Illustration of the ISOP best practice for covariates in Pharmacometrics

Case study using the Theophylline data

Contents

1 Introduction

This document illustrates to some extent the different covariate selection methods presented in the article. Real longitudinal pharmacodynamic data were used: peak expiratory flow rate (PEFR) measurements gathered during a clinical trial with theophylline administration. The document is articulated in sections presenting the data (section 2), the models (section 3), the results of the different covariates selection methods (section 4), some examples to diagnose a covariate effect (section 5), and an example of how to present a covariate model (section 6). The models section presents in details only the published model with covariates that was used as a starting point for this work [1], the derived base model without covariates that was used as a starting point for most the of the covariates selection methods, and the full fixed effect model (FFEM). However, all the code is reported in the appendices of the document: NONMEM models code in the Appendix A, R code in the Appendix B.

The default parameterization of the covariates models can differ between the available implementations of the selection methods.

2 Theophylline data

2.1 Data description

The theophylline data were provided by Nick Holford [1,2]. Details and experimental design are reported elsewhere [2], a summary is provided below. The R code used to prepare the data for the analysis is available in Appendix B.1. The rows with missing information were removed from the data set, the extent of the missing information is provided in the initial publication [1]. Only four covariates (two categorical and two continuous) were retained for this illustration exercise.

Patients with acute airways obstruction eligible to intravenous theophylline treatment were randomized to a double-blind trial of target concentrations of either 10 or 20 mg/L. Initial doses and subsequent adjustment were determined by each patient's physician using scaled reports of actual theophylline concentrations to maintain blinding yet mimic actual clinical practice. PEFR and theophylline concentrations were obtained several times during the first 24 hours and at least daily until discharge.

A numerical summary is provided in Table S1. The average number of sampling times per subjects is 4. The R code used to compute the numerical summary is available in Appendix B.2, more details about the data file are available in Appendix B.3: a description of the data file in Table S9, and an overview of the first ten rows in Table S10.

Table S1: Numerical description of the Theophylline data set

Variables	Categories	Metrics*
Dependent variables (526 observations)		
Theophylline concentrations (mg/L)		13 (8, 18)
Peak expiratory force rate (PEFR, L/min)		220 (160, 300)
Continuous covariates (132 subjects)		
Age (years)		34.5 (21.0, 55.0)
Weight (kg)		66.0 (56.0, 76.0)
Categorical covariates (132 subjects)		
Sex	0: Female	82.0 (62.1%)
	1: Male	50.0 (37.9%)
Diagnosis	1: Asthma	108.0 (81.8%)
	2: COPD	22.0 (16.7%)
	3: COPD and asthma	2.0 (1.5%)

* Median (inter-quartile range) for continuous variables; n (%) for categorical variables.

2.2 Graphical exploration

2.2.1 Dependent variables: spaghettis plots

Spaghettis plots of the observed pharmacodynamic data (i.e. PEFR) vs time are presented in Figure S1, and vs concentration in Figure S2. The R code used for Figure S2 is available in Appendix B.4.

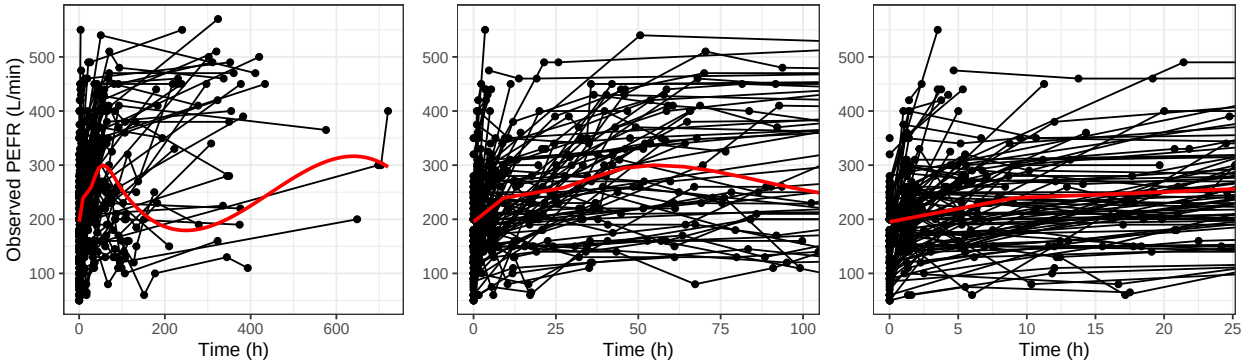

Figure S1: Observed PEFR vs time. Dots represents the observations and the observations of each individual are connected by a line. The red line is a smooth. The data are displayed using a normal scale, with a zoom on the first 100h (medium panel), and a zoom on the first 25h (right panel). PEFR: peak expiratory flow rate.

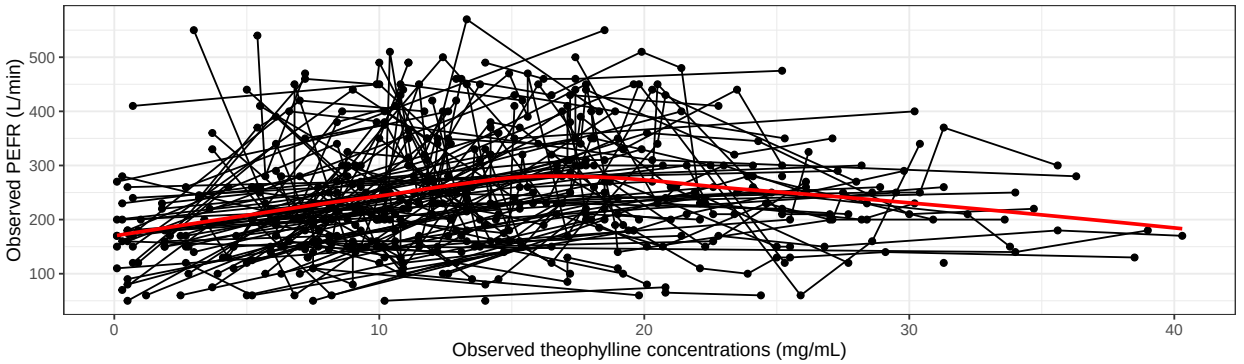

Figure S2: Observed PEFR vs observed theophylline serum concentrations. Dots represents the observations and the observations of each individual are connected by a line. The red line is a smooth. The data are displayed using a normal scale. PEFR: peak expiratory flow rate.

2.2.2 Covariates: scatter plot matrix

The scatter plot matrix (Figure S3) displays the relationships between each pair of covariates. The R code used to generate the scatter plot matrix is available in Appendix B.5.

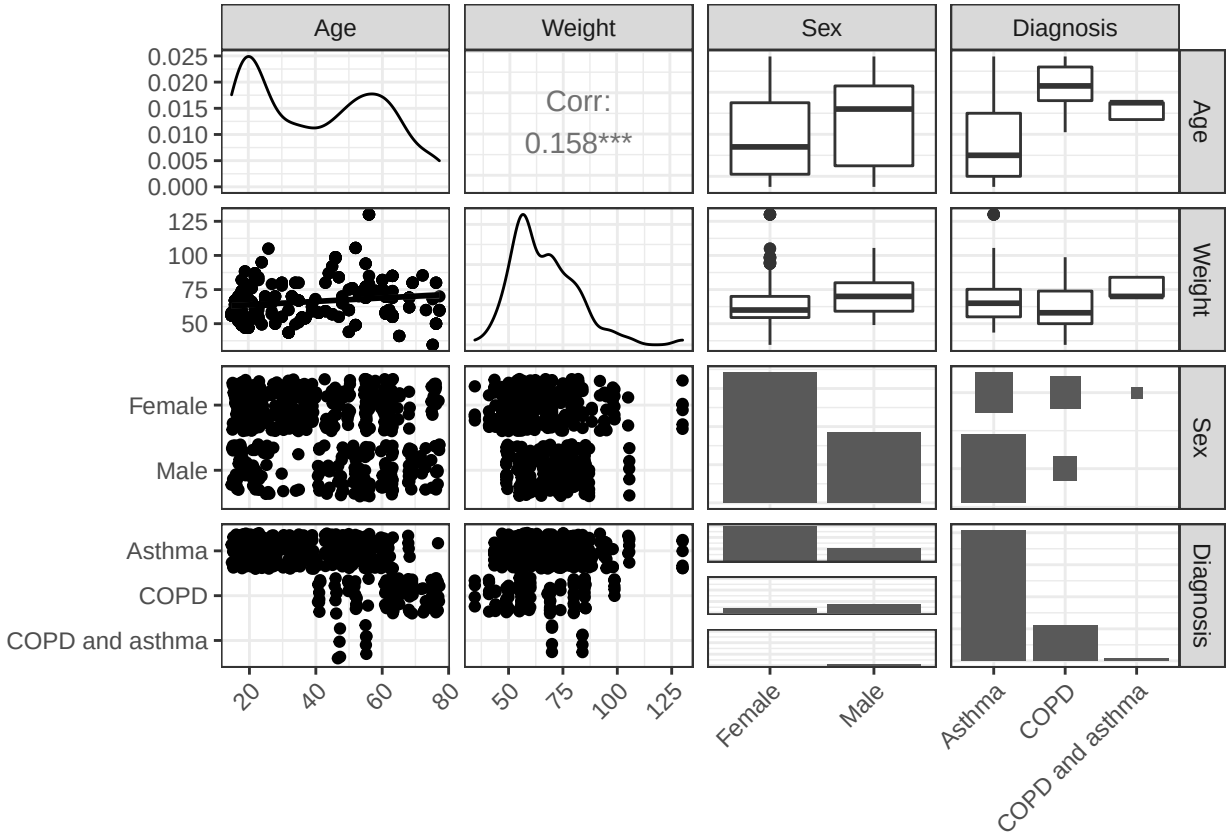

Figure S3: Visual exploration of the relationships between each pair of the available covariates. The diagonal plots represent the distribution of each covariate.

3 Models

A pharmacodynamic model of PEFR was developed based on a previous publication [1] without (run11) or with covariates (run20). From the model without covariates different covariates models were built according to the different covariates selection methods to provide illustrative examples. The different models obtained are presented in the Table S4, together with the final number of parameters.

3.1 Model from the publication

The parametric model previously published in [1] was replicated as run20 from Table S4. The parameters values are presented in Table S2, the NONMEM code is provided in Appendix A.1. The run20 presented here differs from the published model as the dataset was reduced by removing the rows with missing data.

As described in [1], a broncho-constriction factor (BCF) is eliminated by a first-order process characterized by a half-life $T50$ and a value at baseline BCF_0 (Eqn. (1)). $PEFR(t)$, the broncho-constriction due to BCF hence time, is related to the BCF concentration according to an inhibitory pharmacodynamic model described in Eqn (2), where Normal is the PEFR when BCF is zero, and $C50_{BCF}$ is the concentration producing 50% reduction in PEFR from Normal. Given that Base is the PEFR at $t = 0$, the concentration of BCF at this time (BCF_0) is given by Eqn. (3). The theophylline effect on PEFR is restricted to subnormal PEFR and expressed according to Eqn. (4), where $C50_{T\,heo}$ is the theophylline concentration C producing 50% of the potentially recoverable PEFR and $Hill$ is the exponent controlling the steepness of the curve. The combined influences of BCF and theophylline on PEFR can be predicted by a function of time and theophylline concentration according ot Eqn. (5), where ε_{theo} describes the efficacy of the theophylline relative to the time effect mediated by BCF.

All the covariates were added linearly to the Normal typical value. The categorical diagnosis covariate was included as a continuous covariate in the published model, using the numerical value assigned to each category. It was included as a categorical covariate whenever possible throughout the examples presented in this document.

$$BCF(t) = BCF_0 \cdot e^{-ln(2)/T50 \cdot t} \tag{1}$$

$$PEFR(t) = Normal \cdot \left[1 - \frac{BCF(t)}{BCF(t) + C50_{BCF}}\right] \tag{2}$$

$$BCF_0 = C50_{BCF} \cdot (Normal/Base - 1) \tag{3}$$

$$PEFR(c,t) = [Normal - PEFR(t)] \cdot \frac{C^{Hill}}{C^{Hill} + C50_{Theo}^{Hill}} \tag{4}$$

$$PEFR = PEFR(t) + \varepsilon_{theo} \cdot PEFR(c,t) \tag{5}$$

Table S2: Parameters table for the covariate model from the publication (run20).

Parameter	Description	Unit	Value	RSE*	Shrinkage (%)
Base	PEFR at baseline	L/min	136	0.0642	-
Normal	PEFR without BCF†	L/min	481	0.0621	-
$T50$	BCF half-life	h	19.6	0.263	-
$C50$	Concentration producing 50% of the recoverable PEFR	mg/L	8.48	0.268	-
ε_{theo}	Theophylline efficacy relative to BCF effect	-	0.373	0.197	-
Hill exponent	Steepness curve exponent	-	2.59	0.3	-
Diagnosis on Normal	Typical value fraction per category	-	0.202	0.174	-
Female on Normal	Typical value fraction for females	-	-0.258	0.193	-
Years on Normal	Scale factor	-	-0.012	0.12	-
IIV Base		-	0.405	0.256	18.9
IIV Normal		-	0.164	0.364	43.4
IIV $T50$		-	1.55	0.281	29.6
IIV $C50$		-	0.664	0.609	48.2
Proportional RUV		-	0.159	0.123	25.8

IIV and RUV values are reported as standard deviations for the diagonal elements.
RSE for IIV and RUV are reported on the approximate standard deviation scale (standard error/variance estimate)/2. The shrinkage correponds to the standard deviation shrinkage.
PEFR: peak expiratory flow rate, BCF: broncho-constriction factor, RUV: residual unexplained variability, IIV: inter-individual variability, RSE: relative standard error. SIR: sampling intensive resampling.
* RSE obtained with SIR.
† For a typical individual: male, 40kg with asthma.

3.2 Base model without covariates

The covariates effects were removed from the published model (run20) resulting in a base model without covariates that will be used as starting point for most of the covariates selection methods (run11 from Table S4). The parameters values are presented in Table S3, a graphical illustration is provided in Figure S4, and the NONMEM code is provided in Appendix A.2. The RSE values come from the variance-covariance matrix.

Table S3: Parameters table for the pharmacodynamic model for PEFR without covariates (run11).

Parameter	Description	Unit	Value	RSE*	Shrinkage (%)
Base	PEFR at baseline	L/min	143	0.0534	-
Normal	PEFR without BCF	L/min	332	0.0546	-
$T50$	BCF half-life	h	9.54	0.37	-
$C50$	Concentration producing 50% of the recoverable PEFR	mg/L	10.3	0.215	-
ε_{theo}	Theophylline efficacy relative to BCF effect	-	0.414	0.207	-
Hill exponent	Steepness curve exponent	-	3.25	0.25	-
IIV Base		-	0.404	0.229	20.3
IIV Normal		-	0.389	0.221	16.6
IIV $T50$		-	1.22	0.413	34
IIV $C50$		-	0.68	0.638	48.6
Proportional RUV		-	0.145	0.14	30.4

IIV and RUV values are reported as standard deviations for the diagonal elements.
RSE for IIV and RUV are reported on the approximate standard deviation scale (standard error/variance estimate)/2. The shrinkage correponds to the standard deviation shrinkage.
PEFR: peak expiratory flow rate, BCF: broncho-constriction factor, RUV: residual unexplained variability, IIV: inter-individual variability, RSE: relative standard error.
* RSE obtained with SIR.

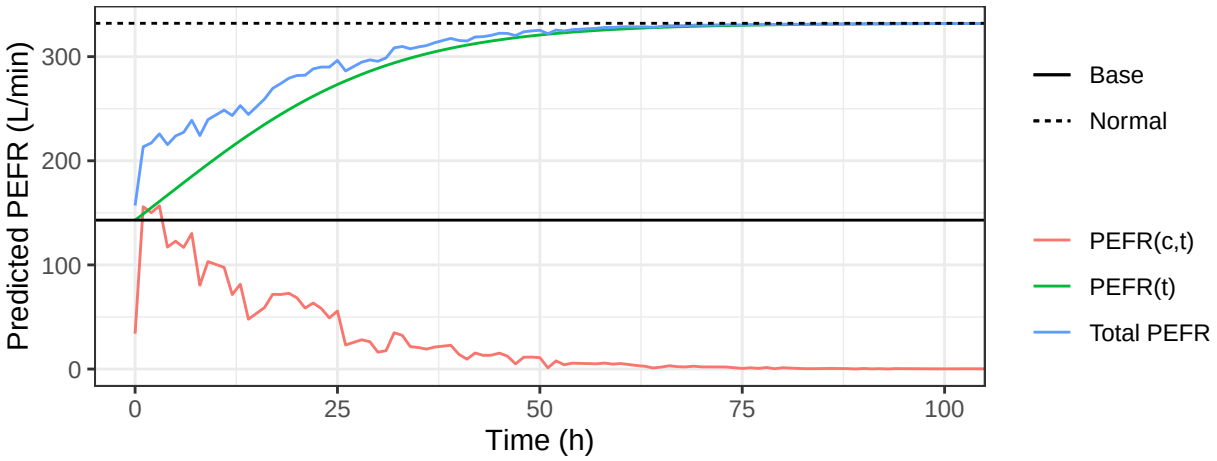

Figure S4: Illustration of the publication model without covariates (run11) using typical values. Base: PEFR at baseline, Normal: PEFR without broncho-constriction factor, PEFR: peak expiratory flow rate, PEFR(t): time dependent PEFR, PEFR(c,t): theophylline concentration and time dependent PEFR.

Table S4: Details of the different models used for the illustrations.

Description	Covariate scope	Run nb.
Base model without covariate		11
Covariate model from the publication		20
FFEM model	Saturated† model on Normal and Base parameters	22
Covariate model based on WAM	Starting from FFEM	24
Covariate model based on SCM	All covariates on all parameters	25
Covariate model based on LASSO	All covariates on Normal	26
FREM final model	All covariates on all parameters	28
GAM analysis	All covariates on Normal	-*

Run nb.: model identifier.
* The covariate-parameters relationships identified were not implemented in a new NONMEM model.
† All the covariates on all the parameters.

3.3 FFEM

A saturated FFEM on two parameters (Base and Normal) was created (run22 from Table S4). The parameters values are presented in Table S5, a graphical illustration of the parameter-covariates relationship is provided in Figure S5, and the NONMEM code is provided in Appendix A.3.

Table S5: Parameters table for the FFEM (run22).

Parameter	Description	Unit	Value	RSE	Shrinkage (%)
Base	PEFR at baseline	L/min	165	0.0742	-
Normal	PEFR without BCF†	L/min	421	0.139	-
$T50$	BCF half-life	h	12.3	0.452	-
$C50$	Concentration producing 50% of the recoverable PEFR	mg/L	10	0.197	-
ε_{theo}	Theophylline efficacy relative to BCF effect	-	0.36	0.124	-
Hill exponent	Steepness curve exponent	-	4.54	0.612	-
Diagnosis on Base	Typical value fraction per category	-	0.0425	2.48	-
Sex on Base	Typical value fraction for females	-	-0.169	0.429	-
Age on Base	Scale factor	-	-0.00534	0.394	-
Weight on Base	Scale factor	-	0.00598	0.606	-
Diagnosis on Normal	Typical value fraction per category	-	0.211	0.294	-
Sex on Normal	Typical value fraction for females	-	-0.193	0.496	-
Age on Normal	Scale factor	-	-0.0111	0.153	-
Weight on Normal	Scale factor	-	0.00643	0.534	-
IIV Base		-	0.375	0.105	17.4
IIV Normal		-	0.285	0.139	25.6
IIV $T50$		-	1.03	0.341	34.5
IIV $C50$		-	0.565	0.158	52.6
Proportional RUV		-	0.153	0.069	27.7

IIV and RUV values are reported as standard deviations for the diagonal elements.
RSE for IIV and RUV are reported on the approximate standard deviation scale (standard error/variance estimate)/2. The shrinkage correponds to the standard deviation shrinkage.
PEFR: peak expiratory flow rate, BCF: broncho-constriction factor, RUV: residual unexplained variability, IIV: inter-individual variability, RSE: relative standard error.

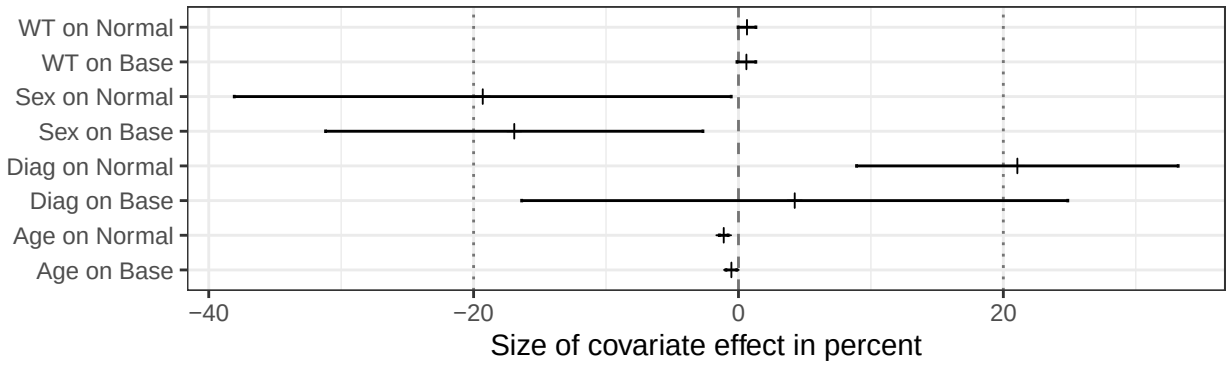

Figure S5: Forest plot representing the covariates fixed effects estimates and uncertainty expressed as size effect in percent for the FFEM (run22).

3.4 Functional forms used for the covariate models

The Figure S6 illustrated the different functional forms for continuous relationships, the corresponding equa-tions are presented in Eqns. 6 to 10. The different models were fitted to the Normal vs age data, the individual Normal values were obtained from the model without covariates (run11).

$$\text{Linear model} \quad y(x) = ax + b \tag{6}$$

$$\text{Exponential model} \quad y(x) = ae^{bx} \tag{7}$$

$$\text{Power model} \quad y(x) = ax^{b} \tag{8}$$

$$\text{Sigmoid model} \quad y(x) = ax/(b + x) \tag{9}$$

$$\text{Piece-wise model} \quad y(x) = \begin{cases} a(x - \bar{x}) + b & \text{if } x \leq \bar{x} \\ c(x - \bar{x}) + b & \text{if } x > \bar{x} \end{cases} \tag{10}$$

Where $\bar{x}$ is the median of the covariate.

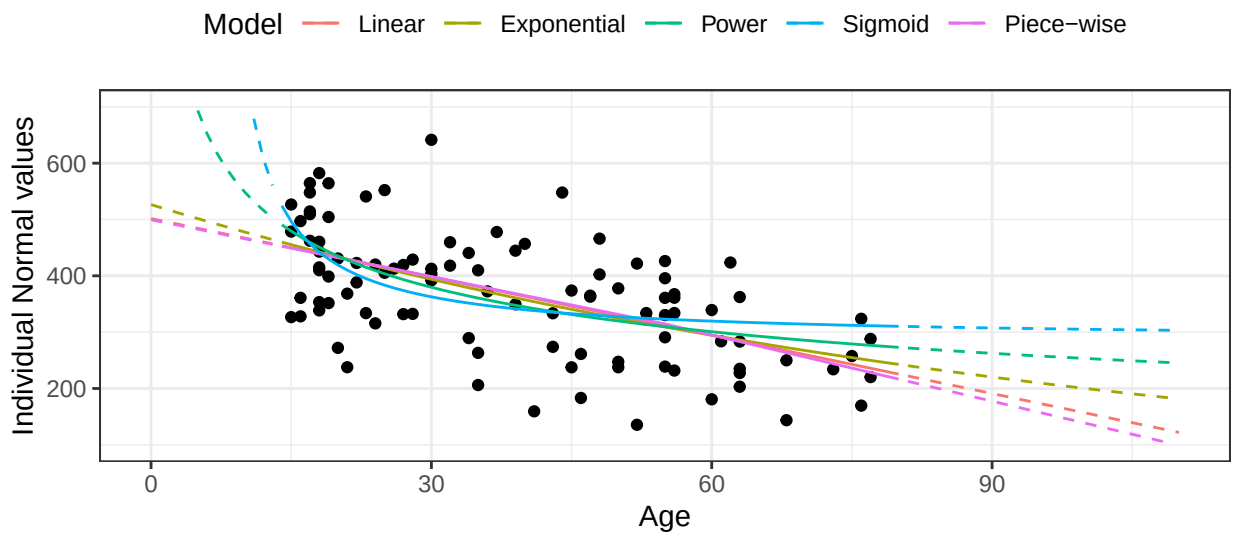

Figure S6: Example of covariate models for continuous covariates fitted to the Normal vs age data. The individual normal values are the EBEs from the model without covariate (run11). The dashed portions indicate extrapolation outside the observed covariate range. EBEs: empirical bayes estimates.

4 Model based covariates screening

4.1 Covariates vs individual random effects

The individual random effects can be reduced by the inclusion of predictors (i.e. covariates). Plotting these individual random effects (EBEs) against the available covariates can help to evidence parameters-covariates relationships and hence suggest which covariate to include on which parameter. Such plot is presented in the left panel of Figure S7 (R code available in Appendix B.7) for the theophylline data, using the model without covariates (run11) presented in Section 3.2.

However, in the presence of an important η shrinkage ($> 20-30$ according to [3]) such plots can indicate false relationships or hide true relationships when used for covariate screening. An alternative is to use random samples from the individual conditional distribution to generate η samples instead of the EBEs, as suggested in [4]. The individual mean and the individual variance for the η samples are available in the NONMEM generated .phi file. The right panel of Figure S7 illustrates the relationships between the covariates and the η samples (R code available in Appendix B.8).

The comparison of the EBEs vs covariates plot and the η samples vs covariates plot for the base model (run11) show that the distribution of the samples is slightly widened. Visually there was no obvious differences in the slope of the linear regression (see Figure S7).

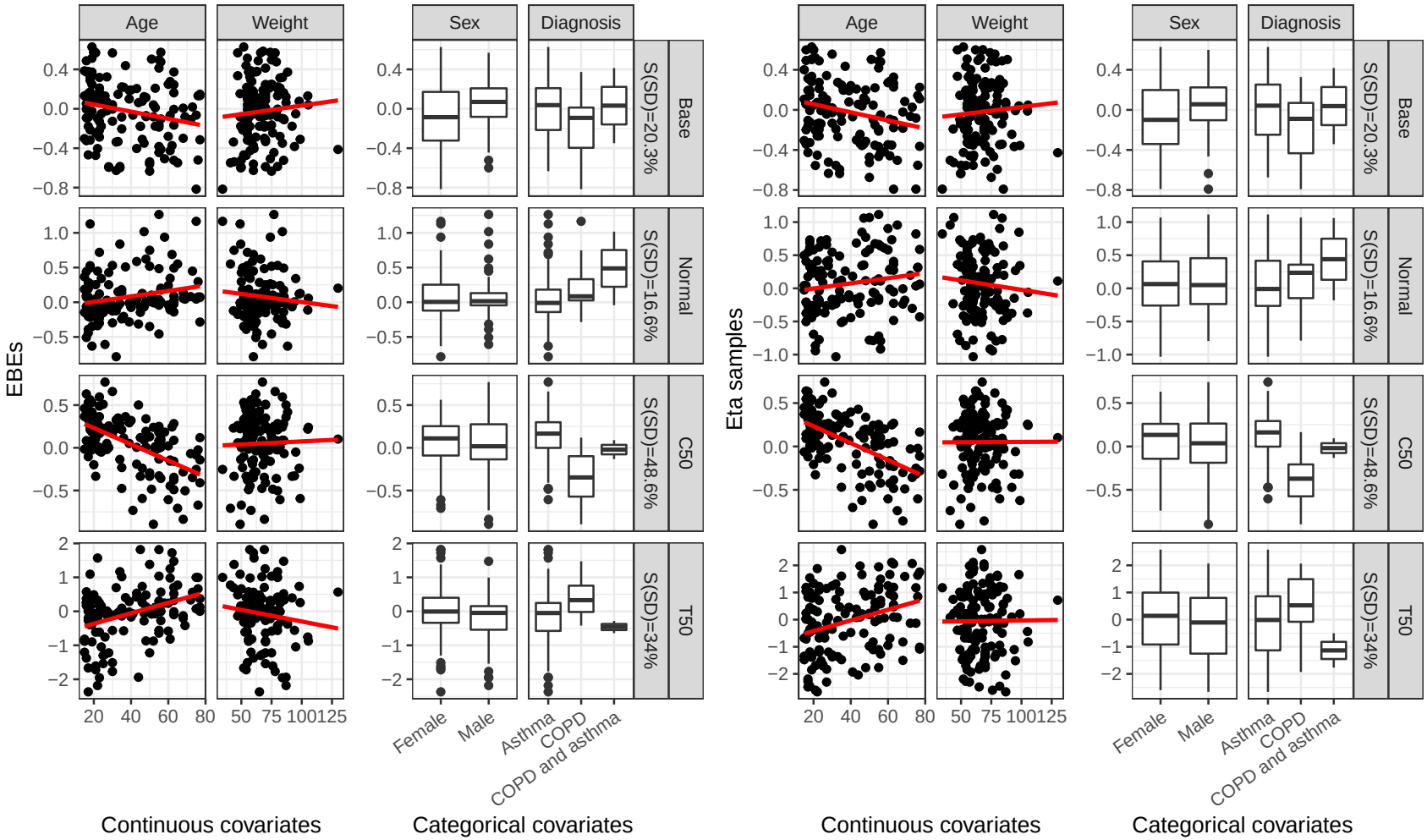

Figure S7: EBEs (left panel) and etas samples (right panel) vs covariates for the base model without covariates (run11). Continuous covariates are displayed using scatter plots on the left panel, the red line is a linear regression. Categorical covariates are displayed using boxplots. S(SD) is the shrinkage computed on the standard deviation. Base, Normal, C50 and T50 are defined in Table S3.

4.2 Generalized additive model (GAM)

Generalized additive model (GAM) is a stepwise search for covariates on a single parameter [5]. One implementation is available in Xpose4 [6,7], in the function `xpose.gam()`, using a stepwise approach. An illustration example is provided below, using the residuals from the base model without covariates (run11 from Table S4).

```
xpdb11 <-  xpose4::xpose.data(runno = 11, # Read base model output with xpose4
                      directory = "models")

xpgam <- xpose4::xpose.gam( # Run GAM
  object = xpdb11,
  parnam = "NORMAL",
  covnams = c("AGE", "WT", "SEX", "DIAG")
)
```

```
## Start:  NORMAL ~ 1; AIC= 1621.422
## Step:1 NORMAL ~ AGE ; AIC= 1575.53
## Step:2 NORMAL ~ AGE + DIAG ; AIC= 1571.575
## Step:3 NORMAL ~ AGE + SEX + DIAG ; AIC= 1566.629
## Step:4 NORMAL ~ ns(AGE, df = 2) + SEX + DIAG ; AIC= 1566.61
```

```
## Call:
## gam(formula = NORMAL ~ ns(AGE, df = 2) + SEX + DIAG, data = gamdata,
##     trace = FALSE)
##
## Degrees of Freedom: 131 total; 126 Residual
## Residual Deviance: 991643.3
```

The command `xpose.gam()` screened the following covariates: age and weight (continuous), and sex and diagnosis (categorical) on the Normal parameter. The covariates were added in a stepwise manner using the AIC criteria. The `ns` option is a non-linear relationship characterized by the number of degrees of freedom indicated. In this case, AGE is included first, with a linear relationship first, then DIAG and SEX successively, then two degrees of freedom are allowed for AGE. A more detailed summary can be accessed with the function `xp.summary()` as presented below, including parameters estimates.

```
xpose4::xp.summary(xpgam)
```

```
##
## SUMMARY
## Call: gam(formula = NORMAL ~ ns(AGE, df = 2) + SEX + DIAG, data = gamdata,
##     trace = FALSE)
## Deviance Residuals:
##     Min      1Q  Median      3Q     Max
## -174.33  -64.82   10.59   51.91  280.08
##
## (Dispersion Parameter for gaussian family taken to be 7870.185)
##
##     Null Deviance: 1620295 on 131 degrees of freedom
## Residual Deviance: 991643.3 on 126 degrees of freedom
## AIC: 1566.61
##
## Number of Local Scoring Iterations: 2
##
## Anova for Parametric Effects
##                  Df Sum Sq Mean Sq F value    Pr(>F)
## ns(AGE, df = 2)   2 493384  246692 31.3451 8.936e-12 ***
## SEX               1  29667   29667  3.7696  0.054425 .
## DIAG              2 105601   52800  6.7089  0.001703 **
```

```
## Residuals       126 991643    7870
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## PATH TO FINAL MODEL
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## NORMAL ~ 1
##
## Final Model:
## NORMAL ~ ns(AGE, df = 2) + SEX + DIAG
##
## Scale: 12368.67
##
##   From              To Df  Deviance Resid. Df Resid. Dev      AIC
## 1              <start>                   131  1620295.4 1621.422
## 2                  AGE -1 -493034.2       130  1127261.2 1575.530
## 3                 DIAG -2  -65921.9       128  1061339.3 1571.575
## 4                  SEX -1  -54412.6       127  1006926.7 1566.629
## 5  AGE ns(AGE, df = 2) -1  -15283.4       126   991643.3 1566.610
##
## COEFFICIENTS
##      (Intercept) ns(AGE, df = 2)1 ns(AGE, df = 2)2             SEX1
##         75.16201       -187.79854        -80.36751         43.67164
##            DIAG2             DIAG3
##       -106.94693        -48.29317
##
## PRERUN RESULTS
## Dispersion:
##
## DATA
## Subset expression:
## Only first value of covariate considered
## for each individual: TRUE
## Covariates normalized to median: TRUE
```

An illustration of the stepwise search is provided in Figure S8 where all the residuals models tried are sorted according to their Akaike information criteria (AIC). An illustration of the correlation between the selected covariates and the residuals of the base model (run11) for the screened parameter is provided in Figure S9.
Numbers correspond to the unique subject identifier and allow to spot outliers. R code used to generate the two figures is available in Appendix B.9.

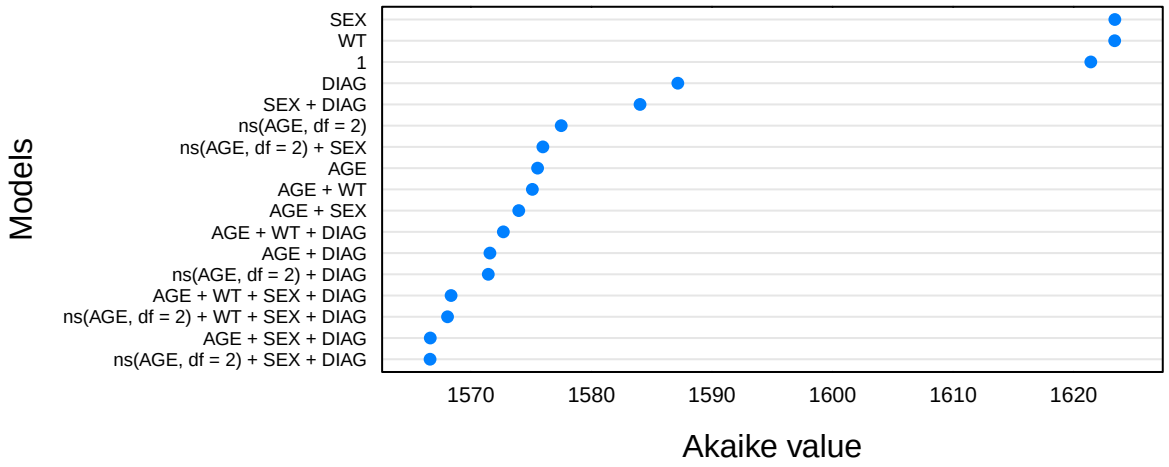

Figure S8: Akaike information criteria values for each model of the stepwise search on Normal for the base model (run11). The model named 1 is the model without covariates.

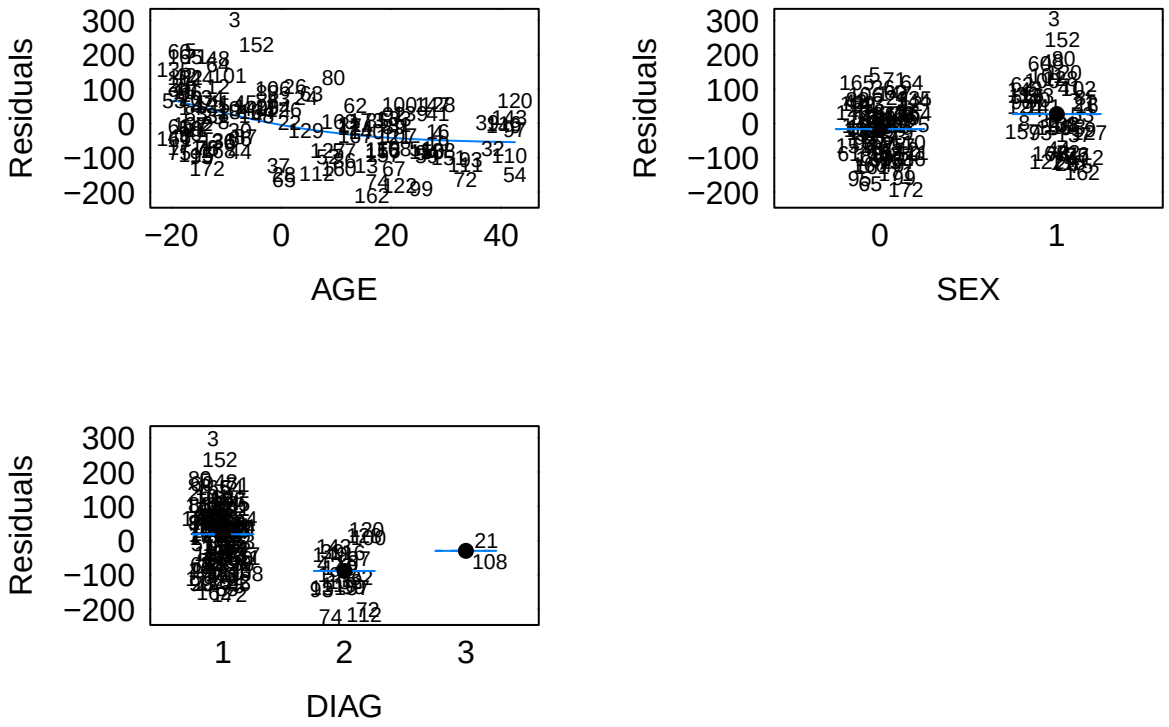

Figure S9: Residuals of the Normal parameter from the base model (run11) vs. covariates for the selected covariates by GAM. The numbers correspond to the unique subject identifier.

4.3 Wald's approximation method (WAM)

The starting point of WAM is the FFEM which contains all the considered parameters-covariates relationships modeled as fixed effects. The covariates effects are illustrated in the Figure S5. As mentioned before, the FFEM was restricted to a saturated covariate model on Base and Normal. The NONMEM code is provided in Appendix A.3.

The uncertainty of the covariates fixed effects is obtained from the NONMEM covariance step. Using these two information, covariates fixed effect estimates and uncertainty, it is possible to approximate the difference in OFV between each reduced submodel and the FFEM. The Wald's approximation method is therefore an approximation of the likelihood ratio test (LRT). The model selection could then be done according to the aLRT (approximated LRT), picking the model with the lowest aLRT value and the fewer parameters. However because the aLRT is an approximation, the Schwarz's Bayesian criterion (SBC) was suggested instead as selection criteria to favor a more parsimonious model [8], picking the model with the highest SBC value. It is then advised to check the approximation assumptions by running the 15 best models suggested by WAM to compare and check the correlation between the real LRT and SBC to the approximated ones.

The R code used to compute the Wald's approximation of the LRT statistic for all the covariate submodels of the FFEM (run22) is available in Appendix B.10. The results are presented in Table S6. Only the first 15 rows (i.e. the 15 best models according to the SBC criteria) of the table are shown out of 256 rows corresponding to the 2^k possible covariates models from that FFEM.

The resulting covariate model suggested by SBC, ranked first in Table S6, was run in NONMEM (run24 from Table S4, NONMEM code available in Appendix A.4), resulting in an increased OFV of +23 points for a model where 4 covariates effects were removed, compared to the FFEM model. The corresponding χ^2 threshold is 9.49 (α=0.05).

Table S6: Results from the WAM selection method using the fit of the FFEM (run22). Hypothetical restricted models are sorted based on SBC, only the first 15 rows of the table are shown.

Rank	Included covariates parameters	aLRT	SBC
1	Sex on Base, Age on Base, Diag on Normal, Age on Normal	9.74	-51.86
2	Sex on Base, Age on Base, Age on Normal	16.61	-52.16
3	Sex on Base, Age on Base, WT on Base, Diag on Normal, Age on Normal	5.94	-53.09
4	Sex on Base, Age on Base, WT on Base, Age on Normal	12.74	-53.36
5	Sex on Base, Age on Base, Diag on Normal, Sex on Normal, Age on Normal	6.66	-53.45
6	Sex on Base, Diag on Normal, Age on Normal	20.16	-53.94
7	Diag on Normal, Sex on Normal, Age on Normal	20.34	-54.03
8	Sex on Base, Age on Normal	26.77	-54.11
9	Sex on Base, Age on Base, Diag on Normal, Age on Normal, WT on Normal	8.16	-54.20
10	Diag on Normal, Sex on Normal, Age on Normal, WT on Normal	14.76	-54.37
11	Age on Base, Diag on Normal, Sex on Normal, Age on Normal, WT on Normal	8.61	-54.43
12	Age on Base, Diag on Normal, Sex on Normal, Age on Normal	15.27	-54.62
13	Diag on Base, Sex on Base, Diag on Normal, Age on Normal	15.46	-54.72
14	Sex on Base, Age on Base, Age on Normal, WT on Normal	15.63	-54.81
15	Diag on Base, Sex on Base, Age on Base, Diag on Normal, Age on Normal	9.42	-54.83

aLRT: approximated likelihood ratio statistic, SBC: Schwarz's bayesian criterion.

4.4 Stepwise covariate modelling (SCM)

The SCM (stepwise covariate model) method includes a forward selection and a backward elimination of covariates effects to a model. In short, one model for each relevant parameter-covariate relationship is tested in an uni-variate manner. In the first step the model that gives the best fit of the data according to some criteria is retained and taken forward to the next step. In the following steps all remaining parameter-covariate combinations are tested until no more covariates meet the criteria for being included into the model. The Forward Selection can be followed by Backward Elimination, which proceeds as the Forward Selection but reversely, using stricter criteria for model improvement. An implementation is available in PsN [9,10] in the function `scm`. An application example using the PsN implementation is given below.

As many options are possible to run SCM, it is impractical to specify them all on the command line, hence PsN requires the user to write a dedicated configuration file saved with the `.scm` extension. The SCM configuration can then be run using the following command:

```
scm -config_file=theophylline.scm
```

An example of configuration file for the theophylline data, called `theophylline.scm`, using the base model without covariates (run11 from Table S4) as starting point to screen all the available covariates on all the fixed parameters, is provided below. More information can be found in the PsN documentation (https://uupharmacometrics.github.io/PsN/docs.html).

```
## model=run11.mod
## search_direction=both
## p_forward=0.05
## p_backward=0.01
##
## continuous_covariates=AGE,WT
## categorical_covariates=SEX,DIAG
##
## [test_relations]
## BASE=AGE,WT,SEX,DIAG
## NORMAL=AGE,WT,SEX,DIAG
## T50=AGE,WT,SEX,DIAG
## C50=AGE,WT,SEX,DIAG
##
## [valid_states]
## continuous = 1,2,3,4,5
## categorial = 1,2,3,4,5
```

PsN outputs the results from the SCM selection under multiple formats, among them two were selected to illustrate the method. The first illustration represents the drop in OFV achieved by each step during the forward search as shown in Figure S10 (R code available in Appendix B.11.1). The second output selected is the text file summarizing the whole SCM search (`scmlog.txt`) and the final parameters-covariate relationships included in the final model, it is available in Appendix B.11.2.

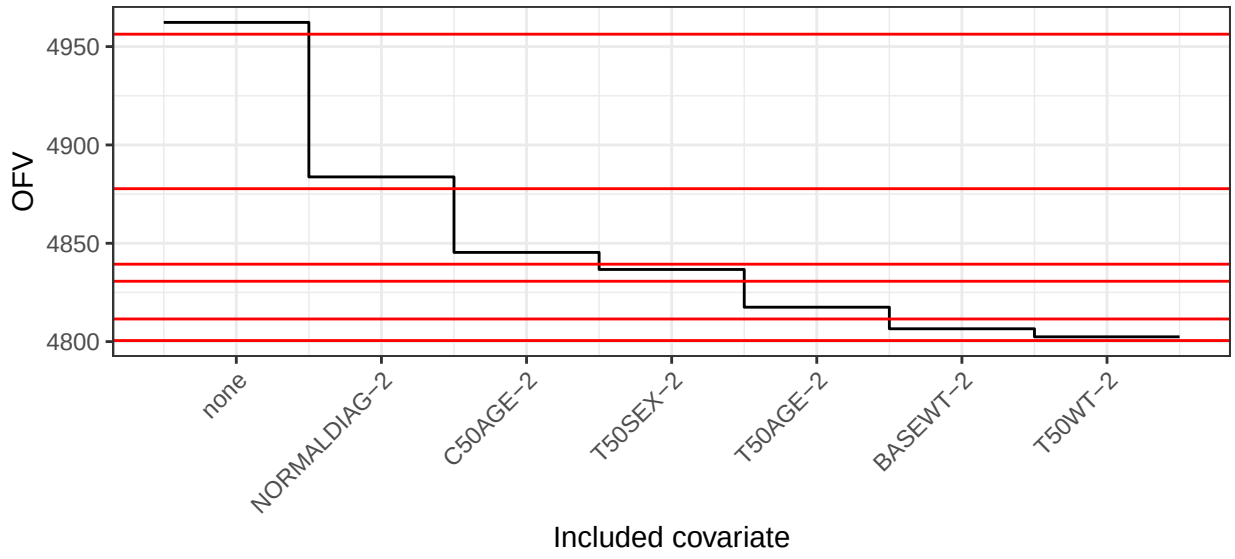

Figure S10: Drop in OFV for each step of the SCM forward search, associated with the selected parameter-covariate selection. Base, Normal, C50 and T50 are defined in Table S3. Red lines correspond to the significance threshold of each step.

The code below shows the lines added in the final SCM model compared to the base model without covariates, hence how the covariates were included into the model. The full NONMEM code is available in Appendix A.6.

```
## > $THETA  (-0.016,0.00557403,0.032) ; BASEWT1
## > $THETA  (-0.024,0.0226863,0.051) ; C50AGE1
## > $THETA  (-1,-0.496745,5) ; NORMALDIAG1
## >  (-1,-0.2698,5) ; NORMALDIAG2
## > $THETA  (-0.024,0.0442177,0.051) ; T50AGE1
## > $THETA  (-1,-0.780861,5) ; T50SEX1
## > $THETA  (-0.016,-0.0129475,0.032) ; T50WT1
## > $OMEGA  0.118039  ;   IIV_BASE
## >  0.0044012  ; IIV_NORMAL
## >  2.8273  ;    IIV_T50
## >  0.328448  ;    IIV_C50
## > $SIGMA  0.0249899  ;   RUV_PROP
## > T50WT = ( 1 + THETA(13)*(WT - 66))
## > IF(SEX.EQ.0) T50SEX = 1  ; Most common
## > IF(SEX.EQ.1) T50SEX = ( 1 + THETA(12))
## > T50AGE = ( 1 + THETA(11)*(AGE - 34.5))
## > T50COV=T50AGE*T50SEX*T50WT
## > IF(DIAG.EQ.1) NORMALDIAG = 1  ; Most common
## > IF(DIAG.EQ.2) NORMALDIAG = ( 1 + THETA(9))
## > IF(DIAG.EQ.3) NORMALDIAG = ( 1 + THETA(10))
## > NORMALCOV=NORMALDIAG
## > C50AGE = ( 1 + THETA(8)*(AGE - 34.5))
## > C50COV=C50AGE
## > BASEWT = ( 1 + THETA(7)*(WT - 66))
## > BASECOV=BASEWT
## > TVBASE = BASECOV*TVBASE
## > TVNORMAL = NORMALCOV*TVNORMAL
## > TVT50 = T50COV*TVT50
## > TVC50 = C50COV*TVC50
```

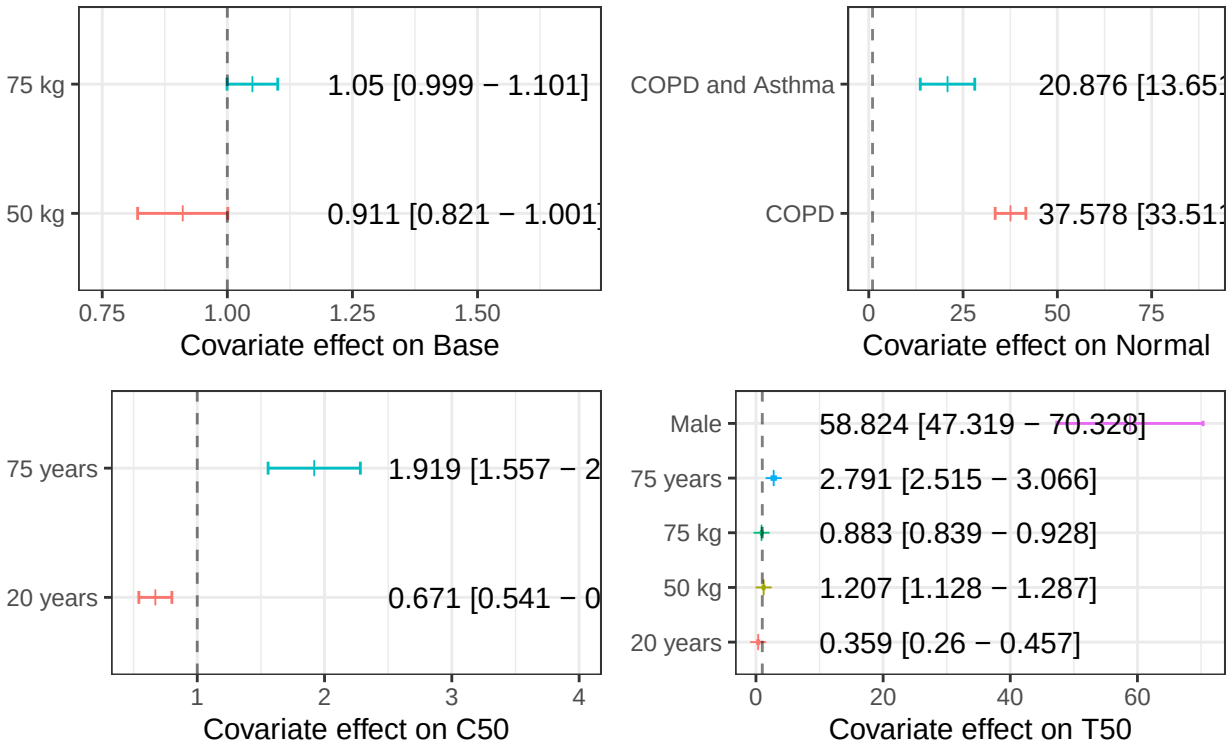

Figure S11: Forest plot representing the impact of the covariates on the Normal parameter based on the published model (run20), relative to a typical participant in the data set (40 years old male with asthma), represented by the vertical dashed black line. The bar symbols represent the median Normal value for the applicable sub population category, and the whiskers represent the 95% CI of the median values based on SIR imprecision estimates. The corresponding values are presented on the right.

4.5 The Least absolute shrinkage and selection operator (LASSO)

Covariate models for population pharmacokinetics and pharmacodynamics are often built with a stepwise covariate modelling procedure. When analyzing a small data set this method may produce a covariate model that suffers from selection bias and poor predictive p erformance. The LASSO tool [11] is a method suggested to remedy these problems. It may also be faster than the SCM tool and provide a validation of the covariate model.

In the LASSO tool all covariates must be standardized to have zero mean and a standard deviation of one. Subsequently, the model containing all potential covariate–parameter relations is fitted w ith a restriction: the sum of the absolute covariate coefficients must be smaller than a value, t. The restriction will force some coefficients towards ze ro wh ile th e ot hers ar e es timated wi th sh rinkage. Th is me ans in pr actice th at when fitting the model the covariate relations are tested for inclusion at the same time as the included relations are estimated. For a given SCM analysis, the model size depends on the p-value required for the selection. In the LASSO tool the model size instead depends on the value of t which can be estimated using cross-validation. The LASSO method is implemented in PsN [9,10], including the cross-validation step required to estimate the t value. An application example using the PsN implementation is given below. The Figure S12 illustrates how the optimal t value is selected based on the cross-validation results.

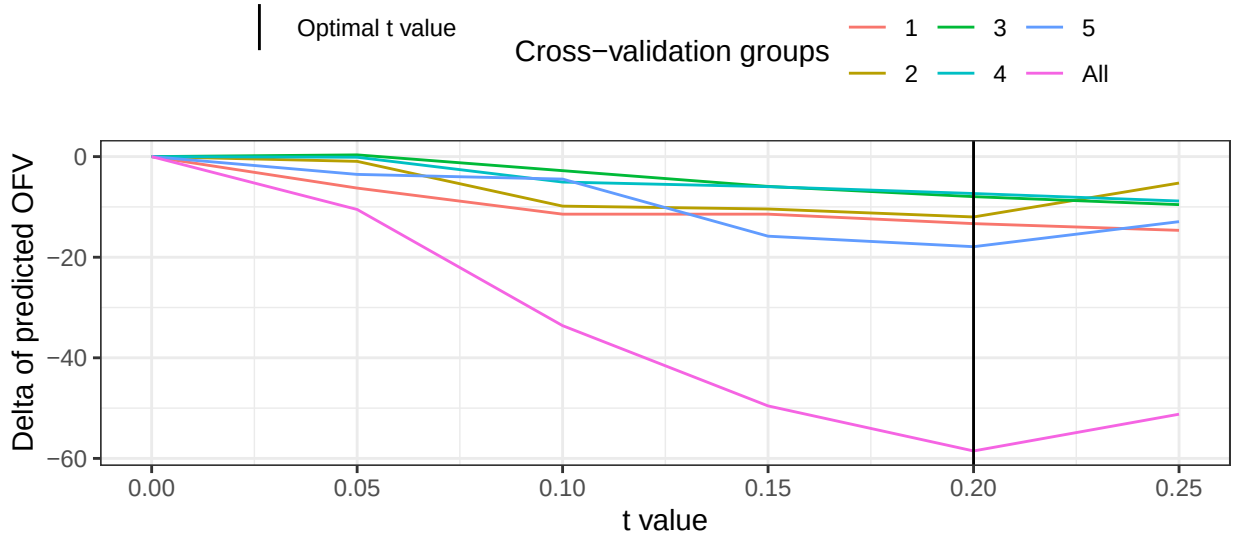

Figure S12: Predictive OFV in each of the five cross-validation groups and sum across the five groups.

Each parameter, e.g. NORMAL, listed with the mandatory option `-relations` must have the typical value encoded as TV parameters in the model e.g. TVNORMAL. The covariate effects w ill b e a dded multiplicatively to the TV variables. In the code below, age and weight (continuous, indicated with -2 in the command), and sex and diagnosis (categorical, indicated with -1 in the command) are tested on the normal parameter of the base model (see section 3.2).

```
lasso run11.mod -dir=lasso_run11_normal -relations=NORMAL:AGE-2,WT-2,SEX-1,DIAG-1
```

The main output of this command is shown in Table S7.

Table S7: Final coefficients of the parameters tested on the parameter of interest (NORMAL) with LASSO.

NORMALAGE	NORMALDIAG2	NORMALDIAG3	NORMALSEX1	NORMALWT	FACTOR	CONVERGED
-0.0512465	-0.1487535	0	0	0	1.000057	1

Coefficients of 0 mean that the covariate was not selected. FACTOR is the correction factor applied to the variables to ensure that the ratio between the absolute sum of the added covariates effects and the t value is equal to 1. CONVERGED indicates that the final covariate model converged.

The code below shows the lines added to the base model without covariates, and how the covariates were included into the model.

```
## > $THETA  (-0.49650,0.0001,0.79798) ; TH7 NORMALAGE
## > $THETA  (-0.44892,0.0001,2.24459) ; TH8 NORMALDIAG2
## > $THETA  (-0.12451,0.0001,8.09297) ; TH9 NORMALDIAG3
## > $THETA  (-0.78384,0.0001,1.28550) ; TH10 NORMALSEX1
## > $THETA  (-0.23442,0.0001,0.44966) ; TH11 NORMALWT
## > $THETA  (-1000000,0.2) FIX ; TH12 T-VALUE
## > $OMEGA  0.164643  ;   IIV_BASE
## >  0.160737  ; IIV_NORMAL
## >  1.59823  ;    IIV_T50
## >  0.41132  ;    IIV_C50
## > $SIGMA  0.0190902  ;   RUV_PROP
## > $PRED
## > ;;; LASSO-BEGIN
## >     TVALUE  = THETA(12)
## >     ABSSUM = ABS(THETA(7))+ABS(THETA(8))+ABS(THETA(9))+ABS(THETA(10))
## >     ABSSUM = ABSSUM+ABS(THETA(11))
## >
## >     RATIO = ABSSUM/TVALUE
## >     IF (RATIO .GT. 5) EXIT 1 1
## >     FACTOR = EXP(1-RATIO)
## >
## >     DIAG2 = 0
## >     DIAG3 = 0
## >     SEX1 = 0
## >     IF (DIAG .EQ. 2) DIAG2=1
## >     IF (DIAG .EQ. 3) DIAG3=1
## >     IF (SEX .EQ. 1) SEX1=1
## >
## >     NORMALAGE = THETA(7)*(AGE-38.78030)/18.97621*FACTOR
## >     NORMALDIAG2 = THETA(8)*(DIAG2-0.16667)/0.37410*FACTOR
## >     NORMALDIAG3 = THETA(9)*(DIAG3-0.01515)/0.12262*FACTOR
## >     NORMALSEX1 = THETA(10)*(SEX1-0.37879)/0.48693*FACTOR
## >     NORMALWT = THETA(11)*(WT-67.22576)/14.71551*FACTOR
## >
```

4.6 Full random effect model (FREM)

Using FREM [12–14] the selection of covariates of interest is made without concern regarding their correlation. Covariates are entered into the data set as observed variables, and their multivariate distribution are modeled as random effects. A full covariance matrix between random effects for parameters and covariates is estimated together with the other model components. The FREM method is implemented in PsN [9,10], an application example using the PsN implementation is given below. The following command was used:

```
frem run11.mod -covariates=SEX,WT,AGE,DIAG -categorical=SEX,DIAG -dir=frem_run11 -bipp
```

The forest plots (Figure S13) illustrate the estimated covariate effect expressed as effect size per parameter in percent with its uncertainty. Such plots are common to illustrate any covariate effect for pre-specified covariates methods where the confidence interval are not subject to selection bias. In this example, the COPD and asthma diagnosis category had only two individuals, resulting in parameters with very high uncertainty (values $> 1e + 20$). These values were not plotted to keep the rest of the plot readable.

The final FREM model code is available in Appendix A.7. The data used to create Figure S13 and Figure S14 were found in the PsN generated folder and processed according to the code provided in Appendix.

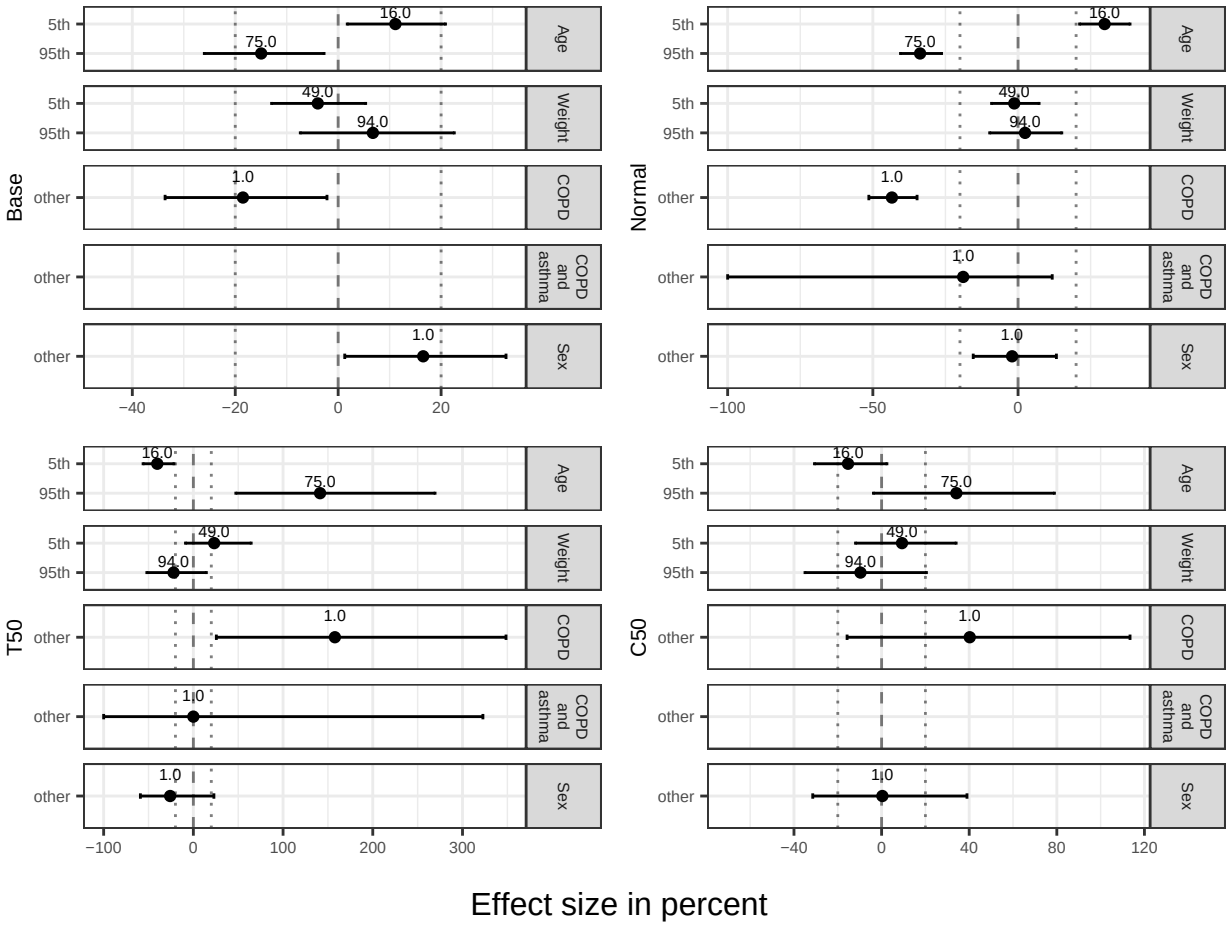

Figure S13: Covariate effects in percent with uncertainty for each parameter and covariate. The figures are the 5th and 95th percentile covariate values, the dashed line is the no effect line and the dotted lines delimit the clinical non-relevance area (±20%). Infinite values were not plotted, resulting in empty facets. Base, Normal, C50 and T50 are defined in Table S3.

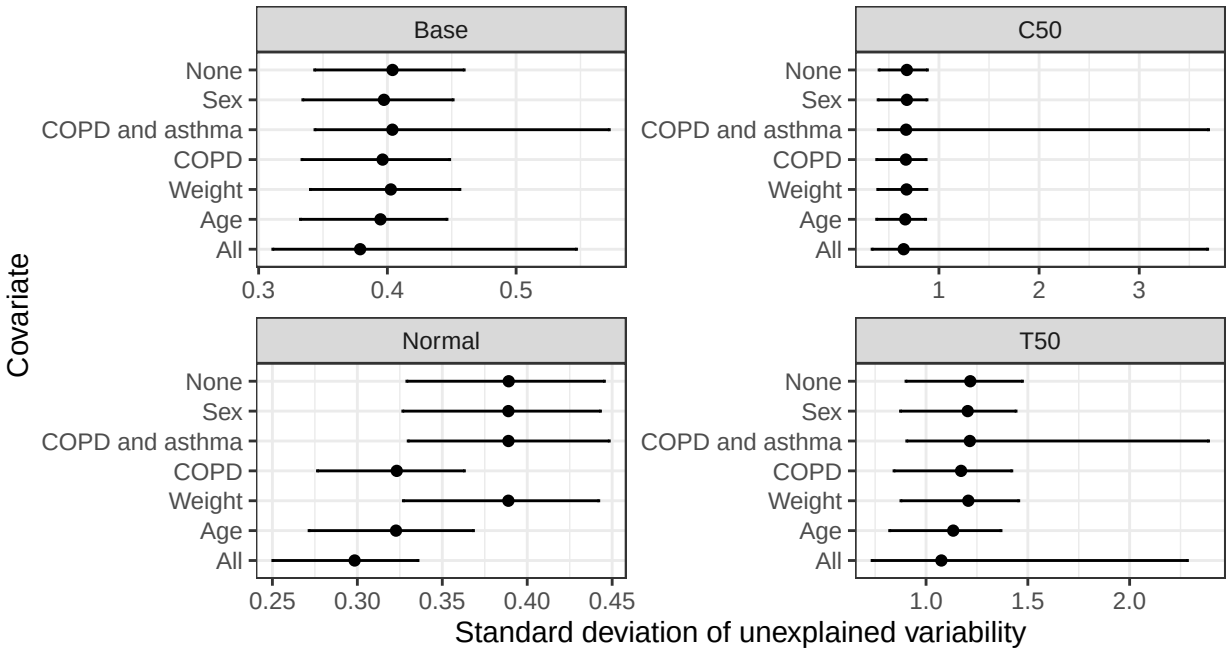

Figure S14: Original unexplained variability with the uncertainty for all parameter and covariate combinations. Dots are the observed standard deviation (SD), and bars represent the 5th and 95th SD uncertainty. Base, Normal, C50 and T50 are defined in Table S3.

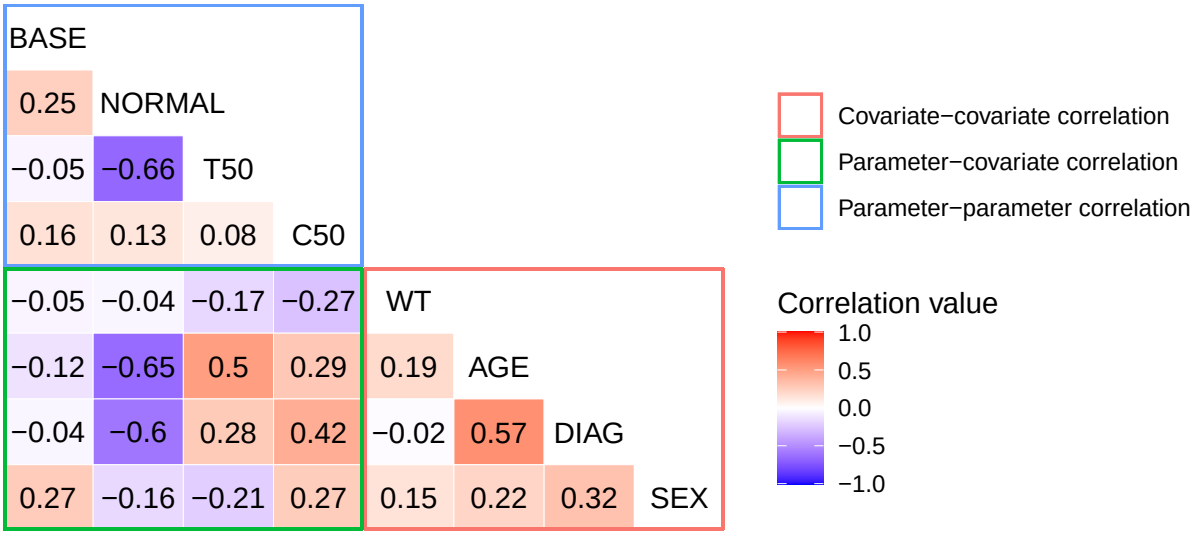

Figure S15: Correlation matrix from the FREM model (run28).

5 Model diagnosis for covariate effect

This section provides examples of how to diagnose covariate effects u sing V PCs a nd E BEs-covariates relationships plots. It can be noted that the VPCs intervals has been suggested to have different names [15,16].

Figure S16 showed an improvement in the VPCs between the model without covariates (run11 from Table S4), and the covariate model obtained by running SCM (run25 from Table S4). In that case the covariate inclusion improved the prediction of the highest observations as the 95% CI of the 95th predicted percentile is narrower, the median of the observations is also better predicted. Figure S17 showed the same VPCs but stratified on the age covariate which was included in the covariate model obtained with SCM, using the median age to create two age categories. The stratification highlighted that the younger individuals had higher PEFR, and that instead of having the median typical observations of each categories at the extreme of the prediction interval, the inclusion of covariates resulted in median prediction intervals better centered around the median observations. Appendix B.14 presented the details on how to create the different VPCs.

Figure S18 showed the PEFR predictions as a function of the age covariate before (left panel), and after (right panel) the inclusion of covariates. It showed that the dependency of PEFR on the age is better predicted after the covariates inclusion. The median prediction interval is better centered around the observations, the PEFR of the oldest individuals and the highest PEFR values are better predicted. Both imprecision and fit improved for the SCM based covariates model in each of the three different VPCs presented.

Figure S19 showed the PEFR predictions as a function of the theophylline concentration before (left panel), and after (right panel) the inclusion of covariates. The SCM covariate effect estimates did not have any clear effect on the median prediction of PEFR but did help move the upper and lower predicted percentiles closer to the observed percentiles. This pattern supports the choice of an Emax pharmacodynamic model for the effect of theophylline but it is not possible to determine which covariate, or combination of covariates identified by SCM, is explaining the between subject variability. Application of just age and diagnosis on T50 which were identified by the FREM method (Figure S13) would be a starting point to see if these two covariates help predict the VPC variability. The T50 parameter describes the speed of recovery from airway obstruction and that would be expected to be longer with older age and COPD.

The difference in the EBEs-covariates relationships before (Figure S7) and after the covariates inclusion can also be inspected. The expected pattern is a decrease in correlation between the EBEs and the covariates. Figure S20 displays the EBEs-covariates relationships after the inclusion of covariates following the SCM method (run25). The linear regression trends are flattened suggesting that the existing correlations are now included in the model. See section 4.4 for details about the SCM covariate model.

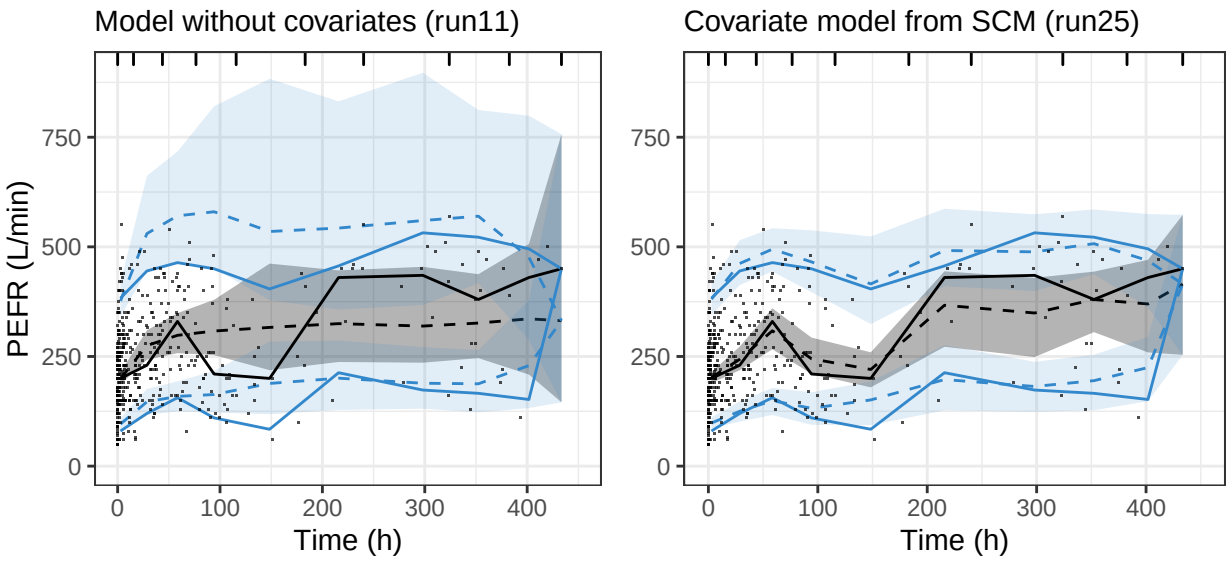

Figure S16: Visual predictive check plots of PEFR versus time profile. The plot is based on 500 simulations. The dashed lines and shaded areas represent the median, 5th and 95th percentiles of the model simulations and their corresponding 95% confidence intervals. The solid lines represent the median, 5th and 95th percentiles of the observed data. PEFR: peak expiratory flow rate.

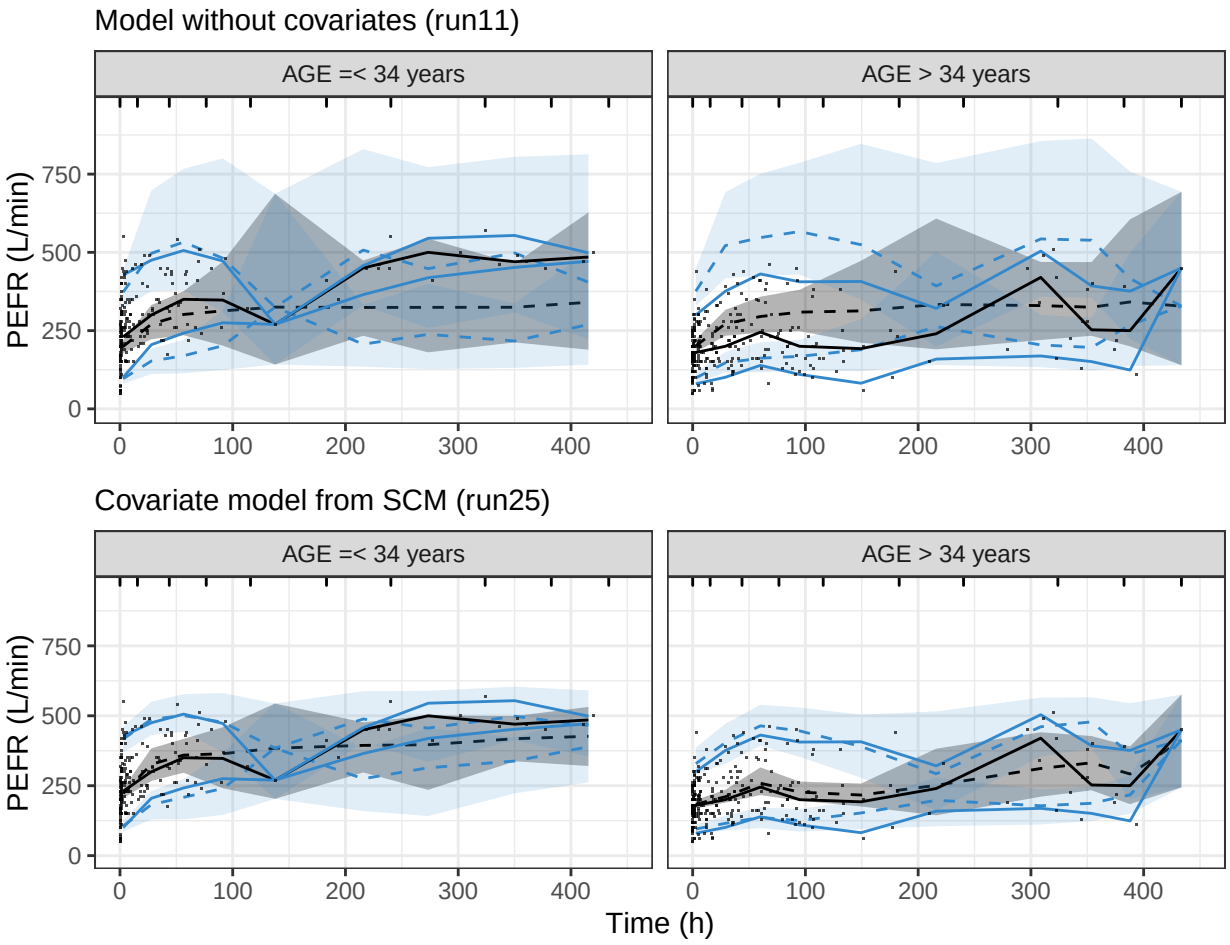

Figure S17: Visual predictive check plots of PEFR versus time profile before (top) and after (bottom) covariate inclusion, stratified based on the age covariate. The plot is based on 500 simulations. The dashed lines and shaded areas represent the median, 5th and 95th percentiles of the model simulations and their corresponding 95% confidence intervals. The solid lines represent the median, 5th and 95th percentiles of the observed data. PEFR: peak expiratory flow rate.

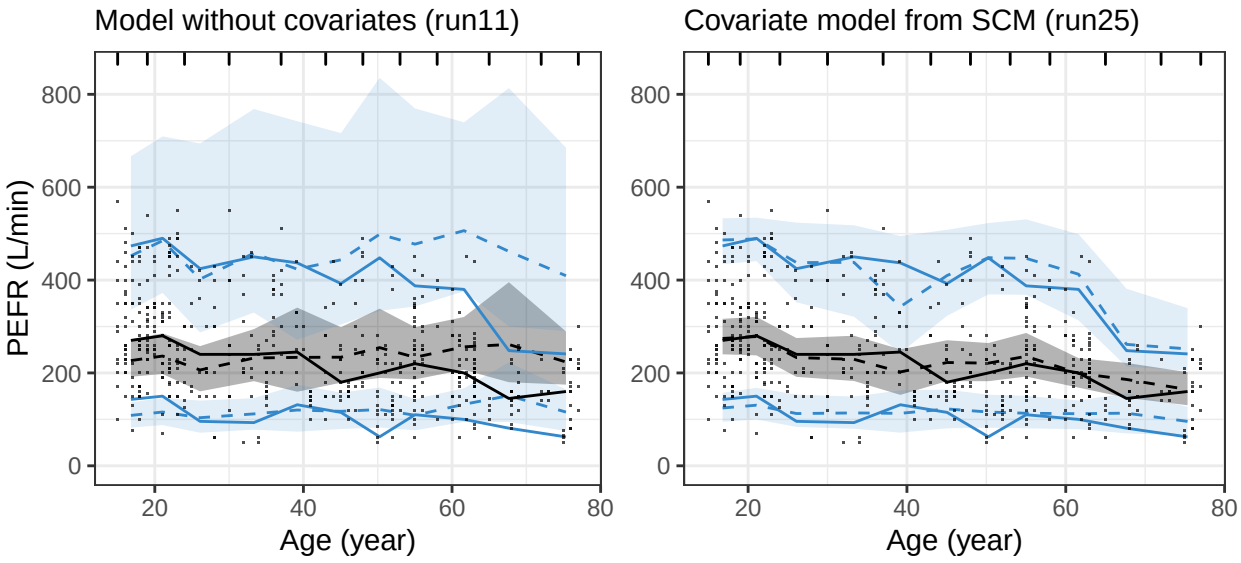

Figure S18: Visual predictive check plots of PEFR versus age. The plot is based on 500 simulations. The dashed lines and shaded areas represent the median, 5th and 95th percentiles of the model simulations and their corresponding 95% confidence intervals. The solid lines represent the median, 5th and 95th percentiles of the observed data. PEFR: peak expiratory flow rate.

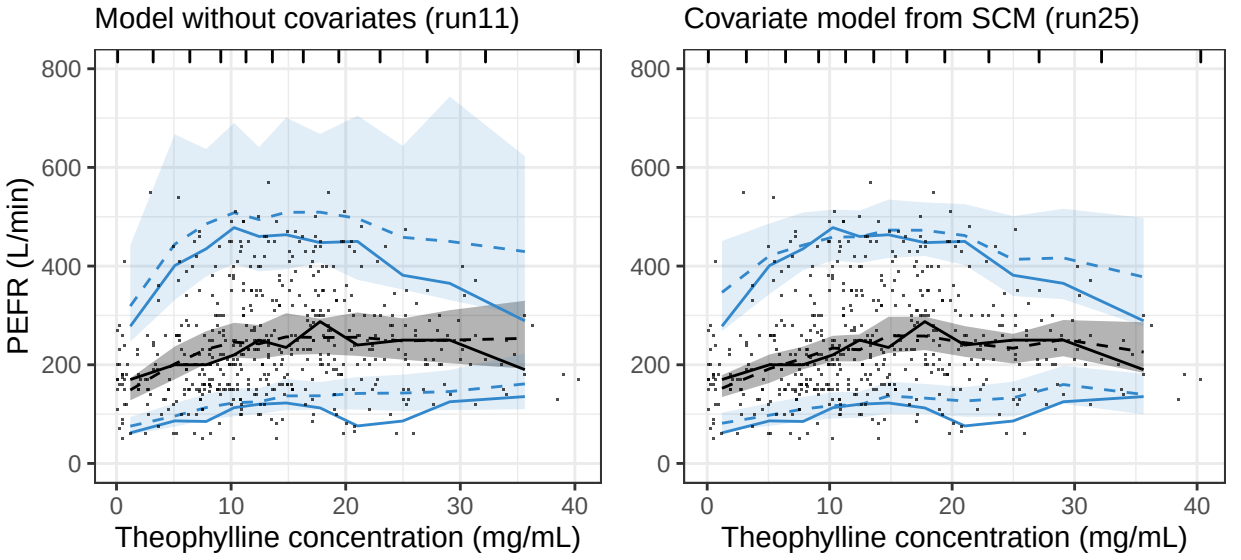

Figure S19: Visual predictive check plots of PEFR versus theophylline concentration. The plot is based on 500 simulations. The dashed lines and shaded areas represent the median, 5th and 95th percentiles of the model simulations and their corresponding 95% confidence intervals. The solid lines represent the median, 5th and 95th percentiles of the observed data. PEFR: peak expiratory flow rate.

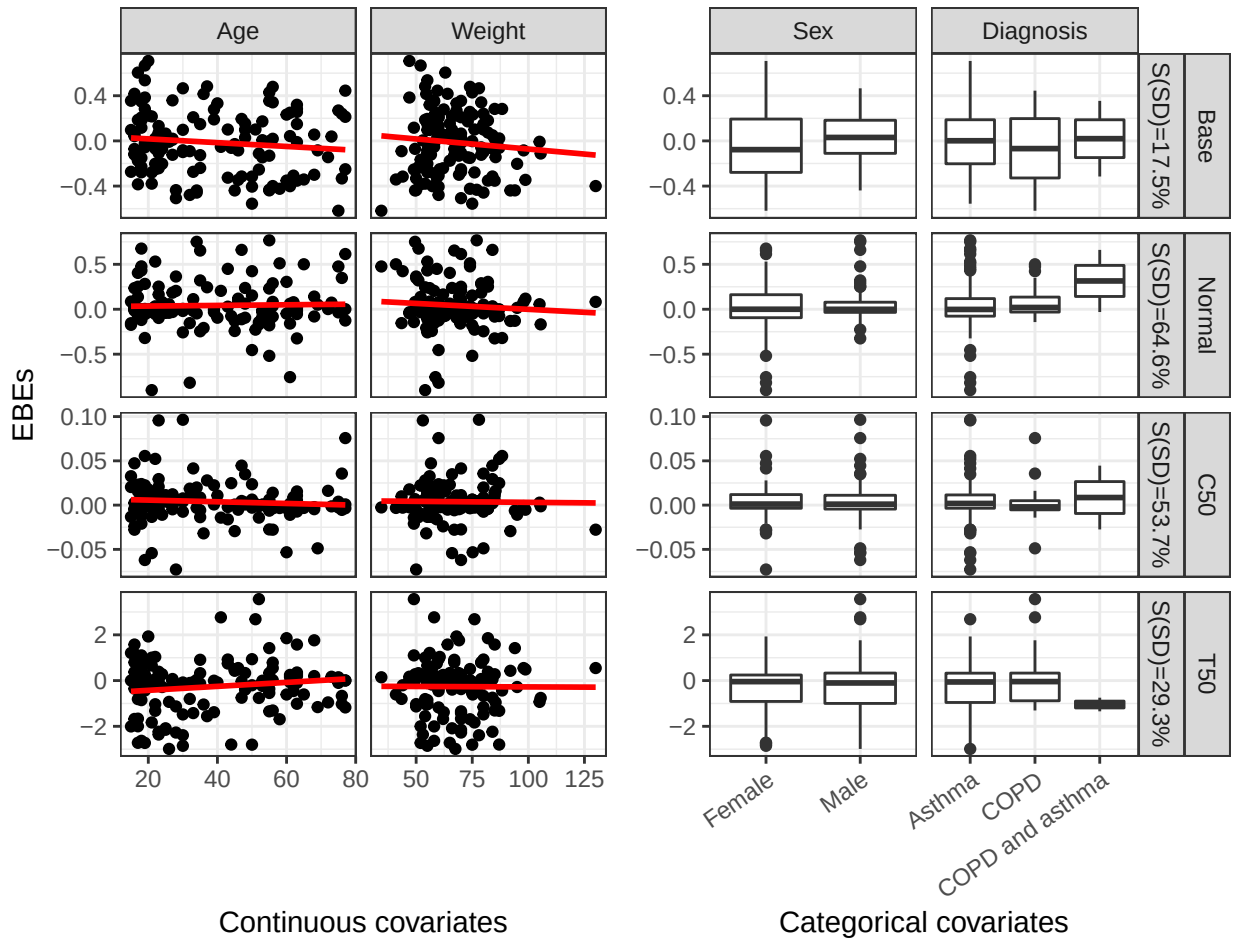

Figure S20: EBEs vs covariates for the covariate model based on SCM (run25). Continuous covariates are displayed using scatter plots on the left panel, the red line is a linear regression. Categorical covariates are displayed using boxplots. S(SD) is the shrinkage computed on the standard deviation. Base, Normal, C50 and T50 are defined in Table S3.

6 Covariate effect summary and reporting

This section aims at providing an example on how to summarize and report the covariates effects included in a model, using the covariate model and results from the publication (run20 from Table S4).

The model obtained after the covariate model building process is presented below. The Table S8 presented the untransformed final estimates of the covariate model (run20), compared to the values of the model without covariates (run11), including RSE and shrinkage for the IIV parameters. The RSE from both models were obtained with SIR (sampling importance resampling). The estimated PEFR on admission to the study (Base) was 136 L/min with a predicted fully recovered value of 481 L/min in a 40-year-old male asthmatic patient. The half-life of the bronchoconstrictor factor was 19.6 hours. The maximum effect of theophylline is to produce 37.3% (ε_{theo}) of the maximum effect associated with time alone. A concentration of 8.48 mg/L achieved 50% of the potential theophylline effect. The steepness of the curve was described by a exponent of 2.59. The model predictions were clearly improved by modification of the Normal parameter with respect to diagnosis, age and gender. The Normal parameter increased with the diagnosis of COPD, was typically lower for females, and was inversely proportional to the age.

The inclusion of the covariates resulted in minor changes in the parameters, with the exception of the Normal parameter, the T50 parameter, and the IIV of the Normal parameter. Changes in the Normal parameter were expected as it is the parameter on which the covariates were added. This parameter represented the typical value of the whole population in the model without covariates, while it represent the typical value of a subset in the covariate model (40 years old males with asthma) because of the normalized relationships. The decrease in the IIV of the Normal parameter was also expected as the covariate inclusion explained part of the parameter variability in the population.

The illustration of the included effects is presented using the forest plot representation in Figure S21. All the covariate effects decrease the PEFR, except young age which was associated to higher PEFR. The code used to create the plot is presented in Appendix B.13.

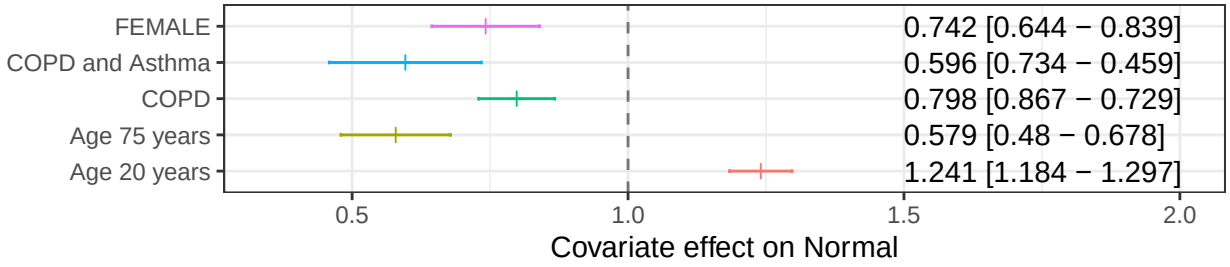

Figure S21: Forest plot representing the impact of the covariates on the Normal parameter based on the published model (run20), relative to a typical participant in the data set (40 years old male with asthma), represented by the vertical dashed black line. The bar symbols represent the median Normal value for the applicable sub population category, and the whiskers represent the 95% CI of the median values based on SIR imprecision estimates. The corresponding values are presented on the right.

Table S8: Parameters table for the base model without covariate (run11) and the model based on the published covariate model (run20).

			Base model without covariates (run11)			Covariate model based on the publication (run20)		
Parameter	Description	Unit	Value	RSE*	Shrinkage (%)	Value	RSE*	Shrinkage (%)
Base	PEFR at baseline	L/min	143	0.0534	-	136	0.0642	-
Normal	PEFR without BCF†	L/min	332	0.0546	-	481	0.0621	-
$T50$	BCF half-life	h	9.54	0.37	-	19.6	0.263	-
$C50$	Concentration producing 50% of the recoverable PEFR	mg/L	10.3	0.215	-	8.48	0.268	-
ε_{theo}	Theophylline efficacy relative to BCF effect	-	0.414	0.207	-	0.373	0.197	-
Hill exponent	Steepness curve exponent	-	3.25	0.25	-	2.59	0.3	-
IIV Base		-	0.404	0.229	20.3	0.405	0.256	18.9
IIV Normal		-	0.389	0.221	16.6	0.164	0.364	43.4
IIV $T50$		-	1.22	0.413	34	1.55	0.281	29.6
IIV $C50$		-	0.68	0.638	48.6	0.664	0.609	48.2
Proportional RUV		-	0.145	0.14	30.4	0.159	0.123	25.8
Diagnosis on Normal	Typical value fraction per category	-	-	-	-	0.202	0.174	-
Female on Normal	Typical value fraction for females	-	-	-	-	-0.258	0.193	-
Years on Normal	Scale factor	-	-	-	-	-0.012	0.12	-

IIV and RUV values are reported as standard deviations for the diagonal elements.
RSE for IIV and RUV, estimate of the variance-covariance matrix from NONMEM, are reported on the approximate standard deviation scale (standard error/variance estimate)/2. The shrinkage correponds to the standard deviation shrinkage.
PEFR: peak expiratory flow rate, BCF: broncho-constriction factor, RUV: residual unexplained variability, IIV: inter-individual variability, RSE: relative standard error.
* RSE obtained with SIR.
† Value corresponding to a typical individual with the following characteristics for the covariate model: male, 40 years old, with asthma.

7 References

1. Holford, N.; Hashimoto, Y.; Sheiner, L.B. Time and Theophylline Concentration Help Explain the Recovery of Peak Flow Following Acute Airways Obstruction. *Clinical pharmacokinetics* **1993**, *25*, 506–515.
2. Holford, N.; Black, P.; Couch, R.; Kennedy, J.; Briant, R. Theophylline Target Concentration in Severe Airways Obstruction—10 or 20 Mg/l? *Clinical pharmacokinetics* **1993**, *25*, 495–505.
3. Savic, R.M.; Karlsson, M.O. Importance of Shrinkage in Empirical Bayes Estimates for Diagnostics: Problems and Solutions. *The AAPS journal* **2009**, *11*, 558–569.
4. Lavielle, M.; Ribba, B. Enhanced Method for Diagnosing Pharmacometric Models: Random Sampling from Conditional Distributions. *Pharmaceutical research* **2016**, *33*, 2979–2988.
5. Mandema, J.W.; Verotta, D.; Sheiner, L.B. Building Population Pharmacokineticpharmacodynamic Models. I. Models for Covariate Effects. *Journal of pharmacokinetics and biopharmaceutics* **1992**, *20*, 511–528.
6. Jonsson, E.N.; Karlsson, M.O. Xpose–an s-PLUS Based Population Pharmacokinetic/Pharmacodynamic Model Building Aid for NONMEM. *Computer Methods and Programs in Biomedicine* **1999**, *58*, 51–64, doi:10.1016/s0169-2607(98)00067-4.
7. Keizer, R.J.; Karlsson, M.O.; Hooker, A.C. Modeling and Simulation Workbench for NONMEM: Tutorial on Pirana, PsN, and Xpose. *CPT: Pharmacometrics & Systems Pharmacology* **2013**, *2*, doi:10.1038/psp.2013.24.
8. Kowalski, K.G.; Hutmacher, M.M. Efficient Screening of Covariates in Population Models Using Wald's Approximation to the Likelihood Ratio Test. *Journal of pharmacokinetics and pharmacodynamics* **2001**, *28*, 253–275.
9. Lindbom, L.; Ribbing, J.; Jonsson, E.N. Perl-Speaks-NONMEM (PsN)—a Perl Module for NONMEM Related Programming. *Computer methods and programs in biomedicine* **2004**, *75*, 85–94.
10. Lindbom, L.; Pihlgren, P.; Jonsson, N. PsN-Toolkit—a Collection of Computer Intensive Statistical Methods for Non-Linear Mixed Effect Modeling Using NONMEM. *Computer methods and programs in biomedicine* **2005**, *79*, 241–257.
11. Ribbing, J.; Nyberg, J.; Caster, O.; Jonsson, E.N. The Lasso—a Novel Method for Predictive Covariate Model Building in Nonlinear Mixed Effects Models. *Journal of pharmacokinetics and pharmacodynamics* **2007**, *34*, 485–517.
12. Karlsson, M.O. A Full Model Approach Based on the Covariance Matrix of Parameters and Covariates PAGE 2012.
13. Ivaturi, V.; Keizer, R.; Karlsson, M.O. A Full Random Effects Model for Characterising Covariate Relations WCOP 2012.
14. Yun, H.; Niebecker, R.; Svensson, E.; Karlsson, M.O. Evaluation of FREM and FFEM Including Use of Model Linearization PAGE 2013.
15. Hu, C. Variability and Uncertainty: Interpretation and Usage of Pharmacometric Simulations and Intervals. *Journal of Pharmacokinetics and Pharmacodynamics* **2022**, 1–5.
16. Jonsson, E.N.; Nyberg, J. A Quantitative Approach to the Choice of Number of Samples for Percentile Estimation in Bootstrap and Visual Predictive Check Analyses. *CPT: Pharmacometrics & Systems Pharmacology* **2022**.

8 Appendix

A NONMEM code

A.1 Covariate model from the publication (run20)

```
;; 1. Based on: 11
$PROBLEM Covariate model from the publication
$DATA      ../data/theo_no_missing.csv IGNORE=@
$INPUT      ID TIME THEO AGE WT SEX RACE DIAG DV
$ESTIMATION METHOD=1 NOABORT MAXEVAL=9999 POSTHOC
$COVARIANCE
$THETA  (0,141.587) ; TV_BASE
 (0,336.984) ; TV_NORMAL
 (.01,7.27855) ; TV_T50
 (.01,11.427) ; TV_C50
 (0,0.463118,1) ; TV_ALPHA
 (.1,2.82371,5) ; TV_HILL
 0.01 ; DIAG
 0.01 ; FEMALE
 0.01 ; FYEARS
$OMEGA  0.160171  ;   IIV_BASE
 0.143395  ; IIV_NORMAL
 1.7514  ;    IIV_T50
 0.487499  ;    IIV_C50
$SIGMA  0.0193434  ;   RUV_PROP
$PRED

; Group variables (FIXED effects)
     FDIAG = 1+THETA(7) * (1-DIAG)
     FSEX  = 1+THETA(8) * (1-SEX)
     FAGE  = 1+THETA(9) * (AGE-40)

     TVNORMAL = THETA(2)
     GRP_NORMAL = FDIAG*FAGE*FSEX*TVNORMAL

; Individual variables (RANDOM effects)
     TVBASE   = THETA(1)
     BASE     = TVBASE*EXP(ETA(1))
     NORMAL   = GRP_NORMAL*EXP(ETA(2))
     TVT50    = THETA(3)
     T50      = TVT50*EXP(ETA(3))
     TVC50    = THETA(4)
     C50      = TVC50*EXP(ETA(4))

     IF (T50.LE.0.01) EXIT 1 3
     IF (C50.LE.0.01) EXIT 1 4

     IF (TIME.LT.0) THEN
          LTZERO=1
     ELSE
          LTZERO=0
     ENDIF

     FTIME=LTZERO+(1-LTZERO)*EXP(-LOG(2)/T50*TIME)
     BCF=(NORMAL/BASE-1)*FTIME
     FEMAX=BCF/(BCF+1)
     PEFRT=NORMAL*(1-FEMAX)
     HILL = THETA(6)
     CPN=THEO**HILL
     C50N=C50**HILL
     PEFRCT=(NORMAL-PEFRT)*CPN/(CPN+C50N)
     ALPHA=THETA(5)
     F = PEFRT + ALPHA * PEFRCT
     IPRED = F
     IRES = DV - IPRED
     IWRES = IRES/IPRED
     Y = F * ( 1 + EPS(1) )
```

```
$TABLE      ID TIME THEO BASE NORMAL T50 C50 ALPHA HILL ETA1 ETA2 ETA3
            ETA4 Y ONEHEADER NOPRINT FILE=run20.dat
$TABLE      ID TIME Y THEO IPRED IWRES NOPRINT ONEHEADER
            FILE=sdtab20
$TABLE      ID TIME THEO BASE NORMAL T50 C50 ALPHA HILL ETA1 ETA2 ETA3
            ETA4 NOPRINT ONEHEADER FILE=patab20
```

A.2 Base model (run11)

```
;; 1. Based on:
$PROBLEM    Theophylline and time on PEFR
$DATA      ../data/theo_no_missing.csv IGNORE=@
$INPUT      ID TIME THEO AGE WT SEX RACE DIAG DV
$ESTIMATION METHOD=1 NOABORT MAXEVAL=9999 POSTHOC
$COVARIANCE
$THETA
 (0,146.22)     ; TV_BASE
 (0,327.575)    ; TV_NORMAL
 (.01,6.06929)  ; TV_T50
 (.01,12.5879)  ; TV_C50
 (0,0.474818,1) ; TV_ALPHA
 (.1,2.63332,5) ; TV_HILL
$OMEGA
 0.164643       ; IIV_BASE
 0.160737       ; IIV_NORMAL
 1.59823        ; IIV_T50
 0.41132        ; IIV_C50
$SIGMA  0.0190902 ; RUV_PROP
$PRED

; Individual variables (RANDOM effects)
     TVBASE   = THETA(1)
     BASE     = TVBASE*EXP(ETA(1))
     TVNORMAL = THETA(2)
     NORMAL   = TVNORMAL*EXP(ETA(2))
     TVT50    = THETA(3)
     T50      = TVT50*EXP(ETA(3))
     TVC50    = THETA(4)
     C50      = TVC50*EXP(ETA(4))

     IF (T50.LE.0.01) EXIT 1 3
     IF (C50.LE.0.01) EXIT 1 4

     IF (TIME.LT.0) THEN
          LTZERO=1
     ELSE
          LTZERO=0
     ENDIF

     FTIME=LTZERO+(1-LTZERO)*EXP(-LOG(2)/T50*TIME)
     BCF=(NORMAL/BASE-1)*FTIME
     FEMAX=BCF/(BCF+1)
     PEFRT=NORMAL*(1-FEMAX)
     HILL = THETA(6)
     CPN=THEO**HILL
     C50N=C50**HILL
     PEFRCT=(NORMAL-PEFRT)*CPN/(CPN+C50N)
     ALPHA=THETA(5)
     F = PEFRT + ALPHA * PEFRCT
     IPRED = F
     IRES = DV - IPRED
     IWRES = IRES/IPRED
     Y = F * ( 1 + EPS(1) )

$TABLE      ID TIME THEO BASE NORMAL T50 C50 ALPHA HILL ETA1 ETA2 ETA3
            ETA4 AGE WT SEX DIAG Y ONEHEADER NOPRINT FILE=run11.dat
```

```
$TABLE      ID TIME Y THEO IPRED IWRES NOPRINT ONEHEADER FILE=sdtab11
$TABLE      ID TIME THEO BASE NORMAL T50 C50 ALPHA HILL ETA1 ETA2 ETA3
            ETA4 NOPRINT ONEHEADER FILE=patab11
$TABLE      ID TIME AGE WT NOPRINT ONEHEADER FILE=cotab11
$TABLE      ID TIME SEX DIAG Y NOPRINT ONEHEADER FILE=catab11
```

A.3 FFEM (run22)

```
;; 1. Based on: 11
$PROBLEM    FCM model
$DATA      ../data/theo_no_missing.csv IGNORE=@
$INPUT      ID TIME THEO AGE WT SEX RACE DIAG DV
$ESTIMATION METHOD=1 NOABORT MAXEVAL=9999 POSTHOC
$COVARIANCE
$THETA  (0,146.22) ; TV_BASE
 (0,327.575) ; TV_NORMAL
 (.01,6.06929) ; TV_T50
 (.01,12.5879) ; TV_C50
 (0,0.474818,1) ; TV_ALPHA
 (.1,2.63332,5) ; TV_HILL
 0.01 ; Diag on Base
 0.01 ; Sex on Base
 0.01 ; Age on Base
 0.01 ; WT on Base
 0.01 ; Diag on Normal
 0.01 ; Sex on Normal
 0.01 ; Age on Normal
 0.01 ; WT on Normal
$OMEGA  0.164643  ;   IIV_BASE
 0.160737  ; IIV_NORMAL
 1.59823  ;    IIV_T50
 0.41132  ;    IIV_C50
$SIGMA  0.0190902  ;   RUV_PROP
$PRED

; Group variables (FIXED effects)
     BDIAG = THETA(7) * (1-DIAG)
     BSEX  = THETA(8) * (1-SEX)
     BAGE  = THETA(9) * (AGE-40)
     BWT  = THETA(10) * (WT-66)
     NDIAG = THETA(11) * (1-DIAG)
     NSEX  = THETA(12) * (1-SEX)
     NAGE  = THETA(13) * (AGE-40)
     NWT  = THETA(14) * (WT-66)

; Individual variables (RANDOM effects)
     COV_BASE = (1+BDIAG)*(1+BAGE)*(1+BSEX)*(1+BWT)
     TVBASE   = THETA(1)*COV_BASE
     BASE     = TVBASE*EXP(ETA(1))
     COV_NORMAL = (1+NDIAG)*(1+NAGE)*(1+NSEX)*(1+NWT)
     TVNORMAL   = THETA(2)*COV_NORMAL
     NORMAL = TVNORMAL*EXP(ETA(2))
     TVT50    = THETA(3)
     T50      = TVT50*EXP(ETA(3))
     TVC50    = THETA(4)
     C50      = TVC50*EXP(ETA(4))

     IF (T50.LE.0.01) EXIT 1 3
     IF (C50.LE.0.01) EXIT 1 4

     IF (TIME.LT.0) THEN
          LTZERO=1
     ELSE
          LTZERO=0
     ENDIF
```

```
     FTIME=LTZERO+(1-LTZERO)*EXP(-LOG(2)/T50*TIME)
     BCF=(NORMAL/BASE-1)*FTIME
     FEMAX=BCF/(BCF+1)
     PEFRT=NORMAL*(1-FEMAX)
     HILL = THETA(6)
     CPN=THEO**HILL
     C50N=C50**HILL
     PEFRCT=(NORMAL-PEFRT)*CPN/(CPN+C50N)
     ALPHA=THETA(5)
     F = PEFRT + ALPHA * PEFRCT
     IPRED = F
     IRES = DV - IPRED
     IWRES = IRES/IPRED
     Y = F * ( 1 + EPS(1) )

$TABLE      ID TIME THEO BASE NORMAL T50 C50 ALPHA HILL ETA1 ETA2 ETA3
            ETA4 AGE WT SEX DIAG Y ONEHEADER NOPRINT FILE=run22.dat
$TABLE      ID TIME Y THEO IPRED IWRES NOPRINT ONEHEADER FILE=sdtab22
$TABLE      ID TIME THEO BASE NORMAL T50 C50 ALPHA HILL ETA1 ETA2 ETA3
            ETA4 NOPRINT ONEHEADER FILE=patab22
$TABLE      ID TIME AGE WT NOPRINT ONEHEADER FILE=cotab22
$TABLE      ID TIME SEX DIAG Y NOPRINT ONEHEADER FILE=catab22
```

A.4 WAM (run24)

```
;; 1. Based on: 22
$PROBLEM    Covariate model based on WAM
$DATA      ../data/theo_no_missing.csv IGNORE=@
$INPUT      ID TIME THEO AGE WT SEX RACE DIAG DV
$ESTIMATION METHOD=1 NOABORT MAXEVAL=9999 POSTHOC
$COVARIANCE
$THETA  (0,164.582) ; TV_BASE
 (0,420.857) ; TV_NORMAL
 (.01,12.3227) ; TV_T50
 (.01,10.0245) ; TV_C50
 (0,0.359849,1) ; TV_ALPHA
 (.1,4.54423,5) ; TV_HILL
 0 FIX ; Diag on Base
 -0.169302 ; Sex on Base
 -0.00533748 ; Age on Base
 0 FIX ; WT on Base
 0.210587 ; Diag on Normal
 0 FIX ; Sex on Normal
 -0.0111277 ; Age on Normal
 0 FIX ; WT on Normal
$OMEGA  0.140716  ;   IIV_BASE
 0.0810301  ; IIV_NORMAL
 1.06199  ;    IIV_T50
 0.318993  ;    IIV_C50
$SIGMA  0.0233773  ;   RUV_PROP
$PRED

; Group variables (FIXED effects)
     BDIAG = THETA(7) * (1-DIAG)
     BSEX  = THETA(8) * (1-SEX)
     BAGE  = THETA(9) * (AGE-40)
     BWT  = THETA(10) * (WT-66)
     NDIAG = THETA(11) * (1-DIAG)
     NSEX  = THETA(12) * (1-SEX)
     NAGE  = THETA(13) * (AGE-40)
     NWT  = THETA(14) * (WT-66)

; Individual variables (RANDOM effects)
     COV_BASE = (1+BDIAG)*(1+BAGE)*(1+BSEX)*(1+BWT)
     TVBASE   = THETA(1)*COV_BASE
     BASE     = TVBASE*EXP(ETA(1))
```

```
     COV_NORMAL = (1+NDIAG)*(1+NAGE)*(1+NSEX)*(1+NWT)
     TVNORMAL   = THETA(2)*COV_NORMAL
     NORMAL = TVNORMAL*EXP(ETA(2))
     TVT50    = THETA(3)
     T50      = TVT50*EXP(ETA(3))
     TVC50    = THETA(4)
     C50      = TVC50*EXP(ETA(4))

     IF (T50.LE.0.01) EXIT 1 3
     IF (C50.LE.0.01) EXIT 1 4

     IF (TIME.LT.0) THEN
          LTZERO=1
     ELSE
          LTZERO=0
     ENDIF

     FTIME=LTZERO+(1-LTZERO)*EXP(-LOG(2)/T50*TIME)
     BCF=(NORMAL/BASE-1)*FTIME
     FEMAX=BCF/(BCF+1)
     PEFRT=NORMAL*(1-FEMAX)
     HILL = THETA(6)
     CPN=THEO**HILL
     C50N=C50**HILL
     PEFRCT=(NORMAL-PEFRT)*CPN/(CPN+C50N)
     ALPHA=THETA(5)
     F = PEFRT + ALPHA * PEFRCT
     IPRED = F
     IRES = DV - IPRED
     IWRES = IRES/IPRED
     Y = F * ( 1 + EPS(1) )

$TABLE      ID TIME THEO BASE NORMAL T50 C50 ALPHA HILL ETA1 ETA2 ETA3
            ETA4 AGE WT SEX DIAG Y ONEHEADER NOPRINT FILE=run24.dat
$TABLE      ID TIME Y THEO IPRED IWRES NOPRINT ONEHEADER FILE=sdtab24
$TABLE      ID TIME THEO BASE NORMAL T50 C50 ALPHA HILL ETA1 ETA2 ETA3
            ETA4 NOPRINT ONEHEADER FILE=patab24
$TABLE      ID TIME AGE WT NOPRINT ONEHEADER FILE=cotab24
$TABLE      ID TIME SEX DIAG Y NOPRINT ONEHEADER FILE=catab24
```

A.5 Final LASSO model (run26)

```
;; 1. Based on: 11
$PROBLEM    Covariate model based on LASSO
$DATA      ../data/theo_no_missing.csv IGNORE=@
$INPUT      ID TIME THEO AGE WT SEX RACE DIAG DV
$ESTIMATION METHOD=1 NOABORT MAXEVAL=9999 POSTHOC MSFO=psn_msf
$THETA  (0,138.879) ; TV_BASE
 (0,333.281) ; TV_NORMAL
 (.01,9.51597) ; TV_T50
 (.01,11.7335) ; TV_C50
 (0,0.535447,1) ; TV_ALPHA
 (.1,1.94753,5) ; TV_HILL
$THETA  -0.00270056670791727 FIX ; TH7 NORMALAGE
$THETA  -0.397630255831705 FIX ; TH8 NORMALDIAG2
$THETA  0 FIX ; TH9 NORMALDIAG3
$THETA  0 FIX ; TH10 NORMALSEX1
$THETA  0 FIX ; TH11 NORMALWT
$THETA  (-1000000,0.2) FIX ; TH12 T-VALUE
$OMEGA  0.158528  ;   IIV_BASE
 0.0615898  ; IIV_NORMAL
 1.70203  ;    IIV_T50
 0.518568  ;    IIV_C50
$SIGMA  0.0247062  ;   RUV_PROP
$PRED
    DIAG2 = 0
```

```
    DIAG3 = 0
    SEX1 = 0
    IF (DIAG .EQ. 2) DIAG2=1
    IF (DIAG .EQ. 3) DIAG3=1
    IF (SEX .EQ. 1) SEX1=1

    NORMALAGE = THETA(7)*(AGE-38.78030)
    NORMALDIAG2 = THETA(8)*(DIAG2-0.16667)
    NORMALDIAG3 = THETA(9)*(DIAG3-0.01515)
    NORMALSEX1 = THETA(10)*(SEX1-0.37879)
    NORMALWT = THETA(11)*(WT-67.22576)

    NORMALCOV = (NORMALAGE+1)*(NORMALDIAG2+1)*(NORMALDIAG3+1)*(NORMALSEX1+1)
    NORMALCOV = NORMALCOV*(NORMALWT+1)


; Individual variables (RANDOM effects)
     TVBASE   = THETA(1)
     BASE     = TVBASE*EXP(ETA(1))
     TVNORMAL = THETA(2)

    TVNORMAL = TVNORMAL*NORMALCOV
     NORMAL   = TVNORMAL*EXP(ETA(2))
     TVT50    = THETA(3)
     T50      = TVT50*EXP(ETA(3))
     TVC50    = THETA(4)
     C50      = TVC50*EXP(ETA(4))

     IF (T50.LE.0.01) EXIT 1 3
     IF (C50.LE.0.01) EXIT 1 4

     IF (TIME.LT.0) THEN
          LTZERO=1
     ELSE
          LTZERO=0
     ENDIF

     FTIME=LTZERO+(1-LTZERO)*EXP(-LOG(2)/T50*TIME)
     BCF=(NORMAL/BASE-1)*FTIME
     FEMAX=BCF/(BCF+1)
     PEFRT=NORMAL*(1-FEMAX)
     HILL = THETA(6)
     CPN=THEO**HILL
     C50N=C50**HILL
     PEFRCT=(NORMAL-PEFRT)*CPN/(CPN+C50N)
     ALPHA=THETA(5)
     F = PEFRT + ALPHA * PEFRCT
     IPRED = F
     IRES = DV - IPRED
     IWRES = IRES/IPRED
     Y = F * ( 1 + EPS(1) )

$TABLE      ID TIME THEO BASE NORMAL T50 C50 ALPHA HILL ETA1 ETA2 ETA3
            ETA4 AGE WT SEX DIAG Y ONEHEADER NOPRINT FILE=run26.dat
$TABLE      ID TIME Y THEO IPRED IWRES NOPRINT ONEHEADER FILE=sdtab26
$TABLE      ID TIME THEO BASE NORMAL T50 C50 ALPHA HILL ETA1 ETA2 ETA3
            ETA4 NOPRINT ONEHEADER FILE=patab26
$TABLE      ID TIME AGE WT NOPRINT ONEHEADER FILE=cotab26
$TABLE      ID TIME SEX DIAG Y NOPRINT ONEHEADER FILE=catab26
```

A.6 Final SCM model (run25)

```
;; 1. Based on: 11
$PROBLEM   Covariate model based on SCM
$DATA      ../data/theo_no_missing.csv IGNORE=@
$INPUT      ID TIME THEO AGE WT SEX RACE DIAG DV
```

```
$ESTIMATION METHOD=1 NOABORT MAXEVAL=9999 POSTHOC MCETA=1
$COVARIANCE
$THETA  (0,146.71) ; TV_BASE
 (0,423.76) ; TV_NORMAL
 (.01,42.4286) ; TV_T50
 (.01,10.2693) ; TV_C50
 (0,0.305535,1) ; TV_ALPHA
 (.1,4.63383,5) ; TV_HILL
$THETA  (-0.016,0.00557403,0.032) ; BASEWT1
$THETA  (-0.024,0.0226863,0.051) ; C50AGE1
$THETA  (-1,-0.496745,5) ; NORMALDIAG1
 (-1,-0.2698,5) ; NORMALDIAG2
$THETA  (-0.024,0.0442177,0.051) ; T50AGE1
$THETA  (-1,-0.780861,5) ; T50SEX1
$THETA  (-0.016,-0.0129475,0.032) ; T50WT1
$OMEGA  0.118039  ;   IIV_BASE
 0.0044012  ; IIV_NORMAL
 2.8273  ;    IIV_T50
 0.328448  ;    IIV_C50
$SIGMA  0.0249899  ;   RUV_PROP
$PRED
;;; T50WT-DEFINITION START
T50WT = ( 1 + THETA(13)*(WT - 66))
;;; T50WT-DEFINITION END


;;; T50SEX-DEFINITION START
IF(SEX.EQ.0) T50SEX = 1  ; Most common
IF(SEX.EQ.1) T50SEX = ( 1 + THETA(12))
;;; T50SEX-DEFINITION END


;;; T50AGE-DEFINITION START
T50AGE = ( 1 + THETA(11)*(AGE - 34.5))
;;; T50AGE-DEFINITION END

;;; T50-RELATION START
T50COV=T50AGE*T50SEX*T50WT
;;; T50-RELATION END


;;; NORMALDIAG-DEFINITION START
IF(DIAG.EQ.1) NORMALDIAG = 1  ; Most common
IF(DIAG.EQ.2) NORMALDIAG = ( 1 + THETA(9))
IF(DIAG.EQ.3) NORMALDIAG = ( 1 + THETA(10))
;;; NORMALDIAG-DEFINITION END

;;; NORMAL-RELATION START
NORMALCOV=NORMALDIAG
;;; NORMAL-RELATION END


;;; C50AGE-DEFINITION START
C50AGE = ( 1 + THETA(8)*(AGE - 34.5))
;;; C50AGE-DEFINITION END

;;; C50-RELATION START
C50COV=C50AGE
;;; C50-RELATION END


;;; BASEWT-DEFINITION START
BASEWT = ( 1 + THETA(7)*(WT - 66))
;;; BASEWT-DEFINITION END

;;; BASE-RELATION START
BASECOV=BASEWT
;;; BASE-RELATION END
```

```
; Individual variables (RANDOM effects)
     TVBASE   = THETA(1)

TVBASE = BASECOV*TVBASE
     BASE     = TVBASE*EXP(ETA(1))
     TVNORMAL = THETA(2)

TVNORMAL = NORMALCOV*TVNORMAL
     NORMAL   = TVNORMAL*EXP(ETA(2))
     TVT50    = THETA(3)

TVT50 = T50COV*TVT50
     T50      = TVT50*EXP(ETA(3))
     TVC50    = THETA(4)

TVC50 = C50COV*TVC50
     C50      = TVC50*EXP(ETA(4))

     IF (T50.LE.0.01) EXIT 1 3
     IF (C50.LE.0.01) EXIT 1 4

     IF (TIME.LT.0) THEN
          LTZERO=1
     ELSE
          LTZERO=0
     ENDIF

     FTIME=LTZERO+(1-LTZERO)*EXP(-LOG(2)/T50*TIME)
     BCF=(NORMAL/BASE-1)*FTIME
     FEMAX=BCF/(BCF+1)
     PEFRT=NORMAL*(1-FEMAX)
     HILL = THETA(6)
     CPN=THEO**HILL
     C50N=C50**HILL
     PEFRCT=(NORMAL-PEFRT)*CPN/(CPN+C50N)
     ALPHA=THETA(5)
     F = PEFRT + ALPHA * PEFRCT
     IPRED = F
     IRES = DV - IPRED
     IWRES = IRES/IPRED
     Y = F * ( 1 + EPS(1) )

$TABLE      ID TIME THEO BASE NORMAL T50 C50 ALPHA HILL ETA1 ETA2 ETA3
            ETA4 AGE WT SEX DIAG Y ONEHEADER NOPRINT FILE=run25.dat
$TABLE      ID TIME Y THEO IPRED IWRES NOPRINT ONEHEADER FILE=sdtab25
$TABLE      ID TIME THEO BASE NORMAL T50 C50 ALPHA HILL ETA1 ETA2 ETA3
            ETA4 NOPRINT ONEHEADER FILE=patab25
$TABLE      ID TIME AGE WT NOPRINT ONEHEADER FILE=cotab25
$TABLE      ID TIME SEX DIAG Y NOPRINT ONEHEADER FILE=catab25
```

A.7 Final FREM model (run28)

```
;; 1. Based on: 27
$PROBLEM   FREM final model (model 4)
$DATA      frem_run11/frem_dataset.dta IGNORE=@
$INPUT      ID TIME THEO AGE WT SEX RACE DIAG DV FREMTYPE
$ESTIMATION METHOD=1 NOABORT MAXEVAL=9999 POSTHOC NONINFETA=1 MCETA=1
$THETA  (0,143.207) ; TV_BASE
 (0,331.877) ; TV_NORMAL
 (.01,9.53588) ; TV_T50
 (.01,10.303) ; TV_C50
 (0,0.413845,1) ; TV_ALPHA
 (.1,3.25128,5) ; TV_HILL
```

```
$THETA  67.2257575758 FIX ; TV_WT
 38.7803030303 FIX ; TV_AGE
 1.19696969697 FIX ; TV_DIAG
 0.378787878788 FIX ; TV_SEX
$OMEGA  BLOCK(8)
 0.1632  ;   IIV_BASE
 0.00157182 0.151386  ; IIV_NORMAL
 0.00491742 0.00473609 1.48168  ;    IIV_T50
 0.0027459 0.00264465 0.00827375 0.462009  ;    IIV_C50
 0.0323345 0.0123791 -0.153176 -0.0670133 1  ;     BSV_WT
 -0.0864764 -0.217291 0.443655 0.145727 0.177913 1  ;    BSV_AGE
 -0.0624184 -0.1921 0.24098 0.153299 -0.0255564 0.572045 1  ;   BSV_DIAG
 0.0729895 -0.0119927 -0.174831 -0.00915787 0.153527 0.225522 0.32921 1  ;    BSV_SEX
$SIGMA  0.0209202  ;   RUV_PROP
$SIGMA  0.0000001  FIX  ;     EPSCOV
$PRED

; INDIVIDUAL VARIABLES (RANDOM EFFECTS)
     TVBASE   = THETA(1)
     BASE     = TVBASE*EXP(ETA(1))
     TVNORMAL = THETA(2)
     NORMAL   = TVNORMAL*EXP(ETA(2))
     TVT50    = THETA(3)
     T50      = TVT50*EXP(ETA(3))
     TVC50    = THETA(4)
     C50      = TVC50*EXP(ETA(4))

     IF (T50.LE.0.01) EXIT 1 3
     IF (C50.LE.0.01) EXIT 1 4

     IF (TIME.LT.0) THEN
          LTZERO=1
     ELSE
          LTZERO=0
     ENDIF

     FTIME=LTZERO+(1-LTZERO)*EXP(-LOG(2)/T50*TIME)
     BCF=(NORMAL/BASE-1)*FTIME
     FEMAX=BCF/(BCF+1)
     PEFRT=NORMAL*(1-FEMAX)
     HILL = THETA(6)
     CPN=THEO**HILL
     C50N=C50**HILL
     PEFRCT=(NORMAL-PEFRT)*CPN/(CPN+C50N)
     ALPHA=THETA(5)
     F = PEFRT + ALPHA * PEFRCT
     IPRED = F
     IRES = DV - IPRED
     IWRES = IRES/IPRED
     Y = F * ( 1 + EPS(1) )

;;;FREM CODE BEGIN COMPACT
;;;DO NOT MODIFY
    SDC5 = 14.7155140713
    SDC6 = 18.976206232
    SDC7 = 0.43579170631
    SDC8 = 0.486933119697
    IF (FREMTYPE.EQ.100) THEN
;       WT  14.7155140713
       Y = THETA(7) + ETA(5)*SDC5 + EPS(2)
       IPRED = THETA(7) + ETA(5)*SDC5
    END IF
    IF (FREMTYPE.EQ.200) THEN
;       AGE  18.976206232
       Y = THETA(8) + ETA(6)*SDC6 + EPS(2)
       IPRED = THETA(8) + ETA(6)*SDC6
    END IF
    IF (FREMTYPE.EQ.300) THEN
```

```
;       DIAG  0.43579170631
        Y = THETA(9) + ETA(7)*SDC7 + EPS(2)
        IPRED = THETA(9) + ETA(7)*SDC7
    END IF
    IF (FREMTYPE.EQ.400) THEN
;       SEX  0.486933119697
        Y = THETA(10) + ETA(8)*SDC8 + EPS(2)
        IPRED = THETA(10) + ETA(8)*SDC8
    END IF
;;;FREM CODE END COMPACT
$COVARIANCE UNCONDITIONAL PRECOND=1
$TABLE      ID TIME THEO BASE NORMAL T50 C50 ALPHA HILL ETA1 ETA2 ETA3
            ETA4 AGE WT SEX DIAG Y NOTITLE ONEHEADER NOPRINT
            FILE=run28.dat
$TABLE      ID TIME Y THEO IPRED IWRES NOPRINT NOTITLE ONEHEADER
            FILE=sdtab28
$TABLE      ID TIME THEO BASE NORMAL T50 C50 ALPHA HILL ETA1 ETA2 ETA3
            ETA4 NOPRINT NOTITLE ONEHEADER FILE=patab28
$TABLE      ID TIME AGE WT NOPRINT NOTITLE ONEHEADER FILE=cotab28
$TABLE      ID TIME SEX DIAG Y NOPRINT NOTITLE ONEHEADER FILE=catab28
$ETAS       FILE=frem_run11/final_models/model_4_input.phi
```

B R code

B.1 Data management

The R code used for the data management.

```
dm <- fread("data/theo_no_missing.csv") # 1. Import data
dm <- subset( dm, select = -c(RACE)) # 2. RACE removed from the analysis
dm$SEX <- as.factor(as.character( dm$SEX)) # 3. Convert numerics to factors
dm$DIAG <- as.factor(as.character( dm$DIAG))
dm$PEFR <- as.numeric( dm$PEFR) # 4. Set numbers to numerics
```

B.2 Numerical summary

Code used to generate the data summary of Table S1.

```
dm_tab <- dm %>%
  mutate( # Rename factors
    Diagnosis = ifelse( DIAG == 1, "1: Asthma",
      ifelse( DIAG == 2, "2: COPD", "3: COPD and asthma")
    ),
    Sex = ifelse( SEX == 0, "0: Female", "1: Male"),
    `Age (years)` = AGE, # Rename variables
    `Theophylline concentrations (mg/L)` = THEO,
    `Peak expiratory force rate (PEFR, L/min)` = PEFR,
    `Weight (kg)` = WT
  )

dp <- data.frame(dm_tab %>% # Extract dependent variables
    select(`Theophylline concentrations (mg/L)`,
           `Peak expiratory force rate (PEFR, L/min)`) %>%
    tbl_summary() %>%
  as_gt())

cov <- data.frame( dm_tab %>% # Extract covariates
    distinct(ID, .keep_all = TRUE) %>% # Keep 1 row per ID
    select(`Age (years)`, `Weight (kg)`, Sex, Diagnosis) %>%
    tbl_summary(digits = list(
      c('Age (years)', 'Weight (kg)', Sex, Diagnosis) ~ c(1))) %>%
  as_gt())

rb <- rbind(dp, cov) %>% # Merge covariates and dependent variable summary
  filter(stat_0 != "") %>% # Filter
  mutate(label = ifelse(row_type %in% c("label"), " ", label))

rb %>% # Create table
  select(variable, label, stat_0) %>%
  dplyr::rename("Metrics*" = stat_0,
         "Categories" = label,
         "Variables" = variable) %>%
  kable(booktabs = T,
        caption = "Numerical description of the Theophylline data set") %>%
```

```
  collapse_rows(columns = 1,
                latex_hline = "none",
                valign = "top") %>%
  kable_styling(latex_options = "H") %>%
  pack_rows("Dependent variables (526 observations)", 1, 2) %>%
  pack_rows("Continuous covariates (132 subjects)", 3, 4, hline_before = T) %>%
  pack_rows("Categorical covariates (132 subjects)", 5, 9, hline_before = T) %>%
  kableExtra::footnote(
    symbol = "Median (inter-quartile range) for continuous variables; n (%) for categorical variables."
    symbol_title = "")
```

B.3 Theophylline data file details

Table S9: Description of the Theophylline data file.

Column	Description	Unit	Values
ID	Unique subject identifier	-	Integers > 0
TIME	Time	hours	Numerics > 0
THEO	Theophylline serum concentration	mg/mL	Numerics > 0
AGE	Age	years	Integers > 1
WT	Body weight	kg	Numerics > 1
SEX	Sexe	-	0: Female, 1: Male
DIAG	Diagnosis (categorical scale)	-	1: Asthma, 2: COPD, 3: COPD and asthma
PEFR	Peak expiratory flow rate	L/min	Numerics > 0

COPD: chronic obstructive pulmonary disease

Table S10: First ten rows of the Theophylline data file.

ID	TIME	THEO	AGE	WT	SEX	DIAG	PEFR
1	0.01	1	27	57	0	1	200
1	1.51	20.7	27	57	0	1	300
2	0.01	0.1	21	76	0	1	110
2	1.42	25.5	21	76	0	1	200
2	4.42	26.2	21	76	0	1	325
3	0.01	0.1	26	67.5	1	1	150

B.4 Spaghettis plots

```
ggplot(dm, aes(y = PEFR, x = THEO)) +
  geom_point(aes(group = ID))  +
  geom_line(aes(group = ID))  +
  geom_smooth(se = FALSE, color = "red") +
  xlab("Observed theophylline concentrations (mg/mL)") + ylab("Observed PEFR (L/min)") +
  theme(legend.position = "none") +
  theme_bw()
```

B.5 Scatter plot matrix

```
dmn <- dm %>%
  mutate( # Rename factors
    Diagnosis = ifelse( DIAG == 1, "Asthma",
      ifelse( DIAG == 2, "COPD", "COPD and asthma")),
    Sex = ifelse( SEX == 0, "Female", "Male"),
    Weight = WT, # Rename covariates
    Age = AGE
  )

cov <- dmn[, c("Age", "Weight", "Sex", "Diagnosis")] # Select covariates columns

ggpairs(cov,
  upper = list(continuous = "cor", combo = "box_no_facet"),
  lower = list(continuous = "smooth", combo = "dot_no_facet")) +
  theme_bw() +
  theme(axis.text.x = element_text( angle=45, hjust=1 ) )
```

B.6 Goodness-of-fit plots

B.7 Covariates vs EBEs

```
#Loading the data
input <- fread("models/run11.dat") # NM output table

### Data management
input <- input[!duplicated(input$ID) , ] # Keep one line per ID

names(input)[names(input) == "AGE"] <- "Age" #Rename interesting columns
names(input)[names(input) == "WT"] <- "Weight"
names(input)[names(input) == "SEX"] <- "Sex"
names(input)[names(input) == "DIAG"] <- "Diagnosis"

ebes <- # Keep only relevant columns
  input[, c("ID", "ETA1", "ETA2", "ETA3", "ETA4", "Age", "Weight", "Sex", "Diagnosis")]
ebes <- melt( # Get long format
  ebes,
  id = c("ID", "Age", "Weight", "Sex", "Diagnosis"),
  variable.name = "ETA",
  value.name = "EBEs"
)
ebes <- melt( # Get long format
  ebes,
  id = c("ID", "EBEs", "ETA"),
  variable.name = "COV",
  value.name = "Covariates"
)

# Change name for plot
ebes$ETA <- as.character(ebes$ETA)
```

```
ebes$ETA[ ebes$ETA %in% c("ETA1") ] <- "Base"
ebes$ETA[ ebes$ETA %in% c("ETA2") ] <- "Normal"
ebes$ETA[ ebes$ETA %in% c("ETA3") ] <- "T50"
ebes$ETA[ ebes$ETA %in% c("ETA4") ] <- "C50"

ebes$ETA <- as.factor(ebes$ETA)
levels(ebes$ETA) <- levels(ebes$ETA)[c(1,3,2,4)]

sum <- xpose::get_summary(xpdb11) # Get etas values
etas <- sum$value[sum$descr %in% c("Eta shrinkage")]
etas <- as.numeric(gsub(" .*", "", unlist(strsplit(etas, "], "))))

# Add the shrinkage information
ebes$Shrinkage <- NA
ebes$Shrinkage[ebes$ETA %in% c("Base")] <- paste0("S(SD)=", etas[1],"%")
ebes$Shrinkage[ebes$ETA %in% c("Normal")] <- paste0("S(SD)=", etas[2],"%")
ebes$Shrinkage[ebes$ETA %in% c("T50")] <- paste0("S(SD)=", etas[3],"%")
ebes$Shrinkage[ebes$ETA %in% c("C50")] <- paste0("S(SD)=", etas[4],"%")

# Divide into continuous and categorical cov
cont <- ebes[ ebes$COV %in% c("Age", "Weight"),]
cat <- ebes[!ebes$COV %in% c("Age", "Weight"), ]

# Create and rename factors
cat$Covariates[ cat$COV %in% c("Sex") &
      cat$Covariates == 0 ] <- "Female"
cat$Covariates[ cat$COV %in% c("Sex") &
      cat$Covariates == 1 ] <- "Male"
cat$Covariates[ cat$COV %in% c("Diagnosis") &
      cat$Covariates == 1 ] <- "Asthma"
cat$Covariates[ cat$COV %in% c("Diagnosis") &
      cat$Covariates == 2 ] <- "COPD"
cat$Covariates[ cat$COV %in% c("Diagnosis") &
      cat$Covariates == 3 ] <- "COPD and asthma"
cat$Covariates <- as.factor(cat$Covariates)

# Plot
contp <- ggplot(cont, aes(y = EBEs, x = Covariates)) +
  geom_point() +
  facet_grid(ETA ~ COV, scales = "free") +
  geom_smooth(method = "lm",
              colour = "red",
              se = FALSE) +
  xlab("Continuous covariates") +
  theme_bw() +
  theme( strip.background.y = element_blank(),
         strip.text.y = element_blank())

catp <- ggplot(cat, aes(y = EBEs, x = Covariates)) +
  geom_boxplot() +
  facet_grid(ETA + Shrinkage ~ COV, scales = "free") +
  xlab("Categorical covariates") +
  ylab(NULL) +
```

```
  theme_bw() +
  theme(axis.text.x = element_text(angle = 35, vjust = 1, hjust = 1))

plot1 <- plot_grid(contp, catp, nrow = 1, align = "h")
```

B.8 Covariates vs η samples

```
#Loading the data
input <- fread("models/run11.dat") # NM output table
phi <- fread( "models/run11.phi" ) # phi file

### Data management
input <- input[!duplicated(input$ID) , ] # Keep one line per ID

names(input)[names(input) == "AGE"] <- "Age" #Rename interesting columns
names(input)[names(input) == "WT"] <- "Weight"
names(input)[names(input) == "SEX"] <- "Sex"
names(input)[names(input) == "DIAG"] <- "Diagnosis"

names(phi) <- gsub("[()]|,", "", names(phi)) # Remove parenthesis and coma from header

# Get eta samples from the individual distributions of the phi file

fun <- function(v1, v2) {
  val <-  rnorm(1,v1,v2)
  return(val)
}

phi <- phi %>%
    mutate( ETAS1 = purrr::pmap_dbl(list( ETA1, ETC11), fun ),
            ETAS2 = purrr::pmap_dbl(list( ETA2, ETC22), fun ),
            ETAS3 = purrr::pmap_dbl(list( ETA3, ETC33), fun ),
            ETAS4 = purrr::pmap_dbl(list( ETA4, ETC44), fun ) )

phi <- merge( input, phi[,c(2,18:21)] ) # Merge etas samples and covariates

etas <- # Keep only relevant columns
  phi[, c("ID", "ETAS1", "ETAS2", "ETAS3", "ETAS4", "Age", "Weight", "Sex", "Diagnosis")]
etas <- melt( # Get long format: stack ETAS
  etas,
  id = c("ID", "Age", "Weight", "Sex", "Diagnosis"),
  variable.name = "ETAS",
  value.name = "Samples"
)
etas <- melt( # Get long format: stack covariates
  etas,
  id = c("ID", "ETAS", "Samples"),
  variable.name = "COV",
  value.name = "Covariates"
)
```

```r
# Change name for plot
etas$ETAS <- as.character(etas$ETAS)

etas$ETAS[ etas$ETAS %in% c("ETAS1") ] <- "Base"
etas$ETAS[ etas$ETAS %in% c("ETAS2") ] <- "Normal"
etas$ETAS[ etas$ETAS %in% c("ETAS3") ] <- "T50"
etas$ETAS[ etas$ETAS %in% c("ETAS4") ] <- "C50"

etas$ETAS <- as.factor(etas$ETAS)
levels(etas$ETAS) <- levels(etas$ETAS)[c(1,3,2,4)]

sum <- xpose::get_summary(xpdb11) # Get etas values
eta_s <- sum$value[sum$descr %in% c("Eta shrinkage")]
eta_s <- as.numeric(gsub(" .*", "", unlist(strsplit(eta_s, "], "))))

# Add the shrinkage information
etas$Shrinkage <- NA
etas$Shrinkage[etas$ETAS %in% c("Base")] <- paste0("S(SD)=", eta_s[1],"%")
etas$Shrinkage[etas$ETAS %in% c("Normal")] <- paste0("S(SD)=", eta_s[2],"%")
etas$Shrinkage[etas$ETAS %in% c("T50")] <- paste0("S(SD)=", eta_s[3],"%")
etas$Shrinkage[etas$ETAS %in% c("C50")] <- paste0("S(SD)=", eta_s[4],"%")

# Divide into continuous and categorical cov
conts <- etas[ etas$COV %in% c("Age", "Weight"),]
cats <- etas[ etas$COV %in% c("Sex", "Diagnosis"), ]

# Create and rename factors
cats$Covariates[ cats$COV %in% c("Sex") &
      cats$Covariates == 0 ] <- "Female"
cats$Covariates[ cats$COV %in% c("Sex") &
      cats$Covariates == 1 ] <- "Male"
cats$Covariates[ cats$COV %in% c("Diagnosis") &
      cats$Covariates == 1 ] <- "Asthma"
cats$Covariates[ cats$COV %in% c("Diagnosis") &
      cats$Covariates == 2 ] <- "COPD"
cats$Covariates[ cats$COV %in% c("Diagnosis") &
      cats$Covariates == 3 ] <- "COPD and asthma"
cats$Covariates <- as.factor(cats$Covariates)

# Plot
contps <- ggplot(conts, aes(y = Samples, x = Covariates)) +
  geom_point() +
  facet_grid(ETAS ~ COV, scales = "free") +
  geom_smooth(method = "lm",
              colour = "red",
              se = FALSE) +
  xlab("Continuous covariates") +
  ylab("Eta samples") +
  theme_bw() +
  theme( strip.background.y = element_blank(),
         strip.text.y = element_blank())

catps <- ggplot(cats, aes(y = Samples, x = Covariates)) +
```

```
  geom_boxplot() +
  facet_grid(ETAS + Shrinkage ~ COV, scales = "free") +
  xlab("Categorical covariates") +
  ylab(NULL) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 35, vjust = 1, hjust = 1))

plot2 <- plot_grid(contps, catps, nrow = 1, align = "h")
```

B.9 GAM plots

```
aic_plot <- xpose4::xp.akaike.plot(gamobj = xpgam) # Akaike plot
res_plot <- xpose4::xp.plot(gamobj = xpgam) # Residuals plot
```

B.10 WAM

- Function used to compute the Wald's approximation of the LRT statistic for all the covariate submodels of the FFEM.

```
wam <- function(k, # no. of covariate thetas in FFEM
                p, # no. of parameters in theta, omega and sigma in full model
                n, # no. of observations in data set
                theta, # Final fixed covariate effect estimates
                cov # Fixed covariate effect estimates var-covar error matrix
                ) {

  # Create a matrix with all possible covariates prm combinations
  kk    <- 2 ^ ((k - 1):0)
  maker <-
    paste("rep(rep(c(0,1),rep(" , kk , ",2))," , rev(kk) , ")" , sep = "")
  x <- apply(matrix(maker), 1, function(x)
    eval(parse(text = x)))
  kkk   <- dim(x)[1] # Nb of columns in the matrix = fixed cov effect
  # i = 1
  ireg <- rep(1, k)
  idel <- 1:k
  c2     <- cov[idel, idel]
  theta2 <- t(theta[idel])
  lrt    <- t(theta2) %*% solve(c2) %*% theta2
  s      <- p - length(idel)
  sbc    <- -0.5 * (lrt + s * log(n))
  results <- data.frame(
    i    = 1:kkk,
    ireg = rep(0, kkk),
    lrt  = rep(0, kkk),
    sbc  = rep(0, kkk)
  )
  results[1, 2:4] <- c(paste(ireg, collapse = ","), lrt, sbc)
  for (i in 2:(kkk - 1))
  {
```

```
    ireg <- (1:k)[x[i, ] > 0]
    idel <- (1:k)[x[i, ] == 0]
    c2     <- cov[idel, idel]
    theta2 <- t(theta[idel])
    lrt    <- t(theta2) %*% solve(c2) %*% theta2
    s      <- p - length(idel)
    sbc    <- -0.5 * (lrt + s * log(n))
    results[i, 2:4] <- c(paste(ireg, collapse = ","), lrt, sbc)
  }
  # i = kkk
  ireg <- (1:k)
  lrt    <- 0
  sbc    <- -0.5 * (lrt + p * log(n))
  results[kkk, 2:4] <- c(paste(ireg, collapse = ","), lrt, sbc)
  results <- results[order(-as.numeric(results$sbc)), ]
  results[, "i"] <- 1:kkk
  results
}
```

- Extraction of the variance-covariance error matrix and final effect estimates of the FFEM for the fixed effect covariate parameters.

```
# Information from the FFEM
ntheta <- 14    # no. of thetas in FFEM
ncovtheta <- 8  # no. of covariate thetas in FFEM
run.no <- 22      # your run number
# Get the variance-covariance error matrix for fixed covariates effects
tc <- # Import NONMEM .cov file of FFEM
  read.table( # .cov file = full variance-covariance error matrix of thetas, sigmas and omegas
    file = paste("models/run", run.no, ".cov", sep = ""),
    skip = 1,
    header = 1,
    row.names = 1
  )
fthetas <- ntheta - ncovtheta # Nb of thetas that are not covariate effects
tcnames <- names(tc)[(tc != 0)[1, ]][(fthetas+1):ntheta] # Names of that thetas

tc <- matrix( # Var cov matrix with only estimated parameters
  tc[tc != 0], ncol = sqrt(length(tc[tc != 0]))
  )
cov <- # Subset fixed covariates effects and add names
  tc[(fthetas+1):ntheta, (fthetas+1):ntheta]
cov[col(cov) > row(cov)] <- 0
cov <- cov + t(cov) - unlist(diag(unlist((diag(cov)))))
dimnames(cov) <- list(theta = tcnames,
                      theta = tcnames)

# Get final estimates for fixed covariates effects
tp <- read.table( # Import NONMEM .ext file of FFEM
   # Raw output file with numerical NONMEM results
    file = paste("models/run", run.no, ".ext", sep = ""),
    skip = 1,
    header = 1
```

```
  )
tp <- # .ext line with final estimates and OFV
  tp[tp$ITERATION == -1000000000, ]
tp <- tp[, tp != 0] # Only estimated parameters
theta <- tp[(1 + fthetas + 1):(1 + ntheta)] # Only fixed covariates effects
thetas <- unlist(strsplit(names(theta), "THETA")) # Keep only the theta number
thetas <- thetas[thetas != ""]
```

- Running the WAM algorithm on the information from the FFEM.

```
results <- wam( # Run the WAM method on the FFEM (run22)
   k = ncovtheta, # no. of covariate thetas in FFEM
   p = 19, # no. of parameters in theta, omega and sigma in full model
   n = 526, # no. of observations in data set
   theta = theta, # Final fixed covariate effect estimates
   cov = cov # Fixed covariate effect estimates var-covar error matrix
   )
```

B.11 SCM

B.11.1 OFV drop plot for SCM

```
scm.short.log <- 'models/scm_run14/short_scmlog.txt'

list_incl_cov <- included_covariates(scm.short.log)
n1 <-
  list_incl_cov$n1 #Number of covariate included after forward step
sign <- list_incl_cov$sign
data <- list_incl_cov$data
name <- list_incl_cov$name

# plot SCM results included covariates
p <- plot_included_covariates(data, sign, n1, name) +
  theme_bw() +
  ggtitle(NULL) +
  theme(axis.text.x = element_text(angle = 45, hjust=1))
print(p)
```

B.11.2 SCM text summary output

```
## Model directory /green/home/USER/estch718/isop_covariates/models/scm_run14/m1
##
## MODEL            TEST     BASE OFV     NEW OFV         TEST OFV (DROP)    GOAL     dDF    SIGNIFICANT PVAL
## BASEAGE-2        PVAL   4962.29196   4952.05523             10.23673  >   3.84150    1        YES!  0.001377
## BASEDIAG-2       PVAL   4962.29196   4954.45130              7.84066  >   5.99150    2        YES!  0.019835
## BASESEX-2        PVAL   4962.29196   4955.05010              7.24186  >   3.84150    1        YES!  0.007122
## BASEWT-2         PVAL   4962.29196   4957.62887              4.66308  >   3.84150    1        YES!  0.030818
## C50AGE-2         PVAL   4962.29196   4898.86897             63.42299  >   3.84150    1        YES!  1.67e-15
## C50DIAG-2        PVAL   4962.29196   4937.24279             25.04916  >   5.99150    2        YES!  0.000004
## C50SEX-2         PVAL   4962.29196   4959.22473              3.06723  >   3.84150    1              0.079886
## C50WT-2          PVAL   4962.29196   4959.05895              3.23301  >   3.84150    1              0.072168
## NORMALAGE-2      PVAL   4962.29196   4904.83901             57.45295  >   3.84150    1        YES!  3.46e-14
```

```
## NORMALDIAG-2     PVAL   4962.29196   4883.73491            78.55705  >   5.99150    2        YES!  8.74e-18
## NORMALSEX-2      PVAL   4962.29196   4962.13938             0.15257  >   3.84150    1             0.696090
## NORMALWT-2       PVAL   4962.29196   4958.22705             4.06491  >   3.84150    1        YES!  0.043783
## T50AGE-2         PVAL   4962.29196   4932.99364            29.29832  >   3.84150    1        YES!  6.20e-08
## T50DIAG-2        PVAL   4962.29196   4927.98398            34.30798  >   5.99150    2        YES!  3.55e-08
## T50SEX-2         PVAL   4962.29196   4958.15679             4.13516  >   3.84150    1        YES!  0.042001
## T50WT-2          PVAL   4962.29196   4956.29366             5.99829  >   3.84150    1        YES!  0.014320
##
## Parameter-covariate relation chosen in this forward step: NORMAL-DIAG-2
## CRITERION              PVAL < 0.05
## BASE_MODEL_OFV        4962.29196
## CHOSEN_MODEL_OFV      4883.73491
## Relations included after this step:
## BASE
## C50
## NORMAL  DIAG-2
## T50
## --------------------
##
## Model directory /green/home/USER/estch718/isop_covariates/models/scm_run14/forward_scm_dir1/m1
##
## MODEL            TEST     BASE OFV     NEW OFV         TEST OFV (DROP)    GOAL     dDF    SIGNIFICANT PVAL
## BASEAGE-2        PVAL   4883.73491   4877.07097             6.66394  >   3.84150    1        YES!  0.009838
## BASEDIAG-2       PVAL   4883.73491   4871.76993            11.96498  >   5.99150    2        YES!  0.002523
## BASESEX-2        PVAL   4883.73491   4872.96706            10.76786  >   3.84150    1        YES!  0.001033
## BASEWT-2         PVAL   4883.73491   4879.35381             4.38110  >   3.84150    1        YES!  0.036340
## C50AGE-2         PVAL   4883.73491   4845.35599            38.37893  >   3.84150    1        YES!  5.83e-10
## C50DIAG-2        PVAL   4883.73491   4881.28372             2.45119  >   5.99150    2             0.293580
## C50SEX-2         PVAL   4883.73491   4880.80471             2.93020  >   3.84150    1             0.086937
## C50WT-2          PVAL   4883.73491   4881.35728             2.37763  >   3.84150    1             0.123080
## NORMALAGE-2      PVAL   4883.73491   4870.69693            13.03798  >   3.84150    1        YES!  0.000305
## NORMALSEX-2      PVAL   4883.73491   4866.78023            16.95468  >   3.84150    1        YES!  0.000038
## NORMALWT-2       PVAL   4883.73491   4878.51777             5.21714  >   3.84150    1        YES!  0.022365
## T50AGE-2         PVAL   4883.73491   4860.40668            23.32823  >   3.84150    1        YES!  0.000001
## T50DIAG-2        PVAL   4883.73491   4881.29294             2.44197  >   5.99150    2             0.294940
## T50SEX-2         PVAL   4883.73491   4866.27328            17.46164  >   3.84150    1        YES!  0.000029
## T50WT-2          PVAL   4883.73491   4879.11279             4.62212  >   3.84150    1        YES!  0.031562
##
## Parameter-covariate relation chosen in this forward step: C50-AGE-2
## CRITERION              PVAL < 0.05
## BASE_MODEL_OFV        4883.73491
## CHOSEN_MODEL_OFV      4845.35599
## Relations included after this step:
## BASE
## C50     AGE-2
## NORMAL  DIAG-2
## T50
## --------------------
##
## Model directory /green/home/USER/estch718/isop_covariates/models/scm_run14/forward_scm_dir1/scm_dir1/m1
##
## MODEL            TEST     BASE OFV     NEW OFV         TEST OFV (DROP)    GOAL     dDF    SIGNIFICANT PVAL
## BASEAGE-2        PVAL   4845.35599   4848.65200            -3.29602  >   3.84150    1                  9999
## BASEDIAG-2       PVAL   4845.35599   4850.96014            -5.60415  >   5.99150    2                  9999
## BASESEX-2        PVAL   4845.35599   4840.26182             5.09417  >   3.84150    1        YES!  0.024006
## BASEWT-2         PVAL   4845.35599   4843.55534             1.80065  >   3.84150    1             0.179630
## C50AGE-3         PVAL   4845.35599   4848.54735            -3.19137  >   3.84150    1                  9999
## C50DIAG-2        PVAL   4845.35599   4848.62926            -3.27327  >   5.99150    2                  9999
## C50SEX-2         PVAL   4845.35599   4850.38852            -5.03253  >   3.84150    1                  9999
## C50WT-2          PVAL   4845.35599   4846.56889            -1.21291  >   3.84150    1                  9999
## NORMALAGE-2      PVAL   4845.35599   4840.72917             4.62682  >   3.84150    1        YES!  0.031476
## NORMALSEX-2      PVAL   4845.35599   4838.39465             6.96134  >   3.84150    1        YES!  0.008329
## NORMALWT-2       PVAL   4845.35599   4850.53065            -5.17467  >   3.84150    1                  9999
## T50AGE-2         PVAL   4845.35599   4841.77384             3.58215  >   3.84150    1             0.058404
## T50DIAG-2        PVAL   4845.35599   4846.75786            -1.40187  >   5.99150    2                  9999
## T50SEX-2         PVAL   4845.35599   4836.68161             8.67438  >   3.84150    1        YES!  0.003227
## T50WT-2          PVAL   4845.35599   4848.67600            -3.32002  >   3.84150    1                  9999
##
```

```
## Parameter-covariate relation chosen in this forward step: T50-SEX-2
## CRITERION              PVAL < 0.05
## BASE_MODEL_OFV        4845.35599
## CHOSEN_MODEL_OFV      4836.68161
## Relations included after this step:
## BASE
## C50     AGE-2
## NORMAL  DIAG-2
## T50     SEX-2
## --------------------
##
## Model directory /green/home/USER/estch718/isop_covariates/models/scm_run14/forward_scm_dir1/scm_dir1/scm_dir1/m1
##
## MODEL            TEST     BASE OFV     NEW OFV         TEST OFV (DROP)    GOAL     dDF    SIGNIFICANT PVAL
## BASEAGE-2        PVAL   4836.68161   4825.92119             10.76042  >   3.84150    1        YES!  0.001037
## BASEDIAG-2       PVAL   4836.68161   4836.10614              0.57547  >   5.99150    2              0.749960
## BASESEX-2        PVAL   4836.68161   4822.16658             14.51502  >   3.84150    1        YES!  0.000139
## BASEWT-2         PVAL   4836.68161   4826.67158             10.01003  >   3.84150    1        YES!  0.001557
## C50AGE-3         PVAL   4836.68161   4841.00847             -4.32687  >   3.84150    1                 9999
## C50DIAG-2        PVAL   4836.68161   4832.32558              4.35602  >   5.99150    2              0.113270
## C50SEX-2         PVAL   4836.68161   4832.13665              4.54496  >   3.84150    1        YES!  0.033016
## C50WT-2          PVAL   4836.68161   4830.37802              6.30359  >   3.84150    1        YES!  0.012049
## NORMALAGE-2      PVAL   4836.68161   4821.43645             15.24515  >   3.84150    1        YES!  0.000094
## NORMALSEX-2      PVAL   4836.68161   4836.79028             -0.10867  >   3.84150    1                 9999
## NORMALWT-2       PVAL   4836.68161   4830.99661              5.68499  >   3.84150    1        YES!  0.017111
## T50AGE-2         PVAL   4836.68161   4817.49657             19.18504  >   3.84150    1        YES!  0.000012
## T50DIAG-2        PVAL   4836.68161   4832.57805              4.10355  >   5.99150    2              0.128510
## T50WT-2          PVAL   4836.68161   4832.42314              4.25847  >   3.84150    1        YES!  0.039055
##
## Parameter-covariate relation chosen in this forward step: T50-AGE-2
## CRITERION              PVAL < 0.05
## BASE_MODEL_OFV        4836.68161
## CHOSEN_MODEL_OFV      4817.49657
## Relations included after this step:
## BASE
## C50     AGE-2
## NORMAL  DIAG-2
## T50     AGE-2            SEX-2
## --------------------
##
## Model directory /green/home/USER/estch718/isop_covariates/models/scm_run14/forward_scm_dir1/scm_dir1/scm_dir1/scm_dir1/m1
##
## MODEL            TEST     BASE OFV     NEW OFV         TEST OFV (DROP)    GOAL     dDF    SIGNIFICANT PVAL
## BASEAGE-2        PVAL   4817.49657   4812.03460              5.46197  >   3.84150    1        YES!  0.019435
## BASEDIAG-2       PVAL   4817.49657   4812.56759              4.92898  >   5.99150    2              0.085052
## BASESEX-2        PVAL   4817.49657   4808.11111              9.38545  >   3.84150    1        YES!  0.002187
## BASEWT-2         PVAL   4817.49657   4806.52627             10.97030  >   3.84150    1        YES!  0.000926
## C50AGE-3         PVAL   4817.49657   4811.26331              6.23326  >   3.84150    1        YES!  0.012537
## C50DIAG-2        PVAL   4817.49657   4810.21025              7.28632  >   5.99150    2        YES!  0.026170
## C50SEX-2         PVAL   4817.49657   4812.85533              4.64124  >   3.84150    1        YES!  0.031213
## C50WT-2          PVAL   4817.49657   4810.32347              7.17310  >   3.84150    1        YES!  0.007400
## NORMALAGE-2      PVAL   4817.49657   4810.18411              7.31245  >   3.84150    1        YES!  0.006848
## NORMALSEX-2      PVAL   4817.49657   4811.63157              5.86499  >   3.84150    1        YES!  0.015445
## NORMALWT-2       PVAL   4817.49657   4812.08371              5.41286  >   3.84150    1        YES!  0.019989
## T50AGE-3         PVAL   4817.49657   4808.20832              9.28825  >   3.84150    1        YES!  0.002306
## T50DIAG-2        PVAL   4817.49657   4810.30762              7.18895  >   5.99150    2        YES!  0.027475
## T50WT-2          PVAL   4817.49657   4807.36205             10.13451  >   3.84150    1        YES!  0.001455
##
## Parameter-covariate relation chosen in this forward step: BASE-WT-2
## CRITERION              PVAL < 0.05
## BASE_MODEL_OFV        4817.49657
## CHOSEN_MODEL_OFV      4806.52627
## Relations included after this step:
## BASE    WT-2
## C50     AGE-2
## NORMAL  DIAG-2
## T50     AGE-2            SEX-2
## --------------------
```

```
##
## Model directory /green/home/USER/estch718/isop_covariates/models/scm_run14/forward_scm_dir1/scm_dir1/scm_dir1/scm_dir1/scm_di
##
## MODEL            TEST     BASE OFV     NEW OFV         TEST OFV (DROP)    GOAL     dDF    SIGNIFICANT PVAL
## BASEAGE-2        PVAL   4806.52627   4804.62739              1.89888  >   3.84150    1              0.168200
## BASEDIAG-2       PVAL   4806.52627   4806.26918              0.25709  >   5.99150    2              0.879370
## BASESEX-2        PVAL   4806.52627   4803.94117              2.58510  >   3.84150    1              0.107870
## BASEWT-3         PVAL   4806.52627   4806.20029              0.32598  >   3.84150    1              0.568040
## C50AGE-3         PVAL   4806.52627   4821.39555            -14.86928  >   3.84150    1                   9999
## C50DIAG-2        PVAL   4806.52627   4803.09786              3.42841  >   5.99150    2              0.180110
## C50SEX-2         PVAL   4806.52627   4806.13601              0.39027  >   3.84150    1              0.532160
## C50WT-2          PVAL   4806.52627   4804.95121              1.57506  >   3.84150    1              0.209470
## NORMALAGE-2      PVAL   4806.52627   4804.05143              2.47484  >   3.84150    1              0.115680
## NORMALSEX-2      PVAL   4806.52627   4806.35661              0.16966  >   3.84150    1              0.680410
## NORMALWT-2       PVAL   4806.52627   4805.41412              1.11215  >   3.84150    1              0.291620
## T50AGE-3         PVAL   4806.52627   4814.38508             -7.85881  >   3.84150    1                   9999
## T50DIAG-2        PVAL   4806.52627   4804.45764              2.06864  >   5.99150    2              0.355470
## T50WT-2          PVAL   4806.52627   4802.43554              4.09073  >   3.84150    1        YES!  0.043119
##
## Parameter-covariate relation chosen in this forward step: T50-WT-2
## CRITERION              PVAL < 0.05
## BASE_MODEL_OFV        4806.52627
## CHOSEN_MODEL_OFV      4802.43554
## Relations included after this step:
## BASE    WT-2
## C50     AGE-2
## NORMAL  DIAG-2
## T50     AGE-2            SEX-2            WT-2
## --------------------
##
## Model directory /green/home/USER/estch718/isop_covariates/models/scm_run14/forward_scm_dir1/scm_dir1/scm_dir1/scm_dir1/scm_di
##
## MODEL            TEST     BASE OFV     NEW OFV         TEST OFV (DROP)    GOAL     dDF    SIGNIFICANT PVAL
## BASEAGE-2        PVAL   4802.43554   4802.08290              0.35264  >   3.84150    1              0.552620
## BASEDIAG-2       PVAL   4802.43554   4802.77135             -0.33581  >   5.99150    2                   9999
## BASESEX-2        PVAL   4802.43554   4801.01227              1.42327  >   3.84150    1              0.232870
## BASEWT-3         PVAL   4802.43554   4798.94602              3.48952  >   3.84150    1              0.061759
## C50AGE-3         PVAL   4802.43554   4802.13061              0.30494  >   3.84150    1              0.580800
## C50DIAG-2        PVAL   4802.43554   4799.46007              2.97547  >   5.99150    2              0.225880
## C50SEX-2         PVAL   4802.43554   4802.25033              0.18521  >   3.84150    1              0.666930
## C50WT-2          PVAL   4802.43554   4802.28782              0.14772  >   3.84150    1              0.700720
## NORMALAGE-2      PVAL   4802.43554   4802.18536              0.25018  >   3.84150    1              0.616950
## NORMALSEX-2      PVAL   4802.43554   4803.01897             -0.58343  >   3.84150    1                   9999
## NORMALWT-2       PVAL   4802.43554   4803.03235             -0.59681  >   3.84150    1                   9999
## T50AGE-3         PVAL   4802.43554   4799.18664              3.24890  >   3.84150    1              0.071472
## T50DIAG-2        PVAL   4802.43554   4800.51755              1.91799  >   5.99150    2              0.383280
## T50WT-3          PVAL   4802.43554   4802.02665              0.40889  >   3.84150    1              0.522530
##
## --------------------
##
##
## --------------------
## Forward search done. Starting backward search inside forward top level directory
## Model directory /green/home/USER/estch718/isop_covariates/models/scm_run14/backward_scm_dir1/m1
##
## MODEL            TEST     BASE OFV     NEW OFV         TEST OFV (DROP)    GOAL     dDF  INSIGNIFICANT PVAL
## BASEWT-1         PVAL   4802.43554   4816.49239            -14.05684  >  -6.63490   -1              0.000177
## C50AGE-1         PVAL   4802.43554   4826.36049            -23.92495  >  -6.63490   -1              0.000001
## NORMALDIAG-1     PVAL   4802.43554   4874.50948            -72.07394  >  -9.21030   -2              2.24e-16
## T50AGE-1         PVAL   4802.43554   4845.45767            -43.02213  >  -6.63490   -1              5.41e-11
## T50SEX-1         PVAL   4802.43554   4848.49816            -46.06262  >  -6.63490   -1              1.15e-11
## T50WT-1          PVAL   4802.43554   4819.83165            -17.39611  >  -6.63490   -1              0.000030
##
## --------------------
```

B.12 FREM

B.12.1 Correlation matrix

```
xpdb28 <- xpose::xpose_data(runno = 28, dir="models") # Read model output
prm28 <- xpose::get_prm(xpdb28) # Get final estimates

omg28 <- data.frame( prm28 %>% # Extract omega values
                      filter(type %in% c("ome")) %>%
                      select(value, name))

lab <- data.frame( prm28 %>% # Get omega matrix labels
  filter(type %in% c("ome"),
         !label %in% "" ) %>%
         select(label))
labs <- as.character(unique(lab$label))

mat <- matrix(0, nrow = 8, ncol = 8) # Create empty omega matrix
mat[upper.tri(mat, diag = TRUE)] <- omg28$value # Fill upper triangle
mat <- t(mat) # Inverse matrix to get lower triangle
rownames(mat) <- lab$label # Assign names to the matrix
colnames(mat) <- lab$label
diag(mat) <- gsub(".*_", "", lab$label) # Assign names to the diagonal

dfmat <- melt(mat) # Melt for plotting
dfmat$text <- as.numeric(dfmat$value) # Create info for plot text
dfmat$text <- ifelse(is.na(dfmat$text), dfmat$value,
                     ifelse(dfmat$text == 0, "", round(dfmat$text, 2)))
dfmat$text <- ifelse( # Prettify for long name
  dfmat$text %in% c("NORMAL"), "      NORMAL", dfmat$text)
dfmat$value <- as.numeric(dfmat$value)

ggplot(dfmat, aes(x = Var2, y = Var1, fill = value)) +
  geom_tile(color = "white") +
  geom_text(aes(label = text)) +
  geom_regon(aes(x0 = 2.5, y0 = 6.5, sides = 4, angle = 0, r = 2.8,
                 color="Parameter-parameter correlation"), size=0.5, alpha=0)+
  geom_regon(aes(x0 = 2.5, y0 = 2.5, sides = 4, angle = 0, r = 2.8,
                 color="Parameter-covariate correlation"), size=0.5, alpha=0)+
  geom_regon(aes(x0 = 6.5, y0 = 2.5, sides = 4, angle = 0, r = 2.8,
                 color="Covariate-covariate correlation"), size=0.5, alpha=0)+
  scale_fill_gradient2( na.value = "white",
    low = "blue", high = "red", mid = "white",
    midpoint = 0, limit = c(-1.01, 1.01)) +
  theme_minimal() +
  # coord_fixed() +
  scale_x_discrete(limits=labs) + # Flip the x axis
  scale_y_discrete(limits=labs[length(labs):1]) +
  theme(
    axis.text = element_blank(),
    axis.title.x = element_blank(),
    axis.title.y = element_blank(),
    panel.grid.major = element_blank(),
```

```
    panel.border = element_blank(),
    panel.background = element_blank(),
    axis.ticks = element_blank(),
    # legend.position = "top"#,
    # legend.direction = "horizontal"
  ) +
  guides(fill = guide_colorbar(
    # barwidth = 7,
    barheight = 4,
    title.position = "top",
    title.hjust = 0.5,
    title = "Correlation value"),
  color = guide_legend(""))
```

B.12.2 Forest plot

- Import data from PsN output folder

```
frem <- read.csv( # Get post processing info from PsN after running frem
    "models/frem_run11/results.csv",
    fill = TRUE,
    col.names = paste0("V", seq_len(8)),
    header = F
  )

# Data management
frem$tbl_id <- # Add unique ID for the different tables
  cumsum(frem$V2 == "")

cov_coef <- # Select lines with info for forest plots
  frem[frem$tbl_id == 3, ] %>%
  slice(-1) %>%
  select(1:6)
names(cov_coef) <-
  lapply(cov_coef[1,], as.character) # rename columns

cov_unex <- # Select lines with info for unexplained variability plots
  frem[frem$tbl_id == 5, ]
names(cov_unex) <-
  lapply(cov_unex[2,], as.character) # rename columns
cov_unex <- cov_unex[-c(1, 2), ]

cov_stat <- # Select lines with cov stat info
  frem[frem$tbl_id == 6, ]
names(cov_stat) <-
  lapply(cov_stat[2,], as.character) # rename columns
cov_stat <- cov_stat[-c(1, 2), ] %>%
  mutate(`5th` = p5, `95th` = p95) %>% # Get long format
  select(covariate, `5th`, `95th`, other) %>%
  gather(key = "condition", value = "cov_stat", `5th`, `95th`, other)
```

- Forest plots

```r
cov <- merge(cov_coef, cov_stat) # Merge cov stat and cov coefficients and uncertainty

cov_p <- cov %>%
  mutate_at(list(as.character), .vars =  vars(2, 4:6)) %>% # Format to numerics
  mutate_at(vars(4:6), list(as.numeric)) %>%
  mutate(
    rp5 = p5 * 100 - 100,
    rp95 = p95 * 100 - 100,
    rmean = mean * 100 - 100
  ) %>% # Convert to % and normalize to 0
  mutate(covariate = recode(
    covariate,
    AGE = "Age",
    SEX = "Sex",
    WT = "Weight",
    DIAG_2 = "COPD",
    DIAG_3 = "COPD and asthma"
  ),
    parameter = recode(
      parameter,
      BASE = "Base",
      NORMAL = "Normal",
      C50 = "C50",
      T50 = "T50"
    )) %>% # Rename
  mutate(condition = fct_relevel(condition, "other", "95th", "5th")) %>% # Reorder
  mutate(covariate = fct_relevel(covariate, "Age", "Weight", "COPD", "COPD and asthma", "Sex"))

cov_p <- cov_p %>% # Remove infinite values from the plot
  mutate(rmean = ifelse(rmean > 2e+20, NA, rmean),
       rp5 = ifelse(rp5 > 2e+20, NA, rp5),
       rp95 = ifelse(rp95 > 2e+20, NA, rp95))

p1 <-
  ggplot(data = cov_p[cov_p$parameter %in% c("Base"), ],
         aes(
           y = condition,
           x = rmean,
           xmin = rp5,
           xmax = rp95,
           label = cov_stat
         )) +
  geom_point() +
  facet_grid(covariate~., scale="free", labeller = label_wrap_gen(width=3))+
  geom_vline(xintercept=0, color="black", linetype="dashed", alpha=.5)+
  geom_vline(xintercept=-20, color="black", linetype="dotted", alpha=.5)+
  geom_vline(xintercept=+20, color="black", linetype="dotted", alpha=.5)+
  theme_bw()+
  geom_text(nudge_y = 0.4, size=2) +
  scale_x_continuous( name=NULL)+
  scale_y_discrete( name="Base")+
  geom_errorbarh(height=.1) + #adds the CIs
  force_panelsizes() +
```

```
  theme(text = element_text(size=8))

p2 <-
  ggplot(data = cov_p[cov_p$parameter %in% c("Normal"), ],
         aes(
           y = condition,
           x = rmean,
           xmin = rp5,
           xmax = rp95,
           label = cov_stat
         )) +
  geom_point() +
  facet_grid(covariate~., scale="free", labeller = label_wrap_gen(width=1.5))+
  geom_vline(xintercept=0, color="black", linetype="dashed", alpha=.5)+
  geom_vline(xintercept=-20, color="black", linetype="dotted", alpha=.5)+
  geom_vline(xintercept=+20, color="black", linetype="dotted", alpha=.5)+
  theme_bw()+
  geom_text(nudge_y = 0.4, size=2) +
  scale_x_continuous( name=NULL)+
  scale_y_discrete( name="Normal")+
  geom_errorbarh(height=.1)+ #adds the CIs
  force_panelsizes() +
  theme(text = element_text(size=8))

p3 <-
  ggplot(data = cov_p[cov_p$parameter %in% c("T50"), ],
         aes(
           y = condition,
           x = rmean,
           xmin = rp5,
           xmax = rp95,
           label = cov_stat
         )) +
  geom_point() +
  facet_grid(covariate~., scale="free", labeller = label_wrap_gen(width=3))+
  geom_vline(xintercept=0, color="black", linetype="dashed", alpha=.5)+
  geom_vline(xintercept=-20, color="black", linetype="dotted", alpha=.5)+
  geom_vline(xintercept=+20, color="black", linetype="dotted", alpha=.5)+
  theme_bw()+
  geom_text(nudge_y = 0.4, size=2) +
  scale_x_continuous( name=NULL)+
  scale_y_discrete( name="T50")+
  geom_errorbarh(height=.1) + #adds the CIs
  force_panelsizes() +
  theme(text = element_text(size=8))

p4 <-
  ggplot(data = cov_p[cov_p$parameter %in% c("C50"), ],
         aes(
           y = condition,
           x = rmean,
           xmin = rp5,
           xmax = rp95,
```

```
          label = cov_stat
        )) +
  geom_point() +
  facet_grid(covariate~., scale="free", labeller = label_wrap_gen(width=2))+
  geom_vline(xintercept=0, color="black", linetype="dashed", alpha=.5)+
  geom_vline(xintercept=-20, color="black", linetype="dotted", alpha=.5)+
  geom_vline(xintercept=+20, color="black", linetype="dotted", alpha=.5)+
  theme_bw()+
  force_panelsizes() +
  geom_text(nudge_y = 0.4, size=2) +
  scale_x_continuous( name=NULL)+
  scale_y_discrete( name="C50")+
  geom_errorbarh(height=.1) + #adds the CIs
  theme(text = element_text(size=8))

gridExtra::grid.arrange(p1,p2,p3,p4, nrow=2,
                        bottom = "Effect size in percent")
```

B.12.3 Unexplained variability

```
cov_u <- cov_unex %>%
  select(1:5) %>%
  mutate_at(list(as.character), .vars =  vars(3:5)) %>% # Convert to numerics
  mutate_at(vars(3:5), list(as.numeric)) %>%
  mutate(  # Rename
    covariate = recode(
      covariate,
      none = "None",
      all = "All",
      AGE = "Age",
      SEX = "Sex",
      WT = "Weight",
      DIAG_2 = "COPD",
      DIAG_3 = "COPD and asthma"
    ),
    parameter = recode(
      parameter,
      BASE = "Base",
      NORMAL = "Normal",
      C50 = "C50",
      T50 = "T50"
    )
  ) %>%
  mutate(covariate = fct_relevel(covariate, # Reorder
  "All", "Age", "Weight", "COPD", "COPD and asthma", "Sex", "None"))

ggplot(data = cov_u,
       aes(
         y = covariate,
         x = sd_observed,
         xmin = sd_5th,
         xmax = sd_95th
```

```
      )) +
  geom_point() +
  facet_wrap(parameter ~ ., scale = "free") +
  theme_bw() +
  scale_x_continuous(name = "Standard deviation of unexplained variability") +
  scale_y_discrete(name = "Covariate") +
  geom_errorbarh(height = .1) #adds the CIs
```

B.13 Forest plots

```
xpdb20 <- xpose::xpose_data(runno=20, dir="models/")
prm20f <- xpose::get_prm(xpdb20)

prm20f$se <- # Values from SIR
  c(8.75124, 29.85375, 5.166212, 2.273961, 0.07364673,
    0.7777551, 0.03512398, 0.04976547, 0.001446226,
    0.04206618, 0.009734037, 0.6732954, 0.268544, 0.003135913)

prmp <- prm20f %>%
  filter(type %in% c("the"), # Select only covariates effects
         !grepl("TV_", label) ) %>%
         select(label, value, se)

for (i in c(20,75)){ # Covariate effect at different ages
prmp <- prmp %>%
  slice(rep(3,1)) %>%
  mutate(label = paste0("Age ", i, " years"),
         value = value * (i - 40),
         se = se * (i - 40)) %>% # SE value from SIR
  add_row(., .before=1, prmp)
}

for (i in c(2,3)){ # Covariate effect for the DIAG categories
prmp <- prmp %>%
  slice(rep(1,1)) %>%
  mutate(label = ifelse(i == 1, "Asthma",
                        ifelse(i == 2, "COPD",
                               "COPD and Asthma")),
         value = value * (1 - i),
         se = se * (1 - i)) %>%
  add_row(., .before=1, prmp)
}

prmp %>% # Create the plot
  filter(!label %in% c("FYEARS", "DIAG")) %>%
  mutate( ci5  = 1 + (value - 1.96*se), # Get the 95 CI
          ci95 = 1 + (value + 1.96*se),
          value = 1 + value, # Covariate equation
          nb = ifelse(value < 1, paste0(round(value, 3), " [", # Text
                      round(ci5, 3), " - ",
                      round(ci95, 3), "]"),
```

```
                      paste0(round(value, 3), " [",
                      round(ci95, 3), " - ",
                      round(ci5, 3), "]"))) %>%
ggplot(., aes( y = label, x = value,
  xmin = ci5, xmax = ci95 )) +
  geom_point(aes(color = label), shape=3) +
  geom_errorbarh(aes(color = label), height=.1) +
  geom_vline(xintercept=1, color="black", linetype="dashed", alpha=.5)+
  geom_text(aes(label=nb, x = 1.5), hjust = 0) +
  scale_x_continuous( name="Covariate effect on Normal",
                      limits = c(0.35,2)) +
  scale_y_discrete( name=NULL) +
  theme_bw() +
  theme(legend.position = "none")
```

B.14 VPCs

B.14.1 Default VPCs

Default VPCs as in the left panel of the Figure S16

- PsN command:

```
vpc run11.mod -samples=500 -auto_bin=auto -rplots=1 -dir=run11_vpc_n500
```

- R code:

```
vpc <- function(n) { # Function to create non stratified VPC
  xpdb <- # Read model information from PsN model folder
    xpose::xpose_data(runno = n, dir = "models")

  p <- vpc::vpc(
    psn_folder = paste0("models/run", n, "_vpc_n500"),
    show = c(
      obs_dv = T,
      sim_median = T,
      pi = T
    ),
    pi = c(0.05, 0.95),
    ci = c(0.025, 0.975),
    xlab = "Time (h)",
    ylab = "PEFR (L/min)",
    title = NULL,
    vpc_theme = vpc::new_vpc_theme(
      update = list(
        obs_shape = ".",
        obs_median_color = "black",
        obs_median_linetype = "solid",
        obs_median_size = 0.5,
        obs_ci_color = "#3388cc",
        obs_ci_linetype = "solid",
```

```
        sim_pi_linetype = 'dashed',
        sim_pi_size = 0.5,
        sim_pi_color = "#3388cc",
        sim_median_fill = "black",
        sim_median_color = "black",
        sim_median_linetype = "dashed",
        sim_median_size = 0.5
      )
    )
  ) + theme_bw() +
    scale_x_continuous(limits = c(0, 450)) +
    scale_y_continuous(limits = c(0, 900))
  return(p)
}

vpc11 <- vpc(11) + # Create VPC for run n
  labs(subtitle = "Model without covariates (run11)")

vpc25 <- vpc(25) +
    labs(subtitle = "Covariate model from SCM (run25)", y = NULL)

grid.arrange(vpc11, vpc25, ncol = 2)
```

B.14.2 Stratified VPCs

VPCs stratified by two categories of a continuous covariate as in Figure S17.

- PsN command:

```
vpc run110.mod -samples=500 -auto_bin=auto -stratify_on=AGEC -rplots=1 -dir=run110_vpc_n500_stratAGEC
```

- Lines added to the NONMEM code to binarize the continuous covariate:

```
> AGEC=0
> IF(AGE.GT.34)AGEC=1
> $TABLE ID TIME SEX DIAG AGEC Y NOPRINT ONEHEADER FILE=catab110
```

- R code:

```
lab <- # Custom strip labels
  c("0" = "AGE =< 34 years", "1" = "AGE > 34 years")

vpc_strat <- function(n, cov) { # VPC function with stratification
  xpdb <- # Read model information from PsN model folder
    xpose::xpose_data(runno = n, dir = "models")

  p <- vpc::vpc(
    psn_folder = paste0("models/run", n, "_vpc_n500_strat", cov),
    show = c(
      obs_dv = T,
      sim_median = T,
      pi = T
```

```
    ),
    stratify = c(cov),
    facet = "rows",
    labeller = labeller(AGEC = lab),
    pi = c(0.05, 0.95),
    ci = c(0.025, 0.975),
    xlab = "Time (h)",
    ylab = "PEFR (L/min)",
    title = NULL,
    vpc_theme = vpc::new_vpc_theme(
      update = list(
        obs_shape = ".",
        obs_median_color = "black",
        obs_median_linetype = "solid",
        obs_median_size = 0.5,
        obs_ci_color = "#3388cc",
        obs_ci_linetype = "solid",
        sim_pi_linetype = 'dashed',
        sim_pi_size = 0.5,
        sim_pi_color = "#3388cc",
        sim_median_fill = "black",
        sim_median_color = "black",
        sim_median_linetype = "dashed",
        sim_median_size = 0.5
      )
    )
  ) + theme_bw() +
    scale_x_continuous(limits = c(0, 450)) +
    scale_y_continuous(limits = c(0, 950))
  return(p)
}

vpc110 <- vpc_strat(110, "AGEC") +
  labs(subtitle = "Model without covariates (run11)", x = NULL)

vpc250 <- vpc_strat(250, "AGEC") +
    labs(subtitle = "Covariate model from SCM (run25)")

grid.arrange(vpc110, vpc250, ncol = 1)
```

B.14.3 VPCs using a covariate as independent variable

VPCs using a covariate as independent variable (Figure S18)

- PsN command:

```
vpc run11.mod -samples=500 -auto_bin=auto -idv=AGE -rplots=1 -dir=run11_vpc_n500_idvAGE
```

- R code:

```
vpc <- function(n) { # Function to create non stratified VPC

    xpdb <- # Read model information from PsN model folder
    xpose::xpose_data(runno = n, dir = "models")

    p <- vpc::vpc(
    psn_folder = paste0("models/run", n, "_vpc_n500_idvAGE"),
    obs_cols = list(idv="AGE"),
    sim_cols = list(idv="AGE"),
    show = c(
      obs_dv = T,
      sim_median = T,
      pi = T
    ),
    pi = c(0.05, 0.95),
    ci = c(0.025, 0.975),
    xlab = "Age (year)",
    ylab = "PEFR (L/min)",
    title = NULL,
    vpc_theme = vpc::new_vpc_theme(
      update = list(
        obs_shape = ".",
        obs_median_color = "black",
        obs_median_linetype = "solid",
        obs_median_size = 0.5,
        obs_ci_color = "#3388cc",
        obs_ci_linetype = "solid",
        sim_pi_linetype = 'dashed',
        sim_pi_size = 0.5,
        sim_pi_color = "#3388cc",
        sim_median_fill = "black",
        sim_median_color = "black",
        sim_median_linetype = "dashed",
        sim_median_size = 0.5
      )
    )
  ) + theme_bw() +
      scale_y_continuous(limits = c(0, 850))
  return(p)
}

vpc11 <- vpc(11) +
  labs(subtitle = "Model without covariates (run11)")

vpc25 <- vpc(25) +
    labs(subtitle = "Covariate model from SCM (run25)", y = NULL)

grid.arrange(vpc11, vpc25, ncol = 2)
```

C Libraries and environment

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] xpose_0.4.13       forcats_0.5.1      stringr_1.4.0      readr_2.1.0
##  [5] tidyr_1.1.4        tibble_3.1.6       tidyverse_1.3.1    summarytools_1.0.0
##  [9] reshape2_1.4.4     purrr_0.3.4        PsNR_0.2.0         plyr_1.8.6
## [13] patchwork_1.1.1    knitr_1.36         kableExtra_1.3.4   huxtable_5.4.0
## [17] ggh4x_0.2.1        ggpubr_0.4.0       ggiraphExtra_0.3.0 ggiraph_0.7.10
## [21] ggforce_0.3.3      GGally_2.1.2       ggplot2_3.3.5      gtsummary_1.5.0
## [25] ftExtra_0.2.0      dplyr_1.0.7        data.table_1.14.2
##
## loaded via a namespace (and not attached):
##  [1] colorspace_2.0-2    ggsignif_0.6.3      pryr_0.1.5
##  [4] ellipsis_0.3.2      sjlabelled_1.1.8    flextable_0.6.10
##  [7] fs_1.5.0            base64enc_0.1-3     rstudioapi_0.13
## [10] farver_2.1.0        bit64_4.0.5         fansi_0.5.0
## [13] lubridate_1.8.0     xml2_1.3.2          codetools_0.2-18
## [16] splines_4.1.2       sjmisc_2.8.7        polyclip_1.10-0
## [19] jsonlite_1.7.2      gt_0.3.1            broom_0.7.10
## [22] dbplyr_2.1.1        compiler_4.1.2      httr_1.4.2
## [25] backports_1.3.0     assertthat_0.2.1    Matrix_1.3-4
## [28] fastmap_1.1.0       cli_3.1.0           tweenr_1.0.2
## [31] htmltools_0.5.2     tools_4.1.2         gtable_0.3.0
## [34] glue_1.5.0          Rcpp_1.0.7          carData_3.0-4
## [37] cellranger_1.1.0    vctrs_0.3.8         svglite_2.0.0
## [40] nlme_3.1-153        broom.helpers_1.4.0 insight_0.14.5
## [43] xfun_0.28           rvest_1.0.2         lifecycle_1.0.1
## [46] rstatix_0.7.0       MASS_7.3-54         scales_1.1.1
## [49] vroom_1.5.6         hms_1.1.1           parallel_4.1.2
## [52] mycor_0.1.1         RColorBrewer_1.1-2  yaml_2.2.1
## [55] pander_0.6.4        gdtools_0.2.3       reshape_0.8.8
## [58] stringi_1.7.5       checkmate_2.0.0     ppcor_1.1
## [61] zip_2.2.0           commonmark_1.7      rlang_0.4.12
## [64] pkgconfig_2.0.3     systemfonts_1.0.2   matrixStats_0.61.0
## [67] evaluate_0.14       lattice_0.20-45     rapportools_1.0
## [70] htmlwidgets_1.5.4   bit_4.0.4           tidyselect_1.1.1
## [73] magrittr_2.0.1      bookdown_0.24       R6_2.5.1
## [76] magick_2.7.3        generics_0.1.1      DBI_1.1.1
## [79] pillar_1.6.4        haven_2.4.3         withr_2.4.2
## [82] mgcv_1.8-38         abind_1.4-5         modelr_0.1.8
## [85] crayon_1.4.2        car_3.0-12          uuid_1.0-3
## [88] utf8_1.2.2          tzdb_0.2.0          rmarkdown_2.11
## [91] officer_0.4.1       grid_4.1.2          readxl_1.3.1
## [94] reprex_2.0.1        digest_0.6.28       webshot_0.5.2
## [97] munsell_0.5.0       viridisLite_0.4.0   tcltk_4.1.2
```