Introduction to PK/PD modelling - with focus on PK and stochastic differential equations

Mortensen, Stig Bousgaard; Jónsdóttir, Anna Helga; Klim, Søren; Madsen, Henrik

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Introduction to PK/PD modelling

with focus on PK and stochastic differential equations

Stig Mortensen, Anna Helga Jónsdóttir,
Søren Klim and Henrik Madsen

November 19, 2008

DTU Informatics

DTU Informatics
Department of Informatics and Mathematical Modeling
Technical University of Denmark

Richard Petersens Plads
DTU - building 321
DK-2800 Kgs. Lyngby
Denmark

ISSN: 0601-2321

Contents

1 Introduction

The development of new medical drugs is driven by progress in many areas. These include medicine, biotechnology, new production equipment and not least, as will be the focus here, the area of mathematical and statistical modelling.

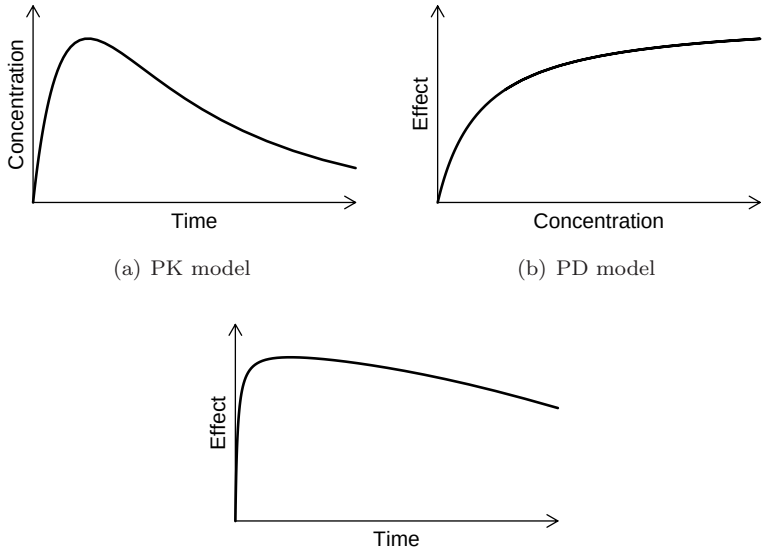
Before a new drug can move from a simple molecule in the laboratory to become a new product in the local pharmacy, there are many questions which must first be answered: Is it safe, also for patients, elderly people, pregnant women, etc.? Does the drug work? In which way should it be given to the patient? Are there any unwanted side-effects? The answer to all these questions require a long series of trials, which must be carefully planned to discover all facets of a new drug candidate.

New drug candidates are in the early development phase initially tested with animals. These tests aim to show if the drug seems to work and also to check for unwanted side effects. The next step is to move to clinical trials with humans. These trials have traditionally been separated into 4 phases. In Phase 1 the drug is given to healthy young male persons mainly to see if it is safe for humans but the sponsor (the drug company) will of course also look for indications of a positive effect. In Phase 2 the drug is given to the target patient group again mainly to show that it is still safe, but also to give indications of a positive effect. When safety has been established and the sponsor believes in the drugs potential, it will be moved to Phase 3, which generally involve the largest and most costly trials. These trials focus on proving the positive effect of the drug to the health authorities. If all goes well, the drug is approved and marketed. In some cases this will be followed by new trials, which is known as Phase 4 studies.

The series of clinical trials is not only very costly but also time consuming. The drug may easily take 10 years to get approved and marketed to the patients. Due to this both health authorities and drug companies are looking of ways to accelerate this project, while still keeping it as safe as possible. The primary target is to insure that no harmful drugs get approved, but at the same time that the good drug are not delayed unnecessarily from reaching the patients, who will benefit from them. It is in this connection that mathematical and statistical modelling has become an important tool, since it may help to give improved understanding of the outcome of clinical trials.

The scientific disciplines concerning mathematical modelling in drug development are called pharmacokinetics and pharmacodynamics, or in brief just PK/PD. In popular terms PK is often described as "what the body does to the drug" and PD as "what the drug does to the body". More specifically PK focuses on modelling how the drug passes through the body, normally by modelling concentrations in various areas of the body as a function of time, see Figure 1(a). PD aims at linking these modelled drug concentration to certain measure of effect through a PD-model. An example of a PD-model is shown

in Figure 1(b). With a combined PK/PD model it is thus possible to give a picture of the expected effect for a given dose as a function of time as illustrated in Figure 1(c).



(a) PK model

(b) PD model

(c) Combined PK/PD model

Figure 1: Illustration of PK/PD modelling.

For more thorough reading on pharmacokinetics and pharmacodynamics the books Gabrielsson and Weiner (1997) and Rowland and Tozer (1997) are suggested.

2 The ADME Model

A complicated process is initiated as soon as a drug enters the body. This process can be divided into four phases, absorption, distribution, metabolism and elimination and hence the acronym ADME.

The absorption phase describes how the drug enters the body, or more precisely how the drug enters the bloodstream. When using intravenous (iv) administration, no absorption phase is present since the drug is injected directly into the bloodstream. The whole dose can be given in one rapid injection, called a bolus dose, or by using a constant rate infusion over a certain period of time. All other dosing methods, that is when the drug is not injected directly into the bloodstream, are called extravascular dosing. Examples of such methods are injections into a muscle or fat tissue and oral dosing. Those methods have

one thing in common, they require an absorption phase since the drug needs to cross some boundaries in the body before it reaches the bloodstream. As an example when administrating a pill (oral dosing), the pill needs to dissolve and cross the gut wall before it reaches the bloodstream.

The distribution phase describes how the drug spreads through the body, into its fluids and tissues, after it has reached the bloodstream. It is in the distribution phase the drug is brought to the place of action trough the bloodstream. The time it takes for the drug to get to the place of action is very dependent on if it is easily accessible by the bloodstream. The heart is, as an example, easily accessible by the bloodstream while the bone marrow is not.

The third phase, metabolism, describes a process where the initial (parent) compound is broken into another compounds, called metabolites. The metabolites can either be inactive, therefore reducing the drugs effect on the body, or they can be active, sometimes more active than the parent compound. The liver plays a leading role in metabolism since it produces many of the enzymes used by metabolism.

The last phase, the elimination phase, describes how the compounds and their metabolites are removed from the body via excretion. Most drugs are eliminated via the kidneys with urine.

The four phases of the ADME model can be summarized as:

Absorption Drug entering the body

Distribution Drug is spreading to different areas of the body

Metabolism Drug is being changed to new chemical compounds

Elimination Drug is removed from the body

3 Fundamental concepts

Concentration is defined as amount per volume and is the most central concept in PK/PD modelling. The reason is that concentration of a drug is relatively easy to measure from a blood sample and at the same time concentration is a key factor when modelling both positive and negative effects of a drug. Concentration is calculated as

$$C = \frac{A}{V} \tag{1}$$

where A is amount of drug and V is the volume of distribution. The amount of drug is either measured as mass (mg) or in number of molecules (mol). The volume of distribution is defined as the volume which the drug has to distribute evenly into in order to reach the measured concentration in the blood, C. If the

the drug only distributes into the blood then the volume of distribution will be equal to the volume of the blood. However, often the volume of distribution will be larger than the volume of the blood. This can happen if the drug distributes into other parts of the body or if the drug is chemically bound in a way where it cannot be measured. The amount of drug is unchanged, but only a smaller part can be measured in the blood. This will result in a larger volume of distribution to reflect the lower measured concentration of the drug.

Another central issue is the timecourse of the elimination of the drug. In the most simple model one assumes that the elimination rate is proportional to the remaining amount of drug, A. The proportionality constant is CL/V, where CL is called clearance and measures the volume of blood which is cleared for drug per time. The rate of clearance is thus $CL/V \cdot A$. CL and V are sometimes referred to as micro constants and are sometimes replaced by

$$K = \frac{CL}{V} \ , \tag{2}$$

where K, the elimination rate constant, is a so called a macro constant. The choice of parameterization will be decided by the identifiability of the parameters based on the given data. In some situations it is possible to estimate both CL and V but some times they cannot be separated and it is thus only possible to estimate K.

A basic measure of the exposure of the drug is called AUC, which stands for area under the curve. AUC measures the area under the curve of concentration vs. time. An important feature of AUC is that it can be evaluated for most types of models and it can even be determined graphically based of a series of concentration measurements.

4 Compartment models

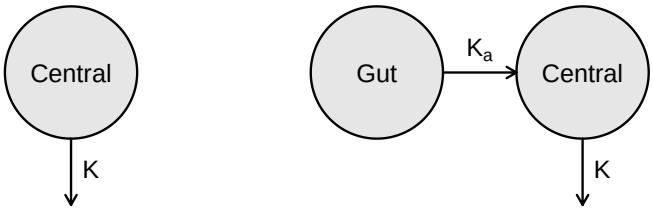
The purpose of PK modelling is mainly to describe how a drug passes through the body by modelling concentrations of the drug in different areas of the body. To measure the concentration of a drug a blood sample is usually taken and the concentration measured. Since the heart is pumping blood constantly it can be assumed that the concentration of the drug is the same within the bloodstream at a given time. This means that as soon as the drug has reached the bloodstream the concentration of the drug is the same throughout the bloodstream. However, it might not be the case that the drug spreads instantly to other parts of the body. Therefore, in order to build mathematical models to describe how the concentration changes with time the body can conveniently be divided into parts, called compartments, where the drug can be assumed to behave in the same manner. This type of modelling is called compartment modelling. The

compartment where the concentration is measured, usually the bloodstream, is of special interest and is called the central compartment.

The most basic model found is the one describing a one compartment system, which only includes the central compartment and a possible absorption compartment. This model is appropriate to use if the drug distributes to accessible areas of the body instantly. If the drug is given directly into the bloodstream (iv) the system only includes a central compartment while in the case of extravascular dosing the system should have an absorption compartment in addition to the central compartment. As an example, if the drug is administrated orally the system should include a gut compartment.

In some cases the one compartment model is not suitable for describing the system and in these cases a multi compartment model may need to be applied (see Sec. 4.2). However, in the remaining part of this report the focus will mainly be on the one compartment model (with an additional compartment if the drug is not administrated directly into the bloodstream) since many systems can be described using that model.

A compartment system can easily be visualized by drawing the compartments as circles and the connections between them with arrows indicating the direction of the flow between the compartments. A figure showing the one compartment system in case of intravenous bolus dose is shown in Figure 2(a) and in the case of oral dosing in Figure 2(b).



(a) 1-compartment system

(b) 1-compartment system with additional gut compartment

Figure 2: Illustration of compartment models.

The transfer rate of a drug from one compartment to another can usually be described using first order kinetics, meaning that the rate of change is proportional to the amount of the drug in the source compartment.

4.1 One-compartment models

The relationship between the rate of elimination and the amount of a drug in a one compartment model with first order kinetics when drug is administrated as

a bolus dose (the system shown in Figure 2(a)) can be written mathematically as:

$$\frac{dA}{dt} = -K \cdot A \tag{3}$$

where A is the amount of the drug and K is the first order elimination rate constant. K is always positive and its size controls the speed of the elimination. The differential equation can be solved resulting in a function describing the amount of the drug in the central compartment at a given time

$$A_{bolus}(t) = A_0 \exp(-K \cdot t) \tag{4}$$

where A_0 is the amount of the drug at time $t = 0$, that is the size of the given dose.

In the case of extravascular dosing other compartments needs to be added to the model. E.g. for oral dosing a extra gut compartment is often sufficient to model the absorption phase (the system shown in Figure 2(b)). Usually the rate of change in the gut compartment can be described with first order kinetics resulting in the following differential equation

$$\frac{dA_{gut}}{dt} = -K_a \cdot A_{gut} \tag{5}$$

where A_{gut} is the amount of the drug in the gut and K_a is the first order absorption constant. It is usually the case that $K_a > K$ meaning the absorption of the drug from the gut into the central compartment is faster than the elimination process. However, in some cases $K_a < K$ which is known as the flip-flop situation. The flip-flop situation is discussed further in Section 4.1.2. The change in amount in the central compartment can now be found by combining (3) describing the elimination, and (5) describing the absorption from the gut resulting in

$$\frac{dA}{dt} = \overbrace{F \cdot K_a \cdot A_{gut}}^{\text{from gut}} - \overbrace{K \cdot A}^{\text{elimination}} \tag{6}$$

where F denotes the bioavailability which is the fraction of the dose that reaches the central compartment. The differential equation can be solved resulting in an expression for the amount of drug in the central compartment for a given time which is a function of both the absorption and the elimination:

$$A_{oral}(t) = \frac{K_a F A_0}{K_a - K} \left(\exp(-K \cdot t) - \exp(-K_a \cdot t)\right) \tag{7}$$

Equations (4) and (7) can now be used to describe the amount of drug in the central compartment for a one compartment system in the case of a bolus

dose and oral dosing, respectively. Usually it is more interesting to model the concentration (C) in stead of the amount since it is the concentration of the drug in the blood that is measured. According to (1) the concentration is found by dividing the amount by the volume of distribution resulting in

$$C_{bolus}(t) = \frac{A_{bolus}(t)}{V} = \frac{A_0}{V} \cdot \exp(-K \cdot t) \tag{8}$$

in the case of a bolus dose, and

$$C_{oral}(t) = \frac{A_{oral}(t)}{V} = \frac{K_a F A_0}{V(K_a - K)} \left(\exp(-K \cdot t) - \exp(-K_a \cdot t)\right) \tag{9}$$

in the case of oral dosing.

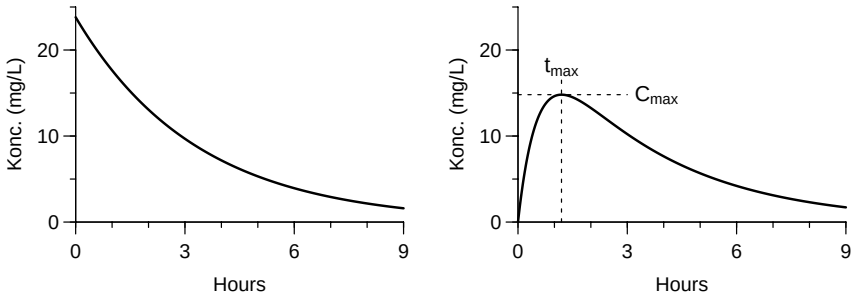
4.1.1 Case study: pain reliever

To illustrate the use of compartment modelling the drug paracetamol will be used as a case study. Paracetamol is the active substrate in a large number of pain relieving drugs on the market although its mechanism of action is still a source of debate. At correct dosages it works well against head ache and fever but at very high dosages it can cause lasting damages on the liver. For adults it is recommended to take doses of 1000mg at most 3-4 times a day and never more than 4g per day.

Paracetamol has been tested in a number of trails and it has been shown that the pharmacokinetics can be adequately described by a multi-compartment model structure, which will be introduced in Section 4.2 in further detail (see also Rawlins et al. (1977)). As a good approximation however, it can be modelled using a one-compartment model with a 1st order elimination from the blood and likewise a 1st order absorption from the stomach. Paracetamol can thus be modelled with the systems shown in Figure 2 for intravenous bolus and oral dosing. Based on Rawlins et al. (1977) the elimination rate constant can be found to $K = 0.28h^{-1}$ and the volume of distribution is $0.60L/kg$ giving $V = 42L$ for an average $70kg$ adult. The absorption is controlled by an absorption rate constant of $K_a = 1.80h^{-1}$ and the bioavailability is found to $F = 0.89$. Note that $K_a > K$ holds here, meaning that paracetamol is absorbed faster than it is eliminated.

Based on the information found for paracetamol it is possible to draw the concentration as a function of time based on (8) and (9). This is called a concentration profile and is shown for both intravenous bolus and oral dosing in Figure 3.

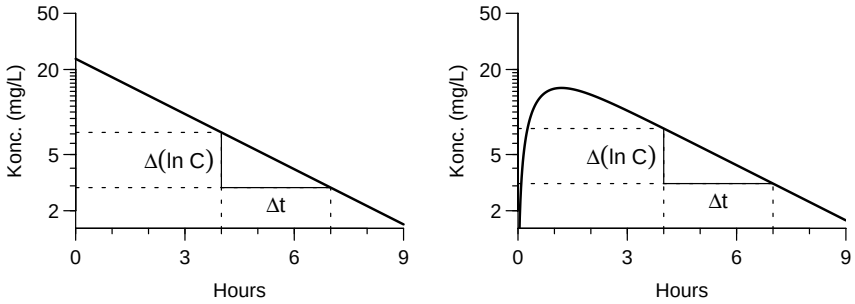
Concentration profiles are also commonly shown with concentration on a log-scale as seen in Figure 4. This makes it possible to directly read off the elimination rate constant K as the slope of the line for intravenous bolus dosing. For oral dosing K is found as the slope of the last part of the profile



(a) Intravenous (iv) bolus dose (b) Oral dose

Figure 3: Concentration profiles for dosing of $1000mg$ paracetamol.

(terminal slope) where it follows a straight line. This holds since $K_a > K$ and the absorption from the gut thus has finished so the drug is only contained in the blood as for intravenous bolus dosing.



(a) Intravenous (iv) bolus dose (b) Oral dose

Figure 4: Concentration profile on log-scale for dosing of $1000mg$ paracetamol.

In this case looking at Figure 4(b) we find approximately

$$K = -\frac{\Delta(\log C)}{\Delta t} = -\frac{\log 3 - \log 7.5}{7h - 4h} \approx 0.305h^{-1} \tag{10}$$

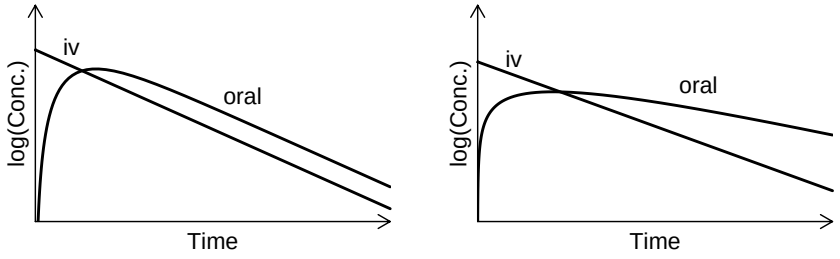
which compares well to the true value of $K = 0.28h^{-1}$.

4.1.2 Flip-flop situation

Cases where the absorption rate constant is larger than the elimination rate constant ($K_a < K$) is called a flip-flop situation. In these situations the absorption will be the so-called rate limiting step in the final phase of the elimination for oral dosing (Gabrielsson and Weiner 1997) and thus the terminal slope for oral dosing will be $-K_a$ instead of $-K$ as shown in Figure 4(b). For intravenous dosing the slope is $-K$ independent of K_a.

In order to be able to decide if it is a flip-flop situation it is necessary to perform both an intravenous bolus dosing study and an oral dosing study. If the terminal slope of the oral dosing concentration profile is parallel to the intravenous bolus concentration profile it is a normal situation, since the final elimination rates are equal. This can be known since the intravenous bolus concentration profile consists only of an elimination phase without an absorption phase.

On the other hand, if it is observed that the terminal slope for oral dosing is less steep than the intravenous bolus profile slope then the final phase must be absorption. The two situations are illustrated in Figure 5.



(a) Terminal phase for oral dosing is elimination

(b) Terminal phase for oral dosing is absorption

Figure 5: Situations with both elimination (normal) and absorption (flip-flop) as terminal phase seen as parallel and non-parallel terminal slopes respectively.

It is possible to determine the remaining rate constant (either K_a in a normal situation and K in a flip-flop situation) using the method of residuals. This is explained in detail in Gabrielsson and Weiner (1997).

4.1.3 Maximum concentration

It is often of importance to know the maximum concentration, C_{max}, of a drug in the blood and the time it takes to reach this maximum, t_{max}. For an intravenous bolus dose the maximum concentration is obtained just after the drug is injected

into the the bloodstream, that is $t_{max} = 0$ and the maximum concentration can be calculated as

$$C_{max,iv} = \frac{A_0}{V} \tag{11}$$

where A_0 is the size of the dose and V is the volume of distribution.

In the case of extravascular administration, the concentration will not peak until after a while because of the absorption step. In general, the time it takes to reach the maximum can be found by differentiating the expression for $C(t)$ for the system, with respect to t, set the derivative equal to zero and finally solve for t_{max}. In the case of oral dosing in a one compartment system following first order absorption and elimination the expression for $C(t)$ is given in (9). Differentiating this expression, setting the derivative to zero and solving for t_{max} gives

$$t_{max} = \frac{1}{K_a - K} \ln\left(\frac{K_a}{K}\right) \tag{12}$$

The resulting maximum concentration at t_{max} then becomes

$$C_{max,oral} = \frac{K_a F A_0}{V(K_a - K)} \left(\exp(-K \cdot t_{max}) - \exp(-K_a \cdot t_{max})\right) \tag{13}$$

which can be simplified to

$$C_{max,oral} = \frac{F A_0}{V} \exp(-K \cdot t_{max}) \tag{14}$$

4.1.4 Half-life of drug

An important property of a drug is its biological half-life, $t_{1/2}$. The half-life is the time it takes for reducing the amount of drug left in the body by 50%. In the case of a bolus dose, in a one-compartment system, the amount of the drug in the body at $t = t_{1/2}$ is

$$A(t_{1/2}) = A_0 \exp(-K \cdot t_{1/2}) \tag{15}$$

according to (4). By definition, half of the given amount (A_0) should be left at in the body at $t = t_{1/2}$ or

$$\frac{1}{2} A_0 = A_0 \exp(-K \cdot t_{1/2}) \tag{16}$$

which can be simplified to

$$t_{1/2} = \frac{\ln 2}{K} \tag{17}$$

4.1.5 Constant rate infusion and multiple dosing

It is often the case that it is not enough for a patient to have effect of a drug during the time span where a single pill or intravenous bolus dose is active in the body. In some cases the solution is simply to give a higher dose, but since this may cause unwanted side effects, this is not always the best way to go.

Another possibility to prolong the effect of a drug is to give it as a constant rate infusion into the vein. This is the best way to control the drug flow into the body and it is easily modelled for a constant rate R_{in} by

$$\frac{dC}{dt} = \frac{R_{in}}{V} - \frac{CL}{V} \cdot C \tag{18}$$

assuming a one-compartment model with first order elimination. The solution is given as

$$C(t) = \frac{R_{in}}{CL}\left[1 - \exp\left(-\frac{CL}{V}t\right)\right] \ . \tag{19}$$

In some cases it is more practical to approximate the constant rate infusion by taking pills with a constant time interval. This is known as multiple dosing and is related to a constant rate infusion by the equation

$$R_{in} = \frac{F \cdot A_0}{\tau} \tag{20}$$

where A_0 is the dose in each pill, F is the bioavailability (the fraction that reaches the blood) and τ is the time interval between dosing. The concentration profile for multiple dosing is a sum of single oral dosing profiles which can be written as

$$C_{MD}(t) = \sum_{n=0}^{N-1} C_{oral}(t - n\tau) \tag{21}$$

where N is the number of doses and C_{oral} is given in (9).

The system is in steady state when the elimination rate equals the infusion rate, that is when

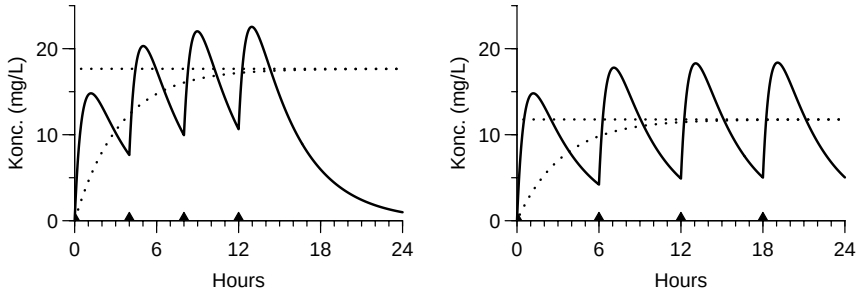
$$\frac{R_{in}}{V} = \frac{CL}{V} C_{SS} \tag{22}$$

which gives a steady state concentration

$$C_{SS} = \frac{R_{in}}{CL} \ . \tag{23}$$

It can be shown by using (19) that 90% of C_{SS} is reached after 3.32 half-lives. This result is independent of the rate of infusion which gives rise to the general rule of thumb that 90% of steady state concentration is reached after 3-4 half-lives (Gabrielsson and Weiner 1997).

Two examples of constant rate infusion and multiple dosing concentration profiles for paracetamol are shown in Figure 6.



(a) 4 hours between oral dosing

(b) 6 hours between oral dosing

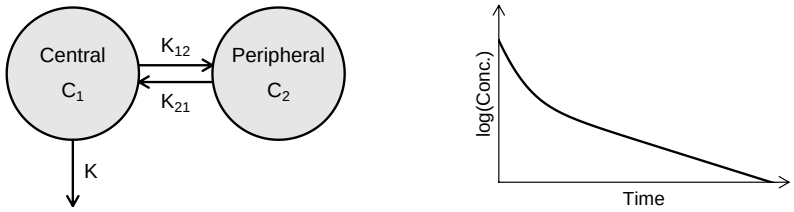
Figure 6: Multiple dosing of 4g paracetamol with 4 oral doses of 1g shown as a thick line. The dotted line is a constant rate infusion at a corresponding rate.

4.2 Multi-compartment models

It may be the case that the one-compartment model is not sufficient to describe the distribution and elimination of a drug. More complicated models where additional compartments are added to the central compartment, resulting in a multi-compartment system, should then be applied. As a consequence, the system consists of a central compartment, representing the bloodstream and rapidly equilibrated organs, one or more peripheral compartments, representing more slowly equilibrating tissues, and finally in the case of extravascular administration, an absorption compartment.

The expression for $C(t)$ for a one compartment system (intravenous dosing) only includes a singe exponential term (Equation (4)). The best way to reveal how many compartments are needed to best describe the time course of the concentration is to plot the concentration on a semi-logarithmic scale. For a multi-compartment system this will most likely look like a piecewise linear function. As a general rule of thumb one compartment is needed for each linear part that is identified. As an example a two compartment system in the case of

a bolus dose and the corresponding concentration profile are shown in Figure 7. It can be seen by looking at the figure that the concentration profile consists of two linear phases, a rapid initial and a slow terminal phase, which is the "fingerprint" of two-compartment systems.



(a) Two-compartment system

(b) Concentration profile for a two-compartment system

Figure 7: Illustration of two-compartment models.

The system shown in Figure 7(a) can be described mathematically using two differential equations, where C_1 and C_2 represents the concentration in the central compartment and the peripheral compartment respectively.

$$\frac{dC_1}{dt} = K_{21} \cdot C_2 - K_{12} \cdot C_1 - K \cdot C_1 \tag{24}$$

$$\frac{dC_2}{dt} = K_{12} \cdot C_1 - K_{21} \cdot C_2 \tag{25}$$

A solution of the differential equations, that is an expression for the concentration in the central compartment, can be written as

$$C = A \cdot \exp(-\alpha t) + B \cdot \exp(-\beta t) \tag{26}$$

Expressions for A, B, α, β given by K_{12}, K_{21} and K can be found in Gabrielsson and Weiner (1997). The half-lives of the two phases can finally be calculated as

$$t_{1/2,\alpha} = \frac{\log(2)}{\alpha} \tag{27}$$

and

$$t_{1/2,\beta} = \frac{\log(2)}{\beta} \ . \tag{28}$$

4.2.1 Case study: pain reliever continued

To illustrate the use of compartment modelling, the intake of paracetamol was modelled using a one compartment model in Section 4.1.1. It has however been shown that paracetamol concentration, after intravenous bolus dosis, follows a

bi-exponential decline indicating that a two compartment model should be used to describe the system. Based on Rawlins et al. (1977), the time course of the concentration following an intravenous bolus dose of 1000mg is given by

$$C(t) = 13.8 \cdot \exp(-2.55t) + 13.0 \cdot \exp(-0.28t) \tag{29}$$

The resulting concentration profile is shown in Figure 8.

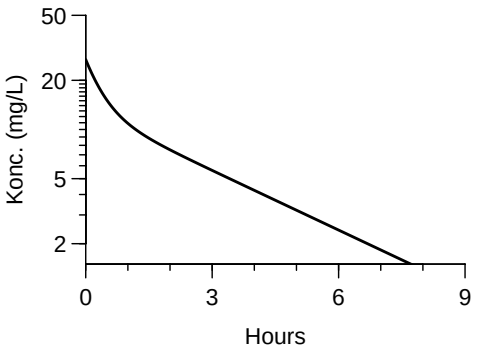


Figure 8: Concentration profile for paracetamol (1000mg intravenous bolus).

5 PD modelling

This section contains a brief introduction to pharmacodynamics (PD). In pharmacodynamics the dependent variable is not always straight forward to define. The dependent variable in pharmacokinetics is amounts or concentrations of the drug but for pharmacodynamics the dependent variable is not so obvious. How is the effect of a drug measured? Drugs against high blood pressure or fewer can be measured on the actual drop in pressure or temperature. Within epilepsy the desired effect of a drug should lower the number of seizures making the PD measurement a count. Within pain relieve the measurement can be a scale where the subject grades the pain and here the response variable will be an ordinal variable.

This wide variety of possible outcomes of a pharmacodynamic trial makes it hard to present a single methodology to handle all cases of PD modelling. This section will focus on the most simple models and only with continuous response variables.

5.1 Receptor Theory

This section includes a quick introduction into receptor theory and how it helps in the understanding of the pharmacodynamic response. The presentation of the receptor theory is highly inspired by Gabrielsson and Weiner (1997).

Before a drug molecule can give rise to a pharmacodynamic effect it needs to interact with the cells. The cell membrane is covered in different receptors each with a specific structure. The structure defines which molecules that can attach to the receptor. In engineering terms a receptor can best be described as a docking station. The drug binding to the receptor initiates a change in the structure of the receptor and thereby changing the cell membrane.

Drugs are divided into two classes according to their function on the receptor. *Agonists* initiates a structural change in the receptor thereby changing the cell resulting in a response. *Antagonists* have a different role by simply binding to the receptor but not inducing any response. By occupying the receptor it blocks the receptor for other molecules. This is also why antagonists are sometimes referred to as blockers.

The set of unoccupied receptors $[R]$ placed on the target cells that potentially can bind with drug molecules $[D]$ and their relation to the bound drug/receptor complex $[DR]$ can be stated as

$$[D] + [R] \quad \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \quad [DR] \tag{30}$$

where D denotes the drug, R the receptors and DR the bound drug/receptor complex.

The equation is based on a reversible receptor that can release the drug and be occupied by a new drug molecule. The constant k_1 indicates the rate of change from unbound to bounded and k_{-1} the other way. The binding property between the drug and the receptor determines the proportion of bound drug at equilibrium. The term affinity, which is often used to describe how good a drug binds to the receptor, is defined as

$$\text{Affinity} = \frac{k_1}{k_{-1}} = \frac{1}{K_d} \ . \tag{31}$$

The inverse of the affinity is denoted the disassociation constant K_d as shown above and is also often used.

The properties of association and disassociation determines the proportions in each state at equilibrium. It is clear that both the willingness to association/bind and the ability to stay associated/connected affects the proportions in equilibrium. One easy way to affect the amount of bound drug is simply to add more drug.

The link from the drug/receptor complex to the pharmacodynamic response can be thought of as a direct link between the receptor occupancy and the response. However often an extra state is included although hard to measure. The extra state extends the occupied state by introducing an occupied and activated state. It splits the assumption that an occupied receptor is automatically active.

$$\overbrace{[D]+[R] \quad \rightleftarrows \quad [DR]}^{\text{Inactive}} \quad \rightleftarrows \quad \overbrace{[DR^*]}^{\text{Active complex}} \tag{32}$$

Equation (32) should be interpreted using the concepts of binding and activation. Now a drug can be specified as both having a property for binding and activation of the receptor. This ability to activate the receptor is denoted *intrinsic activity* and is more difficult to measure (Gabrielsson and Weiner 1997).

5.1.1 Michaelis-Menten model

It is possible to derive a model for the relation between drug concentration and effect based on receptor theory. The effect response E is assumed proportional to the occupancy of the receptors and thus that the maximal effect is achieved if all receptors $[R_{tot}]$ are occupied. This can be stated as

$$\begin{aligned} E &= \alpha[RD] && (33)\\ E_{max} &= \alpha[R_{tot}] && (34) \end{aligned}$$

where $[R_{tot}] = [R] + [RD]$ is the total number of receptors and E_{max} is the maximal effect and α is the proportionality constant.

The steady state conditions can be stated as

$$\begin{aligned} \frac{d[RD]}{dt} &= k_1[R][D] - k_{-1}[RD] = 0 \\ \frac{[R][D]}{[RD]} &= \frac{k_{-1}}{k_1} = K_d \end{aligned} \tag{35}$$

By substituting $[R]$ with $([R_{tot}] - [RD])$

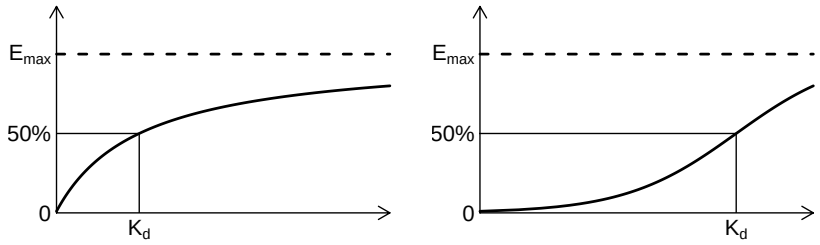
$$\begin{aligned} \frac{[D]([R_{tot}] - [RD])}{[RD]} &= K_d \\ \frac{[RD]}{[R_{tot}]} &= \frac{[D]}{[D] + K_d} \end{aligned} \tag{36}$$

Now by inserting the response assumptions

$$\begin{aligned} \frac{E/\alpha}{E_{max}/\alpha} &= \frac{[D]}{[D]+K_d} \\ \frac{E}{E_{max}} &= \frac{[D]}{[D]+K_d} \\ E &= \frac{E_{max}[D]}{[D]+K_d} \end{aligned} \tag{37}$$

The relationship in (37) is called a Michaelis-Menten relationship between drug concentration and the effect. This derivation demonstrates the motivation for the use of saturable models in pharmacodynamic modelling. Physiologically it also makes sense that at some point increasing the drug concentration will not result in a increased response.

The disassociation constant K_d determines the concentration at $1/2 \cdot E_{max}$ as can be seen from (37). An example of the model on both normal and logarithmic scale is shown in Figure 9.



(a) Response vs. Conc. (b) Response vs. log(Conc).

Figure 9: Michaelis-Menten relationship with response.

5.1.2 Commonly used PD models

The Michaelis-Menten relationship forms the basis of one of the most commonly used models to describe the relation between effect and concentration. This model is called the Sigmoid Emax model and is described by

$$E = E_0 + \frac{E_{max} C^n}{C^n + \mathrm{EC}_{50}^n} \tag{38}$$

where C is the concentration and EC_{50} is the concentration at $E_0 + 1/2 \cdot E_{max}$. The extra parameter n is included to provide a more flexible model. The model is often used with $n = 1$ and then simply called an Emax model.

In order to be able to estimate parameters in the (Sigmoid) Emax model it necessary to have estimates of the effect all the way from E_0 up to a point where maximum effect $E_0 + E_{max}$ seems to have been reached. If this is not the case it is often advisable to use a more simple model such a the linear model

$$E = E_0 + S \cdot C \tag{39}$$

or the log-linear model

$$E = m \cdot \log(C + C_0) \ . \tag{40}$$

Both of the models will in many situations be able to adequately describe the observed concentration-effect relationship. A comparison of all three models are shown in Figure 10. For the Emax model $n = 1$, $E_0 = 3$, $E_{max} = 8$ and $\mathrm{EC}_{50} = 100$ and remaining parameters S, m and C_0 are chosen so they all have the same effect at concentrations of 0 and EC_{50}.

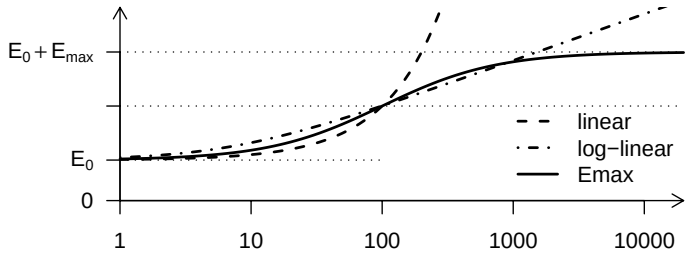


Figure 10: Comparison of standard models in PD analysis.

5.2 Modelling with effect compartments

In many situations it is not enough to directly model the response as a function of systemic concentrations in the PK model. This can happen if maximum effect is delayed compared to the maximum concentration.

This can numerically be handled by adding an additional compartment with concentration C_e representing the near cell tissue. This is called an effect compartment and is assumed to have a negligible volume. There will thus not be any mass transfer from the PK model. Furthermore a rate parameter governing the time delay from systemic concentration C_1 to near cell concentration is needed and for identifiability the same parameter is often used as elimination from the effect compartment. The model for the effect compartment is

$$dC_e/dt = k_{e1}C_1 - k_{e0}C_e \ . \tag{41}$$

The response from the drug is now modelled as a link from the effect compartment to the response. The link can be either a linear, log-linear, Michaelis-Menten or an even more complex relationship.

The response functions is often extended with the use of subject specific covariates as this can increase the accuracy of the model.

5.2.1 Case study: pain reliever continued

This example shows how a PD model can be build on top of a PK model to model the effect after an oral dose of 1000mg paracetamol. The PK-model used is the two-compartment model shown in Sec. 4.2.1. This is used in combination with a first order absorption from the gut. The PK model is defined below in (42) to (44).

$$
\begin{aligned}
dC_{gut}/dt &= -K_a C_{gut} && (42)\\
dC_1/dt &= -k_{12}C_1 + k_{21}C_2 - k_{10}C_1 + FK_aC_{gut} && (43)\\
dC_2/dt &= k_{12}C_1 - k_{21}C_2 && (44)
\end{aligned}
$$

The PD model used in this example is taken from (Gibb and Anderson 2008). It uses an effect compartment with an Emax model to model the effect. The effect compartment model is shown in (41) and is only a hypothetical compartment to introduce a delay of effect. It does thus not influence the PK model.

The compartment structure of the combined PK/PD model is shown in Figure 11. The arrows for the effect compartment are shown with dashed lines to indicate the there is no actual mass transfer.

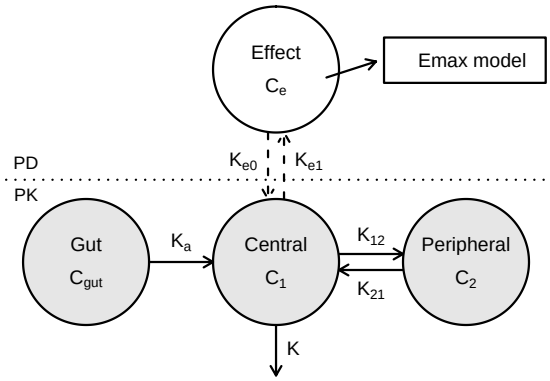


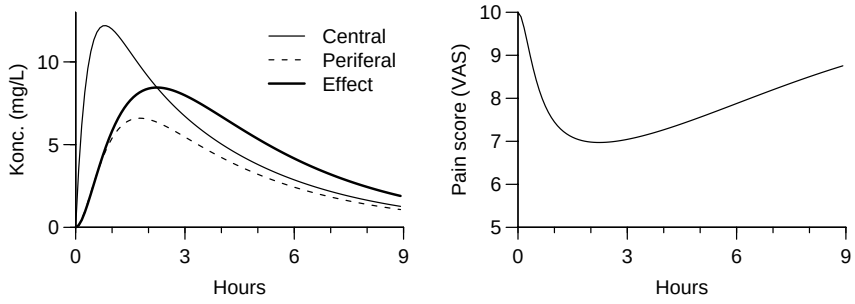
Figure 11: PK/PD model for paracetamol.

The effect is measured on a visual analogue scale (VAS) from 0-10 where reduction below 10 indicates pain relief. The Emax model for the effect is

$$\text{Effect} = 10 - \frac{E_{max} C_e}{\text{EC}_{50} + C_e}$$

with $E_{max} = 5.17$ and $\text{EC}_{50} = 9.98$mg/L. The rate constants for the effect compartment are $k_{e0} = k_{e1} = 0.83h^{-1}$ giving a half-life of 50 min. A simulation of the combined PK/PD model is shown in Figure 12.

By looking at the model for the effect in Figure 12(b) it can be seen that pain relief following a 1000mg oral dose can be expected after 0.5-1 hour and it seems to last around 6 hours or more. With the combined PK/PD model it is now also possible to give estimates of the expected effect at e.g. half or double dose without actually doing the experiment, although results from extrapolation should always be treated with care.



(a) Concentrations after 1000mg oral dose.

(b) Model of pain relief.

Figure 12: Illustration of PK/PD model for paracetamol.

6 Modelling data

To work with real life data a model of a system must be able to handle noisy observations. The models discussed until now has only been concerned with relatively simple deterministic relationships, but real systems naturally contain much more variation than suggested by these simple models. The remaining part of the text will focus on how to extend these models to include a model for the different types of variation found in data. The aim is to enable the use of a reliable statistical framework for consistent inference, simulation, prediction and control by providing a proper description of the variation into the model.

The variation found in data can be caused by a number of sources. The main sources include ordinary measurement variation and variation due to difference

between individuals, sites, occasions, etc. There may also be stochastic fluctuations of the system within an individual or approximations in the applied model, which will also lead to variation in data that must be accounted for. Further more, the model for an individual may depend on an input process (e.g. room temperature) which is sampled continuously together with the response. Measurements of this process is normally assumed to be done without measurement error, but if this is not true it will also lead to variation in the data that must be included in the model.

This section will discuss how to handle all of these different sources of variation.

6.1 Single individual

The structure of data for a single individual is

$$\boldsymbol{y}_j \ , \qquad j = 1...n \tag{45}$$

where $\boldsymbol{y}_j$ is a possibly multi-dimensional response. The sub-index is short hand notation referring to the sampling times $t_0 < t_j < t_n$.

6.1.1 Error model using ODEs

Modelling of PK for single individuals has traditionally been based on ordinary differential equations (ODEs) as for example done in a classical tool like NONMEM (Beal and Sheiner 2004). The observed deviations from the deterministic part of this model is treated as measurement error, which implies that the individual is assumed to follow the model exactly, or stated differently that the model represents the true state of the individual. This class of models can be stated as a state space model, which is written as

$$\begin{aligned} d\boldsymbol{x}_t &= \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\phi})dt & (46) \\ \boldsymbol{y}_j &= \boldsymbol{h}(\boldsymbol{x}_j, \boldsymbol{u}_j, t_j, \boldsymbol{\phi}) + \boldsymbol{e}_j & (47) \end{aligned}$$

where $\boldsymbol{x}_t$ is the state (vector) in the model and the model for the state is given by (46). The parameters in the model are denoted $\boldsymbol{\phi}$, and $\boldsymbol{u}_t$ is a vector of input variables to the system. The second equation (47) in the state space model is the measurement equation defining how the states are observed. In this case an additive error model is chosen, but this is only one of several choices. Both the measurement and state equation can be multi-dimensional.

The states can represent amounts, concentrations, time-varying parameters or other dynamic parts of a system described by a state space model. At any

point in time the state vector contains all the information about the future which is known as the Markovian property. The input variables $\boldsymbol{u}_t$ is a process influencing the system and it is observed only at measurement time points and is often assumed constant in between (known as zero order hold). A typical input variable could be e.g. body temperature or room temperature that may affect the system. The input process is assumed to be known exactly and as a consequence future values of the states in a deterministic model can be predicted without uncertainty by solving the ODE.

In the following a one-compartment ODE model for an intravenous bolus dose will be used as an example. The model is described by

$$dx_t = -kx_t \; dt \tag{48}$$

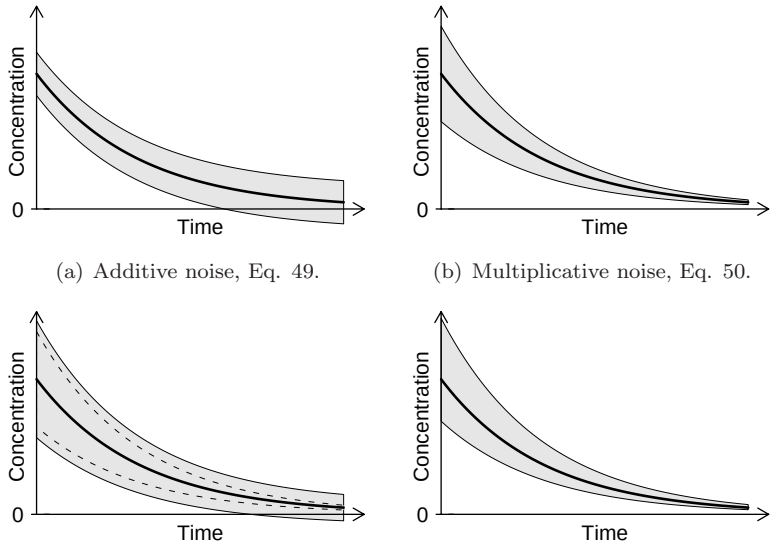
which is also shown in (3). Here x_t represents concentration in the central compartment and k is the elimination rate constant.

Traditionally there are four main types of error models for the measurement equation called additive, multiplicative, additive and multiplicative, and log-normal error. The four error models are shown in (49), (50), (51) and (52) respectively, where e_j, $e_{j,1}$ and $e_{j,2}$ are Normal IID random variables. In Figure 13 the error models are illustrated graphically.

$$\begin{aligned} y_j &= x_j + e_j && (49)\\ y_j &= x_j \cdot (1 + e_j) && (50)\\ y_j &= x_j \cdot (1 + e_{j,1}) + e_{j,2} && (51)\\ y_j &= x_j \cdot \exp(e_j) && (52) \end{aligned}$$

The optimal choice of measurement error model is very dependent on the data. The additive error model (49) is the simplest but may not always be appropriate. Measurements of concentrations are usually more uncertain for higher values, and this is not included in the additive model. However, if the measured concentrations only range over a small interval, the additive model may still be reasonable. When simulating from a model with concentrations close to zero the additive model will easily give negative values, see Figure 13(a), and this is often problematic.

The multiplicative error model (50) takes the increasing uncertainty for higher values into account. In this model the standard deviation of the residuals increases proportionally to the mean of the model. The model can still give negative values in simulations, but it is far less likely than with the additive model. In some cases it is appropriate to use a combination of the additive and multiplicative model, as specified in (51). The model still has increasing residual standard deviation with the mean but also allows for a larger uncertainty for smaller values.



(a) Additive noise, Eq. 49.

(b) Multiplicative noise, Eq. 50.

(c) Additive and multiplicative noise, Eq. 51.

(d) Log-normal noise, Eq. 52.

Figure 13: Error models illustrated with 95% prediction intervals around the model mean for an intravenous bolus dose.

The last error model shown in Figure 13 is the log-normal error model (52). This model resembles the multiplicative model as can be seen by comparing Figure 13(b) and 13(d). Being log-normal the residual distribution is asymmetric and bounded away from zero, which is an advantage for simulation as it will only give positive values. However, the distribution does not include 0, and this is a problem if the observed data has such measurements. The log-normal error model can be achieved with an additive error structure by using a log-transformation of the observations, i.e. $\log y_j = \mu^* + e_j$ where $\mu^* = \log \mu$. An effect of this is that if the model for μ^* is additive then the resulting model for μ will be multiplicative since $\mu = \exp(\mu^*)$.

6.1.2 Error model using SDEs

In most cases it is not reasonable to assume that the variation in time of the concentration for an individual follows the model exactly as it is assumed using an ODE model. As mentioned earlier there may also be some variation due to incorrect model specification, true random biological variation or uncertainty from measuring an input process which cannot be explained or included in the ODE model. A way to describe such sources of errors is to base the state space model on stochastic differential equations (SDEs) instead of ODEs. This

is called a stochastic state space model and is defined as

$$
\begin{aligned}
d\boldsymbol{x}_t &= \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\phi}_i)dt + \boldsymbol{\sigma}_\omega(\boldsymbol{u}_t, t, \boldsymbol{\phi}_i)d\boldsymbol{\omega}_t \qquad (53)\\
\boldsymbol{y}_j &= \boldsymbol{h}(\boldsymbol{x}_j, \boldsymbol{u}_j, t_j, \boldsymbol{\phi}_i) + \boldsymbol{e}_j \ . \qquad (54)
\end{aligned}
$$

The stochastic differential equation in (53) is based on the Standard Wiener process ω_t (Øksendal 1992). This process is characterized by $\omega_0 = 0$, it is almost surely continuous and it has independent normal increments with $\omega_t - \omega_s \sim N(0, t-s)$ for $0 \leq s < t$. The Wiener process can be seen as a process that has the properties expected from the limit of a discrete random walk $\sum_{i=1}^{t/\Delta t} e_i$ with $e_i \sim N(0, \Delta t)$ for $\Delta t \to 0$.

Using the Wiener process the simple one-compartment elimination model in (48) can be extended to a model based on SDEs by writing

$$
dx_t = -kx_t\ dt + \sigma_\omega d\omega_t \ . \qquad (55)
$$

The term $d\omega_t$ is an infinitesimal increment of the Wiener process and the model can thus in simple terms be described as an ordinary differential equation where the evolution in time is perturbed by normal distributed noise. The solution to (55) is

$$
x_t = x_0 e^{-kt} + \int_0^t \sigma_\omega e^{-k(t-s)} d\omega_s \qquad (56)
$$

which can be seen to be the ODE solution plus the integral of Wiener process increments with exponential weights. The process is known as an Ornstein-Uhlenbeck process with zero mean. A simulation of the process is shown in Figure 14.

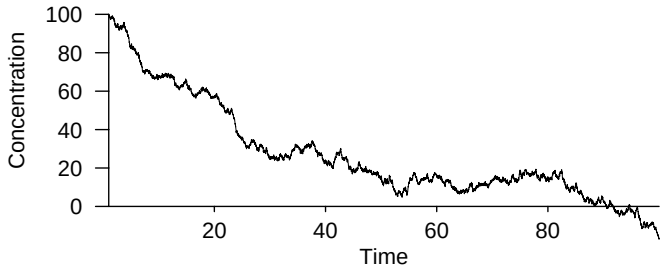


Figure 14: Simulation of simple one-compartment SDE model (55).

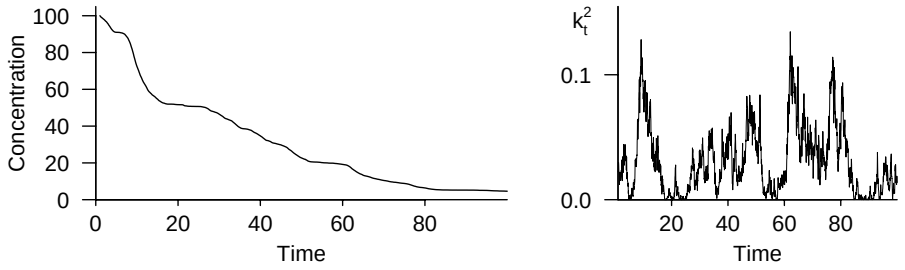
There are a few problems with the model proposed in (55). There is nothing limiting the concentration from increasing when the increments of the Wiener

process are positive. This should normally not happen in the real world as this would require a reverse elimination process. Moreover the model will fluctuate around zero when the initial dose has been eliminated, and the model will thus predict negative concentrations. In fact it can be shown that the unconditional distribution of x_t is normal with mean zero which is not suitable in a model for concentrations.

A more realistic model can be achieved by adding noise to the elimination rate constant k instead of directly to the concentration. A good first choice is to model k_t as an Ornstein-Uhlenbeck process with a non-zero mean $\bar{k}$. However, the elimination rate should never be negative, so the elimination rate used will be k_t^2. The model is given by

$$dx_t = -k_t^2 x_t \, dt \tag{57}$$
$$dk_t = -\gamma(k_t - \bar{k})dt + \sqrt{2\sigma_\omega^2 \gamma} d\omega_t \, . \tag{58}$$

It can be shown that the unconditional distribution of k_t is normal with mean $E[k_t] = \bar{k}$ and that k_t has the autocorrelation function $C_{k_t}(t) = \sigma_\omega^2 \exp(-\gamma t)$. Thus the elimination rate (k_t^2) will be χ^2-distributed with mean $E[k_t^2] = \bar{k}^2 + \sigma_\omega^2$. In Figure 15 a simulation of the model is shown. Since the elimination rate in Figure 15(b) is always non-negative, the concentration curve will be monotonely decreasing towards zero as would be expected in real life.



(a) Concentration as state 1.

(b) Elimination rate as state 2 squared.

Figure 15: Simulation of an extended one-compartment SDE model that fulfills some basic physiological constraints.

6.1.3 Case study: Advantages of using SDEs

In order to illustrate the advantages of modelling using SDEs a small simulation study will be performed.

It is assumed that the drug of interest is eliminated based on the model in (57) and (58). That is, the concentration is assumed to follow a first order

elimination with an elimination rate that varies in a non-deterministic way. The optimal model is thus naturally the SDE model that generates the data, since the non-deterministic variations of the elimination rate cannot be modelled any further. However, the example will focus on illustrating the errors that are introduced if the data is modelled using the simple ODE model shown in (48). The residual error will be assumed to be independently log-normally distributed as defined in (52). Combining this the state space model used for generating data contains two states and one response variable and is thus given by

$$\begin{aligned} dx_t &= -k_t^2 x_t \, dt & (59)\\ dk_t &= -\gamma(k_t - \bar{k})dt + \sqrt{2\sigma_\omega^2 \gamma} d\omega_t & (60)\\ \log y_j &= \log x_j + e_j & (61) \end{aligned}$$

which will in short be denoted the SDE model. The parameters used for the simulation are $\bar{k} = 0.05$, $\gamma = 0.40$ and $\sigma_\omega = 0.15$, which gives an expected elimination rate of $E[k_t^2] = 0.025$ min^{-1} and an expected half-life of 28 minutes. The measurement variation is $e_j \sim N(0, S)$ where $S = 0.20^2$. Dose is 100mg and volume of distribution is 10L giving an initial concentration of 100mg/L. The individual is sampled with 5 minute intervals up to 100 minutes giving 21 observations. The simulated data is shown in Figure 16(a).

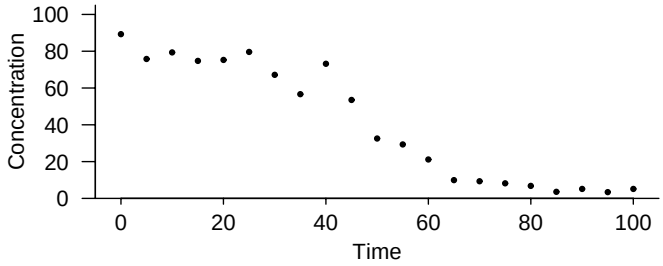
The data is modeled both using the original SDE model and an ODE model where the stochastic part is removed. The ODE model is thus a standard one-compartment model with first order elimination. The ODE model is defined as

$$\begin{aligned} dx_t &= -kx_t \, dt & (62)\\ \log y_j &= \log x_j + e_j & (63) \end{aligned}$$

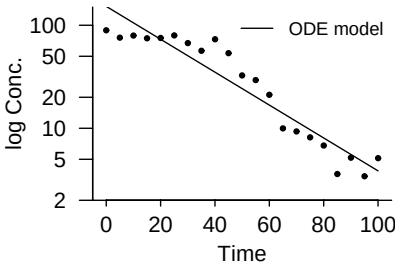
which is equivalent to the SDE model with $\sigma_\omega = 0$. Both the ODE model and SDE models are fitted using the maximum likelihood method, where the likelihood function is evaluated using the Kalman Filter. The estimation method is explained in more detail in Section 6.3. The estimate of the two states (concentration and elimination rate) in the SDE model can be found based on the estimated parameters using the so-called Kalman smoothing estimate. The results of the two model fits as given by the estimated concentration profiles are shown in Figure 16(b) and 16(c).

In order to judge the model fits the residuals for the ODE and SDE models are shown in Figure 16(d) and 16(e) and the auto-correlation functions for the residuals are shown in Figure 16(f) and 16(g).

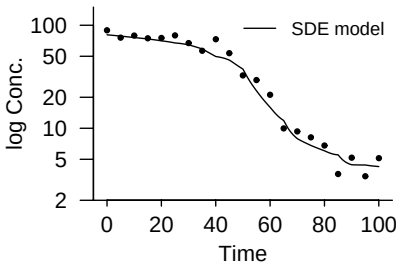
The first obvious observation of the difference between the fits is that the simple ODE model cannot capture the time varying elimination rate which gives



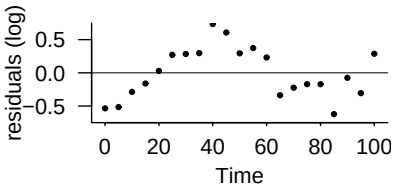
(a) Simulated data from model based on SDEs.



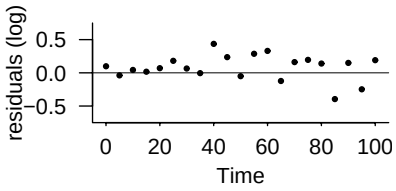
(b) Fitted ODE model on log-scale.



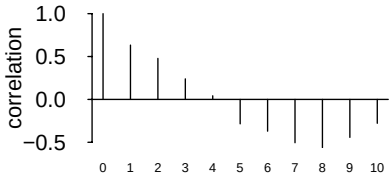
(c) Fitted SDE model on log-scale.



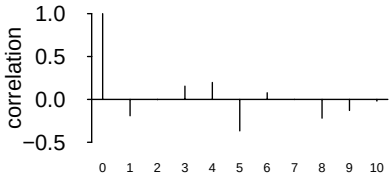
(d) Residuals in ODE model.



(e) Residuals in SDE model.



(f) ODE residual auto-correlation.



(g) SDE residual auto-correlation.

Figure 16: Comparison of fitted ODE and SDE models.

rise to persistence in time of the residuals. This is not the case for the SDE model, which assumes a stochastic elimination rate and estimates it based on the model and data. The result is that the residuals are uncorrelated in time for SDE model whereas they are strongly auto-correlated for the ODE model. This in effect falsifies the ODE model in this case, as both models are based on an assumption of uncorrelated measurement error.

The problem is also apparent from the parameter estimates them selves. The estimates of the measurement variation in the two models are $\hat{S}_{ODE} = 0.39^2$ and $\hat{S}_{SDE} = 0.22^2$ which should be compared to the true value of $S = 0.20^2$. This shows that the time variation of the elimination rate has been included in the measurement error in the ODE model, since this is the only place it allows variation to enter, and hence a wrong interpretation of the measurement error is provided by the ODE model.

It is of interest to see if the estimate in the SDE model of the time varying behavior of the elimination rate (k_t^2) is accurate. Figure 17 shows the outcome of the elimination rate process from the simulation compared to the Kalman smoothing estimate from the SDE model and also the constant estimate from the ODE model. Generally it appears that the Kalman smoothing estimate is fairly close to the true elimination rate, but the accuracy will naturally always be dependent on the kind of measurement noise, sampling rate and appropriateness of the assumed model.

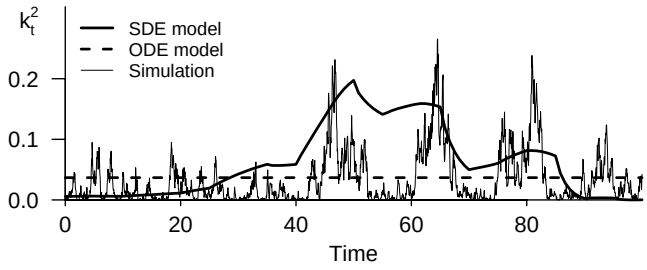


Figure 17: Comparison of estimated elimination rates.

6.1.4 Discussion of SDEs

The previous example in Section 6.1.3 has illustrated some of the issues that arise in a case where an ODE model provides an insufficient error structure. The estimate of the measurement variation was seen to be to high since it also included the stochastic variation of the system, and it resulted in auto-correlated residuals. Such situations calls for the use of an SDE model. In other situations an SDE model may be justified by a need to capture variation in the states introduced by actual stochastic behavior or a too simple deterministic part of a model or by variation caused by measurement error from the input process.

For estimation in the previous example both models were used in a maximum likelihood frame work. From a likelihood perspective, the failure of the ODE model occurs since it is not able to give a sufficient description of the 'true' likelihood function. Such an error cannot be ignored, as it in turn will invalidate e.g. likelihood ratio tests for model reduction and other classical statistical tests that may be used.

In many ways modeling using SDEs seems like an intuitive choice, since it facilitates a way to include dynamic biological variation in the model. Often however, the need for an SDE model is hidden by a sparse sampling scheme with few and distant observations, since it can be hard to detect residual autocorrelation in these situations. This may however change in the future with increasing use of modern frequent-sampling equipment within many areas, which very likely will reveal residual auto-correlation when using standard ODE models.

An example has been shown by Overgaard et al. (2007) using a PK/PD model of effects on thermo-regulation in monkeys. In this case an ODE model is shown to be insufficient, and only a model based on SDEs is able to provide a proper description of the error structure. Due to this the SDE model is able to give realistic simulations and predictions as opposed to the ODE model. Generally, in cases where the model should also be used for control purposes it is important that the model has realistic prediction properties. An example of an application for control of a biological system could be a model for controlling the insulin secretion rate in diabetic patients.

Another application of SDEs in modelling biological systems is to use it as a tool for model development. In this context SDEs can be used as a tool to extract information from data about the appropriate model in situations with a complex underlying deterministic model structure. In Kristensen et al. (2005) a framework is presented for using an SDE model to allow tracking of time variations of parameters to reveal new functional relationships. Although this use aims at improving the deterministic part of a model, it may still result a final model based on SDEs if all stochastic components cannot replaced by deterministic relationships.

6.2 Multiple individuals

A typical data set for PK/PD modelling consists of data from several individuals who all have been exposed to a similar trial. The previous section has been concerned with modelling a single individual, but it is natural to include all individuals in the same model.

A mixed-effects model allows this by assuming the same model for each individual and by further assuming that the parameters in this model can vary between individuals. The general structure for data in a mixed-effects model is

$$\boldsymbol{y}_{ij}, \quad i = 1, ..., N, \quad j = 1, ..., n_i \tag{64}$$

which is an extension of the data structure in (45). The response $\boldsymbol{y}_{ij}$ is a vector of measurements at time t_{ij} for individual i, N is the number of individuals and n_i is the number of measurements for individual i.

In a mixed-effects model the variation is split into intra-individual variation and inter-individual variation, which is modelled by a first and second stage model. The first stage model is the (stochastic) state space model and the second stage model is given by

$$\boldsymbol{\phi}_i = g(\boldsymbol{\theta}, \boldsymbol{\eta}_i, \boldsymbol{Z}_i) \tag{65}$$

where $\boldsymbol{\phi}_i$ are the individual parameters for the first stage model. The random effects $\boldsymbol{\eta}_i$ have the distribution $\boldsymbol{\eta}_i \sim N(\mathbf{0}, \boldsymbol{\Omega})$, $\boldsymbol{\theta}$ are the fixed effects (also called population parameters) and $\boldsymbol{Z}_i$ are possible covariates. The second stage model will often look like $\phi = \theta + \eta$ or $\phi = \theta \cdot \exp \eta$ giving either a normal or log-normal distribution of parameters between individuals.

6.3 Estimation

Parameter estimation in the mixed effects model is most often facilitated by the maximum likelihood method. The population likelihood function is based on the distribution functions for the first and second stage models denoted p_1 and p_2, respectively. The distribution function for the second stage model is simply the normal distribution. For the first stage model based on the stochastic state space model the likelihood function can be evaluated by using the Kalman filter (Overgaard et al. 2005). The population likelihood function for the fixed effects are found by integration over the random effects. This is given as

$$L(\boldsymbol{\theta}) \quad = \quad \prod_{i=1}^{N} \int p_1(\mathcal{Y}_{in_i} | \boldsymbol{\theta}, \boldsymbol{\eta}_i) p_2(\boldsymbol{\eta}_i | \boldsymbol{\Omega}) d\boldsymbol{\eta}_i = \prod_{i=1}^{N} \int \exp(l_i) d\boldsymbol{\eta}_i \tag{66}$$

where $\mathcal{Y}_{in_i}$ are all observations for individual i and l_i is the *a posteriori* log-likelihood function for the ith individual. The population likelihood function in (66) can not be evaluated analytically, and therefore l_i is approximated by a second-order Taylor expansion, where the expansion is made around the value $\boldsymbol{\eta}_i^*$ that maximizes $l_i(\boldsymbol{\eta}_i)$ in the value l_i^*. At this optimum the first derivative $\boldsymbol{\nabla} l_i \big|_{\boldsymbol{\eta}_i^*} = 0$ and the population likelihood function therefore reduces to

$$L(\boldsymbol{\theta}) \quad \approx \quad \prod_{i=1}^{N} \left| \frac{-\boldsymbol{\Delta} l_i^*}{2\pi} \right|^{-\frac{1}{2}} \exp(l_i^*) \tag{67}$$

as shown in Appendix A. The approximation of the 2nd derivative at the optimum $\boldsymbol{\Delta} l_i^*$ is obtained using the First-Order Conditional Estimation (FOCE) method, which is defined as

$$\boldsymbol{\Delta} l_i^* \approx -\sum_{j=1}^{n_i} \left(\boldsymbol{\nabla} \boldsymbol{\epsilon}_{ij}^T R_{i(j|j-1)}^{-1} \boldsymbol{\nabla} \boldsymbol{\epsilon}_{ij} \right) - \boldsymbol{\Omega}^{-1} \quad , \quad \boldsymbol{\nabla} \boldsymbol{\epsilon}_{ij} = \frac{\partial}{\partial \boldsymbol{\eta}_i} \boldsymbol{\epsilon}_{ij} \Big|_{\boldsymbol{\eta}_i^*} .$$

In cases where the fist stage model is non-linear as in the state space model in (53) and (54) the combined model is called a non-linear mixed effects model. When working with SDEs there are only a few software tools available that are able to estimate parameters in this class of models. These tools are listed below.

6.3.1 NONMEM

NONMEM is a software package developed at University of California, San Francisco (UCSF) for use in population PK/PD modelling (Beal and Sheiner 2004). It first appeared in 1979 and its name is an acronym for non-linear mixed effects modeling. NONMEM has become the defacto standard software tool used for PK/PD modelling as it is a very flexible tool and well tested throughout many years of development. NONMEM is however only intended for modelling based on ODEs but it is possible to make it estimate models based on SDEs as shown by Tornøe et al. (2005). This is basically done by including the Kalman filter into the model definition, but the Kalman filter has to be derived and implemented for every new model that is created, and it is thus cumbersome to work with and only feasible for simple models.

6.3.2 CTSM

CTSM is a program for performing estimation of state space models based on SDEs (Kristensen and Madsen 2003, Kristensen et al. 2004). The program is intended for single subject modelling but also handles multiple subjects based on a pooled likelihood without random effects. It has been developed at DTU Informatics. CTSM has previously been used for PK/PD modelling using SDEs in e.g. Tornøe et al. (2004a, 2004b) and Kristensen et al. (2005).

6.3.3 PSM

PSM in an acronym for Population Stochastic Modelling and is a software package developed at DTU Informatics (Mortensen et al. 2007, Klim et al. 2008). It is like NONMEM aimed at non-linear mixed effects modelling but it is focused on modelling using SDEs and as opposed to NONMEM it also directly handles a multivariate response. PSM supports a typical PK data structure with dosing

information, co-variates and also missing observations. PSM is freely available as an extension package for R, which is a free software environment for statistical computing. Instructions for download and installation in R can found at `http://www.imm.dtu.dk/psm`.

Appendix

A NLME log-likelihood function

The Non-linear mixed effects likelihood function is defined as

$$
\begin{aligned}
L(\boldsymbol{\theta}|\mathcal{Y}_{Nn_i}) &= \prod_{i=1}^{N} \int p_1(\mathcal{Y}_{in_i}|\boldsymbol{\theta},\boldsymbol{\eta}_i) p_2(\boldsymbol{\eta}_i|\boldsymbol{\Omega}) d\boldsymbol{\eta}_i && (68)\\
&= \prod_{i=1}^{N} \int L_i(\boldsymbol{\eta}_i) d\boldsymbol{\eta}_i && (69)
\end{aligned}
$$

where L_i is the individual *a posteriori* likelihood function. In most cases the integral cannot be evaluated analytically. For a general evaluation the individual *a posteriori* likelihood function can be approximated by a second order Taylor series expansion of $\log(L_i)$ around the value $\boldsymbol{\eta}_i^*$ which maximizes $\log(L_i(\boldsymbol{\eta}_i))$. Also $l_i = \log(L_i)$, $L_i^* = \exp(l_i^*) = L_i(\boldsymbol{\eta}_i^*)$, $\boldsymbol{\nabla} l_i^* = \frac{\partial}{\partial \boldsymbol{\eta}_i} l_i \Big|_{\boldsymbol{\eta}_i^*}$, $\boldsymbol{\Delta} l_i^* = \frac{\partial^2}{\partial \boldsymbol{\eta}_i \partial \boldsymbol{\eta}_i^T} l_i \Big|_{\boldsymbol{\eta}_i^*}$. It follows that

$$
\begin{aligned}
l_i(\boldsymbol{\eta}_i) &\approx l_i^* + \boldsymbol{\nabla} l_i^{*T}(\boldsymbol{\eta}_i - \boldsymbol{\eta}_i^*) + \frac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\eta}_i^*)^T \boldsymbol{\Delta} l_i^* (\boldsymbol{\eta}_i - \boldsymbol{\eta}_i^*) && (70)\\
&\approx l_i^* + \frac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\eta}_i^*)^T \boldsymbol{\Delta} l_i^* (\boldsymbol{\eta}_i - \boldsymbol{\eta}_i^*) && (71)\\
L_i(\boldsymbol{\eta}_i) &\approx L_i^* \exp\left(-\frac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\eta}_i^*)^T (-\boldsymbol{\Delta} l_i^*)(\boldsymbol{\eta}_i - \boldsymbol{\eta}_i^*) \right) && (72)
\end{aligned}
$$

since $\boldsymbol{\nabla} l_i = 0$ at $\boldsymbol{\eta}_i^*$. Based on the approximation the integral can now be evaluated by moving constants such that the integral is over a Gaussian density with mean $\boldsymbol{\eta}_i^*$ and co-variance $(-\boldsymbol{\Delta} l_i^*)^{-1}$. The result is

$$\int L_i(\boldsymbol{\eta}_i)d\boldsymbol{\eta}_i \approx \int L_i^* \exp\left(-\frac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\eta}_i^*)^T(-\boldsymbol{\Delta} l_i^*)(\boldsymbol{\eta}_i - \boldsymbol{\eta}_i^*)\right) d\boldsymbol{\eta}_i \quad (73)$$

$$\approx L_i^* \left|\frac{2\pi}{-\boldsymbol{\Delta} l_i^*}\right|^{\frac{1}{2}} \int \left|\frac{2\pi}{-\boldsymbol{\Delta} l_i^*}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\eta}_i^*)^T(-\boldsymbol{\Delta} l_i^*)(\boldsymbol{\eta}_i - \boldsymbol{\eta}_i^*)\right) d\boldsymbol{\eta}_i$$

$$\approx L_i^* \left|\frac{2\pi}{-\boldsymbol{\Delta} l_i^*}\right|^{\frac{1}{2}} \quad (74)$$

$$\approx L_i^* \left|\frac{-\boldsymbol{\Delta} l_i^*}{2\pi}\right|^{-\frac{1}{2}} \quad (75)$$

where the step in Eq. (75) is taken to avoid a matrix inversion of the Hessian. The NLME log-likelihood function can now be approximated by

$$L(\boldsymbol{\theta}|\mathcal{Y}_{Nn_i}) \approx \prod_{i=1}^{N} L_i^* \left|\frac{-\boldsymbol{\Delta} l_i^*}{2\pi}\right|^{-\frac{1}{2}} . \quad (76)$$

References

Beal, S. L. and L. B. Sheiner (2004). *NONMEM® Users Guide*. NONMEM Project Group, University of California, San Francisco.

Gabrielsson, J. and D. Weiner (1997). *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications* (Second ed.). Kristianstads Boktryckeri.

Gibb, I. A. and B. J. Anderson (2008). Paracetamol (acetaminophen) pharmacodynamics: interpreting the plasma concentration. *Arch. Dis. Child. 93*(3), pp. 241–247.

Klim, S., S. B. Mortensen, N. R. Kristensen, R. V. Overgaard, and H. Madsen (2008). Population stochastic modelling (PSM) - An R package for mixed-effects models based on stochastic differential equations. Technical report, DTU Informatics. Submitted to *Comput. Methods Programs Biomed.*

Kristensen, N. R. and H. Madsen (2003). Continuous time stochastic modelling - CTSM 2.3 Mathematics guide. Technical report, Technical University of Denmark.

Kristensen, N. R., H. Madsen, and S. H. Ingwersen (2005). Using stochastic differential equations for pk/pd model development. *J. Pharmacokinet. Pharmacodyn.. 32*, pp. 109–41.

Kristensen, N. R., H. Madsen, and S. B. Jørgensen (2004). Parameter estimation in stochastic grey-box models. *Automatica 40*, pp. 225–237.

Mortensen, S. B., S. Klim, B. Dammann, N. R. Kristensen, H. Madsen, and R. V. Overgaard (2007). A matlab framework for estimation of nlme models using stochastic differential equations: applications for estimation of insulin secretion rates. *J. Pharmacokinet. Pharmacodyn. 34*(5), pp. 623–642.

Overgaard, R. V., N. Holford, K. A. Rytved, and H. Madsen (2007). PKPD model of interleukin-21 effects on thermoregulation in monkeys–application and evaluation of stochastic differential equations. *Pharm. Res. 24*(2), pp. 298–309.

Overgaard, R. V., N. Jonsson, C. W. Tornøe, and H. Madsen. (2005). Non-linear mixed-effects models with stochastic differential equations: implementation of an estimation algorithm. *J. Pharmacokinet. Pharmacodyn. 32*(1), pp. 85–107.

Rawlins, M. D., D. B. Henderson, and A. R. Hijab (1977). Pharmacokinetics of paracetamol (acetaminophen) after intravenous and oral administration. *Eur. J. Clin. Pharmacol. 11*(4), pp. 283–286.

Rowland, M. and T. N. Tozer (1997). *Clinical Pharmacokinetics - Concepts and Applications* (Third ed.). Lippincott Williams & Wilkins.

Tornøe, C. W., J. L. Jacobsen, and H. Madsen (2004a). Grey-box pharmacokinetic/pharmacodynamic modelling of a euglycaemic clamp study. *J. Math. Biol. 48*(6), pp. 591–604.

Tornøe, C. W., J. L. Jacobsen, O. Pedersen, T. Hansen, and H. Madsen (2004b). Grey-box modelling of pharmacokinetic/pharmacodynamic systems. *J. Pharmacokinet. Pharmacodyn. 31*(5), pp. 401–417.

Tornøe, C. W., R. V. Overgaard, H. Agersø, H. A. Nielsen, H. Madsen, and E. N. Johnson (2005). Stochastic differential equations in NONMEM®: Implementation, application, and comparison with ordinary differential equations. *Pharm. Res. 22*(8), pp. 1247–1258.

Øksendal, B. (1992). *Stochastic differential equations (3rd ed.): an introduction with applications*. New York, NY, USA: Springer-Verlag New York, Inc.