Correction of the likelihood function as an alternative for imputing missing covariates

Wojciech Krzyzanski and An Vermeulen

PAGE 2017 Budapest

Covariates in Population PKPD Analysis

"Covariates are defined as patient variables, either intrinsic or extrinsic, that attempt to explain between-subject variability in the model parameters."

Attributes:

- Covariates are independent variables (known without a measurement error)
- Covariates differ between subjects (covariate distribution)
- Covariates are not model parameters (not estimated)
- Covariates are used as explanatory variables (predictors) in parameter-covariate relationships

Examples:

- Continuous covariates: body weight, age, creatinine clearance, albumin level
- Categorical covariates: gender, race, diseases stage, CYP450 PM vs EM

Hutmacher and Kowalski, Br J Clin Pharmacol 79:132-147 (2014)

Missing Covariates

Covariate	Median	Minimum	Maximum	Subjects with missing covariate	
Age (years)	65	16	88	0	
Weight (kg)	80	40	147	0	
Creatinine clearance (ml/min)	75	22.4	508[a]	0	
Albumin (g/l)	42	22	58	122	**8.4%**
Bilirubin (umol/l)	0.70	0.07	4.80	22	
AST (IU/l)	21.0	5.0	360	25	
ALT (IU/l)	22.0	3.0	370	23	
ALKP (IU/l)	81.0	21.0	3,278	17	
Potassium (mmol/l)	4.3	3.2	7.8	33	
Magnesium (mmol/l)	1.9	0.96	24.0	27	
Calcium (mmol/l)	9.3	7.4	12.7	23	
Sodium (mmol/l)	140	103	157	23	
Total protein (g/l)	72	32	89	185	**12.8%**

[a] Estimates greater than 150 ml/min were fixed to a value of 150 ml/min
AST Aspartate transaminase, *ALT* alanine transaminase, *ALKP* alkaline phosphatase

Continuous demographics and biochemical markers recorded for the 1,445 patients enrolled in one of 14 phase III efficacy and safety clinical trials of class III anti-arrhythmic drug dofetilide

Tunblad et al., J Pharmacokinet Pharmacodyn 35:503–526 (2008)

Covariate Imputation

Exclusion of subjects with missing covariates from analysis is a waste of data and might lead to a reduction of statistical power. Data imputation assigns a missing data point a specific value

- Single imputation methods: mean imputation, last observation carried forward, random hot deck imputation
- Multiple imputation methods:
  - generating multiple sets of covariates with imputed values
  - analyzing data with one set of imputed covariates at a time
  - combining parameter estimates from each analysis
- Multiple imputation methods give unbiased and precise estimates of population parameters

Johansson and Karlsson, AAPS J 15:1035–1041 (2013)

Maximum Likelihood Method for Missing Covariates

- A regression model with observed covariates is created for the missing covariate
- A random effect is added to the regression model and estimated to take the uncertainty in the model into account
- The variance of the uncertainty distribution is fixed to the estimated value
- The values of the missing covariates are estimated from the fixed distribution and the individuals' observed data
- The ML method has been shown to yield precise and unbiased parameter estimates when the covariates are missing at random

Johansson and Karlsson, AAPS J 15:1232–1241 (2013);
Karlsson et al., J Pharmacokinet Biopharm 26: 207-246 (1998)

Population Model

Consider N subjects, $1 \leq i \leq N$, each with a series of vector observations $\{\boldsymbol{y}_{ij}\}_{1 \leq j \leq n_i}$ at times $\{t_{ij}\}_{1 \leq j \leq n_i}$, and a vector of covariates $\boldsymbol{c_i}$. Let each observation vector be described by the following model

$$\boldsymbol{y}_{ij} = \boldsymbol{f}\left(t_{ij}, \boldsymbol{\psi_i}, \boldsymbol{\theta}, \boldsymbol{c_i}\right) + \boldsymbol{\varepsilon_{ij}}$$

$\boldsymbol{\psi}_i$ is a vector of model parameters specific to individual i
$\boldsymbol{\theta}$ is a vector of parameters common to all subjects

Probability density function for the joint distribution of $\boldsymbol{y}_{ij}$, $\boldsymbol{\psi}_i$ and $\boldsymbol{\varepsilon_{ij}}$

$$p(\underbrace{\boldsymbol{y}_{ij}, \boldsymbol{\psi}_i, \boldsymbol{\varepsilon_{ij}}}_{\text{random variables}}; \underbrace{\boldsymbol{c_i}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{t}}_{\text{deterministic variables}})$$

$\boldsymbol{\gamma}$ is a vector of parameters describing the joint distribution

Maximum Likelihood Estimation of Model Parameters

The joint distribution of all observations $\boldsymbol{y} = \{y_{ij}\}_{1 \le i \le N, 1 \le j \le n_i}$

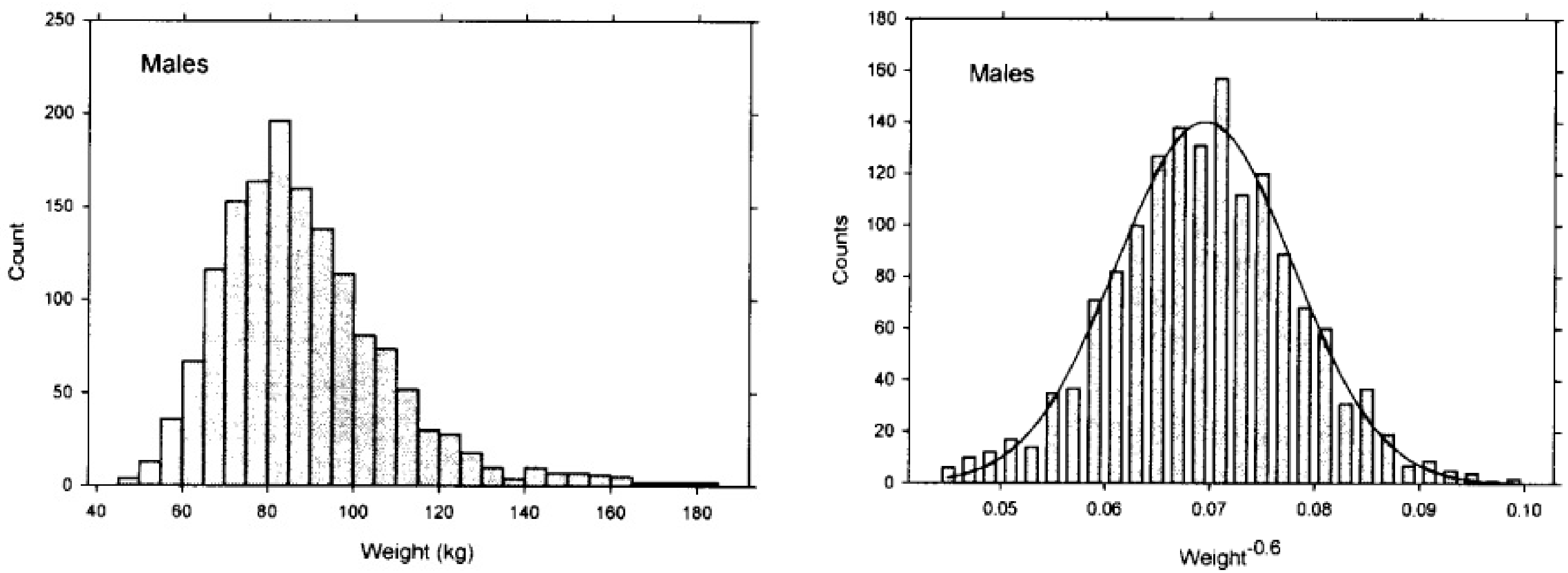
$$p(\boldsymbol{y}; \boldsymbol{c_a}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{t}) = \prod_{i=1}^{N} \prod_{j=1}^{n_i} \int_{\{(\boldsymbol{\psi}, \boldsymbol{\varepsilon}) \in D_i\}} p\left(\boldsymbol{y}_{ij}, \boldsymbol{\psi}, \boldsymbol{\varepsilon}; \boldsymbol{c_i}, \boldsymbol{\theta}, \boldsymbol{\gamma}, t_{ij}\right) d\boldsymbol{\psi} d\boldsymbol{\varepsilon}$$

$\boldsymbol{c_a} = \{c_{ij}\}_{1 \le i \le N, 1 \le j \le q}$ is a vector of all known individual covariates
$\boldsymbol{t} = \{t_{ij}\}_{1 \le i \le N, 1 \le j \le n_i}$
D_i is the domain of individual model and residual parameters for subject i

$$\underbrace{\left(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\gamma}}\right)}_{\text{estimates of population parameters}} = \underset{\boldsymbol{\theta}, \boldsymbol{\gamma}}{\operatorname{argmin}} \left\{-2 log\left(p(\boldsymbol{y}; \boldsymbol{c_a}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{t})\right)\right\}$$

Covariate Distribution Model

- Covariate distribution models are used in clinical trial simulations to generate realistic patient populations
- Covariate distribution model allows for explanation of covariate correlations



Distribution of body weights (left) and its Box-Cox transformation (right) from the National Health and Nutrition Survey data set for males 20-60 years old. The solid line represents the normal p.d.f.

P. Bonate, Pharmacokinetic-Pharmacodynamic Modeling and Simulation. Springer (2011)

Correction of Likelihood Function for Missing Covariates

For subject i with missing covariates $\boldsymbol{c_i}$ we propose that $\boldsymbol{c_i}$ is a random vector of distribution common to all subjects $p(\boldsymbol{c_i};\boldsymbol{\beta})$

Then the joint distribution has the following form:

$$p\left(\boldsymbol{y}_{ij},\boldsymbol{\psi}_i,\boldsymbol{\varepsilon_{ij}}\boldsymbol{c_i};\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\beta},t_{ij}\right) = p\underbrace{(\boldsymbol{y}_{ij},\boldsymbol{\psi}_i,\boldsymbol{\varepsilon_{ij}}|\boldsymbol{c_i};\boldsymbol{\theta},\boldsymbol{\gamma},t_{ij})}_{\text{probability density function conditional on } \boldsymbol{c_i}}p(\boldsymbol{c_i};\boldsymbol{\beta})$$

The joint distribution of all observations $\boldsymbol{y}$:

$$p(\boldsymbol{y};\boldsymbol{c_a},\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{t}) = \prod_{i=1}^{M}\prod_{j=1}^{n_i}\int_{\{(\boldsymbol{\psi},\boldsymbol{\varepsilon})\in D_i\}} p\left(\boldsymbol{y}_{ij},\boldsymbol{\psi},\boldsymbol{\varepsilon};\boldsymbol{c_i},\boldsymbol{\theta},\boldsymbol{\gamma},t_{ij}\right)d\boldsymbol{\psi}d\boldsymbol{\varepsilon}$$

$$\prod_{i=M+1}^{N}\prod_{j=1}^{n_i}\int_{\{(\boldsymbol{\psi},\boldsymbol{\varepsilon},\boldsymbol{c})\in D_i\times C\}} p(\boldsymbol{y}_{ij},\boldsymbol{\psi},\boldsymbol{\varepsilon}|\mathbf{c};\boldsymbol{\theta},\boldsymbol{\gamma},t_{ij})p(\boldsymbol{c};\boldsymbol{\beta})d\boldsymbol{\psi}d\boldsymbol{\varepsilon}d\boldsymbol{c}$$

C is the domain of all possible individual covariates

Case Study

- Simulated data of plasma concentrations for N=80 subjects based on the one compartment model with an IV bolus $Dose$ = 10,000 ng and sampling times $t = 5, 30$ min, 1, 2, 3, 4, 5, 6, 8, 10, 12 h

$$C_p = \frac{Dose}{V_p} exp(-CL/V_p \cdot t)$$

- The body weight (BW) was the only covariate related to CL and V according to the power functions with exponents SCL=0.75 and SV=1

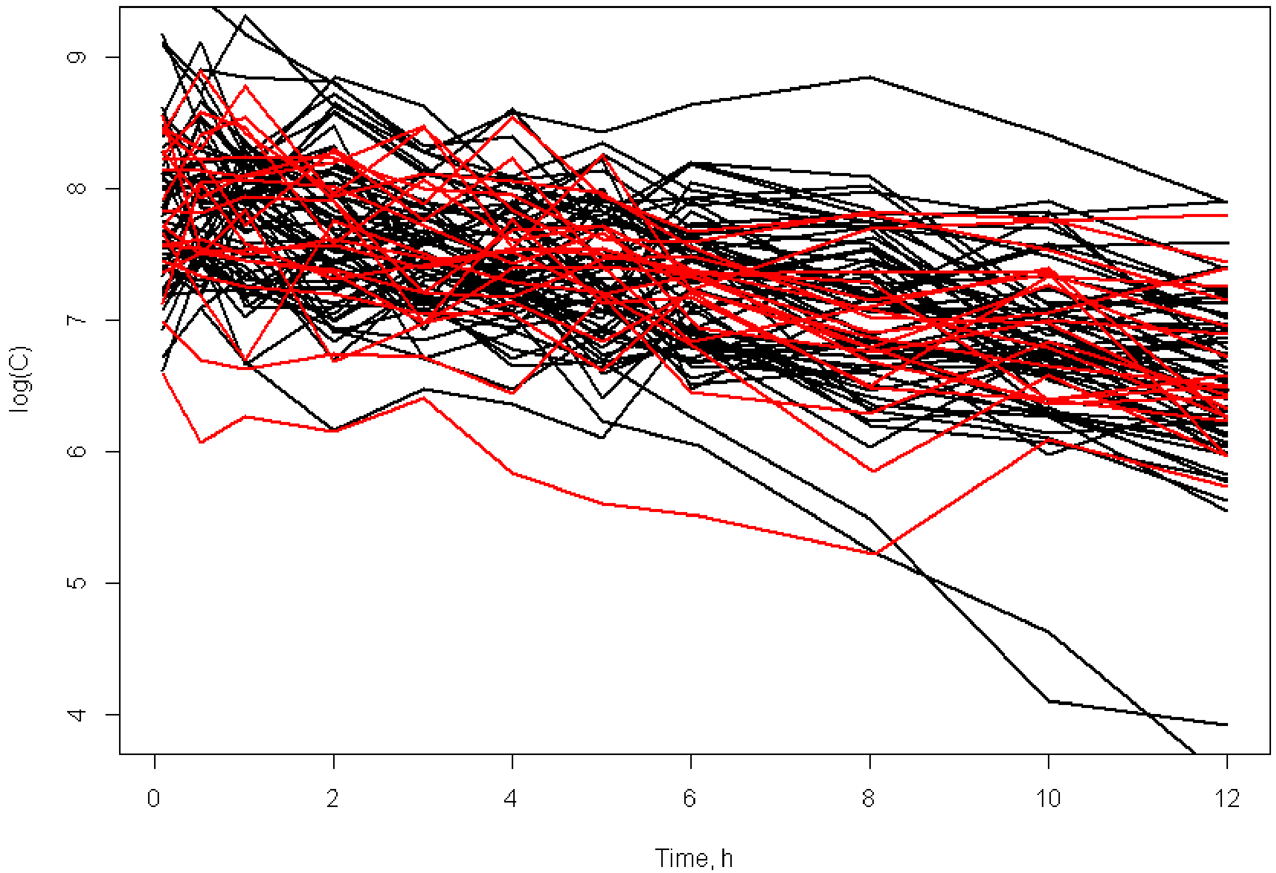
$$CL = CL_0 \left(\frac{BW}{BW_{mean}}\right)^{SCL} exp(\eta_{CL0}) \qquad V = V_0 \left(\frac{BW}{BW_{mean}}\right)^{SV} exp(\eta_{V0})$$

- The plasma concentrations were log-transformed and residual error was added

$$y_{ij} = logC_{pij} + \varepsilon_{ij}$$

- For missing data BWs of 20 subjects were removed from the original dataset
- Parameters were estimated by the FOCE method using NONMEM 7.3.

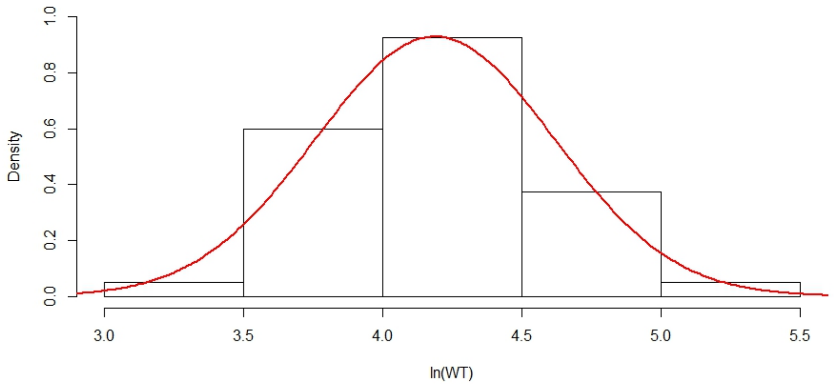
Individual Concentration Time Courses



Black lines represent subjects with BWs and red lines represent subjects without BWs

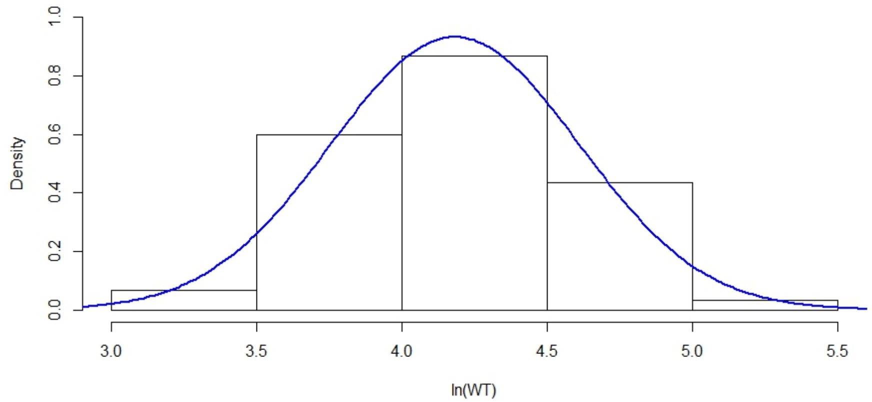
Body Weight Distribution Model

N = 80



$$logBW \sim N(\mathbf{4.19, 0.185})$$

N = 60



$$logBW \sim N(\mathbf{4.18, 0.183})$$

Densities of BW distributions of all subjects (left) and subjects selected to have BW (right)

Data Item MCV

C.ID	TIME	AMT	CMT	DV	MDV	WT	MCV
1	0.000	10000	1	0.00	1	40.99128	0
1	0.083	0	1	9.17	0	40.99128	0
1	0.500	0	1	8.43	0	40.99128	0
1	0.983	0	1	7.94	0	40.99128	0
1	2.000	0	1	8.48	0	40.99128	0
1	3.000	0	1	7.19	0	40.99128	0

C.ID	TIME	AMT	CMT	DV	MDV	WT	MCV
66	0.000	10000	1	0.00	1	0	1
66	0.083	0	1	8.56	0	0	1
66	0.500	0	1	7.52	0	0	1
66	1.000	0	1	7.53	0	0	1
66	2.017	0	1	7.57	0	0	1
66	3.000	0	1	7.39	0	0	1

MCV = 0 Covariate present
MCV = 1 Covariate absent

Coding Missing Covariates in NONMEM

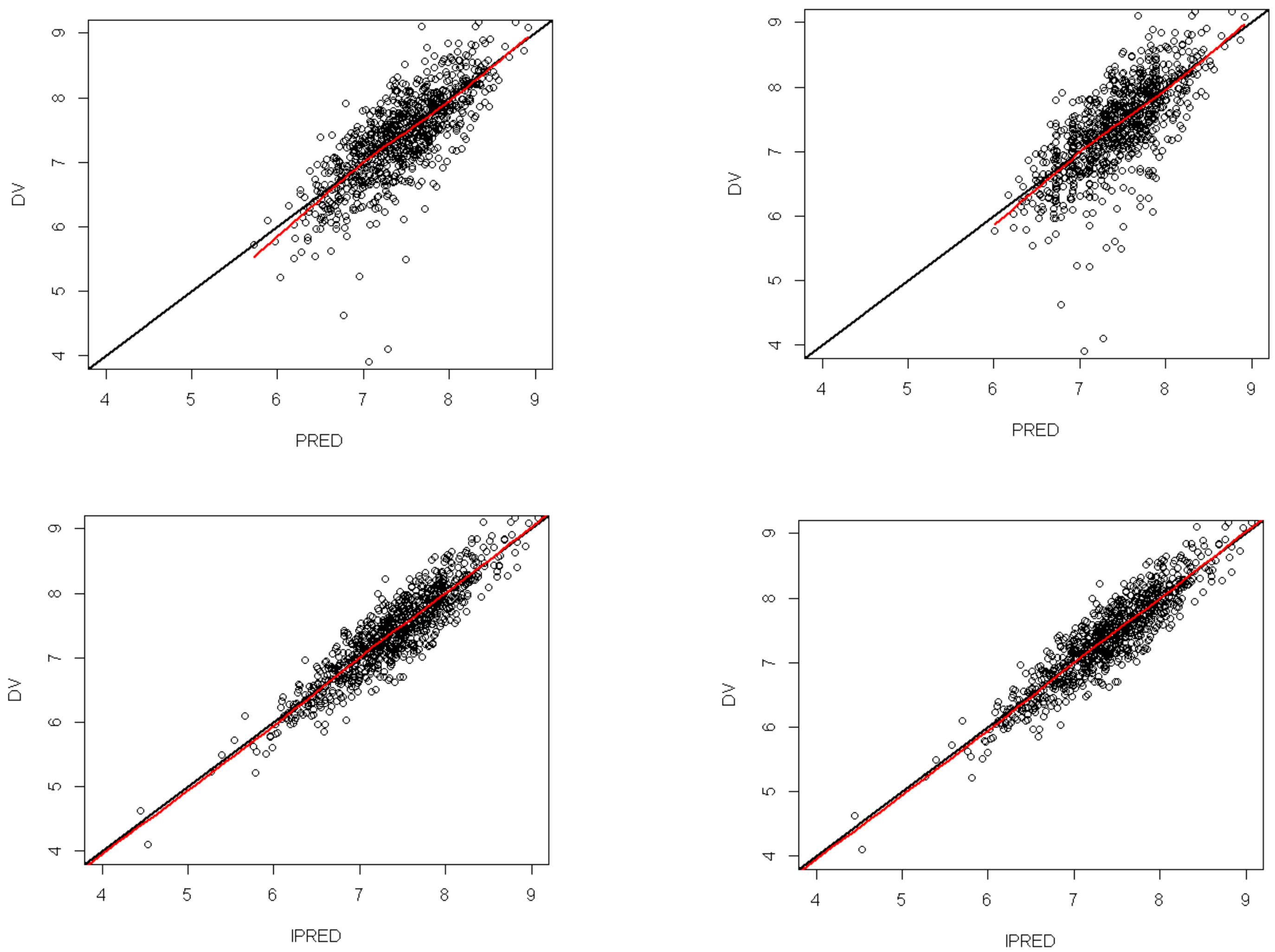
```
V0=THETA(1)
CL0=THETA(2)
SCL=THETA(3)
SV=THETA(4)
mLN= 4.183649
W=exp(mLN)*EXP(ETA(3))

IF (MCV.EQ.0) THEN
 LNCL=log(CL0) +SCL*(log(WT) -mLN)+ ETA(1)
 LNV=log(V0) +SV*(log(WT) -mLN)+ ETA(2)
ELSE
 LNCL=log(CL0) +SCL*(log(W) -mLN)+ ETA(1)
 LNV=log(V0) +SV*(log(W) -mLN)+ ETA(2)
ENDIF

CL=EXP(LNCL)
V=EXP(LNV)
```

```
$OMEGA
0.1
0.1
0.1794029 FIX;Var of LN(WT)
```

Parameter Estimates

Model	All	Covariates	Missing	Covariates
Parameter	**Estimate**	**SE**	**Estimate**	**SE**
V_0, L	4.09	0.160	3.72	0.153
CL_0, L/h	0.394	0.0147	0.363	0.0148
SCL	0.782	0.0782	0.749	0.131
SV	0.961	0.0843	0.956	0.120
ω^2_{CL0}	0.0887	0.0212	0.0968	0.0243
ω^2_{V0}	0.0963	0.0206	0.0960	0.0231
σ^2	0.089	0.00457	0.089	0.0046
$shrink\ \eta_{V0}$, %	14.1		16.2	
$shrink\ \eta_{CL0}$, %	8.4		14.4	
OFV	-960.9		-929.8	

Goodness of Fit Plots



Observed vs. predicted plots for the model with all covariates (left) and with missing covariates (right)

Conclusions

- Presented method is an application of the maximum likelihood method of handling missing covariates
- Correction of the likelihood function for missing covariates is an alternative to imputation
- Covariate distribution model is necessary to account for missing covariates
- Transformation of covariates to ensure normal distribution is required for proper correction of the likelihood function

Acknowledgments

This project was supported by the Graduate Student Fellowship from Janssen Research and Development, a Division of Janssen Pharmaceutica NV, Beerse, Belgium