Pharmacometric Methods and Novel Models for Discrete Data

ELODIE L PLAN
Dissertation presented at Uppsala University to be publicly examined in B41, BMC, Husargatan 3, Uppsala, Friday, May 20, 2011 at 13:15 for the degree of Doctor of Philosophy (Faculty of Pharmacy). The examination will be conducted in English.

Abstract

Pharmacodynamic processes and disease progression are increasingly characterized with pharmacometric models. However, modelling options for discrete-type responses remain limited, although these response variables are commonly encountered clinical endpoints. Types of data defined as discrete data are generally ordinal, e.g. symptom severity, count, i.e. event frequency, and time-to-event, i.e. event occurrence. Underlying assumptions accompanying discrete data models need investigation and possibly adaptations in order to expand their use. Moreover, because these models are highly non-linear, estimation with linearization-based maximum likelihood methods may be biased.

The aim of this thesis was to explore pharmacometric methods and novel models for discrete data through (i) the investigation of benefits of treating discrete data with different modelling approaches, (ii) evaluations of the performance of several estimation methods for discrete models, and (iii) the development of novel models for the handling of complex discrete data recorded during (pre-)clinical studies.

A simulation study indicated that approaches such as a truncated Poisson model and a logit-transformed continuous model were adequate for treating ordinal data ranked on a 0-10 scale. Features that handled serial correlation and underdispersion were developed for the models to subsequently fit real pain scores. The performance of nine estimation methods was studied for dose-response continuous models. Other types of serially correlated count models were studied for the analysis of overdispersed data represented by the number of epilepsy seizures per day. For these types of models, the commonly used Laplace estimation method presented a bias, whereas the adaptive Gaussian quadrature method did not. Count models were also compared to repeated time-to-event models when the exact time of gastroesophageal symptom occurrence was known.

Two new model structures handling repeated time-to-categorical events, i.e. events with an ordinal severity aspect, were introduced. Laplace and two expectation-maximisation estimation methods were found to be performing well for frequent repeated time-to-event models.

In conclusion, this thesis presents approaches, estimation methods, and diagnostics adapted for treating discrete data. Novel models and diagnostics were developed when lacking and applied to biological observations.

Keywords: Pharmacometrics, pharmacodynamics, disease progression, modelling, discrete data, count, ordered categorical, repeated time-to-event, RTTCE, RCEPT, NONMEM, FOCE, LAPLACE, SAEM, AGQ, pain scores, epilepsy seizures, gastroesophageal symptoms, statistical power, simulations, diagnostics

Elodie L. Plan, Department of Pharmaceutical Biosciences, Box 591, Uppsala University, SE-75124 Uppsala, Sweden.

© Elodie L. Plan 2011

ISSN 1651-6192
ISBN 978-91-554-8064-6
urn:nbn:se:uu:diva-150929 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-150929)
“Not everything that can be counted counts and not everything that counts can be counted.”
Albert Einstein (attributed)

In memory of Nicolas Paquereau,
my friend.
(1977-2009)
List of papers

This thesis is based on the following nine papers. They are referred to in the text by their Roman numerals.

I  Elodie L. Plan, Yang Sun, Mats O. Karlsson. Approaches to Likert-Type Ordered Categorical Data Analysis. *In manuscript.*


Reprints were made with permission from the respective publishers: Springer Science and Business Media, American Association of Pharmaceutical Scientists, and Nature Publishing Group.
## Contents

Introduction .............................................................................................................................. 11

1. Pharmacometrics .................................................................................................................. 11
   1.1. Pharmacokinetic-pharmacodynamic relationships .............................................. 11
   1.2. Regression analyses and mixed-effects models .................................................. 12
   1.3. Model-based drug development perspective ..................................................... 13

2. Pharmacodynamics .............................................................................................................. 14
   2.1. Response variable ................................................................................................. 14
   2.2. Discrete approaches ............................................................................................. 16
   2.3. Approaches assumptions ....................................................................................... 17

3. Estimation ............................................................................................................................. 19
   3.1. Likelihood function ............................................................................................... 19
   3.2. Likelihood approximation ...................................................................................... 20
   3.3. Approximation impact ............................................................................................ 22

Aims .............................................................................................................................................. 23

Methods ................................................................................................................................ 25

1. Studied data ......................................................................................................................... 25
   1.1. Continuous data *(Paper II)* ................................................................................. 25
   1.2. Ordinal data *(Papers I & III)* ............................................................................... 26
   1.3. Count data *(Papers IV-VI)* .................................................................................. 26
   1.4. Event data *(Papers VII-IX)* .................................................................................. 27

2. Modeling analyses ............................................................................................................... 28
   2.1. Model development ............................................................................................... 28
   2.2. Continuous model *(Paper III)* ............................................................................. 28
   2.3. Count models *(Papers III-IV & VI-VII)* ............................................................. 29
   2.4. Event model *(Paper VII)* ..................................................................................... 34
   2.5. Model diagnostics ................................................................................................. 35

3. Simulation studies ............................................................................................................... 36
   3.1. Approach evaluation *(Papers I & IX)* ................................................................. 36
   3.2. Algorithm evaluation *(Papers II, V & VIII)* ....................................................... 37

Results ............................................................................................................................................ 39

1. Discrete versus continuous data ....................................................................................... 39
   1.1. Modeling approaches *(Paper I)* ....................................................................... 39
   1.2. Algorithm performance *(Paper II)* ................................................................. 42
   1.3. Novel models *(Paper III)* ................................................................................... 44
2. Count data ................................................................................................................. 46
   2.1. Modeling approaches (Paper IV) ................................................................. 46
   2.2. Algorithm performance (Paper V) ................................................................. 48
   2.3. Novel model (Paper VI) .................................................................................. 50
3. Event data ................................................................................................................... 52
   3.1. Modeling approaches (Paper VII) ................................................................. 52
   3.2. Algorithm performance (Paper VIII) ............................................................. 54
   3.3. Novel models (Paper IX) ................................................................................. 56

Discussion ..................................................................................................................... 59
1. Discrete versus continuous data .............................................................................. 59
2. Count data ............................................................................................................... 61
3. Event data ................................................................................................................. 63

Conclusions ................................................................................................................... 67

Acknowledgements ..................................................................................................... 69

References .................................................................................................................... 73
Abbreviations

AGQ         Adaptive Gaussian quadrature
AR(1)       Autoregressive time series
BAS         Baseline
BIC         Bayesian information criterion
CI          Confidence interval
CPU         Central unit processing
d          Dose
EBE         Empirical Bayes estimates
ED$_{50}$  Dose to reach half of the effect
$E_{\text{max}}$ Maximal effect
FOCE        First-order conditional estimation
GP          Generalized Poisson
h          Hour
i          Indices for individuals
IIV         Interindividual variability
IMP         Importance sampling
IPRED       Individual prediction
j          Indices for observations
kg         Kilogram
m          Severity score
min        Minute
n          Count
NB          Negative binomial
OC          Ordered categorical
OFV         Objective function value
Ovdp        Overdispersion
PD          Pharmacodynamics
PI          Prediction interval
PK          Pharmacokinetics
PS          Poisson
RCEpT       Repeated categorical events per time-intervals
RER         Relative estimation error
RMSE        Root mean squared error
RSE         Relative standard error
RTTCE       Repeated time-to-categorical event
RTTE        Repeated time-to-event
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAEM</td>
<td>Stochastic approximation expectation maximization</td>
</tr>
<tr>
<td>$t$</td>
<td>Time</td>
</tr>
<tr>
<td>$t_{1/2}$</td>
<td>Half-life</td>
</tr>
<tr>
<td>TLESR</td>
<td>Transient lower esophageal sphincter relaxation</td>
</tr>
<tr>
<td>VPC</td>
<td>Visual predictive check</td>
</tr>
<tr>
<td>ZINB</td>
<td>Zero-inflated negative binomial</td>
</tr>
<tr>
<td>ZIP</td>
<td>Zero-inflated Poisson</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Sigmoidicity factor</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Dispersion factor</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>Residual variability</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Interindividual variability</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Typical value of the parameter</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Mean count</td>
</tr>
<tr>
<td>$\pi$</td>
<td>Probability of transition</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Standard deviation of the residual variability</td>
</tr>
<tr>
<td>$\omega$</td>
<td>Standard deviation of the interindividual variability</td>
</tr>
</tbody>
</table>
Pharmacometric Methods and Novel Models for Discrete Data

Introduction

1. Pharmacometrics

Pharmacometric research is a young scientific discipline\(^1\) which has impressively matured since it was introduced more than 30 years ago by L.B. Sheiner. This science is undeniably situated at a cross-road between two fields\(^2\): pharmacology and statistics.

1.1. Pharmacokinetic-pharmacodynamic relationships

Pharmacology is the study of the interactions that occur between a chemical and an organism. The two main areas of pharmacology are pharmacokinetics (PK), which essentially describes how the body affects the drug, and pharmacodynamics (PD), which describes how the drug affects the body. PK mechanisms include absorption, distribution, metabolism, and excretion. PD actions differentiate efficacy, the desirable effects of the drug, and safety, the side-effects. Additionally, understanding baseline time-course, how the disease evolves with time, is crucial for a precise assessment of the above.

Pharmacometrics\(^3\) is used to describe pharmacological responses:
- qualitatively
- quantitatively

and involves models built either empirically or mechanistically in nature, depending on (patho)physiological or biochemical prior understanding and model purpose. Identification and quantification of pharmacokinetic-pharmacodynamic (PKPD) relationships is the ultimate aim of the pharmacometric process. It involves the description of the underlying causal chain of: dose \(\leftrightarrow\) concentration \(\leftrightarrow\) effect.

PKPD modeling consequently holds an important place in drug development\(^4\), classically taking place after the clinical trial by analyzing the collected data. Occasionally, it has a role before the clinical trial by facilitating the design and sporadically it even replaces the clinical trial by enabling simulations\(^5\). The impact of pharmacometrics on decision making in the pharmaceutical industry was confirmed by several surveys\(^6\)-\(^9\). In parallel,
public health agencies like the US Food and Drug Administration (FDA) regularly publish reviews about its impact on regulatory decisions\textsuperscript{10-13}. 

1.2. Regression analyses and mixed-effects models

Statistics is the formalization of relationships between variables in the form of mathematical equations. Mathematical models are developed to describe how a dependent variable, usually concentration in PK, effect in PD, and disease status in baseline time-course, is related to an independent variable. The independent variable in pharmacometric analyses usually is time, and sometimes dose or concentration in PD. Statistical inference is used to define how well observations are mimicked by the generated model predictions\textsuperscript{14}.

The complexity of pharmacometric models ranges from simple linear, over generalized linear, to non-linear functions. Non-linear regressions link the response to an explanatory variable through a non-linear curve. Generalized linear regressions\textsuperscript{15} model the probability of the response as a function of the predictor variable.

Pharmacometrics focuses on population modeling\textsuperscript{16}, in which data from all individuals, \textit{e.g.} involved in a clinical trial, are analyzed simultaneously\textsuperscript{17-19}. The science also lends itself to the analysis of populations pooled across several trials. The primary advantage of this approach is a significant decrease in the number of observations needed per individual for the data to be sufficient to fit a model, since all measurements contribute to the information on which the model is based. Secondly, a population approach enables the understanding of the variability in the population. The main source of variability is usually interindividual variability (IIV). Part of this IIV can be explained by covariates, \textit{i.e.} demographics or other known predictors, the rest is apparently unpredictable, \textit{i.e.} a reflection of current lack of knowledge or true random variation. The other sources of variability are interoccasion variability\textsuperscript{20} and residual variability. Residual variability mainly represents measurement error and unexplained noise.

In modeling of individual data or when using a naïve pooling approach, parameters constituting the model are estimated as single values corresponding to one individual. In population modeling, individual parameters are distributions characterized by a median and a variance. Accordingly, pharmacometrics uses mixed-effects divided into:

- fixed effects
- random effects

For a specific parameter, the fixed effect reflects the parameter value of a typical individual in the population, and the random effects mainly the variability between individuals. A function of both effects defines the individual parameter. The random effect can be associated to the fixed effect in dif-
ferent ways. The relationship can be additive, implying that the variability is constant regardless of the magnitude of the fixed effect. However, this description is now rarely used on physiological parameters as it allows for negative individual parameters. Statistical models used in pharmacometrics are therefore exponential on the normal scale (Equation 1) or additive on the logit-scale (Equation 2):

\[
\phi_i = \mu \cdot e^{\eta_i} \quad \text{(Eq. 1.)} \\
\phi_i = \frac{e^\ln(\frac{\mu}{1-\mu}) + \eta_i}{1 + e^\ln(\frac{\mu}{1-\mu}) + \eta_i} \quad \text{(Eq. 2.)}
\]

where \( \phi_i \) are the individual parameters, \( \mu \) the fixed effects, and \( \eta_i \) the random effects. The \( \eta_i \) are normally distributed with a mean 0 and a variance \( \omega^2 \): \( \eta_i \sim N(0, \omega^2) \). The transformation in Equation 1 yields a log-normal distribution, ensuring non-negative \( \phi_i \) values, and Equation 2 constitutes an expit-logit, ensuring probabilities \( \phi_i \) to remain between 0 and 1.

### 1.3. Model-based drug development perspective

Biological systems are usually very complex; even highly mechanistic models, e.g. physiologically-based pharmacokinetic (PBPK)\(^{21}\) models, are imperfect. It is therefore clear that the goal of pharmacometrics is not to find the true model; the goal is rather to learn how to best use the drug\(^{22}\). The understanding of PK, PD, and baseline time-course can be improved and requires clinical trials to be designed accordingly, but the focus should also be on using the available information in its integrality and in its true nature.

Interestingly, since the field evolved rapidly and is viewed as multidisciplinary, innovative training programs are encouraged\(^{23}\) and proposed\(^{24}\) in academia. In industry, strengths, weaknesses, opportunities, and threats are listed\(^{25}\), the transition to model-based drug development (MBDD) is planned\(^{26}\), and industrialization initialized\(^{27}\). Issues raised are multiple and involve knowledge-sharing and collaboration-need as well as the development of standards and automation of the pharmacometric process.

At the level of model-based analysis, standards or automation may be challenging as modeling is also described as an art\(^{28}\). Nevertheless, good modeling practices should incorporate results from previous analyses and occur in a recursive process. A scheme of the chain leading the data from the (pre-)clinical study to the information learnt from their analysis (Figure 1) suggests the intermediate steps: approach and algorithm selection. These selections are either based on simulations or on prior knowledge, with continuous integration of the modeling outcome along the recursive process.
2. Pharmacodynamics

As described above, pharmacodynamics enables to describe the effect of a drug on the body. The drug effect typically affects a baseline response. In theory, the baseline simply corresponds to the response in the absence of drug; in practice, it is measured in the end of the baseline phase in patients randomized to placebo arm, in order to account for influences directly linked to the clinical trial. The drug effect may evolve with time – be dynamic – due to the time-course of the concentrations. However, the baseline response can also vary with time due to pathophysiological changes; this is usually referred to as disease progression, but can also indicate disease remission. Natural physiological changes, fluctuations, or often cycles, may also affect the baseline. Finally, a placebo effect may be observed. Consequently, describing all these aspects appears crucial in order to correctly characterize and quantify the clinical pharmacology of a drug.

2.1. Response variable

To study the baseline time-course and how drug effect may affect it, a response variable needs to be recorded during a certain period of time. Two types of outcomes have been defined:

- biomarkers
- clinical endpoints

A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention. A clinical endpoint is generally a characteristic or variable that reflects how a patient feels, functions, or survives. Surrogate markers can be a combination of objective and
subjective measures. For the modeling process, the difference resides in whether the data-type of the outcome is:

- continuous
- discrete

A biomarker often consists of a concentration, e.g. cholesterol or glucose, which is therefore seen as continuous. A clinical endpoint, which can be the report of a symptom, e.g. nausea or pain, is commonly discrete.

Discrete data have previously been referred to as odd-type data \(^{32}\). They usually correspond to non-negative integers: 0, 1, 2, … .

The simplest type of discrete data is the binary type, corresponding to event-no event. When there are more than two categories, e.g. the classes mild-moderate-severe, they constitute categorical data. Categorical data are ordinal data when they can be ordered on a scale. Outcomes that are numbers of events of a given type in a given time-interval are count data. Responses consisting of times until an event, such as death, are time-to-event data, or, in this case, survival data. If the event can happen several times during the study period, they are repeated time-to-event (RTTE) data, or, for simplicity, event data.

The most complex types of discrete data are hence: ordinal, count, and event, answering the questions: how much (of the symptom are you experiencing?), how many (symptoms did you experience?), and how long (did it take until you experienced a symptom?), respectively.

![Figure 2. Schematic representation of the different types of data. Observations are represented vs. time, and continuous-time data-types are distinguished from discrete-time data-types.](image)

Another aspect is the following (Figure 2): since event data, like continuous data, can be recorded at any time, they are actually a continuous-time type of data. In contrast, count data are a discrete-time type of data, meaning that the 7 events occurring during the exemplified study period are
summarized into the series 3, 2, 1, 0, and 1, happening per min/hour/day/week/month. In theory, the number of counts is infinite, as opposed to ordinal data where the number of possible categories is defined in advance, i.e. when the scale is chosen. Ordinal data are also a discrete-time type of data in their current use; they often represent a summarized version of an underlying response, e.g. the maximal severity experienced over the past min/hour/day/week/month.

2.2. Discrete approaches

Basic models adapted for discrete data have been introduced before. They have the common property of being based on probabilities, i.e. these models describe the probability of an observation having a specific (discrete) value.

2.2.1. Ordinal approach

An ordinal model\textsuperscript{33} was originally applied in the field of clinical pharmacology by L.B. Sheiner\textsuperscript{34, 35} when analyzing pain relief data. It makes use of a multinomial logistic regression through cumulative probabilities.

If $Y_{ij}$ is the observed symptom severity for the $i^{th}$ individual at the $j^{th}$ measurement, then the probability that it is equal to $m$ in $0, 1, \ldots, M$ is:

$$\text{logit}(P(Y_{ij} \geq m)) = \sum_{k=1}^{m} \alpha_k + \eta_{\alpha_i} \tag{Eq. 3.}$$

$$P(Y_{ij} \geq m) = \frac{e^{\text{logit}(P(Y_{ij} \geq m))}}{1 + e^{\text{logit}(P(Y_{ij} \geq m))}} \tag{Eq. 4.}$$

$$P(Y_{ij} = 0) = 1 - P(Y_{ij} \geq 1)$$

$$P(Y_{ij} = m) = P(Y_{ij} \geq m) - P(Y_{ij} \geq (m + 1))$$

$$P(Y_{ij} = M) = P(Y_{ij} \geq M) \tag{Eq. 5.}$$

where $\alpha_k$ are the fixed effects expressing the cumulative probabilities on the logit scale and $\eta_{\alpha_i}$ the only random effect. In Equation 5, $\sum_{m=0}^{M} P(Y_{ij} = m) = 1$ is ensured. The number of $\alpha_k$ included in the model is equal to $(M-1)$ since $P(Y_{ij} \geq 0) = 1$ by definition. This approach is sometimes referred to as OC in this thesis, for ordered categorical.

2.2.2. Count approach

The first count model was introduced by S.D. Poisson\textsuperscript{36} working on judgments in criminal matters. Later it was also used in biomedical sciences\textsuperscript{37}. It is a probability distribution expressing the probability of a number of events $n$ occurring in a fixed time-interval.

The model was named Poisson and noted $Y_{ij} = PS(\lambda)$:
where \( \lambda \) is a parameter or a function of parameters corresponding to a positive real number equal to the expected number of occurrences during the given interval and its variance.

### 2.2.3. Event approach

Modeling of RTTE data is an extension of time-to-event analysis, also called survival analysis\(^{38} \), and was initially published in the pharmacometric field by E.H. Cox\(^{39} \). It is based on a hazard process, representing the instantaneous risk of an event to occur. Each time-to-event, starting at the beginning of the experiment or after the last event and ending when the next event happens, is modeled using a cumulative hazard process.

The probability that no event occurs between \( t_{j-1} \) and \( t_j \) is:

\[
P(Y_{ij} = 0, t_{j-1} < T < t_j) = e^{-\int_{t_{j-1}}^{t_j} h(t) \, dt} \tag{Eq. 7.}
\]

where \( h(t) \) is the hazard, \( \int_{t_{j-1}}^{t_j} h(t) \, dt = H(t) \), and \( e^{-H(t)} = S(t) \), where \( S(t) \) is the probability of survival. The hazard, which is a function of time, can be constant or time-varying. In the latter case, it is then usually described using an ordinary differential equation (ODE).

### 2.3. Approaches assumptions

All models described above are intended to be applied to their corresponding data-type; in this case they are able to both estimate and simulate the data by respecting their true discrete nature. In the particular case when count data display a large number of possible values, it may be possible to use a continuous distribution. However, some assumptions are inherent to these models that may not always correspond to characteristics met in real data.

In continuous models, the measurement error as well as the unexplained variability is estimated as the residual error. But since discrete models are based on probabilities, the inclusion of such an error is traditionally considered as intractable. Modeling practices advocate against data deletion, therefore all observations are given the same importance, regardless of whether the responses were experimentally recorded in a controlled environment, or subjectively reported in a diary. It is hence assumed that the data collection is accurate and precise.

Another assumption specifically concerns the count approach. The Poisson model presents the property that the expected number of events is assumed to be equal to the variance of the counts within an individual’s observations.
This is called equidispersion. However data presenting a variance either smaller or larger than the mean are common among disease responses.

Both ordinal and count models make the assumption, related to their discrete-time nature, that the reported value is representative of the whole time-interval. For ordinal models, the rated response may in reality be a maximum value across several events during the time-interval, summarizing non-recorded single events. For count models, the count response disregards the time-course of the events during the time-interval and assumes an underlying constant hazard.

![Figure 3. Observed number of epilepsy seizures vs. time for nine patients.](image)

All discrete models described above assume another important characteristic: independence between events. Predicted values depend on the model and the parameter estimates but not on the preceding response. However, representations of real data (Figure 3) sometimes have contradictory characteristics by showing clustering of “low” and “high” response values.

Furthermore, existing models for categorical data handle only one of the above described features of the data at a time. However, one can imagine responses consisting of a number of counts or a time-to-event combined with an ordinal severity.

Finally, diagnostics to explore and evaluate these models tend to be restricted, firstly because they are essentially based on simulations and secondly because the use of discrete models is still limited.
3. Estimation

The pharmacometric process includes the estimation of model parameters by fitting them to the data. The estimation is performed by algorithms embedded within certain software, making informatics the fuel for pharmacometrics\(^4^0\).

Two main classes of estimation methods exist:

- parametric
- non-parametric

where the parametric methods commonly assume Gaussian distribution of the parameters, while the non-parametric does not make any distributional assumptions. Additionally, within both classes, one can distinguish several different methods for obtaining parameter estimates, the main ones being:

- likelihood-based
- bayesian

The most commonly used group of estimation methods in pharmacometrics is the parametric likelihood-based group.

3.1. Likelihood function

The likelihood-based methods aim at maximizing the likelihood of the parameters to describe the data. The individual likelihood in a model without random effects is simply:

\[
L_i(y_i, \Psi) = P(y_i, \Psi) \quad \text{(Eq. 8.)}
\]

where \(P(y_i, \Psi)\) is the data density for continuous data and the data probability for discrete data.

The individual likelihood in a model with random effects can be expressed as the marginal density:

\[
L_i(y_i, \Psi) = \int P(y_i | \eta_i, \Psi) \cdot P(\eta_i, \Psi) d\eta_i \quad \text{(Eq. 9.)}
\]

where \(P(y_i | \eta_i, \Psi)\) is the conditional density of the observations given the individual random effects and \(P(\eta_i, \Psi)\) the population parameter density of the individual random effects. The likelihood of a population of \(N\) individuals is determined by taking the product of the \(N\) individual likelihoods \(L(y, \Psi) = \prod_{i=1}^{N} L_i(y_i, \Psi)\). In many softwares, -2 times the population log-likelihood is calculated as the sum over the \(N\) individual log-likelihoods \(-2 \log L(y, \Psi) = -2 \sum_{i=1}^{N} \log(L_i(y_i, \Psi))\). This transformation makes computations easier for this function to then be minimized to obtain the maximum likelihood estimates of the model parameters.
Equation 8 can be computed exactly whereas the integral in Equation 9 typically has no closed-form solution. This absence of analytical solution is due to the random effects entering the model in a non-linear fashion.

### 3.2. Likelihood approximation

Several algorithms were developed in different softwares to perform maximum likelihood estimation. The first software that emerged for the specific purpose of non-linear mixed effects modeling was NONMEM\(^41\), by L.B. Sheiner and S.L. Beal from the University of California in San Francisco. The first algorithm implemented was a first-order estimation (FO)\(^42\), linearizing the model around the median of the individual parameters, \emph{i.e.} random effects set to zero. The first case studies were essentially PK analyses\(^43\). But over the years the structural models were extended to PKPD analyses and more variability components were added to the statistical model. Along with methodology development, more algorithms were developed\(^44\) and more software have emerged; however NONMEM remains the most commonly used\(^45\).

#### 3.2.1. First-order conditional estimation

Initially proposed by M.L. Lindstrom and D.M. Bates\(^46\), the first-order conditional (FOCE) estimation algorithm approximates the logarithm of Equation 9 by the log-likelihood of a linear mixed-effects model. The linearization is this time done around the current estimates of the fixed effects and the conditional median of the random effects using a first-order Taylor expansion. The random effects are obtained by maximizing the empirical Bayes posterior density of the \( \eta_1 \) using the current estimates of vector \( \Psi \).

The method was originally implemented in NONMEM version 4 as the FOCE method, and in the R software\(^47\) as the nlme module as well. In the case of interaction between the random effects and the residual errors, \emph{i.e.} with proportional or exponential error models, the interaction terms are calculated and included in the likelihood using FOCE INTERACTION.

#### 3.2.2. Laplace approximation

The approximation of Equation 9 is accomplished by a linearization of the model around the conditional estimates of the random effects involving a second-order Taylor expansion of the integrand.

The method is available in the SAS software\(^48\) within the NLMIXED procedure as a special case of the adaptive Gaussian quadrature algorithm, where only one abscissa is defined with a corresponding weight equal to 1. It is implemented in NONMEM as LAPLACE, while LAPLACE INTERACTION exists in versions 6.2.0 and onward.
3.2.3. Adaptive Gaussian quadrature

In adaptive Gaussian quadrature (AGQ), a numerical weighted approximation of Equation 9 is done at predetermined abscissa for the random effects. The abscissa are centered around the conditional mode of the random effects and scaled by the Hessian matrix from the conditional mode estimation, as suggested by J.C. Pinheiro and D.M. Bates.

The adaptive Gaussian quadrature is not available in NONMEM, but it is implemented in SAS within the NLMIXED procedure. This NLMIXED offers the possibility to arbitrarily increase the number of abscissas, i.e. quadrature points (qpoints), the center quadrature point being the empirical Bayes parameter estimate.

3.2.4. Stochastic approximation expectation maximization

Stochastic approximation expectation maximization is an extension of the family of expectation-maximization (EM) algorithms where individual random effects are considered as missing data. It is characterized by complete likelihood maximization through alternation of E and M steps. In the E-step the expectation of the log-likelihood is calculated given the current estimates of the population parameters, and in the M-step new population parameters maximizing the likelihood are computed given the expected likelihood in the E-step. The E-step is often analytically intractable for nonlinear models, and in the stochastic approximation expectation maximization algorithm (SAEM) it is divided, first into a simulation of individual parameters using a Markov Chain Monte Carlo algorithm (S-step), secondly into a stochastic approximation of the expected likelihood (SA-step).

Within the pharmacometric field, the SAEM algorithm was implemented in MONOLIX before been made available in NONMEM version 7.1.0. A number of settings, e.g. the number of samples per subject and the maximum number of iterations in the different steps, can be modified in the two software. Additionally, certain options like simulated annealing, i.e. stochastic variation of the parameters with decreasing tolerance of accepting increases in OFV, and mu-parameterization, i.e. individual parameters defined as the log-transform of a Gaussian random vector to meet with positivity constraints, are default in MONOLIX and user-defined in NONMEM.

3.2.5. Monte Carlo importance sampling

The Monte Carlo importance sampling method (IMP) is also an EM-algorithm. The E-step in the IMP method differs from the SAEM method in that a Monte Carlo integration is performed by sampling around current individual estimates to obtain the expected likelihood. Population parameters are then updated from subjects’ conditional parameters by a single iteration in the M-step.

The method is available in NONMEM version 7.1.0.
3.3. Approximation impact

Although all estimation algorithms use different statistical methods, all are aimed at producing reliable estimates of the model parameters. The complexity of the model and the approximations embedded in the algorithm can potentially lead to poor estimation performance. This estimation performance is typically measured through accuracy and precision.

Biased or imprecise estimation of model parameters may have the consequence that the model cannot replicate original data in simulations\textsuperscript{50, 60}. This is a concern if the model is to be used in clinical trial simulations as a tool for decision making in drug development.

Studies that compared the performance of these algorithms have been performed in the past. They were motivated either by the introduction of a new algorithm\textsuperscript{16, 61}, by a practical application\textsuperscript{62}, or by a survey\textsuperscript{63}.

However, most of these investigations were not supported by a high number of simulations, but rather considered the analysis of only one simulated or real dataset. The methodology to study algorithm performance is of importance to minimize the influence of the stochastic process in the result.

Lately, a large Monte Carlo simulation study compared the performance of several estimation methods for continuous PK data\textsuperscript{64}. This comparison concerned a large number of estimation methods and software, which is rare for this type of investigation.

Nevertheless, there remain few comparative studies into the performance of these algorithms for PD models, although these models are increasingly utilized and usually highly non-linear.

Interest in the accuracy of discrete model parameter estimation is growing. Bias has been reported for models dealing with binary data\textsuperscript{65, 66} and a recent investigation reported characteristics of responses and estimation methods resulting in parameter bias with ordered categorical models\textsuperscript{67}.

However, little is known about the performance of current commonly used estimation methods with other types of PD models.
Aims

The aim of the thesis was to explore methods and models involved in the pharmacometric analysis of discrete data: ordinal, count, and event data.

The specific aims were:

1. Approach evaluation
   - to investigate the benefits of treating ordinal data with a continuous or count approach,
   - to investigate the benefits of treating count data with different Poisson-related count models,
   - to investigate the benefits of treating event data with count or repeated time-to-event approaches.

2. Algorithm evaluation
   - to improve the knowledge concerning the performance of candidate estimation methods for dose-response continuous models,
   - to improve the knowledge concerning the performance of candidate estimation methods for Poisson-related count models,
   - to improve the knowledge concerning the performance of candidate estimation methods for frequency-varying repeated time-to-event models.

3. Modeling analysis
   - to develop novel models adapted for complex ordinal data involving serial-correlated pain scores,
   - to develop a novel model adapted to complex count data involving overdispersed epilepsy seizures,
   - to develop novel models adapted to complex event data involving simulated rated events.
Methods

1. Studied data

In this thesis, many different types of data were studied: continuous, ordinal, count, and event. Fitting a model to these data types aimed either at model evaluation, estimation algorithms comparison, or model development. Because purposes differed, the data were extracted from either real or simulated trials, the latter case meaning that the true models from which they were derived were known.

1.1. Continuous data (Paper II)

The dataset structure used in Paper II mimicked a clinical trial involving 100 individuals and investigating four dose levels of a hypothetical drug: 0, 100, 300 and 1000 mg. Individuals had observations available at all the four dose levels in the rich design, and each individual was randomly allocated to only two of the four dose levels in the sparse design. The continuous response was simulated using a dose-response model based on a sigmoidal $E_{\text{max}}$ function:

$$E_{ij} = E_{0i} + \frac{E_{\text{max}i} \cdot d_j^\gamma}{ED_{50i}^\gamma + d_j^\gamma}$$  \hspace{1cm} (Eq. 10.)

where the sigmoidicity factor $\gamma$ had a value of either 1, 2 or 3. The statistical model included a residual variability representing measurement error:

$$Y_{ij} = E_{ij} + \varepsilon_{ij}$$  \hspace{1cm} (Eq. 11.)

where $\varepsilon_{ij}$ are independent and normally distributed with zero mean and variance $\sigma_{ij}^2$, which was additive: $\sigma_{ij}^2 = \sigma^2$, or proportional: $\sigma_{ij}^2 = E_{ij} \cdot \sigma^2$. Eight dataset scenarios were derived exploring combinations of i. rich (R) and sparse (S) simulation designs, ii. sigmoidicity value $\gamma$, and iii. additive (A) and proportional (P) error models: R1A, R2A, R3A, R1P, R2P, R3P, S3A, and S3P.
1.2. Ordinal data (Papers I & III)

1.2.1. Pain scores from placebo patients

The placebo group of three randomized, double-blind, placebo-controlled parallel-group multi-center Phase 3 clinical trials for an analgesic drug were considered. A total of 231 patients diagnosed with diabetes mellitus and experiencing symptoms of painful distal diabetic neuropathy were randomized into the placebo group. Patients were allowed to use acetaminophen as a rescue medication at any time. In total, 22,492 observations were made during 18 weeks of titration and maintenance phases of the clinical trials. The response variable was neuropathic pain intensity. The patients were provided an 11-point Likert scale as part of their diary and self-rated their pain on a daily basis.

1.2.2. Simulated pain scores

In Paper I, data from an observational trial were simulated with the same design implemented as the trial described above. The response was simulated from a simple ordered categorical model with parameter values determined by fit to real data.

In order to generate a placebo-controlled parallel-group treatment trial, the 231 individuals of the dataset were randomized in silico to one of four dose groups of the analgesic drug: 0, 100, 200, or 300 mg. A dose-driven drug effect was simulated by a linear inhibition implemented on the logit of the cumulative probabilities in a proportional odds model.

1.3. Count data (Papers IV-VI)

1.3.1. Epilepsy seizures from placebo patients

Data from a randomized, double-blind, placebo-controlled, parallel-group, multi-center clinical trial were analyzed in Paper IV. In this trial, epileptic patients suffering from medically refractory partial seizures were included. Only data from the 12-week baseline phase were analyzed considered, during which 551 patients entered the study and were maintained on a common standard epileptic treatment. Daily seizure activity, self-reported in a diary according to instructions, constituted the response.

The dataset structure used in Paper V was derived from this real trial. Six extensions of the Poisson model were first fitted to the baseline data and then used for simulations, with the simulation parameters set to the obtained final estimates. The series of generated responses were therefore mimicking observations from the population described above.
1.3.2. Epilepsy seizures from treated patients

Data in *Paper VI* were collected in three randomized, double-blind, placebo-controlled, parallel-group, multi-center clinical trials, during which the efficacy and safety of pregabalin as an add-on treatment was evaluated. The study population involved 1,053 patients with refractory partial seizures who were also given concomitant antiepileptic co-medication such as carbamazepine, lamotrigine, phenytoin, valproic acid, and topiramate. All studies had a nominal 8-week baseline phase followed by a 12-week treatment phase. A total of 61,198, 83,515, and 134,411 daily seizure counts were reported by the patients from the three respective studies. The evaluated doses of pregabalin were 50, 150, 300, and 600 mg daily.

1.4. Event data (*Papers VII-IX*)

1.4.1. Gastroesophageal symptoms from treated dogs

The preclinical study in *Paper VII* consisted of cross-over experiments on adult male and female Labrador retrievers to study the compound WIN55251-2, a cannabinoid receptor agonist. Nine dogs, on which cervical esophagostomies had been performed, received gastric infusion of an acidified liquid nutrient followed by air insufflation. All the dogs received an intravenous dose of 0.5 mL/kg of vehicle 0.9% NaCl, and four of them were also intravenously administered 0.015 mg/kg of WIN55251-2. Specifically defined transient lower esophageal sphincter relaxations (TLESRs) were measured and their time of occurrence was recorded during 45 min. A total of 294 TLESRs were detected in 32 vehicle experiments and 66 TLESRs in 15 drug experiments. In 8 of the 15 compound experiments, plasma concentrations were measured in samples collected at predetermined time-points up to 180 min.

1.4.2. Simulated gastroesophageal symptoms

Simulations studied in *Paper VIII* were based on a design consisting of 120 individuals observed over 12 days with possible event-records every hour. A repeated time-to-event model with a constant hazard and an exponential interindividual variability was used to simulate the response. The fixed effect was varied so that the frequencies of events resulted in a range of 5 to 100% of individuals experiencing at least one event. Different random effect values were also studied, corresponding to coefficients of variation ranging from 0 to 200%. A total of 45 scenarios, exploring rare to frequent reporting of events and low to high variability within the population, were simulated.
A hypothetical Phase IIa clinical trial studying the efficacy of a treatment for heartburns was simulated in Paper IX. The design included 72 individuals equally allocated to placebo or one of the five treatment dose levels: 10, 50, 100, 200 or 400 mg. Doses 100 and 400 mg were ignored during the estimation step and retrieved for resimulations in order to explore interpolation and extrapolation properties, respectively. Individual pharmacokinetic parameters derived from a one-compartment first-order absorption model were simulated separately and then used to generate observations. The observed response, recorded during a period of 12 h, was the time of occurrence of the heartburns and also their degree of severity, which was recorded as either absent, mild, moderate or severe.

2. Modeling analyses

All real data were analyzed with a mixed-effects approach. The models were based on the models described above, i.e. logistic, Poisson, RTTE, but were extended by implementing new components to handle specific data characteristics.

2.1. Model development

Analyses were performed in the software NONMEM versions 6.2.0 or 7.1.0. Estimation algorithms that were used were FOCE for the continuous model and LAPLACE for the discrete models. Model building was conducted sequentially, with the structural and the statistical components of the baseline model being developed first on placebo data, followed by the implementation of time-course functions. Finally, drug effects were explored in datasets containing dose administrations. Model discrimination between hierarchical models was based on the likelihood ratio test applied to minimum objective function values (OFV). This means that the inclusion of an additional parameter corresponding to one degree of freedom is considered as an improvement when the chi-squared distributed difference of OFV is statistically significant. Hierarchical models are nested models that can be obtained by restricting a parameter in a more complex model to be zero. Model selection was conducted together with model diagnostics, described below.

2.2. Continuous model (Paper III)

Pain scores rated on a Likert scale by patients on placebo treatment were modeled by either a continuous or a count model. The structure of the continuous model involved a logit-transformation and serial correlation.
2.2.1. Logit-transformation

Because the data were constrained between 0 and 10, and the predictions of a continuous model can potentially be infinite, a logit-transformation was applied and rescaled:

\[
Y_{ij} = 11 \cdot \frac{e^{\ln\left( \frac{\lambda_i}{1-\lambda_i} \right) + \varepsilon_{ij}}}{1 + e^{\ln\left( \frac{\lambda_i}{1-\lambda_i} \right) + \varepsilon_{ij}}} - 0.5 \quad \text{(Eq. 12.)}
\]

where \( \lambda_i \) was the individual prediction and \( \varepsilon_{ij} \) the residual variability additive on the logit-scale. Such a transformation can easily be transferred to another interval.

2.2.2. Serial correlation

The observations were suspected to be dependent between subsequent measures. One way to address this characteristic in continuous models is to introduce auto-correlation between residual errors\(^70\). A model that was previously suggested and successfully implemented\(^71, 72\) is the autoregressive time series AR(1):

\[
\text{Corr}(\varepsilon_{ij}, \varepsilon_{ik}) = \exp\left(-\frac{\ln(2)}{\text{AR} t_{1/2}}(t_k-t_j)\right) \quad \text{(Eq. 13.)}
\]

where the correlation between two subsequent residual errors \( \varepsilon_{ij} \) and \( \varepsilon_{ik} \) at times \( t_j \) and \( t_k \) is exponentially decreasing with time according to a half-life \( \text{AR} t_{1/2} \). If observations are simultaneous, the correlation is assumed maximal and the more distant in time they are, the less correlated.

2.3. Count models (Papers III-IV & VI-VII)

Count models were applied to almost all of the data described above: pain scores, epilepsy seizures, and TLESRs. However specific characteristics that were encountered in the data often necessitated adaptations to the model.

2.3.1. Truncation

Due to interval-constraints (Figure 4), the count model applied to Likert data was truncated:

\[
P(Y_{ij} = m) = \frac{P(Y_{ij} = n)}{\sum_{n=0}^{10} P(Y_{ij} = n)} \quad \text{(Eq. 14.)}
\]

where \( P(Y_{ij} = n) \) is the discrete probability distribution allowing an infinite number of counts \( n \), and \( P(Y_{ij} = m) \) is restricted to a finite number of scores \( m = 0, 1, ..., 10 \). This transformation ensures \( \sum_{m=0}^{10} P(Y_{ij} = m) = 1 \) as opposed to a right-truncation without normalization.
2.3.2. Underdispersion

The equidispersion assumption described above was violated with the Likert data, partly as a consequence of the truncation. Underdispersion\textsuperscript{73}, defined as a variance smaller than expected from a Poisson distribution with the predicted $\lambda$: $E[m] > Var[m]$, was observed for the majority of the patients (Figure 5). Extended versions of the Poisson model were then explored.

- The generalized Poisson model\textsuperscript{74} $Y_{ij} \sim GP(\lambda, \delta)$ permits to account for underdispersion:

$$P(Y_{ij} = m) = \frac{\lambda_i (1 - \delta_i) \cdot (\lambda_i (1 - \delta_i) + m \cdot \delta_i)^{n-1} \cdot e^{-\lambda_i (1 - \delta_i) + m \cdot \delta_i}}{m!}$$

(Eq. 15.)

where $\delta$ is a dispersion factor defined in $[\max(-1, -\lambda/(10 - \lambda)), 1]$, which in turn governs the variance $\text{Var}[m] = \lambda_i / (1 - \delta_i)^2$. This additional parameter specifically accounts for underdispersion when it is negative and makes the model collapse to a Poisson model when equal to zero.

Figure 4. Distribution of pain scores raw data. Solid lines represent arithmetic means over all patients, and shading intensity corresponds to the frequency density.

Figure 5. Dispersion in pain scores raw data. The dots are the individual score variance vs. individual score arithmetic mean. The dashed line is a loess smooth curve.
2.3.3. Overdispersion

In contrast to the pain scores, the epilepsy seizures, like many other clinically observed symptoms, presented a phenomenon of overdispersion\textsuperscript{75}, $E[n] < Var[n]$. Several models addressing this have been introduced and were investigated in this thesis (Figure 6).

![Figure 6](image)

_Figure 6._ Epilepsy seizures raw data from _Paper IV_, as well as resimulations from Poisson, zero-inflated Poisson, and negative binomial models. Circles represent individual variance vs. mean number of counts $\lambda$ pairs; the black solid and grey dashed lines show the identity line and a smooth line through the data, respectively.

- The GP model can handle overdispersion when $\delta$ is greater than zero.

- The zero-inflated Poisson model $Y_{ij} \sim ZIP(\lambda, p_0)$ model is a mixture model adapted to data presenting an excess of zero values and therefore needing an adaptation of the distribution that is otherwise extremely skewed:

$$P(Y_{ij} = n) = \begin{cases} p_{0i} + (1 - p_{0i}) \cdot e^{-\lambda_i} & n = 0 \\ (1 - p_{0i}) \cdot e^{-\lambda_i} \cdot \frac{\lambda_i^n}{n!} & n > 0 \end{cases} \quad (Eq. 16.)$$
where $P_0$ is the probability of the dependent variable to be zero. The variance is then changed to $\text{Var}[n] = (1 - P_{0i}) \cdot \lambda_i$ for $n = 0$ and $\text{Var}[n] = (1 - P_{0i}) \cdot \lambda_i \cdot (1 + P_{0i} \cdot \lambda_i)$ for $n > 0$.

- The negative binomial model\textsuperscript{76} $Y_{ij} \sim \text{NB}(\lambda, \text{Ovdp})$, sometimes also referred to as the inverse binomial model, is used when there is overdispersion due to heterogeneity in the responses:

$$P(Y_{ij} = n) = \left[ \frac{\Gamma(n + \frac{1}{\text{Ovdp}_i})}{\Gamma\left(\frac{1}{\text{Ovdp}_i}\right) \cdot n!} \right] \cdot \left( \frac{1}{1 + \text{Ovdp}_i \cdot \lambda_i} \right)^{\frac{1}{\text{Ovdp}_i}} \cdot \left( \frac{\lambda_i}{\lambda_i + \frac{1}{\text{Ovdp}_i}} \right)^n$$  \hspace{1cm} (Eq. 17.)

where $\text{Ovdp}$ is the parameter accounting for the degree of overdispersion and $\Gamma(\cdot)$ denotes the gamma distribution. The variance is $\text{Var}[n] = \lambda_i \cdot (1 + \text{Ovdp}_i \cdot \lambda_i)$.

- The ZIP and the NB model can also be combined into a zero-inflated negative binomial model $Y_{ij} \sim \text{ZINB}(\lambda, P_0, \text{Ovdp})$.

All models presented above collapse to a Poisson model when the additional parameter is equal to zero.

2.3.4. Serial correlation

Both epilepsy seizures and pain scores presented serial correlation between consecutive observations, as new types of diagnostics highlighted: Figure 7 and Figure 8, respectively. For the modeling of discrete data, serial correlation is handled by including Markov components\textsuperscript{77, 78} in the model. Markov components can act as covariates; they consist of variables related to previous observations which act as predictors for the current observations.

![Figure 7](image.png)

\textit{Figure 7.} Epilepsy seizures raw data from \textit{Paper VI}: concordance plot of current seizure counts vs. previous seizure counts. The darker and bigger dots represent higher frequency of such combinations.
When Markov components relate to only the preceding observation, they are called first-order, when they relate to the two preceding observations, it is second-order, and so on. In this thesis, only first-order Markov components were implemented, either because components of a higher-order were found to be non-significant predictors, or for simplicity.

However, types of Markov components studied in the literature are of limited complexity, although there are several aspects of the preceding observation that can be of some predictive value: the absence of event, the number of events, the difference between the values of the two observations, the time since the preceding observation, and others. In this thesis, the first three types were applied to real data in three papers. Although the structural model may not be considered a regular Poisson, equations 18-20 adopt this simple case.

Markov components, based on the presence or absence of an event at the preceding observation, were introduced for the epilepsy seizures studied in Paper IV. It demanded the introduction of additional parameters, a first set being applied when there were no seizures the preceding day, a second one when there was at least one:

\[
P(Y_{ij} = n | Y_{i(j-1)} = 0) = \frac{\lambda_{1i}^n \cdot e^{-\lambda_{1i}}}{n!}
\]

(Eq. 18.)

\[
P(Y_{ij} = n | Y_{i(j-1)} > 0) = \frac{\lambda_{2i}^n \cdot e^{-\lambda_{2i}}}{n!}
\]

where either \(\lambda_{1i}\) or \(\lambda_{2i}\) enter the model depended on a dichotomized interpretation of the preceding day.
Markov components based on the number of responses observed the preceding day were implemented in the modeling of epilepsy seizures carried out in *Paper VI*. The implementation was a proportional E\textsubscript{max} function:

\[
P(Y_{ij} = n | Y_{i(j-1)}) = \frac{\text{BAS}_i \left(1 + \frac{P_{\text{max}} \cdot Y_{i(j-1)}}{\text{PC}_{50} + Y_{i(j-1)}} \right)^n \cdot e^{-\text{BAS}_i \left(1 + \frac{P_{\text{max}} \cdot Y_{i(j-1)}}{\text{PC}_{50} + Y_{i(j-1)}} \right)}}{n!}
\]

(Eq. 19.)

where \(P_{\text{max}}\) is the maximum influence of preceding count on \(\lambda\), and \(\text{PC}_{50}\) the preceding count for reaching 50% of \(P_{\text{max}}\).

Markov components corresponding to the transition value, *i.e.* the difference between the current observation and the preceding one, were used for modeling the pain scores in *Paper III*. The probability of the absolute value of certain possible transitions was estimated, which inherently implies symmetry in the probability of transitions between subsequent scores. This model is equivalent to a transition-inflated model:

\[
P \left( Y_{ij} = n \mid |Y_{ij} - Y_{i(j-1)}| = z \right) = \pi_z + \left(1 - \sum_{k=0}^{K} \pi_k \right) \cdot \frac{\lambda_j^n \cdot e^{-\lambda_j}}{n!}
\]

(Eq. 20.)

where the discrete distribution is modeled for each value of transition \(z\) in \(k = 0, 1, \ldots, K\) and the corresponding transition probability parameters estimated are \(\pi_z = P \left( |Y_{ij} - Y_{i(j-1)}| = z \right)\).

Time-effects can also be introduced in Markov components. This was done when calculating the probability of transitions equal to zero in the case described above. Accordingly, \(\pi_0\) was a function of time.

**2.3.5. Interval construction**

Since count data are defined as the number of events happening during a specific time-interval, the duration of time-intervals is of importance. It determines the resolution of the data, and impacts the modeling. The TLESRs measured in *Paper VII* were represented as counts calculated over different time-interval lengths. The experiment duration was 45 min, and the datasets were established as TLESRs reported for either 5-min or 1-min intervals.

**2.4. Event model (Paper VII)**

The preclinical experiments studying gastroesophageal symptoms were also analyzed with an event model. Numerous TLESRs were observed during the experiment, thus an RTTE approach was applied, with the cumulative hazard reset at each event. The censoring at the end of the experiment was handled by assuming no event had occurred between the last observed event and the 45\textsuperscript{th} min.
2.5. Model diagnostics

Numerical model evaluation was carried out through the assessment of model stability, shrinkage in random effects, and uncertainty of parameter estimates. The stability of the estimates was systematically checked by comparison of convergences with perturbed initial values. Another way to verify model stability is to re-estimate simulations as described below. The shrinkage in random effects can be calculated or obtained from the output of NONMEM version 7.1.0. This phenomenon of shrinkage, which describes the empirical Bayes estimates (EBE) shrinking towards the population median, usually concerns sparse data, rare in discrete reporting. The uncertainty of parameter estimates is typically measured as the relative standard error (RSE). Standard errors are reported in the NONMEM output after implementation of the covariance step. In version 7.1.0, there is also the possibility of obtaining more accurate estimates of standard errors by using IMP. An even better method is bootstrapping, a non-parametric method. It is however computational-intensive, due to repeated random sampling and re-estimation.

Graphical model evaluation has an important place in this thesis. Use of prediction-based, residual-based and EBE-based diagnostics is very common for continuous models evaluation. For evaluating the sole continuous model in this thesis, conditional weighted residuals (CWRES) were used. CWRES provide the difference between the observations and the predictions weighted with the error magnitude and based on the FOCE method. Until to date, they are the most valued residual-based diagnostics, together with normalized prediction distribution errors (NPDE) to evaluate mixed-effects models. Simulation-based diagnostics, on the other hand, are the primary choice to evaluate discrete model evaluation. It is informally recognized that single-simulation diagnostics can be sufficient during model building, but that computational-intensive simulation-based diagnostics are required for the evaluation of the final model. Single-simulation diagnostics are called mirror plots and are seldom presented in publications; graphical computer-intensive simulation-based diagnostics are essentially visual predictive checks (VPC). The implementation and display of a VPC varies between schools and has evolved over the years. The VPCs presented in this thesis were mostly based on the comparison of the observed median, 5th, and 95th percentiles and the simulated 95% confidence interval (CI) for these descriptors, or the simulated 90% prediction interval (PI). They were facilitated by the R-based package Xpose. Simulations are compared to observations and the agreement is conventionally evaluated on graphics representing the dependent variable versus the independent variable, usually time. Less traditional VPCs of proportions of specific ordinal, count, or
transition values versus time were created, as well as variance versus mean. Moreover, the Kaplan-Meier plots comparing the observed survival probabilities to predicted using an RTTE model for the time to the first events were created using the VPC technique.

3. Simulation studies

When the true model from which simulations are performed is known, investigations concerning selection of suitable models or evaluation of estimation algorithms can be more easily made. So-called simulation studies often include stochastic simulations after which the model is estimated. The obtained estimates can then be used for subsequent stochastic resimulations. Such simulations studies, which can be automated using the Perl-coded program PsN, typically create a large number of copies of a specific trial, at least one hundred are recommended.

3.1. Approach evaluation (Papers I & IX)

Often candidate models have different structures, which makes it difficult or impossible to compare parameter estimates. This was the case in Papers I & IX. Additionally, because the models are non-hierarchical, the OFV cannot be used in a strict likelihood ratio test, but an extended likelihood-based criterion, e.g. the Bayesian Information Criterion (BIC), must be used. However, comparing how well the models are able to mimic the data, and therefore enable to detect a drug effect, is of more practical clinical relevance.

3.1.1. Simulation properties

In this thesis, the proportions of each possible score, eleven in the case of pain and four in the case of heartburns, were computed. The agreement between the originally simulated “true” proportions and the resimulated ones can be evaluated through metrics, graphs like correlation plots or stacked bars, (visual) predictive checks based on number/extent of transitions, or days/individuals without an event, to name a few.

3.1.2. Statistical power

As the power of a study depends on the statistical significance criterion, the investigated effect size, and the sensitivity of the data, it is interesting to investigate, for a given trial design, the study power associated with different models. Any gain in power that can be achieved using a more informative analysis increases the potential to identify a true drug effect, decreases
the study size without loss of power. The difference in OFV between runs incorporating, and ignoring a drug effect in the model, was used to define the power reached with the different models, in which a 1% confidence level chi-square test was applied. In Paper IX, several predetermined study sizes were explored, and in Paper I, all possible sizes were evaluated with a rapid new type I error calibrated method, named Monte Carlo mapped power (MCMP). This method is based on sampling of individual OFVs obtained from the simulation and re-estimation of one large dataset, instead of performing multiple simulation and re-estimation cycles.

3.2. Algorithm evaluation (Papers II, V & VIII)

Algorithms treated in this thesis are all based on different ways to determine the maximum likelihood, therefore it would be incorrect to compare these values. Efforts were made to evaluate runtimes between the different algorithms. However investigations in Papers II, V, & VIII were often imperatively performed on computers with different central unit processing (CPU) frequencies. Stability of the estimation algorithms was also evaluated by the proportion of runs converging and producing parameter estimates. But aspects considerably more important in algorithm evaluation are the accuracy and the precision with which parameters are estimated and the impact of potential bias and imprecision on simulations.

3.2.1. Parameter estimation

Several statistical computations were performed for each parameter in order to evaluate the performance of each estimation algorithm: relative estimation errors (RER), relative bias, and relative root mean squared error (relative RMSE or RRMSE). Typically, the RERs are evaluated for each simulation re-estimation cycle; box-plots of RER show both bias, through the mean, and imprecision, through the width of the box and whiskers. The relative bias describes the deviation of the mean of the estimated parameters from their true value. The relative RMSE summarize both the bias and the variability in estimates. The Standardized RRMSE was constructed for each parameter and each approach as the RRMSE divided by the lowest RRMSE value obtained across all approaches for that parameter.

\[
\text{RER}_k(\psi_{p,a}) = \left( \frac{\hat{\psi}_{p,a} - \psi_{p,*}}{\psi_{p,*}} \right) \cdot 100 \quad \text{(Eq. 21.)}
\]

\[
\text{Relative bias}(\psi_{p,a}) = \frac{1}{K} \sum_{k=1}^{K} \text{RER}_k(\psi_{p,a}) \quad \text{(Eq. 22.)}
\]
where \( p \) are the parameters composing the vector \( \Psi \) and \( a \) the different algorithms tested over \( K \) datasets. All these statistical metrics scrutinized the estimated \( \hat{\Psi} \) compared to the “true” \( \Psi^* \).

### 3.2.2. Prediction properties

To assess the effect of the potentially biased and imprecise parameter estimates, the distribution of events in the original simulated datasets was compared with the distribution in the datasets simulated based on the estimated parameters. To minimize the effect of random sampling within the stochastic simulations, each set of parameter estimates was simulated several times, and the results were concatenated to form one large dataset. Characteristics of the estimated models that were implemented to handle the specific data features were investigated by calculating the variance and the mean of the count data (Paper V) and the normalized proportion of individuals without events (Paper VIII).
Results

1. Discrete versus continuous data

1.1. Modeling approaches (Paper I)

The baseline response simulated with an ordered categorical modeling approach was successfully estimated utilizing three modeling approaches: ordered categorical, count and continuous. The count consisted of a truncated generalized Poisson distribution and the continuous of a logit-transformed function. These two models involved four and three parameters, respectively, as opposed to twelve needed with the ordered categorical. The original simulated mean score of 5.2 with the standard deviation of 2.3 were accurately retrieved by all three approaches. Scores were also recovered and their proportions in each resimulated dataset were compared to those in the analyzed datasets. While resimulations from the ordered categorical model were in perfect agreement with the original simulations, small imprecisions were in evidence with the count model resimulations, and a noticeable bias was present for some scores resimulated with the continuous model, especially at the outer regions of the scale. The RMSE of these proportions confirmed this tendency, which ranked the various methods as: ordered categorical, count, and continuous, from the lowest to the highest error value, respectively.

Data simulated from the ordered categorical dose-response model enabled the testing of the inclusion of drug effect in the count and the continuous modeling approaches. The ordered categorical model structure with a linear inhibition on the logit scale, which was used for the original simulations, was kept for estimation. The count and the continuous approaches best described the data profile with a second-degree polynomial function which included two constant coefficients and a common IIV. The polynomial drug effect was added on the logit scale in both cases, the logit of the mean score λ for the count model and the logit of the mean prediction for the continuous model. The simulated distribution, which became more positively skewed with increasing dose levels, was adequately described with the tested approaches. The correlation between resimulated and original proportions of scores was strong for each dose arm with the “true” ordered categorical model (Figure 9). With the count model, it was more variable but still cen-
tered on the line of identity, while with the continuous model extreme scores were slightly overpredicted and central scores underpredicted.

Figure 9. Proportion of scores after resimulation vs. original simulation. Proportions are stratified by dose arm; dose levels from top to bottom are 0, 100, 200, and 300 mg. Resimulations were obtained from the ordered categorical (OC), the count (PO), and the continuous (CO) dose-response models. The ellipses represent the 95% confidence regions of the simulated proportions (grey dots). From the dots (0) to the solid line (10), the longer the dash of the ellipse the higher the score.
The type I error rate associated with the PO model was by far the largest, and therefore substantially decreased the power associated with this model. The power of significantly detecting the drug effect simulated in the treatment trial correlated with the number of individuals included in the trial (Figure 10). For a confidence level of 1%, all approaches reached a power greater than 80% with a study size including more than 50 individuals. The CO model appeared to be associated with the highest power, also when the treatment trial included only two dose arms, when the simulated drug effect was diminished to half or when the variance in the drug effect was increased to double. The PO model showed lower power performances than OC, although similar with low variability and four dose arms. With a larger IIV in drug-effect or a more restrictive design, PO achieved a much poorer power than OC.

Figure 10. Statistical power to detect a drug effect at the significance level of 1% for the different study sizes. The squares represent the power for data analysis with the ordered categorical approach (OC), the circles with the count approach (PO), and triangles with the continuous approach (CO). Power calculated in the “Default conditions”, i.e. a simulated inhibitory drug effect (IIV%) of 0.045 (30%) detected in a treatment trial with 4 dose arms, “Half dose arms”, i.e. treatment trial with 2 dose arms (0 and 100 mg), “Half drug effect”, i.e. simulated inhibitory drug effect of 0.0225, and “Double variance”, i.e. simulated inhibitory drug effect IIV% of 60%.
1.2. Algorithm performance (Paper II)

Each of the eight sets of hundred simulated datasets was estimated with various maximum likelihood estimation algorithms. The eight scenarios corresponded to sigmoid $E_{\text{max}}$-derived models exploring rich and sparse simulation designs, low to high sigmoidicity factor, and additive and proportional error models. The same model from which simulated datasets were generated was used for estimation. Each dataset was analyzed twice, firstly with true initial conditions, i.e. starting estimate values set to original parameter values on which simulations were based, and secondly with altered initial conditions. The explored estimation algorithms were FOCE in NONMEM 7.1.0 and R 2.9.1, LAPLACE in NONMEM 7.1.0 and SAS 9.2, AGQ in SAS 9.2, and SAEM in NONMEM 7.1.0 and MONOLIX 3.1. They were mostly utilized with the default settings with which they are available in the different studied software, except for both the SAEM approaches which were ran with default and modified settings. They were referred to as FOCE_NM, FOCE_R, LAP_NM LAP_SAS, AGQ_SAS, SAEM_NM, SAEM_NM_its, SAEM_MLX, and SAEM_MLX_tun.

![Figure 11](image_url).

Figure 11. Relative estimation error for the parameter ED$_{50}$, for the 8 scenarios R1A, R2A, R3A, R1P, R2P, R3P, S3A, and S3P referring to simulation designs rich (R) or sparse (S), Hill factor values 1, 2, or 3, and residual error models additive (A) or proportional (P), with the estimations from true initial conditions and altered initial conditions. The boxplot represents the median (middle bar) and the interquartile range (box limits), with points for the mean (black) and the outliers (grey).
Completion was maximal for all algorithms except FOCE_R when analyses started from true conditions. With altered initial conditions, less than 50% of the analyses converged for five scenarios with FOCE_R, and one scenario with SAEM_NM. Runs were fastest with FOCE_NM and LAP_SAS, around 10 billion instructions. The slowest algorithm was AGQ_SAS, which was about 80 times slower than FOCE_NM. According to the RER calculated from estimates obtained with the investigated estimation algorithms (Figure 11), the parameter ED50 was accurately estimated under true conditions, but presented a lower precision for scenarios with $\gamma = 1$. The highest and most consistent biases were observed with FOCE_R, for the few scenarios that had sufficient converging runs. On the sparse design, ED50 was best estimated with AGQ_SAS, LAP_NM and FOCE_NM. When starting from altered conditions, most of the methods estimated ED50 similarly to when starting from true values on the rich design. However, all the SAEM results changed, sometimes drastically for the versions with default settings. On the sparse design, most of the methods obtained biased estimates, with the exception of AGQ_SAS, SAEM_NM, and FOCE_NM. The mean standardized RRMSE (Figure 12) was below 1.5 for most of the scenarios with all methods except FOCE_R for the rich design under true conditions. Under the sparse design, the Laplacian methods, AGQ_SAS, and SAEM_NM were the only methods having mean standardized RRMSEs below 1.5 for both error models. With altered initial conditions, the lowest mean standardized RRMSEs were obtained with FOCE_NM, and AGQ_SAS. With LAP_SAS and SAEM_MLX tun they were below 1.5 for all but one scenario; with SAEM_NM, SAEM_NM_its, SAEM_MLX and LAP_NM, they were elevated for about half the scenarios; and with FOCE_R they were above 1.5 for all scenarios.

![Figure 12. Mean standardized RRMSE obtained with each approach for the 8 scenarios and 2 initial conditions, on a semi-log scale. The star symbol (*) represents the R3A estimate from SAEM_NM_its above 300 units.](image-url)
1.3. Novel models (*Paper III*)

Two novel models were developed for the analysis of pain scores measured on a Likert scale by neuropathic patients.

**Markov count model**

A truncated generalized Poisson model was designed to address the interval-constraints as well as the underdispersion characteristics met in the data. The observed serial correlation (*Figure 8*) was accounted for by including Markov components. This introduced a transition-inflation element as described in Equation 20. Inflation probabilities of absolute transition values $k$ from 0 up to $K = 3$ were significant. Increasing transition magnitude was associated with decreasing estimated probabilities. Probability of stable scores, *i.e.* $k = 0$, was expressed as a piece-wise linear function defined from 12% to 55% depending on the preceding score value and with a time-effect of 0.01 days$^{-1}$ (RSE 7.8%).

**Autoregressive continuous model**

A logit-transformed continuous function was developed and predictions of scores were obtained by scaling and discretization of the inverse-logit. Autocorrelation was implemented by using an AR(1) model defined in Equation 13. The residual error, additive on the logit scale, had an estimated standard deviation of 1.8, with acceptable precision and epsilon shrinkage. The correlation between two consecutive observations regressed over time with a half-life of 0.93 days (RSE 15.7%), corresponding to a magnitude of 47% after one day.

A time-course expressed as a decrease of the scores over time was attributable to a placebo effect. It was described in the same manner in the two models where the expected mean score was an exponentially decaying function:

\[
\lambda_{ij} = \text{BAS}_i \cdot \left(1 - \text{PE}_{\text{max}}_i \cdot \left(1 - e^{-\frac{\ln(2)}{\text{PE}_{\text{t}_{1/2}}}_i} \right) \right) \quad \text{(Eq. 25.)}
\]

where BAS was the baseline Likert score estimated slightly above a score of 6, PE$_{\text{max}}$ the maximum placebo effect found to be 19%, and PE$_{\text{t}_{1/2}}$ the half-life with a value of 28 days according to the models.

The longitudinal pattern of the proportion of each score was accurately mimicked by scores generated with the models (*Figure 13*). Proportions of specific transitions between consecutive scores plotted over time (*Figure 14*) also showed agreement between observations and simulations. The CWRES
diagnostics (not shown) demonstrated the absence of any remaining correlation between consecutive residuals for the continuous model.

**Figure 13.** Visual predictive check of each of the scores time-courses. Simulations realized from the Markov count model. Observed lines are compared to generated 95% confidence intervals.

**Figure 14.** Visual predictive check of the transitions directions between scores. Simulations realized from the Markov count model. Observed lines are compared to generated 95% confidence intervals. The panels describe the scores and whether the following one is the same (“stable”), lower (“decrease”), or higher (“increase”).
2. Count data

2.1. Modeling approaches (Paper IV)

Epilepsy seizures during the baseline phase of a trial served as data to investigate count modeling approaches. Eight models in total were fitted to the data, exploring two characteristics: overdispersion and serial correlation. A Poisson (PS) model was compared to a zero-inflated Poisson (ZIP), a negative binomial (NB), and a zero-inflated negative binomial (ZINB) models in order to address overdispersion. The introduction of Markov components following Equation 18 was suggested and evaluated on all these models for its ability to handle serial correlation.

As expected, simulations from a fit with a PS model lacked overdispersion, illustrated by comparing the mean of counts to their variance, and serial correlation, displayed by too many transitions between days without any event and days with at least one: 25,129 vs. 14,682. The inclusion of a Markov component resulted in a decrease in the objective function value by 2,812 points and the number of transitions to get closer to the observed: 15,247 vs. 14,682 (Table 1). However this did not mimic the overdispersion.

Table 1. Posterior predictive check performed on the base population, and on 100 datasets simulated from the Poisson (PS), zero-inflated Poisson (ZIP), and negative binomial (NB) models. Tlap is defined as the elapsed time between days with seizures.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Raw</th>
<th>Simulated mean (10th, 90th percentiles)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PS</td>
</tr>
<tr>
<td>Number of transitions</td>
<td>14682</td>
<td>15247 (14892, 15632)</td>
</tr>
<tr>
<td>Mean Tlap (days)</td>
<td>3.7</td>
<td>3.25 (3.1, 3.4)</td>
</tr>
<tr>
<td>Number of days with</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 seizure</td>
<td>30209</td>
<td>26562 (25841, 27307)</td>
</tr>
<tr>
<td>1 seizure</td>
<td>8942</td>
<td>11503 (11236, 11764)</td>
</tr>
<tr>
<td>2 seizures</td>
<td>3714</td>
<td>4753 (4496, 4977)</td>
</tr>
<tr>
<td>3 seizures</td>
<td>1861</td>
<td>2210 (2038, 2373)</td>
</tr>
<tr>
<td>4 seizures</td>
<td>898</td>
<td>1146 (1020, 1283)</td>
</tr>
<tr>
<td>5 seizures</td>
<td>542</td>
<td>631 (545, 723)</td>
</tr>
<tr>
<td>6-10 seizures</td>
<td>971</td>
<td>845 (680, 1019)</td>
</tr>
<tr>
<td>11-15 seizures</td>
<td>300</td>
<td>101 (44, 158)</td>
</tr>
<tr>
<td>&gt;15 seizures</td>
<td>300</td>
<td>31 (1, 79)</td>
</tr>
</tbody>
</table>
Resimulations from fits with the ZIP, the NB, or the ZINB models without Markov components described, on the other hand, overdispersion, but not serial correlation. Therefore Markov components were added to the ZIP, the NB, and the ZINB models. This resulted in a decrease in the OFV by 6,775 and 10,751, for the ZIP and the NB models respectively, the ZINB model showing no further improvement compare to the NB. In this way, overdispersion, and serial correlation, were both addressed, since the number of transitions predicted with the ZIP and the NB models went down to: 14,509 and 15,259 vs. 14,682, respectively. Other results from the posterior predictive check (Table 1) showed agreement between the original data and the data simulated from the investigated models, e.g. days between seizure days or days with a certain number of seizures. The serial correlation appeared to be equally well handled by the ZIP and the NB models, although the OFV drop was larger with the latter. A visual predictive check of the variance vs. the mean of the events per individual (Figure 15) confirmed that the median of the NB simulations more closely resembled the median of the observations.

Figure 15. Visual predictive check represented in logarithmic scale and obtained from 20 simulation studies based on the best zero-inflated (ZIP) model and the selected negative binomial (NB) model. Solid lines in black correspond to the quartiles of simulated data; dashed lines correspond to the median of simulated (black) and raw (white) data. The area in grey covers the 75% PI of the raw data.
2.2. Algorithm performance (*Paper V*)

Six distributions were used to investigate candidate estimation algorithms for count models: a Poisson (PS) model, a Poisson model with Markov elements (PMAK), a Poisson model with a mixture distribution for individual observations (PMIX), a zero inflated Poisson (ZIP) model, a generalized Poisson (GP) model, and a negative binomial (NB) model. All were first fitted to the real case described above, and the final estimates were then used to simulate 100 datasets. Simulated data were then estimated with the same models, and initial conditions set to the nominal parameter values on which simulations had been based. Estimations were completed with the Laplace approximation in NONMEM VI (LAPLACE), the Laplace approximation in SAS 9.1 (GQ 1 point) and the Gaussian quadrature in SAS 9.1 (GQ 9 points).

Completion rates were higher in NONMEM, 100% for all models, than in SAS: 77-100% with GQ 1 point and 82-100% with GQ 9 points. Runtimes for AGQ methods increased by a factor of 13 when the number of quadrature points increased from 1 to 9. The relative bias was low for all fixed effects and for all estimation algorithms, but it was large for some random effects with LAPLACE. More specifically, parameters of PS, PMAK, and PMIX models estimated with LAPLACE presented a relative bias with an absolute value < 3% whereas parameters of ZIP, GP, and NB models obtained with the same estimation algorithm had a relative bias with an absolute value ranging from 0.09 to 25.87%.

![Figure 16. Relative estimation error obtained with LAPLACE (left) and Gaussian Quadrature 9 points (right) for ZIP, GP and NB models.](image-url)
A common tendency was identified for the ZIP, GP, and NB models (Figure 16): although in all cases the structural parameters and the variability in the \( \lambda \) parameter were accurately estimated (bias 0.14-8.24%), the estimation of the variability in the additional parameter \( P_0 \), \( \delta \), and Ovdp was poor (bias -25.87%, -15.73%, and -21.93%, respectively). Precision however was similar for all parameters. Estimation of these models with GQ 1 point gave identical results for ZIP and GP, and similar results for NB, where negligible differences were attributed to the approximation of the factorial and the gamma function. Upon increasing the number of quadrature points to 9, an improvement of the estimation of the variability of the additional parameter for all three models was seen: the relative bias in the random effect of \( P_0 \), \( \delta \), and Ovdp was reduced to -5.55%, 1.67%, and 4.18%, respectively.

Simulations from the final estimates obtained with LAPLACE and GQ 9 points were displayed graphically. The VPCs, as illustrated with the ZIP model for which the bias of the additional parameter was the highest (Figure 17), showed good accordance between observed and simulated data. When plotting variance versus mean the previously highlighted bias had little impact and was only noticeable at the tails of the areas presented.

![Figure 17](image)

*Figure 17. Visual predictive check of log(variance) vs. log(mean) estimated with LAPLACE and Gaussian quadrature 9 points (GQ) for the ZIP model. The grey polygon represents the quartiles of the primary simulated data, the white line the median, and the dashed lines the median and the quartiles of the initially simulated data.*
2.3. Novel model (Paper VI)

A model was developed to analyze the large dataset of epilepsy seizures and its specific characteristics. As previously observed, this type of response data showed overdispersion, exemplified by an abundance of days with no symptoms as well as instances of extremely high numbers of seizures resulting from acute repeated seizures. Therefore the chosen approach was a negative binomial distribution, as described in Equation 7. Consequently the structural parameters were $\lambda$ describing the mean seizure count and Ovdp characterizing the degree of overdispersion. However $\lambda$, estimated to 0.385 daily seizures (RSE 11.7%), was found to be influenced by a number of aspects:

$$\lambda_{ij} = \text{BAS}_i \cdot f(Y_{i(j-1)}) \cdot f(t_j) \cdot f(d_i) \quad \text{(Eq. 26.)}$$

where $Y_{i(j-1)}$ is the preceding count, $t_j$ the time, and $d_i$ the dose.

**Preceding count effect model**

Markov elements based on the preceding count were incorporated as defined in Equation 19 which reduced the OFV by 1,037 points.

**Time effect model**

Time effect was best modeled with an exponential model, which was stable, significant, and parsimonious compared to a linear or an asymptotic model:

$$f(t_j) = e^{KT \cdot t_j} \quad \text{(Eq. 27.)}$$

where $KT$ is the exponential constant for daily increase in baseline of 0.272 year$^{-1}$ (RSE 26.8%).

**Drug effect model**

A mixture in the population was identified classifying patients as responders or non-responders. Within the responders, the pregabalin effect was captured as a decrease in the daily count of seizures. The predictor of the drug effect was the dose which was best associated to the mean seizure through an exponential model:

$$f(d_i) = e^{-KD \cdot d_i} \left\{ \begin{array}{ll}
P_{\text{mix}} = P_{\text{res}} \quad \text{(responders)} \\
P_{\text{mix}} = 1 - P_{\text{res}} \quad \text{(non - responders)} 
\end{array} \right. \quad \text{(Eq. 28.)}$$

where $P_{\text{res}}$ was 66% and the exponential constant for pregabalin dose effect $KD$ was 0.00108 mg$^{-1}$, which corresponds to an ED$50$ of 642 mg if $E_{\text{max}}$ is 100%.

The VPC of the mean count (Figure 18) showed the observations to be mostly within the 90% PI of the simulations. Slight underpredictions were revealed at the early and late phases of the study, whereas the PI bow-tie shape was due to the relatively fewer number of patients at those stages. The categorical VPC showed the proportion of observations in each category to be
within the 95% CI for each dose. As an example, the VPC for 600 mg is shown (Figure 19).

![Mean visual predictive checks stratified by dose](image)

**Figure 18.** Mean visual predictive checks stratified by dose. The solid line represents the observed mean and is dotted at each observation. The shaded area and gray dashed line represent the 90% prediction interval and the median from the simulations.

![Categorical visual predictive check for dose 600 mg](image)

**Figure 19.** Categorical visual predictive check for dose 600 mg. The dotted line and shaded area specify the proportion of observed data and 95% confidence interval, respectively.
3. Event data

3.1. Modeling approaches (Paper VII)

Three approaches were applied to TLESR events collected from nine dogs: a count approach with events being counted in 5-min intervals, a count approach with events being counted in 1-min intervals, and a repeated time-to-event approach. Although the dataset structure was different between the three approaches, the main element for each, $\lambda$ for the counts and the hazard $h$ for the RTTE, took the same expression for all models:

$$\lambda_i = h_i(t_j) = BAS_i \cdot f(C(t_j)) + f(t_j)$$  \hspace{1cm} (Eq. 29.)

where serial correlation could not be identified, neither could interoccasion variability. The only parameter for which interindividual variability could be determined was the baseline. Baselines estimated with the 1-min count and the RTTE models were identical, with a value of 0.14 min$^{-1}$ (IIV 12%), and lower than that obtained with the 5-min count model with a value of 0.69 min$^{-1}$ (IIV 12%).

A time effect was apparent when visualizing the raw data; it consisted of a high event occurrence time-interval and could be attributed to the experimental set-up. A surge function previously described$^{89-91}$ and improved for shape estimation$^{92}$ was used within all three approaches:

$$f(t_j) = \frac{SA}{\left(\frac{t_j - PT}{SW}\right)^2 + 1}$$  \hspace{1cm} (Eq. 30.)

where SA is the surge amplitude, SW half of the surge width, PT the peak time, and $\gamma$ the shape parameter. PT was estimated to be $\sim$10 min by all approaches, whereas SW was estimated to be 2.5 min by both the 1-min count and the RTTE models versus 5 min by the 5-min count model.

![Figure 20. Visual predictive checks of frequency of events vs. time, comparing the observed data (line) to the 95% confidence interval of the simulated data (area) with the RTTE model, both for the vehicle (left panels) and the compound (right panels).](image)
The PK profile of the compound WIN55251-2 followed a one-compartment model with linear elimination, and corresponding individual PK parameters were imputed for the construction of the PKPD model. The PD profile was characterized as a function of WIN55251-2 concentrations:

\[ f(C(t_j)) = 1 - \frac{C(t_j)}{C(t_j) + IC_{50}} \]  

(Eq. 31.)

where \( I_{\text{max}} \), the maximal inhibitory effect, was fixed to 100%, and \( IC_{50} \), the concentration at half of the maximal inhibitory effect, was estimated. Estimates for \( IC_{50} \) were similar between the 1-min count, RTTE, and 5-min count approaches: 2.43, 2.20, and 2.53 nmol.L\(^{-1}\), respectively.

Simulations facilitated to determine agreement between observations and confidence intervals obtained by simulations with the tested approaches. With RTTE, a VPC of the time-course profile of TLESRs (Figure 20) enabled the visualization of the surge, both in dogs on vehicle and on drug. Kaplan-Meier plots (Figure 21) highlighted that all dogs on vehicle experienced at least 5 events, whereas some of the dogs on drug had no more than 2 events.

![Figure 21. Kaplan-Meier plots for time to the first eight events, comparing the observed data (line) to the 90% prediction interval of the simulated data (area) with the RTTE model, both for the vehicle (left panels) and the compound (right panels).](image-url)
3.2. Algorithm performance (*Paper VIII*)

The 45 simulated scenarios exploring rare to frequent reporting of events and low to high variability within the population, were used to evaluate the performance of LAPLACE, SAEM, and IMP in NONMEM. The population model was mu-parameterized according to the NONMEM guide and fitted to each of the datasets using the true parameter values, *i.e.* the parameter values used in the simulations, as initial estimates.

Overall, the stability of the estimations was high, inferred from the rate of successful minimizations with LAPLACE or convergence with SAEM and IMP. Stability was close to 100% in most scenarios, however, it was only 90% with SAEM for scenarios with IIV equal to 0. Runtimes were short, within seconds for most of the runs, but faster with LAPLACE with an average of 0.3 s, than with SAEM and IMP with an average of 19 and 23 s, respectively. The longest runtimes, over 100 s, were obtained with SAEM for scenarios with IIV equal to 0. The RMSE results (*Figure 22*) revealed that all three estimation methods performed similarly when the frequency of individuals with events was greater than 62%.

![Figure 22. RMSE for the three estimation methods vs. frequency of individuals with events, stratified on the magnitude of inter-individual variability. The top panel displays the relative RMSE of the fixed effect parameter $\theta$. The bottom panel displays the RMSE of the random effect parameter $\omega^2$. LP=Laplace and IMP=Importance sampling.](image-url)
When the frequency of individuals with events was less than 43%, LAPLACE resulted in considerably higher RMSEs than SAEM and IMP, between which no obvious distinctions could be made. The bias in fixed effect was negative when present, higher with LAPLACE, but absent when IIV was equal to 0. The bias and imprecision with the three algorithms were more pronounced for the random effect. The random effect was extremely poorly estimated with LAPLACE when the variability was low, and/or when the frequency of individuals with events was low. The SAEM and IMP methods exhibited a positive bias in the variability parameter at 30 and 50 CV%. When the frequency of individuals with events was < 30% the distribution of RERs from SAEM overlapped zero while it was skewed for IMP.

Simulations based on the estimated parameters indicated that, in general, the parameter biases did not affect the outcome of events. When comparing the normalized proportion of individuals without events (Figure 23), all investigated scenarios were centered around or close to the value 1, indicating that the original proportion of subjects with events and the resimulated proportion of subjects with events were in agreement, hence no appreciable effect by the parameter estimation errors was noted.

![Figure 23](image)

**Figure 23.** Normalized proportion of individuals without events, stratified on nominal frequency of subjects with events (i.e. when $\sigma^2=0$) and on inter-individual variability. A value = 1 indicates that the proportion of events are the same in the average of the original data and in the individual datasets simulated from true or estimated parameters; a value > 1 indicates fewer events in original data than in resimulated data. SIM=original simulations, LP=Laplace and IMP=Importance sampling.
3.3. Novel models (Paper IX)

In this thesis, two novel models enabling the simultaneous handling of event frequency and rated severity are introduced.

Repeated Time-To-Categorical Events (RTTCE) model

The RTTCE model is designed for data consisting of recordings of the exact time of occurrence of the event as well as its severity. The model combined a RTTE component and an OC component. The RTTE component captures the frequency of the events, and essentially is Equation 7, which corresponds to the probability of not having an event since the last reported one. The OC component is transformed so that the probability of an event to be of a severity m is conditioned on the probability of an event to occur:

\[
P(Y_{ij} = m) = \left( P(Y_{ij} \geq m) - P(Y_{ij} \geq (m + 1)) \right) \cdot P(Y_{ij} = 0) \cdot h(t_i) \quad (\text{Eq. 32.})
\]

which is the equation to apply during analysis, whereas when simulations are performed, the equation should be:

\[
P(Y_{ij} = m) = \left( P(Y_{ij} \geq m) - P(Y_{ij} \geq (m + 1)) \right) \cdot (1 - P(Y_{ij} = 0)) \quad (\text{Eq. 33.})
\]

where the fact that each possible time-point is contained in the dataset is taken into account.

Repeated Categorical Events per Time-interval (RCEpT) model

The RCEpT model is a solution to rated events summarized as maximal scores across time-intervals. The RTTE component remains unchanged, although it should here be interpreted as the probability of not having an event within a time-interval of predefined length. The OC component contains a count distribution, like a Poisson, determining the expected number of event occurrences \( \lambda \) during predetermined-length time-intervals:

\[
P(\max(Y_{ij}) = m) = \sum_{n=1}^{+\infty} \binom{n}{m} \cdot \frac{\lambda^n \cdot e^{-\lambda}}{n!} \cdot (P(\max(Y_{ij}) \leq m)^n - P(\max(Y_{ij}) < m)^n) \quad (\text{Eq. 34.})
\]

where the maximum severity m is observed over n number of events experienced during time-intervals and \( \lambda = H(t) \).

A study was conducted to evaluate different approaches; the RTTCE approach was used to simulate 500 datasets of reported heartburns. They were estimated with RTTCE, RCEpT, and OC models and these models were then used to resimulate data. The drug effect implementation used during simulation was composed of an inhibitory non-linear effect on the hazard, and a linear drug effect in the logit of the probability. The drug effect was driven by the concentrations following a 1-compartment first-order absorption model. Correlation between the two drug-effects was implemented.
Figure 24. Stacked-bar plots of original simulations (SIM), simulations with RTTCE, RCEpT, and ordinal (OC) models, stratified on dose level. Proportions of none, mild, moderate and severe maximums over 12h were obtained from 100 sets of parameter estimates and simulated on a large sample.

Only resimulated data from the RTTCE and RCEpT models were able to retrieve outcome measures similar to their original form. The OC model could simply generate the maximal heartburn rates per patient computed over each time-interval or the whole observation period. When compared to resimulations from the other models (Figure 24), this diagnostic however showed poor prediction properties from OC in cases where a drug effect was also expected to change event frequency. A response rate (RR) was defined as heartburn severity not greater than mild across the events occurring during the three last hours. The statistical power to detect the drug effect obtained with the RR approach was compared to the power with the OC model and the two novel models (Figure 25). The number of individuals needed to achieve 90%-power dropped more than 7-fold with all longitudinal modeling approaches compared to RR. Furthermore, with RTTCE or RCEpT models, all information contained in the data was used, which led to a further 25% decrease in the patients needed, compared to the OC approach.

Figure 25. Statistical power of detecting a given drug effect at the level of significance 1% plotted for RTTCE, RCEpT, OC, and response rate (RR) approaches vs. the number N of individuals included in the study.
1. Discrete versus continuous data

The choice of a modeling approach in accordance with the type of data encountered should be the first decision when starting a modeling analysis. Modeling options to treat ordinal data actually depend on the number of categories involved. Data analyzed in the past with an ordinal model predominantly consisted of 5-point pain relief scores\textsuperscript{93}, but more common and more reliable\textsuperscript{94} types of pain scales include a higher number of categories. In this thesis, Likert pain scores including 11 categories were used, and alternatives to an 11-parameter ordinal model were proposed. One proposed approach to model these data is by using a logit-transformed continuous function. Continuous models have previously been used for the analysis of pain scores, but for data recorded on a visual analog scale (VAS)\textsuperscript{95-97}, whose 101 categories can easily be regarded as continuous. Moreover, previously applied continuous models were not logit-transformed, although this feature ensures a constrained range. The second proposed approach is a truncated generalized Poisson distribution, which, to our knowledge, has never previously been used for score analysis. Results imply that these two modeling approaches are suitable for treating Likert data and recovering fitted scores. However a drug-effect simulated with one approach can not necessarily be identically retrieved with approaches of different types. Nevertheless, after expit-transformation of the concerned functions and logistic regression in the case of the ordinal model, the response was found to be very similar across approaches.

Once the approach has been selected, a specific algorithm implemented in a specific software needs then to be considered. While this choice may be highly influenced by tradition or availability, the pharmacometrician should consider the large panel of estimation methods that is available today. Estimation methods have been regularly compared to each other for more than a decade\textsuperscript{98, 99}, but the specificity from the results to the model and the emergence of new techniques justify additional studies. The first comparison in this thesis involved nine different estimation approaches represented by four algorithms often implemented in two of the four investigated software. The results obtained for the continuous model exemplify how different estimation methods can perform compared to each other and depend on model,
design and initial conditions particularities. Among the aspects related to an estimation method that the pharmacometrician may consider is robustness; in the studied case, it was clearly poorest with FOCE in R (nlme). Another aspect is speed, and although the difference in convergence criteria as well as the possibility of stopping rules for EM-based methods complicates comparisons, trends suggest that FOCE is fastest and AGQ slowest. The speed is driven by the extent of the likelihood function simplification, which for AGQ is linked to the number of quadrature points. But aspects that are of greater interest for the pharmacometric analysis outcome are accuracy and precision. Both are highest with AGQ and poorest with nlme. In NONMEM, SAEM performance decrease when the prior knowledge for parameter values is poor; performing an iterative two-stage approximation as a first step partly solves the problem. In MONOLIX, the SAEM algorithm requires the number of iterations and chains to be increased in order to achieve accuracy and precision when the design is sparse.

While the modeling analysis is performed, identifying and describing specific data characteristics linked to the disease or the design is crucial. One problem arising with an 11-parameter ordinal model is that it requires all categories to be present in the population. When this is not the case, the number of categories included in the model must be decreased, which can be misleading in situations of extrapolation or design testing. Another problem is the increased complexity of the inclusion of elements handling serial correlation, which are then often disregarded. Indeed, in the case where there is a continuous underlying process and frequent observations, serial correlation between measurements is expected; it was apparent in the study of daily Likert scores in patients with neuropathic pain, a chronic syndrome. In this thesis, serial correlation was handled through autocorrelation of the residual errors in the continuous model and through the implementation of Markov components in the generalized Poisson model. A first characteristic of the serial correlation was that it decreased over time between two observations, which was handled by the autoregressive autocorrelation model. A second characteristic was an increase in the incidence of consecutive scores of equal value along the trial, which can be expected both from the placebo effect plateau and trend pattern stabilization. This characteristic was not accounted for in the continuous model, but was described through the Markov inflated probability of zero transition in the generalized Poisson model. Developed models have the potential for serving as platform models, i.e. models characterizing symptom time-profiles in the absence of drug and incorporating consistent underlying features of the response. Appropriate platform models lead to increased-power clinical trial PD analyses. The major limitation met with the continuous model was the extremely long runtime, suggesting the exploration of stochastic differential equations (SDE) to account for the autocorrelation.
Besides methodology to handle data characteristics, diagnostic plots were presented in this thesis. They were tailored to illustrate model adequacy for specific features of the data, especially serial correlation.

2. Count data

Now that the potential of Poisson-based count models has been highlighted for ordinal data, more thorough explorations of this modeling approach remain to be carried out for count data. The second example of clinical observations presented in this thesis was epilepsy seizures. Most of the very few count approaches reported in literature were applied to this clinical area. In all reported cases, a Poisson distribution had been used to describe the data. In these analyses, an assumption of equidispersion was made, which is usually not valid. In fact, many count outcomes, and particularly refractory partial seizures, display larger intraindividual variability than predicted by the Poisson model. Several alternative approaches to the Poisson model were provided in this thesis and successfully implemented as mixed-effects models; they were a zero-inflated Poisson, a negative binomial, and a zero-inflated negative binomial. Nevertheless, to properly detect overdispersion features in count data, a suitable diagnostic was needed; consequently a plot relating the within-individual variance to the mean was proposed. This plot can be made from raw data for detection purposes, or from simulated data for model diagnostic purposes. One limitation of this representation in cases where the whole time-span of the study is taken into account at once is the assumption of stationarity of the mean count over time. The mean count is likely to vary with time due to placebo effect or active treatment; however, only the baseline phase of the clinical trial was included in the present study. Moreover, studies lasting for only a few months are unlikely to exhibit disease progression patterns. However, the problem can be solved by splitting the data into shorter time periods and producing several plots.

As previously described, the algorithm used in a modeling analysis together with the selected approach must present adequate performance for the specific type of approach. The second performance investigation in this thesis was completed in a similar manner to a previous study focusing on ordinal models. Six Poisson-based models fitted to the baseline phase data were explored with the Laplacian method, which is the recommended algorithm for discrete models in NONMEM, the most common software. Although accuracy and precision were excellent for most of the parameters in the majority of the models, a bias was observed with the zero-inflated Poisson, the negative binomial, and the generalized Poisson models. Hence, the bias concerned only the models adapted for overdispersion, and among their parameters, only the random effect of the additional dispersion parameter. When
bias is detected, it can be related to a poor estimation method, to a property inherent in the model, or to a combination of both. A property common to the investigated models concerned the distribution of the individual estimates for the dispersion parameter, which was shrunk and skewed. The eta-shrinkage was approximately of the same magnitude as the negative bias of the mean of the empirical Bayes estimates. The asymmetry revealed typical dispersion parameter values closer to low individual dispersion patterns than to high individual dispersion patterns. Unbiased results were achieved with AGQ, the alternative estimation method considered. The approximation of the maximum likelihood is more precise with AGQ and governed by the number of qpoints. The present investigation was performed with 9 qpoints, but one limitation resides in the long runtimes. Therefore with AGQ, a reasonable balance between speed and accuracy must be found.

Because the impact of the bias obtained with Laplace remained negligible in simulation studies, the algorithm was subsequently used for the analysis of count data from a clinical trial with active treatment. The investigated drug pregabalin, a structural but not functional analogue of the inhibitory neurotransmitter $\gamma$-aminobutyric acid (GABA)$^{108-110}$, is indicated as an adjunctive medication for the treatment of partial seizures. The efficacy of pregabalin compared to placebo had already been shown using standard endpoints$^{111,112}$ or using response ratio and responder rate$^{113}$, but these traditional statistical analyses were not optimal for characterizing exposure-response relationship nor variability in seizure frequency. A modeling approach using a Poisson count model had also already been performed for seizure counts$^{114}$, but using monthly counts, which limited the possibility of exploring time trends and serial correlation. The model-based clinical trial data analysis in this thesis was based on daily seizure counts and consisted of a negative binomial model combined with Markov elements handling serial correlation. The Markov elements, based on the number of seizures on the previous day and describing a strong tendency for more seizures to occur when there were many the previous day, corresponded to the notion of epileptogenic seizures$^{115}$. This feature was accompanied with an increase in the mean seizure tendency, which has been previously observed$^{116}$. The pregabalin effect was characterized and supported the presence of non-responders to pregabalin add-on therapy, a phenomenon called refractory epilepsy$^{117}$. The $ED_{50}$-converted drug effect parameter demands cautious interpretation, since the maximal effect was not observed during the study and since the drug effect was investigated only on the mean count parameter. Moreover, the pregabalin effect depends on the definition of non-responders and could be confounded with placebo effect or time effect. Additionally, the best predictor of the drug effect was dose, since dose was correlated to concentrations, i.e. linear with concentrations. Consequently daily average concentrations, not collected in the pregabalin study, added no model benefit compared to dose as a predictor.
Finally, several diagnostics suitable for count data were presented. The model development described in this thesis can serve as an example to assess the efficacy of antiepileptic drugs more adequately by simultaneously considering baseline status, time effect, drug effect, and non-responders.

3. Event data

One of the emphases of this thesis is that certain types of data can be treated with different approaches. In most of the presented studies, the focus is on count models, exploring their use in the analysis of ordinal data and overdispersed count data. A final type of data on which count models were tested is event data. The relevant example in this thesis for this data type was gastroesophageal symptoms, or more specifically TLESRs, detected as rapid prolonged decreases in lower esophageal sphincter pressure in the absence of swallowing.\textsuperscript{118, 119} The traditional analysis approach is based on the number of events counted over the entire experimental time-span.\textsuperscript{120} From this viewpoint, a modeling count approach is a refinement of the data information. This was further illustrated by exploring different time-interval lengths in which events were counted; the shorter the time-intervals, the more precise the information. In addition, the count model based on the minimum time resolution possible was demonstrated to be equivalent to a repeated time-to-event model, \textit{i.e.} treating data where the exact time of occurrence is reported. There are many limitations to the traditional approach, including the inability of describing any time-course, the imprecise or inaccurate characterization of the drug-effect, the lack of simulation properties, and the low statistical power to detect a drug effect. In contrast, the different tested modeling approaches enabled the depiction of a surge-shaped time-course, representing an increase in event frequency entirely attributable to the experimental set-up. However, the longer the time-intervals, the less accurate the surge-width estimate. Furthermore, all modeling approaches permitted the use of a dynamic representation of the exposure by a description of the concentration time-course using a PK model, as opposed to using the dose as the predictor of the drug effect.\textsuperscript{121} TLESR occurrences were highly inhibited by an effect of WIN55251-2. Moreover, several diagnostics were constructed, all being simulation-based VPC-type representations of the investigated data. They focused on confidence intervals around predicted medians given the limited quantification of the interindividual variability with nine dogs. This approach comparison therefore highlighted the importance of the underlying hazard in the characterization of event data. In the case of a time-constant hazard, counts in any time-interval length result in a similar description of the data to a repeated time-to-event approach, whereas in the case of a time-
varying hazard only counts with the highest resolution give the possibility to describe events equivalently to a repeated time-to-event approach. The count approach, however, has as advantages its simplicity and speed compared to repeated time-to-event approaches.

The increased use of the repeated time-to-event approach and the release of EM-based methods in NONMEM motivated an investigation of the estimation properties of Laplace, SAEM and IMP for the analysis of repeated time-to-event models. As described above, the Laplacian method displayed a bias in cases where the data information content was low, i.e. scenarios where less than 20% of the individuals within the population had an event. The bias mainly concerned the random effect parameter which was extremely inflated in many of the estimated datasets. Additionally, the fixed effect of the hazard was often underestimated. This highlighted the fact that the Laplacian method is highly dependent on the empirical Bayes estimates, and requires more information to allow the separation between the fixed and the random effects. However, because of the strong correlation between the estimates of the typical parameter and the variance parameter, there is a net effect on the outcome measurements, leading simulations to be close to the true data distribution. The two EM algorithms, SAEM and IMP, performed equally well with no appreciable bias in the baseline hazard. Nevertheless, a positive bias in the variance parameter for scenarios where few individuals experienced several events occurred. Parameters obtained with the EM methods were closer to the true values than those from Laplace. Furthermore, main features of the data such as average number of events per individual and normalized proportions of individuals without events were, relatively well recollected in the model for all algorithms.

Finally, attention was concentrated on a type of data constituted of two elements: an event element, with a time of occurrence, and an ordinal element, with a severity attributed to the symptom. Modeling approaches treating this type of data usually ignore the categorization of the data, dichotomize the information, or report the maximal severity within a specified time-interval. This represents a substantial simplification of the data. In this thesis, a novel model was introduced, the repeated time-to-categorical event (RTTCE) model, handling both aspects of the data simultaneously. A second novel model, the repeated categorical events per time-interval (RCEpT) model, was built in the same fashion to fit maximal severity grades per time-interval, but with the advantage over the ordered categorical model to consider the expected number of occurrences within the time-interval by using a Poisson distribution. Clearly, not only do the RTTCE and RCEpT models analyze the data in their true nature, but they are also able to reproduce realistic data profiles in simulation. An appealing aspect of the configuration of the RTTCE related models is the possibility to characterize a treatment effect either on the hazard of the events, on the probabilities of their grades, or on both. A drug may differentially affect the frequency and the severity of the
symptoms. Furthermore, this attractive feature also employs the advantage of simultaneous estimation by not conditioning one aspect on the other one and by allowing correlation between the two, i.e. frequency and severity.

In addition, diagnostics of these combined models for complex data structures were presented that depicted the dose-response curve of the proportions of scores as well as the number of events.
Conclusions

Methods involved in the pharmacometric analysis of ordinal, count, and event data were explored, and models evaluated or newly developed.

Specific conclusions are:

1. Modeling approaches
   - ordinal data can accurately be treated with a continuous or a count approach, both needing fewer parameters than an ordinal approach;
   - overdispersed count data can successfully be treated with a generalized Poisson, a zero-inflated Poisson, or a negative binomial model;
   - event data can equally be treated with a high resolution count or a repeated time-to-event approach.

2. Algorithm performance
   - the performance of nine estimation methods was assessed for eight dose-response continuous model scenarios;
   - Laplace and adaptive Gaussian quadrature are adequate for the estimation of equidispersed and overdispersed Poisson-related count models;
   - SAEM and importance sampling may be more adapted than Laplace for the estimation of low frequency repeated time-to-event models.

3. Novel models
   - novel continuous and count models handling underdispersion and serial correlation were developed for pain scores analysis;
   - a count model handling overdispersion and serial correlation was developed for epilepsy seizure analysis;
   - two novel models, RTTCE and RCEpT, simultaneously handling event frequency and rated severity were developed.
“By going beyond empiricism and stressing understanding, not data collection, we not only answer our first question, but we also gain far more.”
L.B. Sheiner

Pharmacometrics continues to evolve as a science and learning phases in drug development continue to gain in accuracy and efficiency through contributions about methods and models. Contributions in this thesis mainly focused on discrete data and were accompanied by guidance on model development and model diagnostics.
Acknowledgements

The work presented in this thesis was carried out at the *Department of Pharmaceutical Biosciences, Uppsala, Sweden.*

Financial support for my PhD studies was provided by *UCB Pharma SA, Braine-l’Alleud, Belgium.*

Educational traveling for courses and conferences was mostly funded by *Apotekarsocieten, Smålands nation* and the *European Union.*

I have been very fortunate to benefit from all kinds of support from numerous people, whom I would like to acknowledge here. Many, many thanks to:

**Prof. Mats Karlsson,** my supervisor and mentor. I am forever grateful for the trust, guidance, and inspiration I had the privilege to receive from you while I was learning to become a good scientist. Thank you for all the stimulating projects we worked on together and all the opportunities you shared with me. I hope this was only the beginning of a long friendship and fruitful collaboration!

**Dr. Laura Sargentini-Maier** and **Dr. Armel Stockis,** my co-supervisors. Thank you for giving me the chance to start a PhD and to be sponsored by UCB. I immensely appreciated your encouragements and understanding during the different phases of my Swedish stay and Belgium go/no-go.

**Dr. Mia Kjellsson,** my PhD tutor for the first few months. You eased my landing in the research group and made me feel good with wise advices.

**Prof. Sven Björkman, Dr. Lena Friberg** and **Dr. Andy Hooker,** my teaching tutors. I was gratified by the experience with the undergrads and the independence you gave me.

**Prof. Margareta Hammarlund-Udenaes, Dr. Siv Jönsson** and **Dr. Jörgen Bengtsson,** my Erasmus instructors. Thank you for starting all this!

**Prof. France Mentré** and **Prof. Nick Holford,** my half-time opponents. I really enjoyed our discussion and your valuable comments.
Dr. Raymond Miller, for welcoming me on-board on several projects and boosting me each time we would meet at a conference.

My co-authors, Dr. Kristin Karlsson, for great collaboration on Papers VIII & IX, Prof. Iñaki Trocóniz, for showing interest in my proposals for Papers IV & V, Dr. Jae Eun Ahn, for excellent work on Paper VI and making it get accepted on time, Al Maloney, for brilliant Gaussian quadrature contribution to Papers II & V, Yang Sun, for being an amazing master student and participating in Paper I, Dr. Jan-Peer Elshoff, for good input and feedback on Paper III, Dr. Julie Bertrand, for precious help in writing Paper II for a statistical journal, Dr. Guangli Ma, Dr. Mats Någård and Dr. Jörgen Jensen, for remarkable efforts and patience on Paper VII.

Dr. Kajsa Harling, Pontus Pihlgren and Magnus Jansson, for technical support. Without outstanding assistance on cluster, computers and posters, this thesis would be much shorter. Thanks for including my PsN suggestions.

Dr. Paul Westwood and Dr. Ron Keizer, for your exceptional proof-reading, tremendous kindness and helpfulness.

My roommates, Dr. Rada Savić, for your great support during the first 2 years, and Ami Mohamed, for nice office-sharing during the last 2 years.

ALL colleagues, past and present, for making the department such an interesting place to work in and creating this unique atmosphere; especially Jocke Nyberg, you’re welcome to our future Lipari consulting business 😊.

The international postdocs/researchers, Dr. Jan-Stefan van der Walt, for always being there, bringing chocolates and talking psychology, Dr. Rocío Lledó-García, Dr. Stefanie Hennig and Dr. Joe Standing, for fun dinners, birthday parties, disputation-after-parties at my place, baby-sittings, trips to The North Pole, to Oslo, to Moscow, etc...

La French mafia, Dr. Paul Baverel, for making me feel like a big sister to you, Dr. Émilie Hénin, Brigitte Lacroix and Camille Vong, for sometimes much needed cheering, caring and comforting with a sensitive touch.

My dearest friends, Martin Bergstrand, for not only being a clever co-worker but also a marvelous soul, and I actually think you’re very funny too, and Angelica Quartino, for lovely dinners in your kitchen, much aid with Swedish matters, and, last but not least, admirable initiation to Lindy Hop.

Seb, for entering my foolish life, understanding my silly moments and following me in crazy adventures.
My friends from home, Lola, Anne, Aurélie, Angélique, Marie, Claire, Cécile et Capucine, for not forgetting me despite the distance, for remembering important events in my life, for reminding me the true person I am, for being such good examples of smart, simple and successful women, for getting married and making babies so that I had to come back to celebrate, and so much more…

My family, Jul', Mamie, Grand-Papa, So', Anne, Frantz, Isa, Éric et Domie, for always making sure I would make it for Christmas and for keeping Martizay and Kerjuanno the most peaceful places on earth I know!

My parents, Maman et Papa, for sometimes telling me you are proud of me.

Elodie-
References


_Elodie L. Plan_


57. *MONOLIX (MOdèles NOn LInéaires à effets miXtes).* Orsay, France; 2008.


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy 145

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”.)