

Outline

(Re-Structured)

1. Overview and Motivating Example.
2. Model formulation.
3. Sources of variability and Base Model.
4. Covariate model.
5. Estimation.
6. Assumptions in the Modeling Process.
7. Precision of Estimates and Confidence Intervals..

4.5 Estimation

Specific Learning Objectives:

1. Explain the applicability and main features of the FO and FOCE-LB estimation methods.

Estimation NLME methods available in Phoenix

- Naïve pooled*
 - FO*
 - FOCE – LB*
 - FOCE ELS
 - Laplacian
 - IT2S-EM
 - QRPEM
- 
- Marginal Likelihood Approximation

*Discussed in more length.

See the NLME Phoenix User's Guide for details.

Naïve Pooled Method

- Applicable to Normally distributed data.
- Treats all observations as if they came from a single subject so does not produce predictions for η_i or estimates for Ψ .
- Only θ, σ^2 are estimated.
- Can be applied to a single individual or to subjects separately as a series of individual fits.
- Only one iteration is required.

Marginal Likelihood Methods

- Because the random effects are unobserved quantities, maximum likelihood estimation is based on the marginal density of the responses Y , which is calculated as:

$$g(y_i; \beta, \sigma^2, \Psi) = \int_{\eta} g(y | \eta; \beta, \sigma^2) g(\eta; \Psi) d\eta,$$

where

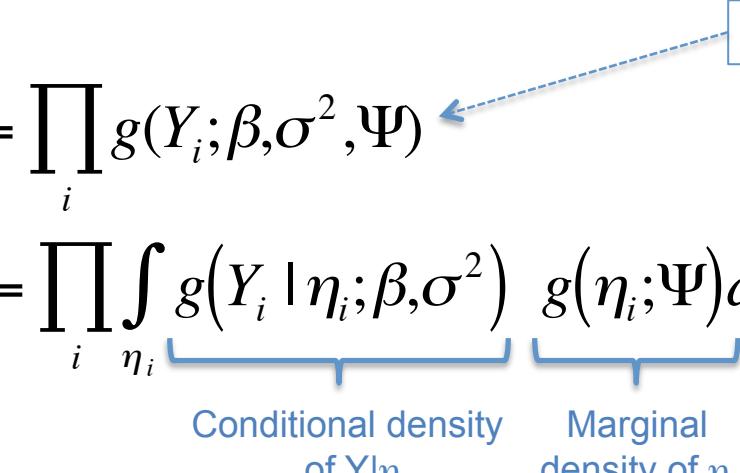
$g(y | \eta; \beta, \sigma^2)$ is the conditional density of Y given the random effects η ,
 $g(\eta; \Psi)$ is the density of η .

Notation: using $g()$ so we don't confound it with the $f()$ describing the structural non-linear model.

Estimation via Maximum Likelihood Methods

- The marginal likelihood function has the following form:

$$\begin{aligned} L(\beta, \sigma^2, \Psi) &= \prod_i g(Y_i; \beta, \sigma^2, \Psi) \\ &= \prod_i \int g(Y_i | \eta_i; \beta, \sigma^2) g(\eta_i; \Psi) d\eta_i. \end{aligned}$$



- The goal is to minimize the marginal log-likelihood.
- The integral generally does not have a closed form because of the non-linearity of the structural component of the NLME.
- Furthermore, to make the numerical optimization tractable, different types of approximations to the marginal density have been proposed.

What is a First Order Taylor Series Approximation to the Non-Linear Model?

E.g. IV Plasma Concentration (Y) model specification

$$Y_{ij} = f(x_{ij}; \beta_i) + \varepsilon_{ij};$$

$$\beta_i = \begin{bmatrix} \beta_{1i} \\ \beta_{2i} \end{bmatrix} = \begin{bmatrix} \theta_1 + \theta_2 Age_i + \eta_{CL_i} \\ \theta_3 + \eta_{V_i} \end{bmatrix}, \quad \text{where } \beta_{1i} = CL_i, \beta_{2i} = V_i;$$

$$x_{ij} = [t_{ij} \ Dose_i \ Age_i].$$

Suppose the sub-models can be expressed as follows:

$$\beta_i = d(a_i, \theta, \eta_i),$$

$$\beta_i = d(a_i, \theta, \eta_i) = a_i \theta + \eta_i, \quad a_i = \begin{bmatrix} 1 & Age_i & 0 \\ 0 & 0 & 1 \end{bmatrix}; \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}; \quad \eta_i = \begin{bmatrix} \eta_{CL_i} \\ \eta_{V_i} \end{bmatrix}.$$

So in matrix notation and after re-expressing the residuals, the model is equivalent to:

$$Y_i = f(x_i; d(a_i, \theta, \eta_i)) + \varepsilon_i^*; \quad \text{Var}(\varepsilon_i^*) = \sigma^2 I$$

$$\varepsilon_i^* = \sigma I \zeta_i, \quad \zeta_i \sim N(0, I), \quad \eta_i \sim N(0, \Psi).$$

What is a First Order Taylor Series Approximation to the Non-Linear Model?

A First Order Taylor Series Expansion⁽¹⁾ around $\eta=0$

$$Y_i \approx f(x_i; d(a_i, \theta, 0)) + Z_i(\theta, 0) \eta_i + \varepsilon_i^*;$$

$$Z_i(\theta, 0) = F_i(\theta, 0) \Delta_{\eta_i}(\theta, 0); \text{ where}$$

- $F_i(\theta, 0)$ is the matrix of derivatives of $f(x_i; \beta_i)$ with respect to β_i evaluated at $d(a_i, \theta, 0)$,
- $\Delta_{\eta_i}(\theta, 0)$ is the matrix of derivatives of $d(a_i, \theta, \eta_i)$ with respect to η_i evaluated at $\eta_i = 0$.

$$\varepsilon_i^* = \sigma I \zeta_i$$

In more complex residual models, ε^* can be function of $d(a, \theta, \eta)$ in which case ε^* would be approximated using a 1 term expansion.

- Note that the approximation above has the form of a LME model.
- It follows that the marginal distribution of Y is approximated by a Normal distribution with expectation and variance:

General expression of the First Order Taylor Approximation:

$$h(x) \approx h(x_0) + \frac{dh}{dx}(x - x_0)$$

$$E(Y_i) \approx f(x_i; d(a_i, \theta, 0)),$$

$$\text{Cov}(Y_i) \approx Z_i(\theta, 0) \Psi Z_i(\theta, 0)^T + \sigma^2 I.$$

The FO Method (Sheiner & Beal, 1982)

- First order Taylor series approximation for the non-linear model at $\eta=0$: “Linearizes the NLME model to LME”.
- Applicable to Normal data only.
- FO gains speed by not having to estimate η 's and θ, σ^2, Ψ are estimated in one single step.
- Converges almost always.
- Adequate for some compartmental models with dense sampling and small BSV.
- Will break down and give biased results for models with higher degrees of non-linearity such as sigmoid Emax with a large Hill coefficient (see Bonate, pg.255).
- Its use is suggested to obtain initial estimates for more accurate estimation methods.

A First Order Taylor Series Expansion around $\eta = \eta^*$

$$Y_i \approx f(x_i; d(a_i, \theta, \eta_i^*)) + Z_i(\theta, \eta_i^*)(\eta_i - \eta_i^*) + \varepsilon_i^*;$$

$$Z_i(\theta, \eta_i^*) = F_i(\theta, \eta_i^*) \Delta_{\eta_i}(\theta, \eta_i^*); \text{ where}$$

- $F_i(\theta, \eta_i^*)$ is the matrix of derivatives of $f(x_i; \beta_i)$ with respect to β_i evaluated at $\eta_i = \eta_i^*$.
- $\Delta_{\eta_i}(\theta, \eta_i^*)$ is the matrix of derivatives of $d(a_i, \theta, \eta_i)$ with respect to η_i evaluated at $\eta_i = \eta_i^*$.

$$\varepsilon_i^* = \sigma I \zeta_i$$

- Again, the approximation above has the form of a LME model.
- The marginal distribution of Y is approximated by a Normal distribution with expectation and variance:

$$E(Y_i) \approx f(x_i; d(a_i, \theta, \eta_i^*)) - Z_i(\theta, \eta_i^*) \eta_i^*,$$

$$\text{Cov}(Y_i) \approx Z_i(\theta, \eta_i^*) \Psi Z_i(\theta, \eta_i^*)^T + \sigma^2 I.$$

The FOCE-LB Method (Lindstrom and Bates, 1982)

- Suggested for additive or log-additive errors (ε_{ij} 's), therefore applicable for Normally or LogNormally distributed data.
- First order Taylor series approximation for the non-linear model at $\eta=\eta^*$.
- Iterative procedure:
 1. Conditional step: for each subject, optimal η_i values are found denoted by η^* , corresponding to the current θ, σ^2, Ψ estimates. This is done via ML methodology in Phoenix.
 2. Substitute η^* in the linear approximation and compute an approximation of the marginal log likelihood function.
 3. Solve the marginal log-likelihood function in (2) to obtain new estimates of θ, σ^2, Ψ . Return to step 1 until the difference between starting and final optimal objective function values are less than a tolerance value (0.001 in Phoenix).

Other methods

- FOCE-ELS
 - Extended Least Squares
 - Normally distributed data
 - Minimizes an extended least squares objective function:
when $\varepsilon_{ij} \sim N(0, \sigma^2 \Lambda)$ as opposed to $\varepsilon_{ij} \sim N(0, \sigma^2 I)$ – this case not seen here.
 - Less computationally intensive than FOCE
- Laplacian
 - For Normal data and data through user supplied likelihoods such as count, categorical and time to event data.
 - More computationally complex than FOCE-ELS.
 - Based on approximating the marginal log-likelihood with a Laplacian approximation.
 - It is iterative.
- In terms of accuracy and computer time:
$$\text{Laplacian} > \text{FOCE-ELS} > \text{FOCE-LB} > \text{FO}$$

Other methods

- IT2S-EM
 - Iterative two-stage & Expectation – Maximization methods.
 - All data types, with user-specific log-likelihood functions for non-Normal data.
 - Not properly a full EM estimation procedure as it incorporates the two-stage maximization.
- QRPEM
 - Quasi-Random Parametric EM
 - Member of a general class of NLME EM estimation procedures.
 - No numerical procedure is required to optimize the likelihood (or its approximation).
 - Relies on numerical integration (for the Expectation step).

- Under very specific conditions, one algorithm may be better than other, but in general there is little difference between algorithms and software packages.
- Since differences may still be present in some cases, it is recommended that they are evaluated on a case by case basis.

4.6 Assumptions in the Modeling Process

Specific Learning Objectives:

1. Explain the 10 main assumptions involved in the NLME modeling process.
2. Graphically assess the distributional assumptions in the NLME model in Phoenix.

10 General Set of Assumptions

(See Bonate for more)

1. *Parameters are estimable and unique*

- Determined during model development since non-identifiable models usually fail to converge successfully or to produce reasonable parameter estimates.

2. *Excluded data did not affect the analysis.*

- Some reasons:
 - ✓ Missing data (intermittent or complete loss to follow-up, missing covariate values).
 - ✓ Outliers.
- Considered safe if there is less than 5% of data excluded.
- Suggestion: run two analyses, with and without the data.
Differences between results should be discussed.

Testing assumptions.

3. *Structural model is appropriate.*

- For descriptive models: Does the model characterize adequately the data?
 - ✓ Checked via residual plots, histograms of the distribution of predicted random effects.
- For predictive models: Is the model useful for predictive purposes?
 - ✓ Checked via simulation and predictive assessments.

4. *Significant interactions between sub-model covariates are included.*

- By plotting residuals vs. the product of covariates from a model without interactions, it can be concluded that the interaction is:
 - ✓ Not present, if the residuals are distributed around zero.
 - ✓ Present, if there is a positive/negative trend.

Testing assumptions.

5. *The distribution of the random effects (η 's) is adequately modeled.*
 - Recall that the distribution of the η 's is unknown but an approximate estimate can be made from the EBs (predicted η 's):
 - ✓ Rich data: EBs are in close approximation to the true values given a correct structural and variance model.
 - ✓ Sparse data: EBs are more dependent on the model assumptions than the data.
 - E.g. if the distribution of the η 's was modeled as Normal but a plot of the EBs distribution shows right skewness, they should be modeled as log-Normal.

Testing assumptions.

6. *Inclusion of a correlation term between η 's is appropriate.*

- Adding a covariance is done by examination of the distribution of
 - ✓ The EBEs for PK parameters : e.g. predicted CL vs. V.
 - ✓ The EBEs for η 's: e.g., predicted η_{CL} vs. η_V .
- This examination is done through scatter plots and/or a correlation analysis (regression, statistical correlation tests)
- It has been shown that including a non-real covariance term is generally more forgiving than not including a real one.
- Post-hoc modeling testing via LRT tests is recommended.

Testing assumptions.

7. *The residual variance model is adequate.*

- Standard residuals assumptions for maximum likelihood estimation:
 - ✓ Independence
 - ✓ Normally distributed
 - ✓ Mean zero
 - ✓ Constant variance
- These assumptions are verified through the usual tools, such as histograms, QQ Normal plots, residuals vs. covariates.

Residual analysis will be discussed in more detail shortly.

Testing assumptions.

8. *The estimation method is adequate.*

- Approximations for the likelihood integral used in software packages are adequate.
- They perform well under most circumstances in pharmacokinetics and any bias introduced is of acceptable magnitude.
- It is suggested to fit the final model with a more accurate estimation routine.

E.g., Fit the final model with the Laplacian method after developing the model with FOCE. Parameters should not be substantially different if the FOCE approximation was working properly.

Note that using Goodness of Fit measures (LRT, AIC) between estimation procedures is not adequate.

Testing assumptions.

9. *The global minimum is found.*

- Usually true if the model development process:
 - ✓ Involved many different starting values
 - ✓ Was carried out from simpler to more complex models.

Testing assumptions.

10. *The model can be used for simulating real-world data.*

- Usually true if the same conditions are employed for simulation:
 - ✓ The final model specification.
 - ✓ Inputs such as dosing history, subject characteristics, etc.
 - ✓ Features of the experimental design (randomization scheme, sampling design, etc.)
- To assess this assumption:
 - Plots of concentration-time simulated data should be indistinguishable from the real data.
 - Particular attention should be paid to the variance of the simulated vs. real data.

Conditional Weighted Residuals

- In linear regression, a weighted residual is one that is standardized by the square root of the Mean Squared Error, which is the estimate of $\text{Var}(Y)=\sigma^2$.
- They are a result of extending the standard assumption of constant variance where the estimation method becomes a weighted version of OLS: Weighted Least Squares or WLS.
- In general, weighted residuals have the following form:

$$\frac{Y_i - \hat{Y}_i}{\sqrt{\text{Var}(Y_i)}} \sim N(0,1)$$

is estimated by plugging in an estimate of the variance

$$\frac{Y_i - \hat{Y}_i}{\sqrt{\hat{\sigma}^2}} \stackrel{\text{approx}}{\sim} N(0,1).$$

- In NLME, before conditional estimation methods were introduced, weighted residuals obtained from the FO method were used (WRES), based on:

$$\frac{Y_i - \hat{Y}_i^{(FO)}}{\sqrt{\text{Var}^{(FO)}(Y_i)}}$$

Conditional Weighted Residuals

- Conditional Weighted Residuals (CWRES) have the same form as WRES, only that the predicted Y and its variance are obtained through the FOCE method:

$$\frac{Y_i - \hat{Y}_i^{(FOCE)}}{\sqrt{Var^{(FOCE)}(Y_i)}}$$

- It has been shown that CWRES is a better diagnostic measure than WRES⁽¹⁾.
- As before, no systematic trends should appear in scatter plots of
 - CWRES vs. time,
 - CWRES vs. predicted population PK parameters (PRED).
- Histograms of CWRES should be consistent with a N(0,1): symmetry around zero and about 95% values concentrated within the range -/+2.

Hooker, A.C., Staatz, C.E. and Mats, O.K. Conditional Weighted Residuals (CWRES): A Model Diagnostic for the FOCE Method. *Pharmaceutical Research* (2007) Vol. 24, No. 12.

Conditional Weighted Residuals

- CWRES plots can be useful in assessing whether the specified residual variance model is appropriate (additive, proportional, exponential, etc).
 - E.g., a funnel-like pattern in the CWRES vs. PRED plot indicates that a better residual variance model may be multiplicative instead of additive.
- Recall the Phenobarbital Data where the PK parameters logCL and logV were modeled with Weight as a covariate.

We will examine the fit and residual plots CWRES vs. PRED in Phoenix, for the following two models:

Additive residuals model:

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

Proportional residuals model:

$$C_{ij} = \left(\sum_{d:t_{id} < t_{ij}} \frac{D_{id}}{\exp(\ln \theta_3 + \theta_4 Wt_i^*) e^{\eta_{v_i}}} \exp \left\{ -\frac{\exp(\ln \theta_1 + \theta_2 Wt_i^*) e^{\eta_{CL_i}}}{\exp(\ln \theta_3 + \theta_4 Wt_i^*) e^{\eta_{v_i}}} (t_{ij} - t_{id}) \right\} \right) (1 + \varepsilon_{ij}),$$

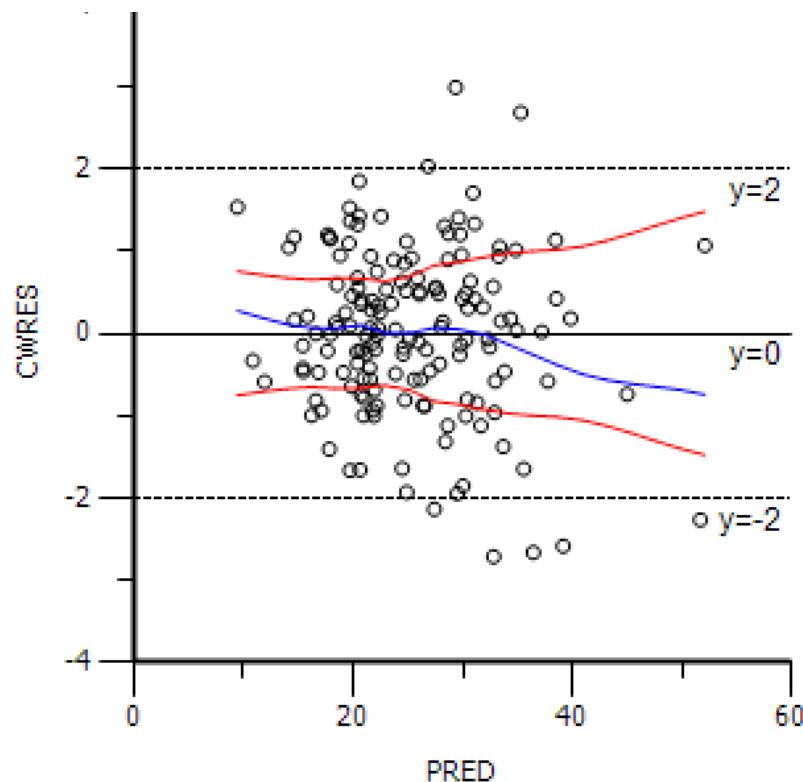
$$Wt_i^* = Wt_i - \text{Median}(Wt_i); \quad \begin{bmatrix} \eta_{CL_i} \\ \eta_{v_i} \end{bmatrix} \sim BVN(0, \Psi), \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2), \quad \Psi = \begin{bmatrix} \psi_{CL}^2 & \psi_{CL,v} \\ \psi_{CL,v} & \psi_v^2 \end{bmatrix}.$$

Conditional Weighted Residuals

Additive

Residual Error:
 C CObs CEps = Additive BQL?

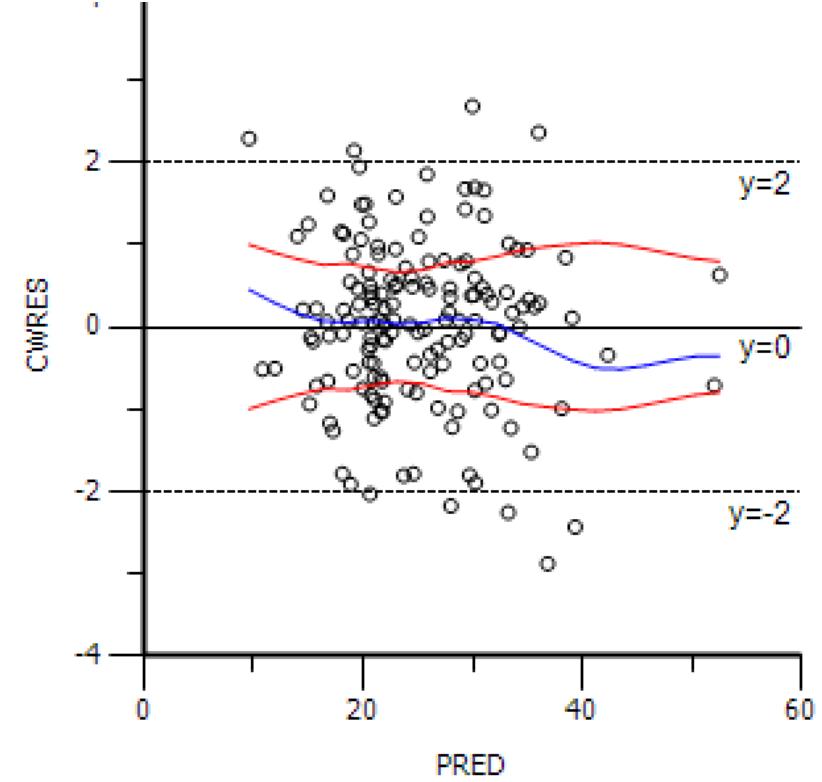
Parameter	Estimate	Units	Stderr
tvV	0.205613		0.025790308
tvCl	-5.13		0.051301072
dVdwgt	0.528646		0.038660051
dCldwgt	0.625466		0.071340732
stdev0	2.86519		0.25443074



Proportional

Residual Error:
 C CObs CEps = Multiplicat BQL?

Parameter	Estimate	Units	Stderr
tvV	0.202608		0.024062003
tvCl	-5.12622		0.057447367
dVdwgt	0.560291		0.04018333
dCldwgt	0.659877		0.083097533
stdev0	0.113446		0.011271646



Omega Output from Model Comparer

Additive

	Label	nV	nCl
	Omega		
nV	0.017282294		
nCl	0.027029676	0.042294607	
	Correlation		
nV		1	
nCl	0.99976435		1
Shrinkage	0.11409427	0.11414186	

Proportional

Omega		
nV	0.020118023	
nCl	0.024995363	0.052113588
Correlation		
nV		1
nCl	0.7719538	1
Shrinkage	0.12790435	0.18803049

$$\text{BSV in V} = \sqrt{0.01728} \times 100\% = 13.14\%$$

$$\text{BSV in CL} = \sqrt{0.02703} \times 100\% = 16.44\%$$

$$\text{BSV in V} = \sqrt{0.02012} \times 100\% = 14.18\%$$

$$\text{BSV in CL} = \sqrt{0.02499} \times 100\% = 15.81\%$$

Shrinkage indicates how much EBEs shrink towards population mean.
The lower the better.

Overall Output from Model Comparer.

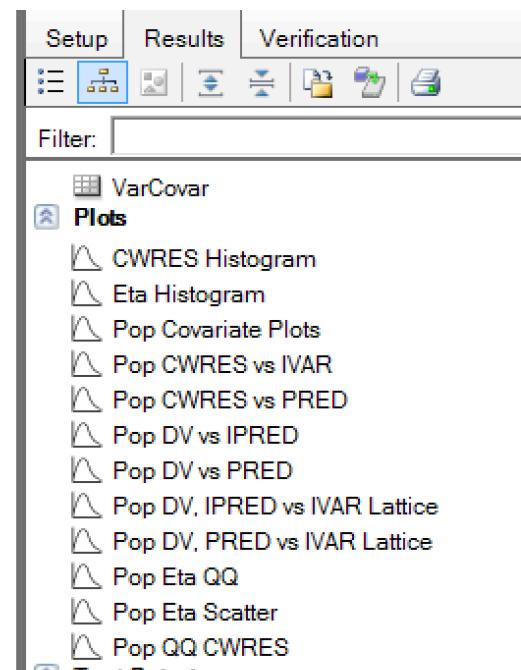
First row: additive

Second row: proportional

	RetCode	LogLik	-2LL	AIC	BIC	nParm	nObs
1	1	-433.30893	866.61786	882.61786	906.96526	8	155
2	1	-432.71151	865.42302	881.42302	905.77042	8	155

Graphical Model Assessment Phoenix Implementation Phenobarbital Data Example

Before running any model, Phoenix allows to select the required plots in the Plots tab:



Main residuals and
EBE plots

Residuals Plots, Additive vs. Proportional Residuals Models

The Phoenix Model Comparer allows to compare the residual plots of two or more models.
(Insert -> Phoenix Modeling -> Model Comparer)

To check for:

- Constant variance
- Distributed around zero
- Outliers

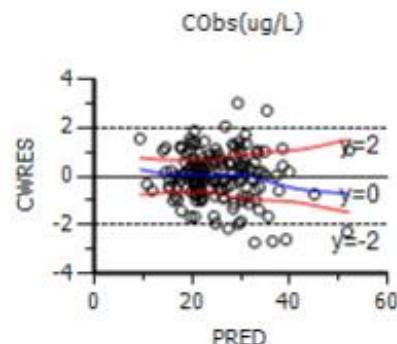
Both models have roughly same amount of outliers, smooth lines show slightly better trend in latter model, with respect to constant variance assumption.

To check for normality:

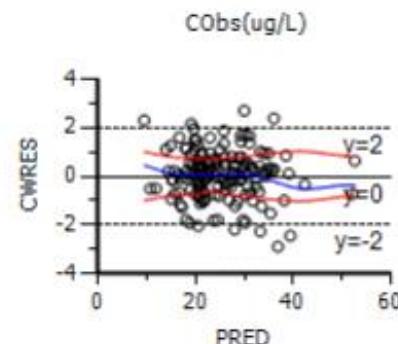
- Aligned,
- No heavy tails.
- Not skewed.
- Outliers.

Additive model shows more outliers.

Additive Residuals

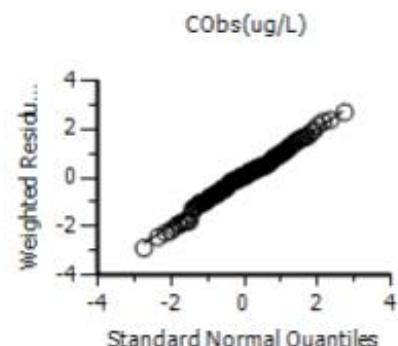
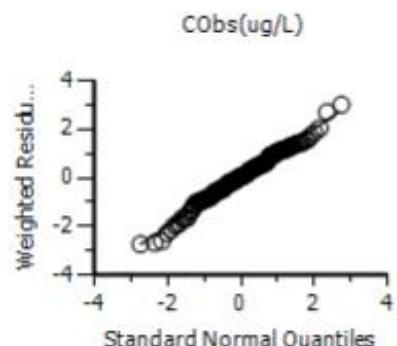


Proportional Residuals



Pop CWRES vs PRED

Pop QQ CWRES

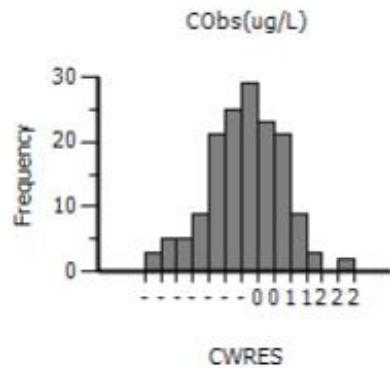


Residuals Plots

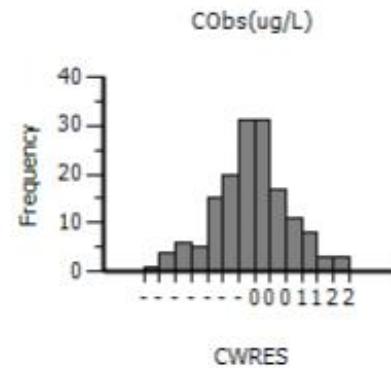
Same as QQ plots:
Symmetry
No heavy tails
Not skewed.
Proportional residual model slightly more symmetric.

CWRES Histogram

Additive Residuals



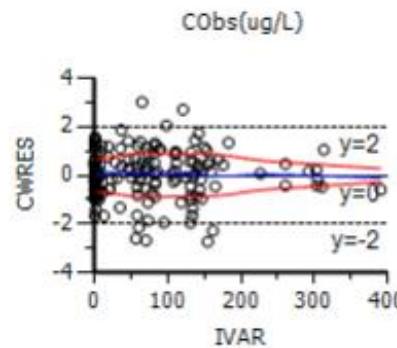
Proportional Residuals



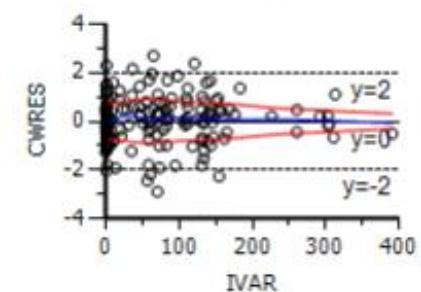
No trends= Uncorrelated (i.e., independence of residuals). Both models are OK.

Pop CWRES vs IVAR

IVAR = Time



CObs(ug/L)

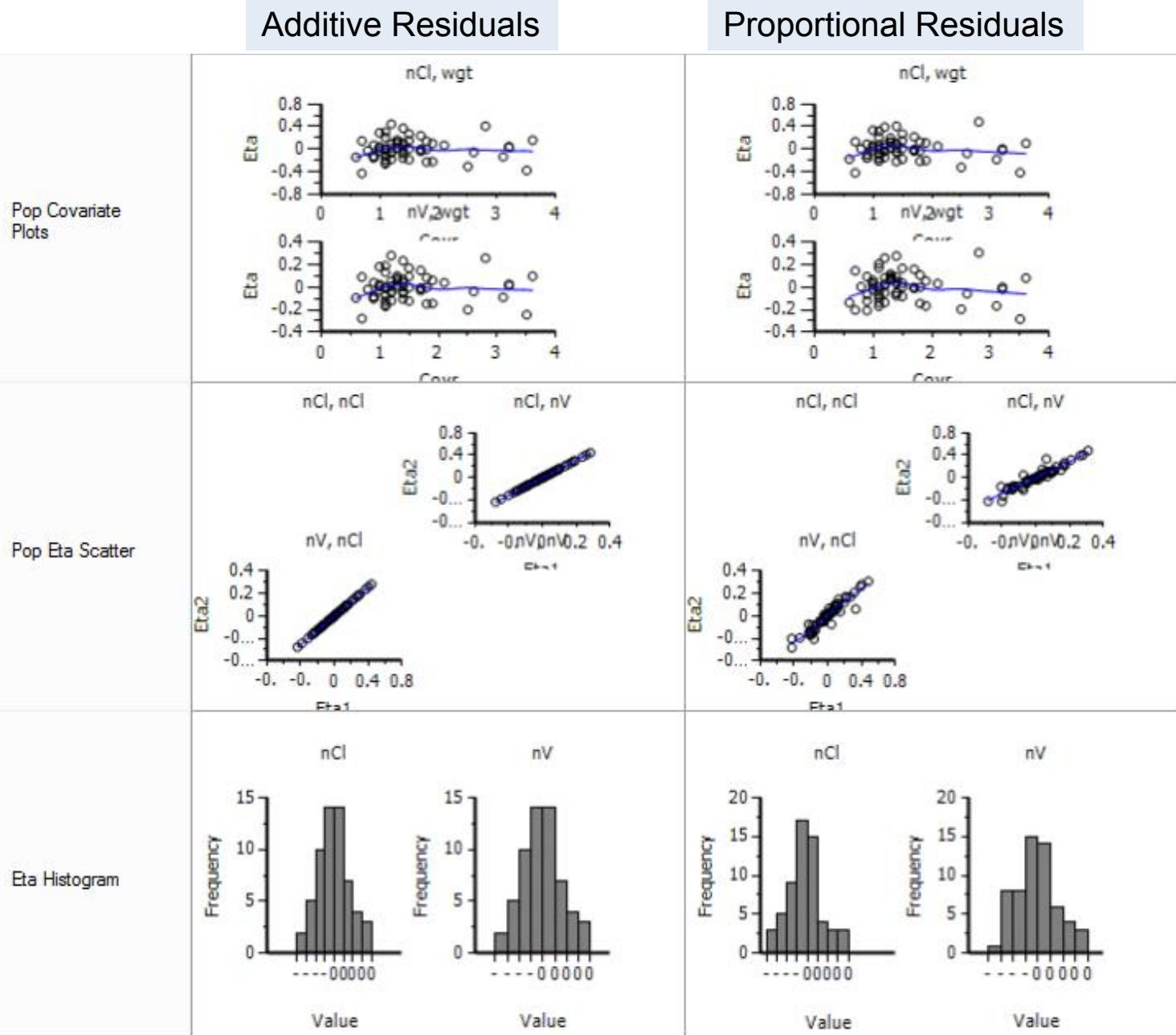


Predicted Random Effects Plots

EBe vs. Weight:
No apparent trend indicates that the relationship between Wt and PK parameters ($\log CL, \log V$) is being accounted for in the models.

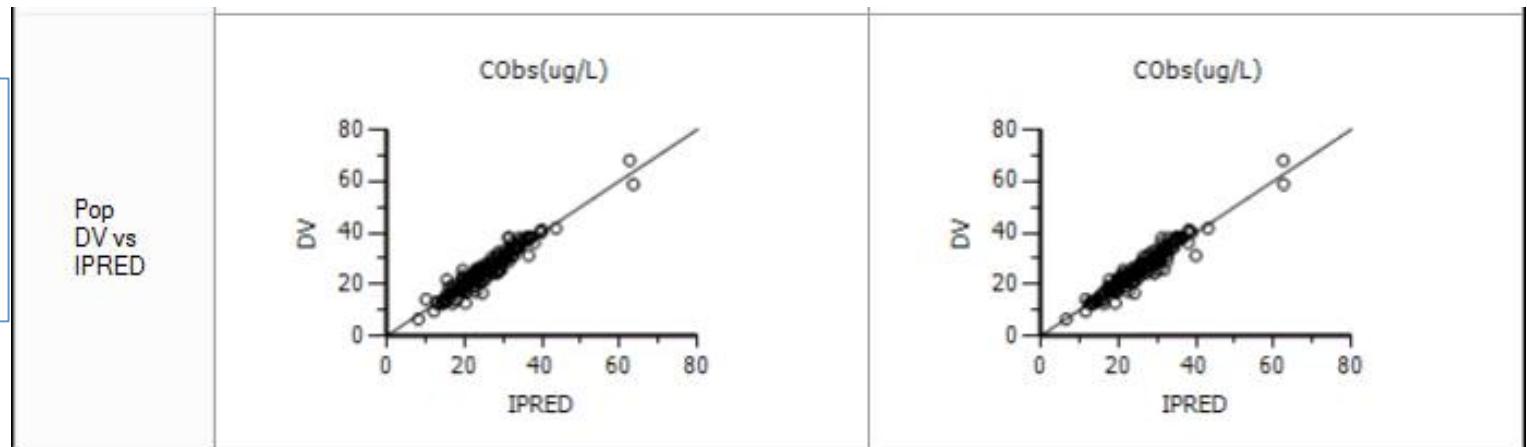
Scatterplot predicted η_{CL} vs. η_V indicate additive residual model has perfect correlation (est corr=1) vs. proportional model has more dispersion around the line (est corr=0.77)

Normality of EBe:
proportional model appears to have less symmetry.



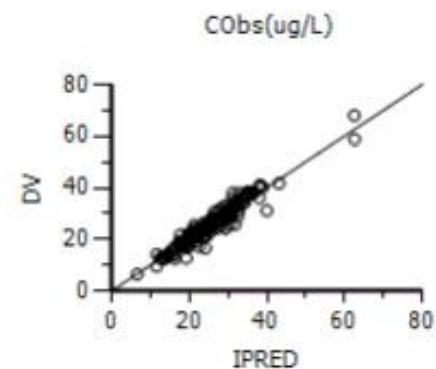
Observed vs. Predicted

Additive Residuals

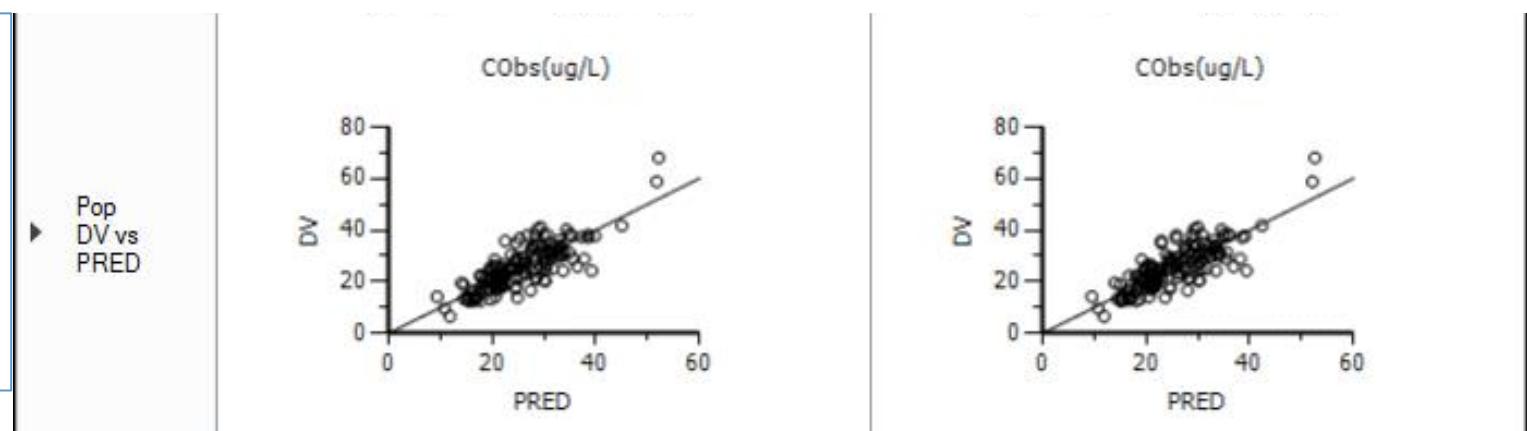


Individual prediction
(including the EBEs).
Ideally should fall
close to the line of
unity.

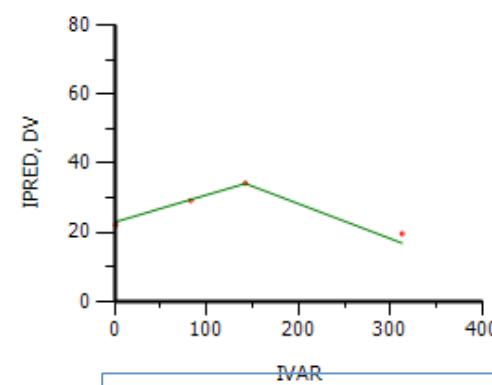
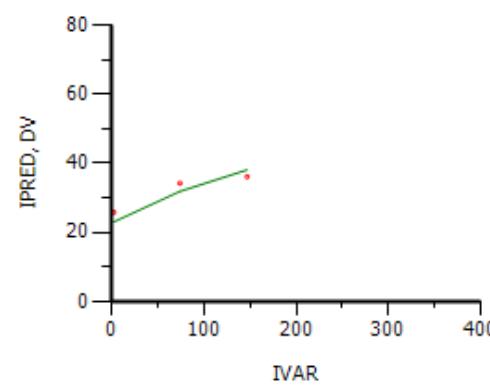
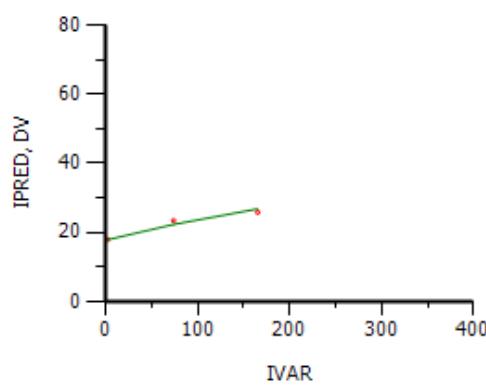
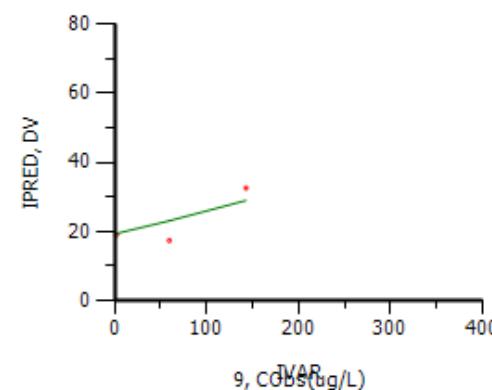
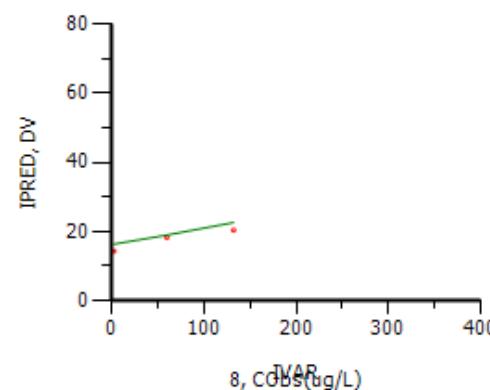
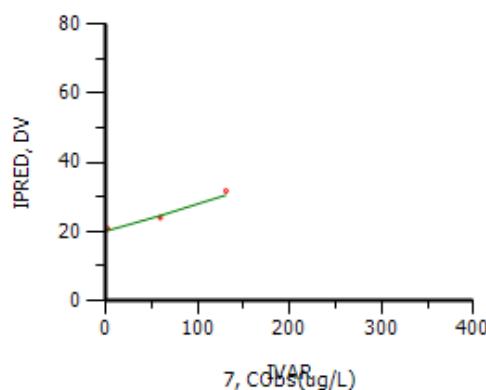
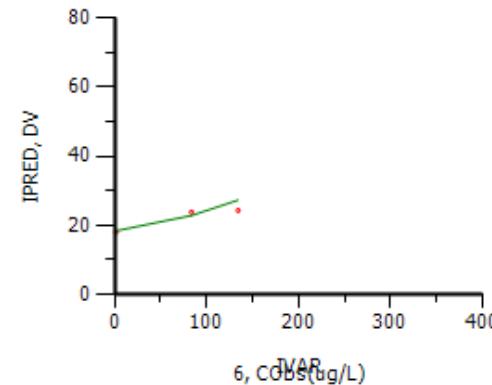
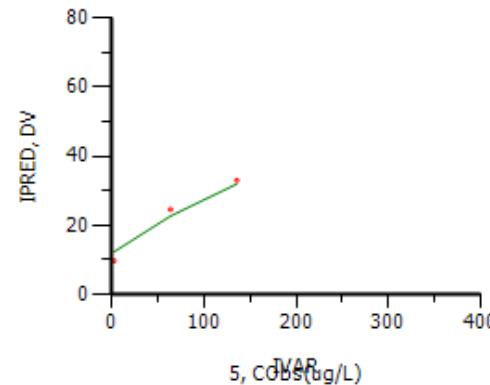
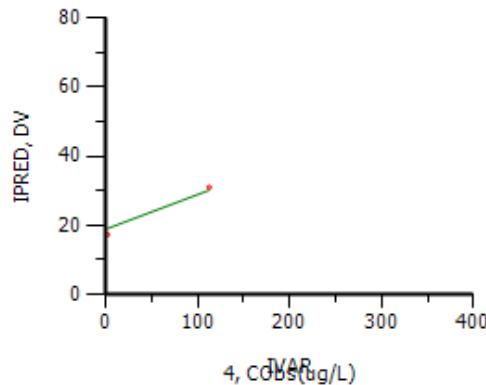
Proportional Residuals



Population
predictions (setting
 η 's to zero). Since
these are typically
less accurate, larger
deviations around
the unity line are
expected.



Pop DV, IPRED vs IVAR Lattice: INDIVIDUAL PREDICTIONS

1, C₀obs(ug/L)2, C₀obs(ug/L)3, C₀obs(ug/L)

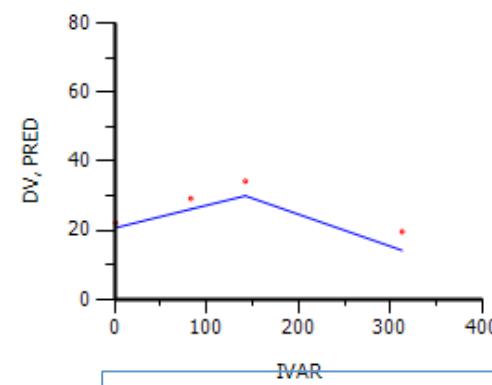
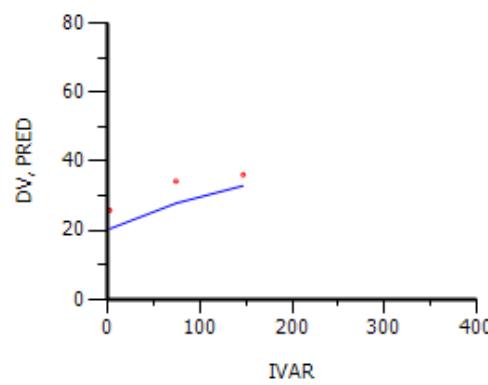
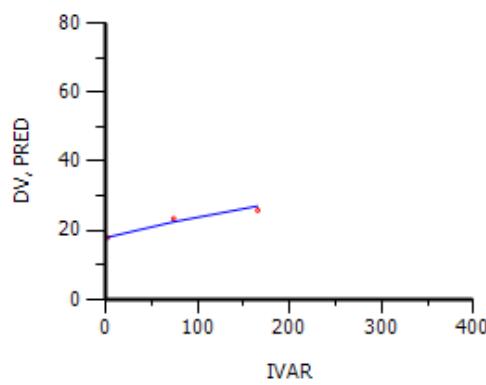
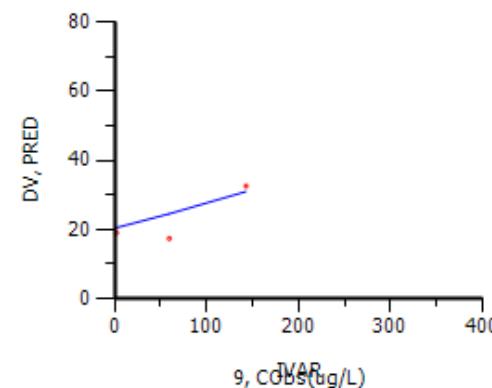
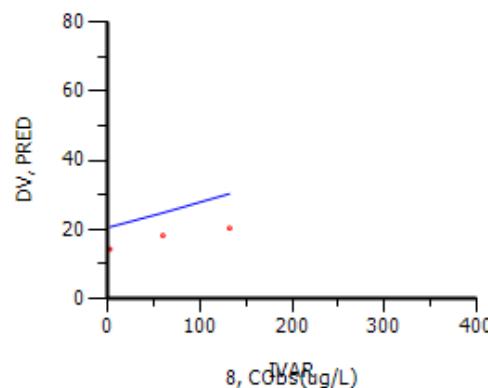
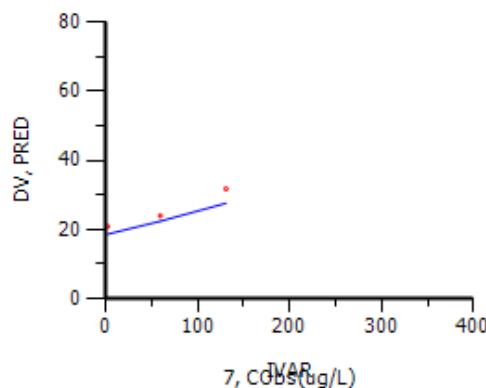
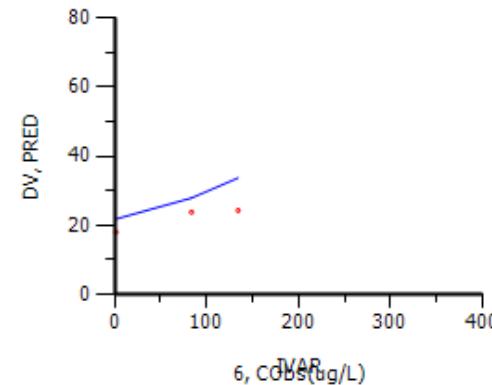
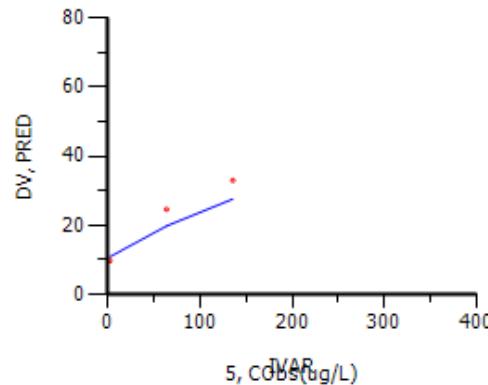
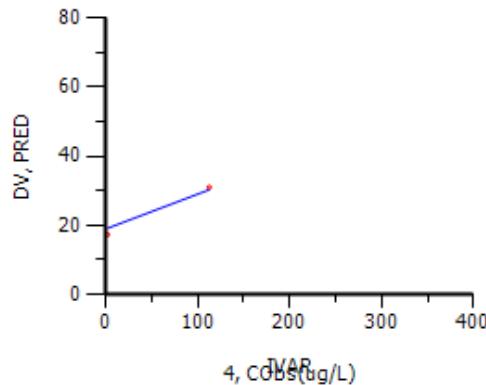
Additive Residuals Model

Pop DV, PRED vs IVAR Lattice: POPULATION PREDICTIONS

1, C0bs(ug/L)

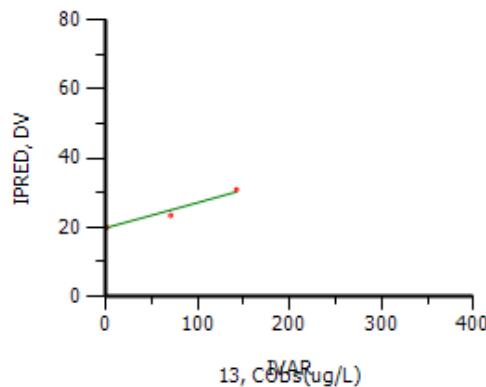
2, C0bs(ug/L)

3, C0bs(ug/L)

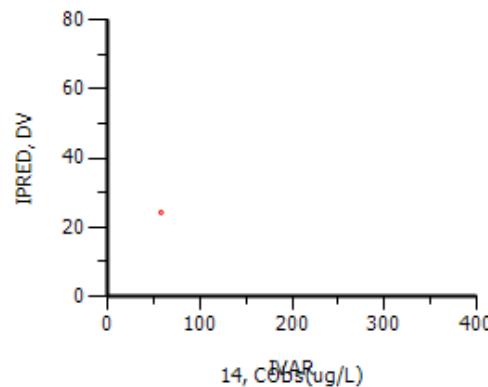


Additive Residuals Model

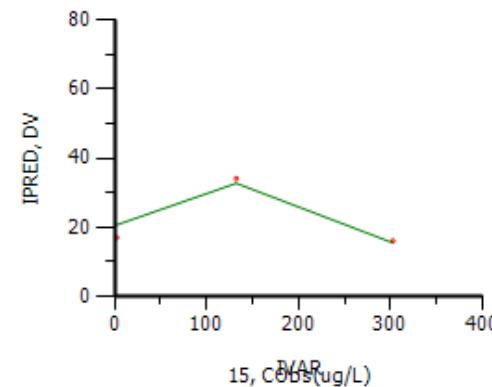
10, CObs(ug/L)



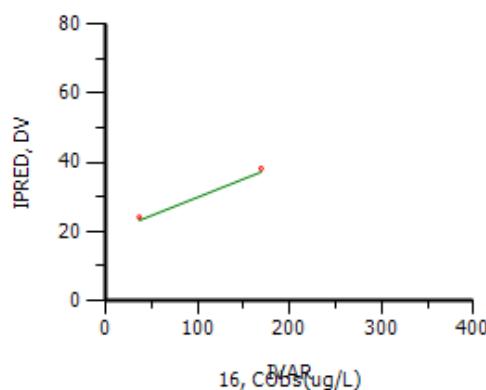
11, CObs(ug/L)



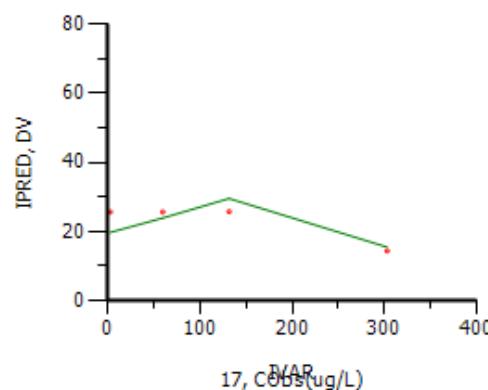
12, CObs(ug/L)



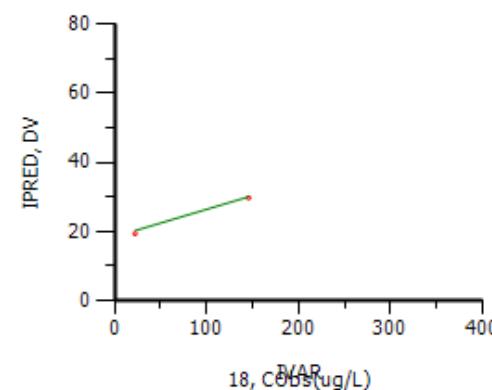
13, CObs(ug/L)



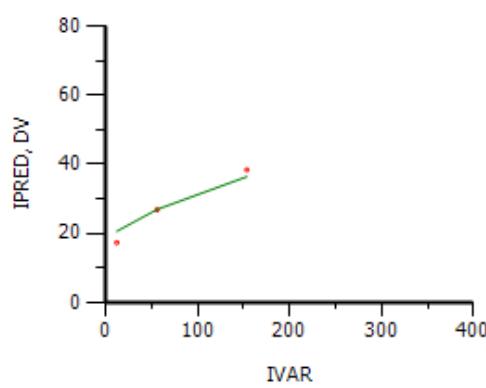
14, CObs(ug/L)



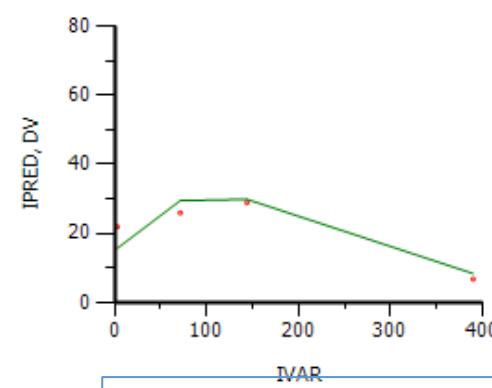
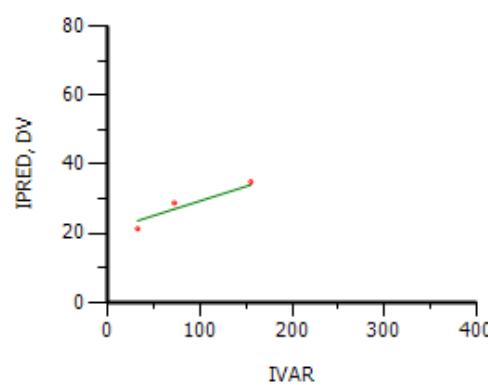
15, CObs(ug/L)



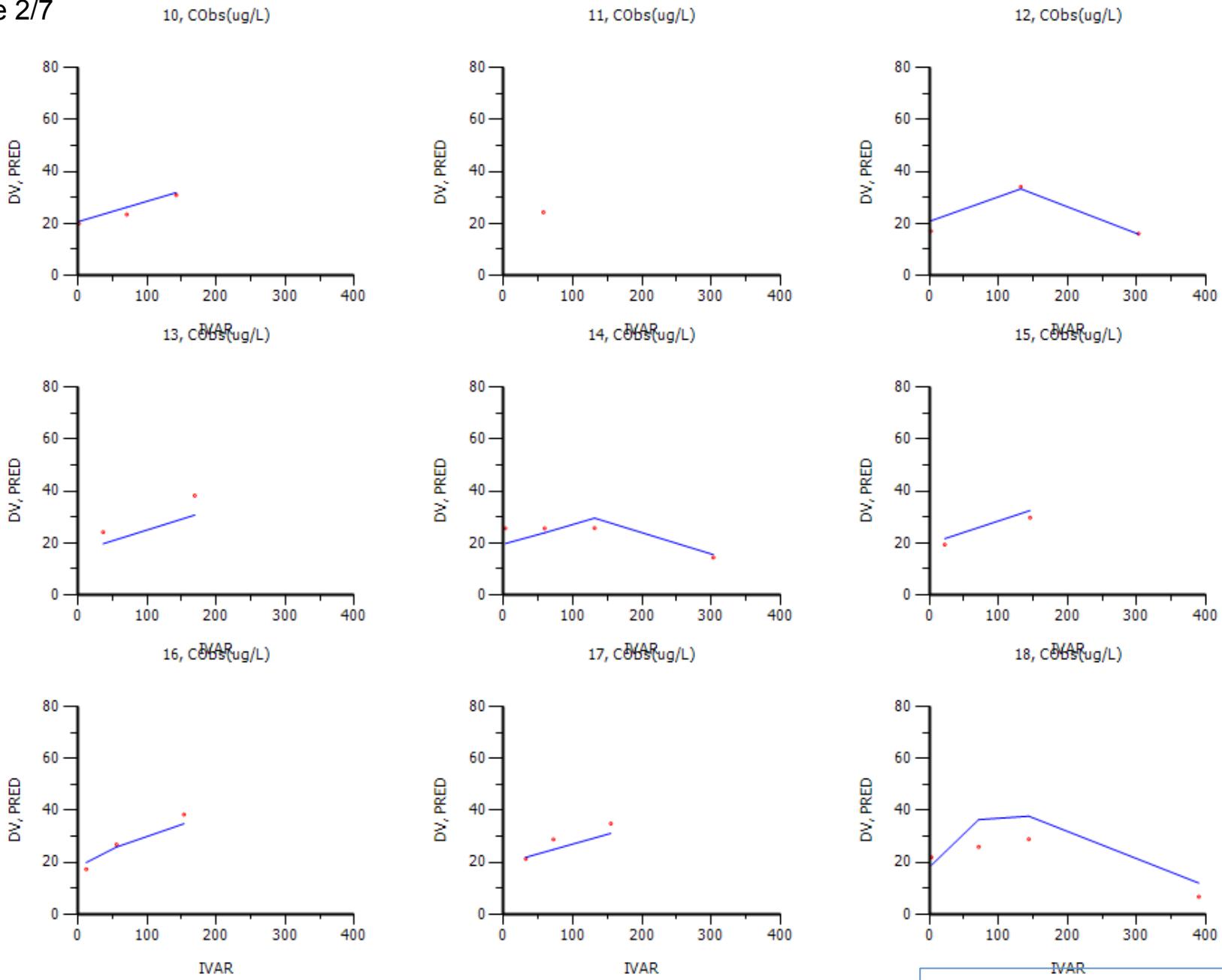
16, CObs(ug/L)



17, CObs(ug/L)



Additive Residuals Model



Additive Residuals Model

4.7 Precision of Estimates and Confidence Intervals

Specific learning objectives:

1. Implement Bootstrap to obtain NLME estimates in Phoenix.
2. Explain how the Bootstrap estimates are obtained.
3. Reproduce Bootstrap SE's and CI's obtained in Phoenix through manual calculation.
4. Perform graphical assessment of Bootstrap results in Phoenix.

Precision of the Parameter Estimates $\hat{\theta}, \hat{\Psi}, \hat{\sigma}_{\varepsilon}^2$

- Recall that in the preliminaries section we talked about the sample mean \bar{Y} :
 - As an estimate of the population mean μ .
 - Having a sampling distribution.
 - And that as a result, we can obtain $SE (= s / \sqrt{n})$.
- Small SE 's are generally indicative of good parameter estimation.
- CI's and SE 's explain the uncertainty in the parameter estimates.
- In case of using a regression model with parameters β, σ^2 , the above is extended to the sampling distribution of the estimates of β and σ^2 and their SE .
- In our NLMEs, we want to estimate $SE(\hat{\theta}), SE(\hat{\Psi}), SE(\hat{\sigma}_{\varepsilon}^2)$.
 - However, sometimes covariance estimation is not successful because of model complexity and also likelihood methods may not always be useful.

Estimation of the SE's of the Model Parameters

- Usually done via standard maximum likelihood theory, assuming:
 - Sample size is large
 - Random effects and within-subject errors are Normally distributed
 - Then, the estimates of the θ 's are asymptotically normally distributed.
 - The asymptotic ML based $(1-\alpha)100\%$ CI for θ is given by

$$\hat{\theta} \pm z_{1-\alpha/2} SE(\hat{\theta}),$$

where

$\hat{\theta}$ is the final parameter estimate,

$z_{1-\alpha/2}$ is the $\alpha/2$ -th percentile under $N(0,1)$
(e.g., for $\alpha = 0.05$, $z_{1-\alpha/2} = 1.96$),

$SE(\hat{\theta})$ is the standard error of $\hat{\theta}$.

Precision of the Estimates and Confidence Intervals

- Drawbacks of the maximum likelihood based SE's:
 - The assumption that the CI's are symmetric is not necessarily true, especially for variance estimates of $(\hat{\Psi}, \hat{\sigma}_\varepsilon^2)$.
 - Asymptotic theory applies better to $\hat{Var}(\hat{\theta})$.
 - Samples are not always large.
- Alternatives for estimation of SE's are
 - Nonparametric Bootstrap.
 - Jackknife.
 - Log-likelihood profiling.

Non-parametric Bootstrap

- Is a re-sampling technique, that consists of randomly selecting with replacement the subjects in the original sample, to conform many replicated samples.
 - As with the original sample, since the units must be selected randomly and without replacement:
 - The units (subjects) within samples are independent.
 - Samples are independent of one another.

Non-parametric Bootstrap



Briefly, sampling without replacement:

- Imagine a bag with 5 red balls and 5 blue balls.
 - Take one ball at random, note the color and set it aside.
 - Now the bag has one less ball. Take a second ball, note color and set aside.
 - Repeat until all balls are selected.
 - Note that the probability of selecting, e.g., one blue ball in any draw depends on whether we've selected a blue or red ball in previous draws and the number of draws.
 - This means that the outcome in one step is not independent of the outcome of the preceding step.

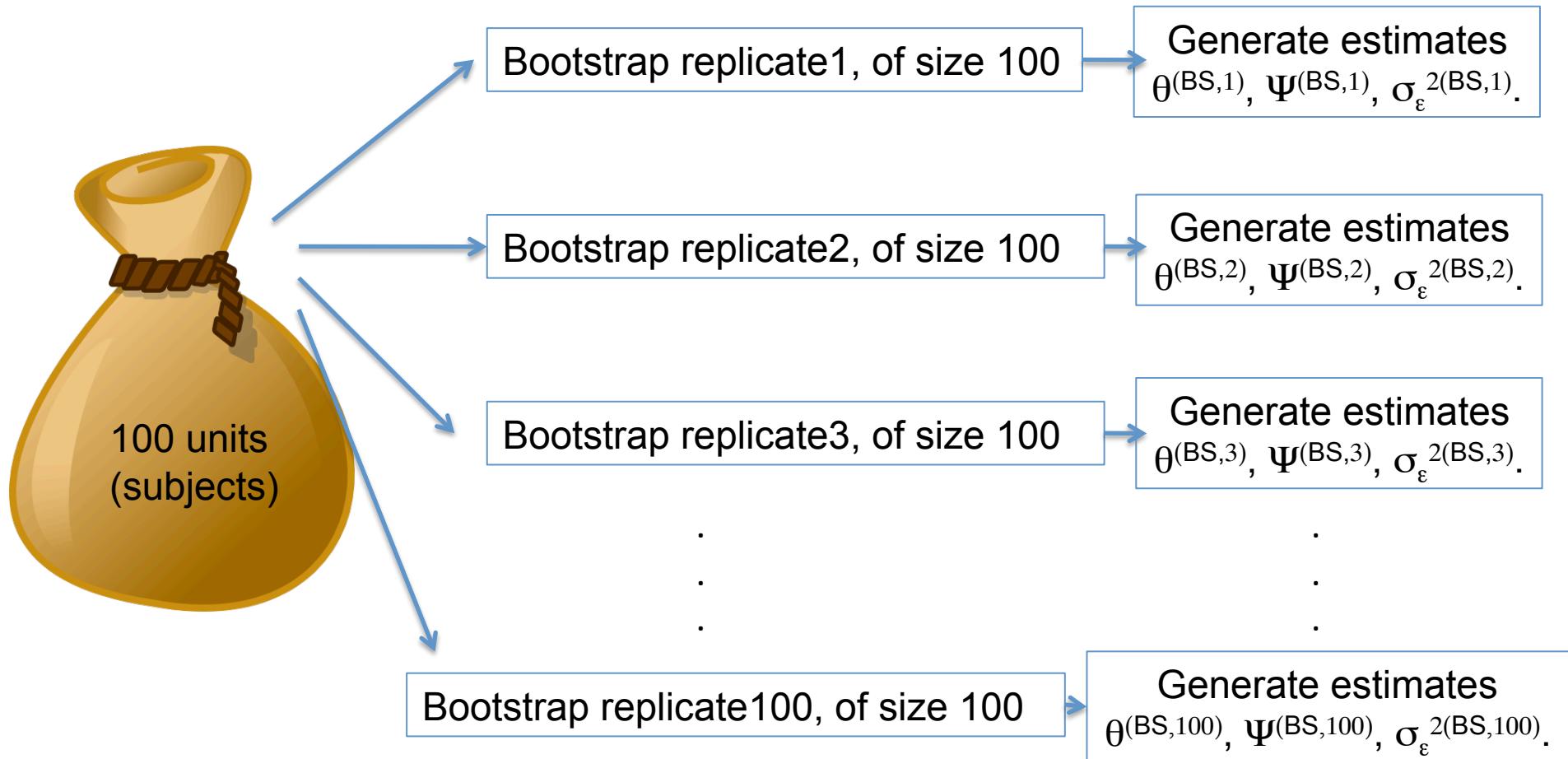
Non-parametric Bootstrap



Briefly, sampling with replacement:

- Again, imagine a bag with 5 red balls and 5 blue balls.
 - Take one ball at random, note the color and put it back.
 - Now the bag has the same number of balls. Take a second ball, note color and put it back.
 - Repeat as many times as you want (100,1000).
 - Note that the probability of selecting, e.g., one blue ball no longer depends on the outcomes from preceding draws. This means that the selection in one step is independent of the preceding step.
 - This leads independent random samples, given the original sample.

Back to Non-Parametric Bootstrap...



Bootstrap Standard Errors and 95% CI's

Let $\{\theta^{(BS,1)}, \theta^{(BS,2)}, \dots, \theta^{(BS,m)}\}$ be the set of estimates of θ from m bootstrap replicates. Then the bootstrap SE of $\hat{\theta}$ is given by:

$$SE(\hat{\theta})^{(BS)} = \sqrt{\frac{\sum_{i=1}^m (\theta^{BS,i})^2 - \bar{\theta}^{(BS)}}{m-1}}$$

Simply the standard deviation of the bootstrap estimates.

A 95% bootstrap CI for θ is given by the 2.5th and 97.5th percentiles of the distribution of $\{\theta^{(BS,1)}, \theta^{(BS,2)}, \dots, \theta^{(BS,m)}\}$.

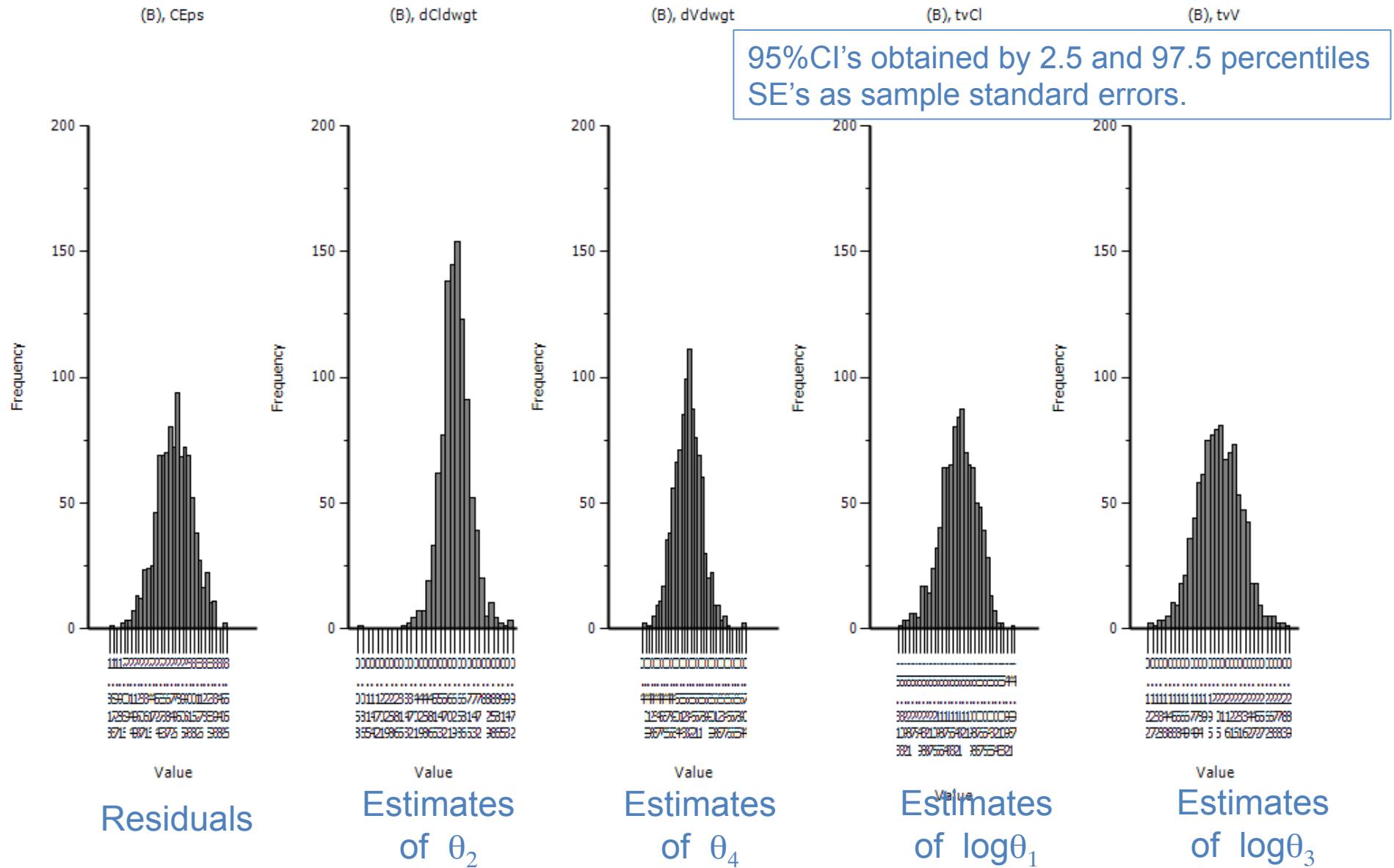
Same procedure applies for the estimates of Ψ and σ_ϵ^2 .

Bootstrap for SE's and CI's
Phoenix Implementation
Phenobarbital Data Example
Additive Residuals Model

Bootstrap implementation

Boot Theta Histogram

Distribution of Residuals and θ Estimates over 1000 replicates



Theta and sigma estimates
100 and 1000 Bootstrap sample replicates

Additional output

- BootSubj(B).csv
- Rawout.csv

Output Data

- BootOmega
- BootOmegaStacke
- BootOmegaStderr
- BootOverall
- BootSecondary
- BootTheta**
- BootThetaStacked
- BootVarCovar
- ConvergenceData
- Doses
- Eta
- EtaCov
- EtaCovariate
- EtaCovariateCat
- EtaEta
- EtaStacked

100 replicates

Parameter	Mean	Stderr	CV%	Median	2.5%	97.5%
1 tvV	0.20361726	0.022630417	11.114194	0.20560448	0.15812688	0.24017658
2 tvCl	-5.1325671	0.047987267	-0.93495646	-5.1315082	-5.2125244	-5.03555862
3 dVdwgt	0.53512312	0.046168148	8.627575	0.53153091	0.44702768	0.62387523
4 dCldwgt	0.6357446	0.090607201	14.252138	0.63183992	0.48292192	0.8596789
5 stdev0	2.7669529	0.31104556	11.241448	2.7675446	2.1191085	3.3060223

1000 replicates

Scenario	Parameter	Mean	Stderr	CV%	Median	2.5%	97.5%
1 (B)	tvV	0.20510926	0.02557059	12.466814	0.2050598	0.15502417	0.25376573
2 (B)	tvCl	-5.1362868	0.052981086	-1.0315056	-5.1338852	-5.2476882	-5.0460638
3 (B)	dVdwgt	0.53242017	0.040219055	7.5540066	0.53321541	0.45403916	0.61331946
4 (B)	dCldwgt	0.63016721	0.086337321	13.7007	0.62964773	0.4655107	0.79998451
5 (B)	stdev0	2.7746148	0.26134749	9.4192352	2.783022	2.2426234	3.2758979

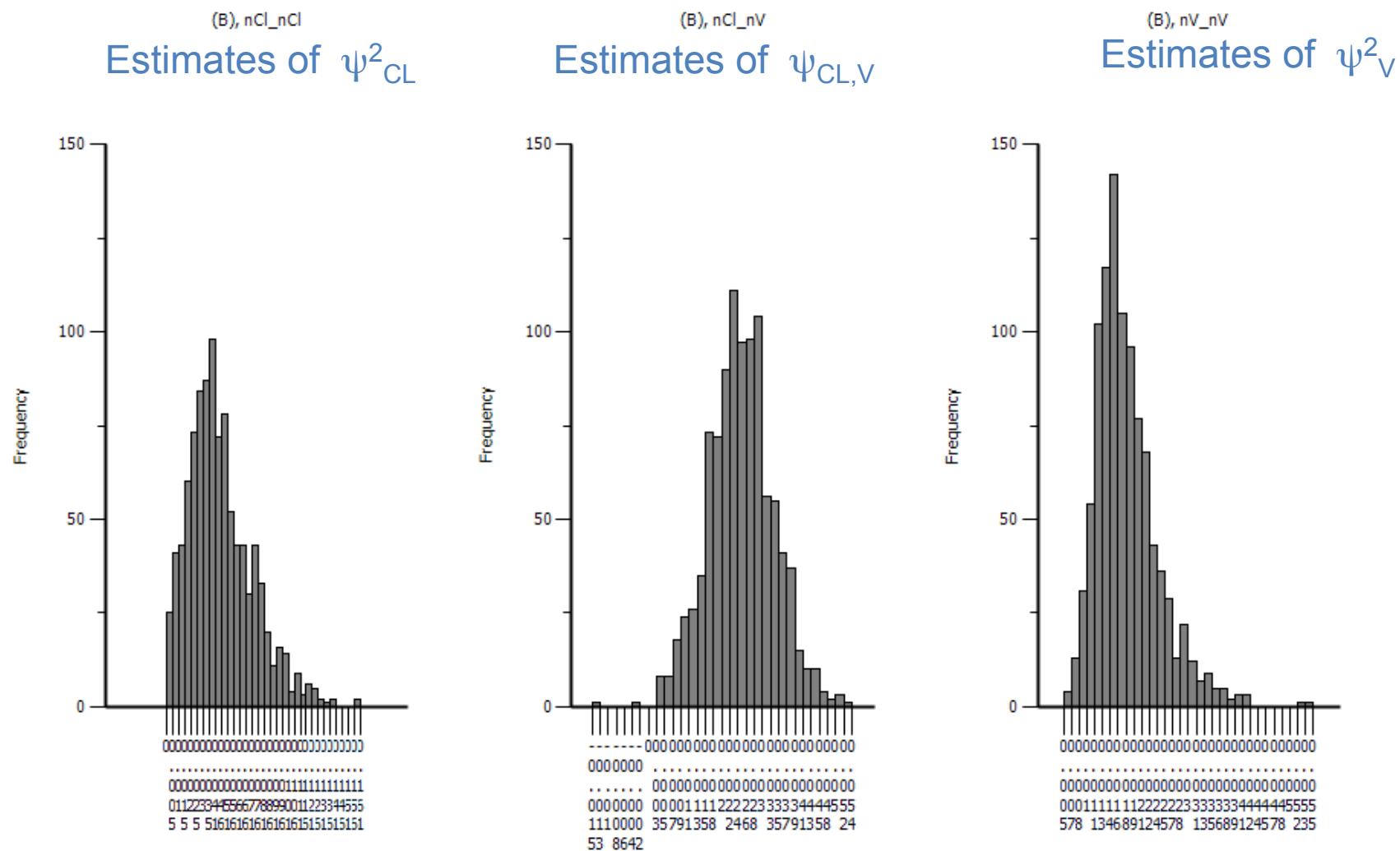
Original sample

Parameter	Estimate	Units	Stderr	CV%	2.5% CI	97.5% CI
1 tvV	0.205613		0.025787594	12.541811	0.15465099	0.25657501
2 tvCl	-5.13008		0.051285476	-0.99970128	-5.2314315	-5.0287285
3 dVdwgt	0.528645		0.038654883	7.3120683	0.45225438	0.60503562
4 dCldwgt	0.625467		0.071350543	11.407563	0.4844625	0.7664715
5 stdev0	2.86519		0.25433836	8.8768409	2.362561	3.367819

Bootstrap implementation

Boot Omega Histogram

Distribution of Estimates of Ψ over 1000 replicates



SE's of each estimate obtained as sample standard deviations

Bootstrap implementation

Sample means from each distribution

100 replicates

1000 replicates

Original sample

Filter:

Additional output

- BootSubj(B).csv
- Rawout.csv

Output Data

- BootOmega
- BootOmegaStacke
- BootOmegaStderr
- BootOverall

	Scenario	Label	nV	nCl
1	(B)	Omega		
2	(B)	nV	0.01816294	
3	(B)	nCl	0.024665443	0.043920062
4	(B)	Correlation		
5	(B)	nV		1
6	(B)	nCl	0.87330185	1

	Scenario	Label	nV	nCl
1	(B)	Omega		
2	(B)	nV	0.017750024	
3	(B)	nCl	0.024812636	0.044776883
4	(B)	Correlation		
5	(B)	nV		1
6	(B)	nCl	0.88012936	1

Label	nV	nCl
Omega		
nV	0.017281474	
nCl	0.02702969	0.042296656
Correlation		
nV		1
nCl	0.99976437	1
Shrinkage	0.1141333	0.11418106

Bootstrap implementation

100 replicates

Precision of ψ 's estimates:
Sample standard deviations
of bootstrap replicate
estimates.

	Scenario	Label	nV	nCI
1	(B)	nV	0.0059859685	
2	(B)	nCI	0.0078295184	0.026220776

1000 replicates

	Scenario	Label	nV	nCI
1	(B)	nV	0.0062373178	
2	(B)	nCI	0.0085854471	0.026370033

Original sample

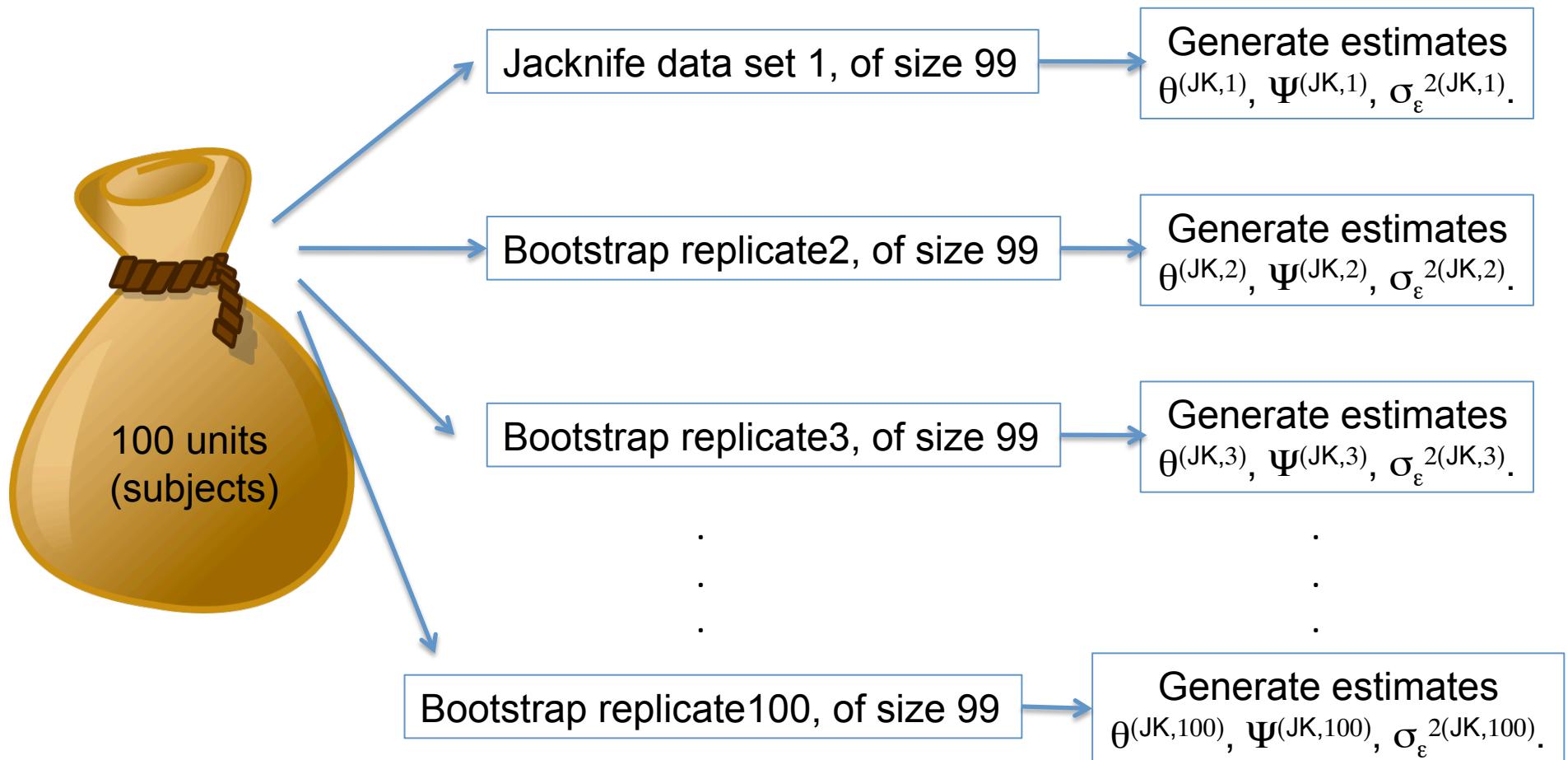
	Scenario	Label	nV	nCI
1		nV	0.0041559911	
2		nCI	0.0068980781	0.011364493

Jackknife

- Similar to bootstrap with respect to the re-sampling idea.
- Two kinds of Jackknife: “delete 1” and “delete 10%”.
- “Delete 1”:
 - Remove one subject at a time from the data set to create m new Jackknife data sets.
 - The model is fit to each data set and the jackknife parameter estimates are calculated.
 - For large data sets “delete 1” may be impractical.
- “Delete 10%”:
 - Similar as before but the removal regards to 10% of the units in the sample.
- In general, Bootstrap is preferred to Jackknife as it produces overall better results.

Jackknife

Jackknife



Jackknife Parameter Pseudovalue Delete 1 Method

Let

$\hat{\theta}^{(JK,i)}$ be the Jackknife estimate of θ from i -th jackknife data set
 $\hat{\theta}$ be the estimate under the original data set.

Then the “Jackknife pseudovalue” corresponding to θ is given by:

$$P_i^{(JK)} = n\hat{\theta} - (n-1)\hat{\theta}^{(JK,i)}$$

And the Jackknife estimator of θ is the mean of the pseudovalue: \bar{P} .
The SE of $\hat{\theta}$ is then:

$$SE(\hat{\theta}) = \sqrt{\frac{\sum_{i=1}^n (P_i^{(JK)} - \bar{P})^2}{n(n-1)}}$$

Same procedure applies for the estimates of Ψ and σ_ε^2 .

Log-Likelihood Profiling

- Also known as “objective function profiling” or “objective function mapping”.
- Serially fix all the model parameters at their final parameter estimates.
- Incrementally change (increase or decrease) the value of the parameter of interest until the objective function value (OFV) changes by a pre-determined amount.
- The cut-off values where the OFV change determine the limits of the CI for that parameter.
- Some drawbacks (warnings):
 - Sensitivity to the starting values of estimates: often requires “fine-tuning” of the initial estimates sometimes for each evaluation, by the user.