Journal of Pharmacokinetics and Pharmacodynamics (2019) 46:43–52
https://doi.org/10.1007/s10928-018-09616-1

ORIGINAL PAPER

Application of new measures of nonlinearity to parameter estimation and simulations in individual pharmacokinetic analyses

Leonid Khinkis[1] · **Milburn Crotzer**[1]

Received: 29 September 2017 / Accepted: 26 December 2018 / Published online: 3 January 2019

Abstract

Since simulation studies are widely used in pharmacokinetics, it is necessary to ascertain their validity. An important, well-documented, concern that may negatively impact the validity of estimated parameters in pharmacokinetic models is existence of multiple minima of the criterion used in estimation. A presence of multiple minima causes instability of the estimates, dependence of the parameter estimates on the initial values, and bimodal (or, generally, multimodal) distributions of the estimated parameters. This paper offers a method to identify when these issues may occur by applying two new measures of nonlinearity.

Keywords Pharmacokinetic models · Nonlinear parameter estimation · Simulation · Compartmental models · Michaelis–Menten model

Introduction

The theoretical and computational challenges in the least squares estimation of parameters in nonlinear regression models are well documented in statistical literature. A major feature that distinguishes nonlinear regression models from their linear counterparts is a potential existence of multiple minima of the criterion used in estimation [1, 2]. The ordinary least squares (OLS) criterion is typically used under assumption of normal and homogeneous errors in the model. It is replaced with the weighted least squares (WLS) criterion when the errors are heterogeneous. In linear models, the least squares criterion has a unique global minimum, and there are no other local minima. In nonlinear models, multiple minima could occur causing instability of the estimates, dependence of the parameter estimates on the initial values, and bimodal (or, generally, multimodal) distributions of the estimated parameters. This phenomenon results in significant computational difficulties in estimation of the parameters and leads to incorrect confidence regions and confidence intervals. Additionally, it raises a natural question whether the final iteration of any sum-of-squares minimization routine results in a true least-squares estimate. The question represents a serious theoretical problem since existing software finds a local, rather than a global, minimum, meaning that it needs to be studied further if any such minimum is also a global one. Cobelli and Salvan [3] noted that the initial guess in the parameter space may materially affect the parameter estimates. The relevant recommendation often cited in the literature is to use a number of different starting points with any minimization routine. While this technique increases the likelihood of finding true parameter estimate, it still does not guarantee that such estimate will be found. Additionally, it may significantly slow down execution of a minimization procedure, which is particularly relevant in simulation studies. Ritz and Streibig discuss this in [4] where self-starter functions are offered as a viable solution. Chavent [5] and Demidenko [1, 6] obtained criteria (sufficient conditions) ensuring that a local minimizer of the residual sum of squares is also a global minimizer. In the absence of a general global function minimization algorithm, sufficient conditions, like in the work cited above, may help to identify a true parameter estimate. The limitations of these criteria are illustrated in [7]. Pronzato and Walter [8] advocate usage of D-optimal

---

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10928-018-09616-1) contains supplementary material, which is available to authorized users.

✉ Leonid Khinkis
khinkis@canisius.edu

[1] Canisius College, Buffalo, NY, USA

designs as the way to avoid multiple local minimizers. Specific methods aiming to solve the posed problem in a narrower class of nonlinear models, such as models with a single nonlinear parameter, have been proposed [2, 9].

Since the effects outlined above are specific to nonlinear models and do not present themselves in their linear counterparts, the commonly held argument is that these effects will be negligible if a nonlinear model is "close to linear". Such argument is often used to justify some methods based on linear (and higher order, if needed) approximations. While it is intuitively clear what the notion of "closeness" entails, any useful quantitative measure of nonlinearity needs, in our opinion, not just quantify the degree of nonlinearity but also help to explain the relationship between nonlinearity and statistical properties of a model. A number of attempts have been made in the statistical literature to quantify the degree of nonlinearity of a statistical model . However, none of these is universally applicable. Ratkowsky [10, 11] identifies a nonlinear model as "close to linear" if its estimated parameters are unbiased and their distribution is close to normal. SAS/STAT(R) 14.1 User's Guide [12] references the cited Ratkowsky's work as well as Box's bias [13] and Hougaard's skewness [14] as measures of departure from such "closeness". Reparameterization is often recommended in the literature and may improve distributional properties of the estimated parameters. However, it does not affect the concern outlined above since reparameterization may reduce parametric curvature but not intrinsic curvature.

Beale [15] and Bates and Watts [16, 17] were the first to apply some fundamental notions and powerful methods of differential geometry to the study of nonlinear regression models. Specifically, Beale applied the notion of curvature to improve approximate confidence regions, while Bates and Watts identified intrinsic and parametric curvatures and developed inferential procedures based on these measures. This work was further advanced in a number of publications including [18, 19]. One needs to note that the referenced curvature measures are local since they are computed from the partial derivatives of the model evaluated at a specified parameter vector. Thus, these measures reflect local behavior of the model in a neighborhood of the specified parameter vector and may lack an ability to capture the model's global behavior. This motivates development of methods based on a technique other than the Taylor's expansion in a neighborhood of the specified parameter vector.

Liu and Rong Li [20] surveyed various existing local measures of nonlinearity (MoNs), listed some of their deficiencies, and proposed their own global MoN. Pronzato and Pazman [7] introduced a new, global MoN, that they called an extended measure of intrinsic nonlinearity due to its relationship to the intrinsic curvature of Bates and Watts [16]. Khinkis et al. [21] have recently introduced new, geometrically appealing, MoNs reflecting the model's global behavior. They advance the methodology which uses the notion of equidistant function originally introduced in [22] and further developed in [21]. These measures can be used in parameter estimation and in the design of simulation studies.

Simulation studies represent an important methodology in pharmacokinetics (PK) since they are used routinely to explore different aspects of PK models. Such ubiquitous use motivates the analysis of simulation studies themselves in order to ensure their reliability. As parameter estimation is commonly embedded into a simulation study, the validity of the results of any simulation study cannot be a priori guaranteed for the reasons discussed above. This paper addresses specifically parameter estimation in PK models obtained via simulation. Metzler [23] alerted the pharmacological community that statistical statements based on fitting one, or a small number of data sets (individual fits) should be regarded with caution since much of mathematics coded into any statistical software depends heavily on approximations and asymptotic results valid in large samples. Accordingly, these approximations may or may not be satisfactory in small samples. Additionally, Metzler identified the following eight factors affecting statistical properties of the estimated parameters: the nonlinearity of the structural component of the model, the number of observations, the values of predictor variable, X (aka experimental design), the distribution of errors, the true (assumed) parameter value(s) used to generate the data, the method of estimation, parameter transformations, and initial parameter estimates. While these factors have been explored to a certain extent in the published literature, a deficiency of a satisfactory, practical guidance to a practitioner of simulation studies is apparent. No promising newer statistical techniques, including those featuring global methods, have yet been incorporated into commercially available software for nonlinear estimation, and it is not yet clear whether any of them will eventually be used by the practitioners, since the theory is not sufficiently developed and the computations are often prohibitive.

As reported in [24, p. 317], "in the two compartment model simulations, bias appeared in the estimators to the extent that bimodal sampling distributions of the estimators were observed as the random error of the model was increased". However, the bimodality of the parameter distribution may not be unique to the two compartment models and, in theory, could be exhibited under certain conditions (e.g. large variability of the errors) by any PK model.

The purpose of this paper is to familiarize the pharmacological community with two recently developed global

MoNs and to illustrate how they can help to answer the following three questions:

(a) whether the obtained parameter estimate is the true estimate;

(b) given a nonlinear model, what is the range of the errors within which no multiple minima of the OLS or WLS criterion occurs if (b_1) true parameter values are specified, or (b_2) a region of feasible parameter values is specified.

To this end, three popular models from PK literature are used. The computational formulas for the mentioned measures are given in "Appendix". One of this measures, $t(\theta)$, is developed mathematically in [21] while the mathematical derivations for the other, R_{max}, are omitted as this paper is intended for pharmacological, rather than mathematical or statistical, readership. The necessary mathematical details are given in [21, 22]. It must also be noted that the paper strives to illustrate the viability of the proposed MoNs in the specific models using specific parameter values and does not expand its scope beyond that stated goal.

Methods

Structural components

The following equations introduce structural components of the three popular models used in this paper.

Model 1

Michaelis–Menten (MM) model,

$$V(C) = \frac{V_{max} \cdot C}{K_m + C}. \tag{1}$$

The model is used to describe the velocity V of an enzymatic process as a function of C, the plasma concentration of the drug that drives the process, and contains two parameters, V_{max}, the maximum rate of the process that occurs when the enzyme is saturated, and K_m representing the concentration which results in half of V_{max}.

Model 2

The structural component of the Model 2 is a one compartment open model with a first order input given by the expression

$$C(t) = \frac{k_a D}{V_c(k_a - k_e)}\left(e^{-k_e t} - e^{-k_a t}\right), \tag{2}$$

where $C(t)$ is the concentration of drug in the body at time t, D is the absorbed mass of drug, V_c is the apparent volume of distribution of the central compartment, k_a is the absorption rate constant, and k_e is the elimination rate constant. Thus, Model 2 contains three parameters, V_c, k_a, and k_e. Since $\gamma = 1/V_c$, a parameter directly related to V_c, enters Model 2 in a linear fashion, both statistical and computational complexity of parameter estimation is reduced if γ, k_a, and k_e are used as the model parameters. Following [24], this will be the primary parameterization referenced in this paper. An alternative parameterization of Model 2 replaces k_e with CL, the clearance, and uses V_c, k_a, and CL as the model parameters. Since $CL = k_e \cdot V_c$, Model 2 parameterized this way is obtained when k_e is replaced with CL/V_c in Eq. (2).

Model 3

The structural component of Model 3 is a well-known two compartment model

$$C(t) = \frac{k_a D}{V_c}\left[\frac{\alpha - k_{21}}{(\alpha - \beta)(k_a - \alpha)} e^{-\alpha t} + \frac{k_{21} - \beta}{(\alpha - \beta)(k_a - \beta)} e^{-\beta t} - \frac{k_a - k_{21}}{(k_a - \alpha)(k_a - \beta)} e^{-k_a t}\right] \tag{3}$$

where, similarly to the Model 2, $C(t)$ is the concentration of drug in the body at time t, D is the absorbed mass of drug, V_c is the apparent volume of distribution of the central compartment, k_a is the absorption rate constant. The other parameters are k_{21}, the rate constant from the peripheral to the central compartment, and k_{12}, the rate constant from the central to the peripheral compartment. The parameters are related to each other and to k_e, the elimination rate constant, by the following two equations:

$$\alpha + \beta = k_e + k_{12} + k_{21}, \quad \text{and} \quad \alpha\beta = \mathrm{k_e k_{21}}.$$

These equations relate macro rate constants α and β to the first-order micro rate constants k_{12}, k_{21}, k_e. For the reasons given above, $\gamma = 1/V_c$ is used as one of the parameters. Following [24], γ, k_a, α, β, and k_{21} will be regarded as five independent parameters of the statistical model associated with the structural Model 3 that will be estimated. Alternatively, a set of more physiologically meaningful parameters, k_a, V_c, V_t, CL, and CL_d, also known as primary pharmacokinetic parameters, could be used to parameterize Model 3. Both referenced parameterizations use k_a; V_t is the volume of the peripheral (tissue) compartment, CL is the systemic plasma clearance, CL_d is the intercompartmental distribution clearance. Equation (4) establish a one-to-one correspondence between the two sets of parameters:

$$\gamma = \frac{1}{V_c}, \quad k_e = \frac{CL}{V_c}, \quad k_{12} = \frac{CL_d}{V_c}, \quad k_{21} = \frac{CL_d}{V_t}. \tag{4}$$

Thus, Model 3 will be parameterized using primary pharmacokinetic parameters if the three micro rate constants are replaced with their expressions given by (4).

It needs to be noted that both measures of nonlinearity presented in this paper do not depend on the model parameterization as these measures quantify intrinsic rather than parametric curvature of the model.

Statistical models

The three statistical models used in this paper are formed by adding an error term, e, to a structural component. The errors are assumed to be additive, unbiased, and normally distributed. A power model for the error variance is commonly used in the pharmacokinetics literature. In this paper, the errors are assumed to either have a constant variance or a constant coefficient of variation. Thus, the statistical models are described by the following equations:

$$y(C) = V(C) + e(C) \tag{5}$$

when Model 1 is used as structural component, and

$$y(t) = C(t) + e(t) \tag{6}$$

when one of the Models 2 and 3 is used as the structural component.

As noted, the variance structures are given by

$$Var(V(C)) = m^2 \tag{7}$$

and

$$Var(V(C)) = m^2[V(C)]^2 \tag{8}$$

with some constant m for Model 1, and by

$$Var(C(t)) = m^2[C(t)]^2 \tag{9}$$

with some constant m, for each of the Models 2 and 3 under the assumption of constant coefficient of variation (CV).

The following values were used in [24] to completely specify the variance structure in (9): $m = 0.01$, $m = 0.05$, $m = 0.1$.

The parameter estimation performed with the weighted least squares (WLS) under Eq. (9) uses weights based upon true (assumed) parameter values.

Designs and parameter values

Since all three models have been extensively published, we used parameter values, designs, and/or data sets referenced in the literature and listed below. Additionally, we used popular D-optimal designs [7, 25] with Model 1.

Model 1

Design 1A consists of the 6 design points, $\{0.2, 0.222, 0.286, 0.4,\ 0.667,\ 2.0\}$ (measured in umol). The 6 points were used in duplicates in [26] where the entire data set is published. The MM model with homogeneous errors (Eqs. (1), (7)) was used in the analysis of the data and has resulted in the vector of the estimated parameters $\hat{\theta} = (0.1058, 1.7077)$. Design 1B also consists of 6 design points (three replicates of two D-optimal design points). According to [25], D-optimal design for Model 1 with homogeneous errors consists of two design points, one at the highest experimentally attainable concentration, $C_2 = C_{max}$, and the other at $C_1 = (V_{max} \cdot C_{max})/(2V_{max} + C_{max})$.

Design 1B assumes that $C_{max} = 20$ meaning that the 6-point D-optimal design replicates each of the two D-optimal concentrations, $(1.7077 * 20)/(2 * 1.7077 + 20) = 1.459$ and 20, three times, and equals to $\{1.459, 1.459, 1.459, 20, 20, 20\}$. When a constant CV is applicable, one of the two D-optimal points remains at $C_2 = C_{max}$ while the other uses the smallest practically reasonable concentration, $C_1 = C_{min}$. Thus, design 1C $=\{0, 0, 0, 20, 20, 20\}$ was obtained under the assumption that [0, 20] is a viable concentration range.

Model 2

This model was studied via simulations in [24] with true (assumed) parameter values of $\gamma = 75/D1$ (meaning $D/V_c = 75\,\text{deq/l}$), $k_a = 1.5\ h^{-1}$ and $k_e = 0.4\ h^{-1}$. Simulated values depend on $\gamma \cdot D$ for any fixed input D, which is why this quantity is listed here rather than individual values of γ and D. Consequently, D may assume any plausible value. The listed parameter values are retained in this paper together with the two designs used in [24], Design 2A = $\{0.5, 1.0, 1.5, 3.0, 7.0\}$ and Design 2B = $\{0.5, 1.0, 1.5, 2.0, 3.0, 5.0, 7.0,\ 9.0,\ 11.0\}$ (times are measured in hours).

Model 3

This model was studied via simulations in [24] where the true (assumed) parameter values were specified as $\gamma = 0.441/D1$ (so that $D/V_c = 0.441\,\text{meq/l}$), $k_a = 2.389\,\text{h}^{-1}$, $\alpha = 0.914\,\text{h}^{-1}$, $\beta = 0.0462\,\text{h}^{-1}$, and $k_{21} = 0.25\,\text{h}^{-1}$. Design 3A from [24] was used with Model 3 in this paper. It consists of 16 time points 0.25, 0.5, 0.75, 1.0, 1.5, 2.0, 3.0, 6.0, 9.0, 12.0, 15.0, 18.0, 21.0, 24.0, 36.0, and 48.0 h.

Following [21], regions in the data space satisfying the “no multiple minima” property will be called regions of readily identifiable global minima (RIGM). These regions are of particular practical interest since, as long as the data belong to an RIGM region, no undesirable effects discussed in the Introduction occur. In a sense, a nonlinear model behaves like a linear one when the data fall within RIGM. A formula expressing the largest radius of an RIGM ball, R_{max}, is presented in "Appendix". This value of R_{max} translates into the largest error variance that satisfies, with a specified probability, the RIGM property. Indeed, under assumptions concerning the errors listed above, $\|\varepsilon\|_W^2/m^2$ follows the χ^2 distribution with N degrees of freedom. Here $\varepsilon = (e(C_1), \ldots, e(C_N))$ or $\varepsilon = (e(t_1), \ldots, e(t_N))$ is an N-dimensional vector of model errors introduced in (5) and (6), $\|\varepsilon\|_W$ is a norm of a vector $\varepsilon \in \mathbb{R}^N$ defined in "Appendix". Then

$$\mathbb{P}\left(\frac{\|\varepsilon\|_W^2}{m^2} > \chi^2_{p,N}\right) = p,$$

is a probability statement consistent with the distribution of $\|\varepsilon\|_W^2/m^2$ identified above and using a cut-off point of $\chi^2_{p,N}$. Then

$$\mathbb{P}(\|\varepsilon\|_W^2 > R_{max}^2) = \mathbb{P}(\|\varepsilon\|_W > R_{max}) = p \text{ when } \mathrm{m} = \frac{\mathrm{R_{max}}}{\sqrt{\chi^2_{\mathrm{p,N}}}},$$

and

$$\mathbb{P}(\|\varepsilon\|_W > R_{max}) < p \text{ when } \mathrm{m} < \frac{\mathrm{R_{max}}}{\sqrt{\chi^2_{\mathrm{p,N}}}}.$$

This is a fundamental inequality establishing a connection between the structural model (like the ones defined by Eqs. (1)–(3) of this paper) and the stochastic component. Due to this inequality, one can assert that the probability of observing data outside of an RIGM ball with the radius R_{max} is small if a small p was used and $m \leq R_{max}/\sqrt{\chi^2_{p,N}}$. Thus, R_{max} serves as a measure of nonlinearity of the model that could be useful for planning simulation studies.

Another measure, $t(\theta)$, is introduced (Eq. 18) to help to decide whether a found parameter is indeed a global minimum of the least squares criterion used in the parameter estimation. An example of such application is given in "Appendix".

It is worth noting that in linear regression models both R_{max} and $t(\theta)$ are infinite which is equivalent to both $1/R_{max}$ and $1/t(\theta)$ equal 0. The latter expressions can be viewed as curvatures accounting for global properties of a statistical model. In nonlinear regression each R_{max} and $t(\theta)$ lies between 0 and infinity, with smaller values indicating larger degrees of nonlinearity and larger values conforming to smaller degrees of nonlinearity.

Results

As noted in the introduction, Formula (19) allows to compute the largest radius, R_{max}, of the RIGM ball. The values of R_{max} and the corresponding values of $m = R_{max}/\sqrt{\chi^2_{p,N}}$ (listed in parenthesis) are presented in Table 1 for Model 1 with three different designs, 1A–1C, and variance structures (7) and (8). Similarly, the values of R_{max} and the corresponding values of $m = R_{max}/\sqrt{\chi^2_{p,N}}$ are presented in Table 2 for Model 2 with two different designs, 2A and 2B, and a variance structure assuming constant CV (9); such values are listed in Table 3 for the Model 3 and a variance structure (9). The value of $p = 0.005$ was used in all three tables. The m values listed in the tables are the maximal values which guarantee that no undesirable issues will arise in simulations. Table 1 illustrates that $R_{max} = \infty$ for both designs 1B and 1C with either variance structure. This means that no undesirable issues discussed above will arise regardless of the magnitude of the error variances. It is proven in "Appendix" that both R_{max} and $t(\theta)$ equal infinity for any nonlinear model with any assumed true parameter values when the number of observations, N, equals to the number of estimated parameters q. This property is preserved when N is a multiple of q, i.e. when q-point design is replicated. D-optimal designs typically consist of q points. When replicated, such D-optimal designs lead to infinite R_{max} and $t(\theta)$, as shown in Table 1 in the case of MM model. However, when design 1A is used, the table alerts that the mentioned undesirable effects may appear if the model's standard deviation is larger than 0.011 or CV is larger than 0.332, respectively. Table 2 illustrates that the m value of the CV increases as more time points are added to the design in the specific case of one-compartment model. As m becomes larger and exceeds the corresponding value listed in the relevant table, the distributions of some estimated parameters evolve from unimodal and approximately normal into bimodal. The distributions of the estimated parameters for Model 3 were obtained via simulations for each of $m = 0.025$, 0.05 and 0.1. Each simulation used 1000 runs. While unimodal distributions are

Table 1 Results for Michaelis–Menten Model 1

Variance structure	Design	
	Design 1A	Design 1B and 1C
Constant variance	0.048 (0.0111)	$\infty(\infty)$
Constant CV	1.432 (0.332)	$\infty(\infty)$

R_{max} and m are shown in the table

Table 2 Results for one-compartment Model 2

Design	R_{max}	m
2A	0.566	0.138
2B	0.721	0.149

Table 3 Results for two-compartment Model 3

Design	R_{max}	m
3A	0.158	0.027

observed for each of the five estimated parameters with $m = 0.025$ that is just under $m = 0.027$ listed in the Table 3, bimodality in the distributions of γ, k_a and α occurs when $m = 0.05$, getting even more pronounced (larger role of an alternative mode) when $m = 0.1$. This is consistent with findings reported in [24] and directly translates in bimodality of each k_a and V_c if primary pharmacokinetic parameters are used.

In order to get a better sense of how the true (assumed) parameter values affect both R_{max} and m when a design is fixed, 12 additional sets of parameters $(\gamma, k_a, \alpha, \beta, k_{21})$ were created and used as true (assumed) parameters for Model 3. These parameter sets were formed by combining each of the three values, $0.2205/D$, $0.441/D$, and $0.882/D$ of γ with each of the two values, 1.195 and 4.778, for k_a, and with each of the two values, 0.457 and 1.828, for α. The values of $\beta = 0.0462$ and $k_{21} = 0.25$ were the same as the true (assumed) values of β and k_{21} originally specified with Model 3. Note that each of the values of (γ, k_a, α) created this way equals to 0.5 times and 2 times of the value of the corresponding parameter ($0.441/D$, 2.389, and 0.914, respectively) used above. These parameter values were used together with the 16-point design for Model 3 specified above. The values of R_{max} varied across the 12 parameter sets from the smallest $R_{max} = 0.049$ when the true (assumed) vector of parameters was ($0.2205/D$, 1.1945, 0.457, 0.0462, 0.25) to the largest $R_{max} = 0.551$ when the true (assumed) vector of parameters was ($0.882/D$, 4.778, 1.828, 0.0462, 0.25), which translates in the corresponding values of $m = 0.0083$ and $m = 0.094$ respectively. As these extreme values of m differ by a factor larger than 11, this means that, for example, when the CV of the model errors, m, equals 0.09, the undesirable properties associated with the model's nonlinearity will be exhibited when the true (assumed) parameter vector is ($0.2205/D$, 1.1945, 0.457, 0.0462, 0.25) but not when it is ($0.882/D$, 4.778, 1.828, 0.0462, 0.25). This serves as an illustration of how calculation of R_{max} can be incorporated into a simulation study when a multi-point design covering the entire response curve has been decided upon, and the combined effect of the true (assumed) parameter values and the variability of the errors is of interest.

The function $t(\theta)$ (Eq. (18)) helps to decide whether a given candidate parameter estimate $\hat{\theta}$ is the true least square estimate. This is done by comparing $t(\hat{\theta})$ to $\sqrt{S(\hat{\theta})}$; the inequality $t(\hat{\theta}) < \sqrt{S(\hat{\theta})}$ guarantees that $\hat{\theta}$ is the true least squares estimate. Importantly, as shown in "Appendix" (Eqs. (24), (29), (30)), in the presence of a linear parameter, both $t(\theta)$ and $R_{max} = R_{max}(\bar{\theta})$ are proportional to the linear parameter. Each of the Models 1–3 contains a linear parameter, such as V_{max} in Model 1 and γ in Models 2 and 3. Figure 1 presents a graph of $h(\theta^{nl}) = h(K_m)$ for Model 1 with design 1A when a constant variance error structure (7) was used. The function $h(\theta^{nl})$ enters Eq. (24) and is defined in Eq. (25). Figure 1 allows to calculate $t(\theta) = t(V_{max}, K_m)$ for any parameter (V_{max}, K_m) by multiplying V_{max} by $h(K_m)$ obtained from the Fig. 1. Figure 1 shows that in the presented case $h(K_m)$ monotonically decreases with increase of K_m.

Discussion

This paper illustrates a novel statistical approach allowing to predict potential undesirable properties of the estimated parameters in a PK (generally, nonlinear regression) model with additive errors. Such illustration is done by using three popular models, one compartment and two compartment PK models and MM model, all widely cited in the literature [24, 26, 27]. Both the model designs and error structures used in the referenced literature are preserved in this paper. It was noted in [24] that bimodality in some parameters occurs in a two compartment, but not in a one compartment model when the error variances are large. While this is true, the major point that this paper strives to convey is that bimodality of some parameter distributions may occur in any nonlinear model when the error variances are

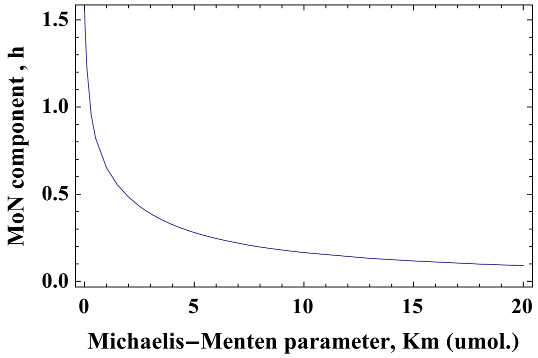

Fig. 1 Plot of $h(V_{max})$ (vertical axis) versus K_m (horizontal axis). The measure of nonlinearity $t(V_{max}, K_m)$ is computed in Eq. (24) as $t(V_{max}, K_m) = V_{max} \cdot h(K_m)$. Model 1 with variance structure (7) and design 1A is used

sufficiently large. Such large variances are not practical when a model is only moderately nonlinear. Thus, we do not observe the bimodality phenomenon in some models because the error variances used in simulations are below the threshold causing bimodality. The m values shown in Table 2 explain why bimodality was not reported in [24] for Model 2. Two MoNs allowing to quantify what is meant by "sufficiently large variances" are presented in this paper. Their calculation is not included in any currently available commercial software, but it is strongly recommended that R_{max}, along with the corresponding parameter(s) of the variance structure of the model, is computed prior to performing simulations. The notion of "sufficiently large variances" can't be expressed in absolute terms: a formula like the one presented above which expresses m in terms of R_{max} establishes a proper reference. Notably, an intuitively appealing design may not achieve unimodality of parameter estimates even with relatively small error variances if a model is highly nonlinear. Model 3 together with the 16-point design that nicely covers the entire concentration curve over the time interval [0, 48] serves as an illustration of that. One would expect such a design to result in unimodal parameter distributions. However, it doesn't achieve that even with the coefficient of variation m as small as 0.05. Conversely, a relatively sparse design might do the job in a less nonlinear model with larger error variances. D-optimal designs and, generally, replicates of designs with the number of design points equal to the number of estimated parameters, eliminate intrinsic nonlinearity of the model and as such were advocated in the literature [8]. However, alleviating nonlinearity concerns is just one of many objectives that a good design needs to satisfy. Additionally, sparse designs may lead to issues related to parameter identifiability. For these reasons, less sparse designs are often used, including range-covering designs.

Our code to compute R_{max} is written in MatLab and will be available to the readers. It should be noted that this calculation requires global minimization, among other mathematical operations. Generally, this is not a straightforward task since an output of any optimization routine depends on initial values of the variables over which optimization is performed. Availability of formulas like (Eqs. (29), (30)) significantly expedite these calculations.

As explained above, $t(\theta)$ (or its modified version, minimum of $t(\theta)$ over its nonlinear parameters), can be used to ascertain that the true least-squares parameter estimate has been found.

Conclusion

This paper presents novel formulas expressing two measures of nonlinearity of a PK model. By comparing one of these measures, R_{max}, to the stochastic component, specifically, variances of the errors, one can predict whether any undesirable features associated with nonlinearity will occur. The formulas are illustrated using three popular models with specified parameter values, which is why the validity of the relevant presented findings is limited to the considered cases. D-optimal designs eliminate intrinsic nonlinearity so that intrinsic curvature of a nonlinear model associated with either of the two measures of nonlinearity is zero, just like it is in a linear model. It is hoped that the proposed approach is further developed and used in the design and analysis of PK experiments.

Appendix

In this appendix we present formulas needed to perform calculations used in this paper. We use vector notation common in statistical literature and, specifically, that used in [28].

The model, information matrix, and projection on the tangent plane

Following [28], we consider a nonlinear regression model given by

$$y = \eta(\theta) + \varepsilon; \quad \theta \in \Theta, \mathbb{E}(\varepsilon) = 0, Var(\varepsilon) = m^2 W, \tag{10}$$

where $\theta = (\theta_1, \ldots, \theta_q)^t$ is the vector of unknown parameters. Thus $\theta = (V_{max}, K_m)$ in Model 1 (Eq. (1)), $\theta = (\gamma, k_a, k_e)$ in Model 2 (Eq. (2)), and $\theta = (\gamma, k_a, \alpha, \beta, k_{21})$ in Model 3 (Eq. (3)).

It is assumed that $\theta \in \Theta$, where Θ is a subset of $\mathbb{R}^q$ that serves as the set of feasible parameter values. In (10), $y \in \mathbb{R}^N$ is the vector of observed data, $\varepsilon \in \mathbb{R}^N$ is the vector of errors that are assumed to be normally distributed, m is the parameter of the variance component which may, but need not, be known, W is a known positive semi-definite matrix. The mapping $\eta : \Theta \rightarrow \mathbb{R}^N$ is a known, twice continuously differentiable mapping on $int(\Theta)$. Clearly, (10) is consistent with the statistical models described by the equations in the theoretical sections above. The parameter m equals to the standard deviation of the error of an individual observation under the assumption of a constant variance in which case $W = I$, the $N \times N$ identity matrix (this is consistent with Eq. (7)). W is an $N \times N$ diagonal matrix with $(C(t_1))^2, \ldots, (C(t_N))^2)$ along the main diagonal when

Eq. (9) is used, and is calculated similarly for Eq. (8). Let $\bar{\theta}$ denotes the vector of true (assumed) parameters.

The weighted least squares estimate (WLSE) of θ is defined as

$$\hat{\theta} = \hat{\theta}(y) := argmin_{\theta \in \Theta} S(\theta), \tag{11}$$

where $S(\theta) = \|y - \eta(\theta)\|_W^2$ is the objective function (criterion) and where the definition of the squared norm $\|a\|_W^2 := a^t W^{-1} a$ is used. This norm corresponds to the inner product $\langle a, b \rangle_W := a^t W^{-1} b$.

Formulas (12)–(14) define some matrices used to calculate $t(\theta, \theta_a)$ and $R(\bar{\theta}, \theta, \theta_a)$ (Eq. (19) below), namely, the Jacobian matrix (12), the information matrix (13), and the projector matrix onto a plane tangent to the expectation surface at θ, (14):

$$J(\theta) = \left[\frac{\partial \eta(\theta)}{\partial \theta_1} \cdots \frac{\partial \eta(\theta)}{\partial \theta_q} \right], \tag{12}$$

$$M_W(\theta) = J^t(\theta) W^{-1} J(\theta), \tag{13}$$

$$P(\theta) = J(\theta) M_W^{-1}(\theta) J^t(\theta) W^{-1}. \tag{14}$$

One can then define $l_1(\theta, \bar{\theta})$ and $l_2(\theta, \theta_a)$ as

$$l_1(\theta, \bar{\theta}) = (I - P(\theta))(\eta(\theta) - \eta(\bar{\theta})), \tag{15}$$

$$l_2(\theta, \theta_a) = (I - P(\theta))(\eta(\theta_a) - \eta(\theta)). \tag{16}$$

The equidistant function

The equidistant function $t(\theta, \theta_a)$ is derived in [21] as

$$t(\theta, \theta_a) = \frac{\|(\eta(\theta_a) - \eta(\theta))\|_W^2}{2\|l_2(\theta, \theta_a)\|_W}. \tag{17}$$

The reader is referred to [21] for the relevant mathematical details. A related quantity is

$$t(\theta) = \min_{\theta_a \in \Theta} t(\theta, \theta_a). \tag{18}$$

The same symbol used to denote two different functions (left-hand sides of Eqs. (17) and (18)) should not cause any confusion since these two functions depend on a different number of clearly identified variables.

Radius of the RIGM ball

The expression $R(\bar{\theta}, \theta, \theta_a)$ is calculated as

$$R(\bar{\theta}, \theta, \theta_a) = \sqrt{\|P(\theta)(\eta(\theta) - \eta(\bar{\theta}))\|_W^2 + \frac{\left(\|\eta(\theta_a) - \eta(\theta)\|_W^2 + 2\langle l_1(\theta, \bar{\theta}), l_2(\theta, \theta_a) \rangle\right)^2}{4\|l_2(\theta, \theta_a)\|_W^2}} \tag{19}$$

where $\theta \in int\, \Theta, \theta_a \in \Theta$, and

$$R_{max} = R_{max}(\bar{\theta}) = \min_{\theta \in int\, \Theta, \theta_a \in \Theta} R(\bar{\theta}, \theta, \theta_a). \tag{20}$$

The meaning of the quantities calculated in Eqs. (17)–(20) is explained in theoretical section. The models examined in this paper share a common feature, namely, some of the parameters enter these models in a linear fashion, which leads to simplification of the corresponding formulas. Such simplified formulas are available and are presented below in the case of a single linear parameter.

Simplified calculation of nonlinearity measures in models with a linear parameter

Assume that

$$\eta(\theta) = \theta_1 \times \phi(\theta_2, \ldots, \theta_q) = \theta^l \times \phi(\theta^{nl}). \tag{21}$$

Then

$$t(\theta, \theta_a) = \frac{\|\theta_a^l(\phi(\theta_a^{nl}) - \theta^l \phi(\theta^{nl}))\|_W^2}{2|\theta_a^l| \|(I - P(\theta))\phi(\theta_a^{nl})\|_W}. \tag{22}$$

Since in the three models considered in this paper the linear parameter, V_{max} or γ, is nonnegative, we assume that $\Theta = [0, \infty) \times \Theta^{nl}$ where $\Theta^{nl} \subset \mathbb{R}^{q-1}$. Under this assumption, the minimum of $t(\theta, \theta_a)$ over $\theta_a^l \in [0, \infty)$ is reached at $\theta_a^l = \theta^l \times \frac{\|\phi(\theta^{nl})\|_W}{\|\phi(\theta_a^{nl})\|_W}$ and equals to

$$\begin{aligned} &\min_{\theta_a \in [0,\infty)} t(\theta, \theta_a) \\ &= \theta^l \times \frac{\|\phi(\theta^{nl})\|_W \cdot \|\phi(\theta_a^{nl})\|_W - \langle \phi(\theta^{nl}), \phi(\theta_a^{nl}) \rangle_W}{\|(I - P(l, \theta^{nl}))\phi(\theta_a^{nl})\|_W} \\ &= \theta^l \times g(\theta^{nl}, \theta_a^{nl}), \end{aligned} \tag{23}$$

where

$$g(\theta^{nl}, \theta_a^{nl}) = \frac{\|\phi(\theta^{nl})\|_W \cdot \|\phi(\theta_a^{nl})\|_W - \langle \phi(\theta^{nl}), \phi(\theta_a^{nl}) \rangle_W}{\|(I - P(l, \theta^{nl}))\phi(\theta_a^{nl})\|_W}.$$

Following (23),

$$\begin{aligned} t(\theta) &= \min_{\theta_a \in \Theta} t(\theta, \theta_a) = \theta^l \times \min_{\theta_a^{nl}} g(\theta^{nl}, \theta_a^{nl}) \\ &= \theta^l \cdot h(\theta^{nl}), \end{aligned} \tag{24}$$

where

$$h(\theta^{nl}) = \min_{\theta_a^{nl}} g(\theta^{nl}, \theta_a^{nl}). \tag{25}$$

Similarly, an expression of $R_{max}(\bar{\theta})$ (Eq. (20)) using only nonlinear components of the parameters can be found under assumption (21). Note that in this case $(I - P(\theta))\eta(\theta) = 0$ and that $P(\theta)$ does not depend on θ^l: $P(\theta) = P(1, \theta^{nl})$.

Since the numerator of (26) is a quadratic function in θ_a^l, the condition that its discriminant is nonnegative needs to be imposed in order for this function to have real roots. Thus,

$$\min_{\theta_a^l \in [0,\infty)} \frac{\left(\|\eta(\theta_a) - \eta(\theta)\|_W^2 + 2\langle l_1(\theta, \bar{\theta}), l_2(\theta, \theta_a)\rangle_W\right)^2}{4\|l_2(\theta, \theta_a)\|_W^2} = \min_{\theta_a^l \in [0,\infty)} \frac{(\theta_a^l)^2\|\phi(\theta_a^{nl})\|_W^2 - 2\theta_a^l\theta^l\langle\phi(\theta_a^{nl}), \phi(\theta^{nl})\rangle_W + (\theta^l)^2\|\phi(\theta_a^{nl})\|_W^2 -}{4(\theta_a^l)^2\|(I - P(1, \theta^{nl}))\phi(\theta_a^{nl})\|_W^2} \frac{2\theta_a^l\bar{\theta}^l\langle(I - P(1, \theta^{nl}))\phi(\bar{\theta}^{nl}), (I - P(1, \theta^{nl}))\phi(\theta_a^{nl})\rangle_W^2}{4(\theta_a^l)^2\|(I - P(1, \theta^{nl}))\phi(\theta_a^{nl})\|_W^2}, \tag{26}$$

provided that

$$0 < \theta^l < \frac{\bar{\theta}^l\langle(I - P(1, \theta^{nl}))\phi(\bar{\theta}^{nl}), (I - P(1, \theta^{nl}))\phi(\theta_a^{nl})\rangle_W}{\|\phi(\theta^{nl})\|_W \cdot \|\phi(\theta_a^{nl})\|_W - \langle\phi(\theta^{nl}), \phi(\theta_a^{nl})\rangle_W}. \tag{27}$$

Introduce

$$g_1(\bar{\theta}^{nl}, \theta^{nl}) = \max_{\theta_a^{nl}} \frac{\langle(I - P(1, \theta^{nl}))\phi(\bar{\theta}^{nl}), (I - P(1, \theta^{nl}))\phi(\theta_a^{nl})\rangle_W}{\|\phi(\theta^{nl})\|_W \cdot \|\phi(\theta_a^{nl})\|_W - \langle\phi(\theta^{nl}), \phi(\theta_a^{nl})\rangle_W}. \tag{28}$$

Standard minimization of $R(\bar{\theta}, \theta, \theta_a)$ over $\theta \in int\,\Theta$, $\theta_a \in \Theta$ leads to the following simplified expressions:

$$R_1(\bar{\theta}) = \min_{\theta^{nl} \in int\,\Theta} \sqrt{\|\bar{\theta}^l P(1, \theta^{nl})\phi(\bar{\theta}^{nl})\|_W^2} = \bar{\theta}^l \cdot \min_{\theta^{nl} \in int\,\Theta} \|P(1, \theta^{nl})\phi(\bar{\theta}^{nl})\|_W;$$
$$R_2(\bar{\theta}) = \min_{\theta^{nl} \in int\,\Theta} \sqrt{\|\bar{\theta}^l g_1(\bar{\theta}^{nl}, \theta^{nl})\phi(\theta^{nl}) - \bar{\theta}^l P(1, \theta^{nl})\phi(\bar{\theta}^{nl})\|_W^2} = \bar{\theta}^l \cdot \min_{\theta^{nl} \in int\,\Theta} \|g_1(\bar{\theta}^{nl}, \theta^{nl})\phi(\theta^{nl}) - P(1, \theta^{nl})\phi(\bar{\theta}^{nl})\|_W. \tag{29}$$

Then,

$$R_{max} = \min_{\theta \in int\,\Theta, \theta_a \in \Theta} R(\bar{\theta}, \theta, \theta_a) = \min\{R_1(\bar{\theta}), R_2(\bar{\theta})\}. \tag{30}$$

This is the value reported in the Tables 1, 2, and 3.

The $R(\bar{\theta}, \theta, \theta_a)$ and R_{max} share the following feature: when a design is replicated l times, each of $R(\bar{\theta}, \theta, \theta_a)$ and R_{max} will be multiplied by $\sqrt{l}$.

Calculation of MoNs when designs replicating *q* distinct points are used

Assume that a design with $N = q$ distinct points is used. Then the inverse of information matrix (13) is computed as

$$M_W^{-1}(\theta) = \left[J^t(\theta)W^{-1}J(\theta)\right]^{-1} = [J(\theta)]^{-1}W[J^t(\theta)]^{-1},$$

so the Eq. (14) becomes

$$\begin{aligned} P(\theta) &= J(\theta)M_W^{-1}(\theta)J^t(\theta)W^{-1} \\ &= J(\theta)[J(\theta)]^{-1}W[J^t(\theta)]^{-1}J^t(\theta)W^{-1} = I, \end{aligned}$$

while Eq. (16) turns into

$$l_2(\theta, \theta_a) = 0. \tag{31}$$

Equations (17), (18), (31) lead to

$$t(\theta) = \infty, \text{ or, equivalently, } \frac{1}{\mathrm{R_{max}}} = 0. \tag{32}$$

Similarly, Eqs. (19) and (20) lead to

$$R_{max} = \infty, \text{ or, equivalently, } \frac{1}{\mathrm{R_{max}}} = 0. \tag{33}$$

Equations (32) and (33) remain valid when $N = l \cdot q$ with some integer l. This follows either from a simple algebraic calculation or by multiplication of the respective MoN corresponding to a non-replicated design by $\sqrt{l}$ (as stated above).

References

1. Demidenko E (2000) Is this the least squares estimate? Biometrika 87:437–452
2. Toulias T, Kitsos C (2016) Fitting the Michaelis–Menten model. J. Comput. Appl. Math. 296:303–319
3. Cobelli C, Salvan A (1997) Parameter estimation in a biological two compartment model—a computer experimental study of the influence of the initial estimate in the parameter space and of the model representation. Math Biosci 33:51–62
4. Ritz C, Streibig JC (2008) Nonlinear regression with R. Springer
5. Chavent G (1990) A new sufficient condition for the well-posedness of non-linear least-square problems arising in identification and control. In: Analysis and optimization of systems, lecture notes in control and inform sci, vol 144. Springer, pp 452–463
6. Demidenko E (2006) Criteria for global minimum of sum of squares in nonlinear regression. Comput Stat Data Anal 51:1739–1753
7. Pronzato L, Pazman A (2013) Design of experiments in nonlinear models. Springer, New York
8. Pronzato L, Walter E (2001) Eliminating suboptimal local minimizers in nonlinear parameter estimation. Technometrics 43(4):434–442
9. Toulias T, Kitsos C (2016) Estimation aspects of the Michaelis–Menten model. REVSTAT 14(2):101–118

10. Ratkowsky DA (1983) Nonlinear regression modeling, a uniform practical approach. Marcel Dekker, New York
11. Ratkowsky DA (1990) Handbook of nonlinear regression models. Marcel Dekker, New York
12. SAS/STAT(R) (2015) 14.1 User's Guide, the NLIN procedure. SAS Institute Inc., Cary, NC
13. Box M (1971) Bias in nonlinear estimation. J R Stat Soc 33:171–201
14. Hougaard P (1985) The appropriateness of the asymptotic distribution in a nonlinear regression model in relation to curvature. J R Stat Soc B 47:103–114
15. Beale EML (1960) Confidence regions in nonlinear estimation. J R Stat Soc 22:41–76
16. Bates DM, Watts DG (1980) Relative curvature measures of nonlinearity (with discussion). J R Stat Soc B 42:1–25
17. Bates DM, Watts DG (1988) Nonlinear regression analysis and its applications. Wiley, New York
18. Hamilton DC, Watts DG, Bates DM (1982) Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions. Ann Stat 10:386–393
19. Hamilton D (1986) Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions. Biometrika 73:57–64
20. Liu Y, Li XR (2015) Measure of nonlinearity for estimation. IEEE Trans Signal Process 63(9):2377–2388
21. Khinkis L, Crotzer M, Oprisan A (2018) Sizing up the regions of unique minima in the least squares nonlinear regression. Math Appl 7:41–52
22. Khinkis L, Crotzer M (2008) A new approach for finding global minima in nonlinear least squares regression. In: Proceedings of the American Statistical Association Biopharmaceutical Section—JSM, pp 2264–2271
23. Metzler CM (1981) Estimation of pharmacokinetic parameters: statistical considerations. Pharmacol Ther 13:543–556
24. Laskarzewski DL, Weiner DL, Ott L (1982) A simulation study of parameter estimation in the one and two compartment models. J Pharmacokinet Biopharm 10(3):317–334
25. Endrenyi L (1981) Design of experiments for estimating enzyme and pharmacokinetic parameters. In: Endrenyi L (ed) Kinetic data analysis: design and analysis of enzyme and pharmacokinetic experiments. Plenum Press, New York, NY, pp 137–167
26. Watts DG (1981) An introduction to nonlinear least squares. In: Endrenyi L (ed) Kinetic data analysis: design and analysis of enzyme and pharmacokinetic experiments. Kluwer Academic Pub, Dordrecht, pp 1–24
27. Westlake WJ (1973) Use of statistical methods in evaluation of in vivo performance of dosage forms. Pharm Sci 62:1579–1589
28. Pazman A (1993) Nonlinear statistical models. Kluwer, Dordrecht