Journal of Pharmacokinetics and Biopharmaceutics, Vol. 24, No. 3, 1996

Parameter Identifiability Is Required in Pooled Data Methods

John A. Jacquez[1]

Received September 26, 1995—Final March 12, 1996

In pooled data methods such as naive pooled data methods and NONMEM, the number of sample points per individual may be less than the number of unknown parameters so that the values of the parameters are not estimable in individuals. However, for the moments of the distributions of the parameters to be estimable, the basic parameters must be identifiable.

KEY WORDS: identifiability; pooled data methods; two-stage methods.

INTRODUCTION

In pharmacokinetics two general classes of methods for the estimation of the distributions of the parameters of models of drug distribution and disposition can be distinguished (2,3). These are the two-stage (TS) and the pooled-data (PD) methods. In the TS methods sufficient data are obtained on each individual to estimate the values of the parameters of the model for each individual. The data on the individuals are then aggregated to obtain the population moments, usually the means and the covariance matrix. In PD methods one usually does not have enough data points on each individual to be able to estimate the parameter values for each individual; these methods estimate the first few population moments directly on the pooled data without first estimating the values of the parameters for individuals.

There has been some confusion over whether or not the basic parameters must be identifiable for the pooled data methods to work. In part, that may be attributed to imprecise use of the terms and ideas about identifiability. In this note, I first present basic theory of identifiability of parameters and define terms. That is followed by a section in which I prove that the

[1]Departments of Biostatistics and Physiology, The University of Michigan, Ann Arbor, Michigan 48109-0622.

0090-466X/96/0600-0301$09.50/0

basic parameters must be uniquely or locally identifiable for pooled data methods to work.

IDENTIFIABILITY

Theory

Suppose the dynamic equations for the system with state variables x_i are

$$\dot{x} = F(x, \theta, t); \qquad x(0) = x_0. \tag{1}$$

Here x is the vector of state variables and θ is the vector of basic parameters which are usually basic kinetic parameters. One observes some functions of the state variables called the observation functions or the response functions. There may be more than one observation function for an experiment but to keep this simple we shall assume only one. The argument carries over directly to multiple observation functions. Let y_i be the observation function for the ith individual.

$$y_i = G(x, \theta_i, t). \tag{2}$$

The actual observations are samples of the observation function at certain times, t_j, with added experimental error, ε_{ij}.

$$z_{ij} = G(x, \theta_i, t_j) + \varepsilon_{ij} \tag{3}$$

The observation functions can also be written in terms of a set of compound parameters, ϕ, that are uniquely determined by the observation function, i.e., uniquely identifiable; these are called the observational parameters (1,4,5)

$$y_i = G(x, \phi_i, t) \tag{4}$$

Suppose there are K observational parameters ϕ_k which are of course functions of the basic parameters. Let Φ_{ki} be the value of the kth observational parameter in the ith individual. Thus we have a set of K simultaneous algebraic equations relating the observational parameters to the basic parameters

$$\Phi_{ki} = \phi_{ki}(\theta_i). \tag{5}$$

As an example, consider the observation function for the one-compartment model for concentration of a drug in plasma as given by

$$y_i = \frac{D_i}{V_i} e^{-(C_i/V_i)t} \tag{6}$$

which has two identifiable parameters. For the ith individual these are ϕ_{1i} and ϕ_{2i}, which are the following functions of the basic kinetic parameters

$$\phi_{1i}=\frac{D_i}{V_i} \tag{7}$$

$$\phi_{2i}=\frac{C_i}{V_i} \tag{8}$$

If D_i is known, C_i and V_i are uniquely determined by the values of ϕ_{1i} and ϕ_{2i}. However, if D_i, C_i, and V_i are unknown, there is no way to obtain unique solutions for D_i, C_i, and V_i from the values of ϕ_{1i} and ϕ_{2i}; in that case, D_i, C_i, and V_i are not identifiable.

Terminology

If the transformation $\phi \mapsto \theta$ is one-to-one, the basic parameters are uniquely or globally identifiable. If the transformation is one-to-many, the parameters are locally identifiable. I use the term *identifiable* to mean locally or globally identifiable. If the transformation has an infinite number of solutions for θ, the basic parameters are unidentifiable. When we examine the properties of the transformation $\phi \mapsto \theta$ we are looking at a priori identifiability (6). I note that it is possible for some of the components of θ to be identifiable and others to be unidentifiable (1,5).

Notice that a priori identifiability has nothing to do with how many samples are taken, it is concerned only with the nature of the transformation from the ϕ to the θ. But, if there are p basic parameters and one takes $m<p$ samples of the observation function, the parameters cannot be estimated even if they are identifiable. That is an entirely different issue and has been called a posteriori or *numerical* identifiability (6). Because of the possibility of confusion, I prefer the term *estimability* for the issues involved in numerical estimation of values for the parameters θ (1,5). Thus the parameters θ must be (a priori) identifiable but if insufficient samples are taken, they may not be estimable (a posteriori identifiable). We distinguish between nonidentifiable parameters (NI), identifiable but not estimable (INE), and identifiable and estimable (IE) parameters. In addition, even if the parameters are IE, the sample may be poorly chosen and thus give poor estimates. That is the issue of optimal sampling design and does not concern us here.

THEORY FOR POOLED DATA METHODS

It is obvious that for the TS methods, the basic parameters must be identifiable and estimable in individuals (IE). Is that also true for PD methods? For PD methods, it turns out that the basic parameters must be

identifiable but need not be estimable in individuals, i.e., can be IE or INE. To see that, consider again the mapping from the basic parameters to the observational parameters. The observational parameters are by definition uniquely identifiable. Suppose we have the distribution function for the ϕ parameters. The theory of transformation of distributions for functions of random variates, such as

$$\phi = f(\theta) \tag{9}$$

requires that the mapping be one-to-one to obtain the distribution function for the basic parameters (7). Otherwise, one cannot obtain the distribution function for the basic parameters and hence their moments are not uniquely defined. But that is the same as the requirement for unique identifiability of the basic parameters. What if the basic parameters are only locally identifiable? The transformation $\phi \mapsto \theta$ is locally one-to-one, so if one picks one of the solutions of the transformation, one can again obtain the distribution function for the basic parameters. The basic parameters must be identifiable!

Remark. An important model for the study of drug distribution is the three-compartment mammillary model in which the central compartment is the amount in the plasma or blood and the peripheral compartments are the amounts in pooled peripheral tissues, usually a fast exchanging group and a slowly exchanging group. For observation of the amount or concentration in the central pool, the parameters are only locally identifiable. But this is a special type of local identifiability which arises because the two peripheral compartments play symmetrical roles in the theory, i.e., they are interchangeable. As a result there are two solutions for the rate constants between the central and peripheral compartments which have the same values but are obtained by exchanging the two peripheral compartments. One need only choose one of the solutions and proceed.

In PD methods, one estimates the moments of the distribution of the parameters but not the parameter values in individuals. Thus the basic parameters must be identifiable but need not be estimable in individuals. To return to our example, from Eqs. (7) and (8), if one of the parameters, C_i, D_i, and V_i is known, the others are identifiable. Then, even if one takes only one sample for each of many individuals, one can determine the moments of the population distribution of the remaining two parameters. However, if all three are unknown, none are identifiable and one cannot estimate the three parameters no matter how many samples of the observation function, Eq. (6), are taken.

CONCLUSIONS

For the PD methods, as well as for the TS methods, the basic parameters must be identifiable for one to be able to estimate the population means

and covariance matrix. However, for PD methods there need not be enough sample points per individual to estimate the values of the parameters in individuals. That distinguishes the PD methods from TS methods.

REFERENCES

1. J. A. Jacquez and T. Perry. Parameter estimation: local identifiability of parameters. *Am. J. Physiol.* **258:**E727–E736 (1990).
2. L. B. Sheiner, B. Rosenberg, and V. V. Marathe. Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *J. Pharmacokin. Biopharm.* **5:**445–479 (1977).
3. J.-L. Steimer, A. Mallet, and F. Mentré. Estimating interindividual pharmacokinetic variability. In M. Rowland, L. B. Sheiner, and J.-L. Steimer (eds.), *Variability in Drug Therapy, Description, Estimation, and Control*, Raven Press, New York, 1985, pp. 65–111.
4. J. A. Jacquez. Identifiably: the first step in parameter estimation. *Fed. Proc.* **46:**2477–2480 (1987).
5. J. A. Jacquez and P. Greif. Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design. *Math. Biosci.* **77:**201–227 (1985).
6. E. Walter. *Identifiability of State Space Models*, Vol. 46, Lecture Notes in Biomathematics, Springer-Verlag, Berlin, 1982.
7. S. S. Wilks. *Mathematical Statistics*, Wiley, New York, 1962.