Mixture Modeling as a Data Imputation Method

S. Willavize[1], J. Fiedler-Kelly[1], W. Hanley[2], L. Mahnke[3], Y. Hang[2], J. Z. Peng[2], L. Phillips[1]

[1]Cognigen Corporation, Buffalo, NY; [2]Merck and Company, Inc., West Point, PA; [3]Formerly at Merck and Company, Inc.

ABSTRACT

Purpose. To demonstrate the use of mixture modeling in population PK analysis to predict drug concentrations for a subset of subjects with missing data for a key categorical covariate.

Methods. Drug X is mainly metabolized by the polymorphic enzyme CYP2D6 and consequently the extent of metabolism is genotype dependent. A population PK model was developed for Drug X with the goal of predicting steady-state concentrations for a subsequent PK/PD analysis of the clinical effects. A core PK model was developed using only those subjects with known CYP2D6 genotype. To estimate the steady-state exposure for subjects with missing genotype, a mixture model was developed. Subjects with known genotypes were assigned to Population 1 and their genotypes were not estimated. The mixture model assigned the subjects with unknown genotype to one of the 4 populations defined by the subjects with known genotype (that is, poor (PM), intermediate (IM), extensive (EM), and ultra-extensive metabolizer (UM)). Population PK modeling was performed using NONMEM® Version 6. Goodness of fit for the core and mixture models was assessed by examination of the following data and figures: the outcome (convergence and covariance); the agreement in scatterplots of population and individual predicted versus measured observations for each classification group; and the scatterplots of weighted residuals versus population and individual predicted observations.

Results. The core model fit the data well; the model converged with 3.3 significant digits and a successful covariance step (standard errors less than 33%). Goodness-of-fit plots show that the model is without bias. The mixture model minimized successfully with 2.5 significant digits. The R matrix for the mixture model was algorithmically singular, and standard errors were obtained from the S matrix. The plots of population predicted versus measured concentrations show that the model provided unbiased predictions for the subjects of unknown genotype classified by the mixture procedure. Scatterplots of weighted residuals versus population and individual predicted concentrations show no trends. Standard errors for the core parameters were less than 20%. This model was considered fit-for-purpose in predicting concentrations for subjects with missing genotype.

Conclusions. Mixture models have been used previously to explain interindividual heterogeneity as arising from the mixing of unknown subpopulations. This case study illustrates the novel and successful use of a mixture model to obtain predicted drug concentrations for a subset of subjects with a missing key categorical covariate.

METHODS

Study Data

- Approximately 100 Phase 1 subjects provided nearly 2500 concentrations and nearly 700 Phase 2 subjects provided over 4000 concentrations.
- Number of samples/subject: 7 to 33 in Phase 1 studies, 4 to 7 in Phase 2 studies.
- In the Phase 2 studies, subjects were recruited without regard to 2D6 genotype.
  - Of those classified: Poor Metabolizers (PM) (6%), Intermediate Metabolizers (IM) (36%), Extensive Metabolizers (EM) (55%), Ultra-Extensive Metabolizers (UM) (3%).
- Genotype data was unavailable for 100 Phase 2 subjects (~12% of all subjects)

Modeling Methods

- NONMEM (V6.2.0) was used to develop a structural model describing and quantifying the mean pharmacokinetic (PK) characteristics, as well as interindividual variability and residual variability in PK.
- Interindividual variability in parameters was modeled using an exponential error model.
- Residual variability was modeled as a constant coefficient of variation.
- The first-order conditional estimation (FOCE) with interaction method was used at all stages of the model development process.
- Assessment of goodness of fit for the core and mixture models was based on
  - Estimation outcome (convergence) and covariance step outcome,
  - Agreement in scatterplots of population predictions versus observed concentrations,
  - Agreement in scatterplots of individual predictions versus observed concentrations, and
  - Lack of trends in scatterplots of weighted residuals versus population and individual predicted concentration values.
- A core base structural model was fitted using data only from subjects with known genotype.
- Mixture modeling
  - Five populations were defined.
  - Subjects with known genotypes were assigned to Population 1, and their genotypes were not estimated.
  - The other 4 populations were used to assign the subjects with unknown genotypes to 1 of the 4 genotype classification groups (that is, PM, IM, EM, UM).
  - It was assumed that genotype data was missing completely at random (independently of any observed or unobserved variables).
  - Details on parameterization and coding can be seen in **Figure 1**.

Figure 1. Parameterization and Coding Details

```
EST=MIXEST

PM=0; IM=0; EM=0; UM=0

IF (MIXNUM.EQ.1.AND.GEN2.EQ.1) PM=1
IF (MIXNUM.EQ.1.AND.GEN2.EQ.2) IM=1
IF (MIXNUM.EQ.1.AND.GEN2.EQ.3) EM=1
IF (MIXNUM.EQ.1.AND.GEN2.EQ.4) UM=1

IF(PM.EQ.1.OR.MIXNUM.EQ.2) THEN
   PM=1
   TK=THETA(4)
ENDIF
IF(IM.EQ.1.OR.MIXNUM.EQ.3) THEN
   IM=1
   TK=THETA(9)
ENDIF
IF(EM.EQ.1.OR.MIXNUM.EQ.4) THEN
   EM=1
   TK=THETA(10)
ENDIF
IF(UM.EQ.1.OR.MIXNUM.EQ.5) THEN
   UM=1
   TK=THETA(11)
ENDIF
   K=TK*EXP(PM*ETA(2)+IM*ETA(3)+EM*ETA(4)+UM*ETA(5))

$MIX

NSPOP=5 ;Known + unknown(PMs,IMs,EMs,and UMs)

;calculate probability for population with known genotype vs. unknown
PKN=0.835962
PUKN=1.0-PKN

;set up indicator for known vs unknown subjects
KNW=1  ;  IF(GEN2.GT.4.5)KNW=0

;P(1) must equal PKN  ;1=P(1)+P(2)+P(3)+P(4)+P(5)

A=EXP(THETA(12))
B=EXP(THETA(13))
C=EXP(THETA(14))
D=1+A+B+C

;pop1 represents 'knowns' only
P(1)=KNW*PKN + (1-KNW)*PUKN*0
;pop2 represents 'unknown'PMs
P(2)=KNW*0 + (1-KNW)*PUKN*A/D
;pop3 represents 'unknown' IMs
P(3)=KNW*0 + (1-KNW)*PUKN*B/D
;pop4 represents 'unknown' EMs
P(4)=KNW*0 + (1-KNW)*PUKN*C/D
;pop5 represents 'unknown' UMs
P(5)=KNW*0 + (1-KNW)*PUKN*(1/D)
```

- The mixture aspects of the model were added to a model of the same form as the core base model and all parameters were re-estimated.
- Using the SAME option in NONMEM, the interindividual random effect terms for subjects of various genotypes were assumed to arise from the same distribution of interindividual variability.[1]

RESULTS

- A 2-compartment model with first-order absorption was found to best fit the classified data. Goodness-of-fit plots for the model without genotype are shown in **Figure 2**.
- Elimination rate was dependent on genotype.
- Parameter estimates for the "Core Base Model" fitted to the subjects with known genotype are listed in **Table 1** and goodness-of-fit plots are shown in **Figure 3**.
- Parameter estimates for the "Base Mixture Model" fitted to all data are listed in **Table 1** and goodness-of-fit plots for genotype-unknown subjects are shown in **Figure 4**.
- Parameter estimates for "Core Base Model" and "Base Mixture Model" were similar, indicating that the unknown genotypes were classified appropriately.
- The proportion of unknowns belonging to each genotype was not estimated with good precision. However, this was not required to adequately predict the concentration profiles for most subjects.

Figure 2. Goodness-of-Fit Plots for Model Without Genotype

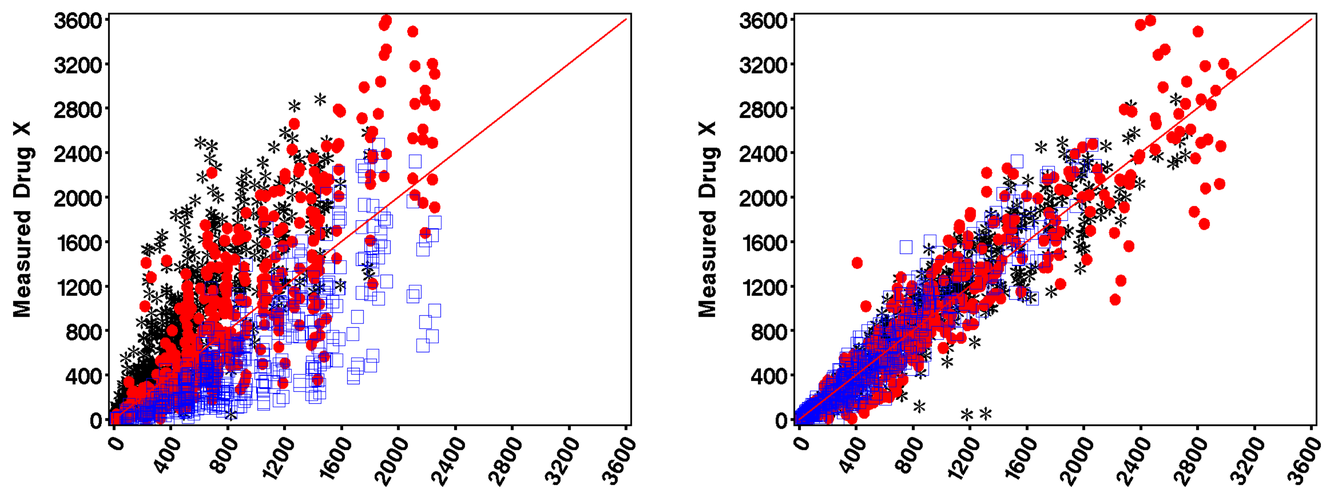

Figure 3. Goodness-of-Fit Plots for Core Base Model (Genotype Known Subjects)

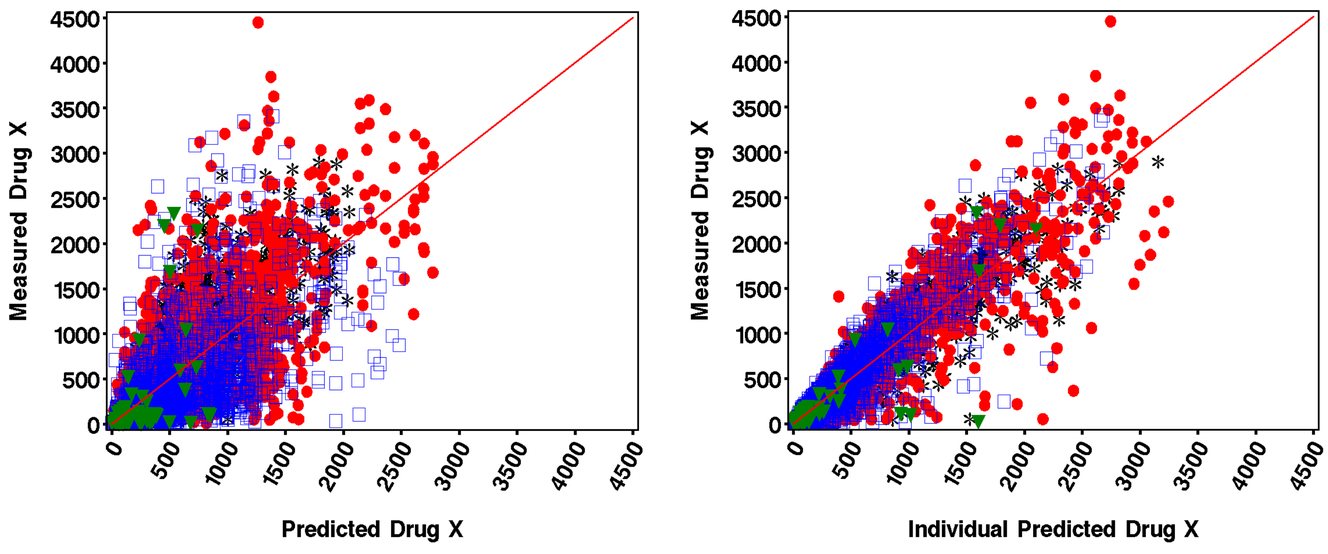

Figure 4. Goodness-of-Fit Plots for Base Mixture Model (Genotype Unknown Subjects)

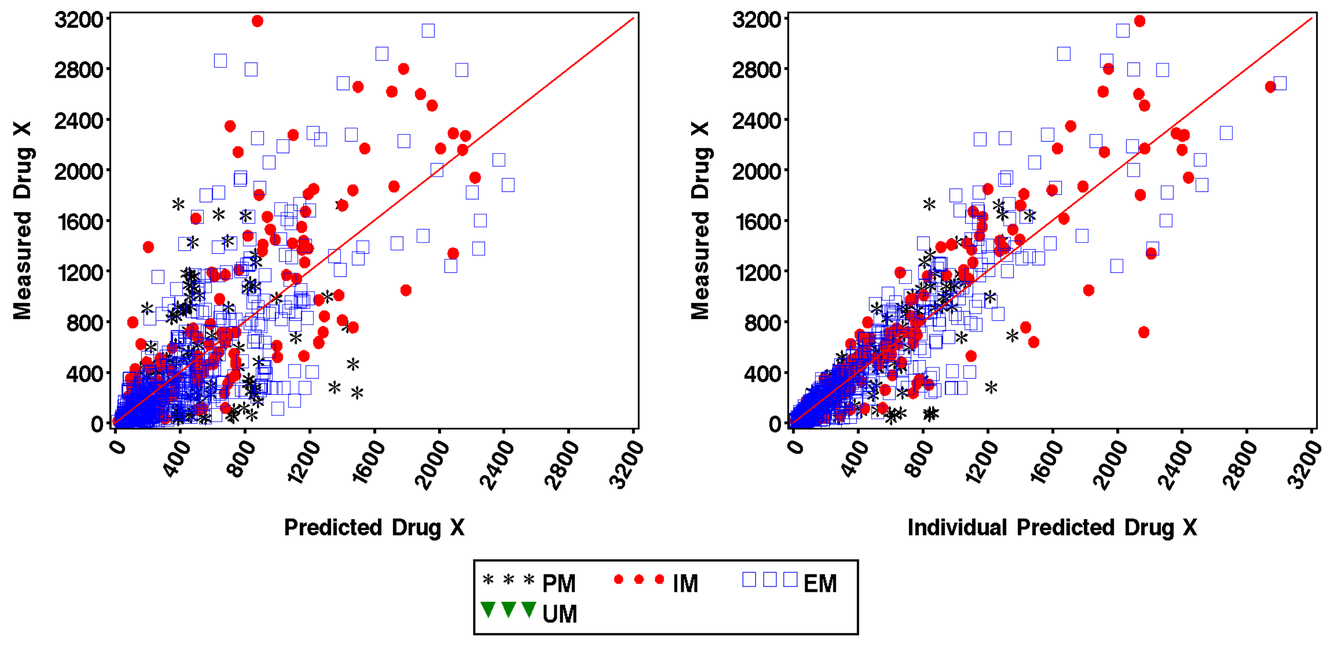

Table 1. Key Parameter Estimates and Standard Errors for the Core Base Pharmacokinetic Model and the Base Mixture Pharmacokinetic Model

	Core Base				Base Mixture			
	Final Parameter Estimate		Magnitude of Interindividual Variability (%CV)		Final Parameter Estimate		Magnitude of Interindividual Variability (%CV)	
Parameter	Population Mean	%SEM[1]	Final Estimate	%SEM[1]	Population Mean	%SEM[1]	Final Estimate	%SEM[1]
Vc: central volume (L)	125	FIXED	66.56	12.7	125	FIXED	63.56	7.1
ka: absorption rate constant (/h)	0.172	2.8	NE	NA	0.167	1.0	NE	NA
K (/h)-PM	0.151	14.7	70.36	11.3	0.150	19.5	68.77	9.9
K (/h)-IM	0.374	8.2			0.392	6.7		
K (/h)-EM	0.706	10.5			0.746	4.9		
K (/h)-UM	1.22	33.0			1.32	19.0		
K12: distributional rate constant 1 (/h)	0.00498	FIXED	NE	NA	0.00498	FIXED	NE	NA
K21: distributional rate constant 2 (/h)	0.0267	FIXED	NE	NA	0.0267	FIXED	NE	NA
ALAG1: absorption lag time (h)	0.343	5.6	NE	NA	0.348	1.4	NE	NA
ln(A): A/D = proportion of unclassifieds that are PM					13.7	3372.3	NE	NA
ln(B): B/D = proportion of unclassifieds that are IM					14.2	3246.5	NE	NA
ln(C): C/D = proportion of unclassifieds that are EM					14.7	3142.9	NE	NA
RV (%CV)	39.62	4.3	NA	NA	40.25	1.2	NA	NA
MVOF	66870.177				77209.074			

Abbreviations: ALAG1, absorption time lag; EM, extensive metabolizer; IM, intermediate metabolizer; K, first-order elimination rate; K12, distribution rate constant from central to peripheral compartment; K21, distribution rate constant from peripheral to central compartment; ka, absorption rate constant; MVOF, minimum value of the objective function; NA, not applicable; NE, not estimated; %CV, percent coefficient of variation; %SEM, percent standard error of the mean; PM, poor metabolizer; RV, residual variability; UM, ultra-extensive metabolizer; Vc, central volume of distribution.
[1]The standard errors were estimated by NONMEM using the S matrix.

CONCLUSIONS

- Genotypic metabolizer status is an important determinant of individual subject plasma concentrations for Drug X.
- Using the subjects with known genotype to anchor the PK parameter estimates provided a means of classifying subjects with unknown genotype and of estimating their PK parameters.
- As indicated by the similar parameter estimates and the concordance of observed and predicted concentrations, the classification of the unknown subjects was relatively unbiased and reasonable estimates of the concentrations for these subjects were obtained.
- Parameters associated with the proportion of unknown subjects in each subpopulation could not be precisely estimated. This result may be a reflection of the small number of unknown subjects (~100) and the rareness of some of the genotypes.
- Mixture models have been used previously to explain interindividual heterogeneity as arising from the mixing of unknown subpopulations.[2,3] This case study illustrates the novel and successful use of a mixture model to obtain predicted drug concentrations for a subset of subjects with a missing key categorical covariate.

REFERENCES

1. Frame B. Mixture modeling with NONMEM V. In: Ette EI and Williams PJ, eds. *Pharmacometrics: The Science of Quantitative Pharmacology*. Hoboken, NJ: John Wiley & Sons, Inc.; 2007:723-757.
2. Vermeulen A, Piotrovsky V, Ludwig EA. Population pharmacokinetics of risperidone and 9-hydroxyrisperidone in patients with acute episodes associated with bipolar I disorder. *J Pharmacokinet Pharmacodyn*. 2007;34:183-206.
3. Feng Y, Pollock BG, Coley K, et al. Population analysis for risperidone using highly sparse sampling measurements from the CATIE study. *British J Clin Pharmacol*. 2008;66(5):629-639.

Willavize S, Fiedler-Kelly J, Hanley W, et al. Mixture Modeling as a Data Imputation Method. Poster presented at: American Conference on Pharmacometrics (ACoP); April 3 - 6, 2011; San Diego, CA.

For additional information, please contact
Susan Willavize, PhD
Assistant Director, PK/PD
Cognigen Corporation
395 South Youngs Road, Buffalo, NY 14221
(716) 633-3463, ext. 252 or susan.willavize@cognigencorp.com