The R Package PK for Basic Pharmacokinetics

Martin J. Wolfsegger[1]

[1] Department of Biostatistics, Baxter AG, Vienna, Austria

Address of the author:
Martin J. Wolfsegger
Department of Biostatistics
Baxter AG
Wagramer Straße 17-19
1220 Vienna
Austria

Summary

Two-phase half-life estimation became popular in the recent years but there are no stand-alone functions in popular statistical standard software. A lot of people who are interested in pharmacokinetic (PK) analyses have a medical background but are less experienced in implementing complex fitting algorithms. To make R [1] more popular in their community a few functions for basic PK analyses are implemented in the R package PK [2]. The available functions for two-phase half-life estimation are presented in a simulation study.

Introduction

A common model for individual half-life estimation is the two-compartmental model assuming a distributive phase and elimination phase. The expected drug level $E(y_t)$ at time t for an individual can be expressed as a linear combination of two exponentials

$$E(y_t) = \alpha_1 e^{-\beta_1 t} + \alpha_2 e^{-\beta_2 t}$$

where β_1 and β_2 must be positive to be physically meaningful and $\beta_1 > \beta_2$ to ensure identifiability. Initial half-life representing the distribution of the drug and terminal half-life indicating the actual degradation of the material are calculated as

$$t_{1/2;1} = \frac{\log(2)}{\beta_1} \quad \text{and} \quad t_{1/2;2} = \frac{\log(2)}{\beta_2}.$$

A common approach is to use nonlinear fitting with the least squares criteria to estimate α_1, β_1, α_2 and β_2. The estimates obtained by this approach may depend on the fitting algorithms and on the starting values used. Lee et *al.* [3] presented a two-phase linear regression model by fitting two simple ordinary least squares regressions on log-linear transformed data based on the two sets of points $\{t_1,\ldots,t_i\}$ and $\{t_{i+1},\ldots,t_n\}$. The method of Lee results in n-3 sets of two regression lines with requesting at least 2 data points in the terminal phase. The estimates of the two sets of regressions are $\hat{\alpha}_{k1}, \hat{\beta}_{k1}$ and $\hat{\alpha}_{k2}, \hat{\beta}_{k2} \; \forall$ k=1,…,n-3.

The total sum of squares residuals is determined by

$$T_k = \sum_{j=1}^{i} \left(\log(y_j) - \hat{\alpha}_{k1} - \hat{\beta}_{k1} t_j\right)^2 + \sum_{j=i+1}^{n} \left(\log(y_j) - \hat{\alpha}_{k2} - \hat{\beta}_{k2} t_j\right)^2$$

In addition, the simple least squares regression equation for all n points is also calculated with

$$T_0 = \sum_{j=1}^{n} \left(\log(y_j) - \hat{\alpha} - \hat{\beta} t_j\right)^2$$

If the regression lines for initial and terminal phase do not join in the interval $[t_i,t_{i+1}]$ (i.e. $\hat{\gamma} \leq t_i$ or $\hat{\gamma} \geq t_{i+1}$) where

$$\hat{\gamma} = \frac{\hat{\alpha}_{k1} - \hat{\alpha}_{k2}}{\hat{\beta}_{k2} - \hat{\beta}_{k1}}$$

then set $T_k=\infty$. Among the n-3 regression tuples and the single-phase model, the one with the $\min(T_k) \; \forall$ k=0,..,n-3 is used to calculate initial and terminal half-life. If a single-phase model is selected ($\min(T_k)=T_0$) the half-life so determined can be utilized as both initial and terminal phase half-life.

Bitman [4] presented a simulation study of the Lee method modified to use the least absolute deviations, Huber-M and non-parametric regression and using the sum of squared residuals to select the best tuple of regression lines. For example,

using the sum of absolute residuals for estimation of regression lines but using the sum of squared residuals for selection of the best tuple of regressions.

Here, the Lee method is modified for ordinary least squares, least absolute deviations, Huber-M and non-parametric regressions using the same criterion for regression line estimation and selection of the regression tuple. A further modification was to require decreasing regression lines in initial and terminal phases to ensure physiologically meaningful results.

Methods Compared

1. Biexponential model as implemented in R function `SSbiexp` (package `stats`) using estimates for β_1 and β_2 obtained by

```
exp(coef(nls(conc~SSbiexp(time, a1, b1, a2, b2)))[c(2,4)]).
```

where the maximum and the minimum of the two estimates obtained by the above function call were used as β_1 and β_2, respectively.

2. Biexponential model (Biexp) fitted by the least squares criteria using the Nelder-Mead algorithm as implemented in R function `optim` with the parameterization

$$E(y_t) = \alpha_1 e^{-\left(\exp\left(\beta_2'\right)+\exp(\delta)\right)t} + \alpha_2 e^{-\left(\exp\left(\beta_2'\right)\right)t}$$

where $\beta_2=\exp(\beta_2')$ and $\beta_1=\exp(\beta_2')+\exp(\delta)$ to ensure $\beta_1>\beta_2>0$. Curve peeling as suggested in Foss [5] is used to get start values for nonlinear model fitting. When no adequate start values (i.e. $\beta_1<\beta_2<0$) are determined by curve peeling, a single exponential model is fitted with start values obtained from an OLS regression on natural log transformed values

$$E(y_t) = \alpha e^{-(\exp(\beta'))t}$$

where $\beta=\exp(\beta')$ to ensure $\beta>0$. The half-life so determined can be utilized as both initial and terminal phase half-life.

3. Lee method using the ordinary least squares (OLS) regression to estimate regression lines with

$$T_k = \sum_{j=1}^{i} e_j^2 + \sum_{j=i+1}^{n} e_j^2 \text{ and } T_0 = \sum_{j=1}^{n} e_j^2.$$

4. Lee method using the least absolute deviations (LAD) regression to estimate regression lines with

$$T_k = \sum_{j=1}^{i} |e_j| + \sum_{j=i+1}^{n} |e_j| \text{ and } T_0 = \sum_{j=1}^{n} |e_j|.$$

5. Lee method using Huber-M regression (Huber-M) to estimate regression lines with

$$T_k = \sum_{j=1}^{i} \rho(e_j) + \sum_{j=i+1}^{n} \rho(e_j) \text{ and } T_0 = \sum_{j=1}^{n} \rho(e_j)$$

where

$$\rho(e) = \begin{cases} e^2 & -k \le e \le k \\ 2k|e| - k^2 & e < -k \text{ or } k < e \end{cases}.$$

Huber M-estimates were calculated by non-linear estimation using the Nelder-Mead algorithm as implemented in the R function `optim`, where OLS regression parameters were used as starting values. The function that was minimized involved k = 1.5*1.483*MAD, where MAD was defined as the median of absolute deviation of residuals obtained by a least absolute deviation (LAD) regression based on the observed data. The initial value of MAD was used and not updated during iterations as recommended by Holland and Welsch [6].

6. Lee method using non-parametric (NPR) regression as suggested in Birkes and Dodge [7] to estimate regression lines with

$$T_k = \sum_{j=1}^{i} rank(e_j) - \frac{i+1}{2} e_j + \sum_{j=i+1}^{n} rank(e_j) - \frac{n-i+2}{2} e_j$$

$$\text{and } T_0 = \sum_{j=1}^{n} rank(e_j) - \frac{i+1}{2} e_j$$

7. Biexponential model fitted on the log-scale (LogBiexp) using R function `nls` (package `stats`) and using the procedure presented in [5] to get start values for non-linear fitting. Estimates for β_1 and β_2 were obtained by

```
coef(nls(log(conc)~log(a1*exp(-b1*time) + a2*exp(-b2*time)),
start=start, control=nls.control(maxiter=1000))).
```

The maximum and the minimum of the two estimates obtained by the above function call were used as β_1 and β_2, respectively.

Simulations

Test data were generated under assumption of log-normal distributed y_t's, which is a plausible statistical distribution for drug levels. Drug levels in blood cannot be negative, while the upper end is open leading to a non-symmetrical distribution. A discussion on the distribution of drug levels including more complex error distributions can be found in [8]. The following two models were used for simulations

1. α_1=50, β_1=1/1, α_2=4 and β_2=1/10
2. α_1=50, β_1=1/4, α_2=4 and β_2=1/8.

Within each simulation run, theoretical half-life of initial and terminal phases were determined after uniformly random variation of parameters (α_1,β_1,α_2,β_2) and time points (0, 1/2, 1, 3, 6, 12, 18, 24, 32, 48) by ±20%. Theoretical concentrations $E(y_t)$ were varied by

$$y_t = E(y_t)e^{\varepsilon_t} = e^{\log(E(y_t))+\varepsilon_t}$$

where ε_t is a normal distributed error with $E(\varepsilon_t)=-V(\varepsilon_t)/2$ and $V(\varepsilon_t)=(0.1^2)\log^2(E(y_t))$ resulting in unbiased log-normal distributed y_t with identical per time point variability of 10% (on the log-scale). With a probability w, theoretical concentrations were contaminated by a cauchy distributed ε_t with 0 as location and $V(\varepsilon_t)$ as scale parameter. The following deviation

$$c = \left(\frac{\log(2)}{\beta_1} - \frac{\log(2)}{\hat{\beta}_1}\right)^2 + \left(\frac{\log(2)}{\beta_2} - \frac{\log(2)}{\hat{\beta}_2}\right)^2$$

between theoretical and estimated parameters was calculated within each simulation run and probability of contamination. The percentage of simulation

runs in which a specific method achieved min(c) was reported. Simulations were performed with 1000 simulation runs per probability of contamination.

Results

Table 1
Proportion of Simulation runs a Specific Method Achieved min(c)

		Method						
Model	**w**	**SSbiexp**	**Biexp**	**OLS**	**LAD**	**Huber-M**	**NPR**	**LogBiexp**
1	0%	0.011	0.015	0.109	0.184	0.070	0.117	0.516
	10%	0.011	0.010	0.102	0.174	0.073	0.124	0.528
	20%	0.019	0.008	0.087	0.214	0.100	0.110	0.488
	30%	0.013	0.014	0.100	0.204	0.090	0.145	0.467
	40%	0.012	0.007	0.081	0.217	0.101	0.141	0.467
	50%	0.015	0.017	0.084	0.219	0.103	0.125	0.459
	75%	0.019	0.008	0.071	0.234	0.101	0.155	0.443
2	0%	0.010	0.022	0.228	0.301	0.132	0.173	0.200
	10%	0.008	0.023	0.217	0.300	0.144	0.170	0.208
	20%	0.012	0.027	0.227	0.281	0.132	0.169	0.206
	30%	0.007	0.030	0.201	0.287	0.171	0.168	0.195
	40%	0.010	0.025	0.168	0.307	0.155	0.189	0.214
	50%	0.010	0.021	0.193	0.307	0.136	0.175	0.213
	75%	0.010	0.018	0.162	0.314	0.176	0.201	0.195

The proportion of simulation runs in which a specific method achieved min(c) is summarized in table 1. Method 1 and 2 were inferior in terms of min(c). Both methods work well under assumption of normal distributed drug levels with identical variances for each time point.

The LogBiexp method showed best results for model 1 followed by the Lee method using LAD regression for all rates of contamination.

Using the specifications of model 2, the Lee method using LAD regression was superior in terms of min(c). The second best method was the Lee method using OLS regression followed by the LogBiexp method for low contaminated data. For higher contaminated data, the LogBiexp method was better than the Lee method using OLS regression. For w=75% the Lee method using NPR regression was superior to the LogBiexp method.

The sub optimal performance of the LogBiexp method under specifications of model 2 (very small difference between initial and terminal half-life) may be due to convergence issues of the non-linear fitting approach. Table 2 presents the proportions of simulation runs where the LogBiexp method did not converge. Under non-contaminated (w=0) conditions the LogBiexp method did not converge in 0.3% and in 43% for model 1 and 2, respectively. Under high-contaminated (w=75%) conditions the LogBiexp method did not converge in 5% and in 46% for model 1 and 2, respectively.

Table 2
Proportion of Simulation runs Without Convergence of Method LogBiexp

	Percentage of contamination						
Model	**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**75%**
1	0.003	0.009	0.014	0.017	0.030	0.030	0.048
2	0.429	0.442	0.440	0.425	0.448	0.428	0.455

Discussion and Conclusion

In summary, the non-convergence of the LogBiexp method in a high percentage of simulation runs under specification of model 2 is a disadvantage in practical medical research as the complete method has to be specified in advance. In case of non-convergence in face of the actual data analysis, a posterior change of the statistical methodology is necessary (e.g. using a different approach to get start values like grid search algorithms).

To avoid a possible posterior change of the statistical analysis which may make the results less convincing to regulatory authorities, the Lee method using LAD regression can be considered as a good alternative for two-phase half-life estimation under log-normal distributed concentrations.

References

[1] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2005. ISBN 3-900051-07-0; URL: http://www.R-project.org.

[2] Wolfsegger MJ, Jaki T. PK: Basic pharmacokinetics. R package version 0.03, 2005.

[3] Lee ML, Poon W-Y, Kingdon HS. A two-phase linear regression model for biologic half-life data. J. Lab. Clin. Med 1990, 115:745–748.

[4] Bitman B, Lee ML, Schroth P. Robust regression methods for biologic half-life data. Royal Statistical Society International Conference, Reading, United Kingdom, 2000.

[5] Foss SD. A method for obtaining initial estimates of the parameters in exponential curve fitting. Biometrics 1969, 25:580–584.

[6] Holland PW, Welsch RE. Robust regression using iteratively reweighed least-squares. Commun. Stat.-Theor. M. 1977, 6:813–827.

[7] Birkes D, Dodge Y. Alternative methods of regression. John Wiley and Sons, New York, 1993.

[8] PharmPK. PharmPK Discussion List Archive last visited at 2005-10-31; URL: http://www.boomer.org/pkin/PK03/PK2003160.html.