Pharmaceutical Research, Vol. 22, No. 8, August 2005 (© 2005)
DOI: 10.1007/s11095-005-5269-5

Research Paper

Stochastic Differential Equations in NONMEM®: Implementation, Application, and Comparison with Ordinary Differential Equations

Christoffer W. Tornøe,[1,2,3,4] Rune V. Overgaard,[2] Henrik Agersø,[1] Henrik A. Nielsen,[2] Henrik Madsen,[2] and E. Niclas Jonsson[3]

Received December 10, 2004; accepted March 16, 2005

Purpose. The objective of the present analysis was to explore the use of stochastic differential equations (SDEs) in population pharmacokinetic/pharmacodynamic (PK/PD) modeling.
Methods. The intra-individual variability in nonlinear mixed-effects models based on SDEs is decomposed into two types of noise: a measurement and a system noise term. The measurement noise represents uncorrelated error due to, for example, assay error while the system noise accounts for structural misspecifications, approximations of the dynamical model, and true random physiological fluctuations. Since the system noise accounts for model misspecifications, the SDEs provide a diagnostic tool for model appropriateness. The focus of the article is on the implementation of the Extended Kalman Filter (EKF) in NONMEM® for parameter estimation in SDE models.
Results. Various applications of SDEs in population PK/PD modeling are illustrated through a systematic model development example using clinical PK data of the gonadotropin releasing hormone (GnRH) antagonist degarelix. The dynamic noise estimates were used to track variations in model parameters and systematically build an absorption model for subcutaneously administered degarelix.
Conclusions. The EKF-based algorithm was successfully implemented in NONMEM for parameter estimation in population PK/PD models described by systems of SDEs. The example indicated that it was possible to pinpoint structural model deficiencies, and that valuable information may be obtained by tracking unexplained variations in parameters.

KEY WORDS: degarelix; Extended Kalman Filter; NONMEM; parameter tracking; population PK/PD modeling; stochastic differential equations, systematic model development.

INTRODUCTION

Traditionally in population pharmacokinetic/pharmacodynamic (PK/PD) modeling, the variability is decomposed into inter-individual and intra-individual variability. The intra-individual (residual) variability accounts not only for the various environmental errors such as those associated with assay, dosing, and sampling errors, but also for errors associated with structural model misspecifications and approximations in the differential equations and variations in model parameters over time. Since most of these errors cannot be considered real sources of uncorrelated measurement noise, the sources of the intra-individual error should preferably be separated (1,2).

Erroneous dosing, sampling history, as well as structural model misspecifications may introduce time-dependent or serial correlated residual errors. Karlsson *et al.* (3) introduce three types of residual error models to population PK/PD data analysis to account for more complicated residual error structures. The study concluded that the inter-individual variability is overestimated while the residual variability is underestimated when serial correlations are present in the residuals but not accounted for. Neither accuracy nor precision of the structural parameters was improved by accounting for serial correlations. Only the variance components were more accurately estimated. Furthermore, time-dependent residual errors due to inaccurate recording of sampling times are very difficult to handle when dosing

[1] Experimental Medicine, Ferring Pharmaceuticals A/S, DK-2300, Copenhagen S, Denmark.
[2] Informatics and Mathematical Modelling, Technical University of Denmark, DK-2800, Lyngby, Denmark.
[3] Department of Pharmaceutical Biosciences, Division of Pharmacokinetics and Drug Therapy, Uppsala University, Box 591, 751 24, Uppsala, Sweden.
[4] To whom correspondence should be addressed. (e-mail: christoffer.tornoe@ferring.com)

ABBREVIATIONS: EKF, Extended Kalman Filter; FOCE, first-order conditional estimation; GnRH, gonadotropin releasing hormone; LLOQ, lower limit of quantification; ODE, ordinary differential equation; PK/PD, pharmacokinetics/pharmacodynamics; RSE, relative standard error; SC, subcutaneous; SDE, stochastic differential equation; Variables: e, measurement error; ϵ, residual error; η, inter-individual variability; K, Kalman gain; Ω, inter-individual covariance; P, state covariance; ϕ, individual parameter; R, output prediction covariance; $\sum$, measurement error covariance; σ_w, diffusion term; t, time; θ, population mean parameter; w, Wiener process; x, state variable; y, measurement.

patterns are more complex than single-dose or steady-state data. Another source of variability is unaccounted variations in model parameters. Part of this variability can sometimes be linked to surrogate variables (e.g., demographic covariates) but the variability is most often not predictable owing to the governing processes not being fully understood or too complex to model deterministically.

More sophisticated methods are therefore required to handle models with structural model misspecifications. This motivates the following questions: (1) How should one formulate a population PK/PD model to account for structural model deficiencies? (2) Given sparsely sampled, incomplete, and noise-corrupted PK/PD data, how should one optimally estimate the parameters in such models?

Grey-box PK/PD modeling provides an attractive approach by combining prior physiological knowledge about the modeled system with information from experimental data. Stochastic state-space or *grey-box* models consist of stochastic differential equations (SDEs) describing the dynamics of the system in continuous time and a set of discrete time measurement equations. The advantage of such models is that they allow for decomposition of the noise affecting the system into a system noise term representing unknown or incorrectly specified dynamics and a measurement noise term accounting for uncorrelated errors such as assay error (4–6).

The approach considered in this article aims to pursue the above mentioned questions by implementing SDEs in a nonlinear mixed-effects modeling setup, as previously investigated by Overgaard *et al.* (7). The intra-individual error is thereby decomposed into two types of noise: uncorrelated measurement noise arising from, for example, assay error, and system noise representing structural model deficiencies. This setup makes identification of structural model misspecification feasible by quantifying the model uncertainty and provides the basis for systematic model development (5,8).

The aims of this analysis are (1) to illustrate how SDEs can be implemented in NONMEM (9) (the most commonly used program for population PK/PD modeling), (2) demonstrate its application to systematic model development using a clinical PK data example, and finally (3) compare the results to common practice using ordinary differential equations (ODEs).

Theory

Nonlinear mixed-effects models based on SDEs extend the first-stage model of the hierarchical structure by decomposing the intra-individual variability into measurement and system noise (7). In the following, the population likelihood function based on SDEs is formulated along with a proposed algorithm for parameter estimation in SDE models. To ease the notation, bold symbols refer to vector or matrix representation. Readers not interested in the theory behind SDEs and state filtering methods can skip these sections without loss of understanding of the remainder of the article.

Nonlinear Mixed-Effects Models Based on SDEs

The first-stage model with SDEs written in state-space form consists of a set of continuous time system equations in Eq. (1) and discrete time measurement equations in Eq. (2) (see [4,5,7]), that is,

$$d\boldsymbol{x}_{it} = \boldsymbol{g}(\boldsymbol{x}_{it}, \boldsymbol{d}_{it}, \boldsymbol{\phi}_i)dt + \boldsymbol{\sigma}_w d\boldsymbol{w}_{it}, \qquad \boldsymbol{w}_{it} - \boldsymbol{w}_{is} \in N(\boldsymbol{0}, |t-s|\boldsymbol{I}) \tag{1}$$

$$\boldsymbol{y}_{ij} = \boldsymbol{f}(\boldsymbol{x}_{ij}, \boldsymbol{\phi}_i) + \boldsymbol{e}_{ij}, \qquad \boldsymbol{e}_{ij} \in N(\boldsymbol{0}, \boldsymbol{\Sigma}) \tag{2}$$

where $\boldsymbol{x}$ is the state vector, $\boldsymbol{y}$ is the observation vector, $\boldsymbol{d}$ is the input vector, $\boldsymbol{\phi}$ is the individual parameter vector, t is the time variable, $\boldsymbol{\sigma}_w d\boldsymbol{w}$ is the system noise, $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{e}$ is the measurement error with mean zero and covariance $\boldsymbol{\Sigma}$. The subscript ij in Eq. (2) refers to the measurement on individual i at time t_{ij}.

The structural model function $\boldsymbol{g}(\cdot)$ is called the drift term, the matrix $\boldsymbol{\sigma}_w$ is a scaling diffusion term, while $\boldsymbol{w}$ is a standard Wiener process also referred to as Brownian motion (10). The standard Wiener process $\boldsymbol{w}$ is a nonstationary stochastic process with mutually independent (orthogonal) increments $(\boldsymbol{w}_t - \boldsymbol{w}_s)$ which are Gaussian distributed with mean zero and variance $|t-s|\boldsymbol{I}$. If the diffusion term $\boldsymbol{\sigma}_w$ is zero, the SDE in Eq. (1) reduces to an ODE. The usual physiological interpretation of the parameters is thereby preserved in the SDE model formulation.

The second-stage model describing the inter-individual variability (IIV) is included in the same way as for ODEs. The individual parameters $\boldsymbol{\phi}_i$ are modeled as

$$\boldsymbol{\phi}_i = \boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{Z}_i)\exp(\boldsymbol{\eta}_i), \qquad \boldsymbol{\eta}_i \in N(\boldsymbol{0}, \boldsymbol{\Omega}) \tag{3}$$

where $\boldsymbol{h}(\cdot)$ denotes the structural type parameter model which is a function of the fixed-effects parameters $\boldsymbol{\theta}$, covariates $\boldsymbol{Z}_i$, and random-effects parameters $\boldsymbol{\eta}_i$ influencing $\boldsymbol{\phi}_i$. The random-effects $\boldsymbol{\eta}_i$ are assumed independent and multivariate normal distributed with zero mean and covariance matrix $\boldsymbol{\Omega}$. The three levels of random-effects $\boldsymbol{w}_{it}$, $\boldsymbol{e}_{ij}$, and $\boldsymbol{\eta}_i$ are assumed mutually independent for all i, t, and j.

The population likelihood function with SDEs can be written as

$$L(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}_w, \boldsymbol{\Omega}) \propto \prod_{i=1}^{N} \int p_1(\boldsymbol{\mathcal{Y}}_{in_i}|\boldsymbol{\eta}_i, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}_w, \boldsymbol{d}_i) p_2(\boldsymbol{\eta}_i|\boldsymbol{\Omega}) d\boldsymbol{\eta}_i = \prod_{i=1}^{N} \int \exp(\boldsymbol{l}_i) d\boldsymbol{\eta}_i \tag{4}$$

where $p_1(\cdot)$ and $p_2(\cdot)$ are the distributions associated with the first- and second-stage models, respectively, $\boldsymbol{\mathcal{Y}}_{in_i} = [\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{in_i}]$ represents all n_i observations of the ith individual, N is the total number of individuals, while $\boldsymbol{l}_i$ is the individual log-likelihood function (7).

Using the first-order conditional estimation (FOCE) method which approximates the likelihood using an expansion around the conditional estimates of the random effects $\hat{\boldsymbol{\eta}}_i$, the population likelihood function with SDEs can be written as

$$L(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}_w, \boldsymbol{\Omega}) \approx \prod_{i=1}^{N} \left|\Delta \boldsymbol{l}_i\right|^{-1/2} \exp(\boldsymbol{l}_i)|\hat{\boldsymbol{\eta}}_i \tag{5}$$

where $\Delta \boldsymbol{l}_i$ is the Hessian of $\boldsymbol{l}_i$. Further information about the FOCE method as implemented in NONMEM can be found in NONMEM Users Guide—part VII (9).

Assuming a Gaussian conditional density for the first-stage distribution density, the individual *a posteriori* log-likelihood function $\boldsymbol{l}_i$ and its Hessian $\Delta \boldsymbol{l}_i$ are given by

$$\boldsymbol{l}_i = -\frac{1}{2}\sum_{j=1}^{n_i}\left(\boldsymbol{\epsilon}_{ij}^T \boldsymbol{R}_{i(j|j-1)}^{-1}\boldsymbol{\epsilon}_{ij} + \log\left|2\pi \boldsymbol{R}_{i(j|j-1)}\right|\right) - \frac{1}{2}\boldsymbol{\eta}_i^T \boldsymbol{\Omega}^{-1}\boldsymbol{\eta}_i - \frac{1}{2}\log|2\pi\boldsymbol{\Omega}| \tag{6}$$

$$\Delta \boldsymbol{l}_i = -\sum_{j=1}^{n_i} \nabla \boldsymbol{\epsilon}_{ij}^T \, \boldsymbol{R}_{i(j|j-1)}^{-1} \nabla \boldsymbol{\epsilon}_{ij} - \boldsymbol{\Omega}^{-1} \tag{7}$$

respectively, where $\nabla \epsilon_{ij} = \frac{\partial \epsilon_{ij}}{\partial \boldsymbol{\eta}}\Big|_{\hat{\boldsymbol{\eta}}_i}$ is the gradient of the one-step prediction error ϵ_{ij} with respect to the random effects $\boldsymbol{\eta}$.

The one-step output prediction $\hat{\boldsymbol{y}}_{i(j|j-1)}$ is the optimal prediction of the jth measurement before that measurement is taken while $\boldsymbol{R}_{i(j|j-1)}$ is the expected covariance for that prediction. As indicated in (7), the Gaussian density is completely described by

$$\hat{\boldsymbol{y}}_{i(j|j-1)} = E\left[\boldsymbol{y}_{ij}\middle|\boldsymbol{\mathcal{Y}}_{i(j-1)}, \cdot\right] \tag{8}$$

$$\boldsymbol{R}_{i(j|j-1)} = V\left[\boldsymbol{y}_{ij}\middle|\boldsymbol{\mathcal{Y}}_{i(j-1)}, \cdot\right] \tag{9}$$

where $\hat{\boldsymbol{y}}_{i(j|j-1)}$ and $\boldsymbol{R}_{i(j|j-1)}$ are the conditional mean and covariance, respectively, of $\boldsymbol{y}_{ij}$ conditioned on all previous measurements up to time t_{j-1} for individual i denoted by $\boldsymbol{\mathcal{Y}}_{i(j-1)} = \left[\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{i(j-1)}\right]$. The subscript notation $i(j|j-1)$ in Eqs. (8) and (9) refers to the jth prediction based on all $j-1$ previous measurements for individual i.

The one-step prediction error are calculated by

$$\boldsymbol{\epsilon}_{ij} = \boldsymbol{y}_{ij} - \hat{\boldsymbol{y}}_{i(j|j-1)}, \quad \boldsymbol{\epsilon}_{ij} \in N\left(\boldsymbol{0}, \boldsymbol{R}_{i(j|j-1)}\right) \tag{10}$$

The one-step prediction $\hat{y}_{i(j|j-1)}$ and the associated covariance $\boldsymbol{R}_{i(j|j-1)}$ are calculated using the Extended Kalman Filter (EKF), which is described in the following section.

Extended Kalman Filter

The Extended Kalman Filter (EKF) (11,12) provides an efficient recursive algorithm to calculate the conditional mean and covariance for the assumed Gaussian conditional densities needed to evaluate the likelihood function in Eq. (5) (10,13).

The EKF equations can be grouped into two parts: prediction and update equations. The prediction equations predict the state and output variables one step ahead (i.e., until the next measurement) while the update equations update the state predictions with the newly obtained measurement.

The one-step *state prediction equations* of the EKF, which are the optimal (minimum variance) prediction of the mean and covariance, can be calculated by solving the state and state covariance prediction equations from measurement time t_{j-1} until t_j, that is,

$$\frac{d\hat{\boldsymbol{x}}_{i(t|j-1)}}{dt} = \boldsymbol{g}\left(\hat{\boldsymbol{x}}_{i(t|j-1)}, \boldsymbol{d}_i, \boldsymbol{\phi}_i\right), \qquad t \in \left[t_{j-1}, t_j\right] \tag{11}$$

$$\frac{d\hat{\boldsymbol{P}}_{i(t|j-1)}}{dt} = \boldsymbol{A}_{it}\boldsymbol{P}_{i(t|j-1)} + \boldsymbol{P}_{i(t|j-1)}\boldsymbol{A}_{it}^T + \boldsymbol{\sigma}_w\boldsymbol{\sigma}_w^T, \quad t \in \left[t_{j-1}, t_j\right] \tag{12}$$

with initial conditions

$$\hat{\boldsymbol{x}}_{i(1|0)} = \boldsymbol{x}_{i0} \tag{13}$$

$$\boldsymbol{P}_{i(1|0)} = \boldsymbol{P}_{i0} = \int_{t_1}^{t_2} e^{\boldsymbol{A}_{it}s}\boldsymbol{\sigma}_w\boldsymbol{\sigma}_w^T\left(e^{\boldsymbol{A}_{it}s}\right)^T ds \tag{14}$$

where t_1 and t_2 are the sampling times of the two first measurements while $\boldsymbol{x}_{i0}$ are the initial states which can be prespecified or estimated along with the other model parameters. The integral in Eq. (14) specifying the initial state covariance $\boldsymbol{P}_{i0}$ is taken as the integral of the Wiener process and the dynamics of the system over the time difference between the first two measurements. This initialization scheme has proven to be successful in other software implementations (6).

The EKF is an exact solution to the state filtering problem for linear systems, while it is a first-order approximative filter for nonlinear systems. Hence, the $\boldsymbol{A}_{it}$ matrix is calculated for nonlinear systems by linearizing the state equations in Eq. (11) using a local first-order Taylor expansion of $\boldsymbol{g}(\cdot)$ about the current state at each time instant t, that is,

$$\boldsymbol{A}_{it} = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}}\bigg|_{\boldsymbol{x}=\hat{\boldsymbol{x}}_{i(t|j-1)}} \tag{15}$$

Next, the EKF one-step *output prediction equations* are calculated by

$$\hat{\boldsymbol{y}}_{i(j|j-1)} = \boldsymbol{f}\left(\boldsymbol{\phi}_i, \hat{\boldsymbol{x}}_{i(j|j-1)}\right) \tag{16}$$

$$\boldsymbol{R}_{i(j|j-1)} = \boldsymbol{C}_{ij}\boldsymbol{P}_{i(j|j-1)}\boldsymbol{C}_{ij}^T + \boldsymbol{\Sigma} \tag{17}$$

where $\boldsymbol{C}_{ij}$ is obtained using a local first-order Taylor expansion of the measurement equation in Eq. (2), that is,

$$\boldsymbol{C}_{ij} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\bigg|_{\boldsymbol{x}=\hat{\boldsymbol{x}}_{i(j|j-1)}} \tag{18}$$

The one-step output prediction covariance $\boldsymbol{R}_{i(j|j-1)}$ is thus the sum of the state covariance associated with the observed states ($\boldsymbol{C}_{ij}\boldsymbol{P}_{i(j|j-1)}\boldsymbol{C}_{ij}^T$) and the covariance of the actual measurement ($\boldsymbol{\Sigma}$). In case of no system noise ($\boldsymbol{\sigma}_w = 0$), the one-step output prediction $\hat{\boldsymbol{y}}_{i(j|j-1)}$ and covariance $\boldsymbol{R}_{i(j|j-1)}$ will reduce to the ODE predictions $\hat{\boldsymbol{y}}_{ij}$ and residual covariance $\boldsymbol{\Sigma}$ typically used in the NONMEM likelihood function.

Finally, the one-step state and state covariance predictions are updated by conditioning on the jth measurement using the EKF *state update equations*, that is,

$$\hat{\boldsymbol{x}}_{i(j|j)} = \hat{\boldsymbol{x}}_{i(j|j-1)} + \boldsymbol{K}_{ij}\left(\boldsymbol{y}_{ij} - \hat{\boldsymbol{y}}_{i(j|j-1)}\right) \tag{19}$$

$$\boldsymbol{P}_{i(j|j)} = \boldsymbol{P}_{i(j|j-1)} - \boldsymbol{K}_{ij}\boldsymbol{R}_{i(j|j-1)}\boldsymbol{K}_{ij}^T \tag{20}$$

Table I. Implementation of Recursive EKF Algorithm in NONMEM VI

Control stream implementation	EKF algorithm	Equation	Description
35–37	$\hat{\boldsymbol{x}}_{i(1\|0)} = \boldsymbol{x}_{i0}$	(13)	Specify initial state conditions
Data file	$\boldsymbol{P}_{i(1\|0)} = \int_{t_1}^{t_2} e^{\boldsymbol{A}_{it}s}\boldsymbol{\sigma}_w\boldsymbol{\sigma}_w^T\left(e^{\boldsymbol{A}_{it}s}\right)^T ds$	(14)	Calculate initial state covariance
	for $i = 1$ to N do		For each of the N subjects
	for $j = 1$ to n_i do		For each of the ith subject's n_i measurements
104–106	$\frac{d\hat{\boldsymbol{x}}_{i(t\|j-1)}}{dt} = g\left(\hat{\boldsymbol{x}}_{i(t\|j-1)}, \boldsymbol{d}_i, \boldsymbol{\phi}_i\right)$	(11)	Compute one-step state prediction at time t_j
108–113	$\frac{d\boldsymbol{P}_{i(t\|j-1)}}{dt} = \boldsymbol{A}_{it}\boldsymbol{P}_{i(t\|j-1)} + \boldsymbol{p}_{i(t\|j-1)}\boldsymbol{A}_{it}^T + \boldsymbol{\sigma}_w\boldsymbol{\sigma}_w^T$	(12)	Calculate one-step state covariance prediction
120	$\hat{\boldsymbol{y}}_{i(j\|j-1)} = \boldsymbol{f}\left(\boldsymbol{\phi}_i, \hat{\boldsymbol{x}}_{i(j\|j-1)}\right)$	(16)	Compute one-step output prediction at time $\hat{\boldsymbol{y}}_{i(j\|j-1)}$
72 and 121	$\boldsymbol{R}_{i(j\|j-1)} = \boldsymbol{C}_{ij}\boldsymbol{P}_{i(j\|j-1)}\boldsymbol{C}_{ij}^T + \boldsymbol{\Sigma}$	(17)	Calculate one-step output covariance prediction variance
73–75	$\boldsymbol{K}_{ij} = \boldsymbol{P}_{i(j\|j-1)}\boldsymbol{C}_{ij}^T\boldsymbol{R}_{i(j\|j-1)}^{-1}$	(21)	Compute the Kalman Gain
78–80	$\hat{\boldsymbol{x}}_{i(j\|j)} = \hat{\boldsymbol{x}}_{i(j\|j-1)} + \boldsymbol{K}_{ij}\left(\boldsymbol{y}_{ij} - \hat{\boldsymbol{y}}_{i(j\|j-1)}\right)$	(19)	Update state prediction with the jth measurement
82–87	$\boldsymbol{P}_{i(j\|j)} = \boldsymbol{P}_{i(j\|j-1)} - \boldsymbol{K}_{ij}\boldsymbol{R}_{i(j\|j-1)}\boldsymbol{K}_{ij}^T$	(20)	Update state covariance
	end for		
	end for		

where $\hat{\boldsymbol{x}}_{i(j\,|\,j-1)}$ is the updated state estimate, $\boldsymbol{P}_{i(j|j)}$ is the updated state covariance, and the Kalman gain $\boldsymbol{K}_{ij}$ is calculated by

$$\boldsymbol{K}_{ij} = \boldsymbol{P}_{i(j\,|\,j-1)}\boldsymbol{C}_{ij}^T\boldsymbol{R}_{i(j\,|\,j-1)}^{-1} \tag{21}$$

The optimal state estimate at time j denoted by $\hat{\boldsymbol{x}}_{i(j\,|\,j)}$ is equal to the best state prediction $\hat{\boldsymbol{x}}_{i(j\,|\,j-1)}$ before the measurement is taken plus a correction term consisting of an optimal weighting value times the difference between the measurement $\boldsymbol{y}_{ij}$ and the one-step prediction of its value. For measurements with a large variance $\boldsymbol{\Sigma}$, the Kalman gain becomes small and the measurement is weighted lightly owing to the little confidence in the noisy measurement. In the limit as $\boldsymbol{\Sigma} \to \infty$, the Kalman gain becomes zero and the infinitely noisy measurement is completely ignored in the update. When the system noise is dominant implying uncertainty in the output of the system model, the measurement is heavily weighted. In the limit when $\boldsymbol{\sigma}_w \to \infty$ and $\boldsymbol{P} \to \infty$, the Kalman gain will approach 1 and the updated state will be equal to the measurement (13).

The EKF algorithm specified above and summarized in Table I is recursive by repeating the calculations of the one-step state and output prediction equations in Eqs. (11)–(18) as well as the state update equations in Eqs. (19)–(21) for each individual measurement.

METHODS

Implementation of EKF Algorithm in NONMEM

The EKF algorithm for parameter estimation in SDE models is implemented in NONMEM by modifying the standard NONMEM data file and control stream as described later for a clinical PK data example. An overview of

Table II. Extract from the Example SDE Data File

ID	ORIG	TIME	AMT	CMT	DV	EVID	MDV
1	–10.19	0.00	·	·	·	2	1
1	0.00	10.19	80000	1	·	1	1
1	0.00	10.19	80000	2	·	1	1
1	7.97	18.16	·	·	8.84	0	0
1	7.97	18.16	·	·	·	2	1
1	7.97	18.16	·	·	·	3	1
1	26.13	36.32	·	·	13.88	0	0
1	26.13	36.32	·	·	·	2	1
1	26.13	36.32	·	·	·	3	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

The data file includes the following records: ID (patient number), ORIG (original sampling time), TIME (new time variable), AMT (amount), CMT (compartment), EVID (event identifier), and MDV (missing dependent variable) record.

the necessary modifications is shown in Table I. The necessary control stream and data file modifications are made automatically using an S-PLUS script, which can be obtained by contacting the corresponding author.

Data File Modifications

The necessary data file modifications are illustrated by an extract from the example SDE data file in Table II. In order for the EKF to work, the initial states $\boldsymbol{x}_{i0}$ and the associated covariance matrix $\boldsymbol{P}_{i0}$ must be specified. The initial states $\boldsymbol{x}_{i0}$ are often zero when dealing with pharmacokinetics of an exogenous compound but might otherwise be specified to some initial amounts or estimated along with the other model parameters. The integral in Eq. (14) specifying the initial state covariance $\boldsymbol{P}_{i0}$ cannot be entered directly into NONMEM. This problem is handled by rewinding the time variable so that the estimation starts at t_{start} calculated by

$$t_{start} = t_1 - (t_2 - t_1) \tag{22}$$

where t_1 and t_2 are the sampling times of the first and second measurements, respectively, which in the example data in Table II are equal to 7.97 and 26.13 h.

The initial conditions $\boldsymbol{P}_{i0}$ are thereby calculated exactly as the integral in Eq. (14) at the time of the first measurement. If t_{start} is negative, it is necessary to create a new time variable since negative time data records are not accepted in NONMEM. For the example data in Table II, the new time variable (TIME) is therefore calculated by subtracting t_{start} from the original sampling time variable (ORIG), that is, TIME = ORIG − t_{start}. This new starting time is implemented in the data file by an event identification (EVID) record equal to 2 and the missing dependent variable (MDV) data item set to 1.

The doses in Table II are specified as usual with a dose event record (EVID = 1). Finally, each observation is duplicated and presented as three records in the data file with an EVID = 0, EVID = 2, and EVID = 3 record. The EVID = 0 makes the one-step prediction as usual, the EVID = 2 record stores the one-step prediction $\hat{\boldsymbol{x}}_{i(j|j-1)}$ and $\boldsymbol{P}_{i(j|j-1)}$ from the EVID = 0 record before the states are reset to zero and updated with the EKF updates in the subsequent EVID = 3 record.

Control Stream Modifications

The NONMEM SDE control stream for a one-compartment disposition model with two absorption components is shown in Appendix. The one-step prediction equations of the EKF are specified in the $DES block of the NONMEM control stream. The $\frac{k(k+1)}{2}$ one-step state covariance prediction equations in lines 108–113 associated with the k one-step state prediction equations in lines 104–106 are calculated using Eq. (12). The one-step state and state covariance prediction equations (both of which are systems of ODEs)

Table III. Summary of Population Pharmacokinetic Parameter Estimates for the Initial Model with First-Order Absorption (**Initial Model**), Tracking Absorption Half-Life Model (**Tracking $t_{1/2,abs}$**), and the Final Model with Two Absorption Components (**Final Model**)

Model		**Initial model**	**Tracking $t_{1/2,abs}$**[a]	**Final model**			
Method		SDE	SDE	SDE		ODE	
OFV		-1809	-323	-3387		-3236	
Parameter	Units	Estimate	Estimate	Estimate	RSE (%)	Estimate	RSE (%)
$t_{1/2,abs}$	h	484	—	—	—	—	—
$t^0_{1/2,abs}$	h	—	245	—	—	—	—
$t_{1/2,fast}$	h	—	—	32.3	9.47	32.5	9.45
$t_{1/2,slow}$	h	—	—	949	3.94	922	4.09
CL/F	$\frac{1}{\text{h}}$	4.64	12.1	5.48	3.69	5.68	7.59
V/F	l	20.9	108	102	8.41	102	13.2
Fr	—	—	—	0.112	3.86	0.107	4.15
F^{20}_{rel}	—	0.511	—	0.71	3.66	0.735	7.84
$\omega_{t_{1/2,abs}}$	%	25.9	—	—	—	—	—
$\omega_{t_{1/2,fast}}$	%	—	—	31.7	10.9	29.7	12.7
$\omega_{t_{1/2,slow}}$	%	—	—	24.7	11.7	31.5	8.40
ω_{CL}	%	—	31.9	—	—	—	—
ω_{V}	%	—	54.1	—	—	—	—
ω_{Fr}	%	—	—	33.8	10.1	36.8	8.59
$\omega_{F_{rel}}$	%	29.8	—	27.8	6.65	28.0	6.18
σ_{prop}	%	29.8	15.7	14.5	3.98	17.6	3.85
σ^{abs}_{w}	$\frac{\mu g}{\sqrt{\text{h}}}$	82.2	0.00991	—	—	—	—
$\sigma^{central}_{w}$	$\frac{\mu g}{\sqrt{\text{h}}}$	0.0[b]	0.0[b]	0.0[b]	—	—	—
$\sigma^{t_{1/2,abs}}_{w}$	$\frac{1}{\sqrt{\text{h}}}$	—	0.0346	—	—	—	—
σ^{fast}_{w}	$\frac{\mu g}{\sqrt{\text{h}}}$	—	—	0.0[b]	—	—	—
σ^{slow}_{w}	$\frac{\mu g}{\sqrt{\text{h}}}$	—	—	244	10.6	—	—

[a] Single-dose data from group 3 only.
[b] Fixed parameter.

are progressed until the time of the next measurement.

The one-step output prediction equation in line 120, which is a function of the central compartment and its volume, is defined in the $ERROR block. The standard deviation of the one-step prediction error is calculated in line 121 as the square root of $\boldsymbol{R}_{i(j|j-1)}$ in Eq. (17).

Updating the state predictions and covariances is slightly tricky in NONMEM, because this involves first a reset of the compartment and subsequently an update to the new value, which is all performed at EVID = 3. Lines 34–68 store the one-step state and state covariance prediction values, and ensure that these values are not overwritten during system reset. Calculation of the measurement updated predictions is based on the stored state values, as performed in lines 71–87. First, the one-step output prediction variance $\boldsymbol{R}_{i(j|j-1)}$ is calculated. Next, the Kalman gain $\boldsymbol{K}_{ij}$ is calculated in lines 74–76, and the state and state covariance updates are calculated in lines 78–87. Finally the execution of the update is performed in lines 91–99.

Example Data

Clinical PK data of GnRH antagonist degarelix is used to illustrate the application of SDEs in population PK/PD modeling.

Study Design

The study was designed as a 6-month, multicenter, open-labeled, equally randomized, parallel group trial investigating the efficacy and safety of three dose regimens of degarelix in prostate cancer patients. The study was performed in accordance with the Declaration of Helsinki and according to Good Clinical Practice (GCP). The appropriate independent ethics committee approved the protocol prior to study initiation. Written informed consent was obtained from all patients prior to participation in the study.

The following dose regimens were included in the study:

Group 1 received 80 mg at days 0 and 3, and 40 mg at day 28 and every 4 weeks thereafter (80/80/40).

Group 2 received 40 mg at days 0, 3, 28, and every 4 weeks thereafter (40/40/40).

Group 3 received 80 mg at day 0, 20 mg at day 28, and every 4 weeks thereafter (80/–/20).

The doses were administered as subcutaneous (SC) injections of the following volume and concentrations:

80 mg as two SC injections of 2 ml with 20 mg/ml.
40 mg as one SC injection of 2 ml with 20 mg/ml.
20 mg as one SC injection of 2 ml with 10 mg/ml.

One hundred and twenty-nine patients were randomized to receive at least one dose of degarelix, with 43 patients in treatment group 1, 46 patients in treatment group 2, and 40 patients in treatment group 3. Blood samples were drawn before dosing at each visit except for visit 1. At visit 2, an additional blood sample was taken 8 h after dosing. Degarelix plasma concentrations were measured using a validated liquid chromatography with tandem mass spectrometric detection (LC-MS/MS) method according to current guidelines for bioanalytical samples (14) with a lower limit of quantification (LLOQ) of 0.5 ng/ml.

Data Analysis

The data analysis was performed using NONMEM version VI beta with subroutines ADVAN6 and ADVAN8 and relative tolerance set to six on a Pentium 4-M 2.0 GHz computer with 512 MB RAM running Windows XP. Three significant digits were requested in the parameter estimates. All models were estimated using the first-order conditional estimation (FOCE) method.

The individual parameter vector $\boldsymbol{\phi}_i$ was assumed to be distributed around the population mean parameter vector $\boldsymbol{\theta}$ according to the exponential IIV model in Eq. (3), that is,

$$\boldsymbol{\phi}_i = \boldsymbol{\theta}\exp(\boldsymbol{\eta}_i), \qquad \boldsymbol{\eta}_i \in N(\mathbf{0}, \boldsymbol{\Omega}) \tag{23}$$

where the random-effects $\boldsymbol{\eta}_i$ influencing $\boldsymbol{\phi}_i$ are assumed multivariate normal distributed with zero mean and covariance matrix $\boldsymbol{\Omega}$. The inter-individual covariance matrix $\boldsymbol{\Omega}$ was specified as a diagonal matrix with ${\omega_k}^2$ in the k^{th} diagonal element. Correlations between all pairs of $\boldsymbol{\eta}_i$ were investigated in the pre-analysis but were found to be non-significant.

RESULTS

To illustrate the application of SDEs in population PK/PD modeling, clinical PK degarelix data are used in the following systematic model development scheme motivated by (8). First, the structural model misspecifications of the initial PK model are pinpointed using the diffusion term estimates. Next, the model is expanded to track unexplained time variations in the entity of the model believed to be deficient, which leads to an extension of the SDE model. Finally, the parameter estimates of the final SDE model are compared with the corresponding ordinary differential equation (ODE) model.

Initial Model

A one-compartment disposition model with first-order absorption and elimination was selected as the initial PK model describing the SC administration of degarelix. From previous studies, it is known that the SC release of degarelix does not follow simple first-order absorption kinetics (15). But in order to demonstrate the use of SDEs for systematic model development, it was chosen as the initial PK model. The system of SDEs governing the initial PK model can be written in matrix notation together with the observation equation as

$$d\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} -k_a A_1 + F_{rel} D(t) \\ k_a A_1 - \dfrac{CL}{V} A_2 \end{pmatrix} dt + \begin{pmatrix} \sigma_w^{abs} & 0 \\ 0 & \sigma_w^{central} \end{pmatrix} d\boldsymbol{w}_t$$

$$\log c = \log \frac{A_2}{V} + e \tag{24}$$

where A_1 and A_2 are the state variables for the amount of drug in the SC absorption and central compartment, respectively, while c is the observed plasma degarelix concentration. The diffusion term $\boldsymbol{\sigma}_w$ was specified as a diagonal matrix to pinpoint possible model misspecifications. The presence of significant parameters in the diffusion term is an indication that the corresponding drift term may be incorrectly specified (8). The system noise $\boldsymbol{\sigma}_w d\boldsymbol{w}_t$ is additive while σ_{prop} is the coefficient of variation for the measurement error.

The model in Eq. (24) was parameterized in terms of CL and V symbolizing the clearance and volume of the central compartment, respectively, while the first-order absorption rate constant k_a was estimated as the absorption half-life $t_{1/2,abs} = \frac{\log 2}{k_a}$ and $D(t)$ is the input from the SC degarelix doses. The bioavailability of SC injected degarelix doses has previously been reported to be dependent on the concentration of the injected suspension (15,16). The relative bioavailability F_{rel} of the 10 mg/ml doses was fixed to 1 while the bioavailability of the 20 mg/ml doses relative to the 10 mg/ml doses was estimated as the parameter F_{rel}^{20}. IIV was estimated for the parameters $t_{1/2,abs}$ and F_{rel} using the exponential model in Eq. (23).

The parameter estimates of the initial SDE model are shown in Table III. The diffusion parameters can be used to pinpoint possible model misspecifications and provide information on how to reformulate the model. The diffusion parameter associated with the absorption compartment σ_w^{abs} was estimated to 82.2 $\frac{\mu g}{\sqrt{hr}}$ with an asymptotic relative standard error (RSE) estimate of 41.1% while that of the central compartment $\sigma_w^{central}$ was insignificant and fixed to zero in the final estimation. This may indicate that the disposition of degarelix is sufficiently explained by a one-compartment disposition model while the first-order absorption model inadequately describes the SC absorption profile of degarelix.

Tracking Absorption Half-Life Model

Since the diffusion term on the absorption compartment was significant in the initial PK model, the point of interest for the following model expansion should be the absorption model. From previous studies, it is known that SC administration of degarelix results in the formation of a gel-like *in situ* SC depot from which the drug is released into the systemic circulation (15). The depot undergoes a maturation stage where the density of the gel increases and the release rate decreases.

The absorption half-life is therefore suspected to change over time. Instead of jumping to conclusions by imposing a structural model for this behavior, we want to track the absorption half-life $t_{1/2,abs}$ and thereby confirm/reject whether it is constant throughout the study. This is done by expanding the model with a state equation for the absorption half-life that fluctuate randomly like a Wiener process. For a given individual, the EKF gives the filtered estimate of the evolution of $t_{1/2,abs}$. The drift term in the state equation for the absorption half-life is set to zero, thereby assuming it is driven only by the random variation $\sigma_w^{t_{1/2,abs}}\ dw_t$. If $\sigma_w^{t_{1/2,abs}}$ turns out to be significant, it is an indication that the absorption half-life is not constant since the variations in $t_{1/2,abs}$ are explained by the diffusion term.

The model for tracking the variations in $t_{1/2,abs}$ can thereby be written as

$$d\begin{pmatrix} A_1 \\ A_2 \\ \log t_{1/2,abs} \end{pmatrix} = \begin{pmatrix} -\dfrac{\log 2}{\exp\left(\log t_{1/2,abs}\right)}\ A_1 + D(t) \\ \dfrac{\log 2}{\exp\left(\log t_{1/2,abs}\right)}\ A_1 - \dfrac{CL}{V} A_2 \\ 0 \end{pmatrix} dt + \begin{pmatrix} \sigma_w^{abs} & 0 & 0 \\ 0 & \sigma_w^{central} & 0 \\ 0 & 0 & \sigma_w^{t_{1/2,abs}} \end{pmatrix} dw_t$$

$$\log\ c = \log \frac{A_2}{V} + e \tag{25}$$

where $t_{1/2,abs}$ is the state variable for the absorption half-life with initial value $t_{1/2,abs}^0$. To constrain $t_{1/2,abs}$ to be non-

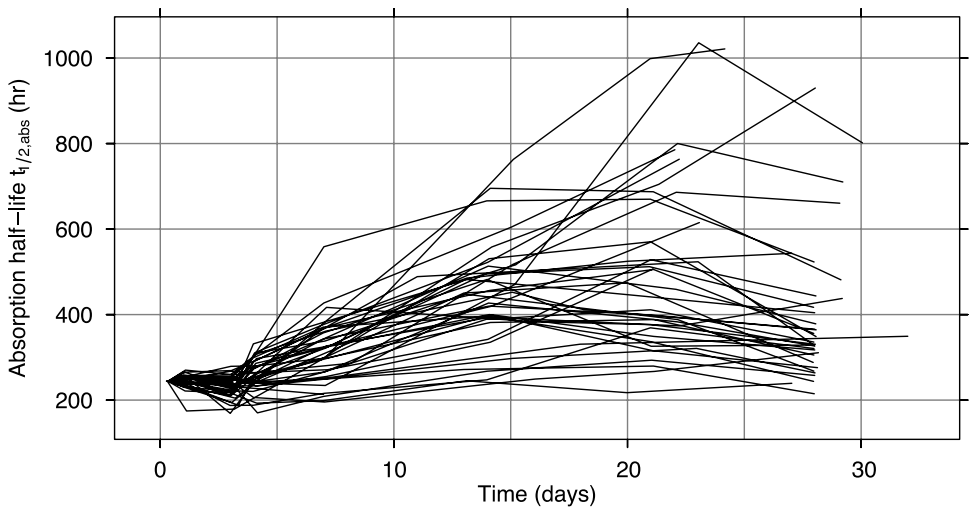

Fig. 1. Tracking absorption half-life $t_{1/2,abs}$ for patients in group 3 (80/-/20) using single-dose data only. Each line represents the tracked absorption half-life for one individual.

negative, multiplicative (i.e., state dependent) system noise is needed. By log-transforming the state equation for the absorption half-life, the multiplicative system noise on $t_{1/2,abs}$ was transformed into additive system noise in Eq. (25) which can be estimated using the EKF based algorithm (7). IIV was estimated for the parameters *CL* and *V* using the exponential model in Eq. (23).

For simplicity reasons in the implementation of this model, only single-dose data from group 3 were included in this estimation since a new depot is formed at each injection which would make it necessary to track the different absorption half-lives from each depot. The estimated parameters for the tracking absorption half-life model are shown in Table III. The diffusion parameter on the absorption compartment σ_w^{abs} was now reduced to 0.00991 $\frac{\mu g}{\sqrt{hr}}$ while $\sigma_w^{t_{1/2,abs}}$ was estimated to 0.0346 $\frac{1}{\sqrt{hr}}$ with an asymptotic RSE estimate of 8.99%. This confirms the suspicion that $t_{1/2,abs}$ is not constant throughout the study since $\sigma_w^{t_{1/2,abs}}$ is significantly different from zero.

Tracking of the absorption half-life is shown in Fig. 1 where the updated EKF estimates of $t_{1/2,abs}$ using Eq. (19) are plotted as a function of time. The initial absorption half-life estimate is 245 h and remains constant until approximately 3 days after drug administration for most of the patients. From 3 to 14 days the absorption half-life increases and seems to reach a new constant or slightly decreasing level thereafter. A reasonable approximation of the pattern seen in Fig. 1 would be a model with two absorption components, that is, an initial fast absorption phase followed by a prolonged slow release phase. This interpretation corresponds well with the clinical observation that SC administration of degarelix results in the formation of a depot out of which the drug diffuses.

Final Model

The final step in this illustrative example is to reformulate the initial PK model with the information obtained by tracking the absorption half-life. The final PK model with two first-order absorption components (see Fig. 2) was modeled by the following system of SDEs and observation equation

$$d\begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} = \begin{pmatrix} -k_{a,1}A_1 + Fr\,F_{rel}D(t) \\ -k_{a,2}A_2 + (1-Fr)F_{rel}D(t) \\ k_{a,1}A_1 + k_{a,2}A_2 - \dfrac{CL}{V}A_3 \end{pmatrix} dt + \begin{pmatrix} \sigma_w^{fast} & 0 & 0 \\ 0 & \sigma_w^{slow} & 0 \\ 0 & 0 & \sigma_w^{central} \end{pmatrix} dw_t \qquad (26)$$

$$\log c = \log \frac{A_3}{V} + e$$

where A_1, A_2, and A_3 are the state variables for the amount of drug in the slow SC absorption, fast SC absorption, and central compartment, respectively. The fraction of the SC dose being absorbed via the fast and slow absorption route were estimated as *Fr* and (1 – *Fr*), respectively.

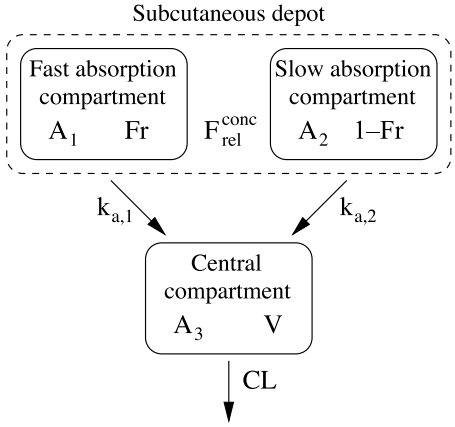

Fig. 2. Schematic illustration of the final PK model for SC injected GnRH antagonist degarelix. The model parameters are explained in the text.

The first-order absorption rate constants $k_{a,1}$ associated with the fast absorption components was estimated as the absorption half-life $t_{1/2,fast} = \frac{\log 2}{k_{a,1}}$. To ensure that $t_{1/2,slow} = \frac{\log 2}{k_{a,2}}$ was constrained to be larger than the $t_{1/2,fast}$ even after taking IIV into account, the following parametrization was chosen.

$$t_{1/2,slow,i} = t_{1/2,slow-fast} \exp\left(\eta_{t_{1/2,slow-fast,i}}\right) + t_{1/2,fast,i} \qquad (27)$$

where $t_{1/2,slow-fast}$ is the typical individual's difference between $t_{1/2,fast}$ and $t_{1/2,slow}$. IIV was estimated for the parameters $t_{1/2,fast}$, $t_{1/2,slow}$, and F_{rel} using the exponential model in Eq. (23). The parameter *Fr* was constrained to be between 0 and 1 by logit-transformation, that is,

$$\rho = \log \frac{Fr}{1-Fr} \qquad (28)$$

and the individual parameter Fr_i was calculated by

$$Fr_i = \frac{\exp(\rho + \eta_{Fr_i})}{1 + \exp(\rho + \eta_{Fr_i})} \qquad (29)$$

The SDE control stream for the final PK model is shown in Appendix A. The run times for the estimation of model parameters in Eq. (26) using SDEs was about three times that of the corresponding ODE model. Since the one-step output prediction variance $\boldsymbol{R}_{i(j|j-1)}$ in the SDE model depends on the individual η's [see Eq. (17)], the FOCE method with interaction (FOCE-INTER) was used. For the corresponding ODE model (i.e., $\boldsymbol{\sigma}_w = 0$), the FOCE method was applied without the INTER option since the intra-individual variance is homoscedastic on the log-scale.

The population PK parameter estimates from the final PK model with two absorption components using ODEs and SDEs are summarized in Table III. The diffusion parameters σ_w^{fast} and $\sigma_w^{central}$ for the fast absorption and central com-

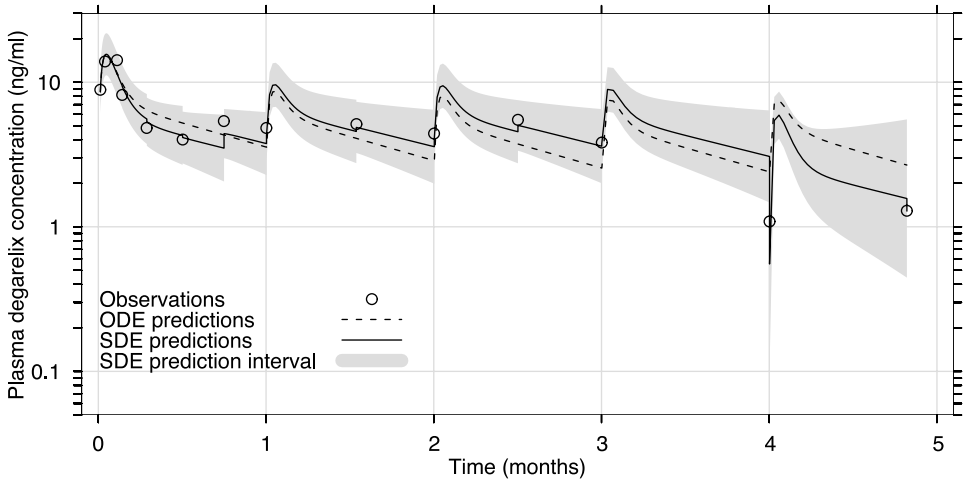

Fig. 3. One-step individual SDE predictions, updates, and prediction interval plotted together with the individual ODE predictions for an illustrative patient. The SDE prediction interval is calculated by $\hat{y}_{j|j-1} \pm \sqrt{R_{j|j-1}}$.

partment, respectively, were fixed to zero in the final run since they were found not to be significantly different from zero. The only diffusion parameter that was estimated was therefore the diffusion parameter on the slow absorption compartment. The estimated standard deviation of $\hat{\sigma}_w^{slow} = 244 \frac{\mu g}{\sqrt{hr}}$ for the additive system noise indicates that there still is a model misspecification in the slow absorption compartment. However, it is relatively small compared to the amount of drug in the slow absorption compartment which ranges from 2.500 to 80.000 μg throughout the study. Other more complex absorption models such as diffusion out of a spherical SC depot have previously been suggested to describe the PK profile of SC administered degarelix (15). For this study, the final PK model with two absorption components representing an initial fast release followed by a prolonged slow release was considered acceptable at describing the observed data.

Since the SDE and ODE models are nested, the significance of including the diffusion parameter σ_w^{slow} can be tested using the likelihood ratio test (LRT) with $\Delta OFV = -3236 - (-3387) = 151 > \chi^2(1)_{0.95} = 3.84$ which is highly significant with a p-value < 0.001. When comparing the SDE and ODE parameter estimates, the parameters which seem affected the most by including system noise on the slow absorption compartment are $\omega_{t_{1/2,slow}}$ and σ_{prop}. The inter-individual variability of $t_{1/2,slow}$ was reduced from a coefficient of variation (CV) of 31.5% to 24.7% while the CV of the proportional measurement error was deflated from 17.6% to 14.5% when using SDEs instead of ODEs. The RSE estimates were more or less the same for the SDE and ODE models.

The one-step individual SDE predictions, updates, and prediction interval are shown together with the individual ODE predictions for an illustrative patient from the study in Fig. 3. The observed discrepancy between the SDE and ODE predictions are due to the SDE predictions being conditioned on all previous observations and therefore updated at each sampling time (visualized by the vertical lines in the SDE predictions).

The ODE predictions are considerably lower than the observed plasma concentrations between 1 and 3 months and above thereafter, which falsifies the statistical assumption of uncorrelated residuals. All except one of the one-step SDE predictions are within the SDE prediction interval

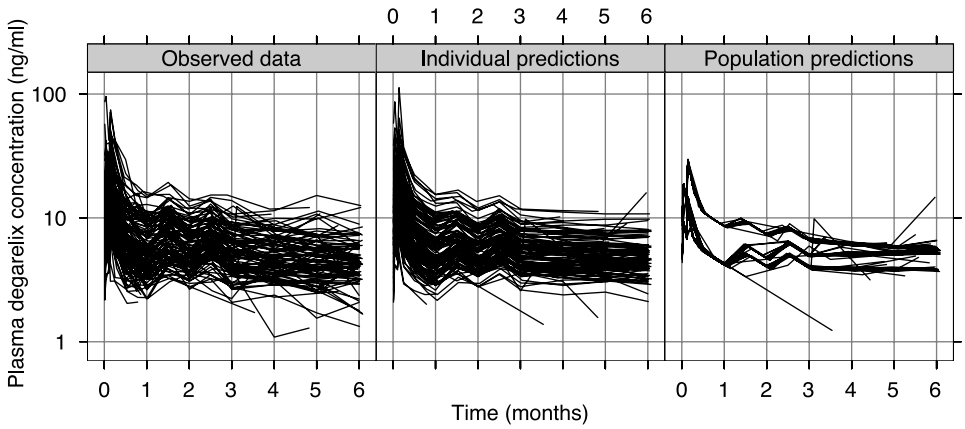

Fig. 4. Observed and SDE predicted plasma degarelix concentration-time profiles plotted on a semilogarithmic scale with each line representing data from one patient.

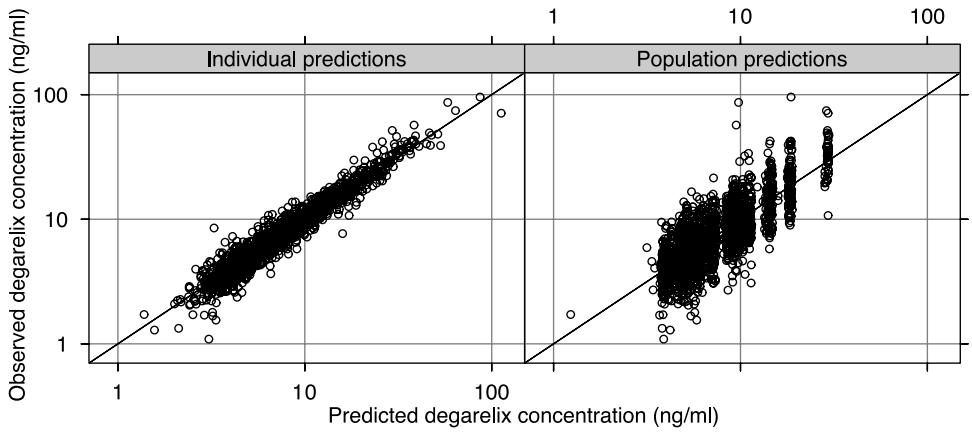

Fig. 5. Observed *vs.* individual (left) and population SDE predicted (right) plasma degarelix concentrations plotted on a double-logarithmic scale.

in Fig. 3 calculated by $\hat{y}_{j|j-1} \pm \sqrt{R_{j|j-1}}$ using Eqs. (16) and (17). The only SDE prediction outside the prediction interval is around 4 months where the observed plasma concentration is 1 ng/ml while the SDE prediction is close to 3 ng/ml, making it hard to falsify this model based on correlation structure or distribution of residuals. The subsequent SDE predictions are again all within the SDE prediction interval. The ODE predictions show systematic deviations from the observed data and exhibit a high degree of autocorrelation.

The observed degarelix concentration–time profiles for all 129 patients in the study are displayed in Fig. 4 together with the one-step individual predictions and the population predictions from the SDE model. The final PK model seems to describe the observed data well with good agreement between predicted and observed plasma concentrations as shown in Fig. 5.

DISCUSSION

The variability in population PK/PD modeling is traditionally divided into inter- and intra-individual variability. This article suggests to further decompose the intra-individual variability into measurement and dynamic noise by considering SDEs instead of ODEs in nonlinear mixed-effects models. The system noise represents structural model misspecifications or true random fluctuations within the system. By quantifying the model uncertainty in the system noise, it is possible to identify which parts of the model are misspecified, and the estimated system noise can be used as a tool for systematic model development. This would not be possible using similar approaches such as an AR(1) model (3) since it acts only on the measurement equation and not the dynamics of the system.

In the present analysis, the recursive EKF algorithm was implemented in NONMEM version VI beta for parameter estimation in SDE models. The EKF corrects for structural model deficiencies by updating the system with the individual measurements at each sampling time, unlike the ODE approach which progresses the system without including this additional information. The algorithm works only with NONMEM version VI beta because it is necessary to have access to the state estimates in $PK which is not possible with NONMEM version V (see lines 49–57 of the SDE control stream example in Appendix).

The application of SDEs for systematic model development was illustrated using a clinical PK data example. Starting from a one-compartment disposition model with first-order absorption, a PK model for SC injected degarelix was systematically developed by pinpointing model deficiencies in the absorption model. By tracking variations in the absorption half-life, it could be shown that the absorption profile could be approximated by a model with two absorption components describing the initial fast absorption phase followed by a prolonged slow release phase.

The iterative framework for systematic model development used in this analysis is summarized below (8).

- Formulate an initial SDE model with a diagonal diffusion term.
- Identify possible model misspecifications by examining the significant diffusion term estimates.
- Extend the model with state equations for the pinpointed model deficiencies.
- Track unexplained variations using the updated EKF state estimates.
- Use the information obtained by tracking variations to reformulate the model.
- Accept/reject model based on the diffusion term estimates of the reformulated model.

In a recent publication by Bayard *et al.* (17), an interacting multiple model (IMM) algorithm for tracking changing PK parameters was presented. The IMM and EKF algorithms are practically identical. The IMM algorithm, however, cannot be used for quantifying model uncertainty owing to the multiple model formulation.

The introduction of system noise in the differential equations gave a more satisfactory description of the observed clinical PK data compared to the ODE model in which the residuals were correlated. The SDE and ODE model population parameter estimates were nearly identical, however, since the system noise was relatively small in the final PK model. For other systems in which the degree of model misspecification is more dominant (e.g., nonlinear pharmacodynamic models of complicated physiological systems), the discrepancies between the SDE and ODE

parameter estimates are expected to be greater since the intra-individual variance decomposition into measurement and system noise resembles the true physiological variations in the modeled system more correctly.

In conclusion, the EKF-based algorithm was successfully implemented in NONMEM for parameter estimation in population PK/PD models described by systems of SDEs. A clinical PK data example was used to illustrate the application of SDEs for systematic model development by quantifying model uncertainty and tracking unexplained variations.

ACKNOWLEDGMENTS

The authors wish to acknowledge the help of Professor Stuart Beal (UCSF), who instructed us on how to implement the EKF algorithm in NONMEM. This work was financially supported by Ferring Pharmaceuticals A/S and Center for Information Technology, Denmark.

REFERENCES

1. R. Jelliffe, A. Schumitzky, and M. Van Guilder. Population pharmacokinetics/pharmacodynamics modeling: parametric and nonparametric methods. *Ther. Drug Monit.* **22**(17–18):354–365 (2000).
2. J. K. Lindsey, B. Jones, and P. Jarvis. Some statistical issues in modelling pharmacokinetic data. *Stat. Med.* **20**:2775–2783 (2001).
3. M. O. Karlsson, S. L. Beal, and L. B. Sheiner. Three new residual error models for population PK/PD analyses. *J. Pharmacokinet. Biopharm.* **23**:651–672 (1995).
4. C. W. Tornøe, J. L. Jacobsen, and H. Madsen. Grey-box pharmacokinetic/pharmacodynamic modelling of a euglycaemic clamp study. *J. Math. Biol.* **48**:591–604 (2004).
5. C. W. Tornøe, J. L. Jacobsen, O. Pedersen, T. Hansen, and H. Madsen. Grey-box modelling of pharmacokinetic/pharmacodynamic systems. *J. Pharmacokinet. Pharmacodyn.* **31**:401–417 (2004).
6. N. R. Kristensen, H. Madsen, and S. B. Jørgensen. Parameter estimation in stochastic grey-box models. *Automatica* **40**:225–237 (2004).
7. R. V. Overgaard, E. N. Jonsson, C. W. Tornøe, and H. Madsen. Non-linear mixed-effects models with stochastic differential equations. Implementation of an estimation algorithm. *J. Pharmacokinet. Pharmacodyn. In press* (2005).
8. N. R. Kristensen, H. Madsen, and S. B. Jørgensen. A method for systematic improvement of stochastic grey-box models. *Comput. Chem. Eng.* **28**:1431–1449 (2004).
9. S. L. Beal and L. B. Sheiner (Eds.). *NONMEM® Users Guides*. NONMEM Project Group, University of California, San Francisco, 1994.
10. A. H. Jazwinski. *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.
11. R. E. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D J. Basic Eng.* **82**:35–45 (1960).
12. R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory. *Trans. ASME, Ser. D J. Basic Eng.* **83**:95–108 (1961).
13. P. S. Maybeck. *Stochastic Models, Estimation, and Control*, Academic Press, New York, 1979.
14. V. P. Shah, K. K. Midha, S. Dighe, I. J. McGilveray, J. P. Skelly, A. Yacobi, T. Layloff, C. T. Viswanathan, C. E. Cook, and R. D. McDowall. Analytical methods validation: bioavailability, bioequivalence and pharmacokinetic studies. Conference report. *Eur. J. Drug Metab. Pharmacokinet.* **16**(4):249–255 (1991).
15. C. W. Tornøe, H. Agersø, H. A. Nielsen, H. Madsen, and E. N. Jonsson. Population pharmacokinetic modeling of a subcutaneous depot for GnRH antagonist degarelix. *Pharm. Res.* **21**:574–584 (2004).
16. H. Agersø, W. Koechling, M. Knutsson, R. Hjortkjær, and M. O. Karlsson. The dosing solution influence on the pharmacokinetics of degarelix, a new GnRH antagonist, after s.c. administration to beagle dogs. *Eur. J. Pharm. Sci.* **20**:335–340 (2003).
17. D. S. Bayard and R. W. Jelliffe. A Bayesian approach to tracking patients having changing pharmacokinetic parameters. *J. Pharmacokinet. Pharmacodyn.* **31**:75–107 (2004).

APPENDIX

NONMEM SDE Control Stream (See Table I)

```
$PROBLEM SDE 1-COMP MODEL WITH 2 ABS COMPONENTS
$INPUT ID HOUR TIME AMT CMT DV EVID MDV
$DATA SDE.dta IGNORE=@
$SUBROUTINE ADVAN6 TOL 6 DP
$MODEL COMP = (ABS1) COMP = (ABS2) COMP = (CENT)
       COMP = (P11)  COMP = (P12)  COMP = (P13)
       COMP = (P22)  COMP = (P23)  COMP = (P33)

$THETA (0,5.50)(0,100)(0,30.0)(0,900)(0,0.10,1)
       (0,0.70,1)(0,0.20)(0,1)(0,250)(0,1)
$OMEGA 0.25 0.25 0.25 0.25
$SIGMA 1 FIX

$PK
  CL   = THETA(1)
  V1   = THETA(2)
  HL1  = THETA(3)*EXP(ETA(1))
  HL2  = THETA(4)*EXP(ETA(2)) + HL1
  KA1  = LOG(2)/HL1
  KA2  = LOG(2)/HL2
  TFR  = THETA(5)
  RHO  = LOG(TFR/(1-TFR))
  FRAC = EXP(RHO+ETA(3))/(1+EXP(RHO+ETA(3)))
  TBIO = 1
  IF(CONC.EQ.20) TBIO = THETA(6)
  BIO  = TBIO*EXP(ETA(4))
  SIG  = THETA(7)
  SGW1 = THETA(8)
  SGW2 = THETA(9)
  SGW3 = THETA(10)
  F1   = FRAC*BIO
  F2   = (1-FRAC)*BIO
; Initialize state and state covariance equations
IF(NEWIND.NE.2) THEN
  AHT1 = 0
  AHT2 = 0
  AHT3 = 0
  PHT1 = 0
  PHT2 = 0
  PHT3 = 0
  PHT4 = 0
  PHT5 = 0
  PHT6 = 0
ENDIF
; Store observations for EKF update
IF(EVID.EQ.0) OBS = DV
; Store one-step predictions for EKF update
IF (EVID.NE.3) THEN
  A1 = A(1)
  A2 = A(2)
  A3 = A(3)
  A4 = A(4)
  A5 = A(5)
  A6 = A(6)
  A7 = A(7)
  A8 = A(8)
  A9 = A(9)
 ELSE
  A1 = A1
  A2 = A2
  A3 = A3
  A4 = A4
  A5 = A5
  A6 = A6
  A7 = A7
  A8 = A8
  A9 = A9
ENDIF
; EKF state update equations
IF(EVID.EQ.3) THEN
; Output cov. R_j|j-1 = C P_j|j-1 C^T + Sig Sig^T
  RVAR = A9/(A3*A3) + SIG**2
; Kalman Gain K_j = P_j|j-1 C^T R_j|j-1^-1
  K1   = A6/(A3*RVAR)
  K2   = A8/(A3*RVAR)
  K3   = A9/(A3*RVAR)
; State update eq. x_j|j=x_j|j-1+K_j (y_j-y_j|j-1)
  A1UP = A1 + K1*(OBS - LOG(A3/V1))
  A2UP = A2 + K2*(OBS - LOG(A3/V1))
  A3UP = A3 + K3*(OBS - LOG(A3/V1))
; State cov. update eq. P_j|j=P_j|j-1-K_j R_j|j-1 K_j^T
  P1UP = A4 - K1*RVAR*K1
  P2UP = A5 - K1*RVAR*K2
  P3UP = A6 - K1*RVAR*K3
  P4UP = A7 - K2*RVAR*K2
  P5UP = A8 - K2*RVAR*K3
  P6UP = A9 - K3*RVAR*K3
ENDIF
; Update states
IF(A_OFLG.EQ.1) THEN
   A_0(1) = A1UP
   A_0(2) = A2UP
   A_0(3) = A3UP
   A_0(4) = P1UP
   A_0(5) = P2UP
   A_0(6) = P3UP
   A_0(7) = P4UP
   A_0(8) = P5UP
   A_0(9) = P6UP
ENDIF

$DES
; State predicton eq. dx_t|j/dt = g(x_t|j,d,phi)
DADT(1) = -KA1*A(1)
DADT(2) = -KA2*A(2)
DADT(3) = -KA1*A(1) + KA2*A(2) - CL/V1*A(3)
; State cov. dP_t|j/dt=A_t P_t|j+P_t|j A_t^T+SGW SGW^T
DADT(4) = -2*KA1*A(4) + SGW1*SGW1
DADT(5) = -(KA1+KA2)*A(5)
DADT(6) = -(KA1+CL/V1)*A(6) + KA1*A(4) + KA2*A(5)
DADT(7) = -2*KA2*A(7) + SGW2*SGW2
DADT(8) = -(KA2+CL/V1)*A(8) + KA1*A(5) + KA2*A(7)
DADT(9) = 2*KA1*A(6)+2*KA2*A(8)-2*CL/V1*A(9)+SGW3*SGW3

$ERROR (OBS ONLY)
  IF (ICALL.EQ.4) THEN
     IF (DV.NE.0) Y = LOG(DV)
     RETURN
ENDIF
  IPRED = LOG(A(3)/V1)
  W     = SQRT(A(9)/(A(3)*A(3)) + SIG**2)
  IRES  = DV - IPRED
  IWRES = IRES/W
  Y     = IPRED + W*EPS(1)
$SIM (1)
$EST METH=1 INTER
$COV
```