

## 4.4 Covariate Model

Specific Learning Objectives:

1. Identify the common types of functional relationships between covariates and PK parameters.
2. Perform graphical assessments to determine the functional relationship between covariates and PK parameters.
3. Interpret the meaning of the estimated fixed parameters in the covariate model depending on their functional relationship with the PK parameters.
4. Interpret the estimated BSV in PK parameters when covariates are used.
5. Write a full NLME model specification and identify its components.
6. Implementation of covariate selection techniques and model building.

## The Covariate Model (Population Level)

- Covariate: any variable that is specific to an individual and may influence the process under study (i.e., the pharmacokinetics or pharmacodynamics of a drug).
- Subject-specific characteristics included in the model will reduce the BSV and residual variability.
  - E.g. A model where Volume of distribution ( $V_i$ ) depends on weight, has both reduced BSV in  $V_i$  and unexplained residual variability.
- Individualized dosing regimen is possible based on significant covariate relationships.

## The Covariate Model (Population Level)

- In general, the covariate's type will determine its functional relationship with the PK parameters:
  - Continuous covariates: linear, exponential and power.
  - Categorical covariates: linear, fractional change, exponential.

## Continuous Sub-model Covariates

- E.g., Weight, height, age.
- Sub-model functional forms:
  - Consider the Phenobarbital Data example with Weight as covariate.

(Suppressing  $\eta_i$ 's for now)

### Linear

$$CL_i = \theta_1 + \theta_2 Wt_i$$

$\theta_1$ : mean CL when Wt=0  
 $\theta_2$ : mean change in CL for unit change in Wt

### Exponential

$$CL_i = \theta_1 \exp(\theta_2 Wt_i)$$

$\theta_1$ : mean CL when Wt=0  
 $\theta_2$ : mean change in ln(CL) for unit change in Wt

### Power

$$CL_i = \theta_1 (Wt_i)^{\theta_2}$$

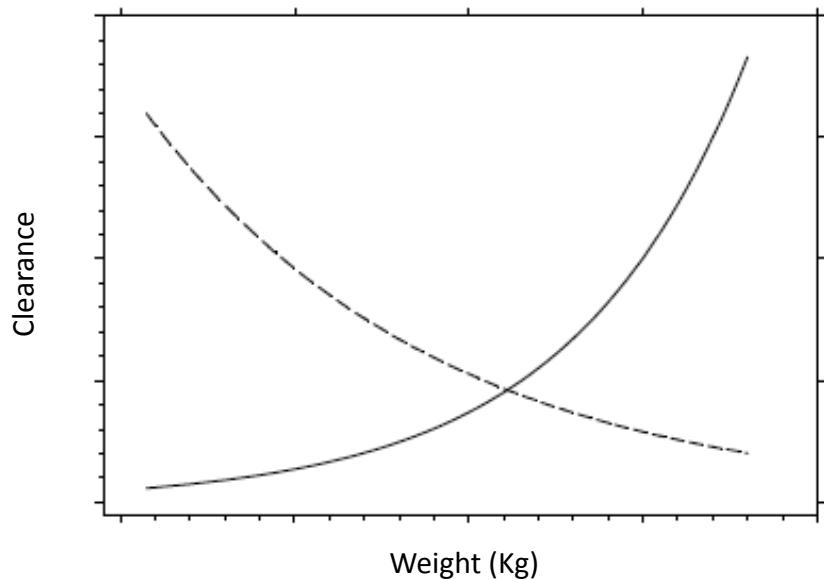
$\theta_1$ : mean CL when Wt=1  
 $\theta_2$ : mean change in ln(CL) for unit change in ln(Wt)

Curvilinear relationship, linear in ln scale.

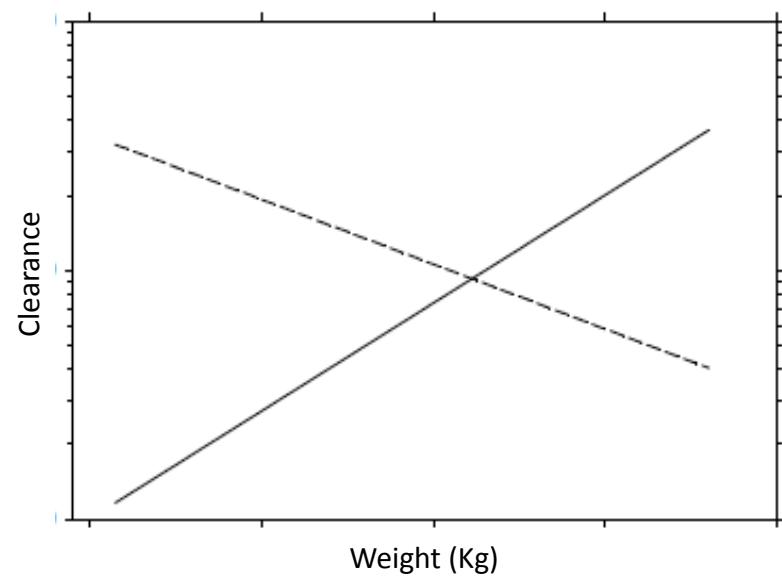
## Continuous Sub-model Covariates

- **Exponential covariate model:** solid line has  $\theta_2 > 0$ , dashed line has  $\theta_2 < 0$ .
- Drawback of this model: if  $\theta_2$  is positive and large it may lead to overflow errors.

$$CL_i = \theta_1 \exp(\theta_2 Wt_i)$$



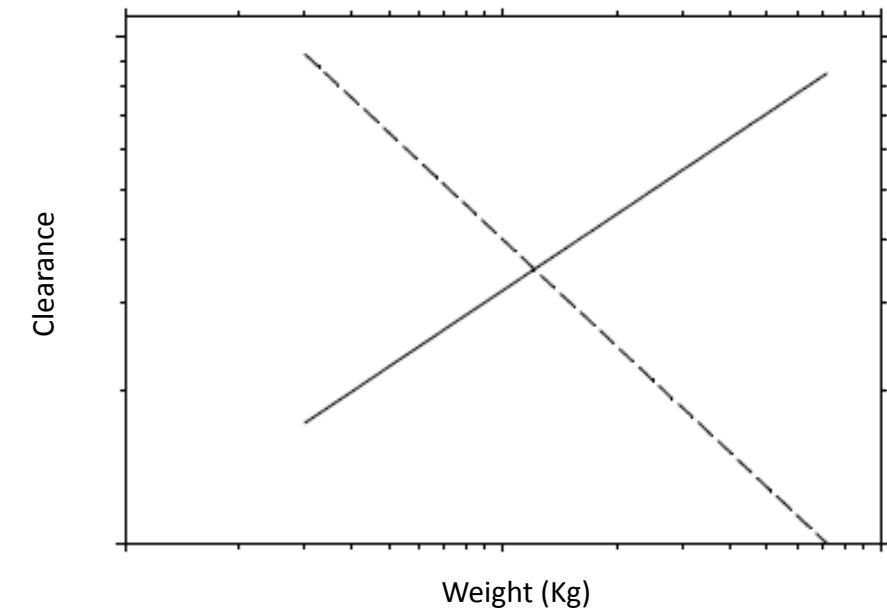
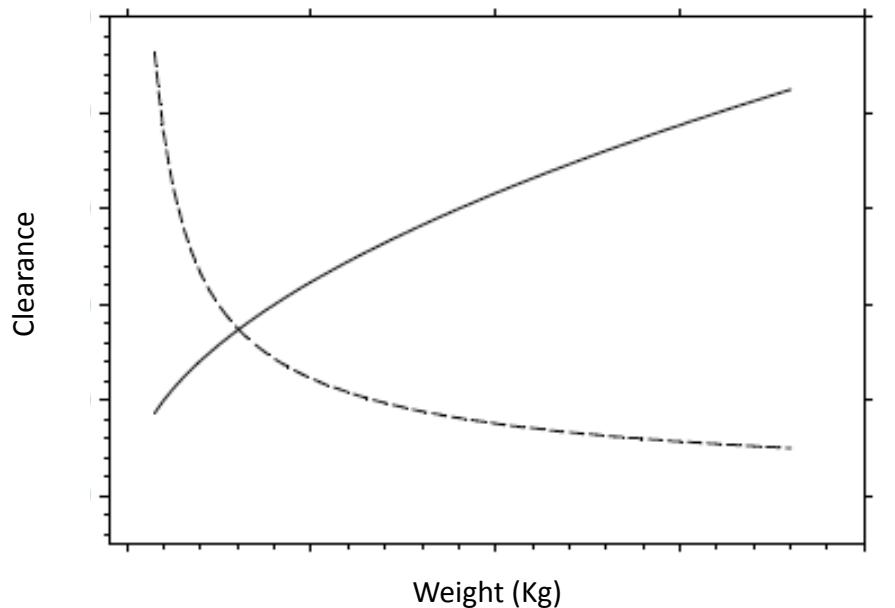
$$\ln CL_i = \ln \theta_1 + \theta_2 Wt_i$$



Plots adapted from Bonate, 2011.

## Continuous Sub-model Covariates

- **Power covariate model:** solid line has  $\theta_2 > 0$ , dashed line has  $\theta_2 < 0$ .
- Most commonly seen in literature.
- Often used because of the allometric relationship between many physiological parameters and body weight.
- When weight or some other measure of size is a covariate then there is a physiological reason for selecting a power model.



Plots adapted from Bonate, 2011.

## Choosing Sub-model functional forms

- Covariate screening:
  - Scatter plots of EBE's of PK parameters (i.e., predicted PK parameters under a model with no covariates) vs. covariates may indicate trend.
- Formal testing:
  - Smallest AIC's among competing models.

## Centering/Scaling Covariates

- Performed based on covariate functional form.
- To improve interpretation of estimates
- To avoid instability during estimation procedures

If the median weight is 1.3 Kg,  $\theta_1$  is the population mean CL for the typical weight.

**Scaled linear**       $CL_i = \theta_1 + \theta_2(Wt_i - 1.3)$

**Scaled exponential**       $CL_i = \theta_1 \exp\left(\theta_2 \frac{Wt_i}{1.3}\right)$

**Scaled power**       $CL_i = \theta_1 \left(\frac{Wt_i}{1.3}\right)^{\theta_2}$

The 0.75 constant can be estimated with the data at hand.

E.g. Scaled Power Allometric function of CL vs. Wt with  $\theta_2=0.75$ :

$$CL_i = \theta_1 \left(\frac{Wt_i}{70}\right)^{0.75}$$

## Continuous Sub-model Covariates

### More on Centering/Scaling Covariates...

- Centering/scaling is done depending on the scale in which estimation of the parameters is preformed.
  - E.g., estimation is done in the log-scale, that is, in the linear form in both the R function `phenoModel()` and the “`exp(Sum+eta)`” option in Phoenix.
  - In this case, `Wt` is to be centered, not scaled.
- After centering (linear model): estimated fixed effects remain the same except the intercept. Units don't change.
- After scaling (exponential, power models): estimated fixed effects change but intercept remains the same. Units do change.
- These transformations are a reparameterization of the model that does not improve the fit of the model nor does it change the data.

## Collinearity

- Bonate (1999) showed that
  - If correlation between covariates is greater than 0.5 then the resulting population parameter estimates begin to show increasing bias and standard error.
  - If one parameter was correlated with another, the bias and increase in standard errors is propagated to the other parameters as well.
- How to deal with collinearity between two variables:
  - Use surrogate variables (e.g. BSA in terms of weight and height.)
  - Use the covariate with the greatest predictive value in the model.

## Categorical Sub-model Covariates

- E.g. Gender, race, disease status.
- Sub-model functional forms:
  - Consider the Phenobarbital data example with the grouping variable:  
 $ApgarInd = \{1 \text{ if Apgar Score} \geq 5, 0 \text{ otherwise}\}$  as covariate.

(Suppressing  $\eta_i$ 's for now)

**Linear**     $CL_i = \theta_1 + \theta_2 Wt_i + \theta_3 ApgarInd_i$

$$CL_i = \begin{cases} \theta_1 + \theta_2 Wt_i, & \text{if } ApgarInd = 0 \\ \theta_1 + \theta_2 Wt_i + \theta_3, & \text{if } ApgarInd = 1. \end{cases}$$

$\theta_3$  is the mean CL difference between Apgar Score groups

## Categorical Sub-model Covariates

(Suppressing  $\eta_i$ 's for now)

### Fractional Change

$$CL_i = (\theta_1 + \theta_2 Wt_i)(1 + \theta_3 ApgarInd_i)$$

$$\begin{aligned}\Delta &= (\theta_1 + \theta_2 Wt_i)\theta_3 = \theta_1\theta_3 + \theta_2\theta_3 Wt_i \\ &= \theta^*_1 + \theta^*_2 Wt_i.\end{aligned}$$

$\Delta = (CL \text{ for } ApgarInd=0) - (CL \text{ for } ApgarInd=1)$   
= is a linear function of  $Wt$ .

### Exponential

$$CL_i = (\theta_1 + \theta_2 Wt_i) \exp(\theta_3 ApgarInd_i)$$

$$CL_i = \begin{cases} (\theta_1 + \theta_2 Wt_i)e^{\theta_3}, & \text{if } ApgarInd = 1, \\ \theta_1 + \theta_2 Wt_i, & \text{if } ApgarInd = 0. \end{cases}$$

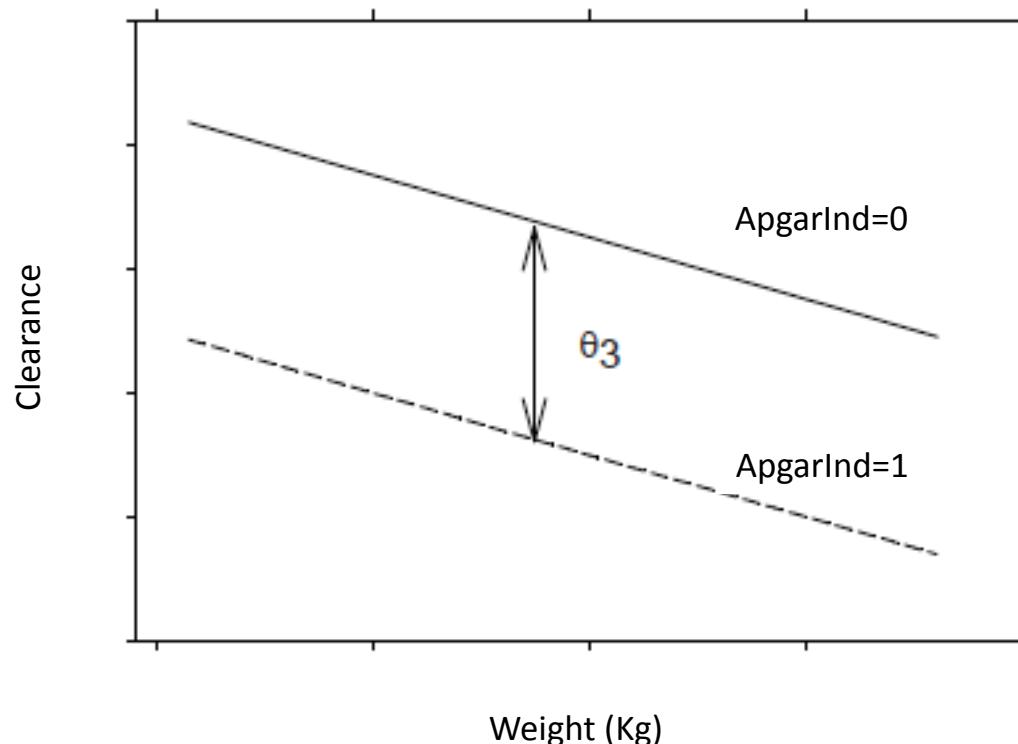
## Categorical Sub-model Covariates

**Linear**

$$CL_i = \theta_1 + \theta_2 Wt_i + \theta_3 ApgarInd_i$$

$$CL_i = \begin{cases} \theta_1 + \theta_2 Wt_i, & \text{if } ApgarInd = 0, \\ \theta_1 + \theta_2 Wt_i + \theta_3, & \text{if } ApgarInd = 1. \end{cases}$$

$\theta_3$  is the mean CL difference between groups.



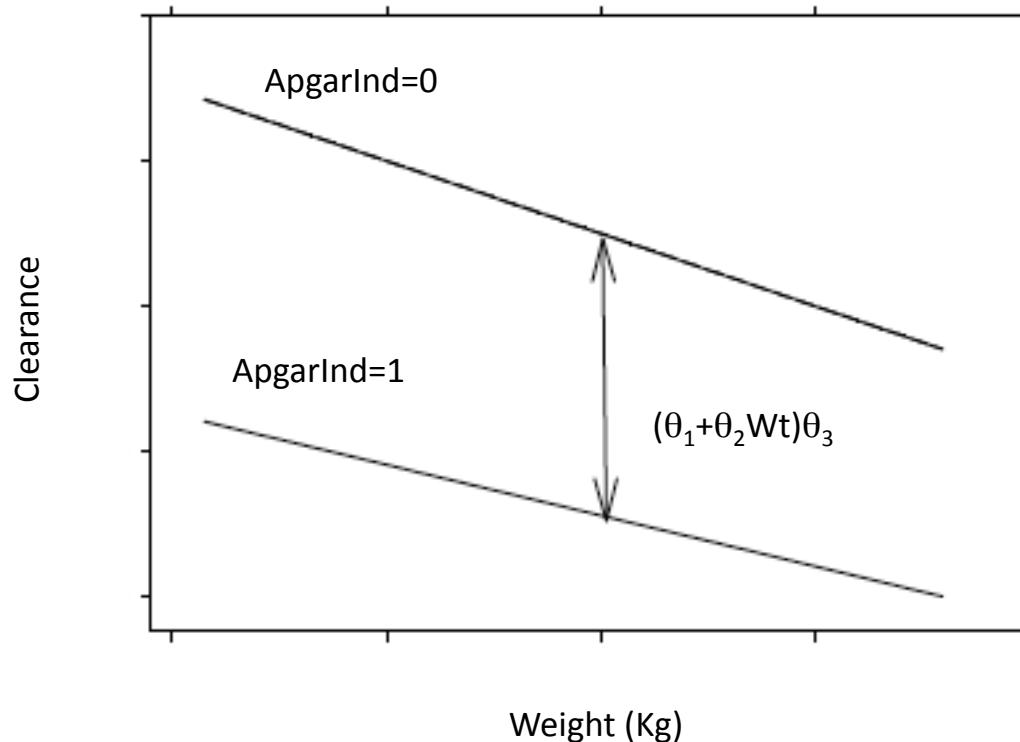
Plot adapted from Bonate, 2011.

## Categorical Sub-model Covariates

**Fractional  
Change**

$$CL_i = (\theta_1 + \theta_2 Wt_i)(1 + \theta_3 ApgarInd_i) \quad \Delta = (\theta_1 + \theta_2 Wt_i)\theta_3.$$

$$\Delta = (CL_{ApgarInd=1}) - (CL_{ApgarInd=0})$$



One problem with this model is that it is possible that  $CL < 0$ ,  
physiologically impossible.

Plot adapted from Bonate, 2011.

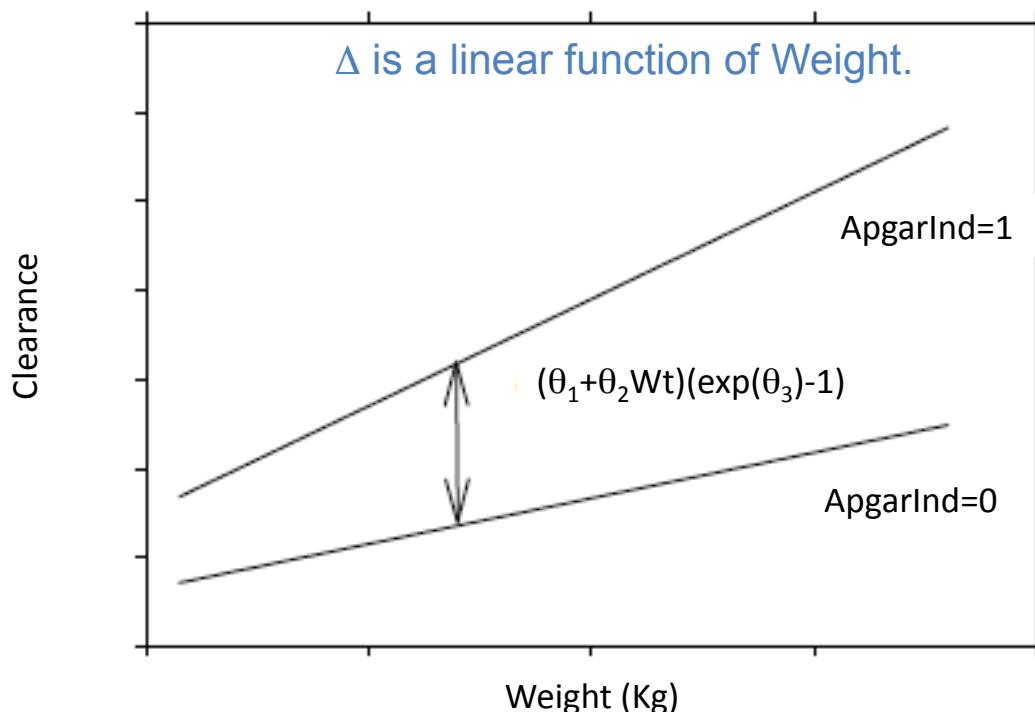
## Categorical Sub-model Covariates

**Exponential**  $CL_i = (\theta_1 + \theta_2 Wt_i) \exp(\theta_3 ApgarInd_i)$   $CL_i = \begin{cases} (\theta_1 + \theta_2 Wt_i) e^{\theta_3}, & \text{if } ApgarInd = 1, \\ \theta_1 + \theta_2 Wt_i, & \text{if } ApgarInd = 0. \end{cases}$

Difference  
between  
Apgar Score  
groups:

$$\begin{aligned}\Delta &= (\theta_1 + \theta_2 Wt_i)(\exp(\theta_3) - 1) \\ &= \theta_1 \exp(\theta_3) - \theta_1 + \theta_2 \exp(\theta_3) Wt_i - \theta_2 Wt_i \\ &= \theta_1 \exp(\theta_3) - \theta_1 + (\theta_2 \exp(\theta_3) - \theta_2) Wt_i \\ &= \theta_1^* + \theta_2^* Wt_i.\end{aligned}$$

Note is constrained  
 $CL > 0$ , and  
 $\theta_3$  is  
unconstrained in  
 $(-\infty, \infty)$ .



Plot adapted from Bonate, 2011.

## Categorical Sub-model Covariates

The models can be generalized for covariates with two or more categories. E.g.,  $Race_i$ : Caucasian, Black, Asian:

$$CL_i = (\theta_1 + \theta_2 Age_i)(1 + \theta_3 Black_i + \theta_4 Asian_i)$$

$$CL_i = \begin{cases} \theta_1 + \theta_2 Age_i & \text{if } race_i = \text{Caucasian} \\ (\theta_1 + \theta_2 Age_i)(1 + \theta_3 Black_i) & \text{if } race_i = \text{Black} \\ (\theta_1 + \theta_2 Age_i)(1 + \theta_4 Asian_i) & \text{if } race_i = \text{Asian} \end{cases}$$

### Categorization of a continuous variable

- In general, this results in loss of information.
- There is slightly greater power when subjects are categorized into equal numbers per group as opposed to equal bin width.
- In some cases this may be useful when assessing a particular subset of values in the variable.

E.g. if there was a question about CL in the smaller preterm babies in the Phenobarbital Data, then

$$WtGroup_i = \begin{cases} 1 & \text{if } Wt_i \geq 1 \\ 0 & \text{if } Wt_i < 1 \end{cases}$$

$$CL_i = \theta_1 [1 + \theta_2 WtGroup_i]$$

$\theta_2 > 0$  with statistical significance indicates a substantial difference in CL between smaller and bigger preterm babies.

### Interpretation of BSV with covariates

- For scaled covariates, the variability associated with the PK parameter is to be interpreted with respect to the typical subject.
- The variability does not refer to the BSV across all individuals. That value would be one where no covariates were used. That is, e.g. if  $\psi_{CL} = 0.50$  for the model:

$$\ln CL_i = \ln \theta_1 + \eta_{CL_i}$$

Then the BSV *across all individuals* is 50%.

- Suppose that  $\psi_{CL} = 0.30$

$$\ln CL_i = \ln \theta_1 + \theta_2(Wt_i - 1.3) + \eta_{CL_i}$$

Then it is said that variability associated with CL is 30% *for the typical subject*. This value refers to subjects with a weight of 1.3 Kg.

## Summary Modeling of NLME elements

Structural model	E.g. I.V. Plasma concentration with first order elimination
Functional form $\eta$ 's vs. population parameters (CL,V)	Additive, exponential.
Functional form $\varepsilon$ 's vs. main response (Y=plasma concentration)	Additive, proportional, exponential, combined. (Residual variance model).
Covariates and functional form of covariate models	Continuous (e.g. Weight): Linear, exponential, power. Categorical (e.g. ApgarInd): Linear, fractional change, exponential.

## Example

### Neonatal Pharmacokinetics of Phenobarbital

(Grasela & Don, 1985)

- Continued -

- Concentration-time-dose history data.
- During first 16 days after birth on 59 preterm infants.
- Initial doses followed by one or more sustaining doses.
- 1-6 concentration measurements at times other than dose times (total of 155 measurements.)
- Additional measurements: birth weight and 5-minute Apgar score.

#### Final BSV Model Formulation, parameterization in R

$$C_{ij} = \sum_{d:t_{id} < t_{ij}} \frac{D_{id}}{V_i} \exp\left\{-\frac{CL_i}{V_i}(t_{ij} - t_{id})\right\} + \varepsilon_{ij}; \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

$$\ln CL_i = \ln \theta_1 + \eta_{CL_i}$$

$$\ln V_i = \ln \theta_2 + \eta_{V_i}$$

$$\Psi = \begin{bmatrix} \psi^2_{CL} & \psi_{CL,V} \\ & \psi^2_V \end{bmatrix}$$

## Phenobarbital Example

```
# model with no covariates for lCl and lV, correlated random effects
# can specify random = pdSymm(lCl + lV ~ 1) or simply:
model2 <- nlme( conc ~ phenoModel(Subject, time, dose, lCl, lV),
                 data = pheno.dat, fixed = lCl + lV ~ 1,
                 random = lCl + lV ~ 1, start = c(-5, 0),
                 na.action = na.pass, naPattern = ~ !is.na(conc) )

> summary(model2)
. . .
Data: pheno.dat
      AIC      BIC      logLik
 985.4241 1003.685 -486.712

Random effects:
Formula: list(lCl ~ 1, lV ~ 1)
Level: Subject
Structure: General positive-definite, Log-Cholesky parametrization
          StdDev     Corr
lCl       0.4878416 lCl
lV       0.3930599  1
Residual 2.8701292

Fixed effects: lCl + lV ~ 1
      Value  Std.Error DF   t-value p-value
lCl -4.993383 0.07362492 95 -67.82192      0
lV   0.332162 0.05398625 95    6.15272      0 . . .
```

## Phenobarbital Example

```
> model2.ranef <- ranef( model2, augFrame = T )
> model2.ranef[1:10, ]
```

	lCl	IV	Wt	Apgar	ApgarInd	time	dose	conc
42	1.2758744	1.0279839	2.8	9	>= 5	54.75	NA	13.60000
28	0.9923130	0.7995461	3.2	9	>= 5	2.00	NA	16.90000
30	0.3338673	0.2690272	1.8	8	>= 5	116.30	NA	17.20000
56	-0.7827877	-0.6307185	0.6	4	< 5	20.00	NA	18.80000
46	-0.2725951	-0.2196409	1.1	8	>= 5	2.00	NA	20.10000
5	0.3224235	0.2597799	1.4	7	>= 5	64.50	NA	17.56667
55	0.1339530	0.1079283	1.1	4	< 5	74.00	NA	20.80000
32	1.4548125	1.1721809	3.6	9	>= 5	38.00	NA	18.66667
43	-0.6168281	-0.4970031	0.9	1	< 5	2.00	NA	22.30000
29	-0.3749222	-0.3020840	1.0	7	>= 5	47.50	NA	22.90000

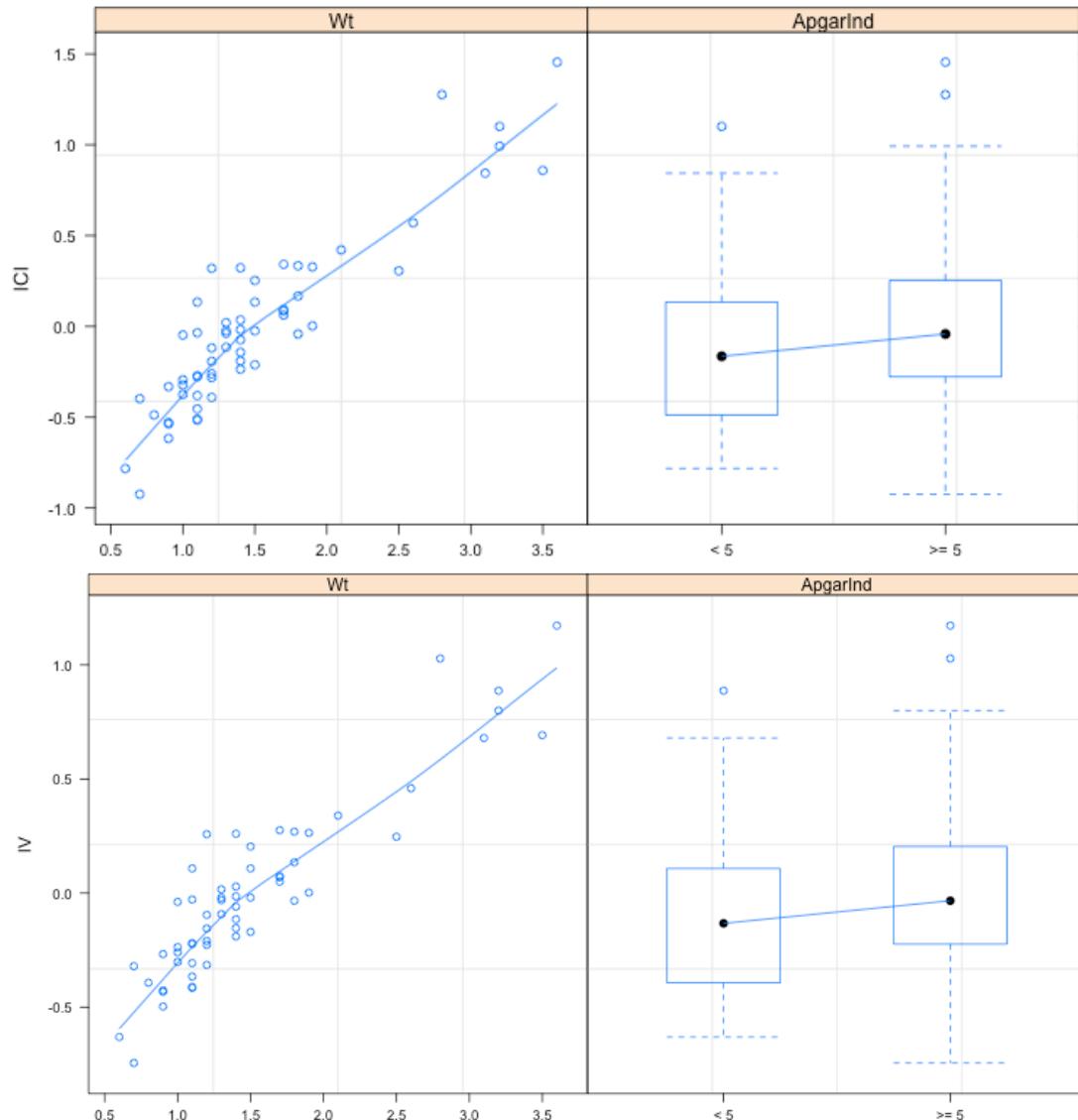
Predicted random effects  
ICI and IV

Potential covariates

Adds columns 3-8

## Plots of predicted logCL and logV random effects vs. Potential Covariates (under model without covariates and correlated random effects)

- Just like in linear regression, and the interpretation of residuals vs. covariates: a linear trend indicates the need of including more sources of variability in a linear way.
- Our “residuals” at a population level are the predicted random effects.
- Here, for the predicted random effects with no covariates, a linear relationship with Weight is apparent (both ICI and IV).
- There may be an effect of the ApgarInd variable on both the predicted logCL and logV random effects although it seems small.



```
plot( model2.ranef, form = lCl ~ Wt + ApgarInd, aspect=1 )
plot( model2.ranef, form = lV ~ Wt + ApgarInd, aspect=1 )
```

Phenobarbital Example

## Phenobarbital Example

### R Code to compute descriptive statistics from a longitudinal data set in R (within-subject fixed covariates)

- Recall that the data set has multiple rows per subject.
- The R function gsummary() will subset the data into one with one row per subject.
- The invariant=T option will retrieve only those variables that are invariant within subjects.

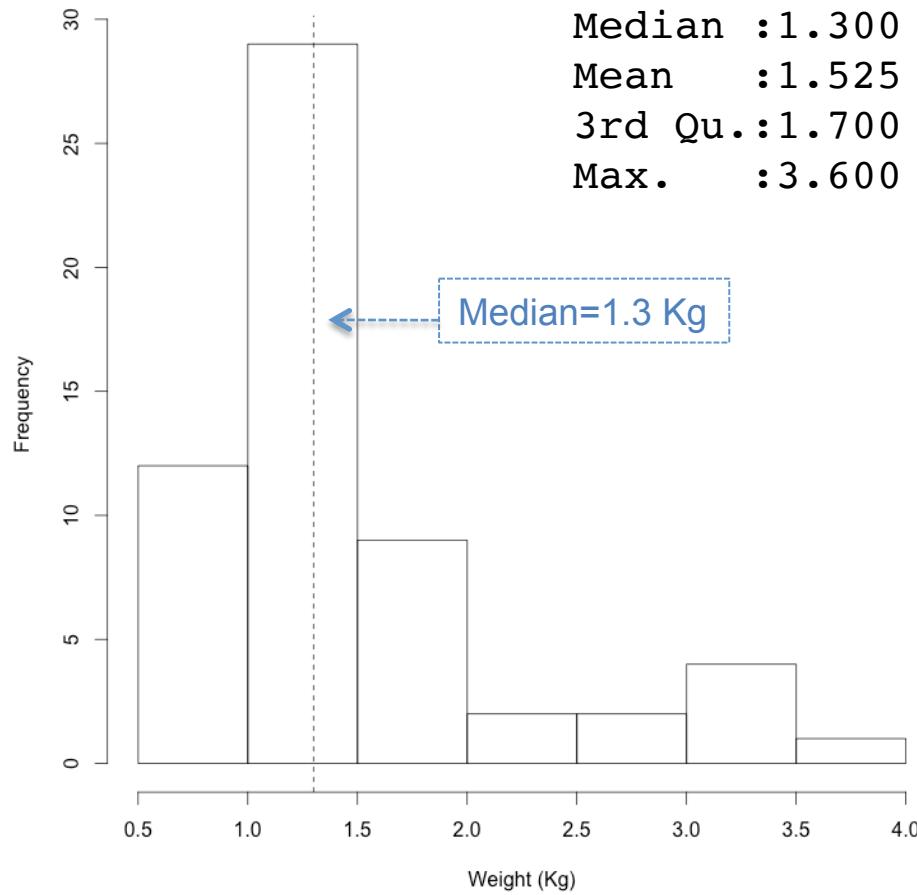
```
> sub.dat <- gsummary(pheno.dat,invariant=T)
> head(sub.dat)
  Subject  Wt Apgar ApgarInd
42      42  2.8     9    >= 5
28      28  3.2     9    >= 5
30      30  1.8     8    >= 5
56      56  0.6     4     < 5
46      46  1.1     8    >= 5
5       5  1.4     7    >= 5
> # Verifying that data set has all 59 subjects:
> nrow(sub.dat)
[1] 59
```

For more details, see Pinheiro & Bates (2010), section 3.4.

## Phenobarbital Example

```
> summary(sub.dat[,c("Wt", "ApgarInd")])
```

Wt	ApgarInd
Min. :0.600	< 5 :10
1st Qu.:1.100	= 5:49
Median :1.300	
Mean :1.525	
3rd Qu.:1.700	
Max. :3.600	



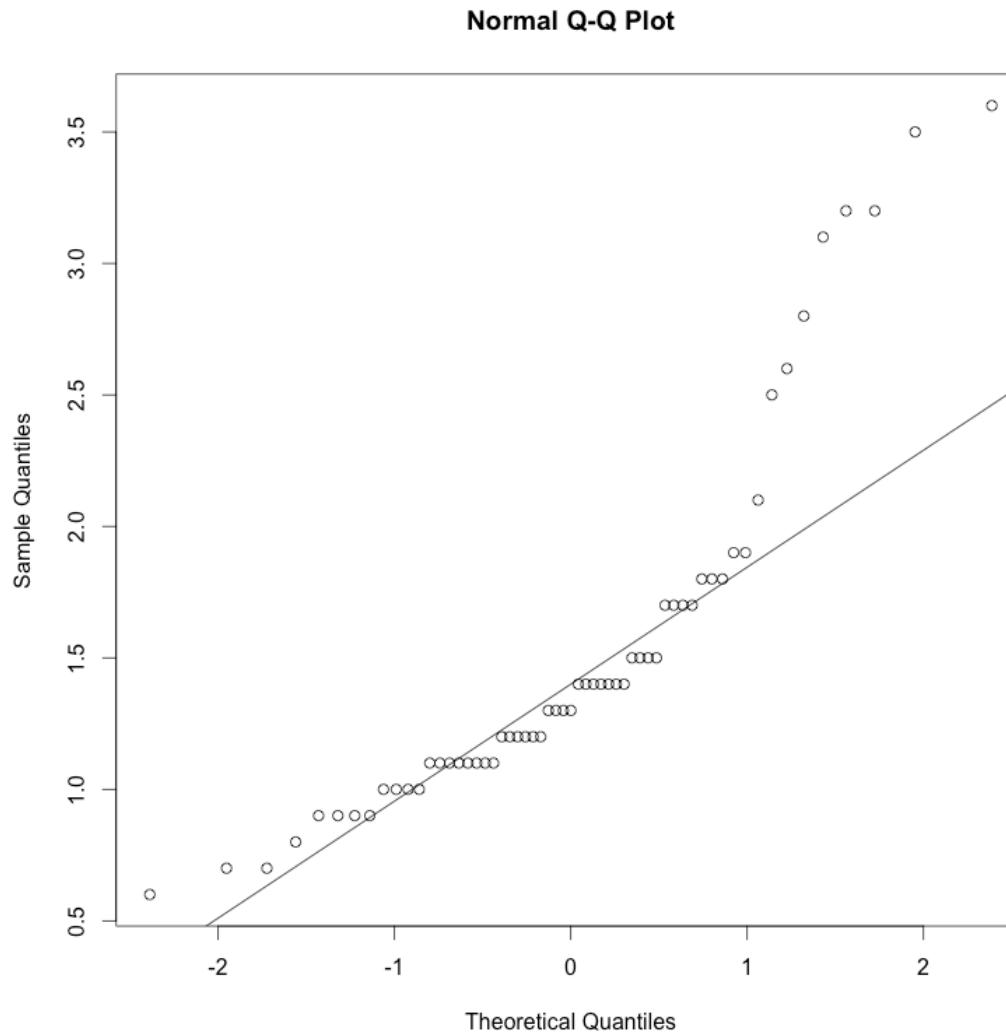
Comparison between Apgar groups must be taken with caution due to unbalanced sample size.

Weight has a skewed distribution, which must be taken into account when centering this variable in the model.

```
hist(sub.dat$Wt,xlab="Weight (Kg)",main=""); abline(v=1.3,lty=2)
```

## Phenobarbital Example

### Distribution of Weight (Kg)



```
qqnorm(sub.dat$Wt)  
qqline(sub.dat$Wt)
```

## Phenobarbital Example

### Covariate Model (sub-model) Specification Used in R

- Based on exploratory analysis of predicted PK parameters logCL and logV, we specify a model that leads to a linear relationship with Wt and ApgarInd:

Exponential Covariate  
Model for Wt and  
ApgarInd and  $\theta_1, \theta_4$  re-  
parameterized to their  
log-transformation.

Leading to the linear  
model:

$$CL_i = \exp(\ln \theta_1 + \theta_2 Wt_i + \theta_3 ApgarInd_i) e^{\eta_{CL_i}}$$

$$V_i = \exp(\ln \theta_4 + \theta_5 Wt_i + \theta_6 ApgarInd_i) e^{\eta_{V_i}}$$

$$\ln CL_i = (\ln \theta_1 + \theta_2 Wt_i + \theta_3 ApgarInd_i) + \eta_{CL_i}$$

$$\ln V_i = (\ln \theta_4 + \theta_5 Wt_i + \theta_6 ApgarInd_i) + \eta_{V_i}$$

## Phenobarbital Example

### R Code: Centering Weight to its Median and Fitting the Model

```
> # median Wt
> sub.dat <- gsummary(pheno.dat)
> sub.dat$Subject <- as.numeric(as.character(sub.dat$Subject))
> sub.dat <- sub.dat[order(sub.dat$Subject),]
> head(sub.dat)
  Subject  Wt Apgar ApgarInd      time      dose      conc       swt
1        1 1.4     7    >= 5 54.98333 5.650000 24.15000 1.0769231
2        2 1.5     9    >= 5 60.32000 4.733333 22.43333 1.1538462
3        3 1.5     6    >= 5 67.37333 5.891667 22.03333 1.1538462
4        4 0.9     6    >= 5 60.77857 3.781818 25.46667 0.6923077
5        5 1.4     7    >= 5 60.96429 5.545455 17.56667 1.0769231
6        6 1.2     5    >= 5 66.26667 4.750000 22.93333 0.9230769
> median(sub.dat$Wt)
> [1] 1.3
> pheno.dat$cwt <- pheno.dat$Wt - median(sub.dat$Wt)
```

```
> # fitting model with cwt and ApgarInd for lC1 and lV
> model.covs1 <- update(model2,
+                         fixed = list(lC1 ~ cwt + ApgarInd, lV ~ cwt + ApgarInd),
+                         start = c(-4.9933, 0, 0, 0.39306, 0, 0),
+                         control = list(pnlsTol = 0.1))
```

We center Wt instead of scaling because estimation is done based on the linear (log) scale.

Starting values:  $\text{start} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)$

lC1 parameters, lV parameters

## Phenobarbital Example

```
> summary(model.covs1)
...
      AIC      BIC    logLik
886.3146 916.7489 -433.1573

Random effects:
Formula: list(lC1 ~ 1, lV ~ 1)
Level: Subject
Structure: General positive-definite, Log-Cholesky parametrization
```

	StdDev	Corr	
lC1.(Intercept)	0.2072855	lC.(I)	$\hat{\psi}_{CL}$ ,
lV.(Intercept)	0.1294963	1	$\hat{\psi}_V$ , $\hat{Corr}(\eta_{CL}, \eta_V)$
Residual	2.8685512		$\hat{\sigma}_\epsilon$

Fixed effects: list(lC1 ~ cWt + ApgarInd, lV ~ cWt + ApgarInd)

	Value	Std.Error	DF	t-value	p-value
lC1.(Intercept)	-5.122725	0.15072299	91	-33.98768	0.0000
lC1.cWt	0.630453	0.07743436	91	8.14177	0.0000
lC1.ApgarInd>= 5	-0.009575	0.15388732	91	-0.06222	0.9505
lV.(Intercept)	0.235269	0.06039444	91	3.89554	0.0002
lV.cWt	0.527490	0.03508085	91	15.03641	0.0000
lV.ApgarInd>= 5	-0.035828	0.06471774	91	-0.55360	0.5812

## Phenobarbital Example

```
> # removing ApgarInd from lCl model
> model.covs2 <- update(model2,
+                         fixed = list(lCl ~ cWt, lV ~ cWt + ApgarInd),
+                         start = c(-4.9933, 0, 0.39306, 0, 0),
+                         control = list(pnlsTol = 0.1))
> summary(model.covs2)
...
AIC      BIC      logLik
884.3158 911.7067 -433.1579
```

Random effects:

Formula: list(lCl ~ 1, lV ~ 1)

Level: Subject

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
lCl.(Intercept)	0.2073912	lC.(I)
lV.(Intercept)	0.1295271	1
Residual	2.8681963	

Fixed effects: list(lCl ~ cWt, lV ~ cWt + ApgarInd)

	value	Std.Error	DF	t-value	p-value
lCl.(Intercept)	-5.131606	0.04851575	92	-105.77197	0.0000
lCl.cWt	0.632401	0.07002963	92	9.03048	0.0000
lV.(Intercept)	0.235279	0.06016313	92	3.91069	0.0002
lV.cWt	0.527079	0.03423004	92	15.39813	0.0000
lV.ApgarInd>= 5	-0.035658	0.06448858	92	-0.55294	0.5816

## Phenobarbital Example

```
> # removing ApgarInd from lV model
> model.covs3 <- update(model2,fixed = list(lC1 ~ cWt, lV ~ cWt),
  start = c(-4.9933, 0, 0.39306, 0),
  control = list(pnlsTol = 0.1))
> summary(model.covs3)
AIC      BIC      logLik
882.641 906.9884 -433.3205

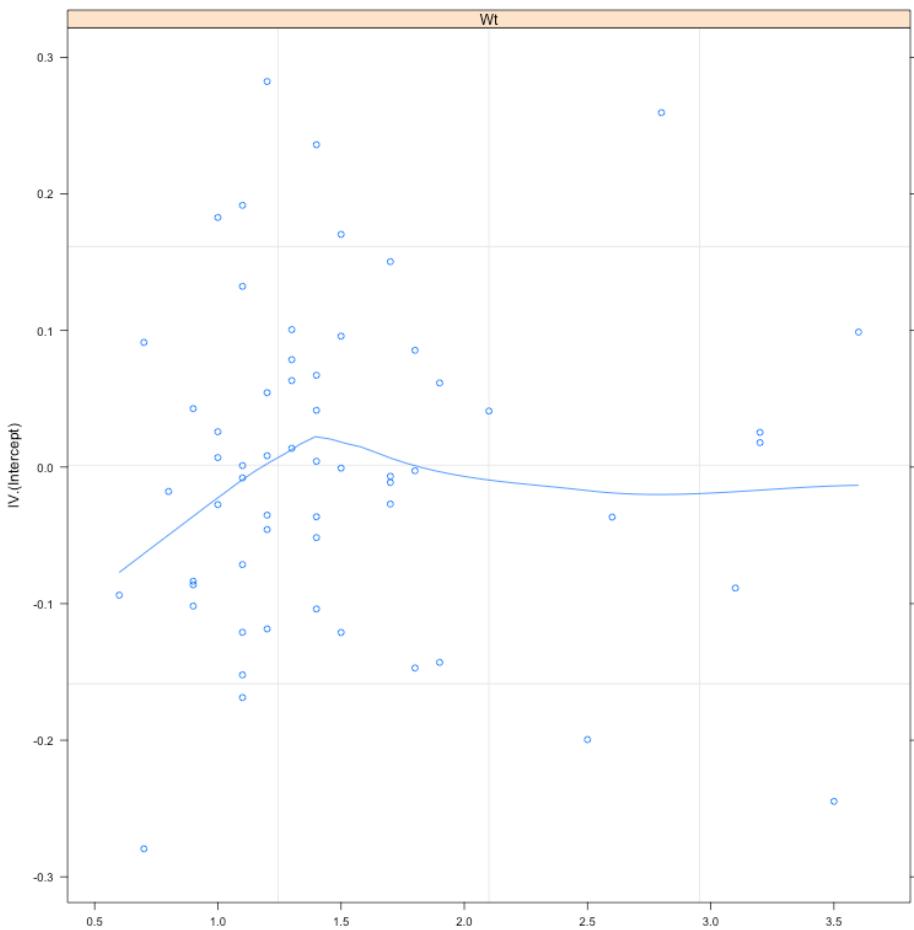
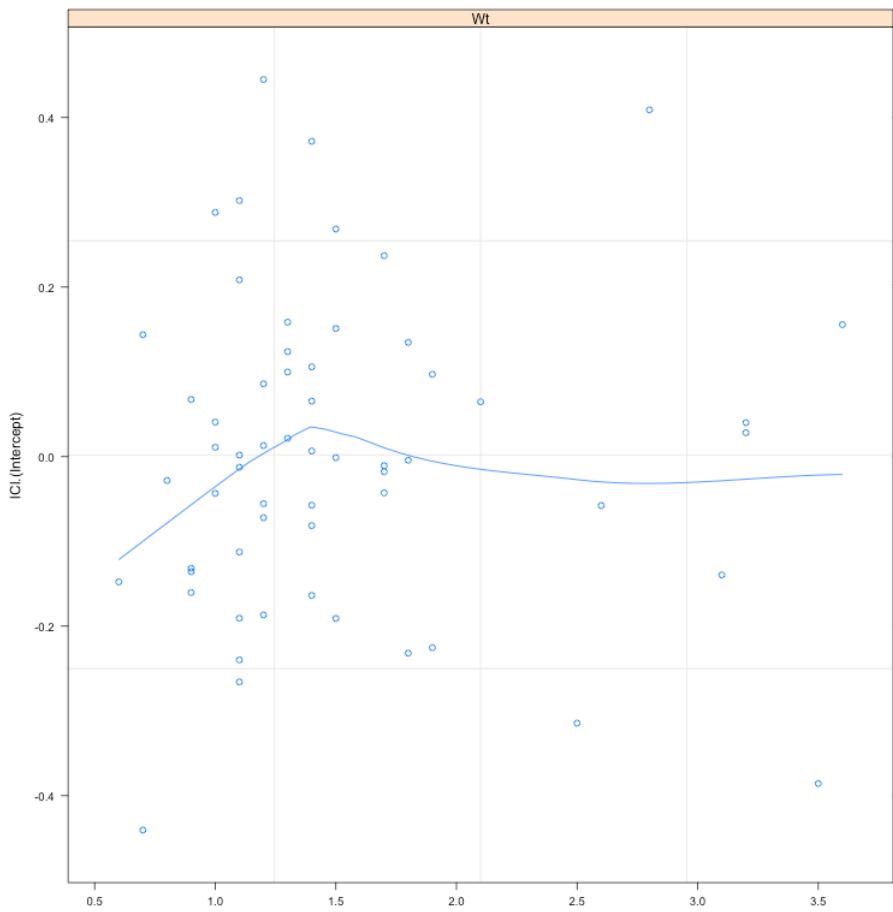
Random effects:
Formula: list(lC1 ~ 1, lV ~ 1)
Level: Subject
Structure: General positive-definite, Log-Cholesky parametrization

          StdDev     Corr
lC1.(Intercept) 0.2064273 lC.(I)
lV.(Intercept)  0.1311690 1
Residual        2.8651241

Fixed effects: list(lC1 ~ cWt, lV ~ cWt)
              Value  Std.Error DF   t-value p-value
lC1.(Intercept) -5.129138 0.04794210 93 -106.98608      0
lC1.cWt         0.625231 0.06754182 93    9.25695      0
lV.(Intercept)  0.205124 0.02494470 93    8.22317      0
lV.cWt          0.528498 0.03408686 93   15.50446      0
```

## Phenobarbital Example

Plots of the predicted random effects for  $lC1$  and  $IV$  vs. Weight show a slight linear trend for Wt up to  $\sim 1.4$  Kg, however has no trend for values above 1.4. `model1.covs3` is better than the model with no covariates.



## Phenobarbital Example

### Final Model Full Specification With log-parameterization of $\theta_1, \theta_3$

$$C_{ij} = \sum_{d:t_{id} < t_{ij}} \frac{D_{id}}{\underbrace{\exp(\ln \theta_3 + \theta_4 Wt_i^*) e^{\eta_{Vi}}}_{\mu_{Vi}}} \exp \left\{ - \frac{\exp(\ln \theta_1 + \theta_2 Wt_i^*) e^{\eta_{CLi}}}{\exp(\ln \theta_3 + \theta_4 Wt_i^*) e^{\eta_{Vi}}} (t_{ij} - t_{id}) \right\} + \varepsilon_{ij};$$

$$Wt_i^* = Wt_i - Median(Wt_i)$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2),$$

$$\begin{bmatrix} \eta_{CL_i} \\ \eta_{V_i} \end{bmatrix} \sim BVN(0, \Psi), \quad \Psi = \begin{bmatrix} \psi^2_{CL} & \psi_{CL,V} \\ & \psi^2_V \end{bmatrix}.$$

## Phenobarbital Example

### Interpretation of PopPK parameters

- The mean CL for typical values of Wt (median=1.3 Kg) is  $\exp(-5.1292)=0.0059$  L/(Kg\*hr).
- The mean V for typical values of Wt (median=1.3 Kg) is  $\exp(0.2051)=1.2277$  L/Kg.
- On average, a 1 Kg increase in Wt would result in a significant increase of  $\exp(0.6252)=1.8687$  times the value in CL (p-val<0.001).
- On average, a 1 Kg increase in Wt would result in a significant increase of  $\exp(0.5285)=1.6964$  times the value in V (p-val<0.001).

Fixed effects: list(lC1 ~ cWt, lV ~ cWt)	Value	Std.Error	DF	t-value	p-value
lC1.(Intercept)	-5.129138	0.04794210	93	-106.98608	0
lC1.cWt	0.625231	0.06754182	93	9.25695	0
lV.(Intercept)	0.205124	0.02494470	93	8.22317	0
lV.cWt	0.528498	0.03408686	93	15.50446	0

The units of Cl and V were obtained through the unit builder in Phoenix. Discussion arose regarding to whether it is reasonable to have e.g., L/Kg units for V (i.e., L/Kg should not vary with Wt as it is a standardized measure by weight.) Question remains, should the units for V be in L?

## Interpretation of PopPK parameters

- The BSV in CL for typical values of Wt (median 1.3 Kg) is  $CV_{CL}=20.6\%$ .
- The BSV in V for typical values of Wt is  $CV_V=13.12\%$

**Random effects:**

	StdDev	Corr
lc1.(Intercept)	0.2064273	lc.(I)
lv.(Intercept)	0.1311690	1

# Implementation in Phoenix

## Phenobarbital Data Example

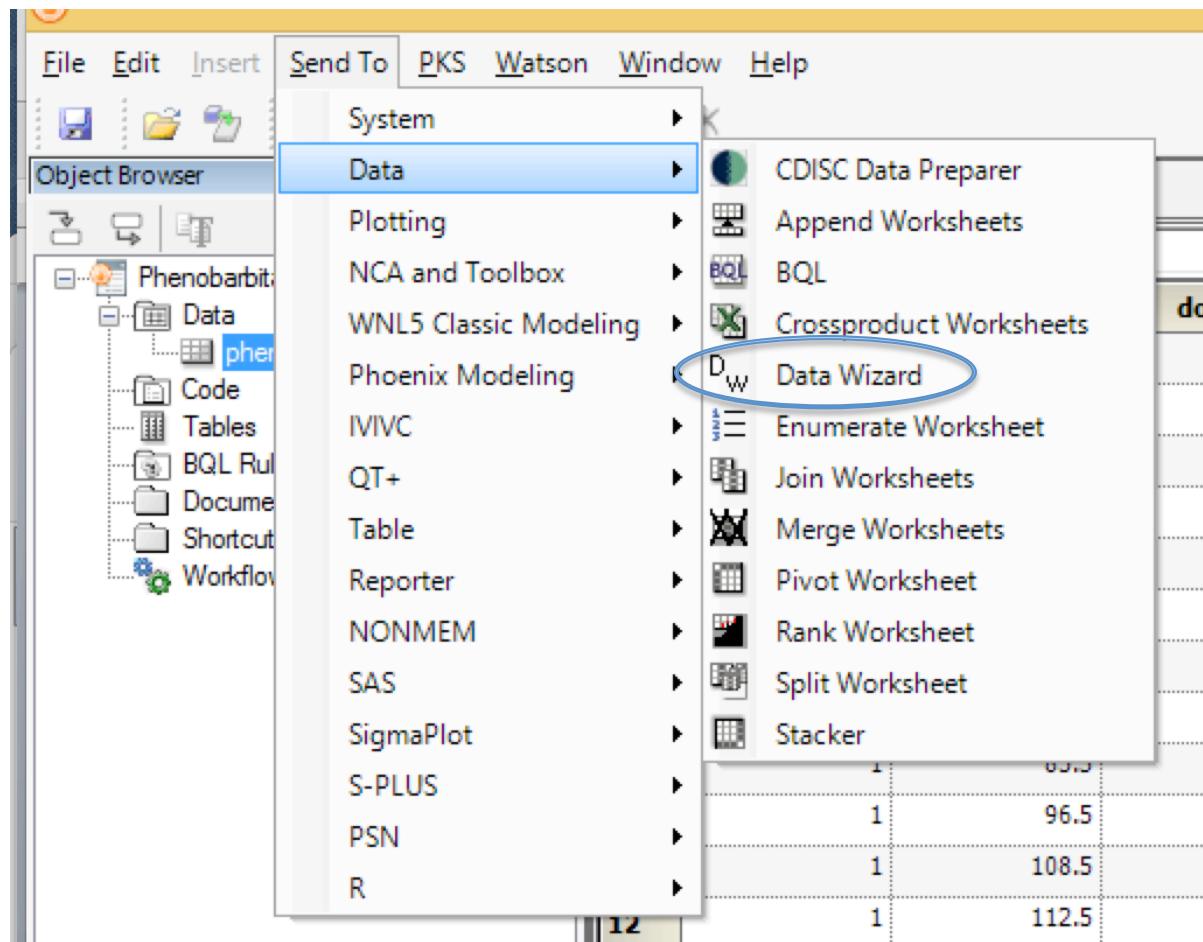
## Creating an indicator variable for Apgar Score

1. After clicking on the data set,

Select *Send To*

-> *Data*

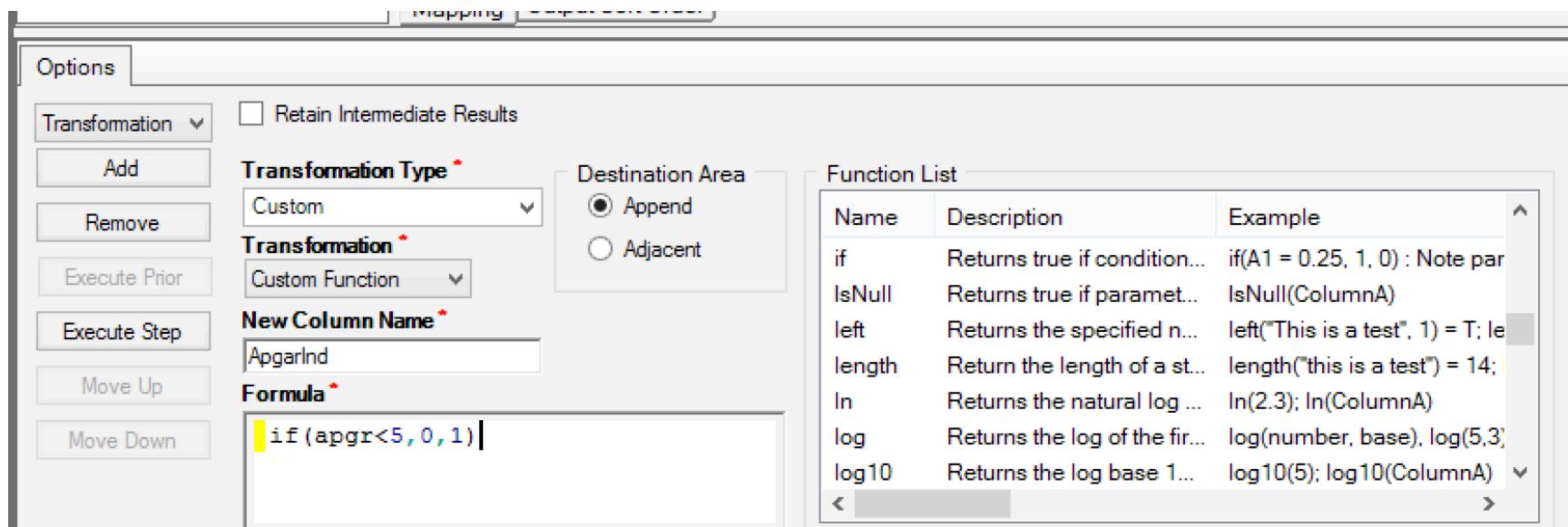
-> *Data Wizard*



## Creating an indicator variable for Apgar Score

Note: if Phoenix presents an error while executing this transformation, create a model and map the units before attempting this again.

2. Select the *Transformation* Option
3. Click on *Add*
4. Select *Custom* in the *Transformation Type* box
5. Give the name of the new variable, e.g. ApgarInd in the *New Column Name* field
6. Type *if(apgr<5,0,1)* in the *Formula* field.
7. Click on the *Execute* icon.



## Creating an indicator variable for Apgar Score

Copy the contents in the new ApgarInd created variable:

7. Right click on new ApgarInd column in the *Data Wizard* and select *Edit/Copy* from the drop down menu.
8. Right click on the “idum1” column in the original data set, select *Insert Column* and name it, e.g. ApgarInd
9. Right click on the empty ApgarInd column and right click/select *Paste*.

	xid	time (hr)	dose (ug/kg)	wgt (kg)	apgr	yobs (ug/L)	ApgarInd	idum1	idum2
1		1	0	25	1.4	7		1	1
2		1	2		1.4	7	17.3	1	0
3		1	12.5	3.5	1.4	7		1	1
4		1	24.5	3.5	1.4	7		1	1
5		1	37	3.5	1.4	7		1	1
6		1	48	3.5	1.4	7		1	1
7		1	60.5	3.5	1.4	7		1	1
8		1	72.5	3.5	1.4	7		1	1
9		1	85.3	3.5	1.4	7		1	1
10		1	96.5	3.5	1.4	7		1	1
11		1	108.5	3.5	1.4	7		1	1
12	R1	1	112.5		1.4	7	31	1	0
13		2	0	15	1.5	9		1	1
14		2	2		1.5	9	9.7	1	0
15		2	4	3.8	1.5	9		1	1
16		2	16	3.8	1.5	9		1	1
17		2	27.8	3.8	1.5	9		1	1

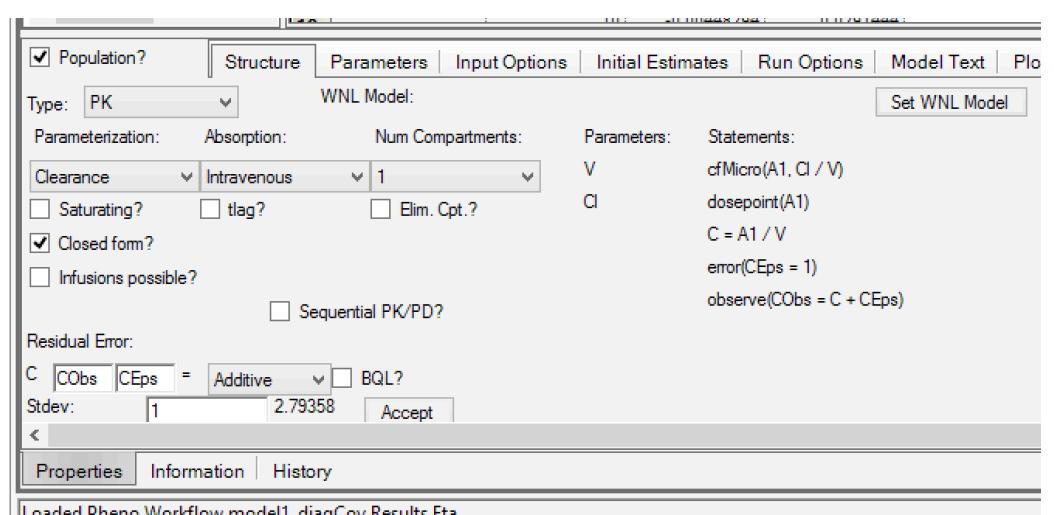
## Fitting a model with no covariates first, to obtain starting values.

1. Create a new model by clicking in the Send To/Phoenix Model/Phoenix model drop down menu, as done in the previous exercise.
2. Map the variables xid->ID, time->Time, dose->A1, yobs->CObs.

Note that the *Population?* Box must be checked in order to be able to map the xid variable.

	None	Sort	ID	A1	Time	CObs
xid	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
time	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
dose	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
wgt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
apgr	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
yobs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

3. Select the Clearance IV PK model as shown below, with additive residuals, in the Structure tab.



## Fitting model with no covariates

3. In the *Parameters/Structural* tab, select the “exp(Sum+eta)” option

The screenshot shows the SPAM software interface with the 'Parameters' tab selected. Under the 'Structural' tab, the 'Covar. Type' section is active. The 'Style' dropdown menu is open, showing options like 'exp(Sum+eta)', 'tvV', and 'nV'. The 'exp(Sum+eta)' option is highlighted. Below this, there are sections for 'SPam' parameters (V and Cl) and 'Covariate' settings (wgt and ApgarInd). The 'Covariate' section includes fields for 'Center', 'Pos?', 'Direction', 'V', and 'Cl', all set to 'No'.

4. Add the wgt and ApgarInd variables under the *Covariate* heading but do not include them in the model. Covariate plots will be produced but a model with no covariates will be fitted.

5. In the *Parameters/Covar.Type* tab, fill in the fields for the ApgarInd variables as shown:

The screenshot shows the 'Parameters/Covar.Type' tab. For the variable 'ApgarInd', two categories are defined: '<5' and '>=5'. The corresponding values are 0 and 1 respectively. An 'Add' button is visible at the bottom.

## Fitting model with no covariates

6. Uncheck that the “Diag” box in the *Parameters/Random Effects* tab is to enable an unstructured variance-covariance random effects matrix. If you are running this model for the first time you will have zero values (as opposed as in the image below) and empty boxes, this is OK.

Structure   Parameters   Input Options   Initial Estimates  
Fixed Effects   Random Effects   Secondary   Scenario  
Freeze   Diag  
0.01 < 0.016  
-6.2 < 0.026   2.04 < 0.042

Population?   Structure   Parameters   Input Options   Initial Estimates   Run Options   Model  
Method: FOCE L-B   Stderr: Central Diff   Run Mode: Simple  
Lindstrom-Bates: Best with Additive or Log-additive error  
Method: Sandwich  
N Iter: 1000   Confidence Level %: 95  
NonParametric   PCWRES?  
Sort Input?   MAP-NP Start?  
Use MPI?   Advanced >  
Max ODE: matrix exponent  
Synthetic Gradients?

7. Select the *FOCE-LB Method* in the *Run Options* tab.
8. Click on the *Execute* icon.



## Fitting model with no covariates

### Results

- The *Theta Output Data* sheet should show the following:

Scenario	Parameter	Estimate	Units	Stderr	CV%	2.5% CI	97.5% CI
1	tvV	0.331733		0.051717502	15.590099	0.22953853	0.43392747
2	tvCl	-4.99269		0.07350102	-1.4721727	-5.137929	-4.847451
3	stdev0	2.85621		0.30313198	10.613085	2.2572172	3.4552028

- Note that the coefficients are very similar to those obtained in R (`model$coef$fixed`).
- The 95% CI's for the intercepts do not include the value of zero. This is consistent with the results obtained via p-values in R: Intercepts are statistically significant.

## Fitting model with covariates

1. Right click on the previous model and copy/paste in the *Workflow* section. Rename it.
2. In the *Parameters/Structural* tab, keep the “exp(Sum+eta)” option

**Population?** **Structure** **Parameters** **Input Options** **Initial Estimates** **Run Options** **Model Text** **Plots** **no warn**

**Structural** **Covar. Type** **Fixed Effects** **Random Effects** **Secondary** **Scenarios** **Edit**

SPam Style Fixef Ran Ranef Code

V      $V = \exp(tvV + (\text{wgt}-\text{median}(\text{wgt})) * dVdwgt + (\text{ApgarInd}=1) * dVdApgarInd1 + nV)$

CI      $CI = \exp(tvCI + (\text{wgt}-\text{median}(\text{wgt})) * dCldwgt + (\text{ApgarInd}=1) * dCldApgarInd1 + nCI)$

**Covariate** **Center** **Pos?** **Direction** **V** **CI**

x <input type="text" value="wgt"/>	median	<input type="checkbox"/>	Forward	Yes	Yes
x <input type="text" value="ApgarInd"/>		<input checked="" type="checkbox"/>	Forward	Yes	Yes

**Add Covariate** **Add From Unused Columns**

3. Add the wgt and ApgarInd variables in the fields under the *Covariate* heading and include them in the model. Center wgt to its median.

4. In the *Parameters/Covar. Type* tab, make sure that the fields for the ApgarInd variables look as shown:

**Population?** **Structure** **Parameters** **Input Options** **Initial Estimates** **Run Options** **Model Text** **Plots** **no warn**

**Structural** **Covar. Type** **Fixed Effects** **Random Effects**

ApgarInd **Type: Categorical**

Cat. Name	Value
<5	0
>=5	1

**Add**

## Fitting model with covariates

5. In the Parameters/Fixed Effects tab, click “Accept” for the suggested initial values for the intercepts.

Population?		Structure	Parameters	Input Options	Initial Estimates	
Structural	Covar. Type	Fixed Effects	Random Effects	Secondary	So	
Fixef	Initial	Lower *	Upper *	Freeze	Estimate	Units *
tvV	0.3317			<input type="checkbox"/>	0.331733	<input type="button" value="Accept"/>
tvCl	-4.992			<input type="checkbox"/>	-4.99269	<input type="button" value="Accept"/>
dVdwgt	0			<input type="checkbox"/>		
dCldwgt	0			<input type="checkbox"/>		
dVdApgarInd	0			<input type="checkbox"/>		
dCldApgarInd	0			<input type="checkbox"/>		

[Accept All Fixed+Random](#)

6. Since this is a copy of the model with no covariates with unstructured variance-covariance matrix, which was fitted via the FOCE-LB method, selected options will remain unchanged. The only thing left is to click on the *Execute* button.



## Main Results

- The *Theta Output Data* sheet should show the following:

	Scenario	Parameter	Estimate	Units	Stderr	CV%	2.5% CI	97.5% CI
1		tvV	0.23566		0.060497107	25.671352	0.11608987	0.35523013
2		tvCl	-5.12249		0.11443339	-2.2339407	-5.348663	-4.896317
3		dVdwgt	0.527758		0.038696253	7.3321964	0.4512764	0.6042396
4		dCldwgt	0.63032		0.085416626	13.551311	0.46149744	0.79914256
5		dVdApgarInd1	-0.0358032		0.063632539	-177.72864	-0.16157038	0.089963982
6		dCldApgarInd1	-0.0104439		0.10791663	-1033.2982	-0.22373683	0.20284903
7		stdev0	2.86861		0.25431477	8.8654354	2.365967	3.371253

- Note that the coefficients are very similar to those obtained in R \ (model\$coef\$fixed)
- The 95% CI's for ApgarInd in the Clearance and Volume PK parameters include the value of zero. This is consistent with the results obtained via p-values in R: ApgarInd is not significant for either PK parameter.

## Main Results

- The *Omega Output Data* results should show the following:

The screenshot shows the 'Omega Output Data' interface. On the left is a vertical navigation bar with icons and labels: Filter, Omega (selected), OmegaStderr, Overall, Residuals, Secondary, StrCovariate, StrCovariateCat, Theta, and ThetaCorrelation. The main area has a table with columns: Scenario, Label, nV, and nCI. The data rows are numbered 1 through 7.

Scenario	Label	nV	nCI
1	Omega		
2	nV	0.016829616	
3	nCI	0.026803517	0.042702125
4	Correlation		
5	nV	1	
6	nCI	0.99983869	1
7	Shrinkage	0.11561057	0.11564214

- Note that the estimates of variance and correlation are very similar to those obtained in R (`VarCorr(model)`).

Exercise: produce a model with only Weight for Clearance and Volume by copying and pasting the model in the *Workflow* section (as done in previous exercises – see LME4 slides for more details.)

Nothing should change but the specification of the new model in the *Parameters/Structural* tab.

## Model Comparer

- Select the *Insert/Model Comparer* drop down menu and an object with the same name should appear in the *Workflow* section, containing information with all the models that you have fit. For example:

The screenshot shows the Model Comparer interface with the following components:

- Top Bar:** Contains tabs for "Setup", "Results", and "Verification".
- Left Sidebar:** Shows a tree view of models under "Model". The expanded node "Pheno Model" contains "Pheno Model\_Diag, no cov" and "Pheno Model\_Unstr,wt\_exp".
- Main Table:** A grid displaying model details. Columns include "Hide", "Compare", "Name", "Sort", "Method", "Description", and "Line". The table lists several models, including "Pheno Model\_Diag, no cova\_exp\_sum\_eta" (FOCE L-B), "Pheno Model\_Unstr, no cova\_product\_exp\_eta" (FOCE L-B), and "Model" (FOCE ELS). Other models listed include various copies and specific parameterizations.
- Bottom Panel:** An "Options" panel with sections for "Columns", "Worksheets", "Plots", and "Plot Options".

  - Columns:** Includes checkboxes for Hide, Compare, Name, Sort, Method, Description, Lineage, LogLik, -2(LL), AIC, BIC, -2(LL)Delta, AIC Delta, BIC Delta, and #Params.
  - Worksheets:** Includes checkboxes for Eta, Omega, Overall, Residuals, and Theta.
  - Plots:** Includes checkboxes for Convergence, Pop DV vs PRED, Pop DV vs PRED Log, Pop DV vs TAD, Pop DV vs IPRED, Pop DV vs IPRED Log, Pop CWRES vs PRED, Pop QQ WRES, Pop QQ CWRES, Pop WRES vs IVAR, Pop CWRES vs IVAR, Pop DV, PRED vs IVAR, Pop DV, IPRED vs IVAR, Pop DV, PRED vs TAD, and Pop DV, IPRED vs TAD.
  - Plot Options:** Includes fields for Height (200), Width (300), Format (Gif), and checkboxes for Hide Selected and Full Names.

- Bottom Navigation:** Buttons for "Properties", "Information", and "History".
- Bottom Status:** Shows "VirtualBoxVM" and other system icons.

## Model Comparer

2. Check the boxes for the two models to compare (and hide the rest): that is, one with Wt and ApgarInd and one with only Wt (both with the same var-covar structure). For example:

Phenobarbital >> Workflow >> Model Comparer

Setup Results Verification

	Hide	Compare	Name	Sort	Method	Description	Lin
Pheno Model_Diag, no cov	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Pheno Model_Diag, no cova_exp_sum_eta		FOCE L-B		
Pheno Model_Unstr,wt_exp	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Pheno Model_Unstr, no cova_product_exp_eta		FOCE L-B		Phe
Model	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Pheno Model_Unstr,wt_exp_sum_eta		FOCE L-B		
	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Copy of Pheno Model_Diag, no cova_product_exp_eta		FOCE L-B		Phe
	<input type="checkbox"/>	<input type="checkbox"/>	Copy of Pheno Model_Unstr, no cova_exp_sum_eta		FOCE L-B		Phe
	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Pheno Model_Unstr,wtMEDIAN_exp_sum_eta		FOCE L-B		Phe
	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Pheno Model_Unstr,wtMEDIAN_exp_sum_etaPROGRES		FOCE L-B		Phe
	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Pheno Model_Unstr,wtMEDIAN_ApgarInd_exp_sum_etaPROGRES		FOCE L-B		Phe
	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Pheno Model_Unstr,wtMEDIAN_ApgarInd_exp_sum_eta		FOCE L-B		Phe
▶	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Model		FOCE ELS		

3. Click the Execute icon.



## Model Comparer

The Results tab should show the most important information from both models pertaining Goodness of Fit measures, predicted random effects, variance-covariance matrix of random effects, residuals and fixed effect estimates.

The screenshot shows the 'Results' tab of the Model Comparer software. The interface includes a top navigation bar with tabs for 'Setup', 'Results' (which is selected), and 'Verification'. Below the navigation bar is a toolbar with various icons. A left sidebar contains a tree view of output data categories: 'Output Data' (selected), 'Comparison Result' (highlighted in blue), 'Eta', 'Omega', 'Overall', 'Residuals', 'Theta', 'Plot tables' (highlighted in red), 'Plot table', 'Text Output', and 'Settings'. A 'Filter:' input field is also present in the sidebar. The main content area displays a table with columns: 'Hide', 'Compare', 'Name', 'Sort', 'Method', 'Description', and a header row. Two rows are listed:

	Hide	Compare	Name	Sort	Method	Description
1	False	True	Pheno Model_Unstr,wtMEDIAN_exp_sum_eta		FOCE L-B	Pheno Model_Unst
2	False	True	Pheno Model_Unstr,wtMEDIAN_ApgarInd_exp_sum_eta		FOCE L-B	Pheno Model_Unst

## Covariate Sub-Model Specification

Selection of covariates and their functional relationship with a PK parameter go hand in hand:

- Once a covariate is identified as being significant to a PK parameter, the question is how best to model their relationship.
- By failing to detect an appropriate functional relationship there is risk of failing to detect any relationship at all.

# Covariate Selection Techniques

1. Covariate screening **estimate best effect**
  - Regression of the EBEs of the  $\eta$ 's (predicted  $\eta$ 's) vs. the covariate.
  - Alternatively, use EBEs of the parameter of interest (e.g. predicted logCL) as dependent variable.
2. Graphical approach
  - Visually determine a possible relationship through plots of predictors of the EBEs of the  $\eta$ 's vs. the potential covariate.
3. Direct testing
  - Perform a test of significance in the NLME model (t-tests, CI's, LRT for more than one regression coefficient).
  - Compare with a candidate model with a different functional relationship AIC (as long as they have the same number of estimable parameters, e.g. linear vs. power.)

## Power and Sample Size with Direct Testing

Ribbing and Jonsson (2004)<sup>(1)</sup> found that:

1. With one covariate in the model, 80% power was achieved with 3 occasions/subject and varying degrees of correlation between covariate and PK parameter:
    - a) 20 subjects, high correlation: ~0.85.
    - b) 100 subjects, medium correlation: ~0.50
    - c) 300 subjects, low correlation: ~ 0.15
  2. The value of the covariate coefficient was overestimated as the number of subjects decreased leading to a false statistical significance (think  $\uparrow$ estimate/SD  $\Leftrightarrow \uparrow$ test statistic,  $\downarrow$  p-value).
- Implications for cases of sparse data on small number of subjects.
  - It is suggested that forward direct covariate screening not to be done on small to moderate sized data sets (<50 to 100 subjects.)

Recall Power= 1-Type 2 error =  $\Pr(\text{Reject } H_0 | H_0 \text{ is false})$ = false significance.

<sup>(1)</sup> Power, selection bias, and predictive performance of the population pharmacokinetic covariate model.  
*Journal fo Pharmacokinetics and Parmacodynamics*, 31:109-134.

## Covariate Selection

Bonate, 2011:

"It must be stressed that there is no universally accepted method for covariate selection in any regression-based model.

In a survey of 324 papers published on PopPK in between the years 2002 and 2004, Dartois et al. (2007)<sup>(1)</sup> reported that:

- 71% used graphical analysis,
- 22% used Generalized Additive Models (GAMs),
- 12% used a univariate test,
- 0.7% used classification and regression trees
- 21% used some other method, and
- 22% did not report which method was used.

It should be noted that the percentages here do not add to 100% because multiple methods for covariate selection were sometimes reported."

<sup>(1)</sup> Overview of model-building strategies in populations PK/PD analyses: 2002-2004 literature strategy.  
*British Journal of Clinical Pharmacology*; 64:603-612.

## Steps for Model Building

1. Determine the structural model.
  - An overparametrized model is preferred, especially regarding random effects.
2. Examine the distribution of the EBEs (predicted  $\eta$ 's) and residuals.
  - EBEs should be consistent with a  $N(0, \Psi)$ .
  - Residuals should be consistent with a  $N(0, \sigma^2 I)$  and uncorrelated.

Note: it is possible to use an Extended NLME, meaning the specification of a covariance structure to the residual variance; however, this is not covered in this course.
3. Select covariates (ie., graphically, by regression or direct testing).
4. Build model and in case of not using direct testing in (2), implement a forward/backward stepwise method using T-tests or LRT.
5. Evaluate the final parameter estimates. Just like in step 2, the distributional assumptions of the random components need to be assessed.