain relief scores as an example. A ranking scale is used when measuring pain relief and ranking scales are the most commonly used categorical scale in drug development. It is used for measuring adverse events in almost all clinical trials performed and it assesses the side-effects as mild, moderate or severe. A large number of analyses have been published where a wide variety of ranking scales have been analyzed, for example dry mouth severity[15], nicotine cravings[16], severity of allergic rhinitis[17], severity of adverse bleeding events[18], grade of neurophenia[19, 20], grades of diarrhoea[21], receptor density[22], change in thrombus size[23], severity of lesions in rats[24], sedation scales[25-28] and pain relief[14, 29, 30]. The proportional odds model is used for the analysis in all of these publications.

If $y_i = [y_{i1}, y_{i2}, y_{i3}, ..., y_{iN}]$ is the vector of ordinal responses for the i^{th} individual with N observations, then the probability that the observation y_{in} is greater than or equal to a score m, of the possible categories $1 \ldots I$, has the following general structure:

$$P\left(y_{in}\geq m\,|\,\eta_i\right)=p_{mi} \quad (7)$$

$$logit\left(p_{mi}\right)=f_m+\eta_i,\ \eta_i\sim N\left(0,\omega^2\right) \quad (8)$$

$$p_{mi} = \frac{e^{f_m+\eta_i}}{1+e^{f_m+\eta_i}} \tag{9}$$

where p_{im} is the individual probability of observing a score $\geq m$. The f_m is a function of baseline conditions and effects of covariates $g(X_i)$, shown in equation (10).

$$\begin{aligned} f_2 &= \alpha_2 + g\left(X_i\right) \\ f_3 &= \alpha_2 + \alpha_3 + g\left(X_i\right) \\ &\ldots \\ f_m &= \alpha_2 + \sum_{k=3}^{m} \alpha_k + g\left(X_i\right) \end{aligned} \tag{10}$$

in which α are the intercept parameters, where α_2 specify the baseline probability of observing score ≥ 2, and all other α's specify the decrease in probability due to observing higher scores and these are bound to (-infinity, 0]. $g(X_i)$ denotes the function describing the effect of the covariate, for example placebo and/or drug effects. The model is commonly referred to as the cumulative odds model with random intercept if it is defined without any covariate relationships.

The corresponding probabilities for each of the scores are shown in equation (11). The intercept parameter for the lowest score is not estimated as the probability for this score is derived from the probability of all scores, which by definition is one.

$$\begin{aligned} P\left(y_{in}=1 \mid \eta_i\right) &= 1 - P\left(y_{in} \geq 2 \mid \eta_i\right) \\ P\left(y_{in}=2 \mid \eta_i\right) &= P\left(y_{in} \geq 2 \mid \eta_i\right) - P\left(y_{in} \geq 3 \mid \eta_i\right) \\ &\ldots \\ P\left(y_{in}=I \mid \eta_i\right) &= P\left(y_{in} \geq I \mid \eta_i\right) \end{aligned} \tag{11}$$

Binary, or dichotomous, data are most commonly analyzed using a model that can be viewed as the proportional odds model with two categories, given in equation (12).

$$\begin{aligned} &logit\left(p_{1i}\right) = \alpha + g\left(X_i\right) + \eta_i \Leftrightarrow p_{1i} = \frac{e^{\alpha+g(X_i)+\eta_i}}{1+e^{\alpha+g(X_i)+\eta_i}} \\ &p_{0i} = 1 - p_{1i} \end{aligned} \tag{12}$$

with p_{1i} and p_{0i} being the individual probability of observing score = 1 and 0, respectively, α the intercept parameter describing the baseline probability of observing score = 1 and $g(X_i)$ the function describing covariate relationships such as placebo and drug effects.

Several assumptions are made with the proportional odds model. Two are of interest for this thesis; *i)* all observations are treated as they are independent of each other and *ii)* furthermore, as the proportional odds model was developed for categorized continuous data, the effect of the covariate is assumed to be the same on all log odds of all categories.

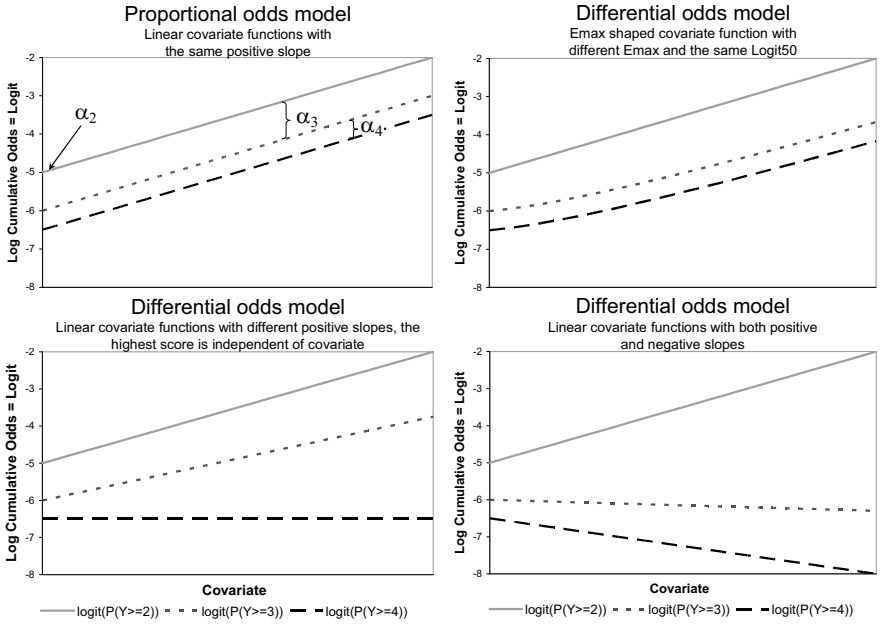

Figure 2. Illustration of the effect on the log odds for different covariate functions. The logits for a categorical scale with four scores are plotted versus the covariate. Examples are given for covariate functions being all equal (upper left), Emax shaped with different Emax values (upper right), linear with positive slopes (lower left) and linear with both positive and negative slopes (lower right).

The differential odds model

The differential odds model was developed as an alternative for analysis of categorical data. This model relaxes the assumption of the proportional odds model that the covariate effects are the same for all log odds of all categories by allowing different functions of the covariates for the different cumulative scores, as seen below (compare with equation (10) for the proportional odds model).

$$
\begin{aligned}
f_2 &= \alpha_2 + g_2\left(X_i\right) \\
f_3 &= \alpha_2 + \alpha_3 + g_3\left(X_i\right) \\
&\ldots \\
f_m &= \alpha_2 + \sum_{k=3}^{m} \alpha_k + g_m\left(X_i\right)
\end{aligned}
\tag{13}
$$

It is important to ensure that $g_2(X_i) \geq g_3(X_i) \geq \ldots \geq g_l(X_i)$ as the model is implemented using cumulative probabilities. Neglecting to do so, will render impossible probabilities, such as the probability of score ≥ 3 being higher than the probability of score ≥ 2. Having different functions describing the $g_j(X_i)$ will allow the covariate effect to be different for different scores; thus this relaxes the assumption of the proportional odds model with parallel slopes on the log odds as is shown in Figure 2.

The proportional odds model is a limiting case of the differential odds model, as the differential odds model collapses into the proportional odds model if all covariate functions are equal. The shape of the $g_j(X_i)$ may vary as long as $g_2(X_i) \geq g_3(X_i) \geq \ldots \geq g_l(X_i)$. Thus, it is possible to have some of the covariate functions being increasing functions, while others are decreasing or even independent of the covariate. Making the function for the covariate effect vary with several covariates is also possible as well as having different models added together. To exemplify the latter, consider the following equation (14) where the covariate function is a step model for scores ≥ 2, a step, and an Emax shaped model for scores ≥ 3 and independent of the covariate for the highest score (score = 4).

$$
\begin{aligned}
g_2\left(X_i\right) &= \left(I_{0/1} \cdot \theta_{step}\right) \\
g_3\left(X_i\right) &= \left(I_{0/1} \cdot \theta_{step}\right) \cdot \left(\theta_{baseline} + \frac{\left(1-\theta_{baseline}\right) \cdot X_i}{X_i + \theta_{Logit50}}\right) \\
g_4\left(X_i\right) &= 0
\end{aligned}
\tag{14}
$$

However, starting the implementation using simple models of $g_j(X_i)$ all acting in the same direction should be considered the primary implementation, in order not to risk over-parameterising the model.

The Markov model

An observation is assumed independent of any other observations with the proportional and differential odds models, thus the observations are only

dependent on the independent variables and the model. If the categories are viewed as states these can be described using the Markov model in which an observations is dependent on the preceding observation as well as the independent variables and the model. The Markov assumption, which is the basis for the Markov model, states that future states depends only on the present state and not any of the past states; thus the transition to a category is conditionally independent of past observations. The Markov model, also called the transition model, is implemented using the transition probabilities between states and these vary depending on between which states the transition takes place. Strictly speaking, this model is a first order Markov model, as the probability of a transition is conditioned only on the value of the state that directly preceded it. The Markov model can also be of higher orders, for example second order with the transition probability influenced by the two preceding states. Markov models of higher order can also be implemented by using covariates including a time aspect. An example of this is given in this thesis for sleep data, where the transition probability is modelled as dependent on the previous observation as well as the time elapsed since last transition.

If $y_{ij} = (y_{ij1}, y_{ij2\ldots} y_{ijN})$ is the vector of observations for the i^{th} individual at the j^{th} occasion with N observations then the probability that the observation at time = n, y_{ijn} is equal to score = m, given that the preceding observation of score = l ($m \neq l$) has the structure given in equations (15), (16), (17) and (18). This probability can also be referred to as the transition probability (TP) from score l to score m.

$$P\left(y_{ijn} = m \mid y_{ijn-1} = l, \eta_i, \kappa_{ij}\right) = p_{ijm|l} \tag{15}$$

$$logit\left(p_{ijm|l}\right) = f_{m|l} + \eta_i + \kappa_{ij}, \quad \eta_i \sim N\left(0, \omega^2\right) \text{and } \kappa_{ij} \sim N\left(0, \pi^2\right) \tag{16}$$

$$p_{ijm|l} = \frac{e^{f_{m|l}+\eta_i+\kappa_{ij}}}{1+e^{f_{m|l}+\eta_i+\kappa_{ij}}} \tag{17}$$

$$f_{m|l} = \alpha_{m|l} + g\left(X_i\right) \tag{18}$$

$p_{ijm|l}$ is the individual transition probability from score = m to score = l. If there is only one measurement per occasion, no IOV can be estimated and thus no κ_{ij} is included in the model. The $f_{m|l}$ is a function of baseline conditions and effects of covariates and for the Markov model, an intercept parameter, $\alpha_{m|l}$ describing the probability of transiting and a function of covariates, $g(X_i)$ are defined for each transition. Thus, the probability of not transit-

ing is one minus the sum of all probabilities of transition from that particular state, see equation (19).

$$P\left(y_{ijn} = m \mid y_{ijn-1} = m\right) = 1 - \sum P\left(y_{ijn} \neq m \mid y_{ijn-1} = m\right) \tag{19}$$

With binary data, the implementation of the Markov model is as described in equation (20).

$$\begin{aligned} &logit\left(p_{1|0i}\right) = \alpha_{1|0} + g\left(X_i\right) + \eta_i \Leftrightarrow p_{1|0i} = \frac{e^{\alpha_{1|0} + g(X_i) + \eta_i}}{1 + e^{\alpha_{1|0} + g(X_i) + \eta_i}} \\ &logit\left(p_{0|1i}\right) = \alpha_{0|1} + g\left(X_i\right) + \eta_i \Leftrightarrow p_{0|1i} = \frac{e^{\alpha_{0|1} + g(X_i) + \eta_i}}{1 + e^{\alpha_{0|1} + g(X_i) + \eta_i}} \\ &p_{1|1i} = 1 - p_{0|1i} \\ &p_{0|0i} = 1 - p_{1|0i} \end{aligned} \tag{20}$$

in which $p_{1|0i}$ and $p_{0|1i}$ is the individual probability of a transition with previous observations of score = 0 and 1, respectively, while $p_{0|0i}$ and $p_{1|1i}$ is the individual probability of no transition. $g(X_i)$ is the function describing covariate relationships such as placebo and drug effects and $\alpha_{1|0}$ and $\alpha_{1|1}$ describe the baseline transition probability from one to zero and from zero to one, respectively.

Estimation of parameters

There is a wide variety of software available for regression of mixed-effects models[31], all of which iteratively finds the parameters of the model that maximize the likelihood of the data. The first software presented for regression of mixed-effects models within the area of PK-PD modelling was NONMEM in 1979[32]. The later introduced programs WinNonMix[33] and nlme (S-plus)[34] apply a similar approach as the methods in NONMEM. In recent years a new approach to solve the integral for the likelihood has been presented and is available in for example NLMIXED (SAS)[35].

A complement to these parametric approaches is the non-parametric likelihood methods, for example NPEM[36], and the Bayesian methods, for example WinBUGS[37]. Even though new software constantly is being presented for mixed-effects modelling, NONMEM is still the most widely used soft-

ware for population non-linear mixed-effects PK-PD modelling, within both academia and the industry[31, 38, 39].

Laplacian method (in NONMEM)

In NONMEM (Icon Development Solutions), three methods for estimation of parameters are available that all maximizes the likelihood $L(y|\theta,\eta,X)$ of the observations given the model; the first-order (FO), the first-order conditional estimation (FOCE) and the Laplacian methods[32]. The likelihood is the product of all individual likelihoods L_i and since the likelihood must account for the random effects on the individual level, the individual likelihood is expressed as the integral over all possible values of η_i as shown in equation (21).

$$L = \prod_i L_i = \prod_i \int l\left(y_i \mid \theta, \eta_i, X_i\right) h\left(\eta_i \mid \omega^2\right) d\eta_i \tag{21}$$

As this integral does not have a closed form expression when the model is non-linear with respect to the η's different approximations are required in order to solve the likelihood. The FO and FOCE methods involve a linearization of the non-linear model using a first-order Taylor expansion around zero and the conditional estimates of the random effects, respectively. The third method available in NONMEM is the Laplacian method in which a closed form of the expression is obtained by using Laplace's approximation of the integral with a second-order Taylor expansion around the conditional estimates of η's for the non-linear model. The conditional estimates or the empirical Bayes estimates (EBE) of η's are the best predictors of the random effects, η_i.

The Laplacian method with the likelihood option is needed to be specified for estimation of categorical models as the coding of these models are done specifying probabilities of the observations, instead of the actual value of the observation as with continuous data.

The mean of the EBE of η's is allowed to be non-zero if the data imply so. In the output from NONMEM, the means of the EBE of η's is given together with the probability that these are zero. As the mean of the random effects of a mixed-effects model as implemented in NONMEM is assumed to be zero, the centering option has been provided forcing the mean of the EBE of η's to be zero.

Gaussian quadrature method (in NLMIXED, SAS)

As oppose to the Laplacian method in NONMEM where the likelihood is approximated, the Gaussian quadrature (GQ) method estimates the likelihood exactly; however with the use of an inexact numerical integration method. As the GQ method is defined in NLMIXED (SAS®), the numerical integration is performed as a weighted average of the likelihood evaluated at predefined points, the Gauss-Hermite abscissas[35]. The integral is centred at either zero (non-adaptive approach) or the EBE of η's (adaptive and non-adaptive scaling approaches). Additionally, the scaling of the abscissas is either performed using the Hessian matrix (adaptive approach) or the variance-covariance matrix (non-adaptive and non-adaptive scaling approaches). The number of quadrature points is selected automatically by evaluation of the log likelihood function at the initial values of the parameters with different numbers of quadrature points until two successive evaluations have a difference less than a predefined value. However, the number of quadrature points can be specified and a special case of the GQ method is to set only one quadrature point, which will render Laplace's approximation.

Back-step method

The back-step method (BSM) is not a method of regression, as it does not maximize the likelihood, merely uses the existing method within the software where it is run. The BSM is nevertheless an iterative method for estimation of parameters, which was developed to provide the unbiased parameter estimates in those cases where the regression method is unable to provide those.

The BSM is schematically depicted in Figure 3. The method searches for the parameters that upon simulation will yield a marginal distribution of the simulated observations that is similar to that observed. The search for these unbiased parameter estimates is an iterative process involving sequential simulations-estimations. The BSM is started for a given model by simulating a new dataset using the parameters that are suspected to be biased, Φ_{bias}. The model is used again for estimations of new parameters, $\Phi_{est}(1)$ based on the simulated data. A summary measurement for the difference between $\Phi_{est}(1)$ and Φ_{bias} is calculated, $\Delta\Phi$, see equation (22) with n=1. If $\Delta\Phi$ is less than a pre-specified value convergence is reached.

$$\Delta\Phi = \sum_{r=1}^{R} \frac{\left|\phi_{bias,r} - \phi_{est,r}(n)\right|}{SE\left(\phi_{bias,r}\right)} \tag{22}$$

in which R is the number of parameters in the model, $\phi_{bias,r}$ is the r^{th} parameter belonging to the set of parameters that were suspected to be biased, Φ_{bias}, while $\phi_{est,r}(n)$ is the r^{th} parameter belonging to $\Phi_{est}(n)$. $SE(\phi_{bias,r})$ is the standard error (SE) estimate of the r^{th} parameter of Φ_{bias}. The expression for $\Delta\Phi$ is scale-variant, however dividing the difference between the biased and r^{th} estimate of a parameter with the SE estimate of the biased parameter was done to reduce the effect of large differences between magnitudes of the parameters in the model. The SE estimates are merely used to get the parameters on approximately the same scale, as the size of the SE is related to the size of the parameter.

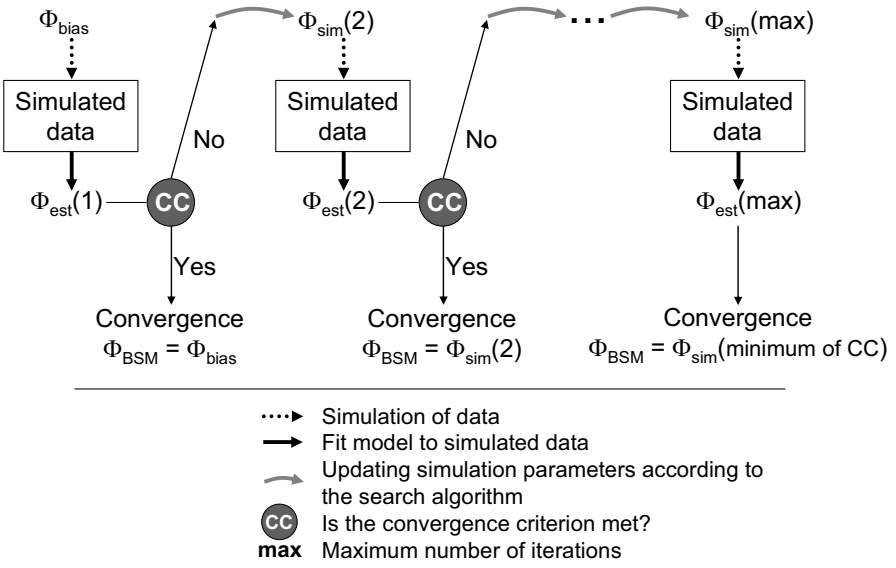

Figure 3. The back-step method involves iterative simulations-estimations, with an update of the parameters in between. This sequence is repeated until the estimated parameters, Φ_{est} fulfil the termination criterion (CC) or the maximum number of iterations is reached. At convergence, the set of parameters used to simulate, Φ_{sim} is the final parameter set, Φ_{BSM}. If the maximum number of iterations is reached, the step in the BSM sequence with the Φ_{est} that is closest to convergence will be used as the final step.

Before a new step in the BSM is taken, the simulation parameters need to be updated according to the search algorithm, equation (23). The new r^{th} simulation parameter, $\phi_{sim,r}(n+1)$ has the same value as the previous r^{th} simulation parameter, $\phi_{sim,r}(n)$ with a percentage change that is the same as the percentage difference between the first r^{th} parameter, $\phi_{bias,r}$ belonging to Φ_{bias} and the latest r^{th} estimated parameter, $\phi_{est,r}(n)$.

$$\phi_{sim,r}(n+1) = \phi_{sim,r}(n) \cdot \left(1 + \frac{\phi_{bias,r} - \phi_{est,r}(n)}{\phi_{est,r}(n)}\right) \quad (23)$$

The process with simulation-estimation followed by parameter update is continued until the $\Delta\Phi$ is less than the pre-specified value of the convergence criterion. If this does not occur within the specified maximum number of iterations, the simulation-estimation step with the lowest $\Delta\Phi$ is considered the final step. Consequently, depending on how strict the convergence criterion is defined all or none of the BSM runs may end with maximum number of iterations being reached.

There are three ways of getting the final BSM parameter estimates, Φ_{BSM}: *(i)* single estimates, or using the *(ii)* serial or *(iii)* parallel procedure. If a single estimate is used, simply use the Φ_{sim} of the final step as the BSM parameter estimates. Both the serial and the parallel procedure will provide estimates with higher precision as these uses an average of several Φ_{sim}. In the serial procedure, the iterative process of a BSM run is extended with u additional steps, and the average of the Φ_{sim} of the v steps with the lowest $\Delta\Phi$ is used as the Φ_{BSM}. In the parallel procedure, several BSM estimations are performed for the same data and the average of the single estimates of these runs is used as the final Φ_{BSM}.

Model selection

When deciding to include a parameter in a model, in particular a parameter describing a covariate relationship, objective measurements are needed to verify that inclusion is relevant. Relevance can be defined as physiological plausibility, clinical impact or statistical significance. A combination of all these is usually used in combination with assessments of the predictive performance of the model. Physiological plausibility as well as clinical impact are depending on the drug and disease and are therefore decided based on subject-matter information[45]. Criteria for inclusion of a parameter based on physiological plausibility and clinical impact need to be defined pre-analysis in order to make these measurements objective.

Statistical comparison

The most commonly used test within PK-PD modelling to assess statistical significance is the likelihood ratio test (LRT). The ratio between the likelihood of the model with the new parameter included (the full model) or excluded (the reduced model), given that the models are nested, is assumed to be χ^2-distributed with the number of differing parameters between the mod-

els as the degree of freedom (df). Models are nested if the full model can collapse into the reduced model. The parameter can thus be included in the model based on a statistical significance level. This level is corresponding to the type I error, the risk of including a parameter that do not belong to the model.

The objective function value (OFV) calculated by NONMEM is proportional to -2log likelihood and the difference in OFV (ΔOFV) between two models is then the likelihood ratio. Thus, the ΔOFV can be used to perform formal statistical testing between nested models. A number of studies has been performed to assess the robustness of the LRT[46, 47], but only one has been performed with categorical data[48]. Investigated in this study was ordinal data, originating from and analyzed with the proportional odds model and it was concluded that the LRT performs well with type I errors equal to or lower than expected.

If the assumptions made with a model are violated, the χ^2-distribution may be distorted giving other type I errors than expected when using the tabulated critical values for the χ^2-distribution. The actual critical values given a type I error can be obtained by performing a simulation hypothesis test[49]. In this test, a given number of datasets are simulated using the reduced model. These datasets are then analyzed using both the reduced and the full model and the $1\text{-}z^{th}$ percentile of the corresponding ΔOFV of the reduced-full model analysis pairs is the actual critical value at a z significance level. The number of datasets simulated is depending on which significance level that is desired.

Predictive performance

Goodness-of-fit graphics are commonly used to aid the decisions of model building. The predictive performance of a model can be assessed using different approaches all involving simulation of new data. At the end of an analysis extensive simulation exercises such as the posterior predictive check (PPC)[40-42] is used to investigate predictive performance of the model. The PPC as suggested by Gelman[40] involves assessing a summary measurement of the data, for example maximum concentration. The posterior predictive distribution of the summary measurement is obtained through simulation of new data using the model. The parameter estimates used for simulations are drawn from the posterior density of the parameters; that is taking the SE of the estimated parameters into account. The summary measurement observed is then compared to its posterior predictive distribution. This method is even though exhaustive also quite computational burdensome and it is only practical to perform for the final model.

The visual predictive check (VPC) as discussed by Holford and Karlsson[43] is a more convenient alternative during the model building. A statisti-

cal summary measurement, for example median, calculated from the distribution of the observed and the simulated data is compared over the range of the independent variable or a covariate. The simulated data are usually obtained using the point estimates of the parameters. The observed data points, the summary statistic calculated based on the observed and the prediction interval of the simulated data are plotted against the independent variable or covariate, to give a visual check of how the model performs over the range of the predictor. Also, the VPC is require quite large computer power and maybe more feasible for each step in the model building are the graphical goodness-of-fit methods for categorical data analysis available in Xpose version 3[44].

Aim of the thesis

The aim with this thesis was to investigate the performance and to improve the use of models and methods for mixed-effects population PK-PD analysis of categorical data.

In particular, the aims were:

- To examine the Laplacian method's performance with polychotomous data and explore and develop alternative methods
- To investigate two of the assumptions made with the proportional odds model with emphasis on the predictive performance and the impact on the likelihood ratio test's reliability and explore and develop alternative approaches
- To illustrate the use of a proposed reduced model building strategy for sleep stages using the Markov model

Methods

A majority of the studies included in this thesis were performed using simulated data. In these studies, the true model is known, making investigation of for example bias (paper I) and the appropriateness of the LRT (paper IV) possible. Simulation of data is also the basis for the BSM (paper II). However, data from real clinical studies were used in papers III and V.

Clinical data

T-cell receptor density (Paper III)

T-cells reacting with the myelin components in the brain as well as in the spinal cord is the cause for the inflammatory lesions seen in the disease multiple sclerosis (MS). This inflammatory reaction leads to demyelination and in the end axon loss, typical for this disease[50, 51]. The drug ATM-027, a humanized antibody, was developed as a treatment for MS as this antibody selectively depletes a subgroup of T-cells, Vβ5.2/5.3+, thought to be associated with the disease. $V\beta5.2/5.3^{+}$ T-cells can be measured in the blood using two-colour flow cytometry.

ATM-027 was investigated in a dose ranging phase I study with 14 patients (7 females and 7 males) with relapsing-remitting or secondary-progressive MS[52]. ATM-027 was administered as a 30 minutes constant rate infusion with initially 6 dose levels: 0.03, 0.3, 3, 30, 100 and 300 mg. The dose was escalated every 72 hours, provided it was safe, until $V\beta5.2/5.3^{+}$ T-cells were depleted. Blood sample for analysis of ATM-027 concentrations were taken before drug administration, at the end of each infusion and at each follow-up visit up to and including 8 months. The blood samples for flow cytometry were drawn 4 weeks, 1-2 weeks and 24 hours before administration of ATM-027, at 2, 24 and 72 hours after start of each infusion and at each follow-up visit up to and including 18 months (1, 2, 4 and 8 weeks, 4, 6, 8, 10, 12 and 18 months after last dose). Since ATM-027 was found to affect both the number of T-cells as well as the expression of receptors on the T-cell surface, the receptor density was reported as a subjectively categorized continuous variable, with the levels: dim, intermediate and bright. The

number of observations of T-cell receptor density is given in Table 1. For further details about the clinical study see Olsson *et al*[52].

An analysis of these data has previously been published using the proportional odds model[22]. The concentrations (C), predicted using the doses and individual PK parameters from a previously performed standard two stage analysis, were used as the covariate in the drug effect model and an E_{max} model was identified as the most appropriate to describe the relationship between concentrations of ATM-027 and T-cell receptor density. Two parameters are estimated in this model: θ_{Emax} the maximum drug effect and θ_{EC50} the concentration at which half of θ_{Emax} is obtained (see equation 24). For further details about the analysis with the proportional odds model see Zingmark *et al*, 2004[22].

$$g(C) = \frac{C \cdot \theta_{Emax}}{C + \theta_{EC50}} \tag{24}$$

Table 1. *Marginal distribution of responses for the clinical data analyzed in this thesis given as percent of total number of observations. Two rows are provided for diarrhoea severity showing (i) observations where irinotecan and SN-38 concentrations are available and (ii) observations where SN-38G concentrations are available. For sleep data, the number of transitions is presented with the category from which the transition takes place given in the second column while the category to which the transition takes place given in the second row.*

Data		Total number of observations	Category						
			0	**1**	**2**	**3**	**4**	**5**	**6**
T-cell receptor density[a]		206	54	6.6	40	*NA*	*NA*	*NA*	*NA*
Diarrhoea severity	*(i)*	143	46	28	20	4.2	2.1	*NA*	*NA*
	(ii)	65	46	25	20	4.6	4.6	*NA*	*NA*
Sedation[b]		21092	*NA*	48	33	11	6.3	1.8	0.35
Sleep[c]	0	40214	-	21	6.0	0[d]	*NA*	13	*NA*
	1		7.3	-	22	0[d]	*NA*	4.8	*NA*
	2		9.1	8.1	-	6.3	*NA*	3.9	*NA*
	3		0.98	9.1	4.4	-	*NA*	0[d]	*NA*
	5		4.5	4.3	10	0[d]	*NA*	-	*NA*

[a] Categories dim / intermediate / bright correspond to 0 / 1 / 2
[b] Definition of sedation categories see Table 2
[c] Categories awake / stage 1 / stage 2 / deep sleep / REM sleep correspond to 0 / 1 / 2 / 3 / 5
[d] The observed values were 1, 15, 1 and 12 for TP 0→3, 1→3, 5→3 and 3→5, respectively.
NA - Not Applicable

Diarrhoea severity (Paper III)

Diarrhoea is the major dose-limiting non-haematological toxicity of the anti-cancer agent irinotecan[53] and this side-effect is commonly reported on the ordinal scale with 5 levels, ranging from 0 (no diarrhoea) to 4 (severe diarrhoea). The drug irinotecan is a topoisomerase I inhibitor with one active metabolite, SN-38[54]. SN-38 is further conjugated to an inactive glucuronic acid conjugate, SN-38G[55, 56].

The diarrhoea data used in this work came from a study with 64 men and 44 women having different malignant solid tumours and the patients were included in the study for 1 or 2 (no = 44) treatment courses[21]. The diarrhoea severity was measured in 104 of the patients. This was done once during a course and measured the worst grade of diarrhoea experienced during a three-week period after drug administration. Blood sample were collected before the drug was administered and at approximately 0.5, 1.5, 1.67, 1.83, 2, 2.5, 3.5, 4.5, 5, 5.5, 6.5, 8, 12, 25.5, 49 and 56 hours post-dose. Plasma concentrations of irinotecan and SN-38 were measured in all patients and SN-38G concentrations were measured in 51 of the patients with diarrhoea severity measurements. The area under the plasma concentration curve from time 0 to 60 hours ($AUC_{(0\text{-}60)}$) for each of the compounds was calculated using the PK parameters for the individuals estimated based on the measured concentrations. The marginal distribution of responses are in Table 1 given for both observations were irinotecan and SN-38 concentrations are available as well as for those where SN-38G concentration measurements are available.

A previously published PK-PD analysis of these data was performed using a proportional odds model[21]. This analysis resulted in statistically significant correlations between the diarrhoea severity and the $AUC_{(0\text{-}60)}$ of irinotecan and SN-38G, but not with $AUC_{(0\text{-}60)}$ of SN-38. A linear model with one parameter, θ_{slope}, was used to describe the relation between the diarrhoea severity and the $AUC_{(0\text{-}60)}$ of the three different compounds, as shown in equation (25). For further details about the clinical study and the analysis with the proportional odds model see Xie *et al*[21].

$$g\left(\mathrm{AUC}_{(0\text{-}60)}\right) = \mathrm{AUC}_{(0\text{-}60)} \cdot \theta_{slope} \tag{25}$$

Sedation (Paper III)

Stroke is a major public health concern, with an annual incidence ranging from 1300 to 4100 in a population of 1,000,000 people[57]. It is the third leading cause of death, behind heart disease and cancer and the leading neurological cause of long-term disability[58]. Even though efforts have been made

to find a treatment for this severe condition, the only approved treatment so far is tissue plasminogen activator. Clomethiazole was after promising findings in animals[59-65] investigated as a potential drug in the treatment of acute ischemic stroke because of its neuroprotective properties. The main dose-limiting side effect of clomethiazole in stroke patients is sedation; especially as stroke itself causes sedation[66].

Sedation is measured on a ranking scale with six levels (Table 2). When the symptom of the disease and the treatment side effects coincide, it is even more important than usual to monitor the effect and an adaptive design was utilised in all three studies used for the analysis of sedation. 1545 patients (777 on clomethiazole and 768 on placebo) with a clinically diagnosed acute stroke within 12 hours after onset were available for the sedation analysis[25]. Clomethizole was administered as a 3-phasic infusion with the rates; 24 mg kg^{-1} $hour^{-1}$ for the first 15 minutes, 4 mg kg^{-1} $hour^{-1}$ up to 8 hours and 1.94 mg kg^{-1} $hour^{-1}$ from 8 to 24 hours, but alterations of this was made depending on the observed sedation. If a score of 4 or greater was observed, the infusion was stopped and optionally, the infusion could be re-started at half the previous rate as the reported score decreased to three or less. Blood samples were taken 15 minutes, any time between 1 and 2 hours and 24 hours after infusion of drug was started. If the infusion was stopped, the last blood sample was taken at the time of stopping the infusion. At the late stage of one of the studies, the sampling times were altered to be any time between 2 and 10 hours and any time between 26 and 36 hours after starting the infusion. The first time the infusion was stopped, an extra sample was also drawn. Sedation was monitored at least every 3rd hour up to 24 hours but if a score of 3 or greater was observed the observations frequency increased to every 15 minutes. The number of observations in each category is given in Table 1.

Table 2. *Sedation is commonly measured on a six levels ordinal scale.*

Category	Definition
1	Fully awake
2	Drowsy but answers when spoken to
3	Answers slowly when spoken to
4	Reacts when spoken to but does not answer
5	Reacts only to pain
6	Does not react to pain

In the previously published PK-PD analysis[25] using the proportional odds model both a placebo/disease progression and a drug model was developed. A sedation non-sensitive sub-population identified for the placebo/disease progression model was described with a mixture model defining different

intercept parameters for the probability of observing sedation scores ≥ 2 for sedation non-sensitive and sensitive populations. The placebo effect/disease progression was best described using an exponential model, with three parameters: θ_{plc1} the maximum placebo effect, θ_{plc2} the maximum change in placebo effect and θ_{plc3} the steepness of the effect onset. The drug effect was a step model, with one parameter: θ_{drug} the magnitude of the drug effect present for individuals treated with clomethiazole and not present for all other individuals, see equation (26). A linear and an E_{max} model were also tried using predicted plasma concentrations based on the PK model for clomethiazole. None of these models were however statistically significant on a 5% level. The marginal distribution of the responses of sedation is given in Table 1. For further details on the clinical study and the analysis with the proportional odds model see Zingmark *et al*, 2003[25].

$$\begin{aligned} g(placebo) &= \theta_{plc1} - \theta_{plc2} \cdot e^{-\theta_{plc3} \cdot time} \\ g(drug) &= I_{0/1} \cdot \theta_{drug} \end{aligned} \tag{26}$$

Sleep stages (Paper V)

Traditionally, six sleep stages are defined for data analysis of sleep: awake, stage 1, stage 2, stage 3, stage 4 and REM sleep. Stage 3 and stage 4 can be recorded together in a stage referred to as deep sleep. This categorization is done using polysomnography (PSG) where the dominating sleep stage is determined for each 30-second interval using the standard definitions by Rechtschaffen and Kales[67]. Based on these sleep stages, quantitative and qualitative measurements of sleep can be calculated, such as latency to persistent sleep (LPS), number of awakenings (NAW) and sleep efficiency (SEF). Insomnia is a condition where the sleep architecture is disturbed with for example a prolonged LPS and larger NAW. Consequently, insomnia may lead to sleep deprivation, which has been shown to increase the risk of for example type 2 diabetes[68], depression[69] and high blood pressure[70]; thus, finding treatments for insomnia is important. The goal when developing a drug for treatment of insomnia is to increase the amount of sleep, especially the deep sleep, decrease the time to fall asleep but at the same time not alter the non-REM versus REM sleep architecture.

Study Design

The drug PD 0200390 is a new class of drugs for treatment of insomnia, a calcium channel $\alpha 2\delta$ subunit-binding compound[71]. The data used in this analysis were obtained in a randomized, double-blind, active- and placebo-controlled, Latin square, 4-way crossover study of PD 0200390, conducted

in 43 patients (25 females and 18 males) with primary insomnia to assess the safety and efficacy of 25 mg and 75 mg of PD 0200390. Zolpidem 10 mg was included as the active control. Insomnia patients, eligible for inclusion into the study, were scheduled for two nights of screening PSG in the sleep laboratory. Each night for screening as well as during the study, patients went to bed at their habitual bedtime and underwent PSG evaluation for eight hours. On the testing nights, subjects took the study medication 30 minutes prior to the start of PSG recording. PSG data were collected digitally and transferred directly to a central reader for scoring. Subjects returned for two consecutive nights of PSG for each treatment, at approximately weekly intervals, until all four study-treatments had been completed.

Number of transitions

Sleep was divided into 6 stages; initial sleeplessness (IS), awake (0), REM sleep (5), stage 1 (1), stage 2 (2), and deep sleep (3). The stage of initial sleeplessness was added based on previous experience with sleep data[72] and was introduced to provide a separate transition probability for the first time a patient falls asleep, since this is different from falling asleep when sleep already has occurred. All patients start in this stage and the only occurring transition from this stage is to stage 1. Furthermore, a patient cannot return to this stage once a transition has occurred.

Since there are 5 stages + initial sleeplessness, 21 transitions are theoretically possible. However, not every one of these transitions occurs physiologically and some transitions are very infrequent. To reduce the number of transitions to model, three criteria were defined to identify the transitions of interest; *(i)* a transition represented at least 1% of all observations, *(ii)* a transition represented at least 10% of all transitions from a stage and *(iii)* a transition represented at least 10% of all transitions to a stage. A transition was modelled, if at least one of the criteria was fulfilled. The marginal distribution of the sleep observations are given in Table 1 with the observations of initial sleeplessness and awake given together.

Data analysis

Models

All models used in this thesis; the proportional odds, the differential odds and the Markov model, are described in detail in the section regarding models for categorical data analysis. The proportional odds model was used in all papers, except paper V, for simulation of data (papers I and II), estima-

tion of parameters (papers I, II and IV) and comparison with another model (paper III). The placebo effects included in these models were described either using step models (papers I and II) or exponential models with time as the covariate (paper III). The drug effect of the same models were; a step model (paper III), a linear model with dose (papers I and II) or $AUC_{(0\text{-}60)}$ (paper III) as the covariate and an E_{max} model with concentrations as the covariate (paper III). For a majority of the models used for simulations and estimations in paper I and for the reduced incorrect model in paper IV, no covariate functions were defined.

The performance of the differential odds model was investigated for different ranking scales; diarrhoea severity, T-cell receptor density and sedation data (paper III). In practice, the differential odds model was implemented in this thesis using differential effect parameters, β as shown in equation (27).

$$\begin{aligned} f_2 &= \alpha_2 + g\left(X_i\right) \\ f_3 &= \alpha_2 + \alpha_3 + g\left(X_i\right)\cdot\beta_3 \\ f_3 &= \alpha_2 + \alpha_3 + g\left(X_i\right)\cdot\beta_3\cdot\beta_4 \\ &\ldots \\ f_m &= \alpha_2 + \sum_{k=3}^{m}\alpha_k + g\left(X_i\right)\cdot\prod_{k=3}^{m}\beta_k \end{aligned} \tag{27}$$

The differential effect parameters, β_j describe how much the covariate effect deviates from parallel slopes on the log odds for the probability of scores $\geq j$, and they have the limits (0, 1). The upper limit was set to one to ensure that $g_2(X_i) \geq g_3(X_i) \geq \ldots \geq g_l(X_i)$ while the lower limit was set to zero as this is the value at which a score is independent of the covariate. The differential odds model was only investigated as affecting the covariate relationship describing drug effects.

The impact of changing the link function from logit to probit or complementary log-log was investigated in paper III for the data where a significant differential effect was found. This was done to examine if changing the link function was sufficient in providing differential odds, without estimating additional differential effect parameters.

The Markov model was in this thesis used for simulation and analysis of di- and trichotomous data (paper IV) and analysis of sleep stages (paper V). In the simulations and analyses performed in paper IV, no IOV was defined. The baseline transition probability was in paper IV implemented using a constant intercept parameter and the only covariate relationship for these models (a placebo effect) was described using a step model. The baseline transition probabilities and covariate relationships for the models describing

sleep stages (paper V) are discussed in more detail in the section regarding developing a model for sleep stages.

Simulation studies

The simulated response in paper I and II was an ordinal variable with the levels: 0, 1, 2 and 3. The generated datasets contained 1000 individuals with four observations each. The simulation values were set to generate data with different distribution of responses between the four categories with varying degree of variability, see Table 3.

Table 3. *Datasets were simulated to have different distribution of responses between the categories with different degree of variability. The variability is given as variance on logit scale.*

Scenario	**Expected proportion of response (%)**			**Variance**
	in category 0/1/2/3 at baseline	***in category 3 placebo***	***in category 3 dose=4 units***	
A	25/25/25/25	*NA*	*NA*	4, 10, 40
B	82.5/10/5/2.5	*NA*	*NA*	4, 10, 40
C	90/5/3/2	*NA*	*NA*	0.5, 2, 4, 10, 40
D	24/26/26/25	30	50	4
E	96.5/1.2/1.4/0.84	3	6	0.5, 4, 40

NA - Not Applicable

One hundred datasets of each scenario was simulated and analyzed using the cumulative odds model with random intercept (scenario A, B and C) or the proportional odds model (scenario D and E). The 1000 individuals in scenario D and E were evenly divided into 4 dose groups, receiving 0, 1, 2 and 4 units of drug, respectively. The estimations were performed using different methods of regression: the Laplacian method with and without centering option, and the GQ method with adaptive, non-adaptive scaling and non-adaptive approaches.

Scenario D and E were also used in paper II to evaluate the performance of the BSM where 100 simulated datasets were analyzed using the iterative method, BSM and the Laplacian method.

The Markov model was in paper IV used for simulation of di- and trichotomous data. Each dataset contained 250 individuals with varying number of observations, evenly spaced 1 hour apart. The transition probability, the number of observations per individual and the variance was varied according to Table 4. Standard settings were defined (highlighted with bold font in Table 4) and these were kept constant when possible.

Table 4. *The number of observations per individual (Nobs), the transition probability (TP) and the variability were varied for both the di- and trichotomous datasets. Standard settings are highlighted with bold font. Variability is given as variances on logit scale.*

Scenario	Nobs	TP (%)	Variance
Dichotomous	6, **12**, 24	**10**, 20, 30, **40**, 80	0.001, 0.01, **0.05**, **0.1**, 0.5, 1
Trichotomous	6, **12**, 24	**5**, 10, 15, **20**, 40	0.01, 0.05, **0.1**, **0.5**, 1, 5

One thousand datasets generated through simulations were analyzed with the correct model, a Markov model and an incorrect model, the proportional odds model. Both the correct and incorrect model was tried with and without the inclusion of a false covariate: a step model describing placebo effects. A schematic description of the simulation study design is given in Figure 4. The type I error was the percent of ΔOFVs between the reduced and full model with a value less than 3.84 (df = 1, χ^2-distribution).

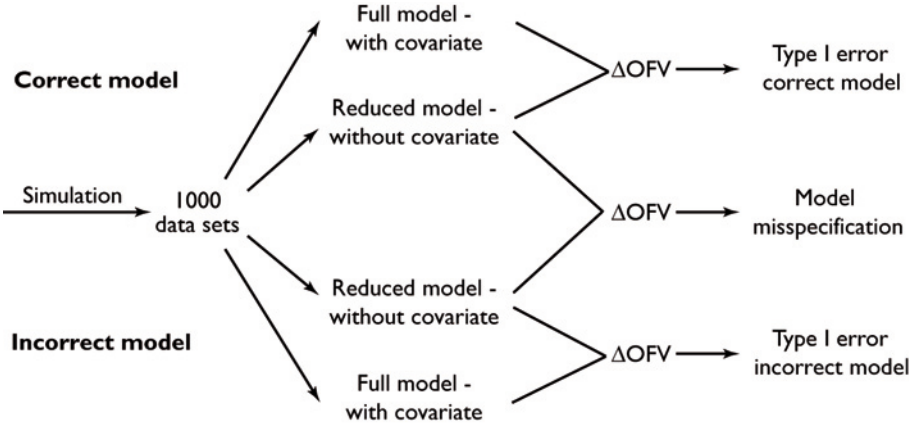

Figure 4. Each simulated dataset was analyzed with four different models to assess the type I error for the correct and incorrect model. The model misspecification was assessed by comparing the OFV for the correct and incorrect model without the false covariate.

Software

The software NONMEM version V, version VIβ and version VI (ICON Development Solutions)[32], with the Laplacian method and the likelihood option with and without the centering option were used for estimation of all models (proportional odds, differential odds and Markov model) and simulation of data using the proportional odds model and the cumulative odds model with random intercept. In paper I the models were also estimated using the GQ method in NLMIXED (SAS Institute Inc.)[35]. Simulation of data using

Markov models, as well as all automated data manipulation was performed in Perl[73]. Some interactions with NONMEM were automated using Perl and the Perl-speaks-NONMEM (PsN) application-programming interface[74]. Bootstrapping was performed using the PsN-toolkit[75].

Model discrimination was based on simulation based goodness-of-fit graphs and changes in the OFV provided by NONMEM. Goodness-of-fit graphics were created using R[76] and S-plus (Insightful Inc.)[77].

Developing a model for sleep stages

Building models for sleep data using data from a clinical study is more demanding than most categorical data analyses. The model choice for sleep data is the Markov model as measuring sleep each 30-second will generate data where the observations are dependent of previous observations. The quite large number of categories in combination with the choice of model, the many observations and the numerous study arms in a clinical study all adds to the complexity of the model building, apart from the complexity the data itself adds. Consecutively, for these models to be used in a clinical drug development setting the model building procedure must be efficient. One of the aims of paper V was to reduce the model building associated with mixed-effects Markov models for sleep data.

As mentioned in the section regarding number of transitions, there were 21 possible transitions, but not all transitions are equally likely. With the help of the defined rules for inclusion of transitions the number of transitions could be reduced to 16, see Figure 5.

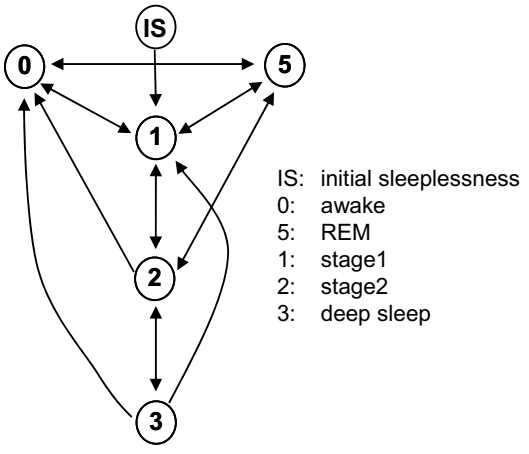

Figure 5. The transitions chosen for modelling are indicated with an arrow.

The main structure of the model for sleep is the Markov model as given in equations (15), (16) and (17). As no parameters are shared between the different transitions, all transitions can be modelled separately. This was done to avoid any problems with NONMEM's limits for maximum allowed number of observations per individual and to decrease the runtimes. The $f_{m|l}$ in equation (28) is a function depending on the baseline probability of a transition $l \rightarrow m$, placebo effects and drug effects, where the baseline probability is depending on relative bedtime and stage time. Relative bedtime is a fraction defined as the time elapsed since actual bedtime divided by the total time elapsed since actual bedtime until the PSG recording is stopped, usually eight hours. Stage time was defined as the time elapsed since last change in sleep stage.

$$
\begin{aligned}
f_{m|l} &= \alpha_{m|l} + g\left(placebo\right) + g\left(PD\ 0200390\right) + g\left(zolpidem\right) \\
\alpha_{m|l} &= \ln\left(\frac{TPB_{m|l}}{1 - TPB_{m|l}}\right) \\
\text{where } TPB_{m|l} &= f\left(relative\ bedtime\right) \cdot f\left(stage\ time\right)
\end{aligned}
\tag{28}
$$

where $TPB_{m|l}$ describes the transition probability at baseline to m given previous observation of l as a function of relative bedtime and stage time. Introducing the $TPB_{m|l}$ into the model as the logit of the expression was done to allow estimation of parameters in the range of standard values of probabilities, that is (0, 1). To ensure that the limits not were exceeded, the upper limits of the parameters of *f(stage time)* were set to the reciprocal of the corresponding upper limits of the parameters of *f(relative bedtime)* and the lower limits of the parameters of *f(relative bedtime)* and *f(stage time)* were set to zero. Keeping the values of the function for relative bedtime in the same ranges as the probability enables interpreting the values of function for stage time as multiplying factors. Thus, if the stage time function is estimated to 2 the transition probability is doubled, and consequently, if the stage time function is estimated to 0.5 the transition probability is reduced to half.

The model building strategy, as depicted in Figure 6, can be described by the following steps *i)* Explore the baseline models using screening sleep data; *ii)* Add placebo data and explore the placebo model; *iii)* Develop the pharmacodynamic models of PD 0200390 and zolpidem in parallel, based on the data for each drug added to the placebo and screening data; and *iv)* Merge all models and re-estimate all parameters with the support of all data. In each step of adding data, the parameters of the previously developed model were fixed. The observed drops in OFV when including a new pa-

rameter in the model represented an overestimate of the drop expected had all parameters in the model been estimated.

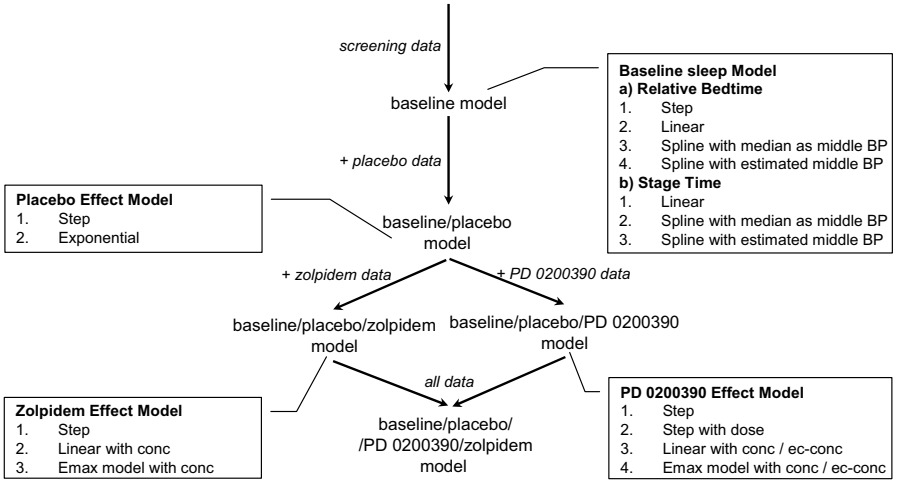

Figure 6. The reduced model building strategy. The four predefined models of relative bedtime and the three of stage time were explored for each transition chosen for modelling using only screening data to define the baseline model. The placebo data was added and the two predefined models for placebo were investigated. The predefined drug models for PD 0200390 and zolpidem were explored in parallel adding the data for the corresponding drug to the screening/placebo data. In the final step, all data and models were merged into one model. BP - breakpoint, conc – predicted concentrations, ec-conc – predicted effect compartment concentrations.

In order to reduce the model building procedure even further, a number of predefined models with increasing complexity were applied to explore the relationships between transition probability and relative bedtime, stage time, placebo and drug effects. For relative bedtime, *f(relative bedtime)*, these models were a step, a linear and a linear spline with three breakpoints. For stage time, *f(stage time)*, the models were a linear and a linear spline with three breakpoints. The middle breakpoint of the spline models were, for both the relative bedtime and the stage time models, either estimated or fixed to the median. The standard models for the placebo effects were a step and an exponential using time as a covariate for the placebo effect. The predefined models for the drug effects were similar: a step, a linear and an Emax. For PD 0200390 two doses were given, hence a step model varying with the dose was also tried for this drug. The covariates for the linear and Emax models were the predicted concentrations and effect compartment concentrations[78] based on previously developed PK and PK-PD models for the two drugs. As concentrations of PD 0200390 were measured, individual predictions could be made of both the concentrations and the effect compartment

concentrations for this drug. However, neither concentration measurements nor a measurement of the effect compartment delay, k_{e0}, was available for zolpidem, thus only the population predicted concentrations were used for predicting the effect of zolpidem.

Bias and precision

The bias in parameter estimates was in paper I as well as in paper II assessed by calculation of the relative estimation error (REE) as described in equation (29).

$$REE = \frac{\phi_{est} - \phi_{sim}}{\left| \phi_{sim} \right|} \tag{29}$$

where ϕ_{est} is the estimate of a parameter, fixed or the variance of the random, and ϕ_{sim} is the value used for the same parameter when generating the data. The REE for each parameter of each simulated dataset was plotted in box-and-whiskers plots to visualize both the bias and the precision. The relative bias is calculated according to equation (30).

$$relative\ bias = \frac{\phi_{est,median} - \phi_{sim}}{\left| \phi_{sim} \right|} \tag{30}$$

in which $\phi_{est,\ median}$ is the median of all estimates of a parameter.

The performance of the BSM was investigated in paper II, with an assessment of the precision of the method. To assess the contribution to the imprecision by the BSM, ten of the performed BSM estimations were analyzed using a mixed-effects model with two levels of random effects: the variability due to that a limited sample of the population is studied and the variability due to the iterative nature of the BSM. This was done for each estimated parameter as either a single BSM estimate or using the serial BSM procedure.

Predictive performance

A PPC with 100 simulations was performed for the developed Markov model for sleep. The point estimates of the parameters instead of the posterior distributions were used for simulations, as this decreased the computational load slightly. The 18 sleep efficacy endpoints, for example LPS and NAW, for each the 4 treatments were used as the summary measurement of the PPC. To get an overview of the performance of the model, the relative

deviation between the observed endpoint and the median, maximum and minimum endpoint as derived from simulated data was calculated for each of the endpoints, according to equation (31).

$$relative\ deviation = \frac{Endpoint_{sim} - Endpoint_{obs}}{Endpoint_{obs}} \quad (31)$$

VPC was utilized to investigate the predictive performance of the used methods/models in papers I, II III and V. The realized design was used for the simulations except for the sedation data in paper III. As this study design was adaptive only the scheduled, observation independent measurement times were used in the simulations for the VPC. Empirical probability was in papers I, II and III the statistical summary measurement and the average time spent in a stage was the summary statistic for sleep data. These summary statistics were plotted versus covariate variables: dose (papers I, II and III) and relative bedtime or stage time (paper V). As dose is a categorical measurement histograms showing the empirical probability of the different doses were created in papers I, II and III. The VPC of sleep data was stratified on treatment. Using the population parameters for simulation assumes zero mean of the variance. In those scenarios of paper I where the parameter estimates were biased the mean of the EBE of η's was non-zero; thus additional simulations using the EBE of η's and the estimated fixed parameters were performed as an alternative method to explore any possible improvement of the predictive performance.

Simulation hypothesis test

The actual critical values for the likelihood ratio test were in paper III and IV assessed using a SHT. A significance level of 1% and 5% was used in paper III and IV, respectively. In paper III, the critical values of including differential effect parameters, β's were investigated by simulating using the proportional odds model and analyzing the data using the proportional and the differential odds model. The actual critical values for the Markov elements were assessed in paper IV by simulating data using the proportional odds model and estimating the parameters using the proportional odds model and the first order Markov model. This was done for dichotomous data. The 99^{th} and 95^{th} percentiles of the difference between the OFV estimates of the reduced model (the proportional odds model) and the full model (the differential odds or Markov model) were used as the actual critical values.

Bootstrapping

Bootstraps[79] of all three datasets analyzed in paper III were performed to investigate the distribution of the estimates of the differential effect parameter. The bootstrap datasets were created by sampling individuals with replacement from the original datasets, as many times as there were individuals in the original dataset, allowing some individuals to be included more than once and some individuals not to be included at all. This procedure was repeated 200 times and analyses of these bootstrap datasets were performed using the differential odds model.

Clinical trial simulations

The developed model for sleep was, apart from simulations to assess the predictive performance, also used for clinical trial simulations. These simulations were performed to assess the effect on the efficacy endpoints of changing the magnitude of dose and time of dosing according to Table 5.

Table 5. *The clinical trial simulations performed using the model developed for sleep.*

Scenario	Dose (mg)	Dosing time (minutes prior to habitual bedtime)
1	25	60
2	25	120
3	50	30
4	75	60
5	75	120

Under the assumption of linear PK characteristics the concentrations following a 50 mg dose of PD 0200390 was calculated as twice the concentrations following a 25 mg dose. The dose dependent magnitude of the step models for drug effects were calculated as a linear interpolation of the magnitudes estimated for 25 mg and 75 mg.

Concentrations, for giving the drug 60 and 120 minutes prior to habitual bedtime were calculated as the concentrations observed when giving the drug 30 minutes prior to habitual bedtime with a shift of 30 and 90 minutes, respectively. The observed concentrations were assumed to only describe elimination 7.5 hours after administration; thus, the individual elimination constant were used to extrapolate the concentrations for the last 30 and 90 minutes of dosing 60 and 120 minutes prior to bedtime, respectively.

Results

Performance of methods for regression (Papers I & II)

The Laplacian regression method performed well in most of the investigated scenarios, with low estimated bias. However, as *(i)* the skewness in the distribution of the responses increased and *(ii)* the variability between the individuals increased an increased bias was observed (Figure 7). Even though using the centering option with the Laplacian method clearly reduced the bias, the bias was still apparent with high variability and large skewness in response distribution, though with reversed sign (Figure 8). The GQ method performed well in all examined scenarios, without bias in parameter estimates (Table 6). No differences in bias or precision were observed when comparing the three different estimation alternatives used within the GQ method: adaptive, non-adaptive scaling and non-adaptive approaches (Figure 8). Using the BSM to estimate the parameters of scenario D and E resulted in stable estimates that were similar to those obtained using the GQ method (Table 6). The imprecision added by the iterative method was on average 35% of the total imprecision with a single estimate but was reduced to 4% using the serial procedure.

Table 6. *Relative bias observed with the different estimations methods; Laplace, Laplace with centering, GQ (adaptive) and BSM, when estimating data from scenario E,* $\omega^2 = 40$.

Scenario	Method	θ_1	θ_2	θ_3	θ_4	θ_5	ω^2
	Laplace	-0.25	-0.21	-0.20	0.37	-0.40	2.03
E	Center	0.24	0.091	0.066	-0.11	0.18	-0.18
$\omega^2 = 40$	GQ	0.007	-0.006	-0.020	0.030	0.002	-0.011
	BSM	0.017	-0.008	0	0.018	0.011	0

Simulations based on the biased population parameters did not produce data with the same marginal distribution of responses as in the observed. Simulat-

ing data using the estimated fixed parameters with the EBE of the ηs improved the predictive performance, especially if the variability was high.

Using the parameters obtained with the GQ method or the BSM for simulations produced data in good agreement with the observed data.

Ignoring assumptions in the proportional odds model (Papers III & IV)

Allowing different covariate effects for the log odds of the categories, by re-analyzing data using the differential odds model instead of the proportional odds model resulted in no improvement of the diarrhoea severity and T-cell receptor density data. These analyses gave a ΔOFV between the proportional and the differential odds model that was larger than the estimated actual critical values needed to conclude a statistically significant difference. Furthermore, the estimated values for the differential effects were one; hence, the proportional odds model was shown to be appropriate to use for these data (Table 7).

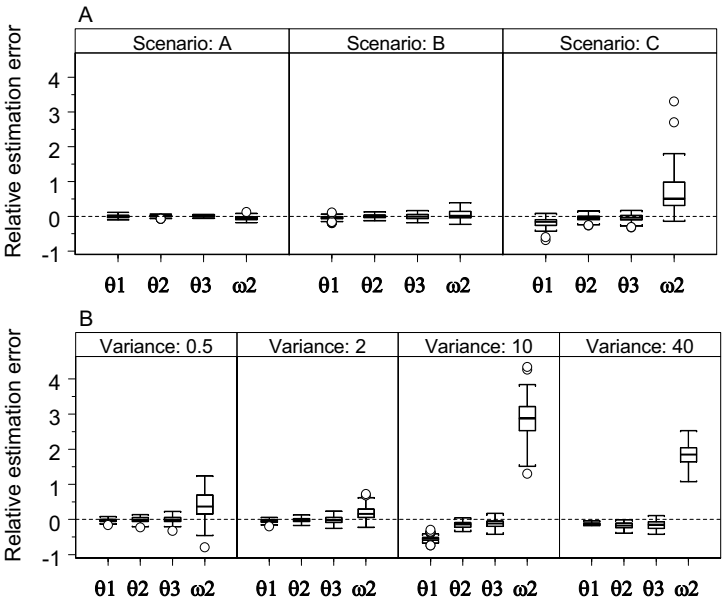

Figure 7. REE for the parameters estimated with the Laplacian method. A. Scenario A, B and C, with $\omega^2 = 4$. B. Scenario C, with $\omega^2 = 0.5$, 2, 10 and 40. θ1, θ2 and θ3 are the intercept parameters and ω2 is the variance. Each box gives the inter-quartile range, with the median indicted as a solid line within the box. The whiskers extend to the maximum and minimum value within 1.5 times the upper and lower quartiles, respectively. Outliers from the whiskers are plotted as circles.

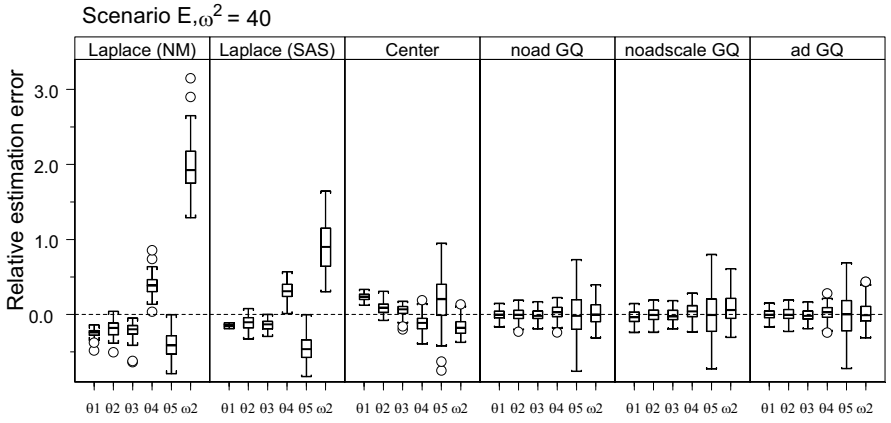

Figure 8. REE for the parameter estimates from scenario E, with $\omega^2 = 40$, estimated with the Laplacian method in NONMEM (NM), in NLMIXED (SAS) and with centering option (Center). REE with the GQ method are also shown with non-adaptive (noad), non-adaptive scaling (noadscale) and adaptive (ad) approaches. θ1, θ2 and θ3 are intercept parameters, θ4 and θ5 describe placebo and drug and ω2 is the variance. Each box gives the inter-quartile range, with the median indicted as a solid line within the box. The whiskers extend to the maximum and minimum value within 1.5 times the upper and lower quartiles, respectively. Outliers from the whiskers are plotted as circles.

Table 7. *Result from analyzing the T-cell receptor density, diarrhoea and sedation data using the differential odds model, with ΔOFV being the difference in OFV between the proportional and the differential odds model. Shown are also the actual critical values to compare the ΔOFV with, assessed using a SHT and the estimates of the differential effects for the different categories for each data.*

Data		**ΔOFV**	**Actual critical value (sign. level p≤0.01)**	**Differential Effects of category 2 / 3 / 4 / 5 / 6**
T-cell receptor density[a]		-4.7	-5.96	1 / *NA* / *NA* / *NA* / *NA*
	Irinotecan	-1.8	-5.70	1 / 1 / 1 / *NA* / *NA*
Diarrhoea	SN-38	-0.5	-6.92	1 / 1 / 1 / *NA* / *NA*
	SN-38G	-0.5	-9.56	1 / 1 / 1 / *NA* / *NA*
Sedation[b]		-131	-9.88	*NA* / 1 / 0.94 / 0.65 / 0

[a] Categories dim / intermediate / bright correspond to 0 / 1 / 2
[b] Definition of sedation categories see Table 2
NA - Not Applicable

However, analyzing the sedation data using the differential odds model gave a statistical improvement, with a ΔOFV much lower than the estimated actual critical value (Table 7). The estimated differential effect for the three highest categories of sedation data revealed that the drug effect for these

categories was not the full magnitude of the step effect estimated with the proportional odds model, which also was supported by the bootstrap distribution of these parameters (Figure 9). The differential effect was for the highest category estimated to zero, that is the probability of observing score = 6 was independent of drug treatment. When expanding on the way the differential effect was implemented, the differential effect was found to vary with concentrations. The final model included a linear and an Emax model describing the concentration effect on the differential effects for category ≥ 4 and category ≥ 5, respectively. The other parameters of the model, such as intercept parameters and placebo effect parameters, were however not altered compared to those estimated using the proportional odds model. Furthermore, the actual critical values corresponding to $p \leq 1\%$ were, for all examined datasets, lower than what is expected when using the number of added parameters as the degree of freedom with a χ^2-distribution (Table 7).

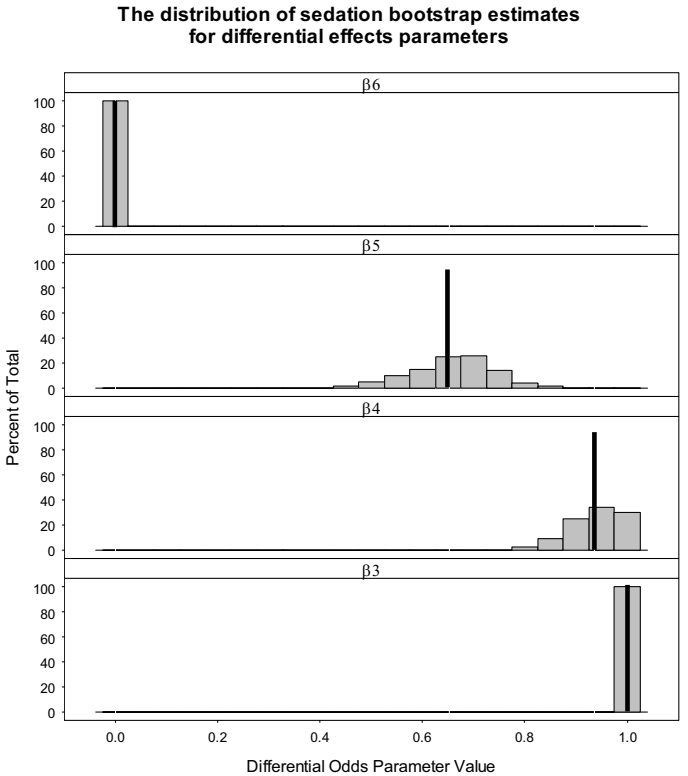

Figure 9. Bootstrap distribution of the differential effect parameters estimated in sedation data. The parameter estimate from the analysis of the observed sedation data is shown as a vertical black line and the distributions are based on 200 bootstrap estimates. No distributions are seen for β3 and β6 as all estimates for these parameters were one and zero, respectively

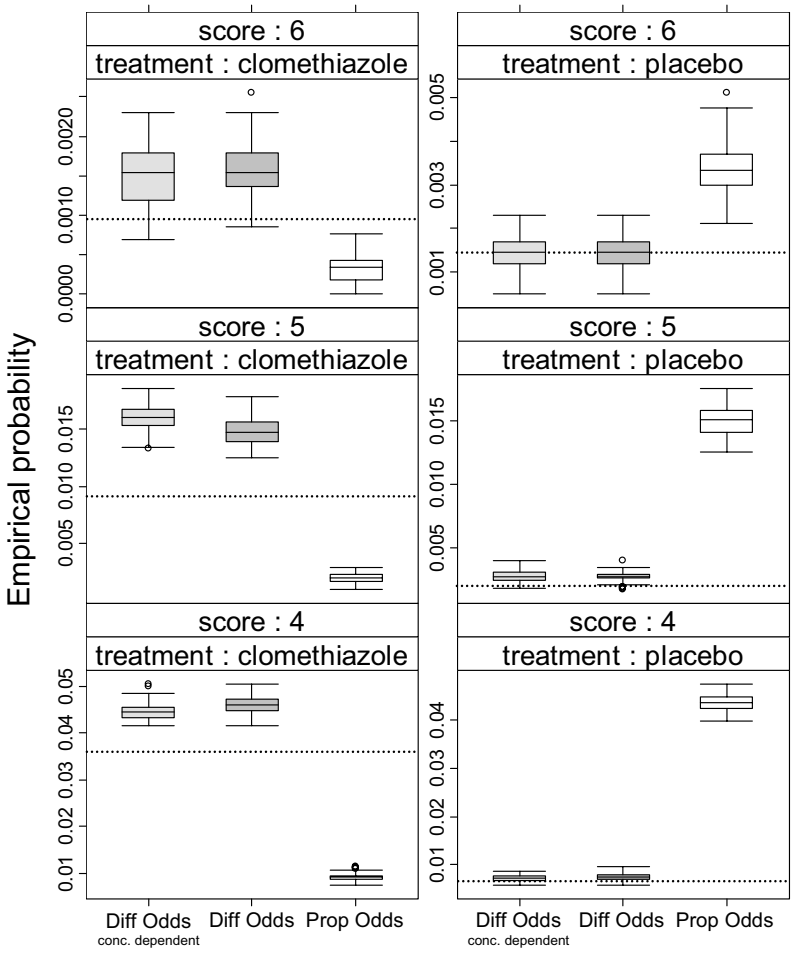

Figure 10. The predicted performance of the proportional (Prop Odds), the differential (Diff Odds) and the concentration dependent differential odds model (Diff Odds conc. dependent) for sedation data shown as the empirical probability for 100 simulations of scheduled, observation independent times, stratified on treatment with the dotted line in each panel showing the corresponding observed empirical probability. Each box gives the inter-quartile range, with the median indicted as a solid line within the box. The whiskers extend to the maximum and minimum value within 1.5 times the upper and lower quartiles, respectively. Outliers from the whiskers are plotted as circles.

Simulating data using the estimated parameters and the scheduled observation times revealed an improvement with both the concentration dependent and independent differential odds model compared to the proportional odds model (Figure 10). As the adaptive design was not fully accounted for by using the scheduled observation times the predictive performance of the

models appears worse than it in reality is and the high discrepancy between the observed and the simulated empirical probabilities was expected for the drug treated group.

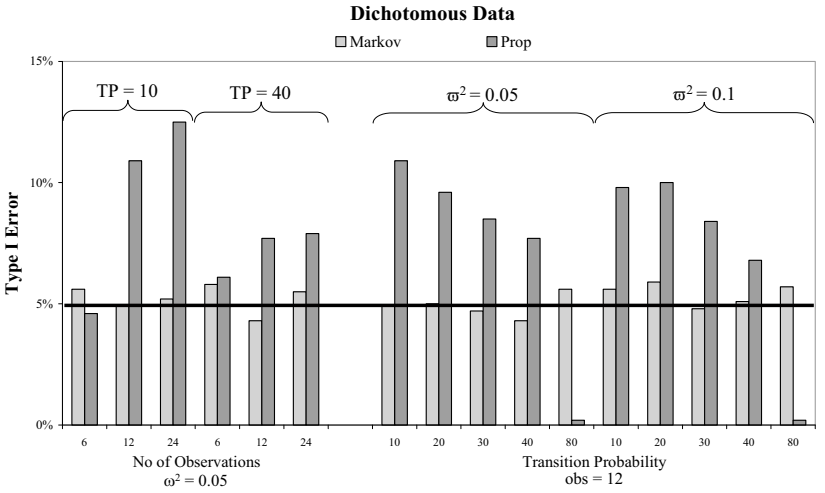

Figure 11. Type I error estimated in the dichotomous data using the Markov (light grey) and the proportional odds model (dark grey). The left part of the histogram shows the effect of changing the number of observations per individual for 2 different transition probabilities (TP): 10% and 40% with the variance, $\omega^2 = 0.05$. The right part of the histogram shows the effect of changing the transition probability with 2 different variabilities, given as variances: 0.05 and 0.1, for 12 observations per individual.

Changing the link function from a logit to the probit and the complementary log-log gave even larger differences in OFV when including the differential effect parameters, -282 and -4959, respectively. The estimates of the differential effects were, however, approximately the same as when using the logit transformation. Thus, changing the link function could not be used instead of the differential odds model to account for non-proportionality on the odds.

Analyzing data as if the observations were independent of each other, by using the proportional odds model, when the observations in fact, are dependent was found to give an increased type I error in most investigated scenarios. The type I error increased with *(i)* increasing number of observations per individual and *(ii)* decreasing transition probability (Figure 11, Figure 12).

The same trends were observed for both di- and trichotomous data. No trends were observed when analyzing the data using the proportional odds model while varying the variance. As a result of too low information content

in the data, the type I error decreased slightly with very low transition probabilities.

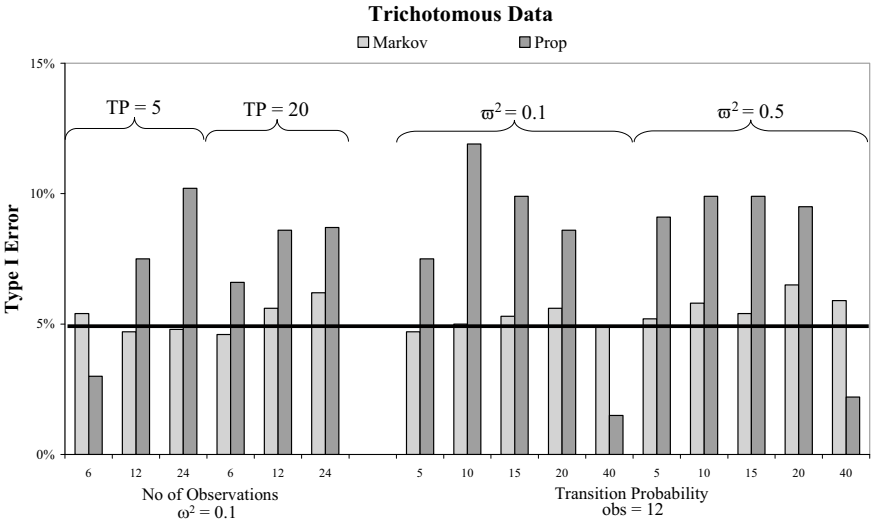

Figure 12. Type I error estimated in the trichotomous data using the Markov (light grey) and the proportional odds model (dark grey). The left part of the histogram shows the effect of changing the number of observations per individual for 2 different transition probabilities (TP): 5% and 20% with the variance, $\omega^2 = 0.1$. The right part of the histogram shows the effect of changing the transition probability with 2 different variabilities, given as variances: 0.1 and 0.5, for 12 observations per individual.

The only observed trend when estimating the data using the correct model (the Markov model) was the expected increased type I error with increasing variance. As the Laplacian method was used for maximizing the likelihood, the parameters will be biased at high variances, with the largest biases seen for the variance and the covariate parameters (paper I); thus leading to an increased risk of finding a spurious covariate relationship.

A reduced model building strategy for sleep stages (Paper V)

The transition probabilities of baseline sleep was characterized as depending on relative bedtime and stage time and a majority of these effects were non-linear, described using three break-points spline functions (Figure 13). The

only linear effect found was on the relative bedtime of TP 3→0. All models included IOV or IIV, but TP 2→5.

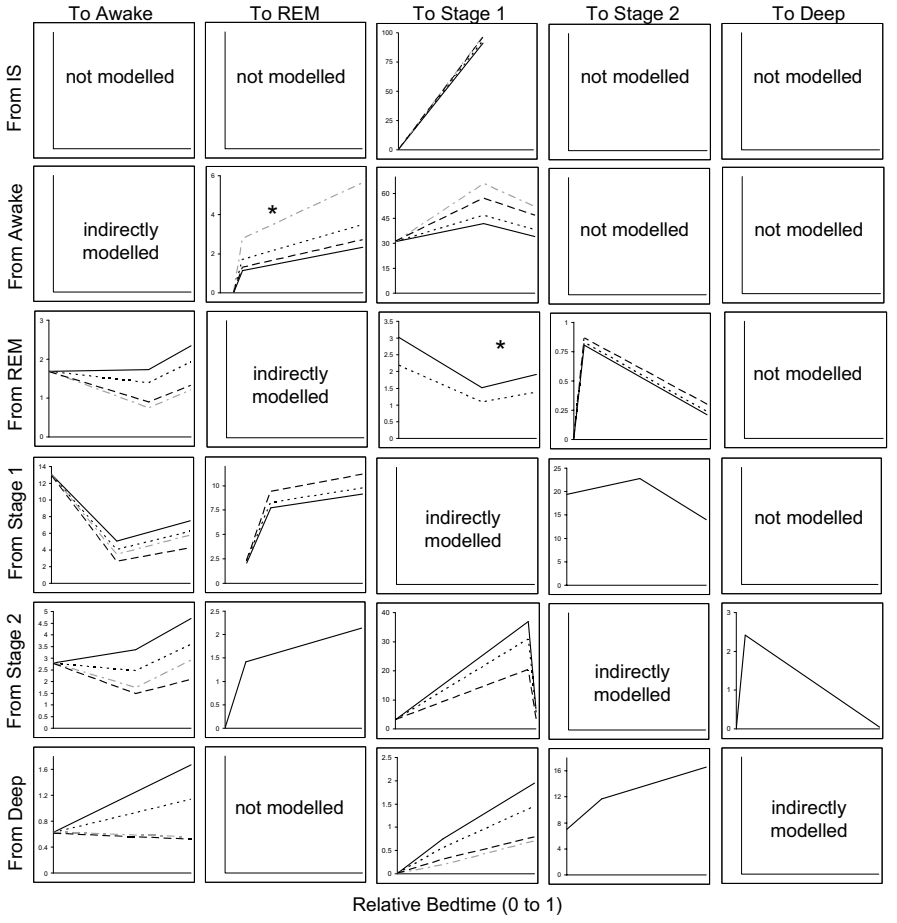

Figure 13. The changes in transition probability (in %) with placebo (black solid line) and drug effects of 25 mg (black dotted line) and 75 mg (black dashed line) of PD 0200390 and zolpidem (grey dot-dashed line). The x-axis shows the relative bedtime, from zero (time of bedtime) to one (end of night), being the same for all panels. Drug effect of PD 0200390 are described with * a step model varying with dose, or if nothing indicated, a linear model with predicted effect compartment concentrations as the covariate. Drug effects of zolpidem are described with a linear model with predicted concentrations as the covariate. IS - initial sleeplessness.

Placebo effects were identified on four transitions, two of which were described with an exponential model; TP IS→1 and TP 1→0, and the other two described using a step model; TP 1→5 and TP 5→1. Only four transitions were unaffected by the drug treatments: TP 1→2, TP 2→3, TP 2→5

and TP 3→ 2. However, TP 2→1, TP 5→2, TP 1→5 and TP 5→1 were only affected by PD 0200390 (Figure 14). Two of the transitions affected by PD 0200390 were described using step models varying with dose: TP 0→5 and TP 5→1. All other drug effects for PD 0200390 were described using a linear model varying with the predicted effect compartment concentrations. The effects of zolpidem were all described using a linear model with the population predicted concentrations as the covariate.

The simulations performed as part of the predictive check showed good agreement with most efficacy endpoints. Only 3 of 72 observed efficacy endpoints were found outside the range of the simulated data: SEF between 6 hours and 8 hours after bedtime for placebo and zolpidem and the NAW for placebo. These endpoints were slightly underpredicted by the model. The figures of average time spent in the different stages, presented in Figure 15, showed that the main trends seen in the observed data was captured.

Three study designs were simulated using the developed model: 50mg PD 0200390 given 30 minutes prior to bedtime and 25 mg and 75 mg PD 200390 given 60 and 120 minutes prior to bedtime. The efficacy endpoints that were the most affected by changing the study design, compared to the actually performed clinical trial, were all latency parameters. Consequently sleep efficiency between 0 and 2 hours after bedtime and total sleep time were also increased. Changing the time of dosing to earlier prior to bedtime had the most impact, even relative to doubling the dose (25 to 50 mg).

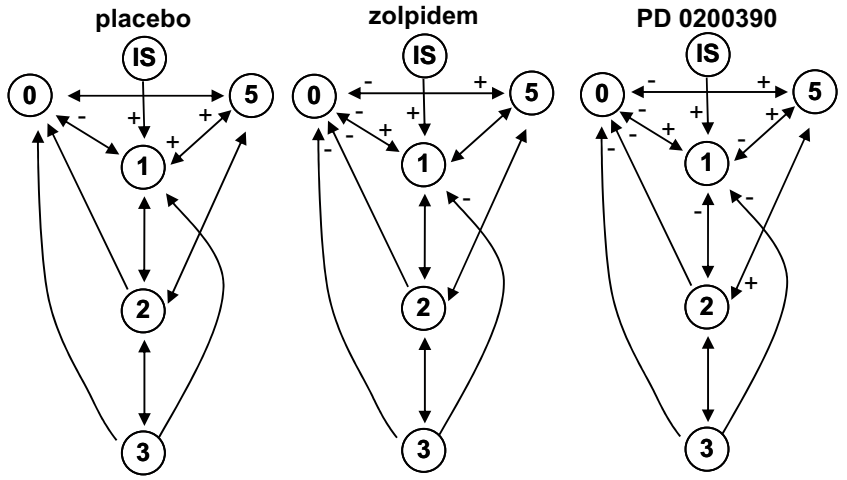

Figure 14. The placebo effects (left panel), the drug effects of zolpidem (middle panel), and PD 0200390 (right panel) found when developing the sleep model using the reduced model building strategy. A plus sign indicates an increased transition probability while a minus sign indicates a decreased transition probability, as a result of the treatment. Definitions of the abbreviations referring to the sleep stages see *Figure 5*.

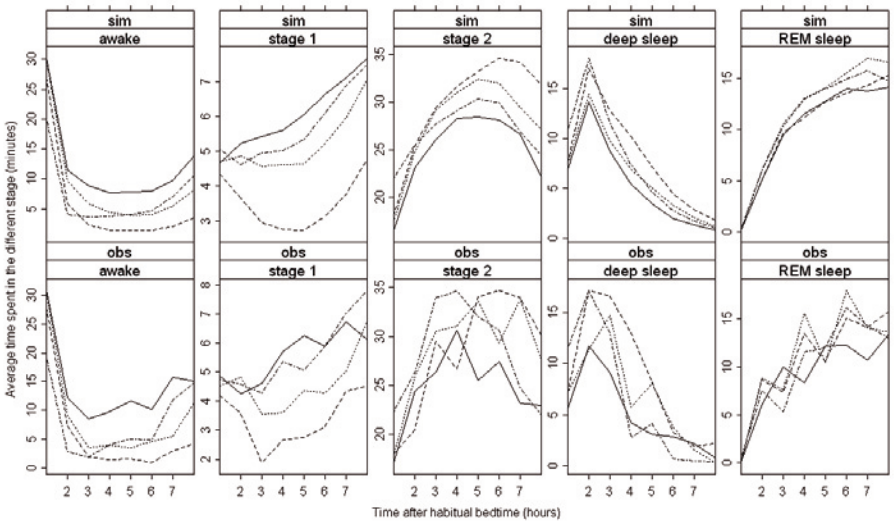

Figure 15. Average time in minutes spent in the different stages, awake (1st column), stage 1 (2nd column), stage 2 (3rd column), deep sleep (4th column) and REM sleep (5th column) at different times after habitual bedtime in hours for different treatments; placebo (solid line), 25 mg of PD 0200390 (dotted line), 75 mg PD 0200390 (dashed line) and 10 mg of zolpidem (dot-dashed line) in the observed (lower row) and in 25 of the simulated datasets (upper row).

Discussion

The aim with this thesis was to investigate the performance and to improve the use of models and methods for mixed-effects population PK-PD analysis of categorical data. This was done by drawing attention to the inability of the Laplacian method to produce unbiased parameter estimates if *(i)* the variability in the data is large or *(ii)* the distribution of the reported responses between the categories is skewed. Solutions to this problem have been suggested in this thesis by using the GQ method instead of the Laplacian method for regression and through the development of the BSM for users of software where the Laplacian method is the only available regression method for categorical data analysis. The BSM provides the unbiased parameter estimates given that the origin of the bias is the method of regression. The performance of proportional odds model has also been investigated. This model was shown to be unsuitable for analysis of data arising from a heterogeneous ranking scale, measuring an effect with several underlying causes. The differential odds model was developed to provide a solution to the problem and it was shown to be an improvement, with regard to statistical significance as well as predictive performance, over the proportional odds model for such data. The appropriateness of the LRT was investigated for situations when the proportional odds model is erroneously used for data with dependent observations. Both an increased and a decreased type I error was observed when the observations are dependent; thus assessing the actual type I error is in such situations wise in order to verify that the appropriate significance level is used. An alternative strategy is to modify the model to handle the dependence adequately. One such possible modification is to incorporate Markov elements. In the case of polychotomous data, such incorporation can involve considerable complexity. A strategy for the reduction of the time-consuming model building with the Markov model and sleep data has been suggested. The strategy aimed at increasing the use of models for sleep data in a clinical drug development setting, where time consumption of an analysis is a crucial issue.

The problems seen with the Laplacian method with data with a skewed response distribution or high variance are related to η shrinkage. The shrinkage in the EBE of η's is about 55% in the most skewed data with the highest variance, calculated by dividing the standard deviation of the EBE of η's

with the estimated ω[80]. As seen in Figure 16 the η shrinkage is also asymmetrical. Two alternative methods handling the bias arising when using the Laplacian method, are presented in this thesis: the GQ method and the BSM. The incidence/severity model is another solution to the same problem provided by Kowalski *et al*[81]. In this model, the regression is divided into two steps; the incidence of responders is first analyzed, followed by assessing the severity among the responders. The model for incidence is considering only one observation per individual, the observation of responder or non-responder. Consequently, the model for incidence is not implemented using random effects to describe the IIV and thus, avoiding the estimation problem with the Laplacian method. By analyzing only the responders in the second step, the skewness in the distribution of responses is reduced and thus the risk of getting biased parameter estimates even though the Laplacian method is used for the regression. Although the incidence/severity model is a simple, yet elegant solution to the problem, it suffers from some inherent deficiencies. As the incidence is dependent on the time interval for which it is assessed extrapolation of the model to shorter or longer time intervals is not possible. Data arising from dose escalation studies cannot be analyzed properly using this model as only one observation per individual is considered in the incidence model and for the same reasons implementation of time varying covariates is not feasible.

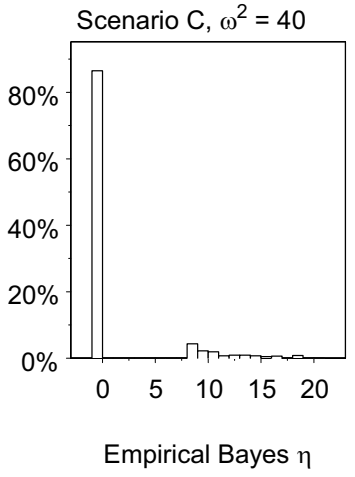

Figure 16. The distribution of the EBE of η's following estimation of data with a skewed distribution of responses (90/5/3/2) and high variance ($\omega^2 = 40$) using the cumulative odds model with random intercept.

PD models for categorical data are usually estimated in conjunction with a larger PK model. The PK and PD models should optimally be fitted simulta-

neously[82]. This is feasible as most models for categorical data analysis are quite fast to run, as the number of random effects parameters defined in the model usually is low. NONMEM has an extensive PK library and a superior implementation of dosing history compared to other software, making NONMEM the preferred software for mixed-effects PK modelling. The BSM was thus developed to give users of NONMEM the alternative not to change the software, but still get an estimate of the unbiased parameters. Noteworthy, if bias in estimates arises from model misspecification, such as censoring, the BSM cannot be used for obtaining the unbiased parameter estimates; as such bias is unrelated to the method of regression. However, the BSM can be used to reduce the bias in parameter estimates arising from the method of regression even thought the data are non-categorical. An example would be bias in parameter estimates due to the use of the FO method with rich data[83].

A meta-analysis of sedation was in paper III re-analyzed using the differential odds model. Separating the natural course of sedation in stroke patients from the sedative effect of clomethiazole was one of the aims with the previously published analysis performed by Zingmark *et al*, 2003[25]. A quite advanced model for placebo using time as the covariate was developed, with a mixture model dividing the patients into a sedation sensitive and non-sensitive population. Even though this complex model was used to describe the time course of stroke-induced sedation, no concentration effect of clomethiazole on sedation could be found when applying the proportional odds model and the drug effect was described with a step model. A concentration effect of clomethiazole on sedation could nevertheless be identified when the differential odds model was applied to the data. Describing the disease progression in detail could in this case not compensate the erroneous assumptions made with the proportional odds model. This shows the importance of applying the appropriate model in order to find all relevant covariate relationships.

The realized design cannot be used to assess the true predictive performance of a model developed using data from an adaptive design, as the design is dependent on the observations. The simulated observations will not be identical to the observed and hence the adaptive rules set up for the design will be violated. Using the scheduled, observation-independent measurement times instead of the full realized design will account for an adaptive design with regard to the monitoring. However, this will not account for an adaptive dosing schedule. The sedation data analyzed in paper III came from studies where both an adaptive monitoring and dosing schedule was utilised and Figure 10 illustrates the predictive performance of the differential and proportional odds model in sedation data using only the scheduled observations accounting for the adaptive monitoring with good predictive performance of

the placebo treated group as the result. The inability of this approach to account for the subjective adaptive dosing schedule is seen in the predictive performance of the drug treated group, where the only category with a good predictive performance is score = 6, which is independent of drug treatment. There are two approaches to account for an adaptive dosing schedule. The first is to simulate the largest possible design, that is with all possible dosing levels at all possible dosing times and then "throw away" data according to the adaptive rules. As the re-starting of infusion was an optional decision for the treating physician in the sedation studies in paper III, this was however not feasible. Another approach is to develop a model describing the adaptive rules, similar to a dropout model. In such model, randomness could be included for the optional re-start of infusion and this randomness would be based on the observed empirical probability of re-start. This "study adaption" model could be fitted simultaneously with the model describing the sedation, reducing any possible risk of bias in parameters due to the adaptive study design.

One of the assumptions made when using the proportional odds model is that the observations are independent of each other. Failing to recognise that the observations in a dataset are dependent, treating them as independent by using the proportional odds model, will most likely affect the LRT with higher type I errors than expected; thus with a increasing risk of including a false effect as the consequence. Assessing whether the observations are dependent, with a transition probability that is less than what is assumed with the proportional odds model, can easily be done by comparing the number of transitions in the observed and the simulated data using the proportional odds model. If there are more transitions made in the simulated data compared to the observed data, this suggests that the observations are dependent. If the observations instead are changing for each new measurement, then maybe it would be prudent to apply a Markov model to investigate whether the observations have a transition probability that is larger than assumed with the proportional odds model. However, finding an effect that would render transition probabilities larger than the level assumed with the proportional odds model seems unlikely.

Additional knowledge, of what effects a drug has on different sleep stages over time, is provided when using a mixed-effects model to explore longitudinal sleep data. Assessing the effect the drug has on the efficacy endpoints, such as LPS and SEF, will only give a crude indication about the underlying drug effects, as these endpoints do not consider the time effect nor provides an understanding of on what stages the drug has its effects. In contrast, mixed-effects Markov models provide which transition probabilities that are changed due to drug treatment as well as the time aspect of these effects. Another advantage of model based analysis of sleep data is that new dosing

regimens can be simulated and used for guidance in designing new studies. In addition, simulations may be performed to predict the outcome of compounds with similar mechanisms of action but a different PK-PD profile, for example a new drug with different potency, or a controlled-release formulation. With the use of clinical trial simulations, different scenarios, such as alternative doses and dosing schedules, can be explored and the results of the clinical trials can act as a support for decisions made in later stages of drug development. A reduced model building strategy has in paper V been suggested, decreasing the number of models to test, but also the run-times by using only parts of the data when defining new covariates relationships. The results of the reduced model building strategy also suggest that the number of predefined models for the relationship between transition probability and relative bedtime and stage time could be reduced in number. Since no relationships between transition probabilities and relative bedtime were described using time-constant models, this step could possibly be removed to reduce the modelling even further. Thus, the suggested reduced model building strategy facilitates the use of these models within drug development, making the advantages provided by mixed-effect modelling readily feasible to use.

The importance of choosing a suitable regression method and the significance of choosing an appropriate model has been addressed in this thesis. On the basis of the results, the following recommendations for mixed-effects categorical data analysis are suggested. Some decisions on appropriate model and suitable method of regression can be made based on simulations using the parameter estimates from an initial fit of the cumulative odds model with random intercept to the data. Fitting the same model as used for the simulations to the simulated data and compare the parameter estimates of the observed and simulated data can be done to decide if the Laplacian method is a suitable method of regression. If the parameter estimates from the fit to the simulated data differs from the corresponding estimates of the observed data, the Laplacian method is an unsuitable method of regression. In these cases, either the method of regression could be changed to the GQ method or the BSM could be used with the final model to provide the unbiased parameter estimates. If the latter suggestion is chosen, it is wise to assess the actual critical value for the LRT through a SHT in order to be confident of the significance level used for inclusion of new parameters. A decision on the proportional odds model vs. the Markov model can be made based on the comparison of the number of transitions in the simulated and observed data. If these numbers are dissimilar, the Markov model is more appropriate to use than the cumulative odds model with random intercept for the analysis. If that is the case, the principle of pre-defining a set of models to be tested with the Markov model, as suggested in paper V, can be utilised

to reduce the model building. If the opposite is concluded and covariates such as placebo or drug effects are added to the model, other link functions than the logit as well as allowing differential odds to be estimated should be investigated to verify that the covariate effects are best described using proportional odds. If the covariate functions of the model are found to be different for different scores the actual critical values to use for inclusion of new parameters should be assess using a SHT. If these recommendations are followed, the risk of including false or excluding true covariates, as well as the risk of over- or underestimating the magnitude of these covariate relationships is reduced.

Perspectives

The use of PK-PD modelling has increased in the drug development, ever since the FDA published recommendations on how to perform population PK modelling[84]. Also, in 2004, the FDA published a paper reviewing some ways that pharmacometrics can be implement to make the drug development process more efficient, including the concept of model-based drug development with drug/disease models that facilitate clinical trial design and optimal dosing[85]. The PK-PD models are now increasing being used in the discussion with regulatory authorities. In addition, the interest for modelling categorical data is increasing, as many efficacy endpoints and side-effects are measured using categorical or ranking scales. More aspects of these models are likely to be explored in the future, as these models will be applied to new therapeutic areas and other measurements than has previously been done. As a natural consequence of this, the models used, as well as the different estimation methods used with these models, will have to be further investigated. The use of the Laplacian method has been shown to produce biased parameter estimates with count data arising from epileptic seizures, in cases where the Poisson, the zero-inflated Poisson and negative binomial model has been used[86]. The use of the GQ method for these data with these models has been shown to perform well without biased parameters; hence, the problem for these models with the Laplacian method seems to be the same as with ordinal data.

The use of mixed-effects models with so-called hidden Markov models is also an interesting focus for future studies. The observations of count data are commonly seen clustered with observations of non-zeros followed by long periods only observing zeros or few counts. The hidden Markov models could for such data improve the analysis, as periods with few or zero counts could be modelled as an inactive state of the disease and periods with many counts as the active state. An example of an analysis using a hidden Markov

model has been published for migraine data[87]. Random effects were however not incorporated into that model.

As computational complexity, being accurately handled by the computers reaches new levels so does the complexity of the models, taking more than the measurements themselves into account. One example is dropout models, which is an area where the research just has started. If the dropout in a clinical study is high, and the reason for dropout is related to the effects of the drug or the disease, analyzing the data without taking the dropout into account may lead to biased parameters, in analogy with below limit of quantification data for individuals with high CL[88]. An example of a situation where the dropout is not completely random is stroke studies, where the dropout most likely is a deterioration of the disease as most patient drops out due to death. Two analyses of stroke have been presented[89, 90] where a model for the dropout was used considering the dropout as non-random, providing additional information to other parts of the model. To develop models taking the expected dropout rate in a population into account would of course also give better predictions using clinical trial simulations.

Nestorov *et al*[29] showed that optimal design theory can be applied with categorical data using fixed effects. However, optimal design for categorical mixed-effects models is yet to be investigated. In optimal design, the uncertainty of the estimated parameters is optimised with regard to dosing, sampling times, etc. and to use this approach to provide better clinical studies early in the drug development program requires knowledge of optimal design for categorical data as the toxicity of a drug is the central concern and most adverse events are measured using categorical scales.

Conclusions

The Laplacian method was shown to be an inappropriate method to use for regression of models for ordinal data when *(i)* the responses are skewly distributed between the categories or *(ii)* the inter-individual variability is high. The Gaussian quadrature method can instead be used in those situations. The unbiased parameter estimates can also be obtained using the back-step method. This iterative new method was developed for analysts that do not have access to software with an appropriate regression method and it was shown to produce stable estimates, without adding much imprecision.

Using the proportional odds model to analyze data with several underlying variables affecting the response may produce models with poor predictive performance. A new alternative model, the differential odds model, was developed. The differential odds model was shown to be superior to the proportional odds model for data with several underlying variables affecting the response, as shown by the likelihood ratio test and the predictive performance of the models.

The likelihood ratio test was shown to be affected when analyzing dependent observations using the proportional odds model, which assumes independent observations. Increased type I error was observed for data with *(i)* a large number of observations or *(ii)* a low transition probability between two observations and a decreased type I error was observed when the transition probability was lower than that assumed with the proportional odds model. The likelihood ratio test was shown to perform well for the same data when analyzed using the Markov model.

The Markov model was used to describe sleep data, including placebo effects and effects of two drugs for the treatment of insomnia. A reduced model building strategy was suggested and shown successful in producing a model that well predicted the observed data.

Populärvetenskaplig sammanfattning

Att utveckla läkemedel är dyrt och studier som görs på människor, s k kliniska prövningar, är ofta den mest kostsamma delen i läkemedelsutvecklingen. Genom att förbättra den statistiska analysen av resultaten från de kliniska prövningarna går det att minska kostnaden för ett läkemedel, vilket är värdefullt för läkemedelsföretagen, patienter och samhället.

En viktigt del i de kliniska prövningarna är att avgöra om det undersökta läkemedlet har någon effekt på sjukdomssymptomen och hur den effekten varierar med läkemedelsdosen. Detta kallas farmakodynamik. Symptomen beskrivs ofta med ett begränsat antal svarsalternativ; s k kategoriska svar. Smärta är exempel på en symptom. En passande fråga är ”hur ont har du?” och patienten får svara ett av följande alternativ: inget, lite eller mycket ont.

Inom läkemedelsindustrin analyseras ofta de kategoriska svaren genom att jämföra svaret före patienten fick läkemedlet med svaret efter patienten fick läkemedlet. Genom att istället beskriva farmakodynamiken med statistiska ekvationer, s k modeller, kan alla svar som erhållits användas, vilket gör det lättare att upptäcka en läkemedelseffekt. Populationsmodeller är en speciell typ av modeller där man dels får en uppskattning av en övergripande effekt, men även en uppskattning på hur mycket patienterna avviker från den övergripande effekten. Dessa effekter kallas populationsestimat. Ett problem med populationsmodeller är att det endast finns ett fåtal modeller att använda för kategoriska svar samt att erfarenheten av dessa är begränsad.

Traditionellt, används mjukvaran *NONMEM* och den proportionella odds modellen för att analysera kategoriska svar med populationsmodeller. I mjukvaran används en metod som heter *Laplace* för att bestämma populationsestimaten. Om patienterna i studien har givit mycket olika svar eller om endast ett fåtal uppgett att de till exempel har mycket ont bör man istället använda *Gaussian quadrature* metoden som finns i mjukvaran *NLMIXED* (delarbete I). Används *Laplace* metoden blir populationsestimaten inte korrekta och risken finns att läkemedelseffekten uppfattas högre än den i verkligheten är.

Vi har dock utvecklat en metod för de analytiker som använder *NONMEM* som ger de korrekta värdena trots att *Laplace* används i denna mjukvara. Metoden kallas *the Back-Step Method* (delarbete II).

För vissa symptom kan det finnas mer än en underliggande orsak. Med proportionella odds modellen antas att läkemedlet har samma effekt på alla svaralternativ oavsett om det är flera orsaker till symptomen. Detta är inte nödvändigtvis sant och därför utvecklades en ny modell, differentiella odds modellen, som tillåter läkemedlet att ha olika effekter på olika svarsalternativ (delarbete III). En jämförelse mellan proportionella och differentiella odds modellerna visade att den senare är bättre för att beskriva trötthet, vilket var ett symptom där flera underliggande orsaker fanns.

Om en patient ofta får samma fråga kommer patienten troligen svara samma sak. Svaren är då beroende av varandra. Den proportionella odds modellen antar att svaren är oberoende av varandra. Vanligen när en populationsmodell byggs används ett statistiskt test som kallas likelihood kvottestet (*LRT*). Med hjälp av *LRT* är det möjligt att statistikt bestämma risken för att ta en läkemedelseffekt för relevant när den inte är det. I delarbete IV undersökte vi vad som händer med *LRT* när beroendet mellan svar ignoreras och vi fann att det fanns en ökad risk för att ta en falsk effekt för sann då beroendet mellan svaren var starkt eller antalet svar per person var stort. Markov modellen är en alternativ modell som tar hänsyn till beroende mellan svar. Risken för att ta en falsk effekt för sann med Markov modellen visade sig vara helt enligt förväntningarna.

Inom läkemedelsindustrin är det vanligt att de analyser som utförs för nya läkemedel måste vara tidseffektiva, eftersom varje dag som passerar utan att läkemedlet kommit ut på marknaden innebär stora kostnader för företaget. Läkemedel för sömnlöshet är ett forskningsintensivt område som potentiellt kan ge stor avkastning, men modellbaserad analys av sömn är tidsödanade. I delarbete V presenterar vi en effektiviserad strategi för hur Markov modellen kan användas för att beskriva sömn och effekter av läkemedel på sömn. Den effektiviserade strategin gav en modell som väl beskrev sömn för patienter som lider av sömnlöshet samt effekterna av två olika läkemedel för behandling av sömnlöshet.

Jag har i den här avhandling givit förslag på hur användandet av farmakodynamiska populationsmodeller för analys av kategoriska svar kan förbättras. Genom att använda rätt metod och rätt modell för olika typer av svar går det att minska risken för att missa en sann eller acceptera en falsk effekt, samt att överskatta eller underskatta effektens storlek. Jag har även givit förslag på hur byggandet av modeller när det finns många svar kan effektiviseras.

Acknowledgements

This thesis was carried out at the Division of Pharmacokinetics and Drug Therapy, Department of Pharmaceutical Biosciences, Faculty of Pharmacy, Uppsala University, Sweden.

Pfizer Central Research, Sandwich, UK, is gratefully acknowledged for financial support for year 2001 and 2002.

I would like to express my many thanks to all who helped with this thesis:

Prof. *Mats Karlsson*, my super supervisor. Thanks for your inspiring passion for pharmacometrics and you enthusiasm to share all your astounding knowledge. You have a way to encourage by making everything seem so easy! Thanks for brilliant supervision with the perfect balance between support and independence.

Dr. *Siv Jönsson*, co-supervisor and co-author. Thanks for answering all my basic questions about PK-PD modelling and for all cheerful calls. Thanks also for reading my thesis and papers. You are the best reviewer ever!

Prof. *Margareta Hammarlund-Udenaes*, head of the division. Thanks for caring for us all and shipping me to Sheffield, England where it all started.

Supervisors of my MSc project, Prof. *Geoff Tucker* and Prof. *Amin Rostami-Hodjegan*, Thanks for inspiring me to continue within the PK-PD area.

Thanks to co-authors for their input: Dr. *Per-Henrik Zingmark,* for sharing the frustration of publishing paper III. Dr. *Niclas Jonsson,* for witty comments and your excellent steering during your time as the captain of the division. *Hanna Silber*, beloved friend, for writing most of paper IV. I am really looking forward to have you at a close distance once again! Dr. *Daniele Ouellet*, Dr. *Brian Corrigan* and Dr. *Raymond Miller*, for your enormous patience as the analysis was delayed due to the arrival of Emma. Many thanks also to *Baerbel Wittke* for taking the time to teach me SAS, Prof. *Nick Holford* for providing housing and inspiring and challenging dinner conversations in New Zeeland and all reviewers of my papers as you have both spurred me to do better and given me valuable comments.

Dr. *Lasse Lindbom*, *Pontus Pihlgren* and *Kajsa Harling*, shepherds of the cluster and PsN. I am not exaggerating saying that this thesis could not have been made without you. Thanks! Sorry for crashing the cluster occasionally… Thanks also to joe, frej, gandalf, and all the others! Thanks, *Magnus Jansson*, for all posters you have printed with unnecessary tight deadlines!

Al Maloney. Thanks for putting my thesis under your PD-PK unbiased magnifying glass and providing helpful comments as a "true" statistician and Englishman.

Kristin, roommate and closest friend. Thank you for being such a support during our undergraduate studies and a rock during our PhD studies. I would not have finished either of them if you had not been there to set my focus when astray. How will I manage without "Tott" at work? Or are you "Knoll"?

Bettan, roommate and confidant. Thanks for constantly letting me bother you with strange questions without being annoyed and for listening when I need to do my logical thinking aloud. You are a truly patient person!

Roommates, *Emma H* and *Ruija*. *Emma* thanks for answering all my PK questions. *Rujia* thanks for teaching me how to start a NONMEM run.

Dr. *Emma Bo*. Thanks for all the times you entered my office with the intent of having some PK discussion, and ended up talking about something completely different. I really enjoyed those breaks from the research.

Elodie and Dr. *Rada*. Thanks for sharing your research with me; it has been a pleasure working with the both of you. Merci! Hvala!

All present and former, internal and external colleagues: *Agneta, Anders G, Anders V, Andy, Angelica, Anja, Anna A, Finkan, Anna T, AnnaSofia, Ann-Catrine, Annika, Ann-Marie, Anubha, Bengt, Birgit, Brigitte, Britt, Christoffer, Christine, Daniël, Doaa, Emma L, Erika, Eva H, Eva V, Fredrik, Grant, Gunnel, Hanna M-C, Jacob, Jakob, Jessica, Jocke, Johan, Jonatan, Jonna, Joy, Justin, Rögen, Kajsa Ham, Karin Tj, Karin Tu, Katti, Kjell, Klas, Kristin C, Lena, Leslie, Lotfi, Ma, Madeleine, Malin, Maria, Marianne, Astra-Mariex2, Marina,* Astra-*Marc/kus*x2*, Martin, MaMa, Paul, Petra, Poppi, Rikard, Rocio, Stefanie, Stina, Sven, Ulrika G, Ulrika S, Ulrika W* and of course all guest researchers. Thank you for interesting lunch discussions, fun travelling and conferences, unplanned office happenings, go'fika and crazy Christmas parties with ormsnaps. I do not think I will find another work place making me wake up in the morning thinking: "Hurra, I get to go to work today!!!"

Till mina vänner och min familj som alla på sitt sätt hjälpt mig med avhandlingen.

Tack till ALLA vänner för att ni ger mig annat att tänka på än forskningen. Speciellt, *Jesper*, *Therese*, *Anton* och *Hanna*, för Happy Friday och roliga semestrar, *Kristina*, *Svante* och *Emelie*, för trevliga simturer och maratonfikor. *Steffi*, *Stefan*, *Greta* und *Jasper*, unsere Nachbarn. Wir fühlen uns wie zu Hause in Freiburg!

Marianne, Kjell, Mie, Jörgen, Arvid, Kristine och Olof. Tack för den avslappning skaraborgsvistelserna ger. Speciellt tack till *Marianne* för all barnpassning och alla trevliga pratstunder vid köksbordet i prästgår'n.

Mamma och *Pappa*, tack för att ni uppmuntrat mig till att bli en vetgirig och frågvis person. *Mamma*, tack för att du alltid uppmanat mig lita på mina kunskaper och för allt ditt stöd, speciellt när det gäller mina utbildningsval. *Pappa*, tack för intresset att förstå vad jag gör och för alla inpirerande naturvetenskapliga diskussioner.
Min stora syster *Åsa,* tack för att du lockade mig till Uppsala och för dina försök att få min naturvetenskapliga hjärna att tänka utanför sina ramar. Jag tror inte alla försök var förgäves... *Anders*, tack för alla recept på trimetoprim och trevliga pratstunder om forskning... *Lova* och *Sara*, ni är så vansinnigt goa.
Mina småsyskon *Micke*, *Malin* och *Matilda*. Tack för att ni påminner mig om hur det är att vara 24, 16 och 14 igen. Det ger mig perspektiv på mitt livet! *Åsa*, tack för din aldrig sinande gästfrihet.

Mina älskade barn *Emma-Pemma* och *Gustav*. Tack för att ni finns och tack för er inspirerande upptäckarglädje. Det är fantastiskt att få komma hem till två så mysiga ungar när man haft en stressig dag på jobbet. Ni får all irritation, stress och frustration att bara rinna av. Emma, visst blev det en fin rosa bok!

Sist och mest, *Martin*, min älskade *Martin*. Utan dig hade jag aldrig haft tiden eller lugnet att få den här avhandlingen klar, även om jag nog hade avslutat mina civilingenjörsstudier lite snabbare. Tack för att du är mitt stöd, mitt cheer-leading team, min clown, mitt samvete, min avkoppling, min gossefilt, min coach, min älskling... För att citera Spice Girls: *You spice up my life!*. Puss puss puss


Mia

References

1. Stevens SS (1946) On the Theory of Scales of Measurement. *Science* **103**:677-680.
2. Peck CC, Barr WH, Benet LZ, Collins J, Desjardins RE, Furst DE, Harter JG, Levy G, Ludden T, Rodman JH and et al. (1992) Opportunities for integration of pharmacokinetics, pharmacodynamics, and toxicokinetics in rational drug development. *Pharm Res* **9**:826-833.
3. Sheiner LB, Beal SL and Dunne A (1997) Analysis of Nonrandomly Censored Ordered Categorical Longitudinal Data from Analgesic Trials. *J Am Stat Ass* **90**:1235-1255.
4. Sheiner LB and Beal SL (1980) Evaluation of methods for estimating population pharmacokinetics parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data. *J Pharmacokinet Biopharm* **8**:553-571.
5. Sheiner LB (1984) The population approach to pharmacokinetic data analysis: rational and standard data analysis methods. *Drug Metab Rev* **15**:153-171.
6. Steimer JL, Mallet A, Golmard JL and Boisvieux JF (1984) Alternative approaches to estimation of population pharmacokinetic parameters: comparison with the nonlinear mixed-effect model. *Drug Metab Rev* **15**:265-292.
7. Karlsson MO and Sheiner LB (1993) The importance of modeling interoccasion variability in population pharmacokinetic analyses. *J Pharmacokinet Biopharm* **21**:735-750.
8. Karlsson MO, Beal SL and Sheiner LB (1995) Three new residual error models for population PK/PD analyses. *J Pharmacokinet Biopharm* **23**:651-672.
9. Karlsson MO, Jonsson EN, Wiltse CG and Wade JR (1998) Assumption testing in population pharmacokinetic models: illustrated with an analysis of moxonidine data from congestive heart failure patients. *J Pharmacokinet Biopharm* **26**:207-246.
10. Armstrong BG and Sloan M (1989) Ordinal regression models for epidemiologic data. *Am J Epidemiol* **129**:191-204.
11. Stromberg U (1996) Collapsing ordered outcome categories: a note of concern. *Am J Epidemiol* **144**:421-424.
12. Sankey SS and Weissfeld LA (1998) A study of the effect of dichotomizing ordinal data upon modeling. *Communication in Statistics - Simulation and Computation* **27**:871-887.
13. Agresti A (2002) *Categorical Data Analysis.* Wiley, Hoboken, NJ, USA
14. Sheiner LB (1994) A new approach to the analysis of analgesic drug trials, illustrated with bromfenac data. *Clin Pharmacol Ther* **56**:309-322.
15. Gupta SK, Sathyan G, Lindemulder EA, Ho PL, Sheiner LB and Aarons L (1999) Quantitative characterization of therapeutic index: application of mixed-effects modeling to evaluate oxybutynin dose-efficacy and dose-side effect relationships. *Clin Pharmacol Ther* **65**:672-684.

16. Gomeni R, Teneggi V, Iavarone L, Squassante L and Bye A (2001) Population pharmacokinetic-pharmacodynamic model of craving in an enforced smoking cessation population: indirect response and probabilistic modeling. *Pharm Res* **18**:537-543.
17. Lunn DJ, Wakefield J and Racine-Poon A (2001) Cumulative logit models for ordinal data: a case study involving allergic rhinitis severity scores. *Stat Med* **20**:2261-2285.
18. Mould D, Chapelsky M, Aluri J, Swagzdis J, Samuels R and Granett J (2001) A population pharmacokinetic-pharmacodynamic and logistic regression analysis of lotrafiban in patients. *Clin Pharmacol Ther* **69**:210-222.
19. Mould DR, Holford NH, Schellens JH, Beijnen JH, Hutson PR, Rosing H, ten Bokkel Huinink WW, Rowinsky EK, Schiller JH, Russo M and Ross G (2002) Population pharmacokinetic and adverse event analysis of topotecan in patients with solid tumors. *Clin Pharmacol Ther* **71**:334-348.
20. Johnston SR, Hickish T, Ellis P, Houston S, Kelland L, Dowsett M, Salter J, Michiels B, Perez-Ruixo JJ, Palmer P and Howes A (2003) Phase II study of the efficacy and tolerability of two dosing regimens of the farnesyl transferase inhibitor, R115777, in advanced breast cancer. *J Clin Oncol* **21**:2492-2499.
21. Xie R, Mathijssen RH, Sparreboom A, Verweij J and Karlsson MO (2002) Clinical pharmacokinetics of irinotecan and its metabolites in relation with diarrhea. *Clin Pharmacol Ther* **72**:265-275.
22. Zingmark PH, Edenius C and Karlsson MO (2004) Pharmacokinetic/pharmacodynamic models for the depletion of Vbeta5.2/5.3 T cells by the monoclonal antibody ATM-027 in patients with multiple sclerosis, as measured by FACS. *Br J Clin Pharmacol* **58**:378-389.
23. Cullberg M, Eriksson UG, Wahlander K, Eriksson H, Schulman S and Karlsson MO (2005) Pharmacokinetics of ximelagatran and relationship to clinical response in acute deep vein thrombosis. *Clin Pharmacol Ther* **77**:279-290.
24. Aarons L and Graham G (2001) Methodological approaches to the population analysis of toxicity data. *Toxicol Lett* **120**:405-410.
25. Zingmark PH, Ekblom M, Odergren T, Ashwood T, Lyden P, Karlsson MO and Jonsson EN (2003) Population pharmacokinetics of clomethiazole and its effect on the natural course of sedation in acute stroke patients. *Br J Clin Pharmacol* **56**:173-183.
26. Knibbe CA, Zuideveld KP, DeJongh J, Kuks PF, Aarts LP and Danhof M (2002) Population pharmacokinetic and pharmacodynamic modeling of propofol for long-term sedation in critically ill patients: a comparison between propofol 6% and propofol 1%. *Clin Pharmacol Ther* **72**:670-684.
27. Olofsen E, Romberg R, Bijl H, Mooren R, Engbers F, Kest B and Dahan A (2005) Alfentanil and placebo analgesia: no sex differences detected in models of experimental pain. *Anesthesiology* **103**:130-139.
28. Gibiansky E and Gibiansky L (2005) Population PK/PD model of GPI 15715 GPI 15715-derived propofol in sedation and comparison of PK/PD models for ordered categorical observations. *PAGE 14* Abstr 735 *[www.page-meeting.org/?abstract=735]*
29. Nestorov I, Graham G, Duffull S, Aarons L, Fuseau E and Coates P (2001) Modeling and stimulation for clinical trial design involving a categorical response: a phase II case study with naratriptan. *Pharm Res* **18**:1210-1219.
30. Mandema JW and Stanski DR (1996) Population pharmacodynamic model for ketorolac analgesia. *Clin Pharmacol Ther* **60**:619-635.

31. Aarons L (1999) Software for population pharmacokinetics and pharmacodynamics. *Clin Pharmacokinet* **36**:255-264.
32. *NONMEM Users Guides* [computer program]. version V, VIbeta and VI Elliot City, MD, USA: Icon Development Solutions
33. *WinNonMix, Users' guide* [computer program]. version 2.0 Mountain View, CA, USA Pharsight
34. Lindstrom ML and Bates DM (1990) Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**:673-687.
35. *The NLMIXED procedure* [computer program]. version 8 Cary, NC, USA SAS Institute Inc.
36. Schumitzky A (1991) Nonparametric EM algorithm for estimating prior distributions. *Applied Mathematics and Computation* **45**:143-157.
37. *WinBUGS Users' guide* [computer program]. version 1.4.3 Cambridge, UK Medical Research Council Biostatistics Unit
38. Pillai GC, Mentre F and Steimer JL (2005) Non-linear mixed effects modeling - from methodology and software development to driving implementation in drug development science. *J Pharmacokinet Pharmacodyn* **32**:161-183.
39. Wade JR, Edholm M and Salmonson T (2005) A guide for reporting the results of population pharmacokinetic analyses: a Swedish perspective. *AAPS J* **7**:45.
40. Gelman A, Carlin JB and Stern HS (1995) *Bayesian data analysis.* Chapman & Hall, London, UK
41. Girard P, Blaschke TF, Kastrissios H and Sheiner LB (1998) A Markov mixed effect regression model for drug compliance. *Stat Med* **17**:2313-2333.
42. Yano Y, Beal SL and Sheiner LB (2001) Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check. *J Pharmacokinet Pharmacodyn* **28**:171-192.
43. Holford NH and Karlsson MO (2008) Model evaluation. Visual predictive checks. *PAGE 17* Abstr 1434 *[www.page-meeting.org/?abstract=1434]*
44. Jonsson EN and Karlsson MO (1999) Xpose--an S-PLUS based population pharmacokinetic/pharmacodynamic model building aid for NONMEM. *Comput Methods Programs Biomed* **58**:51-64.
45. Sheiner L and Wakefield J (1999) Population modelling in drug development. *Stat Methods Med Res* **8**:183-193.
46. Wählby U, Jonsson EN and Karlsson MO (2001) Assessment of actual significance levels for covariate effects in NONMEM. *J Pharmacokinet Pharmacodyn* **28**:231-252.
47. Wählby U, Bouw MR, Jonsson EN and Karlsson MO (2002) Assessment of type I error rates for the statistical sub-model in NONMEM. *J Pharmacokinet Pharmacodyn* **29**:251-269.
48. Wählby U, Matolcsi K, Karlsson MO and Jonsson EN (2004) Evaluation of type I error rates when modeling ordered categorical data in NONMEM. *J Pharmacokinet Pharmacodyn* **31**:61-74.
49. Gisleskog PO, Karlsson MO and Beal SL (2002) Use of prior information to stabilize a population data analysis. *J Pharmacokinet Pharmacodyn* **29**:473-505.
50. Klareskog L and Olsson T (1990) Autoimmunity to collagen II and myelin basic protein: comparative studies in humans and rodents. *Immunol Rev* **118**:285-310.
51. Martin R, McFarland HF and McFarlin DE (1992) Immunological aspects of demyelinating diseases. *Annu Rev Immunol* **10**:153-187.

52. Olsson T, Edenius C, Ferm M, Samuelson P, Torrang A, Wallstrom E, Khademi M, Andersson M and Arfors L (2002) Depletion of Vbeta5.2/5.3 T cells with a humanized antibody in patients with multiple sclerosis. *Eur J Neurol* **9**:153-164.
53. Bleiberg H and Cvitkovic E (1996) Characterisation and clinical management of CPT-11 (irinotecan)-induced adverse events: the European perspective. *Eur J Cancer* **32A Suppl 3**:S18-23.
54. Kawato Y, Aonuma M, Hirota Y, Kuga H and Sato K (1991) Intracellular roles of SN-38, a metabolite of the camptothecin derivative CPT-11, in the antitumor effect of CPT-11. *Cancer Res* **51**:4187-4191.
55. Iyer L, King CD, Whitington PF, Green MD, Roy SK, Tephly TR, Coffman BL and Ratain MJ (1998) Genetic predisposition to the metabolism of irinotecan (CPT-11). Role of uridine diphosphate glucuronosyltransferase isoform 1A1 in the glucuronidation of its active metabolite (SN-38) in human liver microsomes. *J Clin Invest* **101**:847-854.
56. Haaz MC, Rivory L, Jantet S, Ratanasavanh D and Robert J (1997) Glucuronidation of SN-38, the active metabolite of irinotecan, by human hepatic microsomes. *Pharmacol Toxicol* **80**:91-96.
57. Feigin VL, Lawes CM, Bennett DA and Anderson CS (2003) Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *Lancet Neurol* **2**:43-53.
58. Kwan L (2001) Epidemiology of Stroke. *CME J Geriatr Med* **3**:94-98.
59. Cross AJ, Jones JA, Baldwin HA and Green AR (1991) Neuroprotective activity of chlormethiazole following transient forebrain ischaemia in the gerbil. *Br J Pharmacol* **104**:406-411.
60. Cross AJ, Jones JA, Snares M, Jostell KG, Bredberg U and Green AR (1995) The protective action of chlormethiazole against ischaemia-induced neurodegeneration in gerbils when infused at doses having little sedative or anticonvulsant activity. *Br J Pharmacol* **114**:1625-1630.
61. Sydserff SG, Cross AJ and Green AR (1995) The neuroprotective effect of chlormethiazole on ischaemic neuronal damage following permanent middle cerebral artery ischaemia in the rat. *Neurodegeneration* **4**:323-328.
62. Sydserff SG, Cross AJ, West KJ and Green AR (1995) The effect of chlormethiazole on neuronal damage in a model of transient focal ischaemia. *Br J Pharmacol* **114**:1631-1635.
63. Snape MF, Baldwin HA, Cross AJ and Green AR (1993) The effects of chlormethiazole and nimodipine on cortical infarct area after focal cerebral ischaemia in the rat. *Neuroscience* **53**:837-844.
64. Lyden P (1997) GABA and the neuroprotection. In Chaper 10 of *Neuroprotective agents and cerebral ischemia*. Green AR and Cross AJ (eds) AP Ltd, London, UK, 233-258
65. Marshall JW, Cross AJ and Ridley RM (1999) Functional benefit from clomethiazole treatment after focal cerebral ischemia in a nonhuman primate species. *Exp Neurol* **156**:121-129.
66. Hanna JP, Frank JI, Furlan AJ, Sila CA and Secic M (1996) Prediction of worsening consciousness from edema after hemispheric infarction. *J Stroke Cerebrovasc Dis* **6**:25-29.
67. A manual of standarized terminology, techniques and scoring system for sleep stages of human subjects (1968) Los Angles, CA, USA, Access at

68. Mallon L, Broman JE and Hetta J (2005) High incidence of diabetes in men with sleep complaints or short sleep duration: a 12-year follow-up study of a middle-aged population. *Diabetes Care* **28**:2762-2767.
69. Turek FW (2005) Insomnia and depression: if it looks and walks like a duck. *Sleep* **28**:1362-1363.
70. Gangwisch JE, Heymsfield SB, Boden-Albala B, Buijs RM, Kreier F, Pickering TG, Rundle AG, Zammit GK and Malaspina D (2006) Short sleep duration as a risk factor for hypertension: analyses of the first National Health and Nutrition Examination Survey. *Hypertension* **47**:833-839.
71. Dooley DJ, Taylor CP, Donevan S and Feltner D (2007) Ca2+ channel alpha2delta ligands: novel modulators of neurotransmission. *Trends Pharmacol Sci* **28**:75-82.
72. Karlsson MO, Schoemaker RC, Kemp B, Cohen AF, van Gerven JM, Tuk B, Peck CC and Danhof M (2000) A pharmacodynamic Markov mixed-effects model for the effect of temazepam on sleep. *Clin Pharmacol Ther* **68**:175-188.
73. *Comprehensive Perl Archive Network* [computer program]. version http://www.cpan.org/
74. Lindbom L, Ribbing J and Jonsson EN (2004) Perl-speaks-NONMEM (PsN)--a Perl module for NONMEM related programming. *Comput Methods Programs Biomed* **75**:85-94.
75. Lindbom L, Pihlgren P and Jonsson EN (2005) PsN-Toolkit--a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. *Comput Methods Programs Biomed* **79**:241-257.
76. *R* [computer program]. version 2.2-2.4 http://r-project.org/
77. *S-plus 2000* [computer program]. version 6.1 Seattle, WA, USA MathSoft
78. Sheiner LB, Stanski DR, Vozeh S, Miller RD and Ham J (1979) Simultaneous modeling of pharmacokinetics and pharmacodynamics: application to d-tubocurarine. *Clin Pharmacol Ther* **25**:358-371.
79. Efron B (1979) Bootstrap method: Another look at the jackknife. *Ann Stat* **7**:1-26.
80. Savic R and Karlsson MO (2007) Importance of shrinkage in empirical bayes estimates for diagnostics and estimation: Problems and solutions. *PAGE 16* Abstr 1087 *[www.page-meeting.org/?abstract=1087]*
81. Kowalski KG, McFadyen L, Hutmacher MM, Frame B and Miller R (2003) A two-part mixture model for longitudinal adverse event severity data. *J Pharmacokinet Pharmacodyn* **30**:315-336.
82. Zhang L, Beal SL and Sheiner LB (2003) Simultaneous vs. sequential analysis for population PK/PD data I: best-case performance. *J Pharmacokinet Pharmacodyn* **30**:387-404.
83. Booth BP and Gobburu JV (2003) Considerations in analyzing single-trough concentrations using mixed-effects modeling. *J Clin Pharmacol* **43**:1307-1315.
84. Guidance for industry on Population Pharmacokinetics (1999) *Food and Drug Administration*, Rockville, MD, USA, Access at http://www.fda.gov/cder/guidance/1852fnl.pdf.
85. Innovation or stagnation: challenges and opportunity on the critical path to new medical products (2004) *Food and Drug Administration*, Rockville, MD, USA, Access at http://www.fda.gov/oc/initiative/criticalpath/whitepaper.html.
86. Plan E, Maloney A, Trocóniz IF and Karlsson MO (2008) Maximum likelihood estimation methods: performances in count response models population parameters. *PAGE 17* Abstr 1372 *[www.page-meeting.org/?abstract=1372]*

87. Maas HJ, Danhof M and Della Pasqua OE (2006) Prediction of headache response in migraine treatment. *Cephalalgia* **26**:416-422.
88. Beal SL (2001) Ways to fit a PK model with some data below the quantification limit. *J Pharmacokinet Pharmacodyn* **28**:481-504.
89. Jonsson F, Marshall S, Krams M and Jonsson EN (2005) A longitudinal model for non-monotonic clinical assessment scale data. *J Pharmacokinet Pharmacodyn* **32**:795-815.
90. Karlsson KE, Wilkins J, Karlsson MO and Jonsson EN (2007) Modelling disease progression in acute stroke by simultaneously using the NIH stroke scale, the Scandinavian stroke scale and the Barthel index. *PAGE 16* Abstr 1191 *[www.page-meeting.org/?abstract=1191]*

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy 83

Editor: The Dean of the Faculty of Pharmacy

A doctoral dissertation from the Faculty of Pharmacy, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-9333

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2008