Stochastic models

Uppsala Pharmacometrics Research Group
(Ron Keizer, Mats Karlsson)

June 3, 2013

Contents

1 Residual error models

1.1 Interindividual and interoccasion variability in residual error magnitude

The residual error magnitude may differ within individuals, as well as over time. The assumption is usually made that ϵ's are identically distributed between individuals. Violation of this assumption may occur if the residual error magnitude *(i)* varies between subjects, or is dependent on *(ii)* the underlying process (i.e., absorption, distribution, or elimination) or *(iii)* the rate of change in the plasma concentration. Examples of error sources that could create such a nonidentical residual error magnitude are (numbering corresponding to above): *(i)* varying compliance between subjects *(ii)* varying agreement between different parts of the structural model and their respective underlying processes, and *(iii)* sampling time errors.

If the inter-individual variability is assumed to be log-normally distributed the following equation can be used to describe the residual error with inter-individual variability:

$$y_{ij} = \hat{y}_{ij} + \epsilon_{ij} \cdot e^{\eta_i} \tag{1}$$

The difference between the observed value, y_ij, and the subject specific prediction, y_{ij}, is described by the residual error term, ϵ_{ij}, and the exponential of the individual term, η_i. Both stochastic terms are assumed to be normally distributed with a mean of 0 and a variance of σ^2 and ω^2, respectively.

Relevant NONMEM code:

```
; Example IIV on EPS
$ERROR
W = SQRT(THETA(1)**2+THETA(2)**2*IPRED*IPRED) * EXP(ETA(1))
Y = IPRED + W*EPS(1)

; Example IOV on EPS
FLAG1 = 0
FLAG2 = 0
IF(FLAG.EQ.1) FLAG1 = 1
IF(FLAG.EQ.0) FLAG2 = 1
IOV = FLAG1*ETA(1) + FLAG2*ETA(2)
W = SQRT(THETA(1)**2+THETA(2)**2*IPRED*IPRED) * EXP(IOV)
Y = IPRED + W*EPS(1)
```

Key publications:

- Karlsson et al, 1998, J Pharmacokinet Biopharmaceut
- Silber et al, 2009, J Pharmacokinet Pharmacodyn
- Plan et al, PAGE 2011, Abstr. 1983

1.2 Joint residual error between multiple observations

Several situations could occur in which residual errors are correlated between measurements: for example for replicate samples at the same timepoint (from a repeated bio-analysis of the same

sample), or from parent and metabolite measurements at the same timepoint. The residual error for these measurements can be expected to be correlated. In this situation, the difference between the repeated samples is expected to be smaller than between samples at different time points. To account for this correlation between repeated samples, the following structure for the residual error can be used:

$$y_{ijk} = \hat{y}_{ijk} + \epsilon_{ij} + \epsilon_{ijk} \tag{2}$$

The difference between the observed value of replicate k, y_{ijk}, and the subject specific prediction, $\hat{y}_{ij}$, is described by two residual error terms, ϵ_{ij} and ϵ_{ijk}. The first term corresponds to the constant residual error common to all measurements and the second term corresponds to the residual error in the replicate measurements. Both parameters are assumed to be normally distributed with a mean of 0 and a variance of ω^2 and σ^2, respectively.

The L2 data item in NONMEM is used to group together the observations which have the same realization of the residual errors (level-two random effects, epsilons). The group of such observations is called a level-two (L2) record. In NONMEM, data records of an L2 record must be contiguous (and contained within the same individual record).

Relevant NONMEM code (parent - metabolite):

```
$ERROR
IF (TYPE.EQ.1) THEN
  Y = IPRED + EPS(1) ; Observation type 1
ELSE
  Y = IPRED + EPS(2) ; Observation type 2
ENDIF

$SIGMA BLOCK(2)
0.1
0.05 0.1

; Dataset snippet:
  ID   TIME    AMT    DV  TYPE L2
   1    0.   25.0     .   .   1
   1    2.0     .    6.0  1   1
   1    2.0     .   17.3  2   1
   ...
   1  108.5   3.5     .   2   2
   1  112.5     .    8.0  1   2
   1  112.5     .   31.0  2   2
```

Relevant NONMEM code (replicate observations):

```
$ERROR
FLG = 0              ; Replicate 1
IF (TYPE.EQ.2) FLG = 1 ; Replicate 2
Y = IPRED + EPS(1) + (1-FLG)*EPS(2) + FLG * EPS(3)

$SIGMA 0.1  ; "Shared" error
$SIGMA BLOCK (1) 0.05
$SIGMA BLOCK (1) SAME
```

Key publications:

- Karlsson et al, J Pharmacokinet Biopharmaceut, 1995

1.3 Combination of L2 data items with likelihood based inclusion of BLQ data

The joint residual can be implemented as above for continuous observations. However, part of the observations of the dependent variable may be below the limit of quantification, BLOQ (or above the upper limit of quantification, ULOQ). The M3 method described by Beal incorporates the likelihood for observed BLOQ data being BLOQ, given the model and parameter estimates. However, when part of the observations in one L2 data group is BLOQ and the other part are regular continuous observation(s), it is currently not possible in NONMEM to use the M3 method for such an L2 group.

Key publications:

- Beal 2001, J Pharmacokinet Pharmacodyn

1.4 Serial correlation

The AR1 model allows serially correlated errors, as may occur with structural model misspecification. Ignoring this error structure leads to biased random-effect parameter estimates. This situation often occurs in situations where frequent measurements are made. A simple autocorrelation model, where the correlation between two errors is assumed to decrease exponentially with the time between them, provides more accurate estimates of the variability parameters in this case. In the AR1 model the positive correlation between two errors, ϵ_{t1} and ϵ_{t2}, decreases exponentially with the time separating the two observations according to:

$$corr(\epsilon_{t1}, \epsilon_{t1}) = e^{(-|t_1 - t_2|/t_{corr})} \tag{3}$$

where t_{corr} is a constant determining how fast the correlation decreases with time. It may be possible to put inter-individual variability on t_{corr}, but this has not been tried.

Relevant NONMEM code:

```
$PRED  ; or $ERROR
; Autocorrelation
"FIRST
"USE SIZES, ONLY: NO
"USE NMPRD_REAL, ONLY: C=>CORRL2
"REAL (KIND=DPSIZE) :: T(NO)
"INTEGER (KIND=ISIZE) :: I,J
"MAIN
"IF(NEWIND.NE.2)I=0
"IF(MDV.EQ.0)THEN
"   I=I+1
"   T(I)=TIME
"   DO 5 J=1,I
"   C(J,1)=EXP(-CORR*(TIME-T(J)))
"   5 CONTINUE
"ENDIF
CORR = LOG(2)/THETA(4)

Y=IPRED+EPS(1)

$SIGMA .1
```

Key publications:

- Karlsson et al, 1995, J Pharmacokinet Biopharmaceut

1.5 Residual error magnitude varying with covariates and time

Sources of variability such as erroneous dosing and sampling history and model misspecification may introduce errors whose variances are time-dependent. Inappropriately recorded sampling times introduce larger errors when the concentrations are rapidly changing than when they are not. Also, pharmacokinetic absorption models are generally less well specified than disposition models, resulting in potentially larger error magnitudes in the absorption phase. The error variance of PD data can also be time-dependent, an example may be PD data from *in situ* animal models, which tend to be less reliable as time progresses. A step model can be used where the transition time is estimated directly (example 1) or as a function of other parameters (e.g. mean absorption time, example 2).

Relevant NONMEM code:

```
; example 2:
$ERROR
PROP = THETA(1)
ADD  = THETA(2)
FACT = THETA(3)
KA   = THETA(4)
MAT  = 1 / KA ; Mean aborption time
IF(TIME.GT.MAT) FACT = 1
W = SQRT(ADD**2+PROP**2*IPRED*IPRED)*FACT
Y = IPRED + W*EPS(1)

; example 2:
$ERROR
PROP = THETA(1)
ADD  = THETA(2)
FACT = THETA(3)
TT = THETA(5) ; Transition time
IF(TIME.GT.TT) FACT = 1
W = SQRT(ADD**2+PROP**2*IPRED*IPRED)*FACT
Y = IPRED + W*EPS(1)
```

Key publications:

- Karlsson et al, 1995, J Pharmacokinet Biopharmaceut

1.6 Residual error magnitude varying with derivatives of functions w.r.t. to time and parameters

The magnitude of the residual error is often considered to be dependent on the predicted value of the dependent variable ($\hat{y}$). However, it has been found that sometimes additional factors, such as the time after dose (Karlsson et al, JPB, 1995) and $\partial\hat{y}/\partial t$ (Ruppert et al, Biometrics, 1993), are important. It can be investigated whether the residual error magnitude is heterogeneous with respect to $\partial\hat{y}/\partial P$, where P is one of the individual PK parameters. Parameters are associated with specific processes, CL with elimination, V with distribution, and k_a and T_{lag} with absorption. If the description of one process by the PK model is inferior

to that of the others, the residual error magnitude may be expected to be larger when data are highly influenced by that process, i.e., the absolute value of the derivative $\partial\hat{y}/\partial P$ is high. This may occur because the parameters applied are time-averaged, despite knowing that the absorption and disposition characteristics may change from time to time. For a process where time to time variation is particularly large, such time-averaging of parameters will result in more of an approximation than for others. To accommodate these two possible influences on the residual error, the following models can be implemented:

$$\ln(y_{obs,i}) = \ln(\hat{y}_i) + \epsilon \left[1 + \left(\theta_2 \left|\frac{\partial \hat{y}_i}{\partial t}\right|^{+} + \theta_3 \left|\frac{\partial \hat{y}_i}{\partial t}\right|^{-}\right)^2\right] \tag{4}$$

$$\ln(y_{obs,i}) = \ln(\hat{y}_i) + \epsilon \left[1 + \left(\theta_2 \left|\frac{\partial \hat{y}_i}{\partial P}\right|^{+} + \theta_3 \left|\frac{\partial \hat{y}_i}{\partial P}\right|^{-}\right)^2\right] \tag{5}$$

Relevant NONMEM code:

```
$ERROR
DERI  = ABS( K12*A(1)/V - K20*A(2)/V)
IPRED = LOG(.025)
W     = SQRT(THETA(5)**2+THETA(16)**2*DERI**2)
IF(F.GT.0) IPRED=LOG(F)
IRES  = IPRED-DV
IWRES = IRES/W
Y     = IPRED+ERR(1)*W
```

Key publications:

- Karlsson et al, J Pharmacokinet Biopharmaceut, 1998

1.7 Mixture model for residual error

A mixture model for the residual error can e.g. be applied to account for heavily tailed distributions of residual errors. Two distributions of residuals are implemented, one 'regular' group, and one 'outlier' group. The likelihood of individual observations of belonging to either group can then be estimated. A likelihood mixed model approach on the residual, as well as the magnitude (variance) of the residual error of the outliers and the non-outliers is defined:

$$L_1 = \frac{1}{\sqrt{2\pi\sigma^2_{normal}}} e^{\left(-\frac{y_{ij}-\hat{y}_{ij}}{\sigma_{normal}}\right)^2} \tag{6}$$

$$L_2 = \frac{1}{\sqrt{2\pi\sigma^2_{high}}} e^{\left(-\frac{y_{ij}-\hat{y}_{ij}}{\sigma_{high}}\right)^2} \tag{7}$$

$$L = (1 - MP) \cdot L_1 + MP \cdot L_2 \tag{8}$$

where y_{ij} is the observed value, y_{ij} is the subject specific prediction, L is the joint likelihood of all observations, and MP is the fraction of outliers. L_1 and L_2 are the likelihoods for the observations with normal or high residual error, respectively.

Relevant NONMEM code:

```
$PRED
CL = THETA(1)*EXP(ETA(1))
DOSE = 1
MU=LOG(DOSE/CL)
MP=THETA(2)
SIG1=THETA(3)
SIG2=THETA(4)

IW1=(DV-MU)/SIG1
IW2=(DV-MU)/SIG2

LL1=-0.5*LOG(2*3.14159265)-LOG(SIG1)-0.5*(IW1**2)
LL2=-0.5*LOG(2*3.14159265)-LOG(SIG2)-0.5*(IW2**2)
L=(1-MP)*EXP(LL1)+MP*EXP(LL2)
Y=-2*LOG(L)

$THETA (0.1,1) ;1 CL
$THETA (0,.2,1) ;2 MP
$THETA (0,.1)  ;3 SIG1
$THETA (0,1)   ;4 SIG2

$OMEGA 0.1

$EST MAXEVAL=9990 -2LL METH=1 LAPLACE
```

Key publications:

- Silber et al, J Pharmacokinet Pharmacodyn, 2009

1.8 Prior on residual error magnitude

Priors can be used to incorporate prior information into a problem. They can also be used to stabilize estimation of a mixed-effects analysis. Priors can be implemented on both fixed effects and random effects, although NONMEM does not allow direct implementation of priors on $SIGMA. To implement a prior on residual error magnitude, the standard deviation of Σ must be estimated as a fixed-effect, e.g.:

```
$ERROR
IPRED = F
W = THETA(1)
Y = IPRED + W*EPS(1)

$SIGMA
1 FIX
```

The prior can then be implemented on θ_1. See the section 'Prior information on random effects parameters' for NONMEM code for NWPRI and TNPRI routines.

Key publications:

- Gisleskog et al, J Pharmacokinet Pharmacodyn, 2002

1.9 Flexible errors-in-variables models

Errors-in-variables models or *measurement errors models* are regression models that account for measurement errors in the independent variables (usually *time* in pharmacometric applications). Standard regression models assume that those regressors have been measured without error, and account only for errors in the dependent variables. Errors-in-variables models models have so far not been applied in pharmacometrics.

E.g. consider the general non-linear model:

$$y_i = g(x_i^*, \beta_i) + \epsilon_i \tag{9}$$

where x_i^* denotes the true but unobserved value of the regressor. In errors-in-variables models we define this value to be associated with an error:

$$x_t = x_t^* + \eta_t \tag{10}$$

If possible, it would be nice if flexible errors-in-variables could be described by the language, although is currently not handled in NONMEM nor Monolix.

Key publications:

- http://en.wikipedia.org/wiki/Errors-in-variables_models

1.10 Transformation of residual error variables

In the 'Transform-both-sides' (TBS) approach, both the observed dependent variable, and the predicted dependent variable are transformed using the same transformation. In general terms this transformation is described using:

$$h(y_{ij}) = h\{f(x_{ij}, \beta_i)\} + \epsilon_{ij} \tag{11}$$

where h is a monotone transformation, such as log-transform (LTBS), or Box-Cox transform. The Box-Cox transform is a power transform that uses the power parameter λ:

$$y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log y_i, & \text{if } \lambda = 0 \end{cases} \tag{12}$$

Log-transform is a special case of the Box-Cox transform, where $\lambda = 0$.

Relevant NONMEM code:

```
; Box-Cox example
$ERROR
  LAMB = 0.5
  IPRED = (F**LAMB - 1) / LAMB
  Y = IPRED + EPS(1)

; NB: DV in dataset requires the same transformation (i.e. with lamb = 0.5)
```

Estimated transformation

The optimal value for the λ parameter in the Box-Cox transformation can be obtained by evaluating a range of values between 0 and 1. It would however be more appropriate to estimate λ. Current version of NONMEM do not allow this directly, but with user-specified CONTR and CCONTR routines this can be done (at least in version NM6).

```
; Estimated transformation of residuals
; NB: might not work in >NM6
$SUB CONTR=CONTR.TXT CCONTR=CCONTR.TXT
...
$PRED
     W=THETA(2)               ;SD
     CL=THETA(1)*EXP(ETA(1))  ;CL/F
     PRE=1/CL                 ;@ SS ASSUMING INPUT RATE = 1
     LAM=THETA(3)
     Y=(PRE**LAM-1)/LAM+EPS(1)*W
     RES1=(DV-PRE)/W
$THETA
     (0,.1)     ;CL/F
     (0,2)      ;SD ADDITIVE
     (0,.251)      ;BOX COX LAMBDA PARAMETER
...
```

File contr.txt:

```
      subroutine contr (icall,cnt,ier1,ier2)
      double precision cnt
      call ncontr (cnt,ier1,ier2,l2r)
      return
      end
```

File ccontr.txt (NM6):

```
      subroutine ccontr (icall,c1,c2,c3,ier1,ier2)
      parameter (lth=40,lvr=30,no=50)
      common /rocm0/ theta (lth)
      common /rocm4/ y
      double precision c1,c2,c3,theta,y,w,one,two
      dimension c2(*),c3(lvr,*)
      data one,two/1.,2./
      if (icall.le.1) return
      w=y

         y=(y**theta(3)-one)/theta(3)

      call cels (c1,c2,c3,ier1,ier2)
      y=w
      c1=c1-two*(theta(3)-one)*log(y)

      return
      end
```

File ccontr.txt (NM7 or higher):

```
subroutine ccontr (icall,c1,c2,c3,ier1,ier2)
USE ROCM_REAL, ONLY: theta=>THETAC,y=>DV_ITM2
USE NM_INTERFACE,ONLY: CELS
double precision c1,c2,c3,w,one,two
dimension c2(:),c3(:,:)
data one,two/1.,2./
if (icall.le.1) return
w=y(1)

   y(1)=(y(1)**theta(3)-one)/theta(3)



call cels (c1,c2,c3,ier1,ier2)
y(1)=w
c1=c1-two*(theta(3)-one)*log(y(1))

return
end
```

Key publications:

- Box & Cox, J Royal Stat Soc B, 1964
- Oberg & Davidian, Biometrics, 2000
- Frame B et al, J Pharmacokinet Pharmacodyn, 2007
- Slides Bill Frame (http://www.thtxinfo.com/)

2 Random effects models

2.1 Mixture models with estimation of individual probability of belonging to each subpopulation

Mixture models can be used for multimodal distributions of parameters. The fraction of individuals belonging to each of the subpopulations can be estimated, and the most probable subpopulation for each patient is output ($MIXEST_k$). The objective function value (OFV) that is minimized is the sum of the OFVs for each patient (OFV_i), which in turn is the sum across the k subpopulations ($OFV_{i,k}$). The $OFV_{i,k}$ values can be used together with the total probability in the population of belonging to subpopulation k to calculate the individual probability of belonging to the subpopulation (IP_k). A probability estimate such as IP_k provides more detailed information about each individual than the discrete $MIXEST_k$. Individual parameter estimates based on IP_k should be preferable whenever individual parameter estimates are to be used as study output or for simulations.

$$OFV = \sum_{i=1}^{n} OFV_i = \sum_{i=1}^{n} -2\ln(IL_i) \tag{13}$$

$$IL_i = \sum_{k=1}^{m} IL_{i,k} \cdot P_{pop,k} = \sum_{k=1}^{m} \exp(-OFV_{i,k}/2) \cdot P_{pop,k} \tag{14}$$

$$IP_k = \frac{IL_{i,k} \cdot P_{pop,k}}{\sum_{k=1}^{m} IL_{i,k} \cdot P_{pop,k}} \tag{15}$$

OFV_i's can be obtained in a single run in NM7 or later versions. From OFV_i's obtained in a run in which the $P_{pop,k}$ is set to 1 and one in which is set to 0, IP_k can be calculated for all individuals. This functionality is implemented in PsN (pind).

Relevant NONMEM code:

```
$MIX
NSPOP=2
P(1) = THETA(3)
P(2) = 1-THETA(3)

$PK
CL = THETA(1)*EXP(ETA(1))
FACT = THETA(2)
IF (MIXNUM.EQ.1) CL = THETA(1)*FACT*EXP(ETA(1))
```

Key publications:

- Carlsson et al, AAPSJ, 2009

2.2 Nonparametric (discrete) distributions with estimation of individual probability of belonging to each subpopulation

In non-parametric modeling approaches, there is no assumptions as to the shape of the distribution of individual parameters. Instead, a discrete probability density distribution is estimated. In NONMEM, the probability density for all parameters is calculated using support points at discrete intervals over the parameter space. At each support point, the probability observing that parameter value is calculated, given the current model, parameters and studied population. At each separate support point, the probability density can be broken down into individual contributions to the total probability density. These individual contributions can be calculated using PsN, in a similar procedure as outlined in 2.1.

Key publications:

- Baverel et al, J Pharmacokinet Pharmacodyn, 2009

2.3 Semiparametric distributions

To relax the often erroneous assumption of a known shape of the parameter distribution, transformations with estimated shape parameters can be applied to parameter distributions. The logit, Box-Cox, and heavy tailed transformations are examples of such transformations.

Relevant NONMEM code:

```
; Logit transformation
TVCL=THETA(1)
LGPAR1 = THETA(2)
LGPAR2 = THETA(3)
PHI = LOG(LGPAR1/(1-LGPAR1))
PAR1 = EXP(PHI+ETA(1))
ETATR = (PAR1/(1+PAR1)-LGPAR1)*LGPAR2
CL=TVCL*EXP(ETATR)

; Box-Cox transformation
SHP = THETA(2) ; shape parameter
PHI=EXP(ETA(1)) ; Exponent of normally distributed ETAs
PHI2=(PHI**SHP-1)/SHP ; Transformed ETA
CL = THETA(1)*EXP(PHI2)

; And separately (heavy- tailed  distribution)
SHP = THETA(2) ; shape parameter
PHI=ETA(1)*ABS(ETA(1))**SHP ; Transformed ETA
CL = THETA(1)*EXP(PHI)
```

Key publications:

- Petersson et al, Pharm Res, 2009

2.4 Several random effects per structural parameter

If appropriately accounted for in a PK(PD) model, time-varying covariates can provide additional information to that obtained from time-constant covariates. Time-varying covariates can be modeled using extensions to the extended inter-individual variability models. Two examples are given here.

The first model estimates different covariate – parameter relationships for within- and between-individual variation in covariate values, by splitting the standard covariate model into a baseline covariate (BCOV) effect and a difference from baseline covariate (DCOV) effect:

$$P_{pop} = \theta_p \cdot [1 + \theta_{BCOV} \cdot (BCOV_i - BCOV_{median}) + \theta_{DCOV} \cdot \Delta COV] \tag{16}$$

The second model allows the magnitude of the covariate effect to vary between individuals, by inclusion of interindividual variability in the covariate effect:

$$P_i = \theta_p \cdot [1 + \theta_{cov} \cdot e^{\eta_{cov,Pi}} (COV_i - COV_{median})] \cdot e^{\eta_{Pi}} \tag{17}$$

where $\eta_{COV,Pi}$ is a (zero mean, variance ω^2) random variable, which allows the magnitude of the covariate effect to differ between individuals.

Relevant NONMEM code:

```
; model 1
CEFF = THETA(2) ; effect of difference from median COV
CDEFF = THETA(3) ; time-varying covariate effect
DCOV = COV - BCOV ; BCOV = baseline COV
PAR1 = THETA(1) * (1 + CEFF*(BCOV-BCOVM) + CDEFF*DCOV)

; model 2
```

```
CEFF = THETA(2) ; effect of difference from median COV
PAR1 = THETA(1) * (1 + CEFF*EXP(ETA(1)) * (COV-COVM)) * EXP(ETA(2))
```

Key publications:

- Wählby et al, Br J Clin Pharmacol, 2004

2.5 Access to partial derivatives of the predicted response w.r.t. individual random effects

A first-order conditional estimation (FOCE)-based linear approximation can be used e.g. to quickly approximate the change in objective function values of covariate-parameter models. The model is linearized around the empirical Bayes estimates using ($\eta_{\hat{ki}}$) using:

$$y \approx f(\vec{p}, \vec{x_{ij}})|_{Q_0} + \sum_{k=1}^{m} \frac{\partial f}{\partial \eta_{ki}}|_{Q_0}(\eta^*_{ki} - \hat{\eta}_{ki}) + \sum_{k=1}^{n} \frac{\partial f}{\partial g_{ki}}|_{Q_0}(g^*_{ki} - \hat{g}_{ki}) +$$
$$\sum_{l=1}^{t} \epsilon^*_{ijl} \left(\frac{\partial h_{ij}}{\partial \epsilon_{ijl}}|_{Q_0} + \sum_{k=1}^{m} \frac{\partial}{\partial \eta_{ki}} \left(\frac{\partial h}{\partial \epsilon_{ijl}}|_{\vec{\epsilon}=\vec{0}} \right) |_{Q_0}(\eta^*_{ki} - \hat{\eta_{ki}}) + \sum_{k=1}^{n} \frac{\partial}{\partial g_{ki}} \left(\frac{\partial h}{\partial \epsilon_{ijl}}|_{\vec{\epsilon}=\vec{0}} \right) |_{Q_0}(g^*_{ki} - \hat{g}_{ki}) \right)$$

where $Q_0 = \{\vec{\epsilon_{ij}} = \vec{0}, \vec{\eta} = \hat{\vec{\eta}}, g_{ki} = \hat{g}_{ki}\}$, m and n are the total number of η's and covariate functions g in the model. $f(\vec{p_i}, \vec{x_{ij}})|_{Q_0}$ are the model predictions based on $\hat{\vec{\eta}}$, $\vec{\theta}$ and $\hat{\vec{g}}$. The residual function is defined by h_{ij}, with ϵ following a symmetrical distriubtion with variance-covariance matrix Σ. The parameters marked with an $*$ are not known after the minimization of the original nonlinear model, and are estimated in each linearized model in the scm. To perform this transformation to a linear model, access to partial derivatives of the predicted response with respect to individual random effects is necessary, and these can be extracted from NONMEM.

Relevant NONMEM code: (to obtain derivatives)

```
$PK
...
$TABLE ID DV MDV IPRED=OPRED W G011 G021 G031 G041 G051 G061 G071 G081
```

Relevant NONMEM code: (linearized model)

```
$PROBLEM

$INPUT     ID DV MDV OPRED OW D_ETA1 D_ETA2 D_ETA3 D_ETA4 D_ETA5
           D_ETA6 D_ETA7 OETA1 OETA2 OETA3 OETA4 OETA5 OETA6 OETA7
           OGZ_CL OGK_CL OGZ_V OGK_V
$DATA      derivatives.dta IGNORE=@
$PRED
GZ_CL = 1
GZ_V = 1
COV1=D_ETA1*OGK_CL*(GZ_CL-OGZ_CL)
COV2=D_ETA2*OGK_V*(GZ_V-OGZ_V)
CSUM1=COV1+COV2
COV_TERMS=CSUM1
BASE1=D_ETA1*(ETA(1)-OETA1)
BASE2=D_ETA2*(ETA(2)-OETA2)
```

```
BASE3=D_ETA3*(ETA(3)-OETA3)
BASE4=D_ETA4*(ETA(4)-OETA4)
BASE5=D_ETA5*(ETA(5)-OETA5)
BASE6=D_ETA6*(ETA(6)-OETA6)
BASE7=D_ETA7*(ETA(7)-OETA7)
BSUM1=BASE1+BASE2+BASE3+BASE4+BASE5+BASE6+BASE7
BASE_TERMS=BSUM1
IPRED=OPRED+BASE_TERMS+COV_TERMS
Y=IPRED+OW*EPS(1)

$OMEGA BLOCK(2)
 0.1
 0.05 0.1 ; IIV (CL-V)
$OMEGA BLOCK(1)
 1  ;     IIV KA
$OMEGA BLOCK(1)
 0.1;    IOV CL
$OMEGA BLOCK(1) SAME

$OMEGA BLOCK(1)
 0.5;    IOV KA
$OMEGA BLOCK(1) SAME

$SIGMA 0.1
$ESTIMATION METHOD=ZERO FORMAT=s1PE17.10
```

Key publications:

- Khandelwal et al, 2011, AAPSJ

2.6 Possibility to constrain a variance for an individual random effect to be the same value as a residual error variance during estimation

One of several models for the estimation of baseline values (BL_i) uses the magnitude of residual error to model baseline. In this approach, BL_i is estimated to deviate from the individual observed baseline ($BL_{i,o}$) by a random component given by $\eta_{i,RV}$ which is a zero-mean variable that during parameter estimation is constrained to have the same variance as the residual variability (σ^2):

$$BL_i = BL_{i,o} \cdot e^{\eta_{i,RV}} \tag{18}$$

This can be achived in NONMEM by introducing $\eta_{i,RV}$ into the model as $\eta_i \cdot \theta_\epsilon$ where the variance of η_i is fixed to 1, and $\theta_{epsilon}$ is the magnitude of residual error variability (on standard deviation scale) estimated from non-baseline observations (in $ERROR). This approach allows the residual variability in baseline to be varied among individuals, but to be the same within each individual.

Relevant NONMEM code:

```
$PRED
-----Baseline model-----------
IF(TIME.EQ.0)THEN
  OBASE = EXP(ODV)              ; observed baseline response at
ENDIF                           ; time 0

IBASE = OBASE*EXP(ETA(1)*THETA(1))
```

```
...
Y = IPRED + EPS(1) * THETA(1)

$THETA
(0, 0.3) ; 1 RV magnitude (sd)
$OMEGA
1 FIX
$SIGMA
1 FIX
```

Key publications:

- Dansirikul et al, J Pharmacokinet Pharmacodyn, 2008

2.7 Prior information on random effects parameters, specified using different distributions - normal or inverse wishart

Prior information can be used to stabilize estimation of a mixed-effects analysis. The priors can be implemented on both fixed effects and random effects. In NONMEM this can be done by using the NWPRI and TNPRI functions in the $PRIOR subroutine. NWPRI computes a function based on a frequency prior that has a multivariate normal form for THETA, and an inverse Wishart form for OMEGA (independent from the normal for THETA).

TNPRI computes a penalty function based on a frequency prior that has a multivariate normal form for all unconstrained parameters. When used during a Simulation Step, it produces a random value of the vector of all model parameters (whose values are not fixed).

Relevant NONMEM code:

```
...
$PRIOR NWPRI NTHETA=4, NETA=4, NTHP=0, NETP=4, NPEXP=1
$PK
MU_1=THETA(1)
MU_2=THETA(2)
MU_3=THETA(3)
MU_4=THETA(4)
CL=DEXP(MU_1+ETA(1))
V1=DEXP(MU_2+ETA(2))
Q=DEXP(MU_3+ETA(3))
V2=DEXP(MU_4+ETA(4))
S1=V1

$ERROR
Y = F + F*EPS(1)

$THETA 2.0 2.0 4.0 4.0 ; Initial Thetas

$OMEGA BLOCK(4) ; Inital Parameters for OMEGA
0.4
0.01 0.4
0.01 0.01 0.4
0.01 0.01 0.01 0.4

$SIGMA 0.1
; Set the Priors. Good Idea if Doing MCMC Bayesian

$OMEGA BLOCK(4) ; Prior to OMEGA
0.2 FIX
0.0 0.2
0.0 0.0 0.2
```

```
0.0 0.0 0.0 0.2

$THETA (4 FIX) ; Set degrees of freedom of OMEGA PRior
```

Key publications:

- Gisleskog et al, J Pharmacokinet Pharmacodyn, 2003
- NONMEM Guide for version 7.2, 2011

2.8 Several parameters sharing a random effect (same eta) or sharing an eta which is scaled

For some problems, it may be appropriate to incorporate shared inter-individual variability in parameters, instead of model them as separate variability. This can be implemented in NONMEM as two parameters sharing the same η, possibly with different multiplicators for different parameters.

Relevant NONMEM code:

```
PAR1 = THETA(1) * EXP(ETA(1))
FACT = THETA(3)
PAR2 = THETA(2) * EXP(FACT*ETA(1))
```

3 Additional levels of random effects

Additional levels of random effects may be needed to provide a better description of the data. It would be preferable to have a general solution that allows infinite levels of random effects, with no restriction on number of 'IDs' per level. For the additional levels, the following code would be implemented (all models below have been implemented in NONMEM in estimation mode). Inter-occasion variability (IOV) is used here as example, but this could of course also be inter-study-variability (ISV).

3.1 Covariate effects on IOV

The implementation of IOV can be extended by covariate effects influencing the magnitude of IOV, just as one might for IIV and residual variability. An example of this was encountered in a study in both healthy volunteers and patients, where a significantly higher IOV in clearance was found in the latter group (unpublished observation).

Relevant NONMEM code:

```
SPK
FLAG1 = 0
FLAG2 = 0
IF(FLAG.EQ.1) FLAG1 = 1
IF(FLAG.EQ.0) FLAG2 = 1
IF(FLAG.EQ.1.AND.COV.EQ.1) FLAG1 = THETA(3) ; COV = covariate (0/1)
IF(FLAG.EQ.1.AND.COV.EQ.2) FLAG2 = THETA(3)
```

```
CL = THETA(1)*EXP(ETA(1) + FLAG1*ETA(3) + FLAG2*ETA(4))
V = THETA(2)*EXP(ETA(2) + FLAG2*ETA(5) + FLAG2*ETA(6))

$THETA
(0, 10) ; CL
(0, 100) ; V
(0, 2)  ; Factor scaling IOV magnitude between groups

$OMEGA
...
```

Key publications:

- Karlsson et al, J Pharmacokinet Biopharmaceut, 1993

3.2 Interindividual variability (or not) in IOV

Another extension of the IOV model could be to allow the magnitude of IOV to differ between individuals.

Relevant NONMEM code:

```
$PRED
IF(NEWIND.NE.2) PDV=DV
IF(NEWIND.EQ.0) MYID=1
IF(NEWIND.EQ.1) MYID=MYID+1
TTT =TRTT
IF(AZD.EQ.0) TTT=(DAY-9)*TPHS

 ;Baseline values
  B1   =THETA(1)
  B2   =THETA(7)
  B3   =THETA(8)
  B4   =THETA(9)
  B5   =THETA(10)
;----------------IOV------------
OCC1=0
IF(TTT.LT.14) OCC1=1
IOV1=ETA(4)*OCC1+ETA(5)*(1-OCC1)
IOV3=ETA(8)*OCC1+ETA(9)*(1-OCC1)
IF(TPHS.EQ.0) IOV1=ETA(6)
IF(TPHS.EQ.0) IOV3=ETA(7)
;Placebo

PLC  = (THETA(2)+ETA(2)+(THETA(3)+ETA(3)+IOV3)*TTT)*TPHS

;Dose-effect relationship
DRUG= THETA(6)*TTT*TPHS*AZD

;Morning effect
MORN= 0
IF(CLC.LT.0.5) MORN = THETA(4)

;Markov effect
MARK=0
IF(PDV.EQ.1)  MARK = THETA(5)

;Logits for Y>=1, Y>=2, Y>=3
  LGE1 =B1+DRUG+ PLC + MORN + MARK+ETA(1)+IOV1*EXP(ETA(10))
  LGE2 =B1+B2+DRUG+ PLC + MORN + MARK+ETA(1)+IOV1*EXP(ETA(10))
  LGE3 =B1+B2+B3+DRUG+ PLC + MORN + MARK+ETA(1)+IOV1*EXP(ETA(10))
  LGE4 =B1+B2+B3+B4+DRUG+ PLC + MORN + MARK+ETA(1)+IOV1*EXP(ETA(10))
  LGE5 =B1+B2+B3+B4+B5+DRUG+ PLC + MORN + MARK+ETA(1)+IOV1*EXP(ETA(10))
```

```
;Probabilities for Y>=1, Y>=2, Y>=3, Y>=4, Y>=5
 PGE1  =EXP(LGE1)/(1+EXP(LGE1))
 PGE2  =EXP(LGE2)/(1+EXP(LGE2))
 PGE3  =EXP(LGE3)/(1+EXP(LGE3))
 PGE4  =EXP(LGE4)/(1+EXP(LGE4))
 PGE5  =EXP(LGE5)/(1+EXP(LGE5))

;Probabilities for Y=1, Y=2, Y=3, Y=4, Y=5
 P0   =(1-PGE1)
 P1   =(PGE1-PGE2)
 P2   =(PGE2-PGE3)
 P3   =(PGE3-PGE4)
 P4   =(PGE4-PGE5)
 P5   = PGE5
 PTOT=P0+P1+P2+P3+P4+P5
;Select appropriate P(Y=m)
DEL=1D-10
 IF(DV.EQ.0) Y=P0+DEL
 IF(DV.EQ.1) Y=P1+DEL
 IF(DV.EQ.2) Y=P2+DEL
 IF(DV.EQ.3) Y=P3+DEL
 IF(DV.EQ.4) Y=P4+DEL
 IF(DV.EQ.5) Y=P5+DEL
PDV=DV

"LAST
" IF (G(10,1).EQ.0D0) G(10,1)=1D-8

...
$ESTIMATION MAX=9990 PRINT=1 LIKE METH=1 LAPLACE ;NUMERICAL
```

Key publications:

- Karlsson et al, J Pharmacokinet Biopharmaceut, 1993

3.3 IOV with possibility to constrain, or not, IOV to differ across occasions

When coding IOV in NONMEM, we usually constrain the η's for different occasions to be drawn from the same Ω-distribution. Although this constraint improves stability of model estimation, it is not strictly necessary: if supported by the data, the magnitude can be estimated for all separate occasions.

Relevant NONMEM code:

```
; Usual coding of IOV:
$OMEGA BLOCK(1) 0.1
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME

; Estimation of different IOV magnitude across occasions:
$OMEGA BLOCK(1) 0.1
$OMEGA BLOCK(1) 0.1
$OMEGA BLOCK(1) 0.1
```

Key publications:

- Karlsson et al, J Pharmacokinet Biopharmaceut, 1993

3.4 IOV with estimated durations of occasion

While using a classic implementation of IOV, partial derivatives with respect to time are defined, the estimation of occasion length (OCCL) is not possible. However, by using surge functions, with differing peak height and surge width, it is possible to implement a more flexible IOV, and estimate the length of occasions.

$$\kappa = \sum_{n=1}^{\infty} \frac{\kappa_n}{[(T - PT_n)^2/SW^2]^{\gamma} + 1} \tag{19}$$

$$PT_n = (n - 0.5) \cdot L_{occ} \tag{20}$$

$$SW = L_{occ}/2 \tag{21}$$

in which L_{occ} is the occasion length, and γ is the shape parameter of the surge function, which can both be estimated.

Relevant NONMEM code:

```
$PRED
IF(NEWIND.NE.2) PDV=DV
TTT =TRTT
IF(AZD.EQ.0) TTT=(DAY-9)*TPHS

;----------------IOV------------
OCCL=THETA(11)+4
GAM=THETA(12)
SW=OCCL/2
OCC1=ETA(2)/(((OCCL-OCCL/2-DAY)**2/SW**2)**GAM+1)
OCC2=ETA(3)/(((2*OCCL-OCCL/2-DAY)**2/SW**2)**GAM+1)
OCC3=ETA(4)/(((3*OCCL-OCCL/2-DAY)**2/SW**2)**GAM+1)
OCC4=ETA(5)/(((4*OCCL-OCCL/2-DAY)**2/SW**2)**GAM+1)
OCC5=ETA(6)/(((5*OCCL-OCCL/2-DAY)**2/SW**2)**GAM+1)
OCC6=ETA(7)/(((6*OCCL-OCCL/2-DAY)**2/SW**2)**GAM+1)
OCC7=ETA(8)/(((7*OCCL-OCCL/2-DAY)**2/SW**2)**GAM+1)
OCC8=ETA(9)/(((8*OCCL-OCCL/2-DAY)**2/SW**2)**GAM+1)
OCC9=ETA(10)/(((9*OCCL-OCCL/2-DAY)**2/SW**2)**GAM+1)
OCC10=ETA(11)/(((10*OCCL-OCCL/2-DAY)**2/SW**2)**GAM+1)
OCC11=ETA(12)/(((11*OCCL-OCCL/2-DAY)**2/SW**2)**GAM+1)
IOV1=OCC1+OCC2+OCC3+OCC4+OCC5+OCC6+OCC7+OCC8+OCC9+OCC10+OCC11
OCC1=0
IF(TTT.LT.14) OCC1=1
IOV3=ETA(15)*OCC1+ETA(16)*(1-OCC1)
IF(TPHS.EQ.0) IOV3=ETA(17)
;Placebo

PLC  = (THETA(2)+ETA(13)+(THETA(3)+ETA(14)+IOV3)*TTT)*TPHS

;Dose-effect relationship
DRUG= THETA(6)*TTT*TPHS*AZD

;Morning effect
MORN= 0
IF(CLC.LT.0.5) MORN = THETA(4)

;Markov effect
MARK=0
IF(PDV.EQ.1)  MARK = THETA(5)
```

```
;Logits for Y>=1, Y>=2, Y>=3
  LGE1 =B1+DRUG + PLC + MORN + MARK+ETA(1)+IOV1
  LGE2 =B1+B2+DRUG + PLC + MORN + MARK+ETA(1)+IOV1
  LGE3 =B1+B2+B3+DRUG + PLC + MORN + MARK+ETA(1)+IOV1
  LGE4 =B1+B2+B3+B4+DRUG + PLC + MORN + MARK+ETA(1)+IOV1
  LGE5 =B1+B2+B3+B4+B5+DRUG + PLC + MORN + MARK+ETA(1)+IOV1

;Probabilities for Y>=1, Y>=2, Y>=3, Y>=4, Y>=5
  PGE1  =EXP(LGE1)/(1+EXP(LGE1))
  PGE2  =EXP(LGE2)/(1+EXP(LGE2))
  PGE3  =EXP(LGE3)/(1+EXP(LGE3))
  PGE4  =EXP(LGE4)/(1+EXP(LGE4))
  PGE5  =EXP(LGE5)/(1+EXP(LGE5))

;Probabilities for Y=1, Y=2, Y=3, Y=4, Y=5
 P0   =(1-PGE1)
 P1   =(PGE1-PGE2)
 P2   =(PGE2-PGE3)
 P3   =(PGE3-PGE4)
 P4   =(PGE4-PGE5)
 P5   = PGE5

;Select appropriate P(Y=m)
 IF(DV.EQ.0) Y=P0
 IF(DV.EQ.1) Y=P1
 IF(DV.EQ.2) Y=P2
 IF(DV.EQ.3) Y=P3
 IF(DV.EQ.4) Y=P4
 IF(DV.EQ.5) Y=P5
PDV=DV
IPRED=P1+P2*2+P3*3+P4*4+P5*5

$THETA 3.100000 ; 1 BASE
...

$OMEGA 5.010000 ;  1 OVERALL

$OMEGA BLOCK(1) 4.170000
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME

$OMEGA 2.220000 ; 2 PLC ICPT
$OMEGA 0.002430 ; 3 PLC TIME

$OMEGA BLOCK(1) 0.001990 ;     IOV3
$OMEGA BLOCK(1) SAME
$OMEGA BLOCK(1) SAME
```

Key publications:

- Karlsson 2009, ACoP, (Lewis_ACOP_2009.ppt, slide 47)

4 Stochastic differential equations

Intra-individual variability in nonlinear mixed-effects models based on SDEs is decomposed into two types of noise: a measurement and a system noise term. The measurement noise

represents uncorrelated error due to, for example, assay error while the system noise accounts for structural misspecifications, approximations of the dynamical model, and true random physiological fluctuations. Since the system noise accounts for model misspecifications, the SDEs provide a diagnostic tool for model appropriateness. They can also be used for parameter tracking. SDEs can be implemented in NM6 and up, but NM7.2 includes specific functionality to facilitate this.

Relevant NONMEM code:

```
; Example sde9 from NM7.2
$SUBROUTINE ADVAN6 TOL=10 DP OTHER=sde.f90

$MODEL
      COMP = (CENTRAL);
      COMP = (DFDX1)
      COMP = (DPDT11)

$PK
  IF(NEWIND.NE.2) OT = 0
  MU_1  = THETA(1)
  CL    = EXP(MU_1+ETA(1))
  MU_2  = THETA(2)
  VD    = EXP(MU_2+ETA(2))
  SGW1 = THETA(4)


$DES
" FIRST
" REAL*8 SGW(3)
" FIRSTEM=1
DADT(1)=-CL/VD*A(1)
DADT(2)=A(1)/VD
" SGW(1)=SGW1
"LAST
"     CALL SDE_DER(DADT,A,DA,IR,SGW,1.0d+00,1.0d+00)

$ERROR (OBS ONLY)
     IPRED = A(1)/VD
     IRES  = DV - IPRED
     W     = THETA(3)
     IWRES = IRES/W
     WS=1000.0
"LAST
"      CALL SDE_CADD(A,HH,TIME,DV,CMT,1.0D+00,1.0D+00,SDE)
```

Key publications:

- Tornoe et al, Pharm Res, 2005
- NONMEM User Guide for version 7.2, 2011