Optimisation Problems in Pharmacokinetics

Jan Zmazek, University College London

May 29, 2017

1 Introduction

The term *pharmacokinetics* (PK) originates from the ancient Greek words *pharmakon*, which means "drug", and *kinetikos*, which means kinetics or "putting in motion". It is a branch of pharmacology that studies the kinetics of absorption, distribution, metabolism and excretion (ADME processes) of drugs and their corresponding pharmacologic, therapeutic, or toxic responses in a living organism.

The usage of drugs for the purpose of healing goes back to the ancient times when people discovered that many substances have therapeutic effects. However if not administered appropriately, drugs have the potential to cause harm. The medieval physician/alchemist Paracelsus has already stated back in the sixteenth century:

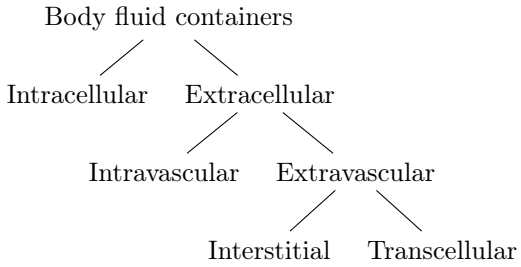
> *"Only the dose makes a thing not a poison."*

But what is the right dose? To answer this question, one must study a relationship between the administered amount of a drug and the amount of the drug measured in the body.

Drugs can be administered in different ways: orally, intravenously, etc. They can vary in volume and concentration. Depending on the type of a drug, they can have stronger or weaker effects. There are many more factors that must be taken into account in order to However, in order to predict the effects with

sufficient accuracy, various mathematical models and methods were developed based on a knowledge about the human body.

The human body and its body fluids are often divided into fluid compartments. An essential partition is between the fluids that are held inside the cells and the fluids elsewhere in the body. The extracellular fluids can then be further divided to the fluids (blood plasma and lymph) inside the vessels and the remaining extravascular fluids that consist of the fluids that surround the tissue cells, and a small amount of fluids in the transcellular compartment.



These containers are an important factor in the study of drug distribution through the body. After administration, the drug is distributed among these fluid containers according to physical laws and chemical properties. For example, a synthetic opiate drug Fentanyl is highly lipophilic, which means that the body fat, which is poorly perfused body fluid container, plays important role in the drug concentration dynamics.

The main purpose of the pharmacokinetics is determining the administered dose D and, for intravenous infusions and oral administration that are performed with multiple dosing, the dosing interval τ. This is done using the optimisation techniques on the pre-defined pharmacokinetics model, dependent on the type of a drug, method of administration and specific physiological parameters of a human body.

The optimisation is also commonly used in the model design. The most commonly used optimisation techniques used in pharmacokinetics are the methods used for the individual parameter estimation.

2 Pharmacokinetics parameters and models

There are various mathematically described pharmacokinetics models used to predict dynamics of a drug concentration. They are derived from the compartmental models and are applicable for various methods of drug administration as they take into account ADME processes, which are

1. absorption: the movement of a drug into the bloodstream,
2. distribution: the reversible transfer of a drug from one location to another within the body,
3. metabolism: the metabolic breakdown of drugs by living organisms,
4. excretion: the process of eliminating the metabolic wastes and other non-useful materials from an organism.

The goal of the pharmacokinetics is to determine the model design parameters

PK Parameter	Symbol
Administered dose	D
Dosing interval	τ

Table 1: Table of the design PK parameters.

Pharmacokinetics parameters that are normally directly measured, are

PK Parameter	Symbol
The peak concentration	$C_{\max}$
Time to reach peak concentration	$t_{\max}$
The lowest concentration reached	$C_{\min}$

Table 2: Table of the directly measured PK parameters.

Other parameters that are derived from other quantities and often simplify the models are

PK Parameter	Symbol and formula
Apparent volume of distribution	$V_{\mathrm{d}} = \frac{D}{C_0}$
Concentration of the drug	$C_0 = \frac{D}{V_{\mathrm{d}}}$
Elimination half-time	$t_{\frac{1}{2}} = \frac{\ln(2)}{k_{\mathrm{el}}}$
Elimination rate constant	$k_{\mathrm{el}} = \frac{CL}{V_{\mathrm{d}}}$
Infusion rate	$k_{\mathrm{in}} = C_{\mathrm{ss}} \cdot CL$
Area under the curve	$AUC_{0\to\infty} = \int_0^{\infty} C \,\mathrm{d}t$
	$AUC_{\tau,\mathrm{ss}} = \int_t^{t+\tau} C \,\mathrm{d}t$
clearance	$CL = V_{\mathrm{d}} k_{\mathrm{el}} = \frac{D}{AUC}$
bioavailability	$F = \frac{AUC_{\mathrm{po}} D_{\mathrm{iv}}}{AUC_{\mathrm{iv}} D_{\mathrm{po}}}$

Table 3: Table of the derived PK parameters.

By knowing fundamental pharmacokinetics parameters such as the apparent volume of drug distribution, the elimination half-life and the elimination rate, and selecting the appropriate model equation, the PK model can then be developed. A choice of the model equation solely depends on the distribution characteristics of the selected drug following its administration. The ADME processes are described by the initial value problem

$$-\frac{\mathrm{d}Y}{\mathrm{d}t} = K_0 Y^n, \qquad Y(0) = Y_0,$$

where y is some variable undergoing the change of value, K_0 some constant, n the order of the process and Y_0 the initial value of the function Y. Zero- and first-order processes are the most common in pharmacokinetics, as they are typical for living organisms. In the compartmental model analysis, mostly first-order processes are used.

The next page contains a table of PK parameter values for the Aspirin, accessible on the website of the US Food & Drug Administration[1], Clinical Pharmacology Biopharmaceutics Review.

[1] `https://www.accessdata.fda.gov/scripts/cder/daf/`

Parameter	**325-mg(1 x 325-mg)**				**650-mg (2 x 325-mg)**			
	Aspirin-PC (n=12)		**Bayer Aspirin® (n=12)**		**Aspirin-PC (n=14)**		**Bayer Aspirin® (n=14)**	
	Mean	%CV	Mean	%CV	Mean	%CV	Mean	%CV
Cmax (μg/ml)	18.0	30	16.8	25	35.1	25	34.1	17
AUC-t (min*μg/mL)	5823	39	5995	47	14465	35	15791	30
AUCinf (min*μg/mL)	6228	45	6200	45	14722	33	15934	30
T1/2 (min)	148	28	161	44	149	22	165	26
Tlag (min)	15		4		14		3	
Tmax (min) Median (max, min)	120 (75, 240)		150 (75, 240)		150 (120, 360)		180 (75, 240)	

Table 4: PK parameters of salicylic acid after administration of Aspirin-PC Capsules and Bayer Aspirin®Tablets at 325-mg (1x 325-mg) and 650-mg (2 x 325-mg) dose levels. The OSI recommended subjects were excluded.

Pharmacokinetics models are divided by the number of compartments (usually one or two compartments) and by the method of drug administration (intravenous bolus, intravenous infusion or oral administration).

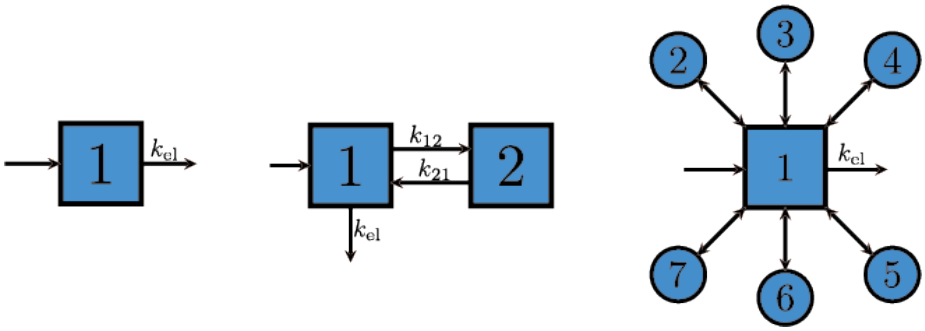


Figure 1: Single-compartment, two-compartment and multi-compartment model representations.

The method of drug administration defines the initial conditions for the compartment differential equations. These initial value problems are of the form

$$\begin{cases} \dot{X}_1 &= F_1(X_1, X_2, \ldots, X_n) \\ \dot{X}_2 &= F_2(X_1, X_2, \ldots, X_n) \\ &\vdots \\ \dot{X}_n &= F_n(X_1, X_2, \ldots, X_n) \end{cases}$$

where X_k is the amount of drug in the k-th compartment at the time t. These problems can be easily solved using Laplace transform. Laplace functions for the central compartment are of the form

$$C_p(s) = \eta(s) \cdot \xi(s),$$

where η is an input function that depends on the route of administration, and ξ is a disposition function that depends on the number of compartments.

The derivations of the input functions are listed below:

Route of Administration	Input Function η
IV Bolus	D
IV Continuous Infusion	$\frac{k_0}{s}$
IV Infusion	$\frac{k_0\left(e^{-T_0 s}-e^{-Ts}\right)}{s}$
Oral	$\frac{fDk_\mathrm{a}}{s+k_\mathrm{a}}$

Table 5: Table of input functions.

Here, T_0 is the time when the infusion starts and T when it is stopped. The derivations of the disposition functions are:

Number of Compartments	Disposition Function ξ
One	$\frac{1}{s+k_\mathrm{el}}$
Two	$\frac{s+k_{21}}{(s+\alpha_1)(s+\beta_1)}$
Three	$\frac{(s+k_{21})(s+k_{31})}{(s+\alpha_2)(s+\beta_2)(s+\gamma_2)}$

Table 6: Table of disposition functions.

These equations hold for α_1 and β_1 coefficients:

$$\begin{aligned}\alpha_1+\beta_1 &= k_\mathrm{el}+k_{12}+k_{21},\\ \alpha_1\beta_1 &= k_\mathrm{el}k_{21},\end{aligned}$$

and these for α_2 and β_2:

$$\begin{aligned}\alpha_2+\beta_2+\gamma_2 &= k_\mathrm{el}+k_{12}+k_{21}+k_{13}+k_{31},\\ \alpha_2\beta_2+\alpha_2\gamma_2+\beta_2\gamma_2 &= k_\mathrm{el}k_{21}+k_\mathrm{el}k_{31}+k_{13}k_{21}\\ &\quad +k_{12}k_{31}+k_{21}k_{31},\\ \alpha_2\beta_2\gamma_2 &= k_\mathrm{el}k_{21}k_{31}.\end{aligned}$$

α and β are hybrid first order constants for rapid dissolution phase and slow elimination phase, which depend entirely on the first order constants k_{12}, k_{21} and k_el

Using inverse Laplace transform, explicit equations for the concentration-time models are obtained. These are of the form:

- Intravenous Bolus Administration is a kind of administration when a relatively large volume of a drug is rapidly injected into a blood vessel in order to magnify a response.
  - Single-compartment model

$$C_p(t) = \frac{D}{V_\mathrm{d}} e^{-k_\mathrm{el} t}$$

  - Two-compartment model

$$C_p(t) = \frac{D}{V_\mathrm{d}} \left[\frac{\alpha - k_{21}}{\alpha - \beta} e^{-\alpha t} + \frac{k_{21} - \beta}{\alpha - \beta} e^{-\beta t} \right]$$

- Intravenous Continuous Infusion administration is the administration of a fluid into a blood vessel, usually over a prolonged period of time. In compartment models, it is represented as a constant flow of a drug into the central compartment.
  - Single-compartment model

$$C_p = \frac{k_0}{k_\mathrm{el} V_\mathrm{d}} \left[1 - e^{-k_\mathrm{el} t} \right]$$

  - Two-compartment model

$$C_p = \frac{k_0}{k_\mathrm{el} V_\mathrm{d}} \left[1 - \frac{k_\mathrm{el} - \beta}{\alpha - \beta} e^{-\alpha t} - \frac{\alpha - k_\mathrm{el}}{\alpha - \beta} e^{-\beta t} \right]$$

- Oral administration is the most common extravascular way of drug administration. It is a route where a substance is taken through the mouth. Medications come in various forms such as tablets, capsules, teas, syrups etc.
  - Single-compartment model

$$C_p(t) = \frac{f D k_\mathrm{a}}{V_\mathrm{d}(k_\mathrm{a} - k_\mathrm{el})} \left[e^{-k_\mathrm{el} t} - e^{-k_\mathrm{a} t} \right]$$

  - Two-compartment model

$$C_p(t) = A e^{-\alpha t} + B e^{-\beta t} + C e^{-\gamma t}$$

These equations are valid for a single dose. For the multiple dose administration models, there exists a simple method. Each exponential function in the model must be converted according to

$$e^{-kt} = \left(\frac{1 - e^{-nk\tau}}{1 - e^{-k\tau}}\right) e^{-kt},$$

where n is dose number and τ dosing interval. Then the new model equations are:

- Multiple IV bolus dose administration (single-compartment):

$$C_p^n(t) = \frac{D}{V_\mathrm{d}} \left(\frac{1 - e^{-nk_\mathrm{el}\tau}}{1 - e^{-nk_\mathrm{el}\tau}}\right) e^{-k_\mathrm{el}t}$$

- Multiple oral dose administration (single-compartment):

$$C_p^n(t) = \frac{fDk_\mathrm{a}}{V_\mathrm{d}(k_\mathrm{a} - k_\mathrm{el})} \left[\left(\frac{1 - e^{-nk_\mathrm{el}\tau}}{1 - e^{-k_\mathrm{el}\tau}}\right) e^{-k_\mathrm{el}t} - \left(\frac{1 - e^{-nk_\mathrm{a}\tau}}{1 - e^{-k_\mathrm{a}\tau}}\right) e^{-k_\mathrm{a}t}\right]$$

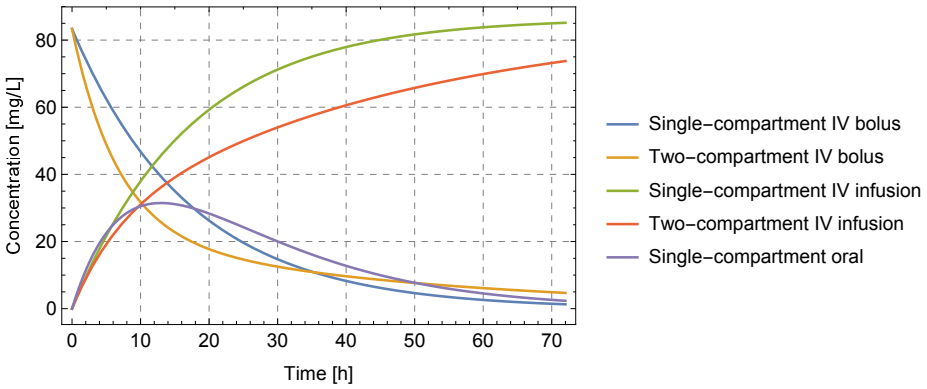


Figure 2: Time dependency of the drug concentration according to various models and ways of drug administration.

As seen in the above figure, single- and two-compartment models are similar but differ in the absorption/elimination rate. For the two-compartment models, two phases take place. The first one is distribution phase, by which a drug is distributed to the secondary (peripheral) compartment. After the equilibrium is reached, the slower elimination phase is visible.

3 PK parameters estimation methods

The most used optimisation techniques in the pharmacokinetics are methods for estimating pharmacokinetics parameters. Pharmacokinetics models are in the functional form

$$C_p(t) = f(t, \boldsymbol{p}, \boldsymbol{c}),$$

where f is some model function dependent on time t, $\boldsymbol{p} = (p_0, p_1, \ldots, p_n)$ are model parameters, and $\boldsymbol{c} = (c_0, c_1, \ldots, c_l)$ fixed constants. Model parameters are usually apparent volume of distribution V_d, elimination constant k_el, absorption constant k_a, etc. After some model is introduced, these parameters must be adjusted so that they fit the experimental data. Since the model function is in general not linear, the proper methods used for adjusting these parameters are nonlinear regression analysis methods.

There are two steps in the parameter estimation process:

1. **Minimisation methods** define the objective function which is minimised when model parameters are adjusted to best fit the data. The oldest and the most commonly used nonlinear regression methods are the Least squares methods. The simplest is the Ordinary Least Squares (OLS) method, which is a special instance of the Weighted Least Squares (WLS) method. Another commonly used method is the Maximum Likelihood Estimation (MLE). Another two (more complex) methods from the family of the least squares methods are Iteratively Reweighted Least Squares (IRLS) and Extended Least-Squares method (ELS). Relatively new minimisation method used in the PK parameters estimation is also Minimum relative entropy (MRE) method, which minimises the Shannon entropy to obtain the appropriate objective function.

2. **Searching algorithms** are algorithms that determine parameter estimations by minimising the objective function. These involve gradient iterative methods (such as most commonly used Gauss-Newton method) direct search methods (e.g. Nelder-Mead or Pattern method), which are normally less effective but appropriate for very general problems as they only depend on the evaluations of the objective function, and similar Random search methods, which are appropriate for even more general problems where objective function is not continuous.

3.1 Minimisation methods

The minimisation methods translate a problem of nonlinear regression (finding function parameters so that the model is best fitted to the measured values) to an optimisation problem. The method is often chosen based on simplicity and whether we know or not the probability distribution of the measurement errors (often normal distribution). Most common methods are presented hereinafter.

3.1.1 Least-squares methods (OLS, WLS)

Let's assume some pharmacokinetics model $C_p(t) = f(t, \boldsymbol{p})$ and some experimentally obtained data $\{u_i \pm e_i\}_{i=0}^{N}$, where u_i is the central value and e_i the measurement error. Then let's define the sum of squared residuals

$$S(\boldsymbol{p}) = \sum_{i=0}^{m} \big(C_p(t_i, \boldsymbol{p}) - u_i\big)^2 = \sum_{i=0}^{m} r_i^2.$$

We often want the error of each measurement to affect the best fit of the model. Then we rather define the weighted sum of residuals

$$S_{\mathrm{w}}(\boldsymbol{p}) = \sum_{i=0}^{m} \big(C_p(t_i, \boldsymbol{p}) - u_i\big)^2 w_i = \sum_{i=0}^{m} r_i^2 w_i,$$

where w_i are the corresponding weights. The weights can be the inverse of the measurement errors. One could also assign greater weights for the initial data in order to magnify their effect.

Let's firstly assume a linear model, $f(t) = p_0 + p_1 t$ and let's write the system of linear equations in the matrix form as

$$\boldsymbol{X}\boldsymbol{p} = \boldsymbol{u},$$

where

$$\boldsymbol{X} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{p} = \begin{bmatrix} p_0 \\ p_1 \end{bmatrix}.$$

It then follows from

$$S(\boldsymbol{p}) = ||\boldsymbol{u} - \boldsymbol{X}\boldsymbol{p}||^2 = 0$$

that

$$(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{p} = \boldsymbol{X}^T\boldsymbol{u} \iff \boldsymbol{p} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{u}.$$

In order to find the best fitting parameters for a general model, one must find the minimum of the squared residuals function. The minimum value occurs when the gradient is zero, i.e.

$$\frac{\partial S}{\partial p_j} = 2\sum_{i=0}^{m} r_i \frac{\partial r_i}{p_j} = 0 \qquad \forall j = 0, 1, \ldots, n.$$

Since the derivatives are still dependent on the time t and and PK parameters $\boldsymbol{p}$, the gradient equations do not have a closed solution. The solution must then be obtained using search algorithms.

The non-linear models can often be transformed to a linear one, such that

$$g(f(t, \boldsymbol{p})) = \alpha(\boldsymbol{p})t + \beta(\boldsymbol{p}),$$

where g is some transform function. The problem is then easily solved by performing the linear least-squares method. Inverse transformation

$$f(t, \boldsymbol{p}) = g^{-1}(\alpha t + \beta)$$

yields the model in its original form.

3.1.2 Maximum Likelihood Estimation (MLE)

This method corresponds to estimation methods in statistics. The likelihood of the true parameters being a certain value (given the data) is the same as the probability of observing the data (given some true parameter values).

For a sample of $\{x_i\}_{i=0}^m$ observations that are independent and identically distributed with (unknown) probability density function f_0 belonging to a family of distributions f, the likelihood function is defined as

$$\mathcal{L}(\boldsymbol{p}; x_1, x_2, \ldots, x_m) = f(x_1, x_2, \ldots, u_m|\boldsymbol{p}) = \prod_{i=0}^{m} f(x_i|\boldsymbol{p}),$$

where $\boldsymbol{p}$ is the vector of PK model parameters. The parameters are then found by maximising the likelihood function. More convenient form of the above equation is

$$\ln \mathcal{L}(\boldsymbol{p}; x_1, x_2, \ldots, x_m) = \sum_{i=0}^{m} \ln \left(f(x_i|\boldsymbol{p}) \right).$$

This transformation does not affect the global maximum as the natural logarithm is strictly monotonically increasing. In order to find the global maximum, one must find the solution to the gradient equations

$$\frac{\partial \ln \mathcal{L}(\boldsymbol{p}; x_1, x_2, \ldots, x_m)}{\partial p_i} = 0 \quad \forall i = 0, 1, \ldots, n.$$

Let's define then the deviation from the model by

$$e_i = u_i - f(\boldsymbol{p}, t_i) \quad \forall i = 0, 1, \ldots, m$$

and let's assume the experimental errors are distributed according the probability distribution $E(e_i)$. The parameters $\boldsymbol{p}$ are found by maximising the likelihood function

$$\mathcal{L}(\boldsymbol{p}; e_0, e_1, \ldots, e_m) = \prod_{i=0}^{m} E(e_i|\boldsymbol{p}).$$

The MLE method is equivalent to the OLE method under the assumption that the errors are normally distributed. However, this is not always the case. One could then apply various error distribution models in order to find the best fit for the parameters.

3.1.3 Iteratively reweighted least-squares method (IRLS)

Suppose that the data is distributed according to

$$u_i = f(t_i, \boldsymbol{p}) + e(t_i, \boldsymbol{p}, \zeta)^{\frac{1}{2}} \varepsilon_i \quad \forall i = 0, 1, \ldots, m,$$

so that the errors ε_i are distributed according to the variance function e. A common choice of e is the power variance function

$$e(t_i, \boldsymbol{p}, \zeta) = f(t_i, \boldsymbol{p})^{\zeta}.$$

The value ζ is appropriate to different distribution types. For example, $\zeta = 1$ corresponds to the data believed to have to come from Poisson distribution. Traditionally in pharmacokinetics, $w_i = u_i^{-1}$ or $w_i = u_i^{-2}$ are used as weights[3] for the weighted least-squares method. When a weighting scheme is applied to a series of data points, low-value data may influence the result more than appropriate. The new weighting is then used to overcome this disadvantage. Therefore, the new least-squares method is called Iteratively reweighted least-squares method. The following nonlinear IRLS procedure may then be defined:

1. Choose initial estimates $\hat{\boldsymbol{p}}$ of $\boldsymbol{p}$ and some value ζ.
2. Compute the set of weights $w = \{w_i\}$ such that $w_i = 1/f_i(\hat{\boldsymbol{p}})^{\zeta}$.
3. Solve the WLS gradient equations

$$\sum_{i=0}^{m} \frac{u_i - f(t_i, \boldsymbol{p})}{f(t_i, \hat{\boldsymbol{p}})^{\zeta}} \frac{\partial f(t_i, \boldsymbol{p})}{\partial p_j} \quad \forall j = 0, 1, \ldots, n$$

   in order to obtain the new estimates $\boldsymbol{p}$.
4. Update the parameters $\hat{\boldsymbol{p}} \leftarrow \boldsymbol{p}$ and return to step 2.

Obviously, the IRLS method with the parameter $\zeta = 0$ yields the ordinary least-squares (OLS) method.

3.1.4 Extended least-squares method (ELS)

Extended least-squares method uses the MLE method principle and can be defined when the power variance function is assumed as

$$\mathcal{L}(\boldsymbol{p}, \sigma^2) = \sum_{i=0}^{m} \frac{X_i^2(\boldsymbol{p})}{\sigma^2} + \sum_{i=0}^{m} \log(\sigma^2 f(t_i, \boldsymbol{p})^{\zeta}),$$

where

$$X_i^2(\boldsymbol{p}) = \frac{(u_i - f(t_i, \boldsymbol{p}))^2}{f(t_i, \boldsymbol{p})^{\zeta}}.$$

The minimisation of the function then yields parameter estimations.

3.1.5 Minimum relative entropy (MRE)

The relative (Shannon) entropy in information systems is based on the Kullback-Leibler divergence, which is defined for the continuous random-variable probability distributions p and q as

$$D(p||q) = \int p(x) \log \frac{p(x)}{q(x)} \, \mathrm{d}x, \qquad D(p||q) \geq 0.$$

It can be extended to the space of positive functions as

$$D^*(p||q) = \int p(x) \left(\log \frac{p(x)}{q(x)} + \frac{q(x)}{p(x)} - 1 \right) \mathrm{d}x, \qquad D^*(p||q) \geq 0 \; \forall p, q \geq 0$$

For the PK parameter estimation, we can then substitute p and q with the data-generating function u and model function f to get

$$S(\boldsymbol{p}) = -\int [u(t) \log f(t, \boldsymbol{p}) - f(t, \boldsymbol{p})] \, \mathrm{d}t.$$

Since we do not have the data-generating function in practice, the relative entropy is given in the discrete form as

$$S(\boldsymbol{p}) = -\sum_{i=0}^{m} [u_i \log f(u_i, \boldsymbol{p}) - f(u_i, \boldsymbol{p})].$$

The PK parameters are then estimated by minimising the relative entropy function.

The MRE method has not yet been used for actual pharmacokinetic analysis[4]. However, it has a firm theoretical background and high estimation power and is comparable to the ELS method.

3.2 Searching algorithms

Several minimisation methods that translate the non-linear regression problem into the minimisation problem were described in the previous section. Let's focus, however, only on solving the (weighted) least-squares method.

3.2.1 Newton's method

The initial values of the PK parameters $\boldsymbol{p_0}$ are chosen. The PK parameters vector is then shifted at each iteration according to the Newton's method

$$\boldsymbol{p}^{k+1} = \boldsymbol{p}^k - \boldsymbol{H}^{-1}\frac{\partial S}{\partial p_j},$$

where

$$\frac{\partial S}{\partial p_j} = 2\sum_{i=0}^{m} r_u \frac{\partial r_i}{\partial p_j}$$

and

$$H_{jk} = 2\sum_{i=0}^{m}\left(\frac{\partial r_i}{\partial p_j}\frac{\partial r_i}{\partial p_k} + r_i\frac{\partial^2 r_i}{\partial p_j \partial p_k}\right)$$

are gradient vector and Hessian matrix.

3.2.2 Gauss-Newton algorithm

The Gauss-Newton algorithm is similar to the Newton's method, but ignores the second-order derivative terms. With the Taylor expansion of the PK model function

$$f(t_i, \boldsymbol{p}) \approx f(t_i, \boldsymbol{p}^k) + \sum_{j=0}^{n}\frac{\partial f(t_i, \boldsymbol{p}^k)}{\partial p_j}(p_j - p_j^k) \approx f(t_i, \boldsymbol{p}^k) + \sum_{j=0}^{n} J_{ij}\Delta p_j$$

and the calculation of the residual

$$r_i = \Delta u_i - \sum_{s=0}^{n} J_{is}\Delta p_s, \quad \text{where} \quad \Delta u_i = u_i - f(t_i, \boldsymbol{p}^k),$$

the gradient equations become

$$-2\sum_{i=0}^{m} J_{ij}\left(\Delta u_i - \sum_{s=0}^{m} J_{is}\Delta p_s\right) = 0,$$

which can be arranged and written in the matrix form as

$$(\boldsymbol{J}^T\boldsymbol{J})\Delta\boldsymbol{p} = \boldsymbol{J}^T\Delta\boldsymbol{u},$$

or in the case of the weighted least-squares method

$$(\boldsymbol{J}^T\boldsymbol{W}\boldsymbol{J})\Delta\boldsymbol{p} = \boldsymbol{J}^T\boldsymbol{W}\Delta\boldsymbol{u},$$

where $\boldsymbol{W}$ is the diagonal weight matrix. The above equations are called normal equations and are the basis for the iterative Gauss-Newton algorithm. The iteration is then

$$\boldsymbol{p}^{k+1} = \boldsymbol{p}^k + \left(\boldsymbol{J}^T\boldsymbol{W}\boldsymbol{J}\right)^{-1}\boldsymbol{J}^T\boldsymbol{W}\boldsymbol{r}(\boldsymbol{p}^k).$$

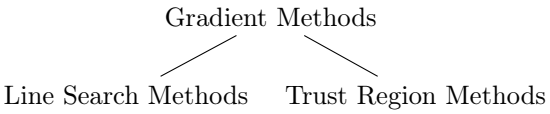
The rate of convergence of the algorithm can approach quadratic, however it is not guaranteed. If a divergence occurs, one can apply function $0 < \alpha < 1$ such that

$$\boldsymbol{p}^{k+1} = \boldsymbol{p}^k + \alpha\Delta\boldsymbol{p}^k.$$

Such approach is called "shift-cutting".

3.2.3 Other gradient methods

Gradient methods are standard iterative methods for solving optimisation problems. They consist of the Line search methods and Trust region methods.



A general **Line search algorithm** consists of 5 steps:

1. Input initial approximation $\boldsymbol{p}$
2. Compute $\Delta\boldsymbol{p} = -\boldsymbol{H}^{-1}(\boldsymbol{p})\boldsymbol{g}(\boldsymbol{p})$
3. Compute $\arg\min_\alpha \frac{1}{2}||\boldsymbol{f}(\boldsymbol{p} + \alpha\Delta\boldsymbol{p})||^2$
4. Update $\boldsymbol{p} \leftarrow \boldsymbol{p} + \alpha\Delta\boldsymbol{p}$
5. Go to step 2

There are many types of Line search algorithms that differ in their choice of the search direction $\Delta\boldsymbol{p}$ and the method used for one-dimensional optimisation along $\Delta\boldsymbol{p}$. Some of these solvers are

- Steepest Descent method, by which we choose $\boldsymbol{H}(p)$ to be an identity matrix,
- Nonlinear Conjugate Gradient, which is a generalisation of the Conjugate Gradient method,
- BFGS method, a generalisation of the Secant method to multiple dimensions,
- LBFGS, which is a limited memory approximation to the full BFGS method.

The other family of methods are **Trust region methods** which consist of these 7 steps:

1. Input initial approximation $\boldsymbol{p}$ and trust region radius μ
2. Solve
$$\arg\min_{\boldsymbol{p}} \frac{1}{2}||\boldsymbol{J}(\boldsymbol{p})\Delta\boldsymbol{p} + \boldsymbol{f}(\boldsymbol{p})||^2 \quad \text{such that} \quad ||D(\boldsymbol{p})\Delta\boldsymbol{p}||^2 \leq \mu$$
for $L \leq \boldsymbol{p} + \Delta\boldsymbol{p} \leq U$
3. Compute
$$\rho = \frac{||\boldsymbol{f}(\boldsymbol{p} + \Delta\boldsymbol{p})||^2 - ||\boldsymbol{f}(\boldsymbol{p})||^2}{||\boldsymbol{J}(\boldsymbol{p})\Delta\boldsymbol{p} + \boldsymbol{f}(\boldsymbol{p})||^2 - ||\boldsymbol{f}(\boldsymbol{p})||^2}$$
4. If $\rho > \varepsilon$ then $\boldsymbol{p} \leftarrow \boldsymbol{p} + \Delta\boldsymbol{p}$
5. If $\rho > \eta_1$ then $\mu \leftarrow 2\mu$
6. Else if $\rho > \eta_2$ then $\mu \leftarrow \frac{1}{2}\mu$
7. Go to step 2

Here, μ is the trust region radius, $D(\boldsymbol{p})$ is a matrix used to define a metric on the domain of $f(\boldsymbol{p})$ and ρ measures the quality of the step $\Delta\boldsymbol{p}$. The most common of the Trust region methods are:

- Levenberg-Marquardt (also known as the Damped least-squares method) can be considered as an interpolation between the Gauss-Newton method and the Gradient descent method. It is commonly used if a divergence occurs even after applying shift-cutting approach to Gauss-Newton method. With the Levenberg-Marquardt approach, normal equations are modified to
$$\left(\boldsymbol{J}^T\boldsymbol{W}\boldsymbol{J} + \lambda\boldsymbol{I}\right)\Delta\boldsymbol{p} = \left(\boldsymbol{J}^T\boldsymbol{W}\right)\Delta\boldsymbol{u},$$
where λ is the Marquardt parameter. The increase of λ changes the direction and the length of the shift vector to the direction of steepest descent.
- Cauchy point calculation, Dogleg method, Conjugated Gradient Steihaug's Method etc.

3.2.4 Direct search methods

Direct search methods only depend on the evaluations of the objective functions and are especially useful when the derivatives are hard (or impossible) to compute. Examples of these methods are

- Nelder-Mead method, which uses a concept of simplex (polytope of $n+1$ vertices in n dimensions) to perform a search to find the minimum. It consists of ordering, reflection, expansion, contraction and shrink steps on the simplex to search across the domain. The steps are performed if certain conditions are met.
- Pattern search method varies one theoretical parameter at a time by steps of the same magnitude. When there is no more improvement in minimising the objective function, the step size is halved. The process is repeated until the termination criterion is met.
- Golden-section search is a technique for finding minimum of a strictly unimodal objective function (having a single minimum) by successively narrowing the range of values inside which the minimum exists. The name comes from the golden ratio $\varphi = \frac{1+\sqrt{5}}{2} \approx 1.618$.

The algorithm for the Golden-section search is the following:

1. Define the initial interval $[a, b]$ and compute $f(a)$ and $f(b)$
2. Compute $c = b - \frac{b-a}{\varphi}$, $d = a + \frac{b-a}{\varphi}$, $f(c)$ and $f(d)$
3. If $f(c) < f(d)$ then update:
   (a) $(b, f(b)) \leftarrow (d, f(d))$, $(d, f(d)) \leftarrow (c, f(c))$ and
   (b) $c \leftarrow b + \frac{a-b}{\varphi}$ and compute $f(c)$
4. Else, update:
   (a) $(a, f(a)) \leftarrow (c, f(c))$, $(c, f(c)) \leftarrow (d, f(d))$
   (b) $d = a + \frac{b-a}{\varphi}$ and compute $f(d)$
5. Go to step 2

3.2.5 Random search methods

Similarly as Direct search methods, Random search methods also don't require gradient of the problem to be optimised. It can be used on functions that are not continuous or differentiable. A general algorithm consists of 5 steps:

1. Input initial approximation $\boldsymbol{p}$
2. If a termination criterion is met, terminate
3. Obtain a new random position $\boldsymbol{p'}$ from a given hypersphere of a given radius surrounding the current approximation $\boldsymbol{p}$
4. If $\boldsymbol{f}(\boldsymbol{p'}) < \boldsymbol{f}(\boldsymbol{p})$ set $\boldsymbol{p} \leftarrow \boldsymbol{p'}$
5. Go to step 2

Various variants of the Random search methods are in use, based on the setting of the radius (step 3):

- Fixed Step Size Random Search (FSSRS) by which a hypersphere is of fixed radius
- Optimum Step Size Random Search (OSSRS)
- Adaptive Step Size Random Search (ASSRS)
- Optimized Relative Step Size Random Search (ORSSRS)

3.3 Solving linear systems

When performing an optimisation algorithms, one must often solve linear systems of equations in the form of

$$A\boldsymbol{x} = \boldsymbol{b}.$$

There are various methods for solving this system. The choice of the method depends on the type of the matrix A and the size of the system. Direct solvers are more appropriate for smaller systems. These methods are e.g.

- Gaussian elimination,
- LU decomposition,
- QR decomposition,
- Cholesky decomposition, etc.

There exist special faster variants for sparse systems and systems of special type, e.g. upper or lower triangular, or tridiagonal system.

Standard procedure for solving very large systems consists of two steps. Firstly, a preconditioner (matrix P) is applied, so that the condition number of the problem

$$P^{-1}A\boldsymbol{x} = P^{-1}\boldsymbol{b}$$

is lower than of the original problem. Preconditioners are e.g.

- Jacobi method: $P = \mathrm{diag}(A)$,
- Gauss-Seidel method: $P = D+E$, where $A = D+E+F$ and D is diagonal matrix, E lower diagonal matrix and F upper diagonal matrix,
- Incomplete LU decomposition: $P = LU$, where L and U are lower and upper triangular matrices such that $A \approx LU$,
- Inclomplete Cholesky decomposition: $P = KK^*$, where K is lower triangular matrix,
- Multigrid methods, etc.

Secondly, the preconditioned system is solved using some iterative method, most commonly Krylov subspace solver. Three of such solvers are

- Generalised Minimal Residual method (GMRES),
- Minimal Residual method (MINRES),
- Conjugate Gradient method (CG)

These methods generate a sequence of improving approximate solutions $\boldsymbol{x}^k$. Their accuracy can be arbitrarily improved for the cost of the number of iterations.

4 Optimisation of administered dose and dosing interval

4.1 Basic principles

Knowing the PK model and measured PK parameters for a specific drug, the goal of pharmacokinetics is then to adjust design PK parameters (D and τ) in order to obtain optimal drug concentration in blood, in terms of being in the therapeutic range. The concentration can be almost perfectly controlled by the intravenous infusion. The initial, loading dose is administrated to rapidly reach the desired concentration. The loading dose is calculated by

$$D_l = \frac{C_p V_d}{FS},$$

where S is the active percentage of a drug and F is bioavailability ($F = 1$ for the intravenous infusion). However, this method is very impractical for common (everyday) use, as it involves direct injection of medication into a vein through an intravenous line, needle, or catheter. Oral administration is much more practical.

A quantity that describes a relationship between the dose of a drug that causes lethal or toxic effects with the dose that causes therapeutic effects of a drug is called therapeutic index (TI) or therapeutic ratio and is for humans defined as

$$TI = \frac{TD_{50}}{ED_{50}},$$

where TD_{50} refers to the toxic dose in 50 % of subjects and ED_{50} to efficacious dose in 50 % of subjects. With higher therapeutic index (which is preferable to a lower one) enables more freedom in choosing the dosing interval τ, which is very desirable.

While therapeutic index generally describes whether concentrations are allowed to vary, more appropriate quantity is therapeutic window, which is a range of doses that produce a therapeutic response without causing any significant adverse effect in patients. It is defined as a range of concentrations between minimum effective concentration (MEC) and the minimum toxic concentration (MTC).

4.2 Proposal of optimisation technique

Optimisation techniques are initialised by defining an objective function S to be minimised. In the case of e.g. Least-squares methods, this function is $S(\boldsymbol{p}) = \sum_{i=0}^{m} r_i^2$, where r_i are residuals. For the optimisation of the administered dose D and dosing interval τ we shall define similar objective function.

Let's define the desired properties of the weight function. It shall be dependent of two quantities; a blood concentration of the drug $C_p(t)$ (which implies the dose D) and dosing interval τ. The blood concentration is ideally in the middle point of the therapeutic window. It must never, however, reach and exceed boundary values of the therapeutic window. The two conditions then imply

$$W_C(C_{\text{ideal}}) = 0 \qquad \text{and} \qquad S_C\left(\left|C_{\text{ideal}} - C_p\right| \geq \frac{C_{\text{range}}}{2}\right) = \infty,$$

where W_C is the concentration component of the weight function. In between the fixed points, the function's derivative should be (strictly) monotonically increasing, i.e.

$$W_C'(C_p) > 0.$$

It is desirable for the dosing interval to be longer, so that the patients don't have to recall on taking the drug as often throughout the day. Then the following properties of the dosing interval part of the weight function W_τ should hold:

$$W_\tau(0) = \infty \qquad \text{and} \quad S_\tau(\infty) = 0.$$

To make the weight function as simple as possible, let's define the concentration part of the weight function as

$$W_C(C_p) = \begin{cases} \infty & C_p < C_{\min}, \\ -A\frac{\left(C_p - \frac{C_{\max}-C_{\min}}{2}\right)^2}{(C_p - C_{\max})(C_p - C_{\min})} & C_{\min} \leq C_p \leq C_{\max}, \\ \infty & C_p > C_{\max}, \end{cases}$$

where A is some adjustable parameter. Similarly, we can define the dosing interval part of the weight function as

$$S_\tau(\tau) = \begin{cases} \infty & \tau < 0, \\ B/\tau & \tau \geq 0, \end{cases}$$

where B is again some adjustable parameter. The concentration part of the weight function is shown in the figure below:

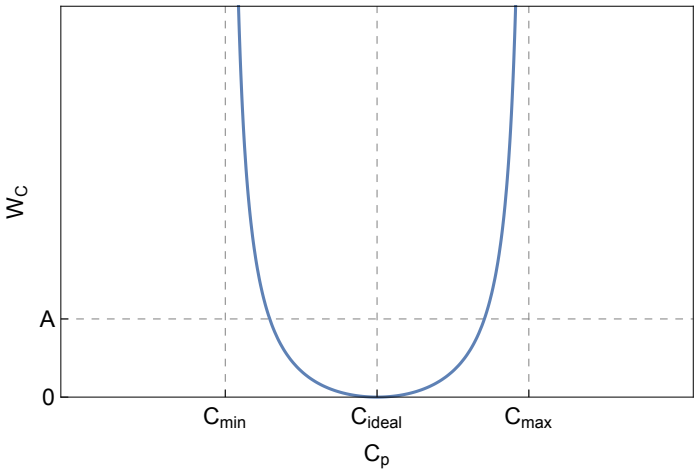


Figure 3: Concentration part of the weight function.

Similarly, the dosing interval part of the objective function is shown below:

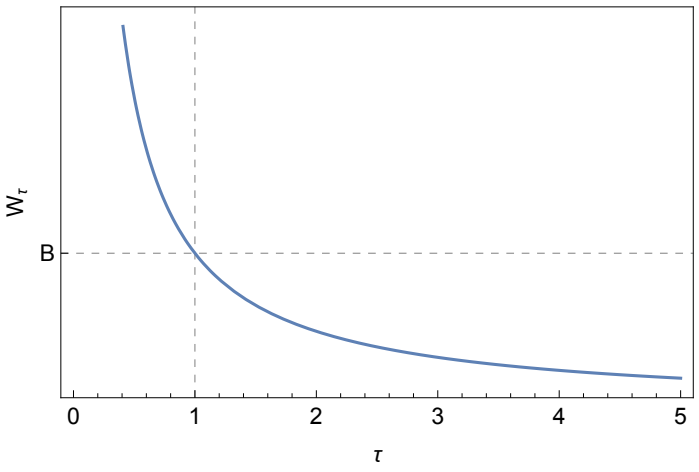


Figure 4: Dosing interval part of the weight function.

The complete weight function can then be defined as

$$W(C_p, \tau) = W_C(C_p) \cdot W_\tau(\tau).$$

Let's firstly assume that correct loading dose D_l is administered every time a patients begins a treatment. With this condition, we can constrain the optimisation problem to just one period of the dosing interval. The objective function can then be defined by

$$S(C_p, \tau) = W_\tau(\tau) \int_{t_0}^{t_0+\tau} W_C[C_p(t)] \, \mathrm{d}t,$$

where t_0 is the initial time (which can be moved due to the absorption lag).

In the case when the loading dose D_l is not administered, the objective function translates to

$$S(C_p, \tau) = W_\tau \int_{t_1}^{N\tau} W_C[C_p(t)] \, \mathrm{d}t,$$

where N is number of doses (e.g. pills, dependent on the dosage interval and dose, i.e. $N = N(D, \tau)$) and t_1 the time to reach $C_{\min}$ (which must be before the second administration).

5 Conclusion

The aim of this project was to closely connect the fields of pharmacokinetics and mathematical optimisation. Pharmacokinetics models describe time dependence of the drug concentration in blood based on the model pharmacokinetics parameters. The first numerical optimisation challenge occurring in pharmacokinetics is non-linear regression fitting of the measured data to the models in order to estimate these PK parameters.

PK parameters estimation is done in two steps. The first step is defining some objective function to be minimised. This is done using minimisation methods, such as e.g. most common Ordinary least-squares method (generalised to Weighted least-squares method, Iteratively reweighted least-squares method and Extended least-squares method), Maximum likelihood estimation method or Minimum relative entropy method. The second step is finding PK parameters that minimise the objective function. This is done using searching algorithms. If the function is differentiable and the partial derivatives are easily computed, then the efficient algorithms are gradient methods, such as Newton's method, Gauss-Newton algorithm, Nonlinear conjugate gradient method, Levenberg-Marquardt method etc. In contrast, if the partial derivatives are hard to compute, or if we don't require the solution to converge fast, then the Direct search methods, which depend only on the evaluations of the objective function, may be used. Examples of such methods are Nelder-Mead method, Pattern search method and Golden-section search method. Even more primitive are Random search methods. They don't require gradient of the problem to be optimised, and can also be used on functions that are not continuous. Such methods require a random number generator and find the minimum of the objective function by setting the parameters randomly until the termination criterion is met.

The second numerical optimisation challenge in pharmacokinetics is determining the design PK parameters D and τ. In practice, this is mostly done experimentally. However, I have presented basic optimisation theory in nonlinear programming and proposed an optimisation technique to determine the model parameters. Defining the weight function (often called Loss function), I have derived an optimisation problem and the objective function to be minimised. The technique involves calculating the integral (area under the curve) of the weighted time-dependent concentration function over the time interval of

the drug treatment.

I believe there is a lot of potential for research in applying mathematical principles to sciences such as pharmacology. This project could be further extended e.g. by studying optimisation techniques in specific drug treatments, or for personalised (precision) medicine, which a growing field in the modern society. To obtain the most effective results in treatment, an application for computing the precise dosing interval (based on individual human's characteristics such as weight, height, metabolism etc.) could be developed for such purposes and could potentially increase general health of the population.

References

[1] Sunil S Jambhekar and Philip J Breen. *Basic Pharmacokinetics*. Pharmaceutical Press 2009. ISBN 978 0 85369 772 5

[2] Edward H. Kerns and Li Di. *Drug-like Properties: Concepts, Structure Design and Methods: from ADME to Toxicity Optimization*. 2008, Elsevier Inc. ISBN: 978-0-1236-9520-8

[3] Takashi Amisaki and Shinto Eguchi. *A Comparison of Methods for Estimating Individual Pharmacokinetic Parameters*. Journal of Pharmacokinetics and Biopharmaceutics, Vol. 27, No, 1, 1999.

[4] Takashi Amisaki and Shinto Eguchi. *Pharmacokinetic Parameter Estimations by Minimum Relative Entropy Method.*

[5] Farhad Khorasheh, Amir Mahbod Ahmadi, and Abbas Gerayeli. *Application of Direct Search Optimization for Pharmacokinetic Parameter Estimation*. J Pharm Pharmaceut Sci (`www.ualberta.ca/ csps`) 2 (3):92-98, 1999

[6] U.S. Food & Drug Administration (FDA)
`https://www.accessdata.fda.gov/scripts/cder/daf/`.
May 2017.