*Journal of Pharmacokinetics and Biopharmaceutics, Vol. 22, No. 5, 1994*

# Comparison of the Akaike Information Criterion, the Schwarz Criterion and the *F* Test as Guides to Model Selection

**Thomas M. Ludden,[1,3] Stuart L. Beal,[2] and Lewis B. Sheiner[2]**

Received May 14, 1993—Final July 18, 1994

*In pharmacokinetic data analysis, it is frequently necessary to select the number of exponential terms in a polyexponential expression used to describe the concentration–time relationship. The performance characteristics of several selection criteria, the Akaike Information Criterion (AIC), and the Schwarz Criterion (SC), and the* F *test ($\alpha=0.05$), were examined using Monte Carlo simulations. In particular, the ability of these criteria to select the correct model, to select a model allowing estimation of pharmacokinetic parameters with small bias and good precision, and to select a model allowing precise predictions of concentration was evaluated. To some extent interrelationships among these procedures is explainable. Results indicate that the* F *test tends to choose the simpler model more often than does either the AIC or SC, even when the more complex model is correct. Also, the* F *test is more sensitive to deficient sampling designs. Clearance estimates are generally very robust to the choice of the wrong model. Other pharmacokinetic parameters are more sensitive to model choice, particularly the apparent elimination rate constant. Prediction of concentrations is generally more precise when the correct model is chosen. The tendency for the* F *test ($\alpha=0.05$) to choose the simpler model must be considered relative to the objectives of the study.*

**KEY WORDS:** Akaike information criterion (AIC); Schwarz criterion; *F* test; model selection.

## INTRODUCTION

Sums of exponentials are often used in pharmacokinetics to provide compartmental (1) or noncompartmental (2,3) descriptions of

This work was supported in part by Grant No. 26676 from the National Institutes of General Medical Sciences, National Institutes of Health.

[1]Division of Biopharmaceutics, Office of Research Resources, Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, Maryland.

[2]Department of Laboratory Medicine, The University of California–San Francisco, San Francisco, California.

[3]To whom all correspondence should be addressed at Division of Biopharmaceutics, Room 13B06, HFD-420, Food and Drug Administration, 5600 Fishers Lane, Rockville, Maryland 20857.

concentration–time profiles. In noncompartmental methods the fitting of exponential equations to data can be considered as a type of smoothing procedure. When using sums of exponentials the data analyst is often confronted with choosing between two or more possible descriptions, for example, a bi- vs. a triexponential equation. Boxenbaum *et al.* (4) described the use of the $F$ test to choose among models with differing numbers of parameters and apply it to the choice of model for the disposition of isoniazid. Yamaoka *et al.* (5) introduced the Akaike Information Criterion (AIC) (6,7) into pharmacokinetic data analysis. They compared the performance of the AIC to the $F$ test and found that the two methods agree for the examples they used. The Schwarz Criterion (SC) (8) can similarly be used in model selection. It has a Bayesian derivation that has been questioned by at least one author (9). Imbimbo and colleagues (10) summarized both the subjective and objective methods used to assist in model selection, including the AIC and SC and $F$ test, and offered a new criterion that chooses a model based on the area between the confidence bounds for model-predicted concentrations. In addition, this group also investigated the influence of the number of sampling times and the difference between the values of exponential constants on the performance of the criteria (11).

Independently, we have also compared the performance of the AIC, SC, and $F$ test using a number of performance measures. Our findings complement and extend those of Imbimbo *et al.* (10,11) and illustrate that application of any of the three criteria can lead to the choice of the wrong model, result in biased or imprecise parameter estimates, or result in poor predictions of true concentrations.

## METHODS

Six cases are studied using Monte Carlo simulation. Three cases (I–III) explore the selection of a mono- vs. a biexponential model, and two cases (V–VI) explore the selection of a bi- vs. triexponential model. An additional biexponential case (IV) is studied.

In describing models the following symbolism is used

$$f(t) = \sum_{i=1}^{N} A_i e^{-\lambda_i t}$$

where $f(t)$ is the predicted concentration, $t$ is time, $N$ is the number of exponential terms and the $A_i$s and $\lambda_i$s are the linear and exponential constants, respectively, with $\lambda_i > \lambda_{i+1}$.

For Cases I–III, the linear constants for the biexponential model, $A_1$ and $A_2$, are fixed at 0.7 and 0.3. The slowest exponential constant, $\lambda_2$, is

fixed to 0.693, so that time units are expressed in the number of half-lives of this terminal log-linear slope. The value of $\lambda_1$ is twice (Case I), three times (Case II), and six times (Case III) the value of $\lambda_2$. In each case the parameters of the simpler monoexponential model are chosen so that the first two moments of the disposition function are the same as those for the biexponential model (12). For Case IV, $A_1$, $A_2$, $\lambda_1$, $\lambda_2$ are set equal to 0.950, 0.05, 6.93, and 0.693, respectively.

For Case V, the $A_1$, $A_2$, and $A_3$ values for the triexponential model are set to 0.7, 0.2, and 0.1, respectively, and $\lambda_3$ is set to 0.693. The $\lambda_2$ value is three times $\lambda_3$, and $\lambda_1$ is three times $\lambda_2$. For Case VI, the $\lambda_2$ value is five times $\lambda_3$, $\lambda_1$ is five times $\lambda_2$, and the $A_i$ values are determined so as to sum to 1 while each exponential term of the model has equal area, i.e.,

$$\frac{A_1}{\lambda_1}=\frac{A_2}{\lambda_2}=\frac{A_3}{\lambda_3}; \qquad \sum_{i=1}^{3} A_i=1$$

For each triexponential model, a competing biexponential model is determined so that its moments are equal to those of the triexponential model (12). These cases are summarized in Table I and Fig. 1. Cases I, II, and V are somewhat pathological in that in each case the differences between the exponential constants are not large.

A standard sampling design of 16 points is used for all cases, and various modified designs consisting of fewer data points are examined in conjunction with specific cases. The standard design and its modifications are summarized in Table II.

**Table I.** Summary of Cases

| Case | Exponential terms | $A_1$ | $\lambda_1$ | $A_2$ | $\lambda_2$ | $A_3$ | $\lambda_3$ |
|---|---|---|---|---|---|---|---|
| I | 1 | 0.885 | 0.947 | | | | |
| | 2 | 0.7 | 1.39 | 0.3 | 0.693 | | |
| II | 1 | 0.752 | 0.977 | | | | |
| | 2 | 0.7 | 2.08 | 0.3 | 0.693 | | |
| III | 1 | 0.543 | 0.902 | | | | |
| | 2 | 0.7 | 4.16 | 0.3 | 0.693 | | |
| IV | 2 | 0.950 | 6.93 | 0.05 | 0.693 | | |
| V | 2 | 0.745 | 3.79 | 0.127 | 0.707 | | |
| | 3 | 0.7 | 6.24 | 0.2 | 2.08 | 0.1 | 0.693 |
| VI | 2 | 0.949 | 6.79 | 0.032 | 0.702 | | |
| | 3 | 0.82 | 17.3 | 0.15 | 3.47 | 0.03 | 0.693 |

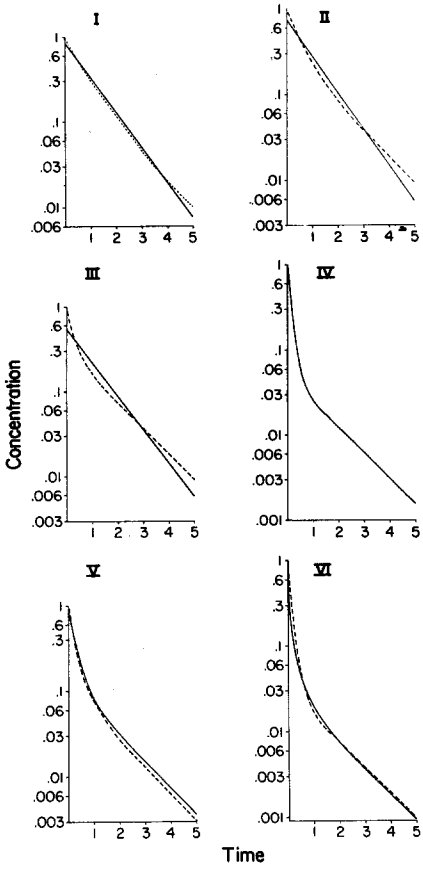


**Fig. 1.** The concentration–time profiles for the two models used for each case (I–VI) are illustrated. The solid line represents the simpler model and the broken line the more complex model. The logarithmic scale makes the models appear more similar than they really are.

Statistical error is simulated according to the approximate constant coefficient of variation ($CV$) error model

$$\ln y = \ln f(t) + \varepsilon$$

This model was chosen because it provides a good description of the measurement error included in $\varepsilon$, generated by most chromatographic analytical methods. Simulation of $\varepsilon$ was realized using the Box–Muller method (13) for producing pseudorandom normal deviates. The standard deviation of $\varepsilon$ was fixed to 0.15. Each case and design was replicated with 250 data sets.

The nonlinear regression analysis program uses ordinary least squares to fit the log-transformed simulated data ($\ln y$) to competing models. A Marquardt search algorithm (14) was used with convergence defined as a

**Table II.** Summary of Sampling Designs

| | Sampling points | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Design | 0.02 | 0.05 | 0.08 | 0.10 | 0.12 | 0.17 | 0.25 | 0.30 | 0.50 | 0.75 | 0.1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 | 5 |
| S | X | X | X | | X | X | | X | X | X | X | X | X | X | X | X | X | X |
| 1 | X | X | X | | X | X | | X | X | X | X | X | X | X | X | X | | |
| 2 | | | X | | X | X | | X | X | X | X | X | X | X | X | X | X | X |
| 3 | | X | X | | X | X | | X | X | X | X | X | X | X | X | X | X | |
| 4 | | | X | | X | X | | X | X | X | X | X | X | X | X | X | | |
| 5 | | X | | | X | X | | X | X | X | X | X | X | X | X | X | | |
| 6 | | | | X | | | X | | X | X | X | X | X | X | X | X | | |
| 7 | | | | X | | | X | | X | X | X | X | X | X | | | | |
| 8 | | | | | | X | | X | X | X | X | X | X | X | X | X | X | X |

change in all parameter values of 0.1% or less. Initial estimates were chosen using a grid search of the parameter space based on the IMSL subroutine ZSRCH.

The expressions for the model selection criteria are

$$AIC = N \ln SS + 2p$$

$$SC = N \ln SS + p \ln N$$

$$F\,\text{ratio} = \frac{SS_j - SS_k}{SS_k} \frac{df_k}{df_j - df_k}; \qquad \text{with } df_j > df_k$$

where $N$ is the number of design points, $p$ is the number of estimated parameters, $SS$ is the sum of squared residuals, and $df = N - p$ is the degrees of freedom. The subscripts $j$ and $k$ represent the simpler and more complex models, respectively.

When the $AIC$ and $SC$ are applied, the model producing the lowest value is chosen. For the $F$ test, the $F$ ratio is compared to a table of critical values for the $F$ distribution with $(df_j - df_k)$, $df_k$ degrees of freedom. Each critical value corresponds to a so-called Type I error, $\alpha$. For all applications of the $F$ test described here $\alpha$ is 0.05.

The performances of the AIC, SC, and the $F$ test are evaluated using (i) percentage of correct model selection; (ii) mean (parameter estimation) error ($ME$) and mean absolute (parameter estimation) error ($MAE$) expressed as a percentage of the true value of the parameter: clearance ($CL$), steady state volume of distribution ($Vss$), terminal slope ($\lambda_z$), and mean residence time ($MRT$); and (iii) prediction error as measured by the mean overall sampling times of the squared deviations between natural logarithms of the predicted and true concentration values.

Since, for all models, unit dose was assumed, clearance is calculated as

$$CL = \frac{1}{\sum_{i=1}^{N} A_i/\lambda_i}$$

$$Vss = \frac{\sum_{i=1}^{N} A_i/\lambda_i^2}{(\sum_{i=1}^{N} A_i/\lambda_i)^2}$$

$$MRT = Vss/CL$$

## RESULTS

### Selection of the Correct Model

The comparative performance of the three criteria in selecting the correct model is illustrated in Table III. The $F$ test ($\alpha = 0.05$) is biased toward selection of the simpler model. The biexponential equation for Case I had exponential constants with a ratio of only 2 and presents a difficult model selection problem. Yet, the AIC and SC perform reasonably well when the standard design is used and the biexponential is the correct model (Table III). However, a less complete sampling profile degrades the performance of these criteria (Case I, design 1). When the ratio of exponential constants is increased to 3 (Case II), the performance of the $F$ test improves. However, the performance of the $F$ test is seriously compromised when the biexponential model is correct and deficient sampling designs are used. Overall, the AIC and SC much more often select the correct model in these situations (Table III). For Case III the ratio of exponential constants is further increased to 6. With the standard design all criteria perform reasonably well, with the AIC exhibiting the worst performance when the monoexponential model is correct. However, it is once again possible to seriously degrade the performance of the $F$ test when the biexponential model is correct by restricting the design points. A similar pattern is revealed for Case IV simulations where the ratio of the exponential constants is 10. The AIC and SC continue to perform well, whereas the $F$ test can fail to select the correct model. The relative performances of the three criteria are similar when applied to the discrimination between bi- and triexponential models. When the exponential constants have ratio 3 (Case V), none of the criteria perform well when the more complex model is correct. When the ratio is increased to 5, the performance of the AIC and SC in this respect improves, but the $F$ test still exhibits very poor performance. For Case III, designs 6 and 7, and for Case IV, all designs, the fits to the biexponential model when the monoexponential

**Table III.** Percentage of Converged Replications for Which the Correct Model Was Chosen[a]

| Case | Sampling design[b] | Correct model[c] | Converged replications | Nonconverged replications with correct model | % correct selection | | |
|---|---|---|---|---|---|---|---|
| | | | | | AIC | SC | *F* test |
| I | S | 1 | 104 | 0 | 80.8 | 88.5 | 100.0 |
| | | 2 | 222 | 28 | 79.3 | 71.6 | 11.3 |
| | 1 | 1 | 88 | 0 | 75.0 | 86.4 | 100.0 |
| | | 2 | 171 | 79 | 53.6 | 40.0 | 0.6 |
| II | S | 1 | 146 | 0 | 80.1 | 93.2 | 100.0 |
| | | 2 | 250 | 0 | 99.6 | 99.6 | 76.0 |
| | 1 | 1 | 147 | 0 | 92.5 | 94.6 | 100.0 |
| | | 2 | 228 | 22 | 95.2 | 92.5 | 2.2 |
| | 2 | 1 | 115 | 0 | 87.0 | 88.7 | 100.0 |
| | | 2 | 249 | 1 | 100.0 | 100.0 | 18.5 |
| | 3 | 1 | 105 | 0 | 87.6 | 94.5 | 100.0 |
| | | 2 | 241 | 9 | 99.2 | 98.8 | 13.3 |
| | 4 | 1 | 121 | 0 | 78.5 | 84.3 | 100.0 |
| | | 2 | 217 | 33 | 95.9 | 93.5 | 0.0 |
| | 5 | 1 | 78 | 0 | 84.9 | 88.5 | 100.0 |
| | | 2 | 237 | 13 | 99.2 | 97.5 | 0.0 |
| III | S | 1 | 128 | 0 | 85.9 | 94.5 | 100.0 |
| | | 2 | 250 | 0 | 100.0 | 100.0 | 98.8 |
| | 6 | 2 | 243 | 7 | 100.0 | 100.0 | 0.0 |
| | 7 | 2 | 227 | 23 | 97.8 | 97.8 | 0.0 |
| IV | S | 2 | 250 | 0 | 100.0 | 100.0 | 100.0 |
| | 4 | 2 | 250 | 0 | 100.0 | 100.0 | 100.0 |
| | 8 | 2 | 250 | 0 | 100.0 | 100.0 | 0.0 |
| V | S | 2 | 202 | 0 | 92.1 | 95.0 | 100.0 |
| | | 3 | 184 | 66 | 37.5 | 26.1 | 0.0 |
| | 1 | 2 | 154 | 0 | 93.5 | 95.5 | 100.0 |
| | | 3 | 153 | 97 | 30.1 | 20.3 | 0.0 |
| VI | S | 2 | 132 | 0 | 85.6 | 93.2 | 100.0 |
| | | 3 | 225 | 25 | 97.8 | 95.6 | 4.4 |
| | 1 | 2 | 143 | 4 | 88.1 | 91.0 | 100.0 |
| | | 3 | 218 | 32 | 95.3 | 92.2 | 0.0 |
| | 2 | 2 | 119 | 0 | 83.2 | 88.2 | 100.0 |
| | | 3 | 215 | 35 | 75.3 | 68.4 | 0.0 |

[a]A converged replication is one for which convergence is achieved with both correct and incorrect models.
[b]See Table II.
[c]Refers to number of exponential terms, See Table I.

is correct often did not converge. Therefore, no results for these situations are included in Table III.

Inspection of Table III indicates that when the more complex model is correct, the AIC selects it more often than does the SC, and when the simpler model is correct, the SC selects it more often than does the AIC. This behavior follows from the fact that in all designs studied, for any data set,

if the SC (AIC) selects the more complicated (simpler) model, then so does the AIC (SC). This fact in turn follows from the proposition proved in the Appendix. With this proposition, results concerning the $F$ test along with the other two criteria also are explainable. Namely, when the more complex model is correct, these other criteria select this model more often, and when the simpler model is correct, the $F$ test selects it more often than do the AIC or SC.

**Accuracy and Precision of Parameter Estimates (Table IV)**

In pharmacokinetic data analysis the choice of the correct model may not be of greatest importance. Rather, accurate estimation of one or more pharmacokinetic parameters, or good predictions of concentrations under similar dosing circumstances may be the eventual goal. To assess the performance of the three model selection criteria for these goals, the bias and precision of various estimated parameters and the mean squared prediction error for the logarithm of true concentration were determined. In the following presentation of results for parameter estimation, specific mention is made of situations for which one or more criteria yield a $ME$ or $MAE$ greater than 10% (Table IV). Situations not described are those for which the $ME$ and $MAE$ are both less than 10%.

*Clearance*

For Cases I, II, and III clearance is accurately and precisely estimated using either the correct or incorrect model and for all designs tested. Therefore, model selection has little influence on the estimation of clearance.

For Case IV, design 8, the $F$ test yields highly biased estimates of clearance ($ME = 67\%$) because the initial exponential phase is essentially ignored. The clearance estimates based on AIC and SC are unbiased ($ME = 0.3\%$), and of modest precision ($MAE = 9.6\%$).

For Case V, accurate and precise estimates of clearance are provided by either model and, therefore, model selection has no influence, in spite of the fact that the $F$ test always chooses the wrong model when the triexponential model is the correct one.

For Case VI, accurate and precise estimates of clearance are provided by either model for designs S and 1. However, for design 2, the clearance estimate based on the $F$ test is biased ($ME = 13.3\%$) when the triexponential model is correct. The clearance estimates based on AIC and SC are unbiased ($ME = -1.4$ and $-0.6$, respectively) but relatively imprecise ($MAE = 13.9$ and 14.3%, respectively).

Overall, estimation of clearance is robust to model selection. Only poor designs yield biased estimates of clearance when the simpler model is chosen by the $F$ test over the more complex correct model.

**Table IV.** Bias and Precision of Parameter Estimates (%)[a]

| Case | Design | Correct model | Mean error | | | Mean absolute error | | |
|---|---|---|---|---|---|---|---|---|
| | | | AIC | SC | *F* test | AIC | SC | *F* test |
| Clearance | | | | | | | | |
| IV | 8 | 2 | −0.3 | −0.3 | 66.6 | 9.6 | 9.6 | 66.6 |
| VI | 2 | 3 | −1.4 | −0.6 | 13.3 | 13.9 | 14.3 | 13.5 |
| Steady state volume of distribution | | | | | | | | |
| I | 1 | 1 | 15.7 | 12.8 | 0.8 | 18.9 | 16.4 | 4.9 |
| | | 2 | 47.9 | 34.8 | −6.5 | 56.5 | 44.7 | 8.0 |
| II | 1 | 2 | 38.3 | 38.2 | −9.0 | 43.8 | 43.9 | 12.5 |
| | 3 | 2 | 13.8 | 13.8 | −2.8 | 18.6 | 18.6 | 9.3 |
| | 4 | 2 | 20.4 | 20.6 | −6.4 | 29.2 | 28.9 | 7.7 |
| | 5 | 2 | 18.8 | 18.7 | −4.2 | 24.3 | 24.3 | 6.5 |
| III | 7 | 2 | 17.4 | 17.4 | −5.4 | 28.3 | 28.3 | 8.5 |
| IV | 8 | 2 | 1.2 | 1.2 | 196.0 | 17.8 | 17.8 | 96.0 |
| VI | 1 | 3 | 2.1 | 2.1 | −0.9 | 11.9 | 11.9 | 6.8 |
| | 2 | 2 | −1.9 | −2.0 | 1.0 | 13.0 | 12.8 | 10.3 |
| | | 3 | 3.1 | 4.5 | 28.0 | 26.6 | 27.3 | 28.2 |
| Terminal slope | | | | | | | | |
| I | S | 2 | 6.0 | 8.4 | 31.7 | 25.7 | 26.3 | 34.6 |
| | 1 | 2 | 16.4 | 22.4 | 44.2 | 40.7 | 41.9 | 45.0 |
| II | S | 2 | −5.7 | −5.7 | 4.0 | 14.0 | 14.0 | 20.1 |
| | 1 | 2 | 0.6 | 1.2 | 50.4 | 26.8 | 27.4 | 51.7 |
| | 2 | 2 | −5.0 | −5.0 | 27.3 | 13.6 | 13.6 | 31.8 |
| | 3 | 2 | −7.1 | −7.0 | 36.6 | 21.1 | 21.1 | 42.7 |
| | 4 | 2 | −2.2 | −1.5 | 48.9 | 26.0 | 26.7 | 48.9 |
| | 5 | 2 | −8.0 | −7.4 | 44.9 | 23.7 | 23.8 | 44.9 |
| III | 6 | 2 | −3.0 | −3.0 | 30.6 | 11.2 | 11.2 | 30.6 |
| | 7 | 2 | −8.0 | −8.0 | 50.0 | 24.0 | 24.0 | 50.0 |
| IV | 8 | 2 | −0.6 | −0.6 | 37.4 | 4.9 | 4.9 | 37.4 |
| V | S | 3 | 2.0 | 4.2 | 10.1 | 12.8 | 12.4 | 11.1 |
| | 1 | 2 | −3.7 | −3.4 | −2.2 | 11.9 | 11.8 | 11.1 |
| | | 3 | 6.6 | 8.8 | 16.1 | 21.0 | 20.7 | 18.8 |
| VI | S | 3 | −3.1 | −2.6 | 14.6 | 8.5 | 8.7 | 15.3 |
| | 1 | 3 | −4.8 | −4.1 | 27.7 | 16.5 | 16.9 | 27.8 |
| | 2 | 3 | −3.7 | −2.8 | 8.8 | 10.7 | 10.9 | 10.2 |

[a]Only situations with one or more values greater than 10% are shown.

*Steady-State Volume of Distribution*

For Case I, design 1, more precise estimates of *Vss* are obtained from the models selected by the *F* test. Both the AIC and SC result in biased ($ME = 13$–$48\%$) and imprecise ($MAE = 16$–$57\%$) *Vss* estimates for this design. The performances of the AIC and SC are worse for the situation

where the biexponential model is correct than for the situation where the monoexponential is correct (Table IV).

For Case II, designs 1, 3, 4, and 5, when the more complex model is correct the AIC- and SC-selected models yield *Vss* estimates that are much less precise ($MAE = 19$–$44\%$) and more biased ($ME = 14$–$38\%$) than the *Vss* values based on *F* test-selected models ($MAE = 6$–$12\%$, $ME = 3$–$9\%$). These designs have one or two of the latest samples deleted. Designs S and 2 did not have missing data at late times, and model selection exhibits little influence on either bias or precision of *Vss* estimates.

For Case III, designs S and 6, there are no important effects of the selection criteria on bias or precision of *Vss* estimates. For design 7, with the correct model being the more complex, the AIC- and SC-selected models yield *Vss* estimates that are biased ($ME = 17.4\%$) and imprecise ($MAE = 28\%$), while the *F* test-selected models yield little bias ($ME = -5\%$) and reasonable precision ($MAE = 8\%$).

For Case IV, design 8, there is a large bias for *Vss* based on the models selected by the *F* test ($ME = 196\%$). This results from choosing the simpler, incorrect model. The AIC- and SC-selected models result in virtually no bias in *Vss* ($ME = 1.2\%$), but precision is rather poor ($MAE = 17.8\%$).

For Case V, there is no major influence of selection criterion on the estimate of *Vss*.

For Case VI, designs S and 1, the results are generally similar to those described for Case V. For design 2, when the triexponential model is correct, the biexponential model selected by the *F* test yields biased estimates of *Vss* ($ME = 28\%$), while the models selected by the AIC and SC are unbiased ($ME = 3$ and $4\%$, respectively). Using each of the criteria, estimates of *Vss* are imprecise ($MAE = 27$–$28\%$).

*Mean Residence Time*

The results for estimates of the *MRT* are essentially the same as those for *Vss*.

*Terminal Slope*

For Case I, design S, when the biexponential model is correct, the *F* test yields biased estimates of $\lambda_z$ ($ME = 32\%$), while the AIC- and SC-based estimates are less biased ($ME = 6$–$8\%$) but still imprecise ($MAE = 26\%$). Results for design 1 are qualitatively similar, but the biases for all selection criteria are greater (Table IV).

Case II, designs 1 through 5, all yield very biased estimates of $\lambda_z$ ($ME = 27$–$50\%$) when the *F* test is applied and the biexponential model is correct. While corresponding models based on the AIC and SC yield $\lambda_z$ estimates with much less bias, they are still relatively imprecise ($MAE = 14$–$27\%$).

Case III, designs 6 and 7, Case IV, design 8, Case V, designs S and 1, and Case VI, designs S and 1, provide results that are qualitatively similar to those described for Case II, designs 1–5 (Table IV).

In general, choice of the simpler model yields biased estimates of the terminal slope when the more complex model is correct. This occurs most frequently with the *F* test.

### Prediction of True Concentrations

The prediction of true concentrations, as measured by the mean squared difference between the logarithm of the predicted and the logarithm of the true concentration, is notably worse when the *F* test chooses the simpler model but the complex model is correct in the following situations: Case II, all designs; Case III, designs 6–7; Case IV, design 8; and Case VI, designs S and 1 (Table V). "Notably worse" here means a two-fold or more difference in the mean squared error. For Cases I and V, the predicted and true concentrations are very similar, and model choice has little influence on this measure of performance.

## DISCUSSION

Our results indicate that when exponential models are distinct and sampling designs are good, all three criteria, the AIC, SC and *F* test, perform well (Table III). These observations are similar to those of Yamaoka *et al.* (5) who introduced the AIC into pharmacokinetics and described some

**Table V.** Sum of Squared Errors for Predicting the Logarithm of the True Concentration[a]

| Case | Design | AIC[b] | *F* test |
|---|---|---|---|
| II | S | 0.099 | 0.238 |
| | 1 | 0.099 | 0.350 |
| | 2 | 0.097 | 0.499 |
| | 3 | 0.096 | 0.402 |
| | 4 | 0.096 | 0.300 |
| | 5 | 0.096 | 0.391 |
| III | 6 | 0.091 | 0.524 |
| | 7 | 0.090 | 0.377 |
| IV | 8 | 0.093 | 3.05 |
| VI | S | 0.137 | 0.383 |
| | 1 | 0.141 | 0.318 |

[a]Only values that differ by at least twofold between AIC and *F* test when the correct model is the more complex model are tabled.

[b]SC results are essentially the same as those for AIC.

limited comparisons of its behavior to that of the $F$ test. Since the $F$ test has as the null hypothesis that the simpler model is correct, it takes reasonably strong evidence for this null hypothesis to be rejected at $\alpha = 0.05$. Thus, if the true and hypothesized models do not differ greatly, or the sampling design is poor, then the tendency of the $F$ test to select the correct, more complex model declines. For example, Case IV is based on a biexponential model with exponential constants that are substantially different, i.e., they differ by 10-fold (see Table I). Design 8, although clearly less informative than the standard design or design 4, still has 3 to 4 data points in the initial, rapid phase of the concentration–time curve. For design 8, the $F$ test always leads to the selection of the incorrect model while the AIC and SC always lead to the selection of the correct model. The power of the $F$ test in this context has been addressed in more detail by Imbimbo *et al.* (11), Burguillo *et al.* (15), and Bardsley *et al.* (16). Bardsley *et al.* (16) conclude that the $F$ test is useful for selecting multiexponential models with up to two or three exponential terms provided that the exponential constants are sufficiently different and the design is good.

Since the ultimate goal of fitting models to pharmacokinetic data is often to estimate parameters or to predict concentrations, the influence of the selection criteria on such estimates/predictions was also investigated. We find that estimates of clearance are robust to the choice of model. A notable bias in the clearance estimate is seen in only two situations (Case IV, design 8 and Case VI, design 2), in both of which sampling designs are very poor. In these cases model selection based on the $F$-test results in modestly biased clearance estimates, while selection based on the AIC and SC does not.

When the more complex model is in fact correct, estimates of *Vss* and *MRT* are often less biased if the simpler, incorrect model is chosen. This is particularly true if the more complex model is not markedly different from the simpler model. However, in one situation where the more complex model is very distinct from a simpler model, selection of the simpler, incorrect model results in highly biased estimates of *Vss* and *MRT* (Case IV, design 8). Therefore, in regard to *Vss* and *MRT*, no particular criterion consistently performs better than all others.

Of all the parameters examined, the terminal slope is the most sensitive to model selection. Choice of the correct model almost always yields the least biased and most precise estimates of $\lambda_z$. As expected, the largest biases are found when the more complex model is correct and model selection is based on the $F$ test, which tends to choose the simpler model.

Prediction of true concentrations is almost always better when the correct model is chosen. The largest differences are again noted when the more complex model is true but the simpler model is chosen. Thus, the $F$ test is

inferior to the AIC and SC by this performance measure, especially when model differences are not trivial and the sampling design is poor.

The performances described here are based on multiexponential, bolus input models fit to data with correct weighting. Other designs, input functions, models, or weighting schemes could yield different results. Consider for example, the simulations performed by Imbimbo *et al.* (11), which differed from ours in several respects. (i) One- and two-compartment open models with either bolus or first-order input were investigated, but not a three-compartment bolus input model. (ii) Whereas Imbimbo *et al.* compared three models when a biexponential model was correct, we compared only two models. (iii) The residual error in their simulations was 5%, whereas we assumed a value of 15%. (iv) The designs used by Imbimbo *et al.* are described as having "sampling prolonged up to four elimination half-lives and sampling times were chosen on a logarithmic basis." It is not clear if the designs with small numbers of samples covered the same time-span as the designs with 15 or 20 samples. We tended to focus primarily on pathological situations caused by small differences in exponential constants and relatively poor designs. Imbimbo *et al.* (11) concluded that overall, the *F* test and their proposed new criterion are better than either the AIC or SC in terms of correct model selection. However, they also found, as we did, that the performance of the *F* test diminished relative to that of the AIC or SC when values of exponential constants differ by a ratio of 2 (11).

It is clear that the *F* test, AIC, and SC do not always agree in their choice of model. When the design is good, and the exponential constants are well separated, the *F* test may be better than the AIC or SC with respect to correct model selection. However, when the competing models are very similar, or when the design is inadequate, the *F* test has a strong tendency to choose the simpler model, even when the more complex model is correct. Imbimbo *et al.* (10,11) also noted that the *F* test performed poorly for ill-conditioned equations. In applications to real data sets (for which, of course, the true model is unknown), and where there is some extra risk associated with choosing the simpler model, i.e., when the best estimate of terminal half-life is desired, it wouild be prudent to use the AIC or SC. This would be especially true when exponential constants appear to be poorly differentiated or the sampling design is thought to be deficient.

## APPENDIX

### Proposition

Let $j$, $k$ refer to the simpler and more complex models, respectively.

(A) Suppose $\ln N > 2$ (i.e., $N > 7$). Then if the SC selects Model $k$, so does the AIC.

(B) Let

$$\lambda_0 = \left(\frac{p_k - p_j}{N - p_k} F_0 + 1\right)^{-N/2}$$

where $F_0$ is the critical value of the $F$ distribution used with the $F$ test. Suppose $-\ln \lambda_0 > p_k - p_j$. Then if the $F$ test selects Model $k$, so does the AIC.

(C) Suppose $-\ln \lambda_0 > 0.5\,(p_k - p_j) \ln N$. Then if the $F$ test selects Model $k$, so does the SC.

**Proof**

(A) Suppose the SC selects Model $k$, i.e.,

$$N \ln SS_k + p_k \ln N < N \ln SS_j + p_j \ln N$$

then

$$N \ln SS_k < N \ln SS_j - (p_k - p_j) \ln N$$

so

$$N \ln SS_k + 2p_k < N \ln SS_j - (p_k - p_j) \ln N + 2p_k$$

by hypothesis

$$-(p_k - p_j) \ln N + 2p_k < 2p_j$$

(B) Let

$$\lambda = \left(\frac{p_k - p_j}{N - p_k} F + 1\right)^{-N/2}$$

where $F$ is the $F$ ratio, so that

$$-2 \ln \lambda = N(\ln SS_j - \ln SS_k)$$

Suppose the $F$ test selects Model $k$, i.e., $F > F_0$. Then $-\ln \lambda > -\ln \lambda_0$. So

$$N \ln SS_j > N \ln SS_k - 2 \ln \lambda_0$$

so

$$N \ln SS_j + 2p_j > N \ln SS_k + 2p_j - 2 \ln \lambda_0$$

by hypothesis

$$2p_j - 2 \ln \lambda_0 > 2p_k$$

(C) Suppose the $F$ test selects Model $k$.

From above

$$N \ln SS_j + p_j \ln N > N \ln SS_k + p_j \ln N - 2 \ln \lambda_0$$

by hypothesis

$$p_j \ln N - 2 \ln \lambda_0 > p_k \ln N$$

## REFERENCES


1. J. G. Wagner. Linear compartment models. In *Fundamentals of Clinical Pharmacokinetics*, Drug Intelligence Publications Inc., Hamilton, Ill., 1975, pp. 57–128.
2. J. G. Wagner. Do you need a pharmacokinetic model, and, if so, which one? *J. Pharmacokin. Biopharm.* **3:**457–478 (1975).
3. D. J. Cutler. Numerical deconvolution by least squares. Use of prescribed input functions. *J. Pharmacokin. Biopharm.* **6:**227–241 (1978).
4. H. G. Boxenbaum, S. Riegelman, and R. M. Elashoff. Statistical estimation in pharmacokinetics. *J. Pharmacokin. Biopharm.* **2:**123–148 (1972).
5. K. Yamaoka, T. Nakagawa, and T. Uno. Application of Akaike's Information Criterion (AIC) in the evaluation of linear pharmacokinetic equations. *J. Pharmacokin. Biopharm.* **6:**165–175 (1978).
6. H. Akaike. Information theory and an extension of the maximum likelihood principle, In B. N. Petrov and F. Csaki (eds.), *2nd International Symposium on Information Theory*, Akademai Kiado, Budapest, 1973, pp. 267–281.
7. H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19:**716–723 (1974).
8. G. Schwarz. Estimating the dimension of a model. *Ann. Statist.* **6:**461–468 (1978).
9. A. C. Atkinson. Posterior probabilities for choosing a regression model. *Biometrika* **65:**39–48 (1978).
10. B. P. Imbimbo, E. Imbimbo, S. Daniotti, D. Verotta, and G. Bassotti. A new criterion for selection of pharmacokinetic multiexponential equation. *J. Pharm. Sci.* **77:**784–798 (1988).
11. B. P. Imbimbo, P. Martinelli, M. Rocchetti, G. Ferrari, G. Bassotti, and E. Imbimbo. Efficiency of different criteria for selecting pharmacokinetic multiexponential equations. *Biopharm. Drug Dispos.* **12:**139–147 (1991).
12. S. L. Beal. Some clarification regarding moments of residence times with pharmacokinetic models. *J. Pharmacokin. Biopharm.* **15:**75–92 (1987).
13. G. E. P. Box and M. Muller. A note on the generation of random normal deviates. *Ann. Math. Statist.* **29:**610–611 (1958).
14. D. W. Marquardt. An algorithm for least squares estimation of nonlinear parameters. *SIAM J.* **11:**431–441 (1963).
15. F. J. Burguillo, A. J. Wright, and W. G. Bardsley. Use of the *F*-test for determining the degree of enzyme-kinetic and ligand-binding data. A Monte Carlo simulation study. *Biochem. J.* **211:**23–34 (1983).
16. W. G. Bardsley, P. B. McGinlay, and A. J. Wright. The *F*-test for model discrimination with exponential functions. *Biometrika* **73:**501–508 (1986).