UPPSALA
UNIVERSITET

Development, Application and Evaluation of Statistical Tools in Pharmacometric Data Analysis

LARS LINDBOM

Dissertation presented at Uppsala University to be publicly examined in B22, BMC, Uppsala, Friday, May 12, 2006 at 09:15 for the degree of Doctor of Philosophy (Faculty of Pharmacy). The examination will be conducted in English.

Abstract

Lindbom, L. 2006. Development, Application and Evaluation of Statistical Tools in Pharmacometric Data Analysis. Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy* 33. 46 pp. Uppsala. ISBN 91-554-6544-7.

Pharmacometrics uses models based on pharmacology, physiology and disease for quantitative analysis of interactions between drugs and patients. The availability of software implementing modern statistical methods is important for efficient model building and evaluation throughout pharmacometric data analyses.

The aim of this thesis was to facilitate the practical use of available and new statistical methods in the area of pharmacometric data analysis. This involved the development of suitable software tools that allows for efficient use of these methods, characterisation of basic properties and demonstration of their usefulness when applied to real world data. The thesis describes the implementation of a set of statistical methods (the bootstrap, jackknife, case-deletion diagnostics, log-likelihood profiling and stepwise covariate model building), made available as tools through the software Perl-speaks-NONMEM (PsN). The appropriateness of the methods and the consistency of the software tools were evaluated using a large selection of clinical and nonclinical data. Criteria based on clinical relevance were found to be useful components in automated stepwise covariate model building. Their ability to restrict the number of included parameter-covariate relationships while maintaining the predictive performance of the model was demonstrated using the antiarrythmic drug dofetilide. Log-likelihood profiling was shown to be equivalent to the bootstrap for calculating confidence intervals for fixed-effects parameters if an appropriate estimation method is used. The condition number of the covariance matrix for the parameter estimates was shown to be a good indicator of how well resampling methods behave when applied to pharmacometric data analyses using NONMEM. The software developed in this thesis equips modellers with an enhanced set of tools for efficient pharmacometric data analysis.

Keywords: pharmacometrics, pharmacokinetics, pharmacodynamics, methodology, statistics, model evaluation, resampling methods

Lars Lindbom, Department of Pharmaceutical Biosciences, Division of Pharmacokinetics and Drug Therapy, Box 591, Uppsala University, SE-75124 Uppsala, Sweden

Papers discussed

This thesis is based on the following papers, which will be referred to by their Roman numerals in the text.

I   *Lindbom L., Ribbing J. and Jonsson E.N.* Perl-speaks-NONMEM (PsN)--a Perl module for NONMEM related programming. *Computer Methods and Programs in Biomedicine.* 75(2); 75-84 (2004).

II   *Lindbom L., Pihlgren P. and Jonsson E.N.* PsN-Toolkit--a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM.
         Computer Methods and Programs in Biomedicine. 79(3); 241-257 (2005).

III   *Tunblad K., Lindbom L., McFadyen L., Jonsson E.N., Marshall S. and Karlsson M.O.* The use of clinical irrelevance criteria in covariate model building with application to dofetilide pharmacokinetic data.
         In manuscript.

IV   *Lindbom L., Wilkins J.J., Frey N., Karlsson M.O. and Jonsson E.N.* Evaluating the evaluations: resampling methods for determining model appropriateness in pharmacometric data analysis.
         In manuscript.

Reprints were made with the permission of the publishers.

Contents

Abbreviations

API	Application programming interface
BCa	Bias and skewness-corrected bootstrap percentile limits
CDD	Case-deletion diagnostics
CL	Clearance
df	degree of freedom
FO	First order
FOCE	First order conditional estimation
FOCE INTER	First order conditional estimation with interaction
IIV	Interindividual variability
IOV	Interoccasion variability
LLP	Log-likelihood profiling
MAE	Mean absolute error
OFV	Objective function value
PD	Pharmacodynamic
PK	Pharmacokinetic
PK/PD	Pharmacokinetic/Pharmacodynamic
PsN	Perl-speaks-NONMEM
RSE	Relative standard error
SCM	Stepwise covariate model
SMP	Shared memory multiprocessor
V	Volume of distribution
ε	The difference between prediction and observation, residual error
η	The difference between an individual parameter and the population parameter
θ	Fixed-effect parameters
$\varkappa$	The difference between occasions for an individual parameter
ϕ	General model parameters, including θ, η and $\varkappa$
σ^2	Variance of the residual error
Π	Covariance matrix of inter-occasion variability
Ω	Covariance matrix of inter-individual variability

Introduction

The purpose of clinical trials is to verify the safe and efficacious use of drugs. Usually very 'simple' statistical analyses are used to evaluate the results from these trials. In pivotal clinical trials, the typical purpose of the statistical analysis is to test if there is a difference in efficacy between drug A and placebo. The type of tests used, e.g. Student's t-test, Pearson's chi-square test or Fisher's exact test, make very few assumptions, which makes it relatively easy to prove that the tests are appropriate. Inferences based on complex designs, such as pharmacometric models, make more assumptions. However, one advantage with model-based inferences is that it may be possible to make predictions for a wider spectrum of situations than those covered by the main objective of a trial.

Pharmacometrics uses models based on pharmacology, physiology and disease for quantitative analysis of interactions between drugs and patients. This involves pharmacokinetics, pharmacodynamics and disease progression with a focus on populations and variability. The purpose of pharmacokinetics is to describe the relationship between dose, formulation and exposure of a drug using the time-course of absorption and disposition processes. Pharmacodynamics describes the drug effect as a function of dose and concentration of drug in plasma or tissues.

Pharmacometric data analysis and modelling are tightly linked concepts in drug development. Modelling can be used simply to summarise pharmacokinetic (PK) and pharmacodynamic (PD) data. It may also be used to make predictions in special populations, to create and test hypotheses, and to support dosage recommendations. Further, models can be used as a means to prove that the mechanisms of pharmacokinetics and pharmacodynamics are adequately understood and addressed. All these aspects of modelling are important in drug development (1, 2).

Population pharmacokinetics is a term closely related to pharmacometrics and was coined in the 1980s. The initial purpose with population pharmacokinetics, was to analyse routine clinical data for individualising dosage regimens (3-7). In drug development the usage of this approach has become much wider, ranging from early human studies in phase I to late clinical trials of phase III and phase IV (8, 9), in addition to the original application area. Population pharmacokinetics and pharmacodynamics describe the mean trends and variability of drug concentrations and/or drug effects in a set of subjects by means of nonlinear mixed-effects models (10, 11).

Lars Lindbom

Decisions in drug development should preferably be based on the best possible information. This information is also important for evaluating efficacy and safety of drug treatment. However, because of the complexities involved, especially in model based analysis, the best possible (or perhaps most correct) information is not always easily obtained. Statistical methods to improve the information content of model based analyses are central to this thesis. Further, these methods are not meaningful to researchers until they are easily applicable to typical problems and there exists a set of examples showing the appropriateness of the methods. This is particularly relevant to pharmacometrics, in which the nonlinearity of most relevant mathematical models makes it difficult to show general validity. Practically, this means that a method must be translated into computer software that can communicate with the programs that researchers already use, handle most common types of applications and be evaluated on a range of real world problems. This thesis focuses on the development, application and evaluation of a set of statistical tools for nonlinear mixed-effects modelling.

Mixed-effects modelling

A mixed-effects model has two (or more) layers of variability. In pharmacometrics, the first layer handles residual variability while the second describes the variability between individuals. Conceptionally, pharmacometric models are often divided into three sub-models: a fixed-effects (structural) model, a random-effects (statistical or stochastic) model and a covariate model.

The structural model

The structural part of a mixed-effects model in pharmacometric data analysis describes the pharmacokinetic and/or pharmacodynamic properties of a drug, which are shared among all individuals in a population. This could for example be a one-compartment model (PK) and an E-max model (PK/PD). The structural model includes fixed effects such as rate of absorption, clearance, maximal effect etc. that belongs to an average individual in the population. Adhering to the well established nomenclature of population pharmacokinetics and pharmacodynamics the structural part (plus residual error) of a mixed-effects model is written

$$y_{ij} = f\left(x_{ij}, \theta, \varepsilon_{ij}\right)$$
$$\varepsilon_{ij} = N\left(0, \sigma^2\right)$$

(Eq. 1)

where y_{ij} is the jth observation in the ith individual, x_{ij} is the jth input for the ith individual, f is the transfer function from input to observations, θ are the model parameters, ε_{ij} is the residual error for the jth observation in the ith individual and σ^2 is the variance of the unexplained residual variability. The residual error is introduced to handle unexplained variability. Errors in dose or sampling time, sample volume, analytical instruments and model misspecification are all included in this term. To describe the pharmacokinetics of a drug, a simple form of f is a one-compartment pharmacokinetic model with a single intravenous administration, a single dose level and an additive error term. This is written as

$$y_{ij} = \frac{Dose}{V} e^{-CL/V \cdot t} + \varepsilon_{ij}$$

(Eq. 2)

The parameter vector θ describing the pharmacokinetic properties of this drug consists of clearance CL and volume of distribution V. Here, these have the same value for all individuals but mixed-effects models allow the model parameters to change between individuals.

11

The stochastic model

People are not identical and consequently the concentrations of a drug and its effect may differ between individuals. Techniques and equipment for chemical analysis as well as routines for analysis of drug concentration in plasma samples also vary over time and within and between laboratories. The stochastic part of a mixed-effects model is used to account for this type of variability.

Interindividual variability

Drug response is in general subjected to substantial interindividual variability (IIV) (12). This variability may partially be described by demographic factors such as sex, weight, age and smoking habits, clinical factors such as liver and kidney disorders or genetic factors that control the sensitivity to a drug effect. The unexplained variability between individuals may still be large even if covariate effects are accounted for and it is therefore important to use an appropriate model structure for the interindividual variability. Let ϕ_i denote the parameters of the ith individual, including covariates z_i, fixed effects parameters θ, and random effects η_i describing the variability of the parameters between individuals.

$$y_{ij} = f\left(x_{ij}, \phi_i, \varepsilon_{ij}\right)$$

$$\phi_i = g\left(z_i, \theta, \eta_i\right)$$

$$\eta_i = N\left(0, \Omega\right)$$

(Eq. 3)

The function g expresses the effect of the covariates and interindividual random variation on the individual parameters ϕ_i. The η_i are random parameters with mean zero and variance-covariance matrix Ω. The one-compartment model above can thus be expanded using formulas describing the variability of clearance and volume of distribution within the studied population in relation to the typical values of the population, *TVCL* and *TVV*. Often, lognormal distributions are used to explain the variability of parameters that have a natural lower boundary (such as zero). This is exemplified for clearance.

$$CL_i = TVCL \cdot e^{\eta_{CL,i}}$$

$$V_i = TVV \cdot e^{\eta_{V,i}}$$

(Eq. 4)

Interoccasion variability

The pharmacokinetics of a drug may change over time within an individual. Concomitant medication of inducers/inhibitors of metabolic enzymes, progression of liver/kidney disease and the development of organ function in neonates are some possible causes for time-varying pharmacokinetics. The underlying processes that control the variation are often poorly understood and difficult to measure. Consequently, a large magnitude of this variability is usually not possible to explain. If the pharmacokinetics changes relatively fast, a time-dependence on the

magnitude of the variability may be needed in the model. If the changes occur over a longer period of time, for example between different treatment periods or occasions, the variability between these periods may have to be described by an interoccasion variability (IOV) term. Neglecting IOV may lead to biased estimates of the pharmacokinetic parameters (13, 14). If there is a need to include a term for the unexplained interoccasion variability the model parameters are determined per individual i and occasion k, ϕ_{ik}.

$$\phi_{ik} = h\left(z_i, \theta, \eta_i, \varkappa_{ik}\right)$$
$$\eta_i = N\left(0, \Omega\right)$$
$$\varkappa_{ik} = N\left(0, \Pi\right)$$

(Eq. 5)

Using this parameterisation, z_i, θ, η_i, and Ω are defined as before while h is the function relating the covariates, IIV and IOV to the individual-occasion specific parameters ϕ_{ik}. $\varkappa_{ik}$ are the random effects of the parameters for individual i and occasion k and Π is the variance-covariance matrix of these random effects. If clearance in the one-compartment model above is subject to IOV the sub-model for clearance would be:

$$CL_{ik} = TVCL \cdot e^{\left(\eta_{CL,i} + \varkappa_{CL,i,k}\right)}$$

(Eq. 6)

The covariate model

Parts of the variability can often be explained by covariates, which are factors whose individual values are measured and recorded in clinical trials. For example, creatinine clearance may help to explain the overall clearance of a drug that is eliminated by the kidneys.

Covariates can be incorporated into a model in various ways depending on their type, shape and range. A categorical covariate can be binary (e.g. sex, smoker (Yes/No), concomitant medication) or it can have multiple categories (e.g. race, scales describing disease states or disease progression). The categories may be non-ordered (race) or ordered (a 4-level scale describing heart failure).

Categories are often assigned numerical values to make it easier to formulate parameter-covariate relationships. For a binary covariate a baseline or a 'No' answer to a question of the type "Do you smoke?" is usually coded as 0, while 1 represents the other state. The effect of smoking habits (SMOK) on clearance may be included in the model as:

$$CL_i = TVCL \cdot \left(1 + \theta \cdot SMOK_i\right)$$

(Eq. 7)

where θ stands for the fractional change in clearance upon smoking. The New York Heart Association (NYHA) has a 4-grade classification scale for heart failure. In

this case, grade 1 is used as the baseline and a new sub-model is included for each new category.

$$CL_i = \begin{cases} TVCL \cdot (1+\theta_1) & if\,(NYHA_i = 2) \\ TVCL \cdot (1+\theta_2) & if\,(NYHA_i = 3) \\ TVCL \cdot (1+\theta_3) & if\,(NYHA_i = 4) \end{cases} \qquad \text{(Eq. 8)}$$

Continuous covariates are included in a similar fashion, and as before, the covariate is often added as a fractional change in a parameter. The effect of the covariate can be expressed relative to its median to make the parameter values more informative. If weight is a good predictor of volume of distribution, it may be added to the model as a linear sub-model according to:

$$V_i = TVV \cdot \left(1 + \theta \cdot \left(WT_i - median\left(WT_i\right)\right)\right) \qquad \text{(Eq. 9)}$$

The parameter-covariate relationship can be assigned any type of linear or nonlinear function. Two common parameterisations for nonlinear models are the piecewise linear model which combines two linear splines, and the exponential model as showed in the equations below.

$$V_i = \begin{cases} TVV \cdot \left(1 + \theta_1 \cdot \left(WT_i - median\left(WT_i\right)\right)\right) & if\,(WT_i < median\left(WT_i\right)) \\ TVV \cdot \left(1 + \theta_2 \cdot \left(WT_i - median\left(WT_i\right)\right)\right) & if\,(WT_i \geq median\left(WT_i\right)) \end{cases} \qquad \text{(Eq. 10)}$$

$$V_i = TVV \cdot \exp\left(\theta \cdot \left(WT_i - median\left(WT_i\right)\right)\right) \qquad \text{(Eq. 11)}$$

Model building

Pharmacometric models range from empirical models to full physiological models. Empirical models use few mechanistic assumptions and are primarily intended to describe data and not to explain it. One example of an empirical pharmacokinetic model is the sum of exponentials. Models that are more mechanistic may be obtained by incorporating knowledge about physiology such as blood flows, organ volumes, receptor affinities and enzyme turnover rates. Full physiological models include blood flows and volumes of all major organs.

Regardless of the level of mechanistic integration, there is often a need of exploring data to gain insight in the pharmacokinetic and pharmacodynamic properties of a drug. This type of modelling, where data drives the structure of a model, is called exploratory data analysis (EDA). Thus, EDA is the task of uncovering information in data. This is often employed in the process of building pharmacokinetic and pharmacodynamic models, and it is a recommended part of the strategy in model development (15). A combination of graphical and statistical tools can be used to find trends in the data, primarily to identify covariates that are good predictors of

model parameters (16, 17). Using EDA may however introduce selection bias since strong signals are explicitly selected (18).

Scatter plots displaying model predictions versus observations and weighted residuals versus predictions/observations or time are often used as means to assess the goodness of fit for a model. The distribution of the weighted residuals is a good indicator of the appropriateness of the residual sub-model. The plots of the residuals may also give the modeller hints of where to find the possible weak spots of the model.

The benefit of using a more complex model structure over a simpler one is relatively easy to deduce when maximum likelihood is used for regression. If the models are nested, the benefit can be evaluated using the ratio of the likelihoods for the two models. The likelihood ratio is χ^2-distributed for nested models and the degrees of freedom (df) for the distribution are determined by the difference in degrees of freedom between the two competing models. Often, as in the case of NONMEM, the likelihood is expressed as an objective function value (OFV), which is equal to minus two times the logarithm of the likelihood. The difference between the OFVs of two competing models is also χ^2-distributed. Consequently, the addition of an extra structural parameter to a model yields a model which better reflects the data if the difference in OFV is larger than 3.84 (χ^2*(1df, p=0.05)*). However, simulation studies have shown that the actual significance levels may not be the expected if inappropriate estimation methods and/or distributional assumptions of the variability are used (19, 20).

Data analysis using models is lined with assumptions about the properties of the data, the statistical methods (regression methods, evaluation methods etc.), the fixed-effects and the stochastic parts of the model (21). Assumptions of particular importance for this thesis are (i) the validity of the likelihood ratio test for different estimation methods and model components; (ii) the distributional properties of the parameter estimates; (iii) model independence of subgroups in the data; (iv) the adequacy of the covariate model building strategy.

Clinical relevance

Clinical relevance in modelling translates the addition of model components into potential benefits for patients. For example, a change in clearance of 50% for a subgroup of patients could be very important for a drug with a narrow therapeutic index while not being relevant for other drugs. The importance of incorporating a covariate that is related to clearance for this subgroup into the model can be calculated from such clinical information.

One measure of the impact of adding a predictor to a model parameter is a reduction in the interindividual variability of this parameter. By reducing the interindividual variability, the average error in the predicted individual parameter value will be reduced. Clinical relevance can also be based on changes in predicted individual parameter values. For example, if the predicted clearance in a patient

drops by 25% after the inclusion of a covariate, this covariate may be regarded as a clinically relevant predictor of clearance. This approach is sensitive to outliers since it only depends on the prediction of a single individual's parameter value. To make a covariate model a bit more robust, the 95th percentiles of the predicted changes of all individuals can be used.

What is regarded as clinically relevant varies between researchers, between physicians and between regulators. A more conservative approach in covariate model building may be to exclude covariates based on clinical irrelevance. The threshold for irrelevance can be set at a low value to be sure not to miss clinically relevant covariates.

Stepwise covariate model building

Stepwise covariate model building is also known as (orthogonal) Forward Selection – Backward Elimination. The method assumes that a structural model has already been defined. The task of stepwise covariate model building is to identify covariates that explain the variability in the parameters of the structural model.

In a first step, each relevant parameter-covariate combination is added and estimated one by one in the structural model. The model with the largest improvement over the starting model is retained as the starting model for the next step. In each subsequent step, the remaining parameter-covariate combinations are tried. This forward inclusion continues until no improvement is gained by adding new model components. The measure of model improvement is usually based on statistical significance. Optionally, the forward inclusion step can be followed by a backward elimination step. This proceeds according to the same general scheme as the forward step, but reversely, using stricter improvement criteria.

This adaptive procedure for covariate model building relies heavily on the validity of the statistics used for model discrimination. Stepwise procedures in general have been shown to exhibit a risk of including false parameter-covariate relationships, of giving rise to biased estimates of the included relationships as well as of yielding too narrow confidence limits (22, 23). Other studies have reported that these problems may not be large for pharmacokinetic models (24, 25). Benefits of stepwise covariate model building are that the procedure is conceptionally simple and easy to understand and that the *a priori* declared criteria make the procedure relatively objective.

Regression

There are a number of different software packages for regression of mixed-effects models. These can be divided into approximate maximum likelihood, exact maximum likelihood, nonparametric maximum likelihood and full Bayesian. The approximate likelihood methods include NONMEM (26), NLINMIX (SAS) (27), NLME (S-PLUS) (28) and WinNonMix (29). The newer exact maximum likelihood methods are: NLINMIX adaptive Gaussian quadrature (SAS), PEM (30) and MCPEM

(31). The nonparametric maximum likelihood methods include NPML (32) and NPEM (33) and SAEM (34). An example of a full Bayesian method is BUGS (35). NONMEM is by far the most used software (36, 37). Some studies have been performed, comparing the different software (38-40) and the recent development of methods using stochastic estimation of the likelihood seems very promising. The studies in this thesis use NONMEM as the regression software.

NONMEM

Papers I & II describe the development of software that uses NONMEM as the engine for regression. Therefore, an introduction to the regression methods used in NONMEM as well as some practicalities of interaction with the software is given below.

Estimation methods

NONMEM uses the maximum likelihood regression procedure. By iteratively adjusting the model parameters it seeks the combination of parameter values that maximises the likelihood $L(y|\phi,x)$ of the observations, given the model and input variables. The likelihood is the product of the likelihood L_i of all individuals in the data. Since the likelihood must acknowledge the random effects on the individual level, the individual likelihood is expressed as an integral over all possible values of η_i according to

$$L = \prod_i L_i\left(y_i \mid \phi_i, x_i\right) = \prod_i \int l_i\left(y_i \mid \theta_i, x_i, \eta_i\right) h\left(\eta_i \mid \Omega\right) d\eta \qquad \text{(Eq. 12)}$$

The parameters ϕ are divided into fixed effects, θ and random effects, η. Most often, no closed form solution of the integral exists for nonlinear mixed-effects models, which means that some level of approximation must be used.

NONMEM uses two versions of a first order approximation by Taylor expansion of y_i. The first, called the first-order estimation method (FO), evaluates the terms of the Taylor series around the expected value of the ηs and εs, which is zero. The second, called the first-order conditional estimation method (FOCE), acknowledges that the individual likelihoods are better approximated if the Taylor series is evaluated around the *conditional estimates* of the ηs, i.e. the posterior Bayes estimates of the parameters in each iteration of the regression procedure. A further refinement of the first order approximation is called FOCE INTER. This method maintains the effect of residual error on the ηs (*interaction* between ηs and εs) when the likelihood is calculated. Interaction will be an issue when the residual error model is allowed to vary between individuals or when a heteroskedastic error model is used. A second order approximation is accessed through the Laplacian method (LAPLACE). This method also computes the likelihood using the conditional estimates of the ηs. Technically, NONMEM minimises the *objective function,* which equals minus two times the natural logarithm of the likelihood, which is reported when the regression converges or terminates.

17

Estimate of standard error

The standard method for calculating standard errors in NONMEM is the robust sandwich method. Using this method the variance-covariance matrix of the parameter estimates is calculated as

$$\mathrm{var}(\hat{\phi}) = R^{-1}SR^{-1} \qquad \text{(Eq. 13)}$$

where R is the hessian (partial second derivatives) of the objective function evaluated at the parameter estimates, and S is a sum of the cross-products of the gradient vectors of the individual objective functions.

Both the S and R matrices should asymptotically be good estimates of the variance-covariance matrix. When normality cannot be assumed, the robust sandwich method may be a better alternative. Wald confidence intervals of the parameters may be calculated as

$$\hat{\phi} \pm SE_{\hat{\phi}} \cdot z_{\alpha} \qquad \text{(Eq. 14)}$$

where $z\alpha$ is the α quantile of the normal distribution (41).

The NONMEM model file

This file contains instructions in ordinary ASCII text that, except for the data, holds all information needed for fitting a non-linear mixed effects model using NONMEM. Typically, a model file contains specifications for a pharmacokinetic and/or a pharmacodynamic model, initial estimates of the model parameters, boundaries for the model parameters as well as details about the location and the format of the data.

The NONMEM data file

In the field of pharmacokinetics and pharmacodynamics parameters of mixed effects models are usually estimated based on data from clinical trials. This information is often held in a text file that typically contains records of subjects included in the study, subject demographics, dosing and sampling times and measurements of drug concentrations and drug effects.

The NONMEM output file

The results of model fits or simulations performed in NONMEM are presented in an output file of a fixed format. Among other things, the output file includes parameter estimates, an estimate of the variance-covariance matrix plus diagnostics from the minimisation of the objective function and the calculation of the variance-covariance matrix.

Model Evaluation

The term "Model validation" is often used when there is a need to demonstrate that a model is suitable for its intended purposes. However this terminology suggests that there do indeed exist valid and, consequently, invalid models. A model is by definition an approximation of a phenomenon that we try to describe. In drug development, we will hardly ever be able to control and measure every aspect of an experimental procedure, and thus there will be parts of a model (especially the stochastic model) where several alternative sub-models provide more or less equivalent results. A better term may therefore be "Model evaluation". Although there may exist a need for criteria to define "good" and "bad" models this will have to be criteria that varies by purpose and application.

A model that fails to meet some or most criteria used in an evaluation may still be useful. It can, for example, serve as a basis for further investigations where the weak parts of the model are studied, be used for predictive purposes acknowledging its flaws, or purely as a way to summarise an idea or a concept (42).

Graphical evaluation

Graphical inspection of the goodness of fit is useful at many stages in the model building and the model evaluation processes. The model building is an interactive method, where each expansion and reduction of the model is followed by an in-progress evaluation. Goodness of fit graphs are easily created using for example Xpose (43), making it a suitable tool for summarising the model status as the modelling process continues.

Consistency

Accuracy and precision

Regardless of the intended use of a model, i.e. to make predictions, to select appropriate dose levels etc., the quality of the model parameter estimates should be considered. If the correct model structure is known, the only sources of inaccuracy and imprecision are the data and the regression procedure. The underlying assumptions of the Wald based confidence intervals, which are usually reported for a population model developed using NONMEM, are that the parameter estimates are symmetrical and that the estimates of standard error are correct. The confidence intervals may be evaluated and further refined by applying bootstrapping (44-47) or log-likelihood profiling (48, 49).

Over-parameterisation

A model should not be more complex than what is supported by the data. This is not only a question about whether parameter estimates are at all possible to get for a certain combination of model and data. Usually, one intention with a model is to summarise the information contained in data. If the model is to represent any

additional value to the data at hand, it should be possible to make some statements about data that does not yet exist. For example, if a data set consisting of 100 observations clearly shows a linear trend of the dependent variable versus an input variable, a second-order polynomial would probably succeed in describing the data since the data is dense. The system will however be ill-conditioned since the trend also can be described using only the linear part of the polynomial. In this case, the linear model is probably preferred for predictive purposes, unless additional evidence supports the nonlinear model.

Correlations between parameter estimates and the condition number of the covariance matrix are two types of diagnostics that can be used to assess the level of over-parameterisation. The bootstrap may also be used to identify over-parameterisation and to get some indication of which parameters that are affected.

Subpopulations

Failure to identify subpopulations can lead to incorrect dosing recommendations of a drug for a certain subpopulation. Therefore, it is important to resolve if a model benefits from the inclusion of a covariate as a predictor of a model parameter. Data not included in the original data set can be used to assess the predictive performance of the original model compared to a model where additional covariates have been added. If the larger model does not predict the external data better than the original model, we can conclude that the original model includes all covariates that are important for prediction. An alternative approach is to use the bootstrap to verify the inclusion or exclusion of covariates to a model (46). Depending on the complexity of the model, this type of evaluation can be very time-consuming.

Identifying external factors that influence the size of the model parameters is also of importance. Continuous external factors that influence the model can be handled by the stochastic parts of the model. The influence of categorical external factors may be evaluated using mixture models (26). Standard statistical tests for model selection can be used to clarify the need for a mixture model.

Data dependence

To be a good description of a system, a model should not be overly dependent on small subsets of data. If the model changes a lot depending on which part of the data that is used for model building it may be a sign of both under- and over-parameterisation. The effect of excluding/including one or several individuals in a pharmacometric data analysis can be investigated through the calculation of the individual contributions to the likelihood (50) or through case-deletion diagnostics (47, 51). The overall dependence of parameter estimates on the entire data set can be evaluated using the bootstrap (46). Histograms of the bootstrap parameter estimates can be used to explore potential multi-modality. The composition of the bootstrap samples within the different modes in the histograms may give an indication of which individuals or groups of individuals that drive the estimate in a particular direction.

Predictivity

If a model is intended for predictive purposes, evaluation of the predictive performance is important. The mean absolute error (MAE) and root mean squared error (RMSE) are two statistical measures, which may be used to summarise bias and the precision of the prediction. Predictivity can be evaluated on the observational level by evaluating the ability of a model to predict concentrations, drug effects or biomarkers. Although this is easy to calculate, the results may be hard to interpret, especially if the predictions are made over a large range of values. It may be wise to calculate the MAE and RMSE on transformed (logarithmic, box-cox, etc.) concentrations to handle heteroskedasticity. Predictivity may also be assessed on the parameter level, for example to get a measurement of how well a model predicts individual clearance values (52). This is most important in the evaluation of the covariate sub-model. A potential problem is the necessary assessment of the individuals' *true* parameter values. Posterior Bayes estimates may be used as surrogates if no exterior validation data exists for this purpose. If the data is relatively dense, this may indeed be a good option but if the data is sparse, the posterior Bayes estimates will be shrunk towards the population mean of the parameters. The MAE and RMSE may in this case be a too optimistic measure of the predictivity.

It should be noted that a measure of the predictivity may not add too much to the perceived quality of the model unless it is evaluated in relation either to the predictivity of another model or to criteria stating what is regarded as sufficient, clinically or otherwise.

Statistical procedures

A number of statistical procedures have been suggested for model evaluation in pharmacometric data analysis (15, 48, 53-55) This section gives a general overview of the algorithms, results and diagnostics for these methods.

Bootstrap

The bootstrap (45, 56) is a general method for measuring statistical accuracy and precision. Briefly, it involves creating "new" data sets by sampling with replacement from the original data and applying the same analysis steps to each of the new data sets as was performed on the original data. In population PK/PD, these analysis steps usually correspond to model fits generating parameter estimates, but it can be any kind of statistical procedure. The results from the new data sets form distributions, which reflect the uncertainty in the original analysis. These distributions can be used to assess covariate selection stability (43), uncertainty of parameter estimates (46) and to correct for certain types of bias (57). The resampling is performed with replacement on statistically independent parts of the data, which in population PK/PD usually corresponds to individuals. Depending on the statistic of interest, different numbers of resampled data sets are needed. Estimation of standard errors typically requires around 200 bootstrap data sets. If the Bias and skewness-corrected bootstrap percentile limits method (BCa) (44) is used to calculate second-

order correct 95% confidence intervals approximately 2000 bootstrap data sets are required.

Log-likelihood profiling

The standard way of estimating the confidence intervals of parameter estimates for nonlinear mixed-effects models is to calculate the interval limits from the standard errors under the assumption that the estimates are normally distributed. Log-likelihood profiling (LLP) is one alternative method where no assumption regarding symmetry of the interval is made (48, 49). Fixing a parameter to values close to the estimate obtained from a maximum likelihood procedure (as the one implemented in NONMEM) and refitting this reduced model generates a likelihood profile. With NONMEM, minus two times the natural logarithm of the likelihood is used and the maximum likelihood then corresponds to the minimum of this quantity. If a parameter is fixed, this model can be regarded as an alternative, competing with the full non-fixed model, in being the most appropriate model for describing the data at hand. A statistically significantly improvement is achieved when the log-likelihood difference is 3.84 for two nested models ($\chi^2(1df, p=0.05)$). Thus, the challenging alternative model with a fixed parameter can be rejected when the log-likelihood difference is above this value. The confidence interval limits for a parameter is then where the log-likelihood is 3.84 higher than at the maximum likelihood estimate. It is however well known that the actual significance levels for log-likelihood differences acquired through NONMEM may not agree with the expected using a χ^2-distribution. For example, it has been shown that the FO gives higher significance levels than the nominal when studying covariate effects (19). Also, tests for non-zero variance components cannot be performed using standard likelihood ratios since the null-hypothesis puts the parameter value on the boundary of the parameter space (zero) (58).

Case-deletion diagnostics

Case-deletion diagnostics (CDD) is a standard method for detecting observations that are the most important for a model fit, e.g. to determine which observation or individual that influences the parameter estimates the most. The data is divided into *k* parts of which *k-1* are used to re-fit the model. This is repeated until all *k* parts have been excluded once. For mixed-effects modelling, this most often means that one individual or group of individuals is excluded at a time. The Cook-score and the covariance ratio may be used to estimate the effect of the removal of one individual on the parameter estimates (59) according to

$$CS_k = \sqrt{\left(\hat{\phi}_k - \hat{\phi}\right)^T \text{cov}\left(\hat{\phi}\right)^{-1}\left(\hat{\phi}_k - \hat{\phi}\right)} \qquad \text{(Eq. 15)}$$

$$CR_k = \sqrt{\frac{\det\left(\text{cov}\left(\hat{\phi}_k\right)\right)}{\det\left(\text{cov}\left(\hat{\phi}\right)\right)}} \qquad \text{(Eq. 16)}$$

$\hat{\phi}_k$ and $\mathrm{cov}(\hat{\phi}_k)$ denote the vector of parameter estimates and the covariance matrix acquired when individual k is removed from the dataset. Analytical methods for case-deletion exist for linear models but for nonlinear mixed-effects models it is necessary to fit the model to data sets in which one individual at a time is excluded (51). This requires a substantial amount of CPU time, especially for large data sets.

Jackknife

The jackknife, first proposed by Quenouille (60, 61) as a means to reduce bias, is very similar to case-deletion diagnostics. Parts of the data are excluded one by one from the training data set, followed by the calculation of a statistic of some kind on each reduced data set. The jackknife estimate of the bias of a parameter estimate for a data set with n subjects is the difference between the jackknife estimate of the mean and the estimate from the full data set scaled by a factor of n-1, i.e.

$$\widehat{bias} = \ n-1)\left(\hat{\theta}_{(.)} - \hat{\theta}\right) \tag{Eq. 17}$$

where

$$\hat{\theta}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(i)} \tag{Eq. 18}$$

Cross-validation

Cross-validation is also very similar to case-deletion diagnostics. It uses the same scheme of excluding a portion *1/k* of the data and refitting the model on the remainder. This is repeated until all *k* parts of the data have been excluded once. However, instead of excluding one individual at a time a larger portion of the data is withhold, often $1/5^{th}$ of the data. The values of the parameters, as they were estimated for each *k-1* sized remainder, are then used to predict the excluded parts of the data. Some statistics of the predictive performance is then calculated and summarised for the *k* predicted sub sets. Possible statistics to calculate using this procedure are cross-validated versions of OFV, MAE and RMSE.

Software development

The academia, the pharmaceutical industry and the regulatory authorities all drive the need for new scientific software for all aspects of modelling and simulation. Guidelines or definitions of 'best practice', formulated by the authorities or others, are based on a fairly pragmatic balance of what is regarded as 'best' or 'appropriate' and what is actually doable. The availability of resources in terms of time, employees and validated tools determines what is regarded as doable. Consequently, there is a need for development of validated software that helps scientists to apply the best available methods.

Pharmacometric software

The dominant position of the NONMEM package for nonlinear mixed-effects modelling affects the format of new software. If the software uses NONMEM models, data or output it should preferably adhere to the standards of NONMEM (26) to be useful to more than a few research groups. NONMEM is distributed as ANSI FORTRAN code and runs on any hardware and software platform that has a FORTRAN compiler for the ANSI standard. A NONMEM run may require several hours or more to finish, depending on the complexity of the model and the amount of data. As discussed earlier, NONMEM uses text files for input and output.

Requirements

Since NONMEM is largely platform independent, it is wise to choose a programming language that is not restricted to one operating system or hardware. The text-processing components of the language are also important. Object orientation is an attractive way of encapsulating blocks of code sharing a common purpose or functionality. In object orientation, the abstraction of code into classes, attributes and methods makes it easy to translate real world problems to software solutions. The runtimes of NONMEM can pose a problem, especially in algorithms that involve many runs. Such software, for example resampling methods, would benefit from executing NONMEM runs in parallel.

Parallel environment

The very frequently cited Moore's Law (62) predicts that the number of transistors on a chip (more or less a good measure of computer speed) doubles about every two years. Lately, there has been a lot of debate about whether this trend is coming to and end or not. Indeed, the speed of the processors that have come to the market after 2001 has not increased at the rate predicted by Moore. Instead, a lot of attention has been given to parallel computing. The idea is to make several processors cooperate and thereby cutting the computing time.

Symmetric Multiprocessing

Symmetric Multiprocessing (SMP) is a computer architecture where two or more processors share the same main memory. It is the most common multiprocessor system type and it is available for many high-end workstations and servers.

Clusters

Clusters are groups of relatively tightly joined computers. The Linux kernel extension openMosix (63) is one example where the total load of the computers in the cluster is balanced by migrating processes to computers where there are free CPU-cycles and memory. This type of system behaves a lot like a standard SMP with the difference that processors on separate computers do not have (direct) access to the same main memory. This puts some, but not many, extra constraint compared to SMPs on the way a parallel computing may be realised. A little simplified, the most apparent adjustment that is necessary to make to get a program to work in parallel on an

openMosix cluster compared to an SMP computer is that it must create processes that have a memory area dedicated to it.

Grids

Grids are collections of loosely joined computers. NorduGrid's Advanced Resource Connector (ARC) (64, 65) and Platform's Load Sharing Facility (LSF) (Platform Computing Inc. Markham, Ontario Canada) are two examples of grid solutions. In a grid, processes are submitted to a queuing system and balancing of the load is done at the launch of each new process. A process that is committed to a particular CPU is not moved from that CPU until the process is finished. Compared to clusters, this solution is less dependent on continuous and steady network connections and it is suitable for large systems with many computing nodes. The simplest way to benefit from a grid environment is to identify pieces of code that can take a long time to run, and which can be executed in isolation.

Validation

A software validation procedure verifies that the software does what the design and performance goals of the development state. Scientific methods are evaluated by careful analysis of their results, and by comparing them to other methods. This will ultimately lead to recommendations about the circumstances under which a method can be used as well as the benefits one method has over another.

To show general applicability, software that uses NONMEM for regression should preferably operate correctly on a large set of models. The set would if possible include multiple estimation methods, different types of data, PK and PD models. Ideally, different modellers should have developed the models since NONMEM coding styles vary.

This thesis in perspective of current research

NONMEM is the first software developed for nonlinear mixed-effects modelling in pharmacometric data analysis, and the first version was distributed in 1979. Since then a number of other software have been developed. Still, NONMEM has been used in the majority of the publications within pharmacometrics during the last decade (425/503) (36). The development of an application programming interface to NONMEM such as Perl-speaks-NONMEM (PsN) is therefore valuable since it enables the development of new tools using NONMEM for regression.

In early 2000, when the work of this thesis was initiated, four papers on the use of bootstrapping in population pharmacokinetics had been published[1]. From 2000 until today, 30 papers have been published. The increased interest for this method may be explained by the Guidance for Industry on population pharmacokinetics, which was released by the U.S. Food and Drug Administration in 1999 (53). This Guidance recommends bootstrapping for model evaluation purposes. Bootstrapping and other techniques that require repeated modifications of the data or the model are cumbersome to carry out using NONMEM as the regression tool. The expansion of PsN to include a set of statistical tools (case-deletion diagnostics, cross-validation, jackknife, log-likelihood profiling, the bootstrap and stepwise covariate model building) makes these tasks a lot easier to perform.

Log-likelihood profiling and case-deletion diagnostics have also been suggested as tools for model evaluation (15, 48, 54, 55). However, there are no general recommendations on when these methods are appropriate to use. Recent examples of model evaluation in population PK/PD vary considerably in focus and methods (Table 1). To clarify the usefulness of different model evaluation methods we used a large set of clinical and nonclinical data. An interesting extension to this investigation would be to make the picture more complete by including the predictive performance of models.

As previously discussed, covariate model building using automated procedures is not without problems. Yet automated procedures are appealing because of their objectivity. As long as the modeller is aware of their deficiencies, the results may be compared between applications. In a stepwise search for a covariate model using statistical criteria for inclusion and exclusion of covariates many small but significant parameter-covariate relationships are often identified. Apart from being a potential

---

1   PUBMED, keywords: bootstrap AND ("population pharmacokinetics" OR NONMEM)

waste of CPU-time, a large number of included covariates may not increase the predictivity of the model. Therefore the gain in external predictivity should ideally be evaluated to determine the appropriate size of the covariate model. Application of clinical relevance criteria may help in limiting the number of included covariates. The inclusion of clinical criteria *a priori* in the automated covariate model building procedure as in Paper III of this thesis has to the authors' knowledge not been described before.

Table 1. Population PK/PD model evaluation techniques reported in the literature. The intention of the examination of the available literature was not to make a complete picture, but rather to give an overview of applied methods and their relative popularity.

Method	Application references	Number of articles
Bootstrap	(47, 55, 66-69)	6
Cross-validation	(70)	1
Data-splitting	(71-76)	6
External validation data set	(77-80)	4
Jackknife (leverage analysis)	(47, 55)	2
Log-likelihood profiling (LLP)	(55)	1
Posterior predictive check (PPC)	(39, 81)	2
Randomization test	(82)	1

The search was performed 2005-09-08 using PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed), with search terms ("model validation"[All Fields] OR "model evaluation"[All Fields]) AND (("pharmacokinetics"[Subheading] OR "pharmacokinetics"[MeSH Terms] OR pharmacokinetics[Text Word]) OR ("pharmacology"[Subheading] OR ("pharmacology"[TIAB] NOT Medline[SB]) OR "pharmacology"[MeSH Terms] OR pharmacodynamics[Text Word])). Articles not related to population PK/PD were not considered.

Aim

The aim of the work presented in this thesis was to facilitate the practical use of available and new statistical methods in the area of pharmacometric data analysis. This involved the development of suitable software tools that allows for efficient use of these methods, characterization of basic properties and demonstration of their usefulness when applied to real world data.

Present investigation

The development of a package of statistical tools for pharmacometric data analysis using NONMEM (Papers I & II)

The joint aim of Papers I and II was to develop a set of statistical tools for pharmacometric data analysis using NONMEM. In Paper I the effort was focused on preparing the ground for software development using NONMEM as a tool for regression. The result was an application programming interface (API) for NONMEM written in Perl. Using this API, we decided to create a common structure for the development of statistical tools. The main goal of this common structure was to make it easy to add new tools and to design it to allow for communication between the tools, i.e. to enable the use of one tool within another. Paper II describes the implementation of this common structure and a set of statistical tools; the bootstrap, case-deletion diagnostics (including jackknife and cross-validation), log-likelihood profiling and automated stepwise covariate model building.

Perl was chosen as the programming language based on (i) speed of text processing; (ii) platform independence; (iii) easy communication with operating systems and with other programming languages and (iv) object oriented programming support.

Object oriented design was used for both the API and the set of statistical tools. The abstraction of object orientation makes it easier to design and describe an application since real world terms are easily translated to code elements and vice versa.

The NONMEM API

The three input/output files of NONMEM, i.e. the model, the data and the output files, were chosen as the main classes for the object oriented design of the API to NONMEM.

Models

The model class of PsN is built around the NONMEM model file. It supports all records and options that are valid as model file specifications for NONMEM versions 5 and 6 with the exception of super-problems (26).

Data

The data class is designed to make it easier to manage NONMEM data files within computer programs. The structure of the data class is subject-centric, recognising that the subjects included in a study often can be regarded as independent. This structure makes it easy to restructure a data set on the subject level, e.g. to prepare for statistical methods based on resampling.

Output

The output class contains routines for parsing NONMEM output files. There is no need to change any aspects of an output file and consequently, the methods of the output class are all accessors reporting the content of a NONMEM output file in a structured manner.

Statistical tools

Each tool is implemented as a separate class, inheriting common routines and attributes from a general tool class (*Figure 1*). The general tool class is responsible for the functionality common to all statistical methods of PsN-Toolkit, e.g. parallel execution. A common structure for workflow is defined within the general tool class as is the ability to employ another tool.

Model fit

To create a general design of the tools of PsN, the execution of normal NONMEM runs was defined as the smallest element in the tool structure. The diagnostics of a NONMEM run is collected through the PsN output class. Some diagnostics are used to categorise a run as successful or non-successful using two levels of criteria within the PsN model fit. The standard criteria are i) a successful minimisation of the objective function and ii) no parameter estimates are close to a parameter boundary. A stricter variant called *'picky'* also requires that the covariance step of NONMEM is successful and that no warnings are reported from this step.

If a NONMEM model fit does not reach a successful minimisation of the objective function, it is often a good idea to try to rerun the fit with slightly perturbed initial estimates of the model parameters. PsN has the capability to rerun a NONMEM model fit automatically using a new set of initial estimates. What is considered a successful run is controlled by the *'picky'* option defined above. The new initial estimates θ_{init} for run i (i=0..N, where 0 is the original run and N is the maximum number of retries) are created as

$$\theta_{init_i} = \theta_{init_0} + rand_{uniform}\left(\pm 0.1 \cdot i \cdot \theta_{init_0}\right)$$

considering parameter boundaries, if any. Thus, the initial estimates for retry number three will differ at most by 30% from the original initial estimates. Similar to the other tools, the PsN model fit can execute parts of its routines in parallel.

Bootstrap

The bootstrap tool can perform a 'normal' nonparametric bootstrap, where confidence intervals are calculated based on bootstrap percentiles. It can also perform a BCa bootstrap where jackknife means of the parameter estimates are used to correct for bias and skewness. PsN runs a jackknife after the bootstrap when the BCa method is requested. Apart from confidence intervals, estimates of standard error and bias are calculated.

Case-deletion diagnostics / jackknife / cross-validation

The case-deletion diagnostics, jackknife and cross-validation are implemented as one class since they share many internal routines. Cook-scores and covariance ratios are calculated for each case-deleted data set. Jackknife estimates of bias are calculated using the parameter estimates from all case-deleted data sets. A cross-validated estimate of the OFV can be calculated from the predictions of the removed parts of the data.

Log-likelihood profiling

Log-likelihood profiling is implemented using a second-order polynomial function to predict the required increase in the OFV. The user can supply initial estimates of the confidence intervals.

Automated covariate model building

Automated stepwise covariate model building is implemented according to the definition stated earlier in this thesis. The user can control the behaviour of this tool to a large extent, for example by specifying the shape of each parameter-covariate relationship, setting initial estimates and defining limits for statistical criteria. It is also possible to supply project-specific code for the implementation of significance criteria.

Interaction with computer systems

Parallel execution

PsN is designed to run on multi-processor systems with a capability to perform many CPU-intensive tasks in parallel. However, NONMEM is not available in a parallelised form yet. Since most of the CPU-cycles within a PsN-program will be spent running NONMEM, the part of PsN that is responsible for executing NONMEM has been most optimised for a parallel environment. Parallel execution of PsN is supported on SMP systems, openMosix clusters, the NorduGrid grid system and Platforms LSF grid system.

Operating systems

PsN has been tested on UNIX and Microsoft Windows platforms. There should however, not be any problem to run PsN on any system that has Perl and NONMEM installed.
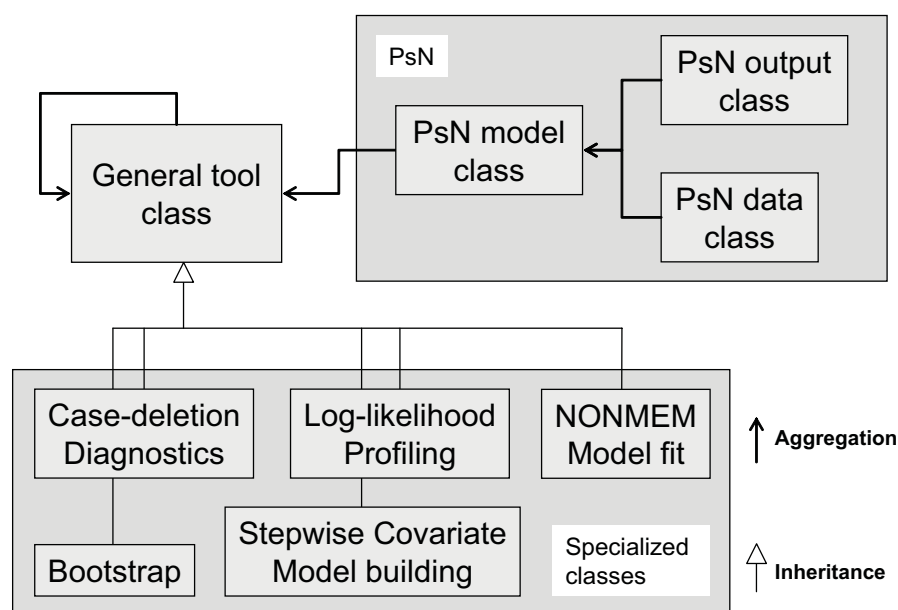
Figure 1. A simplified diagram showing the relation of the PsN classes.

The use of clinical irrelevance criteria in covariate model building with application to dofetilide pharmacokinetic data (Paper III)

The aim of this investigation was to characterise the pharmacokinetics of dofetilide in patients and to identify clinically relevant covariates and their respective contribution to changes in pharmacokinetic parameters. In addition, the consequences of applying three different modelling strategies in covariate model building were investigated using the analysis of dofetilide as an example: 1) using statistical criteria only or in combination with clinical irrelevance criteria for covariate selection, 2) applying covariate effects on total clearance or separately on non-renal and renal clearances and 3) using separate data sets for covariate selection and parameter estimation.

Dofetilide is a selective class III anti-arrhythmic drug with a narrow therapeutic index (66-69). Consequently, knowledge about covariate effects on the pharmacokinetics of dofetilide is valuable since it will increase the predictability of the pharmacokinetics in the patient. To build a predictive PK model the data set can be divided into three parts for model development, parameter estimation and evaluation of the predictive performance, respectively. Using data splitting should increase the predictive performance of the model and decrease bias in the parameter estimates.

Methods

Pooled concentration-time data (1445 patients, 10133 observations) from phase III clinical trials was used to develop a population pharmacokinetic model using NONMEM (26). The pooled data was split, preserving the relative contribution of each trial, in three parts, which were used for covariate selection, parameter estimation and evaluation of predictive performance. Stepwise covariate model building was used to identify important parameter-covariate relationships using the strategies described above. This was accomplished through the stepwise covariate model building tool (SCM) as implemented in PsN-Toolkit version 2.1.8 (Paper II). The statistical significance criterion for inclusion of covariates was set at the 5% level and the 0.1% significance level was used to exclude covariates from the full model. Inclusion and exclusion of covariates using clinical irrelevance was based on reduction in interindividual variability and changes in parameters at the extremes of the covariate distribution. Parametric separation of the elimination pathways was accomplished using creatinine clearance as an indicator of renal function. All covariates were tested on total clearance (joint clearance model) and separately on non-renal and renal clearance (split clearance model). The parameters were estimated using the FO and the FOCE methods. Mean absolute error was used to summarise the predictive performance of the models on the concentration level.

Results and conclusion

A one-compartment model with first order absorption adequately described the data. The base model, the full and the final covariate models all resulted in similar estimates of the structural parameters, the interindividual- and interoccasion variability and the residual error. Using clinical irrelevance criteria rather than only statistical criteria resulted in models containing less parameter-covariate relationships with only a minor loss in predictive power. When the elimination was divided into a renal- and non-renal part additional covariates were found significant although no gain in predictive power could be seen. The FO and FOCE estimation methods gave almost identical final covariate models with similar predictive performance (*Figure 2*). The parameters of the final model were estimated using the full data set and the FOCE method, and clinical irrelevance criteria were used for covariate inclusion. This model included creatinine clearance as a predictor of clearance and weight as a predictor of volume of distribution (Table 2).

In conclusion, in this study clinical irrelevance criteria were valuable for practical reasons since stricter inclusion/exclusion criteria shortens run times of the covariate model building procedure and because only covariates important for the predictive performance were included in the model. For large data sets containing a large amount of covariate information clinical irrelevance is likely superior to statistical significance as criteria in identifying important covariates.
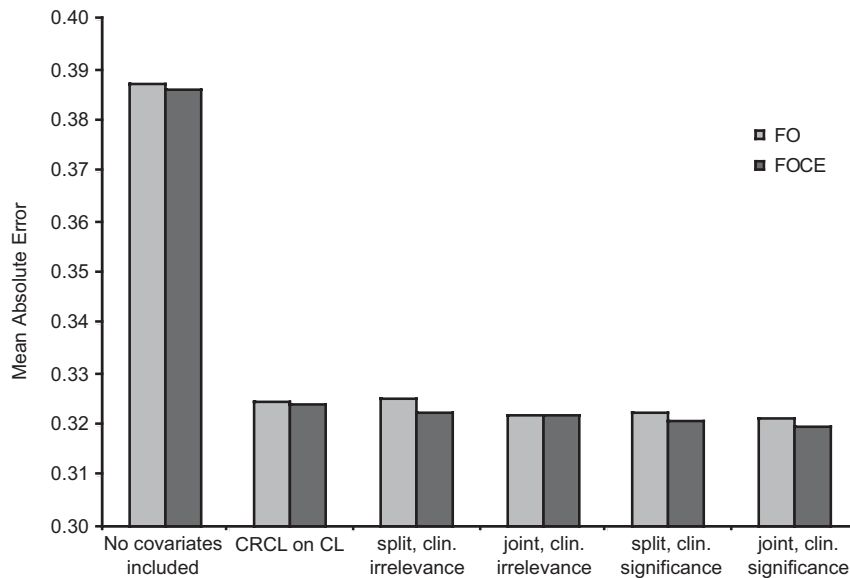
Figure 2. The mean absolute errors on the concentration level for the final models using the prediction data set. The parameters were estimated using the first-order (FO) or the first-order conditional estimation (FOCE) method.

Table 2. Parameter estimates (relative standard error (RSE (%)) for the base model and the final model using the full data set. The parameters were estimated using the first-order conditional estimation (FOCE) method and clinical irrelevance criteria were used for covariate inclusion.

	Base model (FOCE)	Final model Clinical irrelevance (FOCE)
Structural model		
CL/F (L/h)	15.5 (0.71)	15.5 (0.72)
V/F (L)	235 (1.2)	236 (1.2)
Ka (h^{-1})	1.4 (3.5)	1.4 (4.4)
Covariates (%)		
Creatinine clearance on CL/F	+0.89 (3.8)	+0.90 (3.2)
Weight on V/F	-	+0.70 (12)
Interindividual variability (%)		
CL/F	19 (8.4)	19 (8.4)
V/F	28 (9.0)	26 (10)
Correlation CL/F--V/F	49 (13)	57 (11)
Interoccasion variability (%)		
CL/F	19 (9.5)	19 (9.3)
Ka	107 (9.1)	109 (9.0)
Correlation CL/F--Ka	-42 (14)	-41 (14)
Residual error		
Additive on the Box-Cox transformed scale (0.6)	0.29 (2.5)	0.29 (2.5)
Objective function value	-9255	-9335

Evaluating the evaluations: resampling methods for determining model appropriateness in pharmacometric data analysis (Paper IV)

In this study, the objective was to categorise the performance of some of the statistical procedures implemented in PsN. We applied case-deletion diagnostics, log-likelihood profiling and the bootstrap for evaluation of 22 real world clinical and nonclinical data sets. All models and datasets were published, submitted for publication or regarded as finalised and ready for company reports.

Methods

The FO and the FOCE estimation methods were used for all models with a continuous response variable. For models with a heteroskedastic residual error sub-model the INTER variant of the FOCE method was applied. For models with categorical response variables, the Laplacian method was used.

Cook-scores and covariance ratios were calculated per subject for all data sets and estimation methods using case-deletion diagnostics. Wald confidence intervals were computed using estimates of standard error obtained either directly from NONMEM, or through a bootstrap. Bootstrap estimates of parameter confidence intervals were calculated using the nonparametric percentile and the BCa techniques. Log-likelihood profiling was used to acquire confidence intervals based on the decrease in likelihood around the maximum likelihood estimates.

The similarity of the confidence intervals attained through the different approaches was evaluated using both cluster analysis and a normalised mean difference from an index method. The percentile bootstrap estimate of confidence interval limits was chosen as the index method because it handles asymmetric distributions of estimates without relying on assumptions while being reasonably robust to outliers. Its use is also encouraged by regulatory authorities (53).

Results and conclusion

There was a clear tendency towards experiencing problems in the bootstrap and case-deletion diagnostics for models having high covariance matrix condition numbers (the absolute ratio between the highest and the lowest eigenvalues in a normal matrix) in the original analysis (*Figure 3*). Estimation during the bootstrap procedure using the FOCE method was also more prone to minimisation problems as compared to the FO method.

The LLP clearly provided narrower confidence intervals than the bootstrap. Over all parameters, the upper and lower confidence limits were 10% and 17% closer to the maximum likelihood estimate, respectively (*Figure 4*). The LLP also stood out as the most dissimilar method when compared to the rest of the methods using cluster analysis (p=0.05). However, when only models based on FOCE were considered, the LLP was the method that most closely agreed with the index method, especially for fixed-effects parameters (*Figure 5*).

This study shows that the condition number of the model fit covariance matrix is a reliable predictor of model stability to perturbations in the data. For the FO method, log-likelihood profiling resulted in confidence intervals that were narrower than the other methods evaluated in this study. Case-deletion diagnostics was shown to be a useful indicator of model stability in addition to being a tool for detecting influential or outlying individuals or groups of individuals. Based on the clear relation between condition number and the stability of the methods we were able to create some rules of thumb for their use. If the condition number is less than 50, all methods are likely to succeed. If the condition number is between 50 and 1000, successful application of the methods cannot be guaranteed. Therefore case-deletion diagnostics and log-likelihood profiling may be applied before a bootstrap is run (if at all) since these methods are less CPU-intensive. A condition number above 1000 is a clear indication of ill-conditioning and the appropriateness of the model will probably benefit from a re-evaluation.
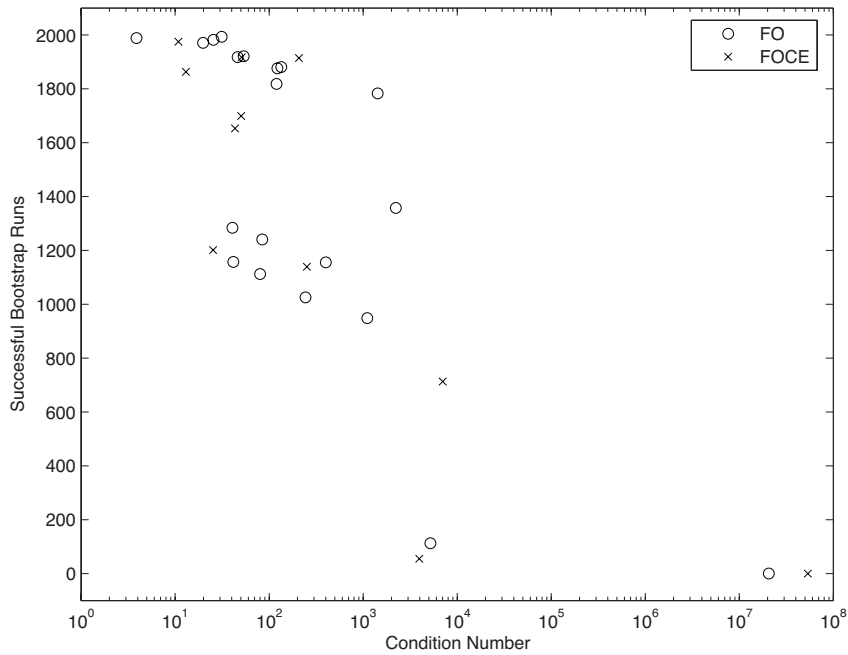

Figure 3. The number of successful bootstrap NONMEM runs, plotted against the condition number of the covariance matrix in the original analysis. All bootstraps included 2000 runs.
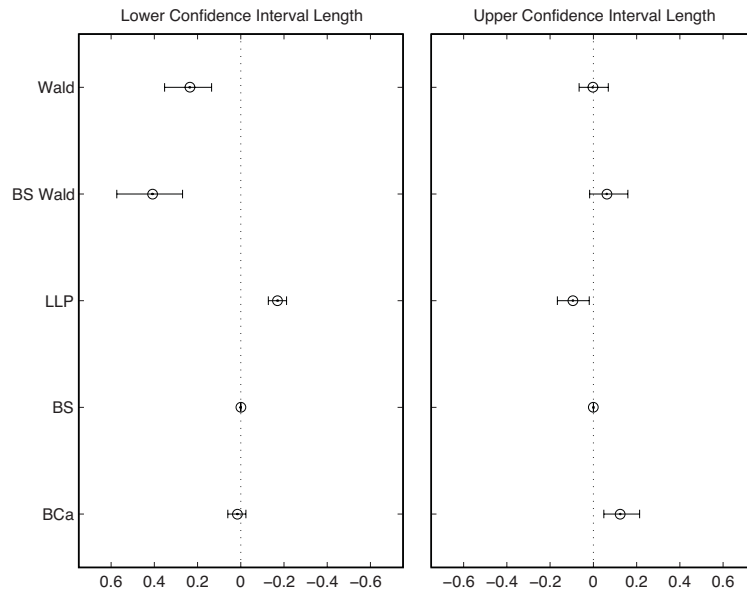
Figure 4. Mean difference for the methods for calculating confidence intervals compared to the percentile bootstrap. The differences are expressed as the relative length of the lower and upper part of the confidence intervals. The whiskers represent the 95% confidence region of the mean difference as computed by a non-parametric percentile bootstrap with 10000 samples.
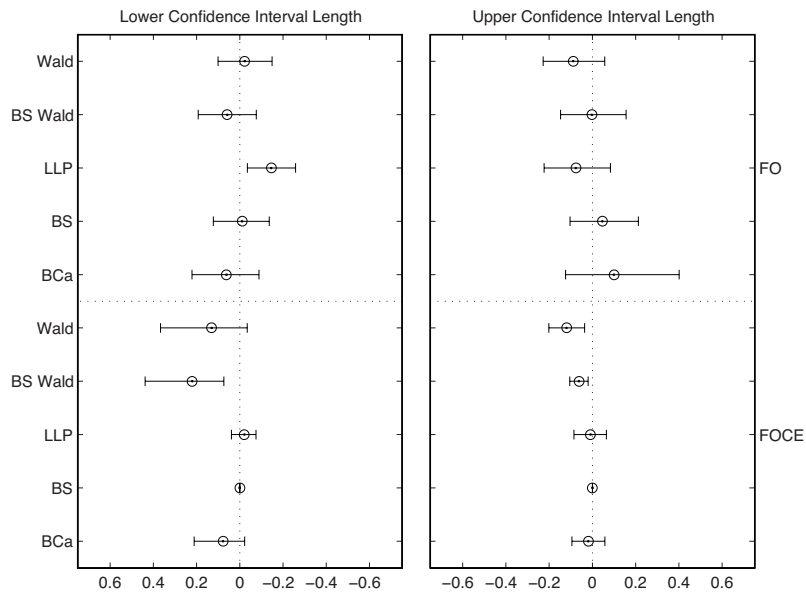

Figure 5. Mean difference for the methods for calculating confidence intervals for fixed-effect parameters compared to the percentile bootstrap using the FOCE estimation method. The differences are expressed as the relative length of the lower and upper part of the confidence intervals. The whiskers represent the 95% confidence region of the mean difference as computed by a non-parametric percentile bootstrap with 10000 samples.

Conclusions

In this thesis, new statistical methods in the area of pharmacometric data analysis were made available. This was done through the development of an API to NONMEM and the subsequent use of this API in the creation of a set of tools implementing various statistical methods. The value of the methods was evaluated using a large set of real world data.

Paper I described the creation of PsN, an API to the software package NONMEM. In paper II, a set of statistical methods including the bootstrap, jackknife, case-deletion diagnostics, log-likelihood profiling and stepwise covariate model building was made available as tools for application in pharmacometric data analysis through the software PsN.

The appropriateness of the methods and the consistency of the software tools were evaluated using a large selection of clinical and nonclinical data. One important result found through these evaluations was that criteria based on clinical relevance are useful components in automated stepwise covariate model building. As demonstrated using the anti-arrhythmic drug dofetilide in Paper III, clinical relevance criteria allowed for the ability to restrict the number of included parameter-covariate relationships while maintaining the predictive performance of the model. The clinical relevance criteria used would have been difficult to apply without PsN or equivalent software without spending a very large amount of time on manual construction of parameter-covariate relationships, running of NONMEM and inspection and evaluation of intermediate results.

In Paper IV, the condition number of the covariance matrix was shown to be a good indicator of how well the bootstrap and case-deletion diagnostics procedures behave when applied to PK/PD data analyses using NONMEM. It is not surprising that a model with a high condition number is likely to experience numerical problems in the regression procedure using case-deleted or bootstrap data sets. Notably, most models in this analysis showed such numerical problems in the regression procedure, despite being regarded as the final by the developers. Log-likelihood profiling can be used as a replacement for the bootstrap to calculate confidence intervals for fixed-effects parameters if the FOCE method is used. It is not recommended to use log-likelihood profiling in combination with the FO method.

In conclusion, the software developed in this thesis equips modellers with an enhanced set of tools for efficient pharmacometric data analysis. PsN is currently being used in research groups in academia and leading pharmaceutical companies as a tool in model development and evaluation.

Lars Lindbom

Acknowledgements

I would like to thank all of you who have contributed to this thesis. Especially I would like to thank:

Niclas Jonsson, min huvudhandledare, för bra stöd genom avhandlingsarbetet, för att du lärt mig så mycket om forskning och för att du är så genuint entusiastisk över alla tekniska/vetenskapliga utfärder. Sen så kan man ju inte förneka att den skrämmande värld som NONMEM erbjuder bleknar när man får insikt i såna saker som "reality dysfunctions, lighthuggers and Vogon poetry".

Mats Karlsson, min bihandledare för mycket stimulerande vetenskapliga diskussioner och uppmuntrande samtal. Till båda mina handledare vill jag säga tack för att ni fick mig att byta fält något och bli "kind of pharmacist".

Margareta Hammarlund-Udenaes, för tillhandahållande av utrustning och lokaler och för att du bryr dig om alla som jobbar på avdelningen.

Annika Scheynius och Ingeborg van der Ploeg, för att ni när jag hade läst 10 poäng cellbiologi och sökte sommarjobb på avdelningen för klinisk immunologi på Karolinska Universitetssjukhuset i Stockholm svarade "Ja, va kul att du vill komma hit och forska!" Året på KS var avgörande för att jag fortsatte med forskning efter min högskoleexamen.

My co-authors: Lynn McFadyen, Scott Marshall, Justin Wilkins and Nicolas Frey, for your contributions. It has been a pleasure working with you all.

Pontus Pihlgren, för att du hjälpt till med att dra PsN dit där det är idag. "suck …, det är du som kodat den här metoden va, Lasse?" Nej, seriöst; mycket, mycket kudos till dig.

Magnus Jansson, för att du har hjälpt till så fantastiskt bra med att få avhandlingen i snyggt skick.

Mama och Jakob, rumskompisar på jobbet och på resor, för hotellbokningar, pingismatcher, löprundor, wowande, långa diskussioner om kryphål, sannolikheter och mål i livet.

Alla på avdelningen. Ni som gör att det känns så kul att gå till jobbet. Man kan inte ha bättre arbetskamrater.

Min mor och far och mina syskon som betyder så mycket för mig, för att ni stöttat mig under hela avhandlingsarbetet och för att ni stått ut med mig och min dator under de senaste åren.

Karin, för att du är du. Älskar dig.

$\mathcal{L}$

References

1. Sheiner LB, Steimer JL. Pharmacokinetic/pharmacodynamic modeling in drug development. Annual Review of Pharmacology and Toxicology. 2000;40:67-95.
2. Balant LP, Gex-Fabry M. Modelling during drug development. European Journal of Pharmaceutics and Biopharmaceutics. 2000 Jul;50(1):13-26.
3. Sheiner LB, Rosenberg B, Melmon KL. Modelling of individual pharmacokinetics for computer-aided drug dosage. Comput Biomed Res. 1972 Oct;5(5):411-59.
4. Sheiner LB, Rosenberg B, Marathe VV. Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. J Pharmacokinet Biopharm. 1977 Oct;5(5):445-79.
5. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetics parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data. J Pharmacokinet Biopharm. 1980 Dec;8(6):553-71.
6. Sheiner BL, Beal SL. Evaluation of methods for estimating population pharmacokinetic parameters. II. Biexponential model and experimental pharmacokinetic data. J Pharmacokinet Biopharm. 1981 Oct;9(5):635-51.
7. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: routine clinical pharmacokinetic data. J Pharmacokinet Biopharm. 1983 Jun;11(3):303-19.
8. Olson SC, Bockbrader H, Boyd RA, Cook J, Koup JR, Lalonde RL, et al. Impact of population pharmacokinetic-pharmacodynamic analyses on the drug development process - Experience at Parke-Davis. Clinical Pharmacokinetics. 2000 May;38(5):449-59.
9. Vozeh S, Steimer JL, Rowland M, Morselli P, Mentre F, Balant LP, et al. The use of population pharmacokinetics in drug development. Clinical Pharmacokinetics. 1996 Feb;30(2):81-93.
10. Davidian M, Giltinan DM. Nonlinear models for repeated measurement data. London: Chapman & Hall; 1995.
11. Sheiner LB, Ludden TM. Population Pharmacokinetics Dynamics. Annual Review of Pharmacology and Toxicology. 1992;32:185-209.
12. Rowland M, Tozer TN. Clinical pharmacokinetics : concepts and applications. 3rd ed. Baltimore: Williams & Wilkins; 1995.
13. Karlsson MO, Sheiner LB. The Importance of Modeling Interoccasion Variability in Population Pharmacokinetic Analyses. Journal of Pharmacokinetics and Biopharmaceutics. 1993 Dec;21(6):735-50.
14. Laporte-Simitsidis S, Girard P, Mismetti P, Chabaud S, Decousus H, Boissel JP. Interstudy variability in population pharmacokinetic metaanalysis: when and how to estimate it? J Pharm Sci. 2000 Feb;89(2):155-67.
15. Sun H, Fadiran EO, Jones CD, Lesko L, Huang SM, Higgins K, et al. Population pharmacokinetics - A regulatory perspective. Clinical Pharmacokinetics. 1999 Jul;37(1):41-58.
16. Ette EI, Ludden TM. Population pharmacokinetic modeling: the importance of informative graphics. Pharm Res. 1995 Dec;12(12):1845-55.
17. Maitre PO, Buhrer M, Thomson D, Stanski DR. A 3-Step Approach Combining Bayesian Regression and Nonmem Population Analysis - Application to Midazolam. Journal of Pharmacokinetics and Biopharmaceutics. 1991 Aug;19(4):377-84.
18. Miller AJ. Selection of subsets of regression variables. J Roy Statist Soc Ser A. 1984;147(3):389--425.

19. Wahlby U, Jonsson EN, Karlsson MO. Assessment of actual significance levels for covariate effects in NONMEM. J Pharmacokinet Phar. 2001 JUN;28(3):231-52.

20. Wahlby U, Bouw MR, Jonsson EN, Karlsson MO. Assessment of type I error rates for the statistical sub-model in NONMEM. J Pharmacokinet Pharmacodyn. 2002 Jun;29(3):251-69.

21. Karlsson MO, Jonsson EN, Wiltse CG, Wade JR. Assumption testing in population pharmacokinetic models: Illustrated with an analysis of moxonidine data from congestive heart failure patients. Journal of Pharmacokinetics and Biopharmaceutics. 1998 Apr;26(2):207-46.

22. Derksen S, Keselman HJ. Backward, Forward and Stepwise Automated Subset-Selection Algorithms - Frequency of Obtaining Authentic and Noise Variables. British Journal of Mathematical & Statistical Psychology. 1992 Nov;45:265-82.

23. Altman DG, Andersen PK. Bootstrap Investigation of the Stability of a Cox Regression-Model. Statistics in Medicine. 1989 Jul;8(7):771-83.

24. Wahlby U, Jonsson EN, Karlsson MO. Comparison of stepwise covariate model building strategies in population pharmacokinetic-pharmacodynamic analysis. AAPS PharmSci. 2002;4(4): E27.

25. Ribbing J, Jonsson EN. Power, selection bias and predictive performance of the Population Pharmacokinetic Covariate Model. J Pharmacokinet Pharmacodyn. 2004 Apr;31(2):109-34.

26. Beal SL, Sheiner LB. NONMEM users guides. San Fransisco, CA: NONMEM project group, University of California; 1992.

27. Wolfinger R. Laplaces Approximation for Nonlinear Mixed Models. Biometrika. 1993 Dec;80(4):791-5.

28. Lindstrom MJ, Bates DM. Nonlinear Mixed Effects Models for Repeated Measures Data. Biometrics. 1990 Sep;46(3):673-87.

29. Pharsight. WinNonMix Users' guide version 2.0. Mountain View, California; 2000.

30. Leary RH, Jelliffe R, Schumitzky A, Port RE. Accurate vs. approximate likelihood methods in population analysis of sparse data sets. Clinical Pharmacology & Therapeutics. 2004 Feb;75(2): P30-P.

31. Ng CM, Guzy S, Bauer RJ. Validation of Monte-Carlo parametric expectation maximization algorithm (MCPEM) in analyzing population pharmacokinetic/pharmacodynamic (PK/PD) data. Clinical Pharmacology & Therapeutics. 2006 Feb;79(2):P28-P.

32. Mallet A. A Maximum-Likelihood-Estimation Method for Random Coefficient Regression-Models. Biometrika. 1986 Dec;73(3):645-56.

33. Schumitzky A. Nonparametric Em Algorithms for Estimating Prior Distributions. Applied Mathematics and Computation. 1991 Sep;45(2):143-57.

34. Kuhn E, Lavielle M. Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics & Data Analysis. 2005 Jun 15;49(4):1020-38.

35. Spiegelhalter D, Thomas A. Graphical modeling for complex stochastic systems: the BUGS project. Ieee Intelligent Systems & Their Applications. 1998 Mar-Apr;13(2):14-5.

36. Mentre F. History and new developments in estimation methods in nonlinear mixed-effects models. PAGE; 2005 16-17 June; Pamplona, Spain; 2005.

37. Aarons L. Software for population pharmacokinetics and pharmacodynamics. Clin Pharmacokinet. 1999 Apr;36(4):255-64.

38. Roe DJ. Comparison of population pharmacokinetic modeling methods using simulated data: Results from the population modeling workgroup. Statistics in Medicine. 1997 Jun 15;16(11):1241-57.

39. Duffull SB, Kirkpatrick CM, Green B, Holford NH. Analysis of population pharmacokinetic data using NONMEM and WinBUGS. J Biopharm Stat. 2005;15(1):53-73.

40. Girard P, Mentre F. A comparison of estimation methods in nonlinear mixed effects models using a blind analysis. PAGE; 2005 16-17 June; Pamplona, Spain; 2005.

41. Wald A. Setting of tolerance limits when the sample is large. The Annals of Mathematical Statistics. 1942 Dec.;13(4):389-99.

42. Hodges JS. 6 (or So) Things You Can Do with a Bad Model. Operations Research. 1991 May-Jun;39(3):355-65.

43. Jonsson EN, Karlsson MO. Xpose--an S-PLUS based population pharmacokinetic/pharmacodynamic model building aid for NONMEM. Comput Methods Programs Biomed. 1999 Jan;58(1):51-64.
44. Efron B. Better Bootstrap Confidence-Intervals. J Am Stat Assoc. 1987 MAR;82(397):171-85.
45. Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman & Hall; 1993.
46. Ette EI. Stability and performance of a population pharmacokinetic model. J Clin Pharmacol. 1997 Jun;37(6):486-95.
47. Rajagopalan P, Gastonguay MR. Population pharmacokinetics of ciprofloxacin in pediatric patients. J Clin Pharmacol. 2003 Jul;43(7):698-710.
48. Sheiner LB. Analysis of Pharmacokinetic Data Using Parametric Models .3. Hypothesis Tests and Confidence-Intervals. Journal of Pharmacokinetics and Biopharmaceutics. 1986 Oct;14(5):539-55.
49. Holford NH, Peace KE. Results and validation of a population pharmacodynamic model for cognitive effects in Alzheimer patients treated with tacrine. Proc Natl Acad Sci U S A. 1992 Dec 1;89(23):11471-5.
50. Sadray S, Jonsson EN, Karlsson MO. Likelihoodbased diagnostics for influential individuals in nonlinear mixed effects model selection. Pharm Res. 1999 Aug;16(8):1260-5.
51. Mandema JW, Verotta D, Sheiner LB. Building Population Pharmacokinetic Pharmacodynamic Models .1. Models for Covariate Effects. Journal of Pharmacokinetics and Biopharmaceutics. 1992 Oct;20(5):511-28.
52. Bruno R, Vivier N, Vergniol JC, DePhillips SL, Montay G, Sheiner LB. A population pharmacokinetic model for docetaxel (Taxotere(R)): Model building and validation. Journal of Pharmacokinetics and Biopharmaceutics. 1996 Apr;24(2):153-72.
53. U.S. Department of Health and Human Services FaDA. Guidance for Industry - Population Pharmacokinetics: U.S. Department of Health and Human Services, Food and Drug Administration; 1999.
54. Mandema JW, Verotta D, Sheiner LB. Building Population Pharmacokinetic-Pharmacodynamic Models. In: D'Argenio DZ, editor. Biomedical Simulations Research Workshop on Advanced Methods of Pharmacokinetic and Pharmacodynamic Systems Analysis; 1993; Los Angeles, California, USA: Plenum Press, New York; 1993. p. 69-86.
55. Ishibashi T, Yano Y, Oguma T. Population pharmacokinetics of platinum after nedaplatin administration and model validation in adult patients. Br J Clin Pharmacol. 2003 Aug;56(2):205-13.
56. Efron B. Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics. 1979 Jan.;7(1):1-26.
57. Hastie TJ, Friedman J, Tibshirani RJ. The elements of statistical learning : data mining, inference, and prediction. New York: Springer; 2001.
58. Stram DO, Lee JW. Variance-Components Testing in the Longitudinal Mixed Effects Model. Biometrics. 1994 Dec;50(4):1171-7.
59. Cook RD. Detection of influential observation in linear regression. Technometrics. 1977;19(1):15--8.
60. Quenouille MH. Approximate tests of correlation in timeseries. J Roy Statist Soc Ser B. 1949;11:68--84.
61. Quenouille MH. Notes on bias in estimation. Biometrika. 1956;43:353--60.
62. Moore GE. Cramming more components onto integrated circuits. Electronics. 1965;38(8):114-7.
63. Knox B. openMOSIX Homepage.   [cited 2006 21/3]; Available from: http://openmosix.sourceforge.net
64. Smirnova O, Eerola P, Ekelöf T, Ellert M, Hansen JR, Konstantinov A, et al. The NorduGrid Architecture and Middleware for Scientific Applications; 2003.
65. NorduGrid. The NorduGrid system homepage.   [cited 2004 17/2]; Available from: http://www.nordugrid.org
66. Jolling K, Ruixo JJ, Hemeryck A, Piotrovskij V, Greway T. Population pharmacokinetic analysis of pegylated human erythropoietin in rats. J Pharm Sci. 2004 Dec;93(12):3027-38.

67. Takama H, Tanaka H, Sudo T, Tamura T, Tanigawara Y. Population pharmacokinetic modeling and model validation of a spicamycin derivative, KRN5500, in phase 1 study. Cancer Chemother Pharmacol. 2001 May;47(5):404-10.

68. Mick R, Ratain MJ. Bootstrap validation of pharmacodynamic models defined via stepwise linear regression. Clin Pharmacol Ther. 1994 Aug;56(2):217-22.

69. Boni JP, Leister C, Bender G, Fitzpatrick V, Twine N, Stover J, et al. Population pharmacokinetics of CCI-779: correlations to safety and pharmacogenomic responses in patients with advanced renal cancer. Clin Pharmacol Ther. 2005 Jan;77(1):76-89.

70. Mulla H, McCormack P, Lawson G, Firmin RK, Upton DR. Pharmacokinetics of midazolam in neonates undergoing extracorporeal membrane oxygenation. Anesthesiology. 2003 Aug;99(2):275-82.

71. Hoppe A, Seronie-Vivien S, Thomas F, Delord JP, Malard L, Canal P, et al. Serum cystatin C is a better marker of topotecan clearance than serum creatinine. Clin Cancer Res. 2005 Apr 15;11(8):3038-44.

72. Mougenot P, Pinguet F, Fabbro M, Culine S, Poujol S, Astre C, et al. Population pharmacokinetics of melphalan, infused over a 24-hour period, in patients with advanced malignancies. Cancer Chemother Pharmacol. 2004 Jun;53(6):503-12.

73. Payen S, Serreau R, Munck A, Aujard Y, Aigrain Y, Bressolle F, et al. Population pharmacokinetics of ciprofloxacin in pediatric and adolescent patients with acute infections. Antimicrob Agents Chemother. 2003 Oct;47(10):3170-8.

74. Klein CE, Gupta E, Reid JM, Atherton PJ, Sloan JA, Pitot HC, et al. Population pharmacokinetic model for irinotecan and two of its metabolites, SN-38 and SN-38 glucuronide. Clin Pharmacol Ther. 2002 Dec;72(6):638-47.

75. Gauvin A, Pinguet F, Culine S, Astre C, Gomeni R, Bressolle F. A limitedsampling strategy to estimate individual pharmacokinetic parameters of vinorelbine in elderly patients with advanced metastatic cancer. Anticancer Drugs. 2002 Jun;13(5):473-80.

76. Phillips L, Grasela TH, Agnew JR, Ludwig EA, Thompson GA. A population pharmacokinetic-pharmacodynamic analysis and model validation of azimilide. Clin Pharmacol Ther. 2001 Oct;70(4):370-83.

77. Mulla H, Pooboni S. Population pharmacokinetics of vancomycin in patients receiving extracorporeal membrane oxygenation. Br J Clin Pharmacol. 2005 Sep;60(3):265-75.

78. Bonate PL, Floret S, Bentzen C. Population pharmacokinetics of APOMINE: a metaanalysis in cancer patients and healthy males. Br J Clin Pharmacol. 2004 Aug;58(2):142-55.

79. Blair EY, Rivory LP, Clarke SJ, McLachlan AJ. Population pharmacokinetics of raltitrexed in patients with advanced solid tumours. Br J Clin Pharmacol. 2004 Apr;57(4):416-26.

80. Chen C. Validation of a population pharmacokinetic model for adjunctive lamotrigine therapy in children. Br J Clin Pharmacol. 2000 Aug;50(2):135-45.

81. Cox EH, Veyrat-Follet C, Beal SL, Fuseau E, Kenkare S, Sheiner LB. A population pharmacokinetic-pharmacodynamic analysis of repeated measures time-to-event pharmacodynamic responses: the antiemetic effect of ondansetron. J Pharmacokinet Biopharm. 1999 Dec;27(6):625-44.

82. Gex-Fabry M, Haffen E, Paintaud G, Bizouard P, Sechter D, Bechtel PR, et al. Population pharmacokinetics of clomipramine, desmethylclomipramine, and hydroxylated metabolites in patients with depression receiving chronic treatment: model evaluation. Ther Drug Monit. 2000 Dec;22(6):701-11.

83. Falk RH, Pollak A, Singh SN, Friedrich T. Intravenous dofetilide, a class III antiarrhythmic agent, for the termination of sustained atrial fibrillation or flutter. Journal of the American College of Cardiology. 1997 Feb;29(2):385-90.

84. Falk RH, DeCara JM. Dofetilide: A new pure class III antiarrhythmic agent. American Heart Journal. 2000 Nov;140(5):697-706.

85. Norgaard BL, Wachtell K, Christensen PD, Madsen B, Johansen JB, Christiansen EH, et al. Efficacy and safety of intravenously administered dofetilide in acute termination of atrial fibrillation and flutter: A multicenter, randomized, doubleblind, placebocontrolled trial. American Heart Journal. 1999 Jun;137(6):1062-9.

86. Suttorp MJ, Polak PE, Vanthof A, Rasmussen HS, Dunselman PH, Kingma JH. Efficacy and Safety of a New Selective ClassIii Antiarrhythmic Agent Dofetilide in Paroxysmal Atrial-Fibrillation or Atrial-Flutter. American Journal of Cardiology. 1992 Feb 1;69(4):417-9.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy 33

Editor: The Dean of the Faculty of Pharmacy