Supplementary Materials

Development of a disease progression model for LRRK2 kinase in Parkinson's disease to inform clinical trial designs

*Corresponding author: Malidi Ahamadi, email: malidi.ahamadi@merck.com

Covariate Analysis

Once the robustness of the base model was established, a series of exploratory covariate analyses were performed to assess: 1) correlation between covariates and 2) correlation between covariates and the disease progression parameters (i.e., baseline, slope). Results from exploratory covariate analyses were combined with prior knowledge of potential predictors of Parkinson's disease (PD) progression to select the combination parameter-covariates to be formally assessed. Covariates tested for inclusion in the model included demographic factors (age, gender, and body weight), disease duration, years of education, LRRK2 mutation, and Concomitant medication. Height was excluded from formal covariate analysis based on its clinical relevance and strong correlation with other demographic factors. The stepwise covariate selection procedure (SCM) [25] as implemented in PSN [24] was used to confirm findings from exploratory analyses. The SCM procedure involved stepwise testing of linear and nonlinear relationships in a forward inclusion (change in objective function value, DOFV, of 6.63, $P < 0.01$, chi-squared with 1 degree of freedom, DF) and backward exclusion (DOFV of 10.8, $P < 0.001$, chi-squared with 1 DF) procedure. The final result from covariate analysis is summarized in the Table S2

Dropout analysis

Exploratory analyses of Kaplan-Meier curves followed by parametric time-to-event model were conducted to: 1) assess the assumption of missing data mechanism, and hence, the disease progression model and 2) account for dropout of subjects during clinical trial simulations. The probability of having a dropout at any given time was described by the hazard associated with the dropout (i.e., event). Different hazard models were evaluated which included exponential, Weibull, log-normal, gamma, log-logistic, and Gompertz. In addition, an univariate covariate analysis was performed to identify predictors of rate. Model selection was guided by the Akaike information criterion (AIC). Information on whether a patient dropped out from the study and the reason was reported in the CPP-database (https://codr.c-path.org/login.do).

Model Evaluation

Assessment of model adequacy and decisions about increasing model complexity were guided by goodness-of-fit criteria which included evaluation of objective function value (OFV) and Bayesian information criterion (BIC) defined as OBJV + np*Ln(N), where np is the total number of parameters in the model, and N is the number of data observations. Visual inspection of diagnostic scatter plots such as observed vs predicted scores, plausibility of parameter estimates, and precision of parameter estimates were used to select the final model. Robustness of the model parameter estimates was assessed by means of a nonparametric bootstrap evaluation. The disease progression parameters were estimated repeatedly by fitting the final model to 1,000 bootstrap datasets sampled from the original dataset with replacement. The median values and 95% CIs of the parameter estimates from these 1,000 bootstrap datasets were compared with the point estimates from the final model. The predictive performance of the final model was assessed using

a VPC approach. A total of 1,000 simulated datasets were generated using the final model. Stratifying by covariates of interest, the observed score data were graphically overlaid with the median values and the 5th and 95th percentiles of the simulated score-time profiles. The performance of the model was deemed adequate if the observed score data were appropriately distributed within the 5th and 95th percentiles of the simulated data.

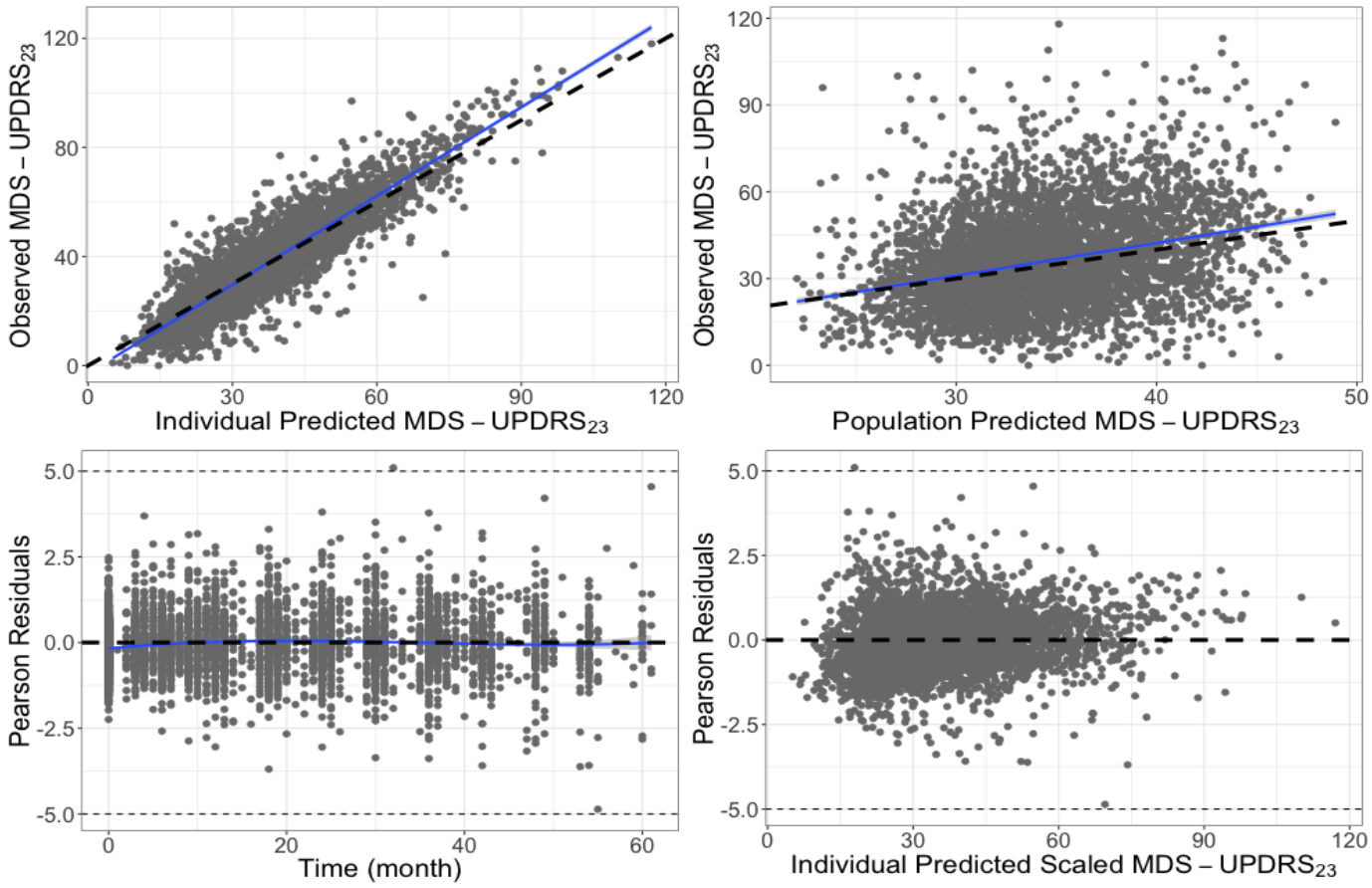

Figure S1. Goodness of fit plot for the final model

Abbreviations: MDS-UPDRS$_{23}$ = Movement Disorder Society-Unified Parkinson's Disease Rating Scale Part II plus Part III score

Note: Dots are individual data; solid lines are smooth lines. In the two plots of the first row, dashed lines are lines of identity, whilst in the two plots of the second row, dashed lines represent zero line. Pearson residuals were calculated as (DV-IPRED)/SQRT(IPRED*(1-IPRED)/(1+τ)), where DV is the dependent variable MDS-UPDRS$_{23}$ score, IPRED is the individual prediction, and τ the summation of the shape parameters of the beta distribution.

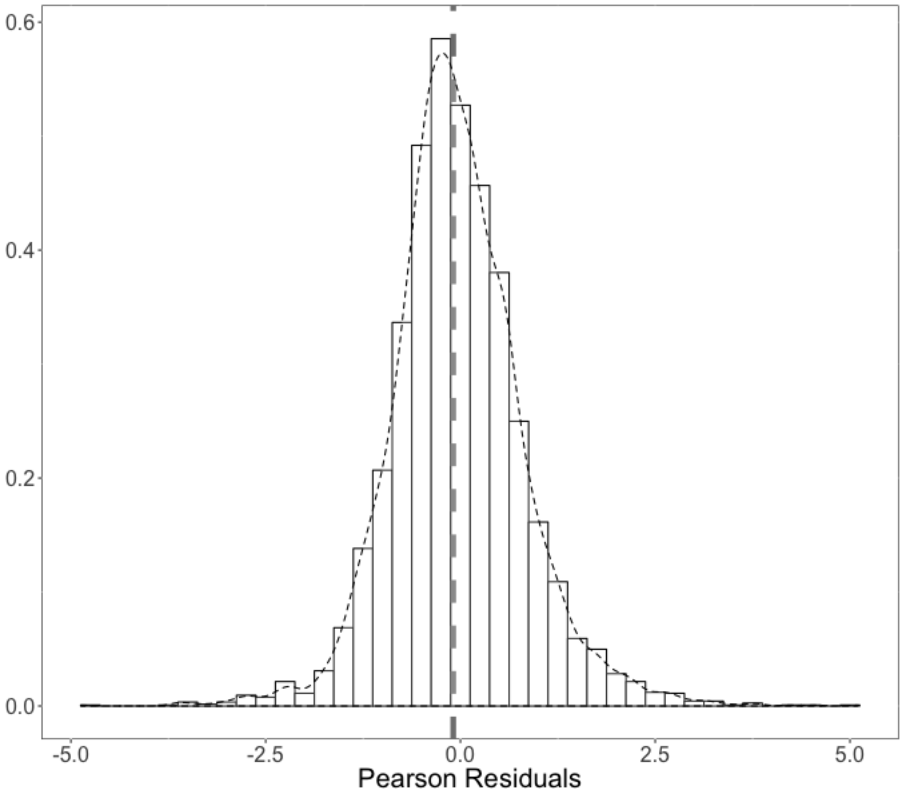

Figure S2. Distribution of the Pearson residual

Note: Dashed vertical line represents the median, which is close to zero. The thin dashed line represents a smoothed representation of the density.

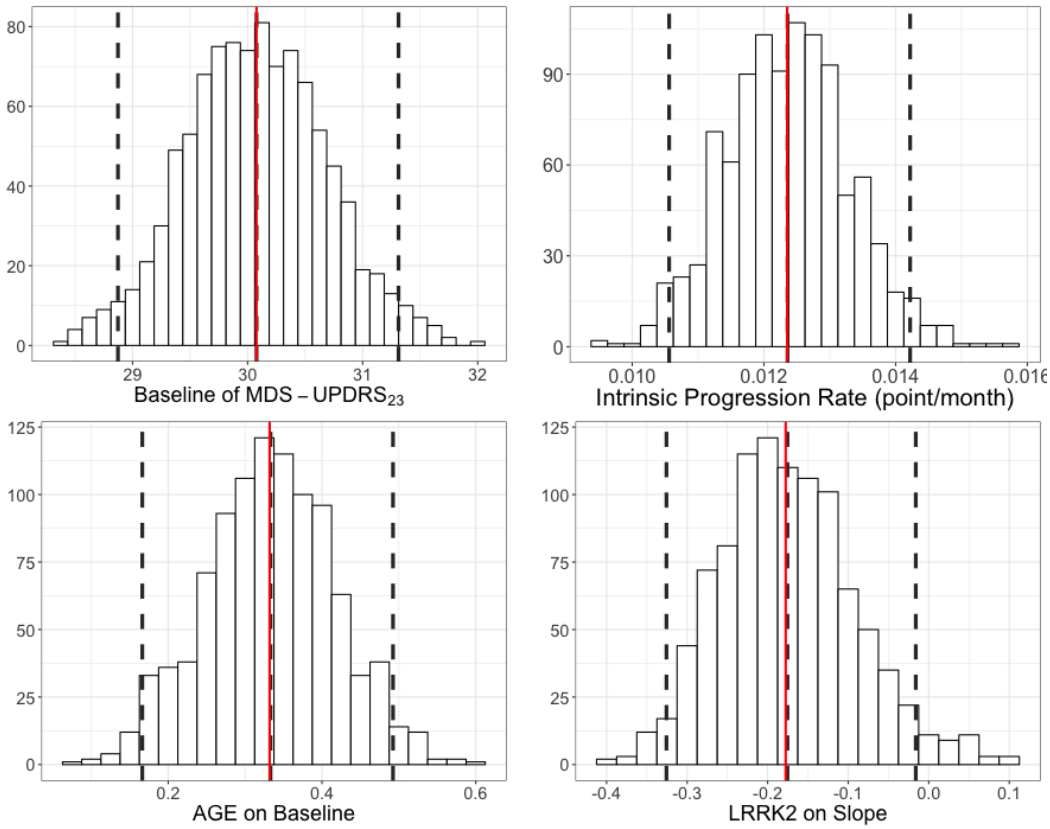

Figure S3. Comparison between parameter estimates and bootstrap results

Abbreviations: LRRK2 = Leucine-rich repeat kinase 2, MDS-UPDRS$_{23}$ = Movement Disorder Society-Unified Parkinson's Disease Rating Scale Part II plus Part III score
Note: Dashed grey vertical black lines are: bootstrap 2.5th, median, and 97.5th percentiles. Solid vertical red line is the original NONMEM estimate.

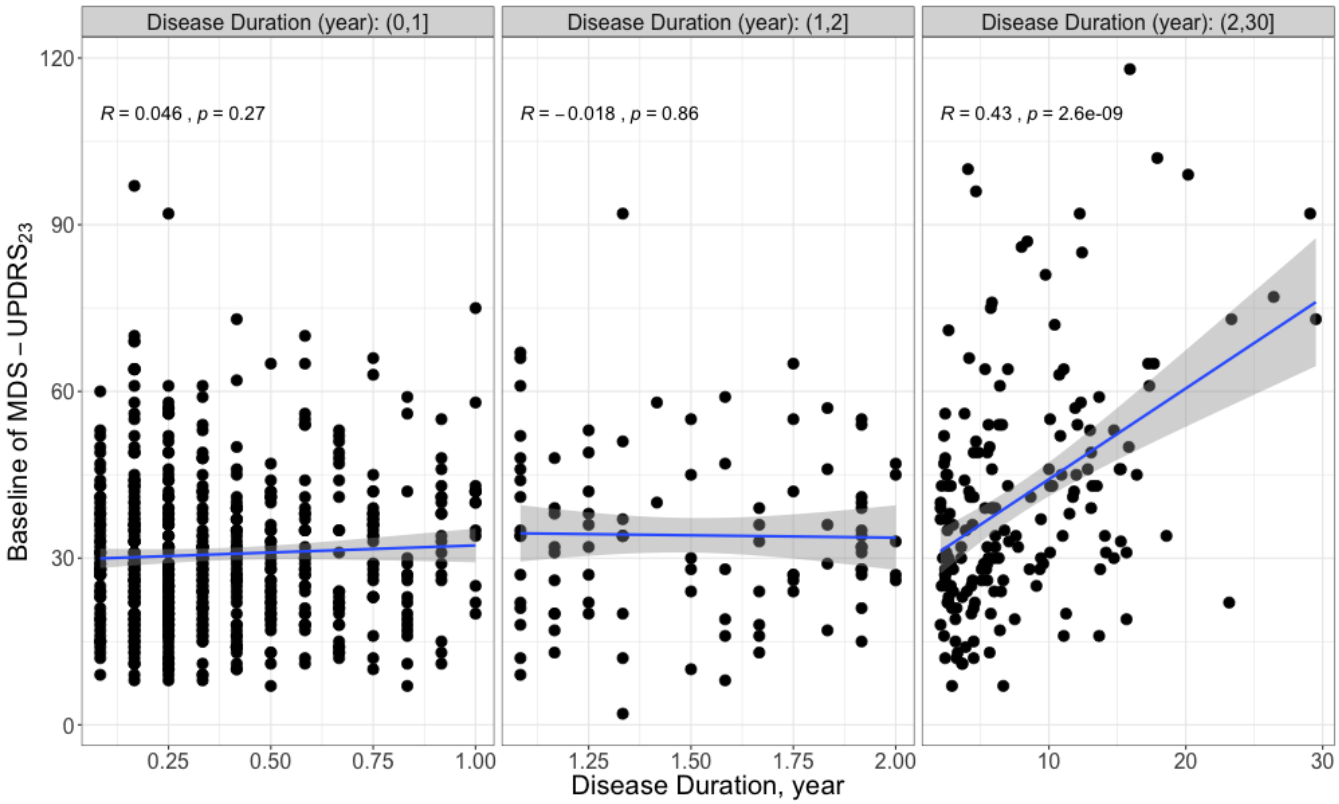

Figure S4: Correlation between disease duration at enrollment and baseline

Abbreviations: R=correlation coefficient, P=P-value

Table S1. Dropout model selection

Model	Number of Parameters	-2 Log Likelihood	Akaike's Information Criterion
Base Model			
Log-normal	2	1028.355	**1036.037**
Gamma	3	1027.401	1038.924
Log-logistic	2	1032.053	1039.735
Weibull	2	1033.072	1040.754
Exponential	1	1039.426	1043.267
Gompertz	2	1038.159	1045.841
Inclusion of covariates			
Log-normal+Other PD medication	3	1005.021	**1016.544**
Log-normal+Age	3	1015.981	1027.504
Log-normal+PD LRRK2	3	1021.732	1033.255
Log-normal+Female	3	1027.605	1039.128

Abbreviations: LRRK2 = Leucine-rich repeat kinase 2, PD = Parkinson's disease, "Log-normal+Other PD medication" will use for clinical trial simulation.

Table S2: Results from Covariate analysis

SCM Steps	Covariate	Parameters	ΔOFV	*df	**P-value
Last forward step	Age	Baseline	27.90178	1	<0.0001
	Gender	Baseline	19.20418	1	<0.0001
	Cohort	Baseline	18.99796	1	<0.0001
	Any PD medication	Baseline	12.35854	1	<0.0001
	LRRK2 mutation	Slope	10.75232	1	<0.0001
	Disease Duration	Baseline	8.084514	1	0.001237
	Gender	Slope	8.795231	1	0.004378
	Body Weight	Baseline	7.809354	1	0.002134
Last backward step	Age	Baseline	28.35721	1	<0.0001
	Gender	Baseline	18.74825	1	<0.0001
	Cohort	Baseline	16.17562	1	<0.0001
	Any PD medication	Baseline	11.94871	1	<0.0001
	LRRK2 mutation	Slope	10.23905	1	<0.0001

*difference in degrees of freedom; ** P-value derived from the chi-square distribution (forwards step acceptance level: 0.01, backward step acceptance level: 0.001).

Table S3: Base model structure

Structural model	Disease progression rate	OFV	*df	BIC
Linear Model $\frac{dScore}{dt} = r$	linear	-8724.428	2	-8707.535
Exponential model $\frac{dScore}{dt} = r * Score$	Non-linear	-8701.345	2	-8684.452
Standard logistic model $\frac{dScore}{dt} = r * Score * \left[1 - \frac{Score}{\max(Score)}\right]$	Non-linear Inflexion point=$\frac{\max(Score)}{2}$	-8743.309	2	-8726.416
Generalized logistic model $\frac{dScore}{dt} = r * Score * \left[\left(1 - \frac{Score}{\max(Score)}\right)^{\beta}\right]$	Non-linear Inflexion point=$\left(\frac{\max(Score)^{\beta}}{1+\beta}\right)^{\frac{1}{\beta}}$	-8742.953	3	-8717.613

Table S4: Parameter estimates of final model and stability assessment using non-parametric bootstrap analysis

Parameters	Base Model	Final Model	Non-parametric bootstrap
	Estimate (%RSE)	Estimate (%RSE)	Mean(95% CI)
MDS-UPDRS$_{23}$ Baseline	30.3 (1.54%)	29.5 (2.30%)	29.5 (28.3 , 30.8)
Age (Centered at 63 years old)	NE	0.376 (22.20%)	0.378 (0.199,0.563)
Female	NE	-0.0968 (27.70%)	-0.0964 (-0.148, -0.042)
Cohort (ICICLE-PD, PPMI Genetic Cohort PD, PPMI Genetic Registry PD)	NE	0.134 (25.30%)	0.133 (0.0689,0.203)
Any PD medications	NE	0.112 (40.40%)	0.111 (0.0217, 0.214)
Intrinsic Progression Rate (per month)	0.00974 (8.39%)	0.0101 (8.91%)	0.0102 (0.0085, 0.012)
LRRK2 mutation	NE	-0.235 (37.40%)	-0.250 (-0.513, -0.0269)
Dispersion factor of Beta Distribution	37.3 (2.53%)	39.2 (2.57%)	39.3 (35, 43.8)
Random Effect			
MDS-UPDRS$_{23}$ Baseline	0.162 (5.87%)	0.15 (6.26%)	0.149 (0.13, 0.168)
Intrinsic Progression Rate	2.67 (19.7%)	0.000236 (9.66%)	0.000237 (0.000185 , 0.000298)
Correlation between Baseline and Intrinsic Progression Rate	0.111 (40.6%)	0.001270 (28.8%)	0.001270 (0.000523, 0.00216)

Abbreviations: ICICLE = Incidence of Cognitive Impairment in Cohorts with Longitudinal Evaluation, LRRK2 = Leucine-rich repeat kinase 2, MDS-UPDRS23 = Movement Disorder Society-Unified Parkinson's Disease Rating Scale part 2 plus part 3 score, NE = not evaluated, PD = Parkinson's Disease, PPMI = Parkinson's Progression Marker Initiative, RSE = relative standard error

Notes: 95% CI = 95% confidence interval of parameter estimate based on bootstrap results. Baseline and the typical progression rate in MDS-UPDRS23 score can be estimated as follows: MDS-UPDRS23 Baseline = 29.5*[(1 + 0.134) if Other Cohort-PD]*[(1-0.0968) if Female]]*[(1+0.112) if Any PD medication]* ((AGE/63)0.376). Intrinsic Progression Rate = 0.0101*[(1 -0.235) if LRRK2 mutation]. The typical progression rate in MDS-UPDRS23 score was estimated to be dScore/dt=r*Score*(1-Score/max(Score)) = 29.5*0.0101*(1-[29.5/118]), which was approximately 0.22 point/per months.

NONMEM 7.3 code for final model

```
$PROBLEM    (Disease Progression Model)
$INPUT
ID                    ; Patient identification
STUDY                 ; Study identification
TIME                  ; Time (days)
DV                    ; Dependent variable
MAXUPDRS              ; Maximum score in the dataset
WGT                   ; Body Weigh (kg)
HT                    ; Height ()
AGE                   ; Age (years)
SEX                   ; Gender
RACE                  ; Race
LRRK2                 ; LRRK2 mutation
DSDUR                 ; Disease duration at enrollment (months)
CONMED                ; PD medication status
ARM                   ; Cohort in the study
MDV                   ; Missing dependent variable


$DATA nmData.csv IGNORE=@

$PRED


;;; BASECONMED-DEFINITION START
IF(CONMED.EQ.1) BASECONMED = 1  ; Most common
IF(CONMED.EQ.0) BASECONMED = ( 1 + THETA(8))
;;; BASECONMED-DEFINITION END

;;; BASEAGE-DEFINITION START
IF(AGE.EQ.-99) THEN
   BASEAGE = 1
ELSE
   BASEAGE = ((AGE/63)**THETA(7))
ENDIF
;;; BASEAGE-DEFINITION END

;;; BASESEX-DEFINITION START
IF(SEX.EQ.2) BASESEX = 1  ; Most common
IF(SEX.EQ.1) BASESEX = ( 1 + THETA(6))
;;; BASESEX-DEFINITION END

;;; BASEARM-DEFINITION START
IF(ARM.EQ.1) BASEARM = 1  ; Most common
IF(ARM.EQ.2) BASEARM = ( 1 + THETA(5))
;;; BASEARM-DEFINITION END


;;; BASE-RELATION START
BASECOV=BASEARM*BASESEX*BASEAGE*BASECONMED
```

```
;;; BASE-RELATION END


;;; SLOPLRRK2-DEFINITION START
IF(LRRK2.EQ.0) SLOPLRRK2 = 1  ; Most common
IF(LRRK2.EQ.1) SLOPLRRK2 = ( 1 + THETA(4))
;;; SLOPLRRK2-DEFINITION END

;;; SLOP-RELATION START
    SLOPCOV=SLOPLRRK2
;;; SLOP-RELATION END


; MODEL DEVELOPMENT

TVBASE=THETA(1)
TVBASE = BASECOV*TVBASE

TVSLOP=THETA(2)
TVSLOP = SLOPCOV*TVSLOP

BASE=TVBASE*EXP(ETA(1))   ; BASELINE
SLOP=TVSLOP +ETA(2)       ; SLOP


TAU      = THETA(3)         ;precision parameter of the beta
distribution (=ALPHA+BETA)


;STRUCTURAL MODEL TO DESCRIBE THE RATE OF DISEASE PROGRESSION
DEN1     = BASE
DEN2     = MAXUPDRS -BASE
DEN3     = EXP(-SLOP*TIME)
DENN     =DEN1 + DEN2*DEN3
MUR      = BASE/DENN

TVDEN1   = TVBASE
TVDEN2   = MAXUPDRS - TVBASE
TVDEN3   = EXP(-TVSLOP*TIME)
TVDENN   =TVDEN1 + TVDEN2*TVDEN3
TVMUR    = TVBASE/TVDENN

;DEFINING IPRED/PRED
F        = MUR               ;INDIVIDUAL (IPRED)
IPRED    = F

; BETA REGRESSION
ALPHA    = MUR*TAU
BETA     = (1-MUR)*TAU
```

```
X1       = ALPHA+BETA
X2       = ALPHA
X3       = BETA

;NEMES APPROXIMATION OF THE LN(GAMMA) DISTRIBUTION
LGAMMAX1 = X1*(LOG(X1)-1) + 0.5*(LOG(2*3.1415)-LOG(X1)) +
(5/4)*X1*LOG(1+(1/(15*X1**2)))
LGAMMAX2 = X2*(LOG(X2)-1) + 0.5*(LOG(2*3.1415)-LOG(X2)) +
(5/4)*X2*LOG(1+(1/(15*X2**2)))
LGAMMAX3 = X3*(LOG(X3)-1) + 0.5*(LOG(2*3.1415)-LOG(X3)) +
(5/4)*X3*LOG(1+(1/(15*X3**2)))

;LOG-LIKELIHOOD OF THE BETA DISTRIBUTION: LN of the probability
density function of the beta distribution
LLBETA   = LGAMMAX1 - LGAMMAX2 - LGAMMAX3 + (ALPHA-1)*LOG(DV) +
(BETA-1)*LOG(1-DV)

; DEFINING RESIDUALS AND Y

;PEARSON RESIDUALS
SOR      = (DV-IPRED)/SQRT(IPRED*(1-IPRED)/(1+TAU))

LLB2=-2.0*LLBETA
Y        = LLB2

;INITIAL ESTIMATES

$THETA
32.52   ; TVBL
0.0096  ; TVSL
33.95   ; TAU
-0.0118 ; SLOPLRRK2
0.0617  ; BASEARM
-0.084  ; BASESEX1
0.2642  ; BASEAGE
-0.0010 ; BASECONMED


$OMEGA BLOCK(2)
 0.0101         ;IIV_BL
 0.00097 0.0808 ;IIV_SL


$ESTIMATION MAXEVAL=9999 PRINT=1 METHOD=COND -2LL LAPLACIAN NOHABORT
FNLETA=0 NSIG=2
SIGL=7 SLOW MCETA=10 NONINFETA=1 FILE=psn.ext


$COVARIANCE MATRIX=R UNCONDITIONAL SIGL=9 SLOW PRINT=E
```