

Tips for the choice of initial estimates in NONMEM

Seunghoon Han^{1,2*}, Sangil Jeon³ and Dong-Seok Yim^{1,2}

¹Department of Clinical Pharmacology and Therapeutics, Seoul St. Mary's Hospital, Seoul 06591, ²PIPET (Pharmacometrics Institute for Practical Education and Training), College of Medicine, The Catholic University of Korea, Seoul 06591, Korea, ³C&R research, Seoul 06199, Korea

*Correspondence: S. Han; Tel: +82-2-2258-7326, Fax: +82-2-2258-7876, E-mail: waystolove@catholic.ac.kr

Keywords

NONMEM,
initial estimate,
pharmacometrics,
pharmacokinetic-
pharmacodynamic model

pISSN: 2289-0882

eISSN: 2383-5427

The importance of precise information and knowledge on the initial estimates (IEs) in modeling has not been paid its due attention so far. By focusing on the IE, this tutorial may serve as a practical guide for beginners in pharmacometrics. A 'good' set of IEs rather than arbitrary values is required because the IEs where NONMEM kicks off its estimation may influence the subsequent objective function minimization. To provide NONMEM with acceptable IEs, modelers should understand the exact meaning of THETA, OMEGA and SIGMA based on physiology. In practice, problems related to the value of the IE are more likely to occur for THETAs rather than the random-effect terms. Because it is almost impossible for a modeler to give a precise IE for OMEGAs at the beginning, it may be a good practice to start at relatively small IEs for them. NONMEM may fail to converge when too small IEs are provided for residual error parameters; thus, it is recommended to give sufficiently large values for them. The understandings on the roles, meanings and implications of IEs even help modelers in troubleshooting situations which frequently occur over the whole modeling process.

Background

The initial estimate (IE) is an essential component of a NONMEM control stream. A single experience with pharmacometric modeling is sufficient for one to realize that the determination of specific IE values is not a simple task and that the final parameter estimates (FPE) and/or NONMEM run time may vary according to the IE values. In particular, beginners in the field of pharmacometric modeling and simulation typically have an inadequate understanding as to how NONMEM works and the meaning of the control stream components; therefore, knowledge regarding the concept, role, and specific value of the IE may be lacking. This problem should not be overlooked because a modeler may obtain irrelevant FPEs if the IE values of model parameters are selected without a complete understanding. In this context, it is important to have precise information and knowledge regarding the IE to achieve an acceptable model fit.

Indeed, it is almost impossible for beginners to thoroughly understand the various aspects of the IE. Because the IE is not

a simple set of values but an essential starting point for estimations based on the model structure, knowledge of the NONMEM estimation algorithm and of concepts of pharmacokinetics (PK) and pharmacodynamics (PD) is essential. Nevertheless, a modeler may avoid spending unnecessary time investigating FPEs that are not valid by determining appropriate IE values with a basic understanding of the IE. This tutorial seeks to provide comprehensive knowledge of the IE at a beginner's level, which has not been cohesively presented in a single article or learning material. By focusing only on the IE, which is typically discussed as a subtopic, this tutorial may serve as a practical guide for modeling efforts related to IE designation.

The need and role

The only and inevitable reason why we should define IE values upon writing a typical NM-TRAN control stream is that NONMEM uses an approach similar to the Newton-Raphson method (a quasi-Newton-type minimization algorithm) for parameter estimation.[1,2] Details of the estimation algorithm are not the focus of this article; however, the fundamental mechanism should be mentioned in discussing the IE. The algorithm starts by searching for the final set of parameter estimates (e.g., clearance, CL , or volume of distribution, V_d) from the IE values.

Copyright © 2016 S. Han

© It is identical to the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>).

© This paper meets the requirement of KS X ISO 9706, ISO 9706-1994 and ANSI/NISO Z.39.48-1992 (Permanence of Paper).

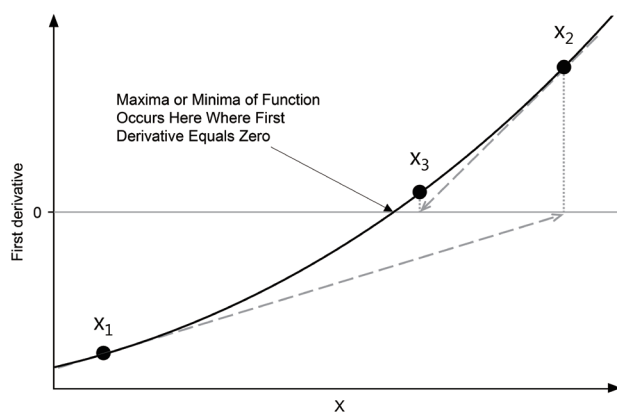


Figure 1. Simple presentation of the Newton–Raphson method. X_1 is the set of initial estimates (starting point) and X_i is the set of parameter estimates at each iteration step. (Redrawn based on the figure in reference [3])

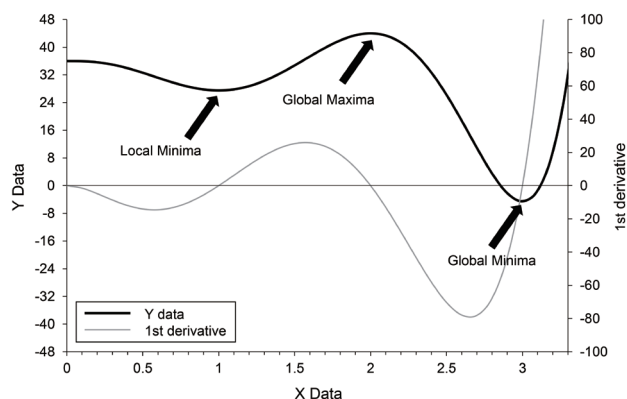


Figure 2. Example of functions having multiple minima. (Redrawn based on the figure in reference [3])

Of course, the final estimates are not obtained in a single step; the process requires multiple steps, referred to as "iterations". In each iteration, the parameter estimates approach the final values gradually, making the first derivative of the NONMEM objective function (OF) near 0 (with a positive second-derivative value). Thus, OF has a minimum value. Numerically, this procedure can be written as follows:

$$x_{i+1} = x_i - \left(\frac{d^2Y}{dx_i^2} \right)^{-1} \left(\frac{dY}{dx_i} \right)$$

where x_i is the IE point and x_{i+1} is the parameter estimates after the i^{th} iteration. See Figure 1 for a simple representation of the idea.

Importance of good IEs

NONMEM requires a 'good' set of IEs rather than arbitrary

Table 1. An example of iterative processes from different starting points

i	x_i	$\frac{dY}{dx_i}$	$\frac{dY}{d^2 x_i}$	x_{i+1}
Starting Value: 0.70				
1	0.70	-13.19	22.16	1.30
2	1.30	17.85	52.24	0.95
3	0.95	-2.72	56.64	1.00
4	1.00	0.09	60.09	1.00
Starting Value: 1.60				
1	1.60	25.80	-7.68	4.96
2	4.96	16,956.16	25,498.56	4.30
3	4.30	5,419.61	10,714.92	3.79
4	3.79	1,696.52	4,601.51	3.42
5	3.42	507.53	2,070.60	3.18
6	3.18	135.68	1,036.75	3.04
7	3.04	26.45	649.51	3.00
8	3.00	2.08	548.97	3.00
9	3.00	0.02	540.07	3.00
10	3.00	0.00	540.00	3.00
11	3.00	0.00	540.00	3.00

values because the IE may influence the location of the OF minimization (FPE values). This is particularly true when the model has greater complexity (i.e., an excessive number of parameters and/or non-linear structures). In this case, the shape of the NONMEM OF and its first derivative may not be simple and may have two or more minima. The following is a simple example of a complex function, suggested by Peter L. Bonate.[3] The shape of the following function and its first derivative are shown in Figure 2.

$$Y = 5x^6 - 36x^5 + \frac{165}{2}x^4 - 60x^3 + 36$$

This function has multiple minima, similar to the OFs of complex models. When the Newton–Raphson method is applied to this kind of function in the search for minima, the process will converge towards different minima (x value) according to the starting point. If the search starts at an x value of 0.70, the final x value will be 1.00, which is not the *global* minimum but a *local* minimum. To find the global minimum, the search process should start from a value closer to the global minimum (e.g., 1.60). The iterative steps are shown in Table 1. Thus, in actual PK-PD modeling practice, the modeler should provide plausible IE values for the corresponding parameters to avoid obtaining an irrelevant FPE. Another consideration in the importance of 'good' IEs is the relative amount of data; this will be discussed subsequently.

Meaning of THETA, OMEGA and SIGMA and the IE their initial estimates

There are two levels of consideration for the IE designation in terms of describing the 'distribution.' The first level is the parameter distribution. As all pharmacometricians are aware, typical population values of PK and/or PD parameters are defined using THETA (θ). This value may be similar to the mean or median of the individual parameter values included in the dataset, and indicates the general extent or the central tendency of the corresponding parameter. The IE values for θ s are given in the \$THETA block. However, the θ value contains no information on the variability or dispersion of corresponding parameters between individuals, which is another essential component to describe a specific distribution. Thus, the IE for the between-subject variability (BSV) of each parameter should be defined separately: this value is ω^2 , defined in the \$OMEGA block. Although one of the basic principles implemented in NONMEM is that each ETA (η , the discrepancy of an individual parameter from the typical population value) is normally distributed with a mean of 0, a modeler may describe a parameter distribution with various types of distribution by defining an appropriate parameter submodel. For example, if the submodel for CL is set as $CL = \theta_n + \eta_n$, then a normal distribution is assumed, whereas a submodel of $CL = \theta_n \times \exp(\eta_n)$ may be used to express a log-normal distribution.

When considering the parameter distribution, individual predictions can be obtained using empirical Bayesian estimates for individual parameters. At this stage, another consideration is the residual error, which is the difference between the individual predictions and the observed values. Several error models may be incorporated and, in some cases, multiple EPSILONs (ϵ) may be used (e.g., $Y = F \cdot (1 + \epsilon_1) + \epsilon_2$). Similar to the PK and/or PD parameters, the distribution should be defined for each ϵ ; however, the central tendency (e.g. mean) of the distribution is not necessarily considered because the sum of each ϵ for all observation points (i.e., $\sum_{i=1}^n \epsilon_{1i}$, where ϵ_{1i} is the value of ϵ_1 for the i^{th} observation when n observations are included in the dataset) is 0, and thus, the mean should be 0 (this is a basic rule implemented in NONMEM). Thus, the dispersion of the normal distribution is the only property that should be specified in the control stream and that is the initial estimate of σ^2 , defined for corresponding ϵ in the \$SIGMA block.

IE for THETAs

In practice, problems related to the value of the IE are more likely to occur for THETAs rather than the random-effect terms (η s and ϵ s). This is because the primary objective of a population analysis is to estimate the typical value of each parameter for the population. Additionally, there are more chances to provide inappropriate IE values for THETAs than OMEGAs and/or SIGMAs. In fact, providing IEs for OMEGAs and/or SIGMAs is more like ordering NONMEM to estimate those random effect parameters, rather than allocating NONMEM certain starting

points of estimation as in the case of the IEs for THETAs.

The IE for a fixed-effect parameter is the starting point of the estimation as well as a rough guess as to the FPE. To provide a 'good' set of IEs for THETAs, the modeler should have an appropriate level of knowledge of PK-PD and understand the meaning of each parameter. An important requirement for FPE is physiological relevance. It is not reasonable to try to obtain physiologically relevant FPE values with IE values that are not. In many cases, the values of CL , V_d , and absorption rate constant (k_a) in a first-order PK model tend to be within the ranges of 1–100 L/h, 10–1000 L, and 0.1–10 h⁻¹, respectively. When dealing with concentration data only for extravascular dosing, the parameter values (except absorption parameters) are actually divided by bioavailability (F , ranging from 0–1), resulting in larger values than the actual ones. Based on this understanding, IE values may be imported from the literature. In doing this, the modeler should consider the population, disease, and/or formulation differences that may influence the PK-PD parameters.

More precise and recommended ways for the determination of IE values include data exploration. The FPE values are the most plausible for the model structure (including the parameters used) given a certain dataset. That is, FPEs are determined by the content of the dataset. Therefore, with careful exploration of a dataset, a modeler may have a better chance of obtaining a good set of IEs. Even a simple time-log concentration plot for the population median concentration at each time point can serve as important information for the values of IEs as well as the model structure. For example, when such a plot for intravascular dosing data suggests that a two-compartment model may be the best structure, a modeler may draw two hypothetical lines explaining the data with the sum of exponentials and obtain values of the macro constants (Fig. 3).[4]

If the modeler is to build a control stream using these macro constants, the values may be used as is. However, to use micro constants or more physiological parameters, such as CL , volume of the central compartment (V_c), volume of the peripheral com-

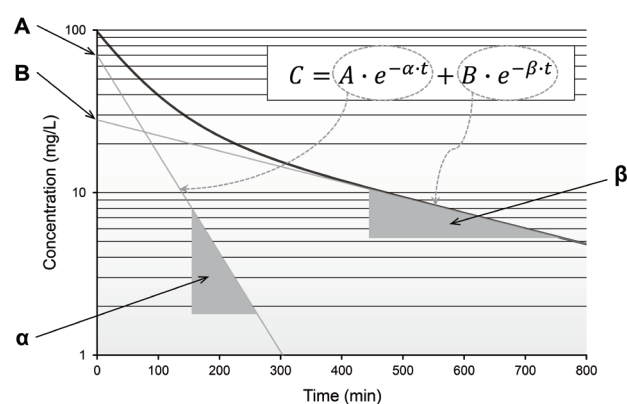


Figure 3. Sum of exponentials for a two-compartment model and macro constants. (Redrawn based on the figure in reference [4])

partment (V_p), and inter-compartmental clearance (Q), the IE values should be calculated from the values of the macro constants as follows:

$$k_{21} = \frac{A\beta + B\alpha}{A+B}, \quad k_{10} = \frac{\alpha\beta}{k_{21}}, \quad k_{12} = \alpha + \beta - k_{21} - k_{10}$$

$$V_c = \frac{\text{Dose}}{A+B}, \quad V_p = V_c \cdot \frac{k_{12}}{k_{21}}, \quad Q = V_c \cdot k_{12} = V_p \cdot k_{21}, \quad CL = V_c \cdot k_{10}.$$

A similar approach may be applied to the PD models. We recommend reading 'Chapter 2. Pharmacokinetic Concepts' and 'Chapter 3. Pharmacodynamic Concepts' in Pharmacokinetic and Pharmacodynamic Data Analysis, 5th ed. by Gabrielsson and Weiner[5] for information on the actual IE values for various PK-PD models.

In many modeling practices, designating the range for each THETA is recommended using the code (lower limit, IE, upper limit) because this prevents NONMEM from searching a THETA value outside the relevant range and helps to enhance practical efficiency (i.e., reduces run time). For steps to handle base models with no covariate, setting the lower limit of the THETA PK parameter as 0 is reasonable and appropriate because no PK parameter can have a negative value. However, care should be taken when covariates are included in a fixed effect. For example, CL between two groups may be modeled as follows (GRP = 0 for Group 1 and 1 for Group 2):

$$CL = \text{THETA}(n) + \text{GRP} \cdot \text{THETA}(n+1)$$

where the difference is the value of $\text{THETA}(n+1)$. If the lower limit of $\text{THETA}(n+1)$ is set at 0, this allows only a positive difference between Group 2 versus Group 1, whereas the data may suggest the opposite. Then, the determination of an acceptable FPE for $\text{THETA}(n+1)$ by NONMEM is likely to fail. A modeler can also handle this problem with careful data exploration and/or removing the limit on the THETA value. An upper limit is not recommended in the initial steps when clues on the parameter values are not sufficient, except for cases where a THETA cannot have a value above a specific limit inherently. As modeling effort progresses, an upper limit may be set with a certain degree of confidence for estimates of THETAs to reduce run time and to avoid irrelevant results.

IE for OMEGAs

Like fixed-effect terms, the value of IEs for BSV terms also serve as the starting point for estimations; however, designating certain initial values for such terms instead of fixing them at 0 is, again, more like ordering NONMEM to estimate the corresponding random effect. Further, they provide little value as a guess for the FPE. That is, it is difficult to determine the extent of between-subject or residual random variability with the dataset; thus, it is almost impossible for a modeler to give a precise IE for the model at the beginning. It may be a good practice to start modeling with relatively small IE values to the

random parameters. When the estimation of a certain ETA is to be estimated for the first time, the IE of ω^2 for a corresponding ETA is given as a very small value (e.g., 0.01, expecting ~10% coefficient of variation). This process is continued until all ω^2 for identifiable ETAs are estimated. Providing a larger value for an existing ETA, based on the results of a previous NONMEM run is not recommended at this stage because the FPE for ω^2 from a model with an insufficient number of ETAs is highly likely to be overestimated. Nevertheless, larger IE is sometimes necessary because assigning too small IE values to the parameters that inherently have a larger BSV (e.g., parameters related to drug absorption) may interfere model convergence (This is more like a trial-and-error issue). In addition to the IEs for each ω^2 , those for covariance terms between ETAs may be provided using the OMEGA BLOCK structure. In this case, a value that leads to a correlation coefficient larger than 1 should not be used. For example, when the ω values are 0.2 and 0.3 for two correlated ETAs, the maximum value allowed for covariance is 0.06 ($\rho = 1$). Thus, similar to IEs for ETAs, it is better to begin with a small value when one is trying to estimate covariance between ETAs.

IE for SIGMAs

For residual error parameters expressed with their variance (σ^2), the characteristics of each parameter determined in the model structure should always be checked. When only one EPSILON is used in the control stream, the value cannot be fixed at 0. If the EPSILON has an additive relationship with the observations (DV), the IE values for the corresponding σ^2 should reflect the magnitude of the DV. For example, a modeler may assume an additive error as large as the LLOQ of a DV. When the LLOQ is 10, the modeler should specify 100 as the IE of σ^2 (10 as the IE of σ); however, 0.01 (0.01 as $\sigma^2 = 0.1$ as σ) is more appropriate when the LLOQ is 0.1. This is one of the steps where beginners most frequently make mistakes. When σ^2 represents the proportional error, this consideration is unnecessary and just the squared value of the expected extent of error may be given (e.g., 0.01 for a 10% proportional error). The characteristic of additive and proportional errors is maintained even though they are used together in the form of combined errors. NONMEM may fail to converge when too small values are provided as IEs for residual error parameters; thus, it is recommended to give sufficiently large values for them.

Sufficiency of data

Sufficiency of the dataset is vital for obtaining reliable FPE. If a dataset is dense enough to estimate all parameters used in the model structure, FPEs of all parameters may be reasonably estimated regardless of the values of the IEs. However, in many cases of modeling, data for the estimation of one or more parameter(s) may be insufficient. This problem arises from insufficient observations or overparametrization. When a given dataset includes dense data in the elimination phase (e.g., >5 observations per subject) while having insufficient information in

the absorption phase (e.g., the first observation is the maximum concentration), parameters related to the absorption process cannot be robust. This is also true for *CL* when the number of observations are too small in the elimination phase. In such cases, a high level of ETA shrinkage may be observed, and more seriously, the FPE for THETAs of the corresponding parameter may be identical or similar to the IE value. When this happens, a modeler should not try to estimate its BSV, and pay more attention to the quality of the IE for THETA. Sometimes, the IE for THETA may be imported from the literature and fixed as the FPE. This procedure may minimize the undesirable influence of inestimable parameters on the overall model fitting and on the other parameter estimates. Of course, this is not a desirable situation, but modelers sometimes encounter such datasets.

When the structure of a PK-PD model is so complex that all parameters cannot be estimated appropriately, the FPE of a certain parameter may shrink to a very small value (e.g., 1/100 of the IE value). An irrelevant IE value of a parameter may cause a problem for another parameter; however, even when the FPE of a problematic parameter is obtained by some adjustment to the IE values, its precision must be low. Thus, the only solution for this situation is to simplify the model structure rather than to repeatedly modify IEs.

Summary

Modelers should use appropriate IE values considering the

intended model structure, the meaning of each parameter, the amount of data, and physiological relevance. With a good set of IEs, the probability of convergence and the efficiency of modeling procedures (e.g. shortening the runtime, avoiding unnecessary efforts) may be improved. The understandings on the roles, meanings and implications of IEs even help modelers in troubleshooting situations which frequently occur over the whole modeling process.

Conflict of interest

The author declared no conflict of interest.

References

1. Beal SL, Sheiner LB. NONMEM Users Guide – Part I. Users Basic Guide. NONMEM Project Group 1989.
2. Choi K, Yim DS. Statistical basis for pharmacometrics: maximum likelihood estimator and its asymptotics. *Transl Clin Pharmacol* 2015;23:8-14. doi: 10.12793/tcp.2015.23.1.8
3. Nonlinear models and regression. In: Bonate PL (ed) *Pharmacokinetic-pharmacodynamic modeling and simulation*, 2nd ed. Springer Science+Business Media, LLC, New York, 2011;101-130
4. Pharmacokinetic Concepts. In: Gabrielsson J, Weiner D. (ed) *Pharmacokinetic & pharmacodynamic data analysis*, 5th ed. apotekarsocieteten, Stockholm, Sweden, 2016;13-198.
5. Pharmacodynamic Concepts. In: Gabrielsson J, Weiner D. (ed) *Pharmacokinetic & pharmacodynamic data analysis*, 5th ed. apotekarsocieteten, Stockholm, Sweden, 2016;199-332.