PopED Manual

Release version 2.10

Introduction

PopED (Population Experimental Design) is a software tool for computing optimal experimental designs. The software has been developed with an emphasis on drug trials based on population models (non-linear mixed effects models).

The purpose of this manual is **NOT** to teach Optimal Design but rather to give an introduction and help to the PopED Graphical User Interface GUI and PopED script version. PopED GUI is a Windows based program written in the language C# .NET 2.0 that will wrap around the script version of PopED (written in Matlab) that performs the calculations needed to get an optimal design. The purpose of the GUI is to get an easy way to build up an experiment and to optimize the design variables in that experiment. There are also tools available for interpretation of the outcome of the optimal design and ways to validate models and simulate models prior to the optimal design. All these tools are accessible via the script version of PopED but then the user needs knowledge about the Matlab programming language and how to set up an experiment in the Matlab environment, therefore the GUI was developed to minimize the skills needed by the user in terms of programming.

The PopED GUI provides model templates and examples that will help the user to set up their own experiments. Some knowledge in the Matlab language might be useful but the model templates should give a good introduction to defining PopED models in Matlab.

In Addition to this manual, the Matlab manual and literature in the fields of Population pharmacokinetic and pharmacodynamic (PK-PD) modeling and Optimal Design is suggested.

PopED is a open source program developed by the pharmacometrics research group, department of Pharmaceutical Bioscience, Faculty of Pharmacy at Uppsala University.

The PopED Development Team

Installation

PopED is developed for a Microsoft Windows environment. PopED is mainly tested on a Windows XP Professional (Service Pack 2) platform.

Requirements

PopED needs the following packages/programs to work perfectly:

1) Microsoft .NET 2.0, use Windows Update or download and install the package from www.microsoft.com. Don't use the SDK for .NET 2.0, (the size is unnecessarily big) instead use the redistributable package.

   or

1) Mono 2.4 (or newer), enables PopED GUI on UNIX-systems / Mac-OS. Download and install it from www.mono-project.com.

2) To run optimization within the Graphical User Interface (GUI), Matlab needs to be installed. The Statistics toolbox and The Symbolic Toolbox is used in some functions. If Matlab is not installed the PopED GUI could be used as a help to build up PopED runs. PopED is developed with Matlab version 2007a 7.4 with a registered COM Server from the same version. However, new features in 2.09 should make it possible to run without the Matlab COM Server.

Installation step by step

1) Fulfill the requirements above.

2) Run the setup file (setup.exe) and then walk through the setup wizard.

3) Open PopED GUI from the Start Menu and Open File/Settings. Set the paths to the different directories.

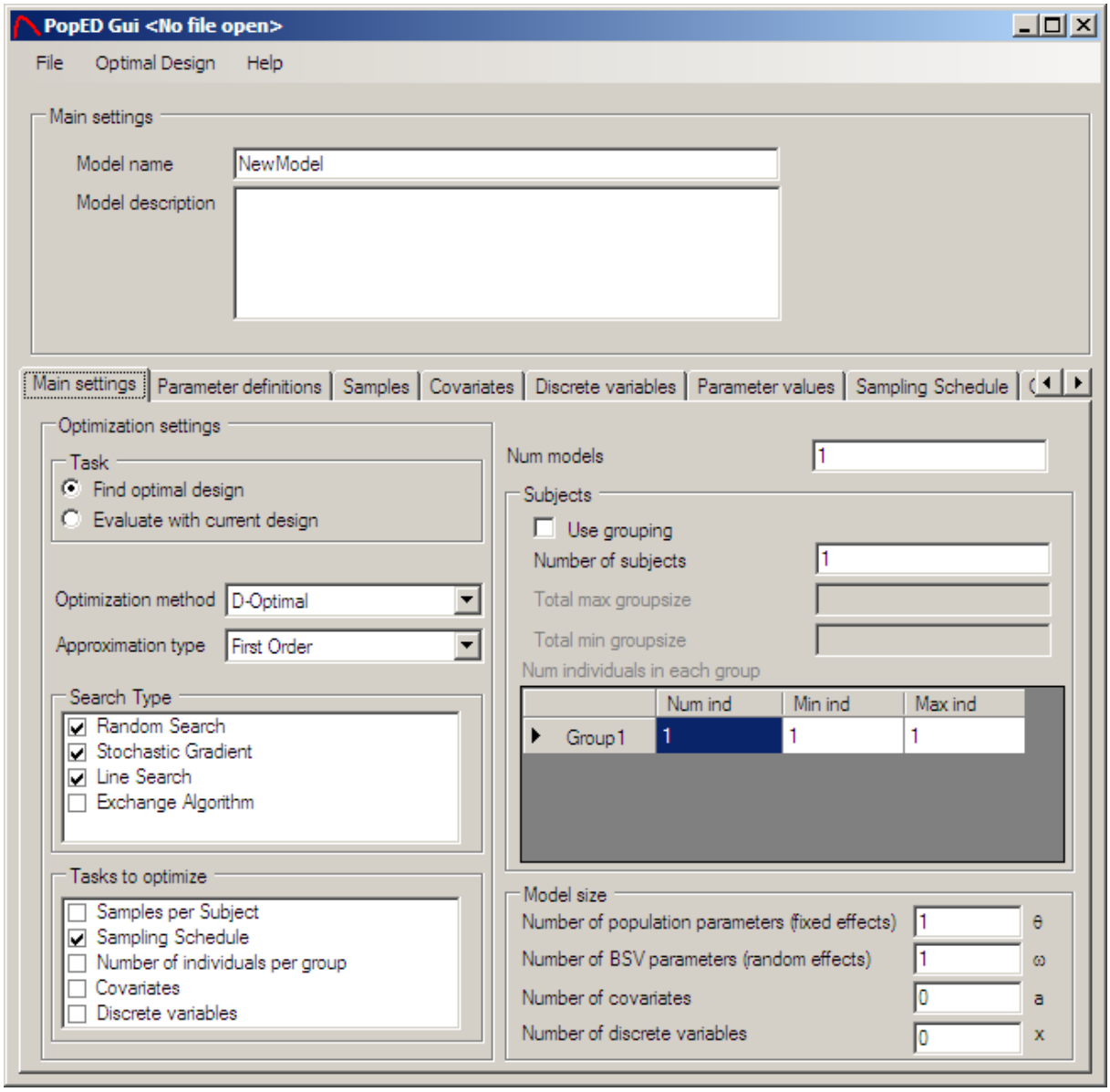
4) If the PopED program is **NOT** installed in the directory *C:\Program Files\PopED*, change the path in the *poped2_0.m* file and the *convert_xml.m* file in the *installation dir\program\xml* directory to your local path.

5) Setup is now completed.

Running PopED

Start PopED GUI from the start menu or run the script version (without graphics) of PopED by starting Matlab. In the script version; add the path to the PopED script version (e.g. *C:\Program Files\PopED\program*) to the working directory of Matlab. This can be done by using the *addpath* command or by changing the directory to PopED and then type *poped2_0('')*. A third way is to use the Matlab *Set Path* command under the File menu in the Matlab GUI.

PopED GUI Main Window



If the PopED GUI version is started the main window will appear. From this window the user can define new PopED GUI settings files, run optimization (if a Matlab COM server is available) and use the diagnostic tools that are available within the GUI.

Creating a new Model with the GUI

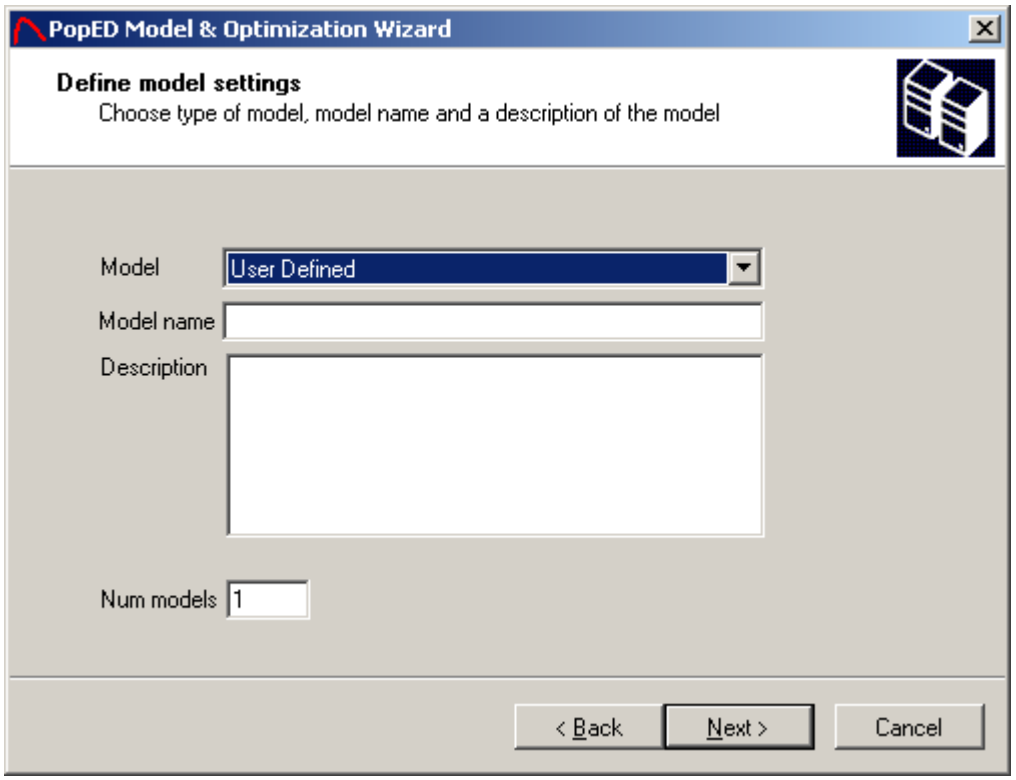
There is several ways to create a new model in the PopED GUI.

1) Under the File Menu choose Wizard and New.
   - Define a user specified model.
   - Use the predefined model templates that are available.

2) Under the File Menu choose New (Ctrl+N).

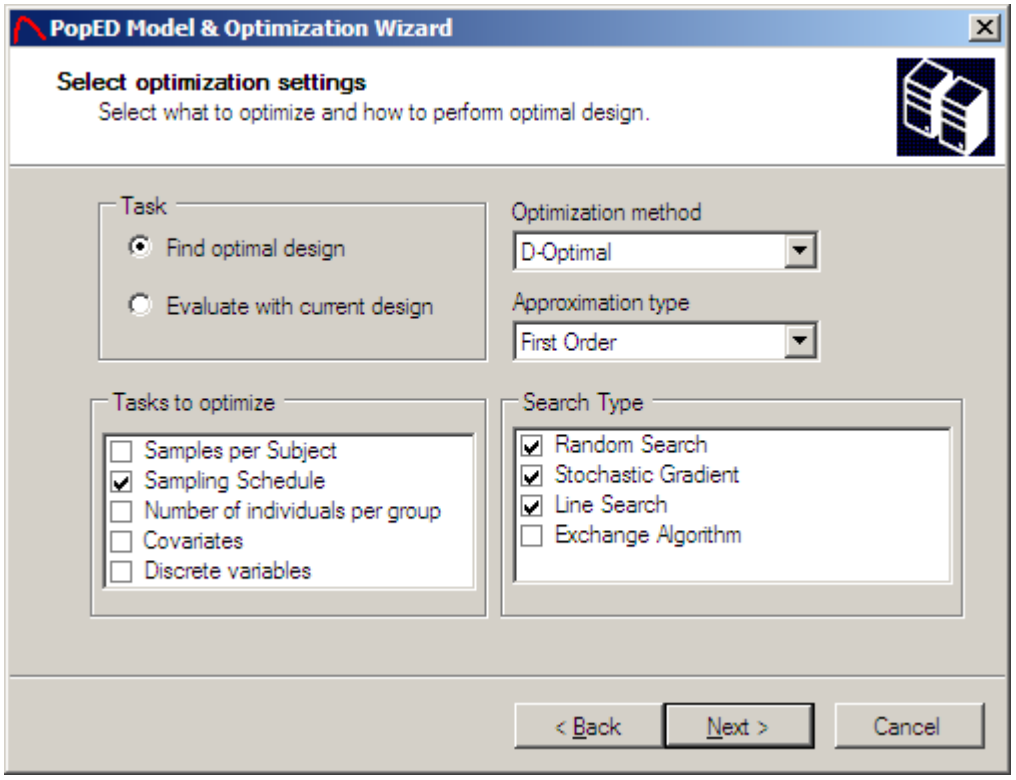
This manual will walk through the creation of a model in the same way as the Wizard does it in PopED.

Define the model settings



Choose a predefined model or choose to create a User Defined model as above. The Model name can be any string (containing spaces). The description can be any string containing white spaces. To define a new line in the description use the Soft Return, e.g. Ctrl+Return. Num models are the number of sub-models (responses) that are going to be used in the optimal design. E.g. Optimizing on the parameters for both a PK and a PD model will be 2 sub-models, concurrent optimizing on n different drug models will be n sub-models etc.

Optimization Settings



Choose the optimization method:

- D-Optimal Design – Optimization by maximizing the determinant of the Fisher Information Matrix (FIM). A prior model is assumed to be known (the population mean, the Inter Individual Variability and the Residual Variability). Fast but sensitive to model misspecification.

- ln D-Optimal Design – Basically the same at maximizing the determinant of the Fisher Information Matrix but the natural logarithm of the determinant of FIM could be a convex function while the surface of the determinant of FIM is non-convex. In these cases the ln D should be used.

- ED-Optimal Design – Optimization by maximizing the expectation value of the determinant of the Fisher Information Matrix. A prior model is assumed to be known but an uncertainty of the model parameters could be taken into account. E.g. assuming a normal distribution around all population mean parameters. Slow but can deal with model misspecification.

- ln ED-Optimal (API) Design - This criterion is not exactly the same as ED optimal (and are likely to give another optimal design) because it optimizes over the expectation value of the natural logarithm of the determinant of FIM. This criterion could e.g. be useful when convexity is an issue and/or when the value of the determinant is high.

Choose the approximation type

- First Order – The model is linearized around the first derivative evaluated at the typical values (b=0). The residual error model is also linearized around the typical values.

- First Order Conditional – The model is linearized around the first derivative evaluated at the individual mixed effects equal to an individual sample. The number of samples could be specified in the search settings. The default is to use 1000 samples per Fisher Information Matrix calculation. The residual variability is linearized around the typical values.

- First Order Conditional Interaction - The model is linearized around the first derivative evaluated at the individual mixed effects equal to an individual sample. The number of samples could be specified in the search settings. The default is to use 1000 samples per Fisher Information Matrix calculation. The residual variability is linearized around the individual samples.

- First Order Interaction - The model is linearized around the first derivative evaluated at the typical values (b=0). The residual error model is also linearized around the typical values and an extra interaction terms is used in the linearization. Could be used when e.g. a proportional error structure is used.

Note that if a conditional approximation method is used the calculation of the Fisher Information Matrix can give different results if the individual samples is not the same. It is also notable that the individual samples during an optimization (several calculation of the FIM) will be the same. I.e. the individual samples are **not** re-sampled between each calculation of the FIM.

In most cases the first order approximation linearized around the typical values (the default option) will yield a very similar design compared to the other approximation methods and it is much faster to calculate.

Choose the different design parameters to optimize over:

- Samples per Subject – Number of samples/measurements per design group or individual.

- Sampling Schedule – Optimize when to take the samples.

- Number of individuals per group – Optimizing over the group size.

- Covariates – Optimize for the optimal covariate value in the design. Could be any design dependent variable, e.g. Dose, Infusion length, Weight, Age, Study length etc.

- Discrete variables – Optimize over discrete variables (list of discrete values). This could be e.g. list of possible sample times, list of possible doses etc.

The sample times and the covariates are assumed to be continuous, while the Samples per Subject and Number of Individuals per group are discrete variables. A group in Optimal Design is a number of individuals that will get the same design, e.g. the same Sampling Schedule.

Search Types available are:

- Random Search (RS) – Does a random search over the whole search space initially but will collapse to an Adaptive Narrowed Random Search as soon as the Random Search finds a good Objective Function Value (OFV). The Adaptive Random Search will search randomly close to the previous best OFV found by the Random Search. The Random Search is global but it will be local with the Adaptive Random Search. Random Search converges towards an optimal OFV when the number of iterations goes towards infinity. The default number of iterations in the Random Search is 300.

- Stochastic Gradient (SG) – Does a stochastic walk in the direction of the best OFV. In D-Optimal design and ED-optimal design with a discrete user defined distribution there is no stochastic value for the derivative of the OFV and therefore the SG will collapse to a steepest descent algorithm. The SG is a local search method. The SG will converge to a local optimum when the number of iterations reaches infinity. The default number of iterations in SG is 150. When using discrete variables the SG only optimizes with the current best discrete value and therefore calculating the gradient of only the continuous variables.

- Line Search (LS) – Line search does a grid search in one dimension, i.e. over one parameter at a time. Line Search will only work when optimizing over variables like sample times, covariates and/or discrete variables. Line Search is a pure global search method but it will not alone converge toward a global optimum even though the number of grid points reaches infinity. This is because the LS could restrict the search space by the order of the design parameter that is used in the LS. The default number of grid points is 50. If optimization is done over a discrete variable the number of grid points will be disregarded for the discrete variables and all possible values will be tested.

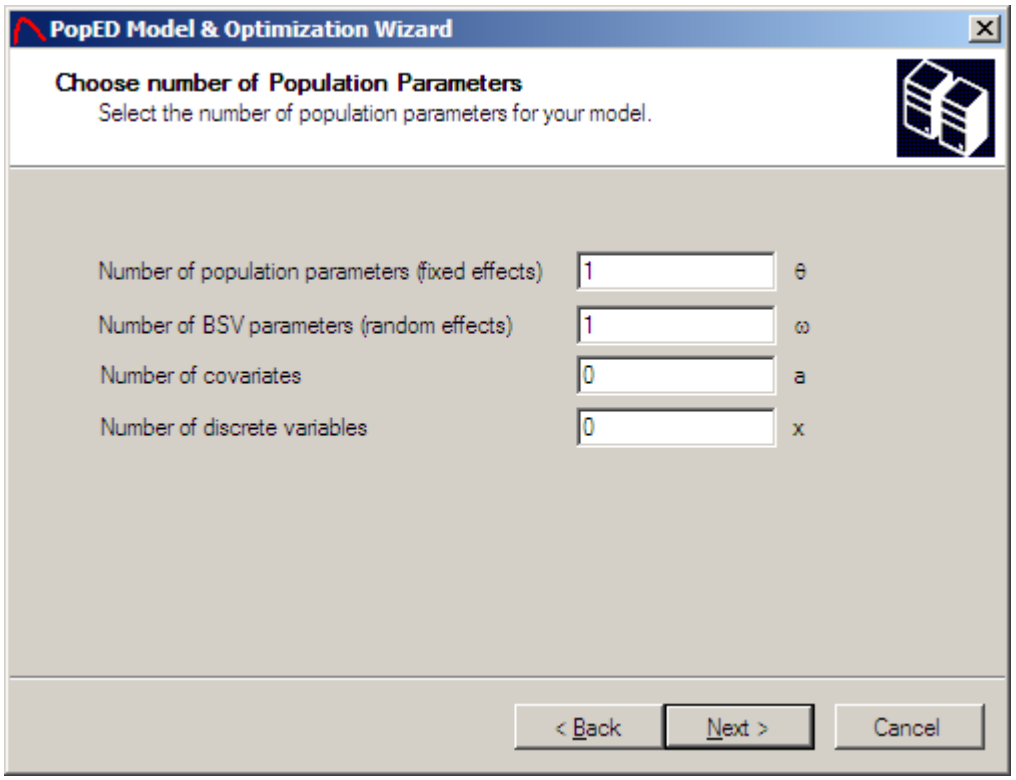
- Modified Fedorov Exchange Algorithm (MFEA) – This algorithm searches, in each iteration, the sample/covariate that changes the OFV the most. That sample is exchanged into the current “optimal” design and a new iteration is performed with the current “optimal” design. The possible sample times/covariates to be exchanged are defined by the max/min sample time/covariate value split up with a step length. When the exchange of a sample time/covariate doesn’t exceed a threshold value <<1. The threshold value represents the number of percent improvement in the current iteration. If optimization is done over a discrete variable the number of grid points will be disregarded for the discrete variables and all possible values will be tested. See *More search settings* for more information.

All the search types (RS, SG and LS) should be used to obtain the best result. The optimization is assumed to have converged if the LS don’t change the optimization

results of the previous SG (or RS if no SG is selected). If the Line Search is not selected the method will search with a predefined number of search iterations, i.e. one search iteration run the RS, SG or both. See Calculation Settings for more information. The default number of Search Iterations when LS is not selected is 10.

The MFEA algorithm couldn't be used with any of the other algorithms. But it can very well be used for each of the optimization tasks above. It is particularly accurate and fast when optimizing over small lists of possible discrete values (using discrete variables).

Population Parameters



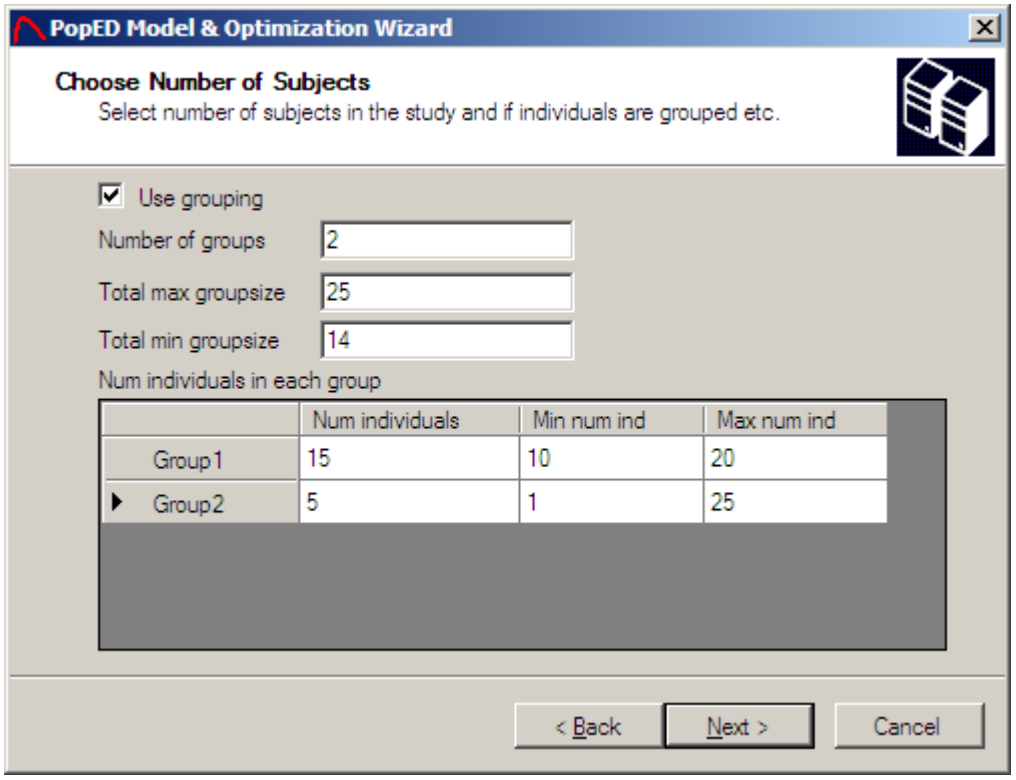
Choose the number of population mean parameters, also called typical values or fixed effects. The number of population mean parameters (bpop) should be the number of typical value parameters for all sub-models. In NONMEM these population parameters are called THETA.

Random effects are the parameters that define the variability between individuals (IIV, BSV). In NONMEM these parameters are called OMEGA. The number of random effects should be the number of random effects for all sub-models.

The numbers of covariates are the number of covariate parameters to have in the model, e.g. Dose, Age, Weight etc. In this version the covariates couldn't have a distribution so they are only continuous variables. The number of covariates should be the number of covariates for all sub-models.

The numbers of discrete variables is the number of discrete variables for all sub-models. This variable could also be used to optimize lists of sample times, in this case, set the number of discrete variables to the number of samples and use *x* in the model instead of *xt* as the time points. See *Define model* and the examples for more information.

Number of Subjects

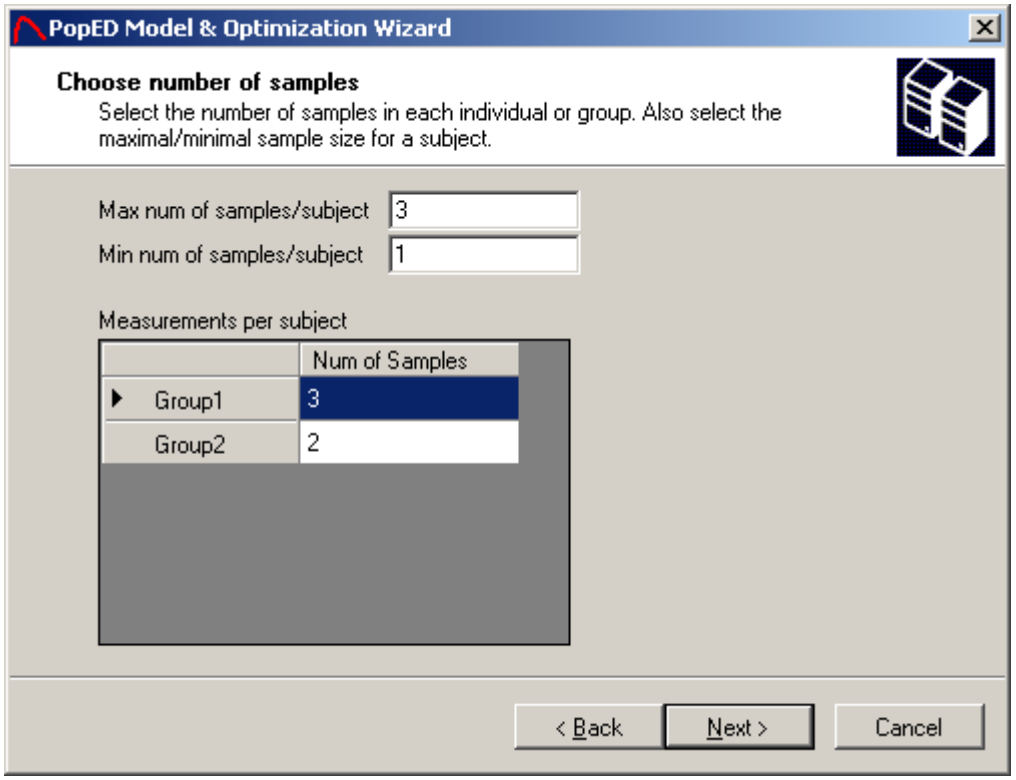


If use grouping are selected there is a possibility to optimize over the number of individuals in each group. In this case it's important to choose the minimal number of individuals in the group and also the maximal number of individuals in the group. If *Use grouping* is not checked, the numbers of individuals in each group are fixed to 1. If *Use grouping* is checked the number of individuals in the group will affect the optimal design even though the optimization procedure only optimize over e.g. the sample times. A group in optimal design is a number of individuals that will have the same design. The *total max groupsize*, i.e. the maximal sum of individuals in all groups and the minimal sum of individuals in all groups must be entered. The sum of initial number of individuals in all groups must be at least equal to the *total min groupsize*. The sum initial number of individuals must also be less (or equal) to the *total max groupsize*.

Example: Optimization of the sampling schedule within a study with at the most, 25 individuals split into 2 groups with initially 15 individuals in group 1 and 5 individuals in group 2. These optimized sampling times will yield one schedule for the 15 individuals and a 2nd sampling schedule for the 5 individuals.

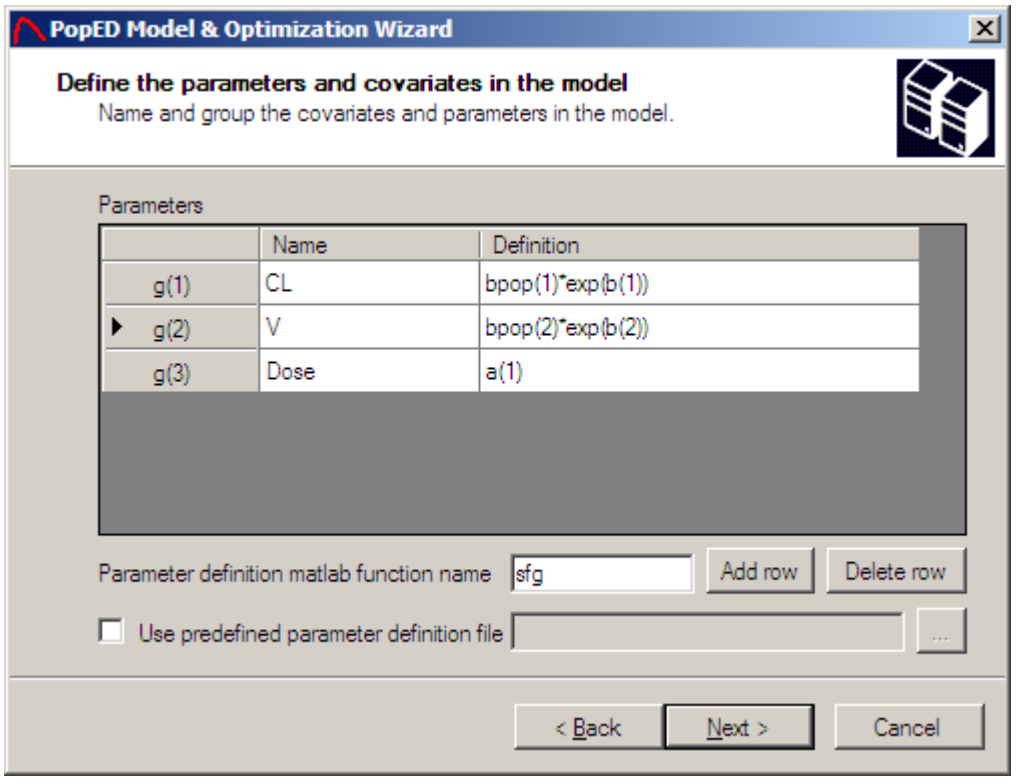
If optimizing on *number of individuals per group*; the optimization procedure will check all possible combinations of individuals divided into the two groups with the restrictions above.

Number of Samples



The number of sample/measurements per subject (group or one individual) could be different for each subject. The minimum number of samples per subject should at least be one. The minimum and maximum numbers of samples per subject are only important when optimizing over the number of samples per subject, except that the maximum number of samples per subject should always be larger or equal to the number of samples in each group.

Parameter Definitions



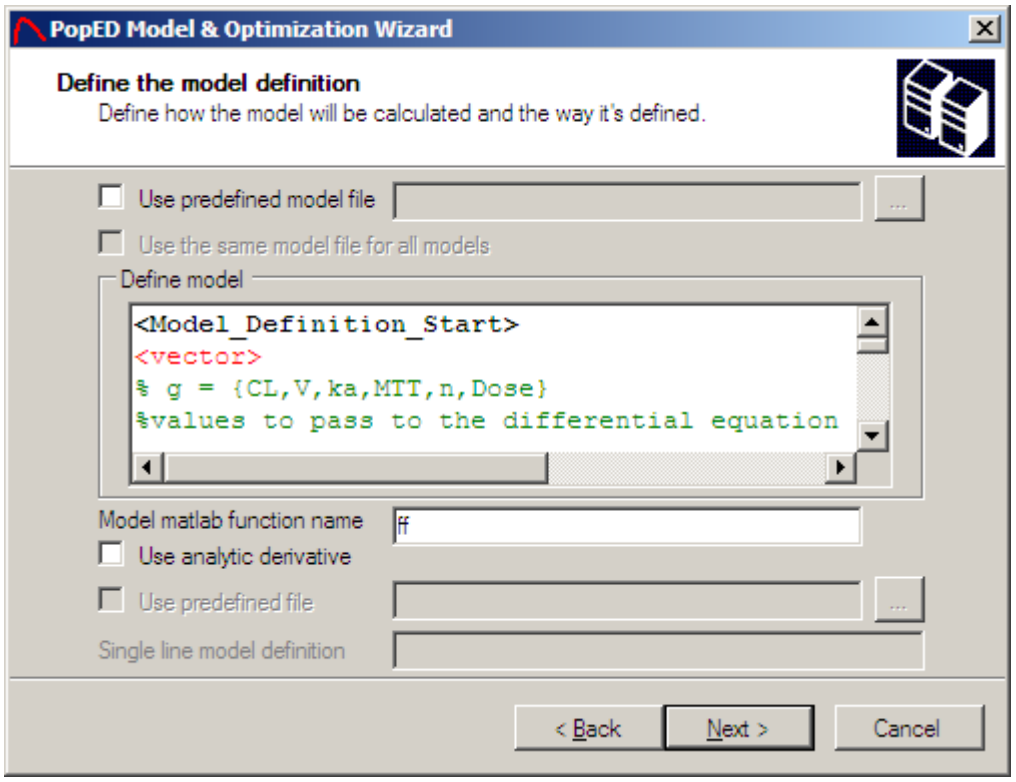
The parameters are defined and stored in a vector called g. The definition of a parameter could be of any type, e.g. normal, exponential etc. The typical values (fixed effects) are a vector called **bpop** and the inter-individual variability is a vector called **b**. In NONMEM the b's are called ETA. The different covariates and discrete variables that should be used in the sub-models should also be defined here. The covariates is a vector called **a** and the discrete variables is a vector called **x**. It's perfectly fine to make transformations to cover other IIV distributions than the normal distribution, e.g. the Box-Cox transformation. The Name column (case sensitive) could be used to define the model in the *model definition*. The number of parameters are unlimited but the number of bpop, b, a and x is limited to the numbers specified in population parameters window. The definition is written in the Matlab language and could therefore use any Matlab function. *The parameter definition Matlab function file name* could be changed to prevent that the GUI overwrites a previous created Matlab parameter definition function. The name should be a valid Matlab function/file name.

It is possible to define a predefined parameter definition file instead of defining the g vector in the GUI. This is done by producing a Matlab function file (*.m) with the following syntax:

```
function pop_params = fg(x,a,bpop,b,bocc)

% -- Parameter definition file --
% -- { Clearance CL, Volume V, Dose (mg) }

pop_params=[ bpop(1)*exp(b(1)); bpop(2)*exp(b(2)); a(1)];
```

```
end
```

Define model



Every sub-model (response) will have a window similar to the one above except that the checkbox *Use the same model file for all models* only will be available at the first sub-model. If this option is checked no more sub-model definitions are needed even though there is more than one sub-model. In the PopED main window only one tab page for all models will be available. Right click in the model definition will popup a menu that enables to switch between g() and the corresponding names defined in the *parameter definitions*.

The sub-models could be defined with a predefined model file or directly in the window by adding Matlab code and tags between the model definition start tag and the model definition end tag. If a differential equation is needed for the current sub-model the code could be added within the differential equation start tag and the differential equation end tag.

The *Model Matlab function name* could be changed to prevent that the previous model file is overwritten. The name should be a valid Matlab function name.

Analytic derivative could be calculated by checking the *Use analytic derivative* option. To calculate the analytic derivative a single line model definition is needed or a predefined analytic derivative file could be specified. Analytic derivative therefore only works for “simple” models and is very slow to use compared to the numerical

derivatives that are used by default. The single line model definition should be written in Matlab code and use the variables mentioned below. Analytic derivatives only work for continuous variable optimization in this version. The predefined analytic derivative file is not fully functional in this version of PopED. Furthermore, there is a restriction that the additive and proportional residual variability must be fixed. The user defined error model could not be used. The Matlab Symbolic Toolbox must be installed to calculate the analytic derivatives.

Predefined model files

The model file could be specified for one sub-model or for all sub-models at the same time (by checking the *Use the same model file for all models* checkbox). If a file is specified only for this sub-model the Matlab function file (*.m) should have this syntax:

```
function [y,globalStructure]=model1(t,g,globalStructure)

% g = {CL, V, E0, Emax, EC50, Dose}
%PK-model
y=g(6)/g(2)*exp(-g(1)/g(2)*t);

end
```

Input should be a scalar value time t and the parameter definition vector g. As output y is set to a scalar value. All sub-models could be defined either by a model file like above or within the GUI independent of the other sub-models.

If a file is specified for all sub-models (even if it's only one sub-model) the file is again a Matlab function file (*.m) with this syntax:

```
function
[y,globalStructure]=model_example(model_switch,xt,g,globalStructure)

%-- Model: One comp IV bolus with direct Emax effect
%-- {CL,V,E0,EMax,EC50,Dose}

y=xt;
for i=1:length(model_switch)
    t=xt(i);
    if (model_switch(i)==1)
        %g = {CL,V,E0,Emax,EC50,Dose}
        %PK-model
        y(i)=((g(6)/g(2))*(exp(-g(1)/g(2)*t)));
    end
    if (model_switch(i)==2)
        %g = {CL,V,E0,Emax,EC50,Dose}
        %PD-model
        pkmodel=(g(6)/g(2))*(exp(-g(1)/g(2)*t));
        y(i) = g(3)+pkmodel*g(4)/(g(5)+pkmodel);
    end
end

end
```

The input to this function should be a vector model_switch that define which sub-model a specific time point i belongs to. The length of the model_switch vector is the same as the length of the xt vector containing all the time points for a design subject. The input g is the parameter definition vector g for the current subject. The last input is the *globalStructure* structure that contains information about all design parameters, all the search settings and information about the differential equation solver and differential equation solver options. See advanced settings.

Models defined in PopED GUI

A model could be defined for every sub-model in the PopED GUI. The models are written in an extended version of Matlab code. The models should be defined within the <Model Definition Start> tag and the <Model Definition End> tag. Differential equations for a sub-model can be specified within the <Differential Definition Start> tag and the <Differential Definition End> tag. Three different tags/reserved word are available for the model specification except the regular Matlab code:

- **<vector>** - Specifies that the definition of this sub-model will be vector wise, i.e. the time vector is called xt and the output vector for all the time points xt is called y. If the vector tag is not present the time will be a scalar value called t and the output will be called y(i). The vector tag can only be used in one sub-model; with several sub-models, use t and y(i) as the time and the output.

- **<diff>** - Specifies a call to the differential equation solver, typically used like this: *amount = <diff>;* The amount will be a vector of the amount for every time point xt if the vector tag is used. Otherwise amount will be the amount at time t. If <diff> is specified; a differential equation must be written in the differential definition part of the sub-model. User defined input parameters to the differential equation are stored in the vector *params* that could be of any length. These values are then accessible in the differential equation. If the differential equation contains several compartments, amount will be a matrix with the amount for all compartments and time points. If e.g. compartment two contains the dependent variable/the output it's accessible by typing *amount(:,2)*. This will be a vector if the vector tag is used, otherwise a scalar. For more information about the differential equation solver, see Advanced Settings.

- **<diff_lin>, <diff_lin_inhomc>** - Specify a call to the differential equation solver (see <diff>) but handles only linear homogeneous and linear time-independent constant inhomogeneous odes. Compartment models with simple rate constants are example of linear homogeneous odes. As long as the rate constants are not non-linear (e.g. Michaelis-Menten elimination, feedback models etc.) this option could be used to speed up the differential equation solver. Look at the typical value plot with <diff> and with <diff_lin>, if they are the same, the <diff_lin> could probably be used. Note that the inhomogeneous solver <diff_lin_inhomc> can only be used with central difference, i.e. not complex differentiation. For an example how to use this, see the distributed example with optimization on 3 drugs.

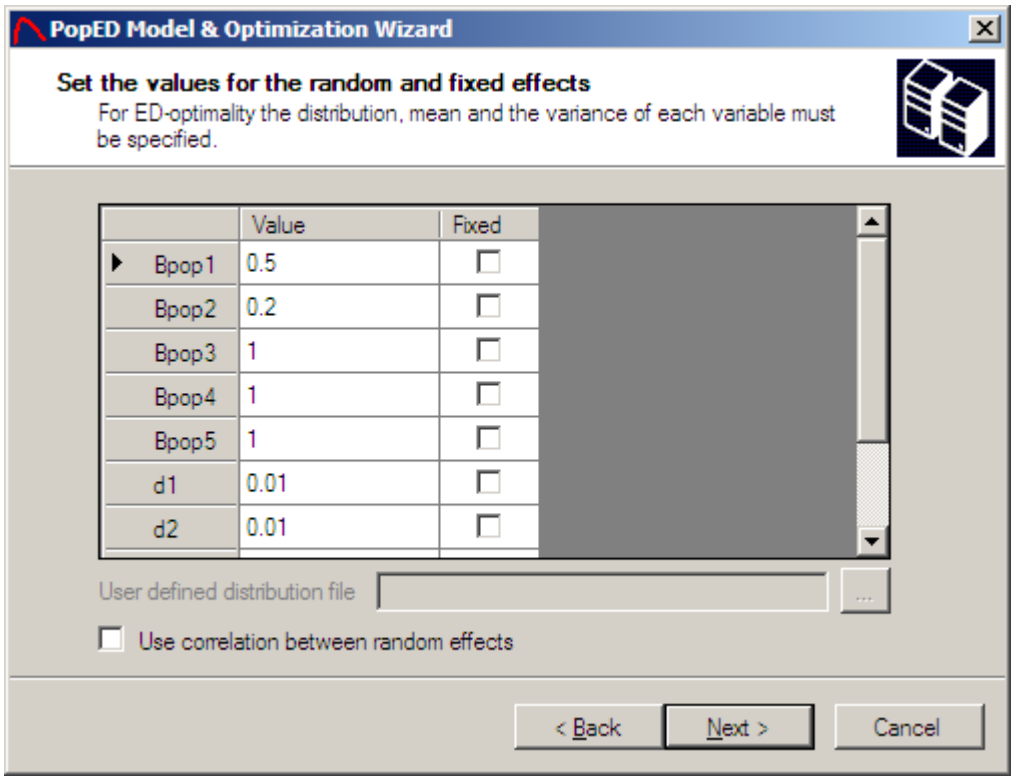
- **<init>** - Specifies a vector of the initial values for the compartments in the differential equation, typically used like this: *<init> = [init1 init2];* where init1 and init2 are the values at compartment one and two at time zero. If the init tag is specified a differential equation definition should be specified.

Example of a GUI defined model:

```
<Model Definition Start>
%Define that operations should be vector wise
<vector>

%g = {ka,ke,V,Dose}
%Store the values needed in the differential equation
params(1) = g(1); %ka
params(2) = g(2); %ke

%The initialization vector for the two compartments (absorption and
central)
<init> = [g(4) 0];

%Call the differential equation
amount = <diff>; %Here the <diff_lin> could be used instead (linear model)
y = amount(:,2)./g(3); %Make the DV a concentration

<Model Definition End>

<Differential Definition Start>

%params(1) = ka
%params(2) = ke
dA(1,:) = -params(1)*A(1);
dA(2,:) = params(1)*A(1) - params(2)*A(2);

<Differential Definition End>
```

See the model templates and the distributed examples for more information about how to code the model files.

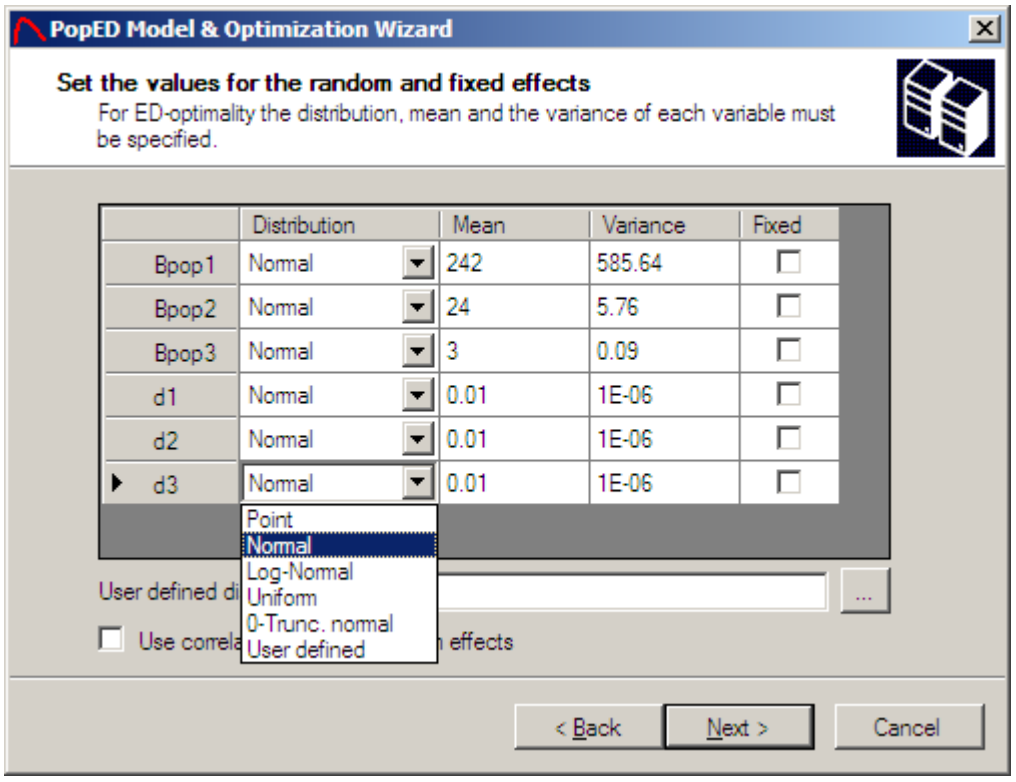
Population parameter values



Enter the model population mean values (bpop, THETA in NONMEM) and the variance of the IIV normal distribution (d). The d variables are stored as variances, hence a variance 0.01 correspond to 10% inter individual variability for an exponential b (ETA in NONMEM), i.e. exp(b). The d is interpreted in the same way as the OMEGA in NONMEM.

If there is correlation between the IIV parameters, this could be addressed by clicking the *Use correlation between random effects* checkbox. This will enable a correlation matrix for the random effects. The matrix will be enabled after clicking on the next button.

If ED-optimal design is going to be used the typical values (bpop) and the IIV variance parameters could have a distribution.

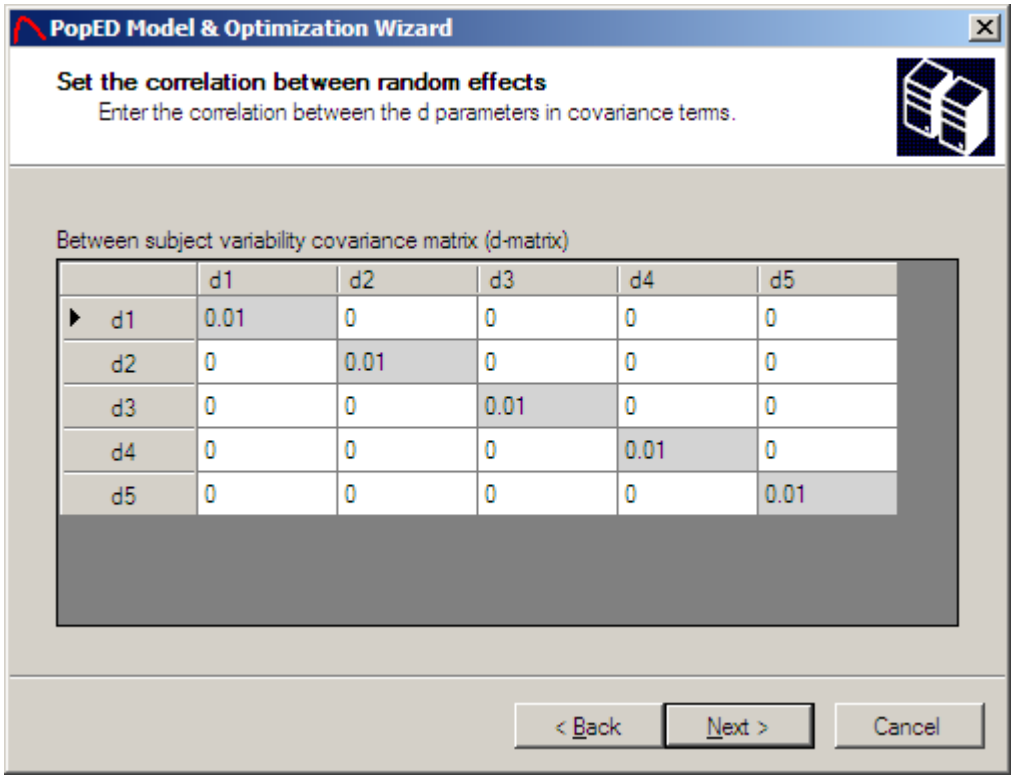


All parameter distributions selected as Point distributions will yield the same result as the D-optimal design. The other available distributions are normal, log-normal, zero-truncated normal and uniformed. There is also possible to define a user specified distribution (stochastic and non-stochastic). To use a user defined distribution; enter the filename pointing to the distribution function. See the examples for more details. In the picture above the variances of the distributions are entered as 10 % but in variance terms. In the uniform distribution the variance is the length of the uniform distribution, **not** the usual definition ($\frac{(length)^2}{12}$).

The zero-truncated normal distribution is left- or right truncated at zero depending on the sign of the mean. I.e. if a negative mean is used the distribution is right truncated at zero and if a positive mean is used the distribution is left truncated at zero.

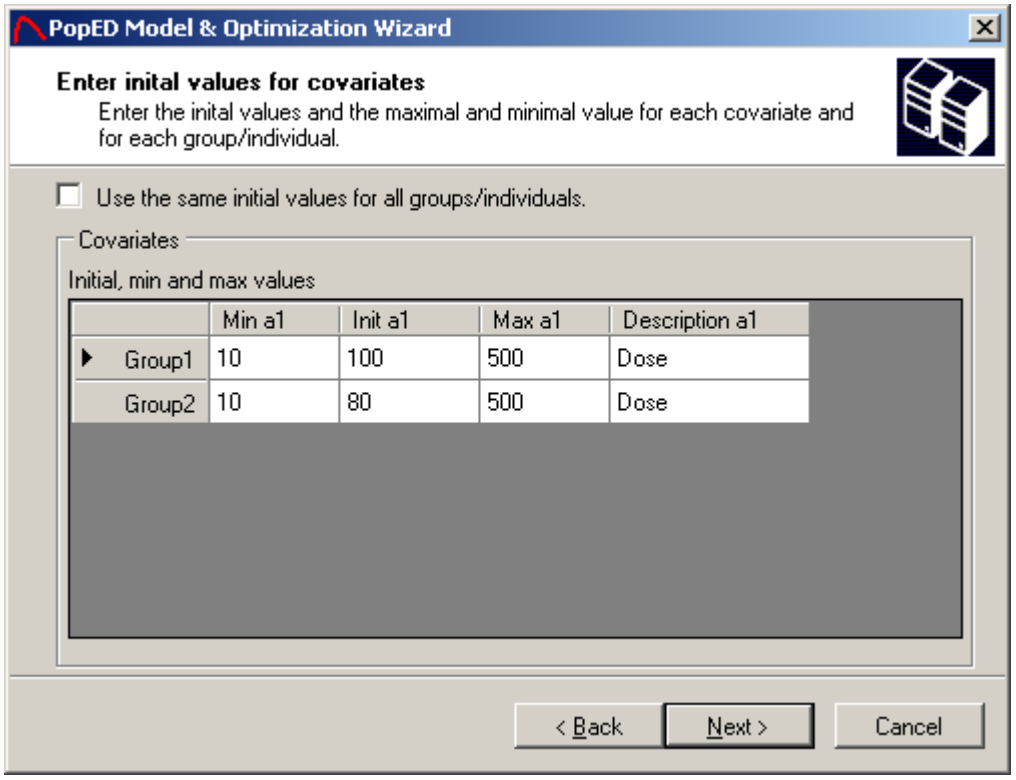
If a parameter is in the model, but is of no interest for estimation, the parameter could be fixed by checking the checkbox corresponding to the variable that is going to be fixed. This means that the variable would not be represented in the Fisher Information Matrix (FIM) and therefore the FIM wouldn't give any expected parameter uncertainty (SE (%)) of that parameter.

Correlation between random effects



Enter the correlation between all random effects expected the diagonal elements, i.e. the actual random effect variance defined in the previous wizard window. The correlation should be entered in covariance terms. A zero means that the random effects are uncorrelated.

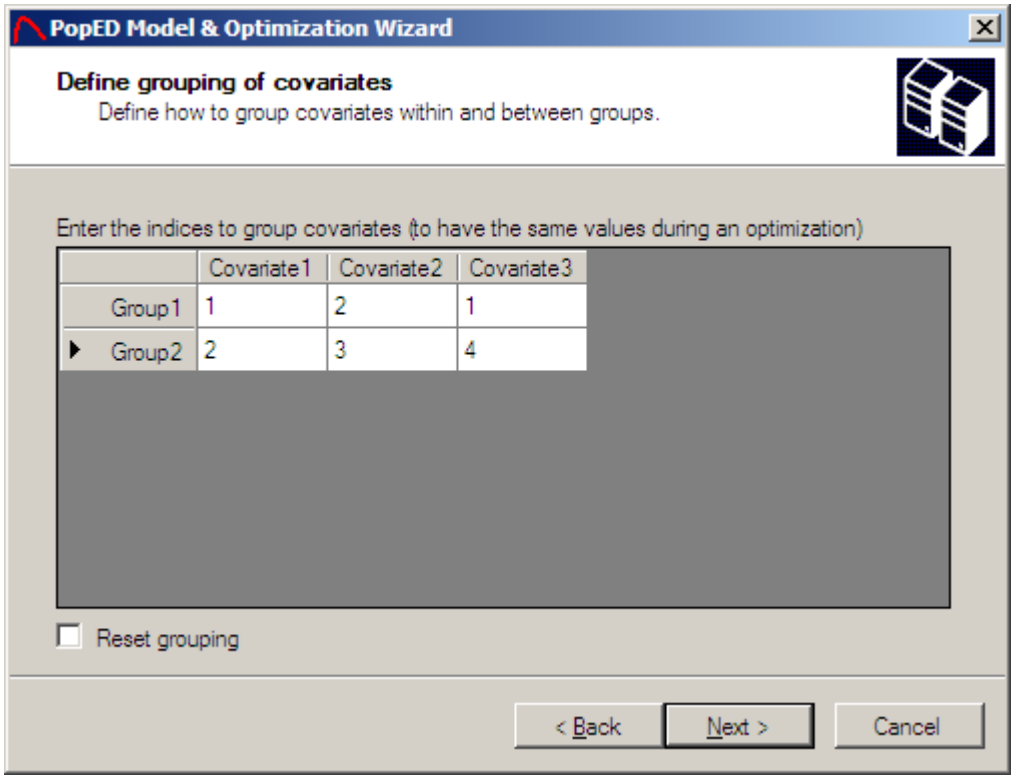
Initial values for covariates



The initial values for the covariates should be entered for all covariates and all subjects (groups/individuals). A minimum value for the covariate and a maximal value for the covariate should be entered as well. The minimum and maximal value is used when optimizing over a covariate. A description for the covariate could be entered as a string. If there are several groups it could be convenient to enter initial, min and max values for all groups at the same time. This is done by checking the *Use the same initial values for all groups/individuals* checkbox. The numbers of covariates are determined by the number specified in the population parameters windows.

A covariate can be fixed even if the optimization method optimizes over the covariate values. To fix a covariate value set the min and max value to the initial value.

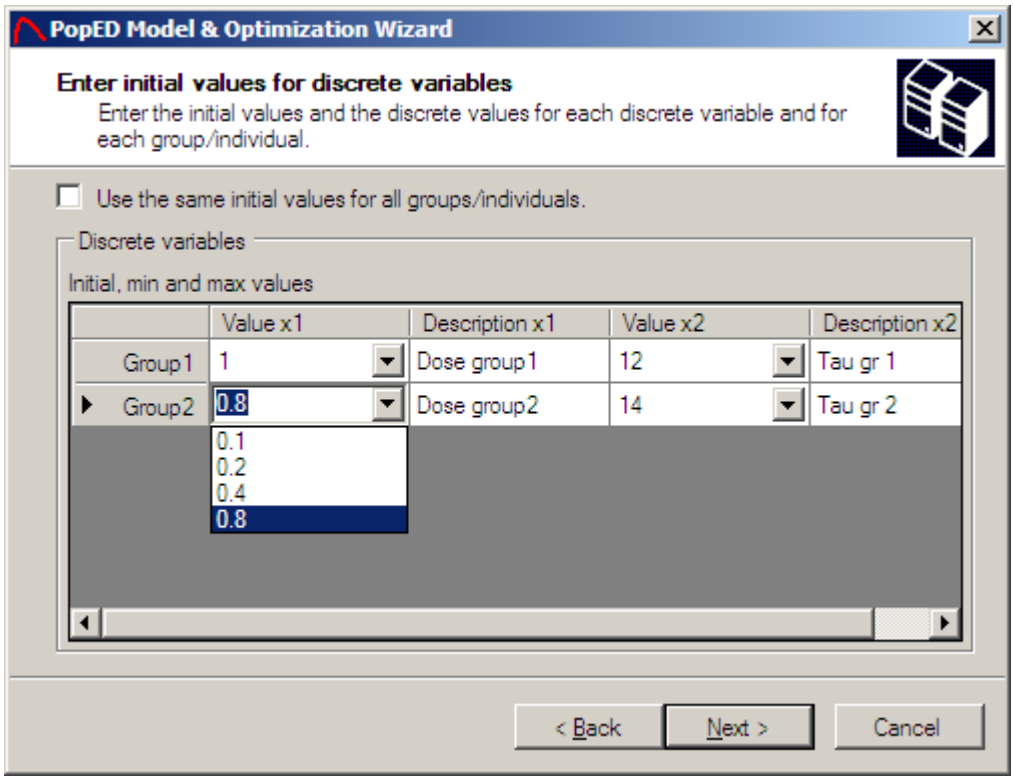
Grouping of covariates



There is a possibility to group covariates between subjects and within a subject. If covariates are said to be grouped it is similar as they will have the same value during an optimization. In the example above Covariate 1 in group 1 will have the same value as the third covariate in group 1. Also covariate 2 in group 1 will have the same value as the covariate 1 in group 2.

Covariates grouped together should have the same initial, max and min value. The grouping should be defined in an increasing order like the example above. Remove the grouping by clicking *Reset grouping*. Pointing with the mouse on a certain covariate will show the grouping for that covariate.

Initial and Possible Values for Discrete Variables

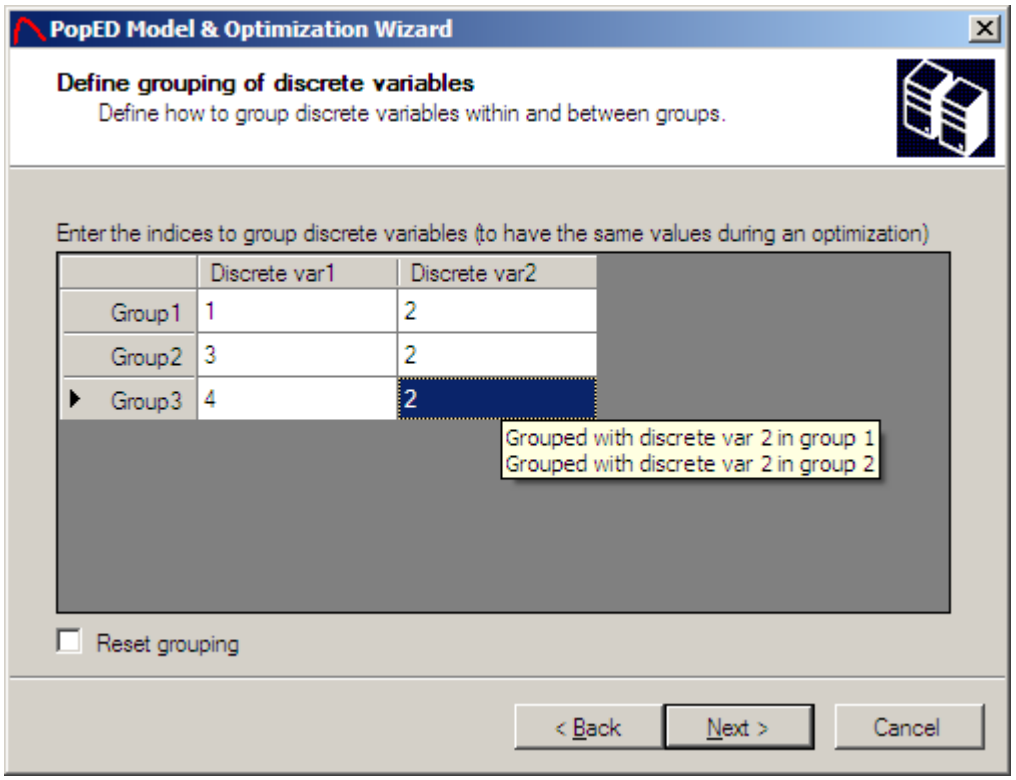


All the possible values for a discrete variable should be entered for each variable and each subject. The initial value for a discrete variable is the value visible in the cell, in the case above 1, 12, 0.8 and 14 are the initial values for the discrete variables. A description for the variable could be entered as a string. If there are several groups it could be convenient to enter initial and possible values for all groups at the same time. This is done by checking the *Use the same initial values for all groups/individuals* checkbox. The numbers of discrete variables are determined by the number specified in the population parameters windows. Note that the number of possible values in each discrete variable doesn't have to be the same, not between different discrete variables within a group and nor between the same discrete variable between two groups (if not grouping of discrete values are used, see below).

A discrete variable can be fixed even if the optimization method optimizes over the discrete variables values. To fix a discrete variable value, have the initial value as the only possible value.

There is a possibility to group discrete variables between subjects and within a subject (see example 1 and below).

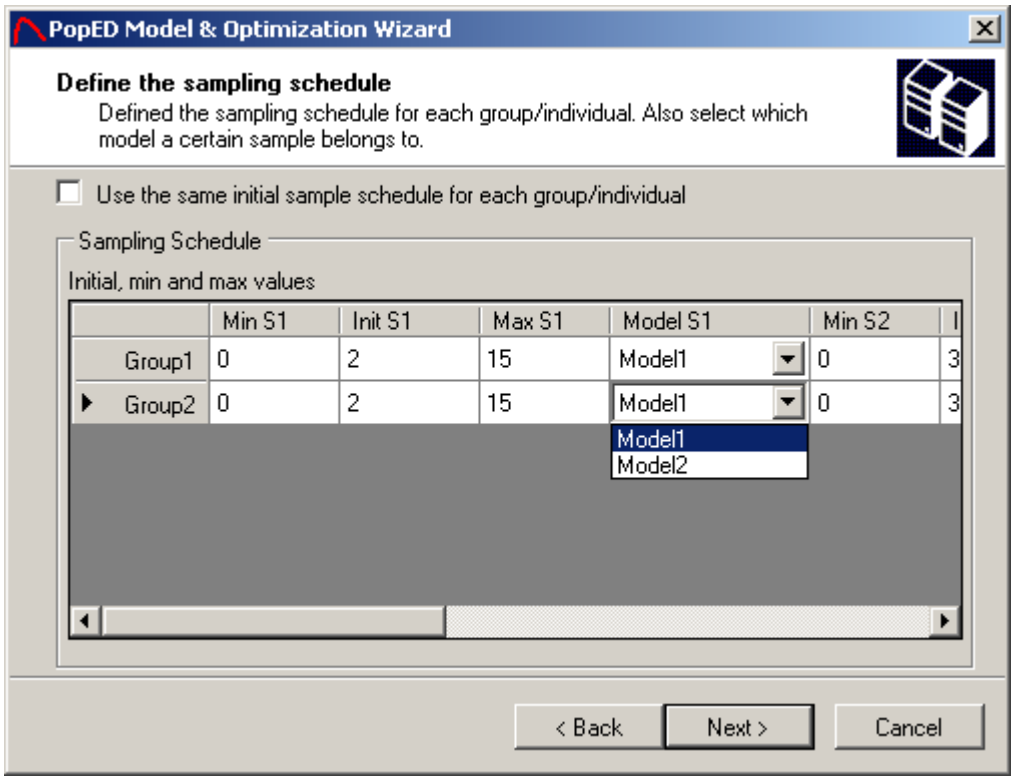
Grouping of Discrete Variables



There is a possibility to group discrete variables between subjects and within a subject. If the discrete variables are said to be grouped it is similar as they will have the same value during an optimization. In the example above the discrete variable 2 is the same for all three groups.

Discrete variables grouped together should have the same initial and possible values. The grouping should be defined in an increasing order like the example above. Remove the grouping by clicking *Reset grouping*. Pointing with the mouse on a certain discrete variable will show the grouping for that discrete variable.

Initial Sampling Schedule

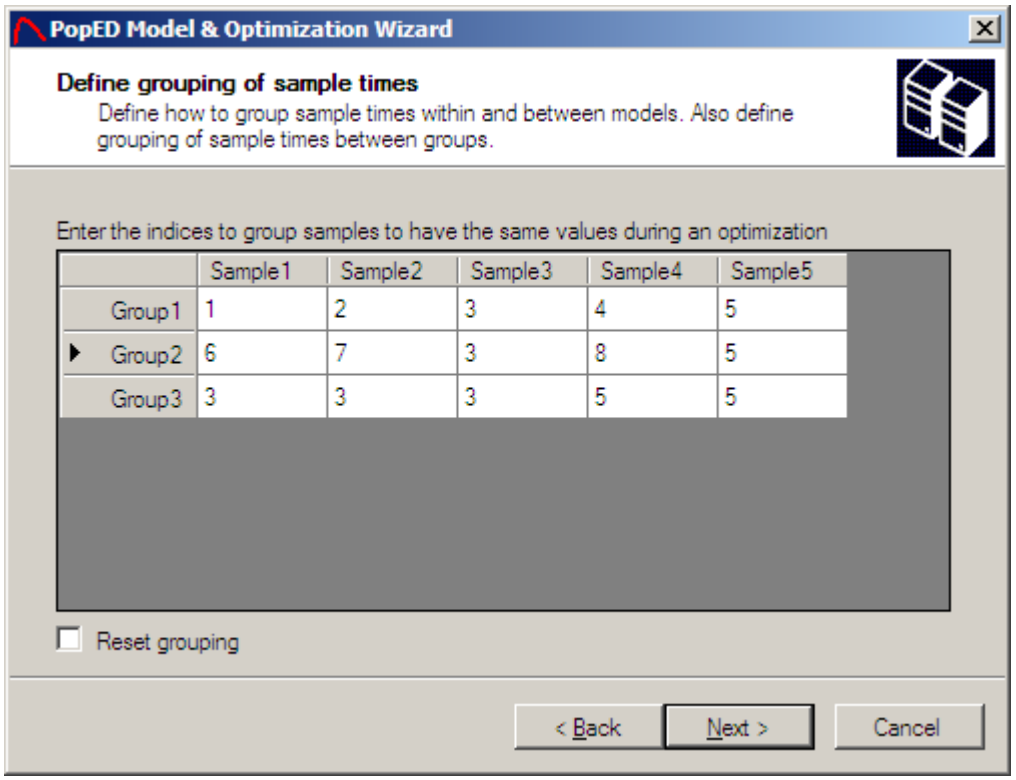


The sampling schedule is defined for all subjects (groups/individuals) and the sample belongs to a sub-model. For all samples; a minimum and maximal sampling time could be specified. This is important when optimizing over the sampling schedule but will not affect the design if the optimization is not performed over the sampling schedule. Number of samples per group will be specified in the number of samples window. If a subject has fewer samples than another subject the subject with fewer samples should define as many samples as the subject with more samples. These samples will not be taken into account when optimizing but are entered to make a consistent matrix of sample times versus subjects. If there are several groups it could be convenient to enter initial, min and max values for all groups at the same time. This is done by checking the *Use the same initial values for all groups/individuals* checkbox.

A sampling time can be fixed even if the optimization method optimizes over the sampling schedule. To fix a sample time set the min and max sample time to the initial sample time.

There is a possibility to group sampling time between subjects and within a subject, e.g. between models (see example 1 and below).

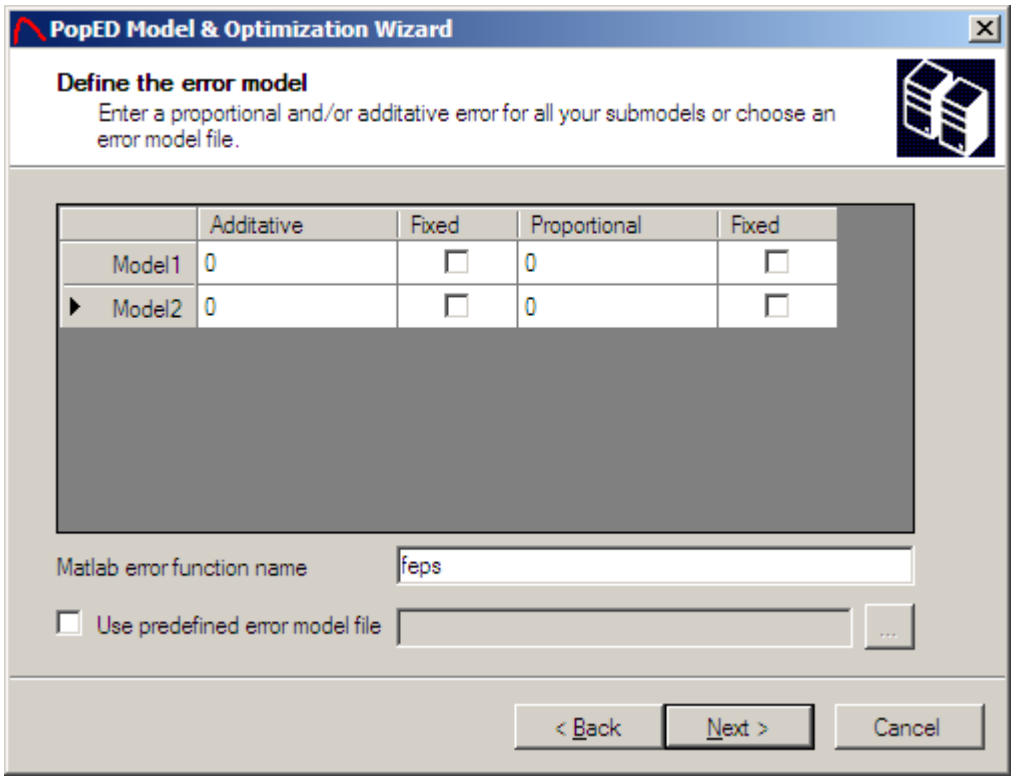
Grouping of samples



There is a possibility to group samples between subjects and within a subject. If the samples are said to be grouped it is similar as they will have the same value during an optimization. In the example above the sample 3 should be the same in all groups. Further the sample 1 and sample 2 in group 3 should be the same as the sample 3 in group 1. Finally sample 5 in all groups and sample 4 in group 3 should be the same as sample 5 in group 1.

Samples grouped together should have the same initial, max and min value. The grouping should be defined in an increasing order like the example above. Remove the grouping by clicking *Reset grouping*. Pointing with the mouse on a certain sample will show the grouping for that sample. It is also possible to group samples that belong to different models.

Define the Residual Variability



The residual variability model could be defined either by an additive part and/or a proportional part or by a predefined error model file. When there is only an additive and/or a proportional part in the error model the error definitions in the PopED GUI works perfectly fine. If more complex error models is going to be used a predefined error model file should be used.

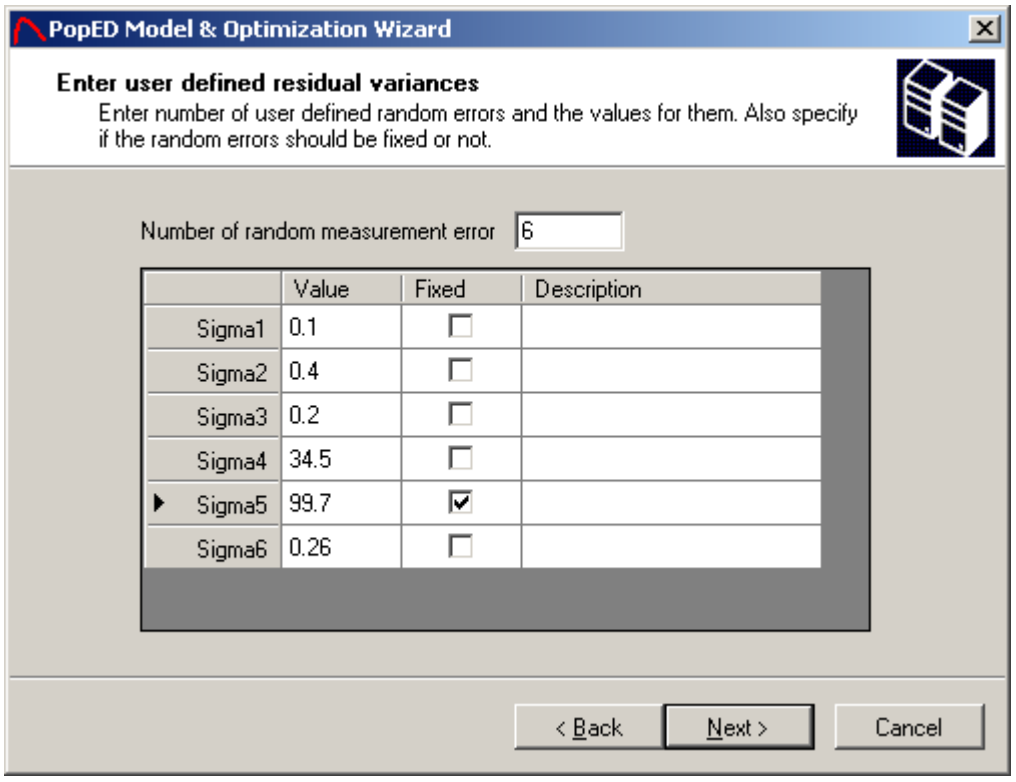
The error values are entered as variances of the residual normal distribution. In NONMEM these variances/variables are called SIGMA.

The *Matlab error function name* could be changed to prevent that the previous error model will be overwritten. This name should be a valid Matlab function name.

If a residual error parameter exists in the model, but is of no interest for estimation, the parameter could be fixed by checking the checkbox corresponding to the variable that is going to be fixed. This means that the variable would not be a row in the Fisher Information Matrix (FIM) and the therefore the FIM wouldn't give any expected parameter uncertainty (SE (%)) of that residual parameter.

In the case with no error model file specified; one additive error and one proportional error could be defined for each sub-model.

If the *Use predefined error model file* is checked a new window were the residual error variances could be entered will be available when pressing the next button.



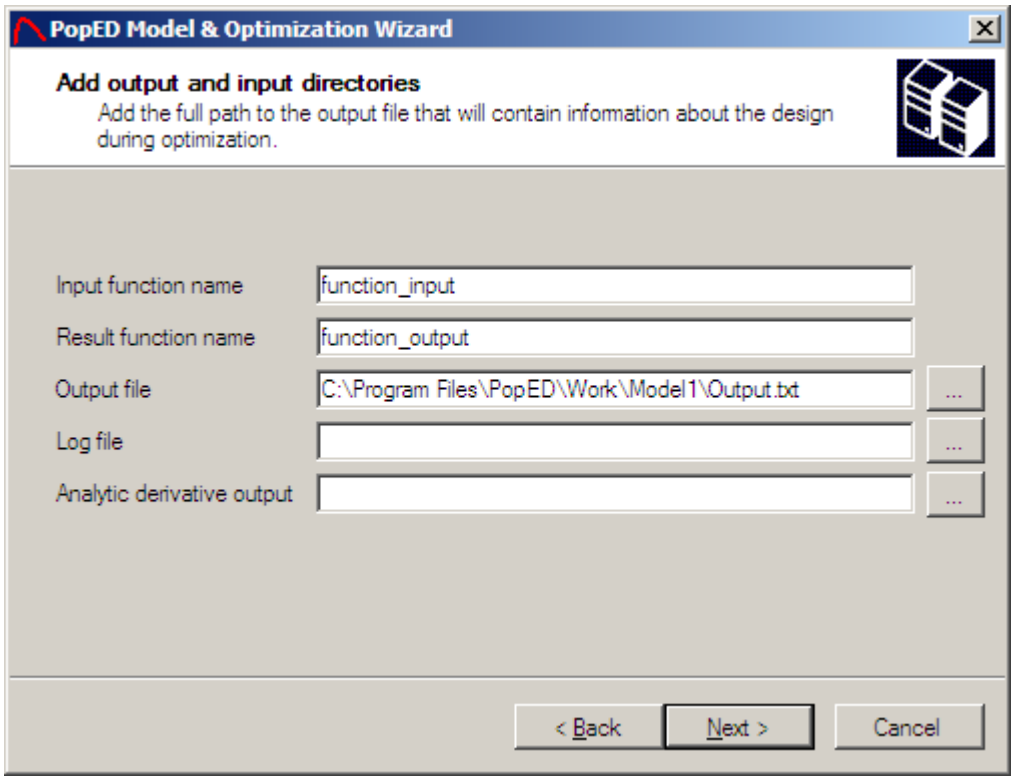
The error model file should be written in Matlab code and have the following syntax:

```
function [y,globalStructure] = feps(model_switch,xt,g,epsi,
globalStructure)

%Error model for model 1 and model 2

[y,globalStructure]=globalStructure.ff_pointer(model_switch,xt,g,globalStru
cture);
for i=1:size(xt,1)
    if (model_switch(i)==1) %The error model for model 1
        y(i) = y(i).*(1+epsi(:,1).*(epsi(:,3).^(epsi(:,4))));
    end
    if (model_switch(i)==2) %The error model for model 2
        y(i) = epsi(:,2) + y(i).*(1+epsi(:,5).*exp(epsi(:,6)));
    end
end

end
```

Input should be a vector with the model switches, the sample time's xt, the parameter definition vector g and the vector with the individual residual error epsi. The last input parameter is a Matlab structure and contains information about the differential equation solver. See Advanced Settings for more information. Epsi is called EPS in NONMEM.

There is also possible to enter a description for the residual error variance parameter. The description could be any string.

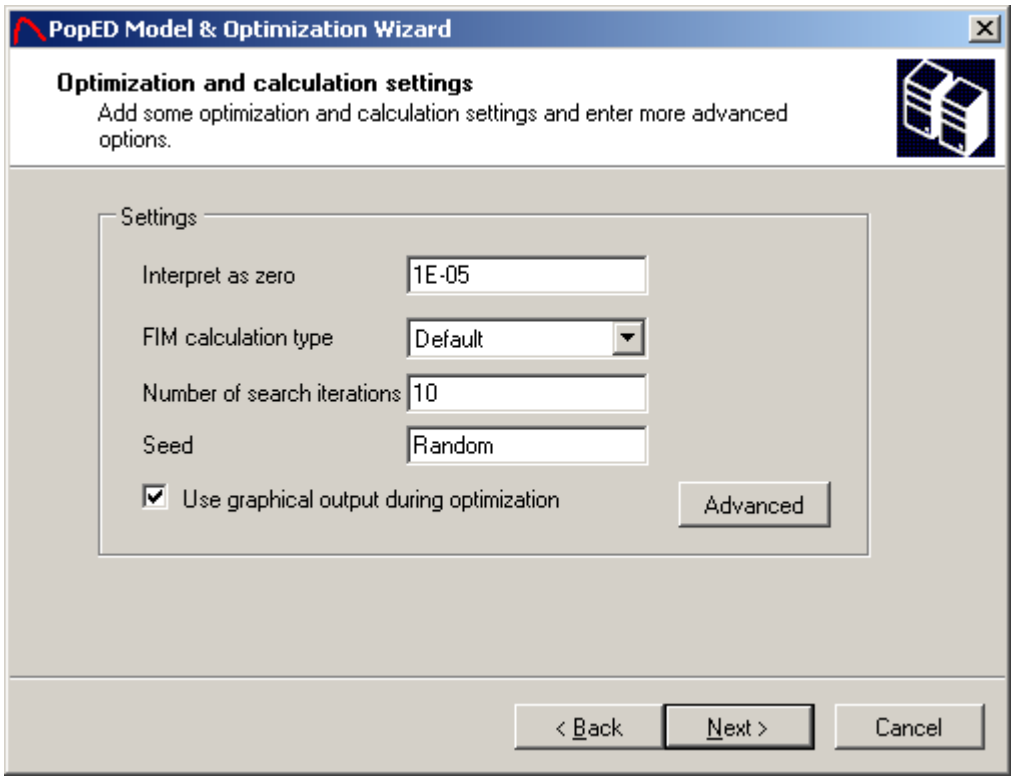
Output and input directories



Enter the name of the Matlab input function and the Matlab result function structure created by Matlab before running PopED. The name should be a valid Matlab function name. The input function will be overwritten each time a new run is started (from the GUI) but the result function will create a new function_output_x where x is the next available number in the running directory.

Also enter the full path to the output file containing information about the current design during optimization. Random Search and Stochastic Gradient will append RS_SG_i.txt to the output path, where i is the search iteration number. If Line Search is used a file with the appendix LS_i.txt, where i is the search iteration, will appear in the output file directory. There must be an output path and file specified to save the optimization settings. The log file is not used in this version of PopED. The Analytic derivate output will store the analytic derivatives for each sub-model in a *.txt file specified by the *Analytic derivative output* path.

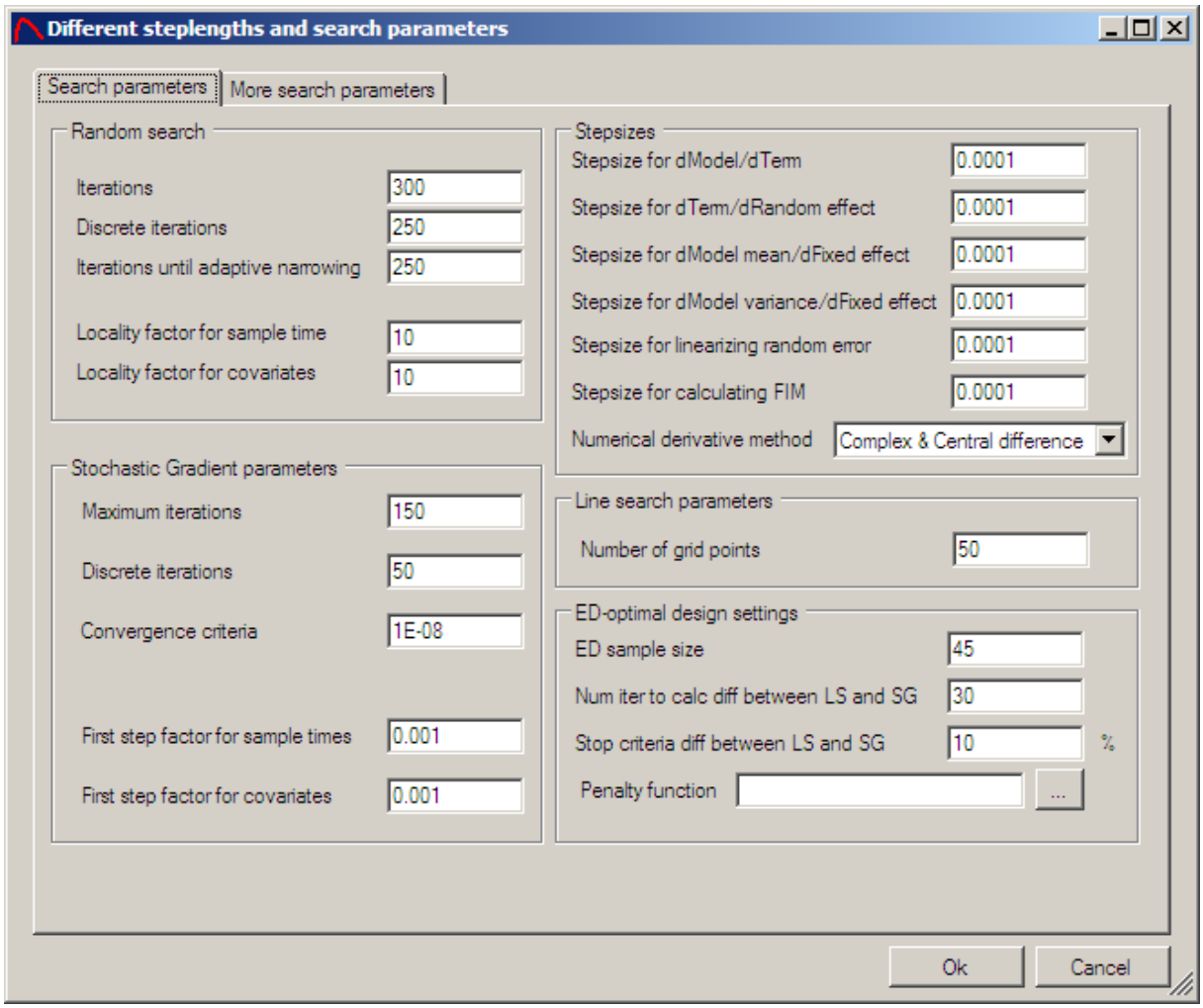
Optimization and calculation settings



When working with numerical problems numerical issues can arise when numbers are exactly zero, therefore PopED needs to define a small number that will be interpret as zero. The Fisher Information Matrix (FIM) calculation type is by default the full FIM calculation but can be changed to a reduced FIM. In the reduced case it is assumed that the derivatives of the variance of the model with respect to the typical values (bpop) are zero. This is a quite ruff approximation but has been shown to give similar design as the full FIM because of the fact that in most models the variance is explained by the inter-individual variability and the residual variability. It's recommended to use the full FIM by default but the reduced FIM can decrease the run time if necessary. The number of search iterations is used when Line Search is not in the search algorithm. See optimization Settings. The seed could be set to a number or bet set to *Random*. A Random seed shouldn't affect the optimization result but could clearly affect the simulation function. See simulate from model. The checkbox *Use graphical output during optimization* will show the current design. In this version it will only show some of the variables, i.e. the sampling times, the covariates and the discrete variables.

Advanced settings for the search algorithms and convergence criterions for ED-optimal design could be entered by clicking the advanced button.

Advanced Settings



A lot of different search settings and step sizes could be entered. Below follows an explanation of the different options.

Random Search Settings

Iterations are the number of random search iterations when optimizing over a continuous variable. The number of discrete iterations is the number of Random Search (RS) iterations to use when optimizing over discrete variables (e.g. Number of samples/Subject, Number of individuals in group). The RS algorithm used is adaptive and therefore it will narrow the search when the number of iterations increases. The adaptive narrowing is started when the Random Search doesn't find a better Objective Function Value (OFV) within the number of iterations entered in *Iterations until adaptive narrowing*. The locality factor for RS will control the localness of the RS algorithm before the adaptive narrowing. If for example a sampling time is in the range of [5-15] hours a locality factor for the sample time of 4 will let the initial RS cover a range of (15-5)/4 = 2.5 hours from the initial sample time. A larger locality factor will produce a better optimization, but then the number of iterations must also be increased.

Stochastic Gradient Settings

Maximum iterations are the maximal number of Stochastic Gradient (SG) search iterations when optimizing over a continuous variable. The number of discrete iterations is the number of SG iterations to use when optimizing over discrete variables (only number of samples per group and group size). The convergence criteria for the SG in ED-optimal design will stop the search when the difference between previous optimal variable value and the variable value after taking a gradient step is less or equal to convergence criteria. In D-optimal design the convergence criteria represent the difference between the current OFV and the OFV of the previous iteration. If the difference is less or equal to the stop criteria the SG search will stop. It's by default set to 1E-08. Stochastic Gradient performs steps in the experiment design variables domain that are guided in direction by the gradient. The magnitude of these steps requires them to be scaled according to iterations history. The first iteration requires a user-defined step, expressed relative to the size of the defined range of the parameters. The required values are the relative values (≤ 1) specifying the size of the first step of the SG algorithm. Increasing the first step factor increases convergence speed, but at the same time increases the probability that the first step moves away from an optimal OFV found by the RS, thus increasing the probability of ending in a local optimum.

Line Search Settings

The number of grid point could be specified in the line-search. That is the number of point each defined range (e.g. a sample time), will be split in. If optimizing over a time point with a range of 5-15 hours and the number of grid points is 50 the OFV will be calculated in 50 points from 5 -15 hours with a step length ≈ 0.2. The best OFV will give the sample time to proceed with. The LS could not be used when optimizing over discrete variables. If LS is used in the optimization the optimization will converge if the LS don't change the OFV from the previous SG or RS. For a discrete variable x the number of points in LS doesn't matter, the number of points will automatically be set to the number of possible values of the discrete variable.

Step sizes

The step sizes are used to define the accuracy when approximating the derivative of different functions numerically. All derivatives used to calculate the Fisher Information Matrix (FIM) are approximated with a central difference approach with a truncation error of $O(h^2)$ where h is the step size. A forward difference approximation is used in the SG algorithm to calculate the derivative of FIM with respect to a continuous variable, e.g. the sample time. This method will have a truncation error of O(h).

ED-Optimal design settings

If the optimization method used is ED-optimal design some ED-optimal settings must be defined. ED sample size is the number of samples that will be drawn from the parameter distribution to calculate the expectation value. One Fisher Information Matrix is calculated for each ED - sample, therefore the execution time will increase linearly with the number of samples. If a user defined distribution is used that are not

stochastic this number should be set to the number of parameters in the user defined distribution and the stop criteria below could be set to a very small number e.g. 0.01%. The number of iterations to calculate the difference between SG and LS is the number of calculations of expectation values that will be used to compare the LS and SG. For example an ED sample size of 45 and the number of iterations to calculate difference between SG and LS set to 30 will yield 45*30 = 1350 calculations of the FIM. When using a user defined distribution that is not stochastic the number of iterations should be one. This is only to have an automated way to look if the search has converged. The ED search converges if the LS don't find a better OFV within the previous SG value multiplied with the percentage value.

If a penalty function is used the objective function value to optimize on could be specified in the penalty function. This enables the user to write e.g. user specified cost functions based on the FIM to be optimized over. This could also be used to optimize over D-optimal cost functions by setting the ED sample size to 0 and write the OFV in the penalty function. Other use of the penalty function is to add penalty to different types of ED-samples and to add restrictions in general to the design.

The penalty function should have a function header defined like this:

```
[ED_fim,ED_ofv,globalStructure]=penalty_function(fim_list,bpop_gen,
                      d_gen_list,docc_gen_list,model_switch,groupsize,ni,xt
                      optn,xoptn,aoptn,bpopdescr,ddescr,covd,sigma,docc,
                      globalStructure);
```

fim_list is a vector with OFV values calculated for each ED sample. If the ED sample size is set to 0, this vector will be empty.

bpop_gen is a matrix with each row of the matrix assigned a vector with the typical values used to calculate the OFV. If the ED sample size is set to 0, the matrix will be empty.

d_gen_list is a list with each element of the list assigned a vector with the inter individual variances used to calculate the OFV. If the ED sample size is set to 0, the list will be empty.

docc_gen_list is a list with each element of the list assigned a vector with the occasion variances used to calculate the OFV. If the ED sample size is set to 0, the list will be empty.

model_switch is a vector defining which sub-model a certain sample belong to.

groupsize is a vector with the current optimal group size for each group.

xoptn is a matrix with the current optimal discrete values for each group.

aoptn is a matrix with the current optimal covariate values for each group.

bpopdescr is a matrix with the mean values, the variances and the distributions of each typical value.

ddescr is a matrix with the mean values, the variances and the distributions of each inter individual variance.

covd is a vector defining with the lower triangular matrix of the covariance between each inter individual variance.

sigma is a matrix defining the residual variances.

docc is a matrix with the mean values, the variances and the distributions of each occasion variance.

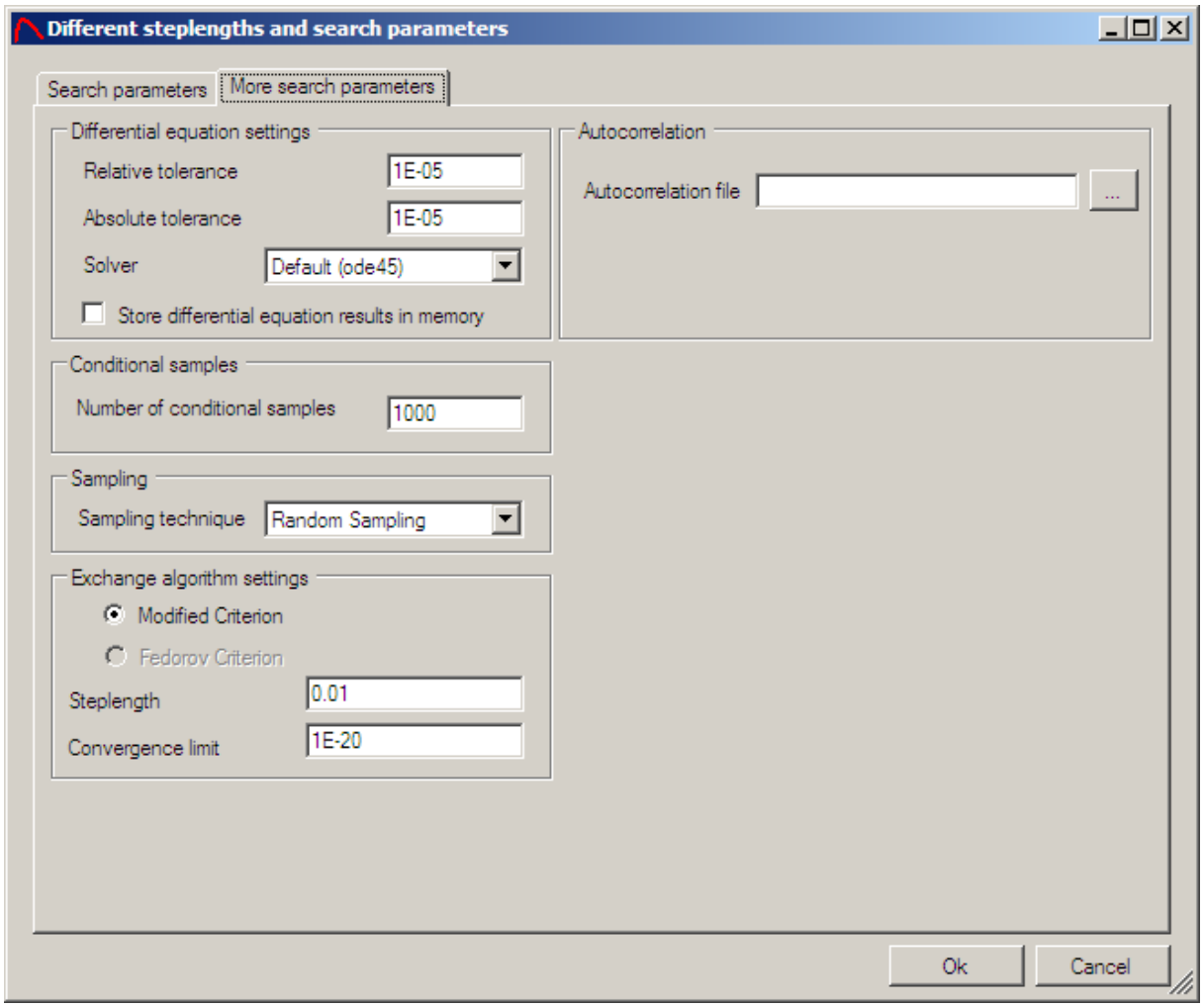
globalStructure is a structure that contains all variable that PopED uses..

ED_fim is a returning value that should contain the Fisher Information Matrix calculated in the penalty function. The matrix should be scaled to the number of ED samples, i.e. it should be the expectation FIM.

ED_ofv is a returning value that should contain the OFV that should be maximized, calculated in the penalty function. The OFV should be scaled to the number of ED samples, i.e. it should be the expectation OFV.

globalStructure is a return structure, see above.

The optimization procedure will try to maximize the OFV based on the penalty function with respect to the design parameters, i.e. the *samples times*, the *discrete variables*, the *covariates*, the *groupsize* and the *number of samples in each group*.



Differential equation settings

For the differential equation solver the relative tolerance (RelTol) and the absolute tolerance (AbsTol) must be set. The solvers will then have an error precision of $e_i = \max\left(relTol \cdot |y_i|, absTol\right)$. The solver method is by default the *ode45* solver but could be changed to a stiff solver (*ode15s*) if the differential equation is stiff. The RelTol, AbsTol and differential equation method could be changed "on the fly" during an optimization. *Store differential equation results in memory* enables the user to store the solutions of the differential equation solver in the memory. This could increase speed a lot when optimizing over number of samples or sampling times. The option should **not** be used when optimizing on continuous covariates.

Numerical derivative method

There are different methods to calculate numerical derivatives. In PopED two different approaches (except analytic derivatives) are available:

- Complex & Central difference – PopED uses complex differentiation when possible, otherwise central difference is used. The complex approach is very robust and accurate for small step lengths and will sometimes be twice as fast as the central

difference approach. The precision for the first order complex difference is $O\left(h^2\right)$ where *h* is the step length.

- Central difference – PopED could use the central difference approach. This could be valid when the function couldn't handle complex numbers or when the complex approach is slow due to complex numbers. The precision for the first order central difference is $O\left(h^2\right)$ where *h* is the step length.

Number of conditional samples

If the approximation type is first order conditional or first order conditional with interaction the number of individuals samples could be specified here. A large number of samples will yield a more robust design against different individuals in a coming study but will take more time to compute.

Sampling technique

The sampling technique is used to generate simulations, to generate sample for ED-optimal design and to generate samples for conditional designs. The normal technique uses standard random numbers (see Matlab manual) while the Latin Hypercube Sampling is more evenly distributed within the sampling region. This could have advantages in ED-optimal design in terms of faster convergence and robustness.

Exchange Algorithm settings

The Exchange Algorithm settings defined the step length and the convergence criterion when the exchange algorithm is used. In the current version only the modified criterion could be used. This criterion will make the search stop if the percentage change of OFV is less than the criterion. The step length defines how the samples times/covariates should be tested. I.e. if the max and min border of a sample time is t=[0,10] and a step length of 0.1 is used, the samples times tested will be [0, 0.1, 0.2, …., 9.9, 10]. If the script version is used the step length could be changed to number of points instead which prevents the search space to be too large when the borders of the sample times and covariates are unequal. If e.g. 50 points are used the sample times/covariates tested are (max-min)/numpoints. When optimizing over discrete variables, i.e. x, the number of points will be the number of possible values for a certain discrete variable.

Autocorrelation file

PopED enables a possibility to define a user defined variance term of the linearized model. This enables features as e.g. optimizing on models with auto correlation. The filename (*.m) and the path of the user defined variance must be specified.

The user defined variance Matlab function should have a header defined like this:

```
var=user_variance(h,model_switch,xt,x,a,bpop,b,
```

```
        bocc,d,sigma,docc,globalStructure);
```

h is a matrix containing the linearized residual model with respect to the residual parameters.

model_switch is a vector defining which sub-model a certain sample belong to for individual/group i.

xt is a vector with the current optimal samples for individual/group i.

x is a vector with the current optimal discrete values for individual/group i.

a is a vector with the current optimal covariate values for ind/group i.

bpop is a vector with the typical values.

d is a matrix with the inter individual variances and covariances.

sigma is a matrix defining the residual variances.

docc is a matrix with occasion variances.

globalStructure is a structure that contains all variable that PopED uses..

var is a return value that should contain the linearized variance model belonging to the residual error, the size should be num_samples x num_samples for individual/group i. Note, *var* is only the linearized variance with linearization purely done for the residual error model. If the complete variance of the model should be user defined, the var must subtract the linearized model with respect to b and possibly the interaction variance if interaction is used.

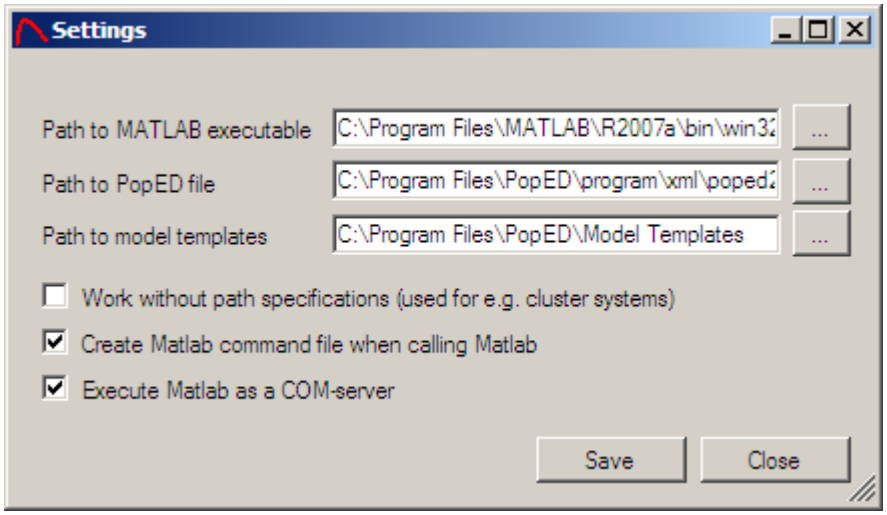
Save the PopED settings

The settings can be saved by using the save function (Ctrl+S) or Save as under the file menu or by walking thru the wizard. If a settings file is saved in the model template directory the model templates will be updated with the new model template. A warning will occur to prevent adding unnecessary files to the model templates directory.

Open a PopED settings file

A PopED Settings file (*.xml) can be open by using Open under the file menu.

PopED GUI Directories Settings



A path to the Matlab executable, usually <Matlab dir>\bin\win32\matlab.exe must be entered to run PopED within the GUI. A path to the PopED script must also be entered to allow for execution of PopED runs within the GUI. This should be <PopED dir>\programs\poped2_0.m. A path could also be set to the model templates directory. If no path should be used in the xml files, the checkbox *Work without path specifications* could be checked. This allows the user to specify all input and output files in a relative manner. This is particularly useful when running the GUI on a separate system than the Matlab installation, i.e. when the GUI is used to set up the design and then run on a separate system using e.g. poped2_0('myfile.xml'). The *Create Matlab command file when calling Matlab* stores every call to Matlab in a Matlab function file, usually *poped_cmd.m* (possible to change in the *config.xml*, see below). The option *Execute Matlab as a COM-server* uses the COM server technology to communicate with Matlab. If this option is not set, the program execute Matlab via command line options and the Matlab process is closed automatically or need to be closed to get the input back to the GUI. This option doesn't use any COM technique and is therefore suitable for running PopED GUI on e.g. Mono or other third party software that enables .NET implementation. When the COM Server option is not used, live log files are visible during the optimization run. All the settings will be stored in the *config.xml* located in the PopED installation directory when pressing save.

Diagnostics before the optimization

Some diagnostics can be used to see that the model works correctly prior to the optimization procedure. The model typical values could be plot and simulation could be performed to validate the model.

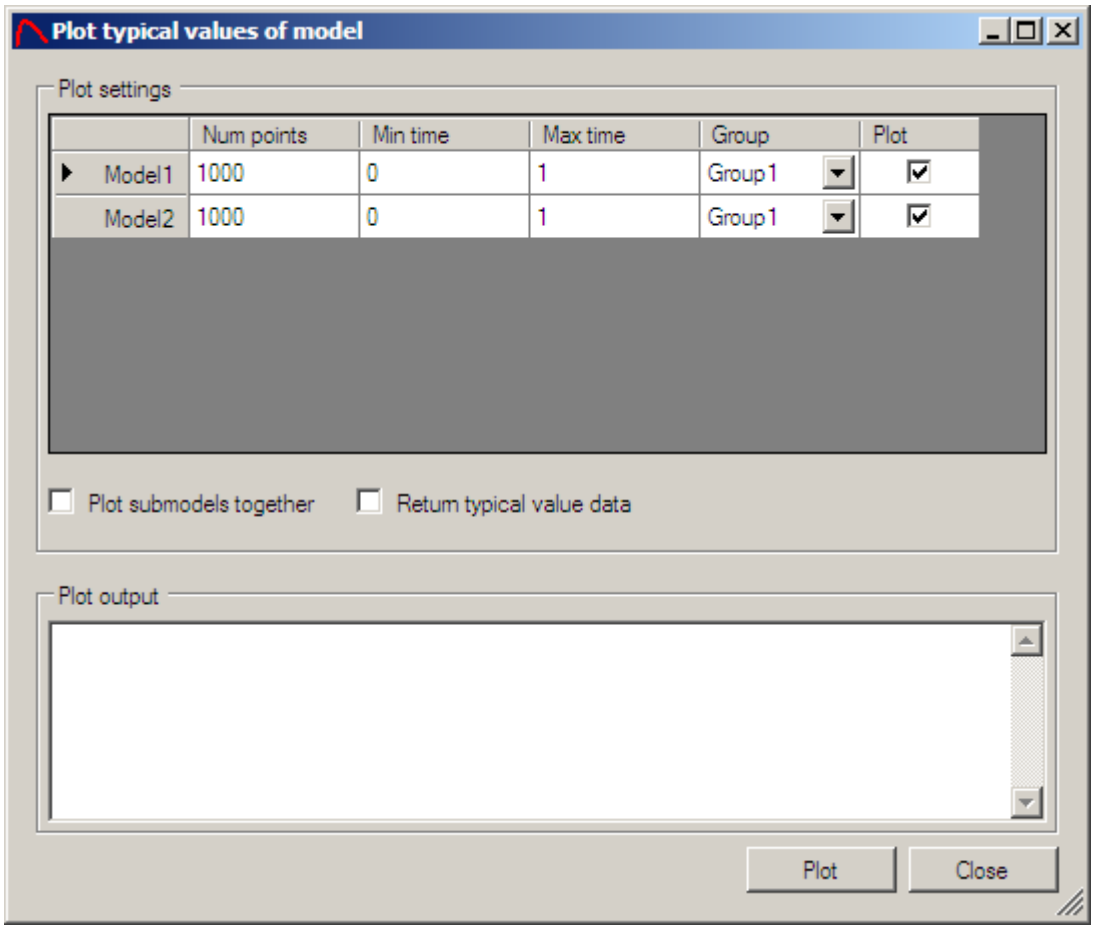
Convert PopED Settings to script version

The *poped2_0* function in Matlab can take the *.xml PopED GUI settings as an input or it could take a Matlab function representing the settings. A Matlab function containing the settings information is created when the function *poped2_0* is called or can be created by a call to the function *convert_xml.m* that will take a *.xml filename as input. The Matlab function containing the settings is often one of the inputs to the diagnostic functions.

Run PopED script commands from the GUI

A number of script commands in PopED can be run from the GUI. The GUI will convert the current settings file to a Matlab function (see above) and call the script command via the Matlab COM Server. This is all done in the GUI code but the Matlab COM Server needs to work properly to run PopED script from the GUI. Otherwise the PopED script version must be used (i.e. run optimization and plots from Matlab command line) or the setting *Use Matlab COM Server* must be unchecked.

Plot Typical Values



PopED GUI has a function that could plot the typical values for all sub-models. The function is available if a PopED setting is saved and open and it's accessible from the Optimal Design menu.

For each sub-model the number of points that will be used to calculate the dependent variable must be entered. For example if the number of time points are 1000 and the time interval are 0-1 the dependent variable (DV) will be calculate every 0.001 time unit. It is also possible to plot the typical value from a certain group (the covariate values can affect the plot, e.g. different doses in different groups). A checkbox indicates whether this sub-model should be plotted or not. If *Plot sub-models together* is checked all the sub-models will be plotted in the same window/figure. The plot output will be the DV for all the sub-models that are plotted and figure windows will show all the sub-models. In the figure window; the data, titles, axis settings etc. could be modified and saved. If the *Return typical value data* is checked the output from the plot (the typical values) will be presented in the Plot output.

The script version of plotting the typical values is:

```
model_values = plot_model(bPlotModel,model_num_points,model_minxt,model,
               maxxt,groups_to_plot,popedInput,bShowGraph,bPlotInSame)
```

bPlotModel is a vector with the length of number sub-models indicating if this sub-model should be plotted or not. 1 equals plot, 0 equals not plot.

model_num_points is the number of time points used to plot the sub-model. This is a vector of the length of the number of sub-models.

model_minxt is a vector with the minimal time point to plot for all sub-models. The length is equal to the number of sub-models.

model_maxxt is a vector with the maximal time point to plot for all sub-models. The length is equal to the number of sub-models.

groups_to_plot is a vector with the group number to plot the sub-model with.

popedInput is the PopED settings function created from the xml file (see above). The function is often accessible by typing *function_input()*.

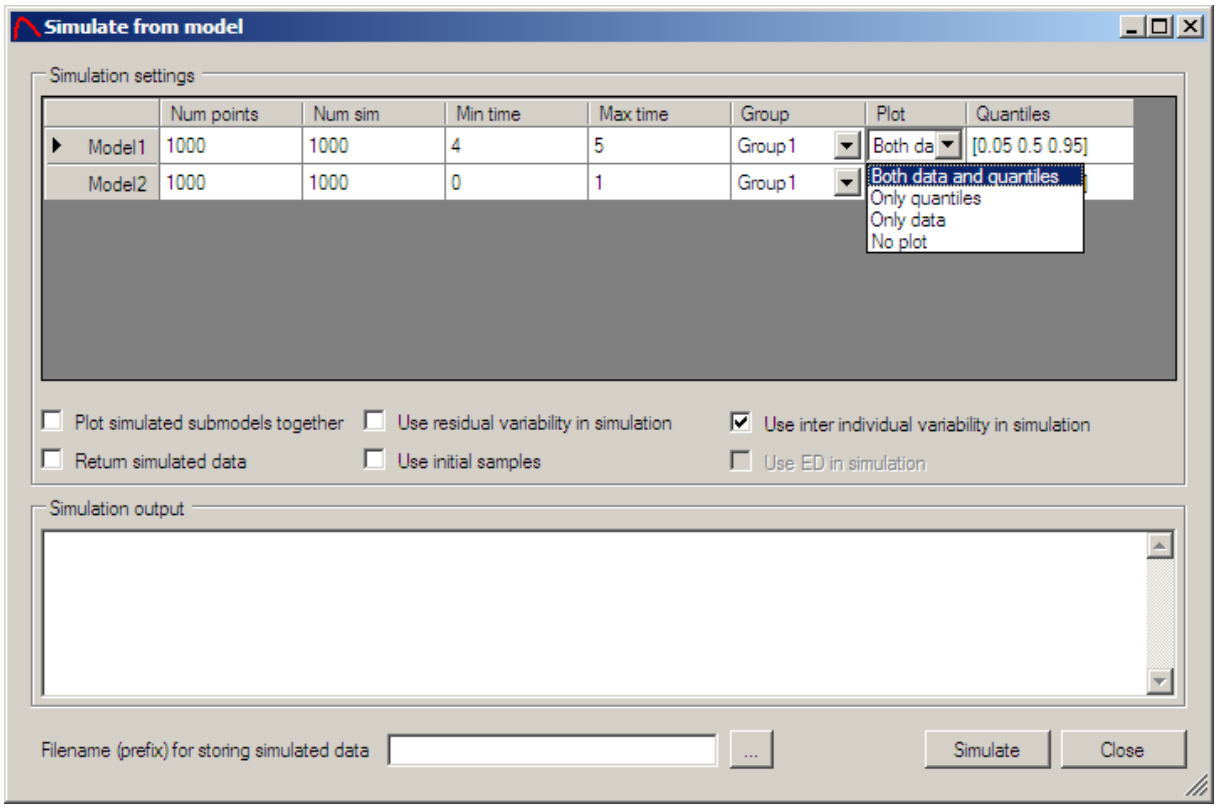
bShowGraph indicates whether a figure will be created or not. If it's set to false or 0 the model_values will be calculated and returned but no figure will be shown. This option could be useful to get the typical values within a script.

bPlotInSame is set to true or 1 if all sub-models should be plotted in the same figure. Otherwise it is set to false or 0.

Example: This will produce the same plot as above:

```
plot_model([1 1],[1000 1000],[0 0],[1 1],[1 1],function_input(),1,0)
```

Simulation of Model



PopED GUI has a function that can simulate data and plot the result for all sub-models. The function is available if a PopED setting is saved and open and it's accessible from the Optimal Design menu.

For each sub-model the number of points that will be used to calculate the dependent variable (DV) is specified, furthermore the num of simulations (number of individual curves to simulate) are specified. All simulations are simulated within a minimum and a maximum time and the group used to simulate can also be specified (the covariates in a group could affect the simulation, e.g. the dose). There is also a plot option that allows for different plots to be created. For each sub-model the plot options are:

- Both data and quantiles – Plot the data from all simulated individual curves and the quantiles specified in the quantiles column.
- Only quantiles – Plot only the quantiles specified in the quantiles column.
- Only data – Plot only the simulated individual curves.
- No plot – Don't plot this sub-model.

In the quantiles column all quantiles that should be plotted from this sub-model could be specified. The quantiles should be specified within brackets ([]) and it should be a value between 0-1. E.g. [0.01 0.5 0.99] will plot the 1 and 99 percentile and the median value of the model given the simulated data. The figure windows produced by the plot could be used to change the plot properties and axis-settings etc. The output returned by the plot is the simulated data but only for the last sub-model that don't have the Plot option – No plot. The data is presented only when the *Return simulated data* is checked. All the sub-models could be plotted in the same window by checking the *Plot simulated submodels together.* The simulation can be done with or without the residual variability by clicking *Use Residual variability in simulation*. If *Use initial samples* is checked the function will only simulate from the current design (the initial samples) and the *Number points* column is ignored. The simulation could also be done with or without inter individual variability by checking the *Use inter individual variability in simulation*. It is also possible to simulate with or without ED-variability by checking the *Use ED in simulation* if the design is defined as an ED design (uncertainty around the parameter values). The sample method used in the simulation could be either Latin Hypercube Sampling or normal sampling, see advanced settings. A filename prefix for storing the simulated data could be defined, one file for each sub-model is created with the prefix name followed by the sub-model number, e.g. my_data_1.csv, my_data_2.csv etc. If no name is specified no data file will be created. The data file is stored as a comma separated file (csv) and the first column represent the individual number, the second column the time and the third column the dependent variable, e.g. the concentration or the response.

The script version of the simulate function is:

```
sim_dat =  plot_simulation(bUseResiduals,bUseIIV,bUseED,
                      model_num_simulations,model_num_points,
                      model_minxt,model_maxxt,
                      groups_to_plot,popedInput,sim_quantiles,bShowVector,
                      bPlotInSame,bUseInitialDesign,strFileName)
```

bUseResiduals is set to true (1) or false (0) if the residuals should be used in the simulation.

bUseIIV is set to true (1) or false (0) if the inter individual variability should be used in the simulation.

bUseED is set to true (1) or false (0) if the ED uncertainty should be used in the simulation.

model_num_simulations is vector with the length of number of sub-models indicating the number of individuals that should be simulated for this sub-model.

model_num_points is the number of time points used to plot the sub-model. This is a vector of the length of the number of sub-models.

model_minxt is a vector with the minimal time point to plot for all sub-models. The length is equal to the number of sub-models.

model_maxxt is a vector with the maximal time point to plot for all sub-models. The length is equal to the number of sub-models.

groups_to_plot is a vector with the group number to plot the sub-model with.

popedInput is the PopED settings function created from the xml file (see above). The function is often accessible by typing *function_input()*.

sim_quantiles is a Matlab cell structure with the length of number of sub-models. The cell structure contains a vector for all sub-models with the quantiles that should be plotted for a sub-model.

bShowVector is vector with the length of the number of sub-models. The show vector will contain information how each sub-model should be plotted and if it should be plotted. The option for the show vector is 0 – Plot data and quantiles, 1 – Plot only quantiles, 2 – Plot only data, 3 – Don't plot this sub-model.

bPlotInSame is set to true or 1 if all sub-models should be plotted in the same figure. Otherwise it is set to false or 0.

bUseInitialDesign is set to true or 1 if the initial design should be used in the simulation. Otherwise it is set to false or 0.

strFileName is set to the path and filename prefix of the csv file where the simulated data should be stored. To skip generate data files, set the value to the empty string ''.

Example: This will produce the same simulation plot as above:

```
plot_simulation(0,1,0,[1000 1000],[1000 1000],[4 0],[5 1],[1 1],
           function_input(),{[0.05 0.5 0.95] [0.05 0.5 0.95]},[0 0],0,0,'')
```

Optimize with PopED

Optimization from the GUI can be performed under the Optimal Design menu, Run optimization. To run the optimization from the script version type poped2_0('file_to_optimize.xml') or call the function with a transformed input function (see Convert PopED settings to script version).

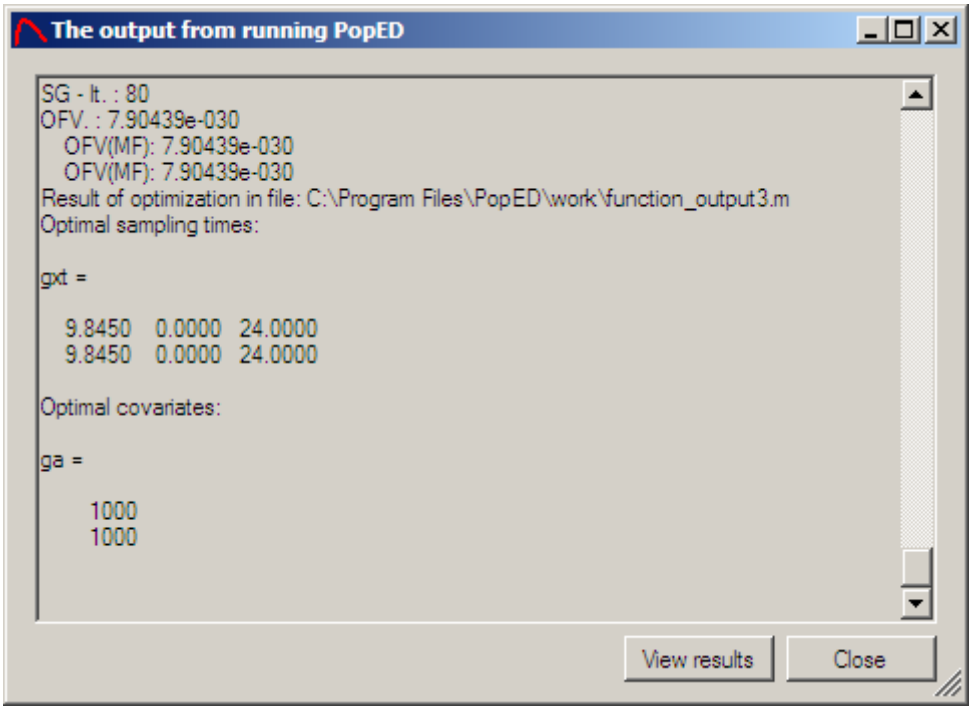
Example:

```
           poped2_0('MyFile.xml');
           poped2_0(function_input());
```

During the optimization, Matlab figure windows will show the optimization current status, i.e. the OFV and the continuous variable values in the optimization. If optimizing over discrete variables no figure windows will be shown in this version. Furthermore no figure windows will be shown if the *Use graphical output during optimization* option is unset (see Optimization and calculation settings).

During the optimization, output to the Matlab command window is outputted in the script version of PopED and the output from the different search algorithms are stored in the output file specified in the PopED settings file (*.xml or the input function *function_input.m* (see Output and input directories)). Stochastic Gradient and Random Search will store their result in the output file with an extension _RS_SG_i.txt, where i is the search iteration number. Line Search will have an extension _LS_i.txt where i is the search iteration number.

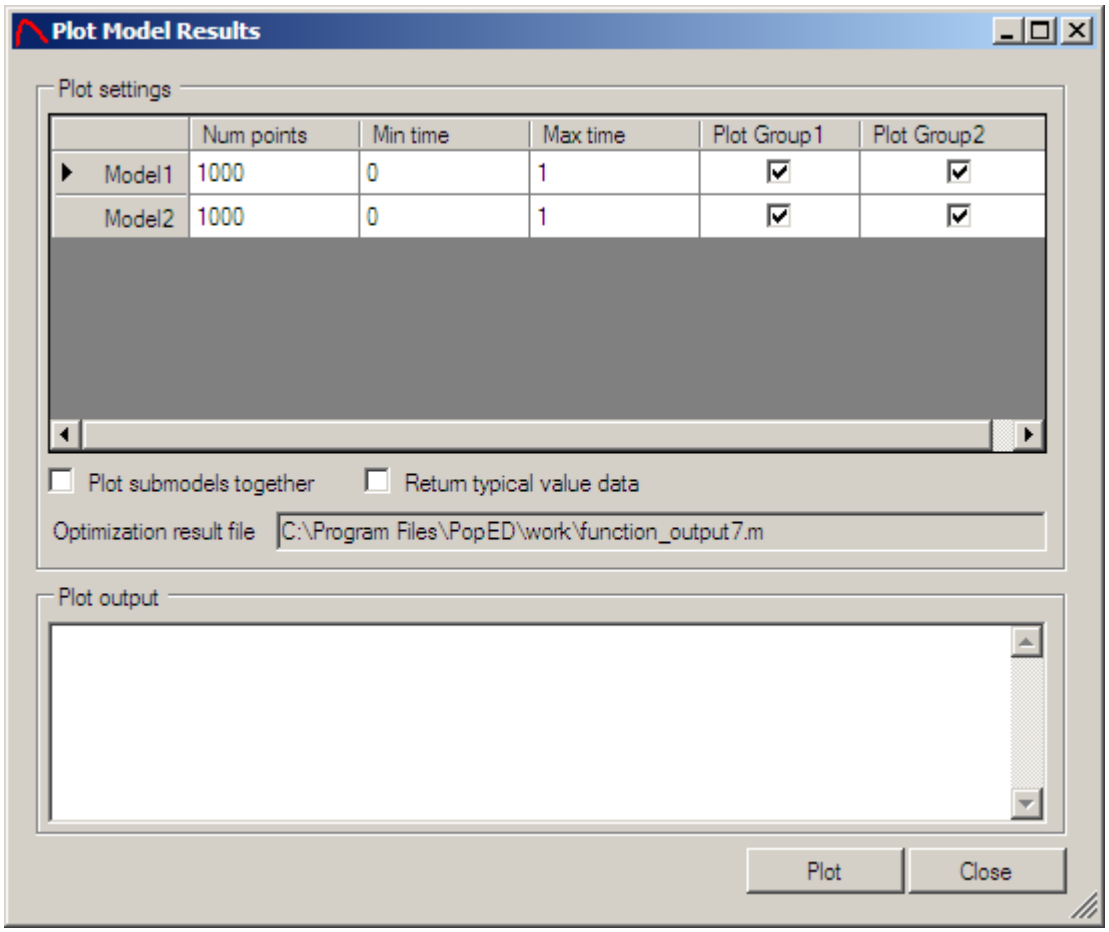
When the optimization are done or if an error occurs, output from the Matlab command window will be shown in an optimization output window for the GUI version and in the Matlab command window in the script version. A Matlab function file with the results from the optimization is also written, the output function is called: *function_outputi.m* (see Output and input directories) where i is the lowest number to create a non-existent filename in the current directory. The same function is also returned from the poped2_0 function as a structure with the same form as the function_output.m.



The continuous optimized variables are outputted as text and the result function filename is also visible. If the optimization didn't terminate due to an error the View results button will be visible.

Plot model results

The optimized result for continuous optimization could be viewed with the plot model result function available in the GUI from the Optimal Design menu. To be able to run the function from the GUI an output function corresponding to the optimization of the open PopED GUI Settings file (*.xml) must be specified.



The function works similar to the plot model function and plots the typical values for the different sub-models. The main difference is that the optimal sample times are marked in the plot with red rings. Like in the model plot function the number of time points to calculate the model with is specified in the num point's column. The time interval to plot over is also specified besides which group in each sub-model to plot. If the *Plot submodel together* checkbox are checked the entire plot will appear in the same figure window. If the *Return typical value data* is checked the typical value are presented in the Plot output.

The script version of plot model result function is:

```
[model_values opt_value opt_time opt_group opt_model] =
plot_model_results(bPlotModel,model_num_points,model_minxt,model_maxxt,pope
dInput,bShowGraph,bPlotInSame,bPlotGroups,popedOutput,clustValue)
```

bPlotModel is a vector with the length of number sub-models indicating if this sub-model should be plotted or not. 1 equals plot, 0 equals not plot.

model_num_points is the number of time points used to plot the sub-model. This is a vector of the length of the number of sub-models.

model_minxt is a vector with the minimal time point to plot for all sub-models. The length is equal to the number of sub-models.

model_maxxt is a vector with the maximal time point to plot for all sub-models. The length is equal to the number of sub-models.

popedInput is the PopED settings function created from the xml file (see above). The function is often accessible by typing *function_input()*.

bShowGraph indicates whether a figure will be created or not. If set to false or 0 the model_values will be calculated and returned but no figure will be shown. This option could be useful to get the typical values within a script.

bPlotInSame is set to true or 1 if all sub-models should be plotted in the same figure. Otherwise it is set to false or 0.

bPlotGroups is matrix with the length of number of groups multiplied with the number of sub-models. E.g. the matrix contains a 1 in the first row and 2nd column if the 2 group should be plotted in the first sub-model.

popedOutput should be a PopED output function e.g. function_output1.

clustValue is a value to determine if a point will be considered clustered or not. If the distance between two optimal samples is within the clustValue, a value with the number of points that are clustered will be written in the plot, next to the clustered points. If no clustering should be checked the value should be set to 0.

model_values is a output matrix containing the data from all plotted sub-models. The model_values does not necessarily contain the optimal sampling points. See below.

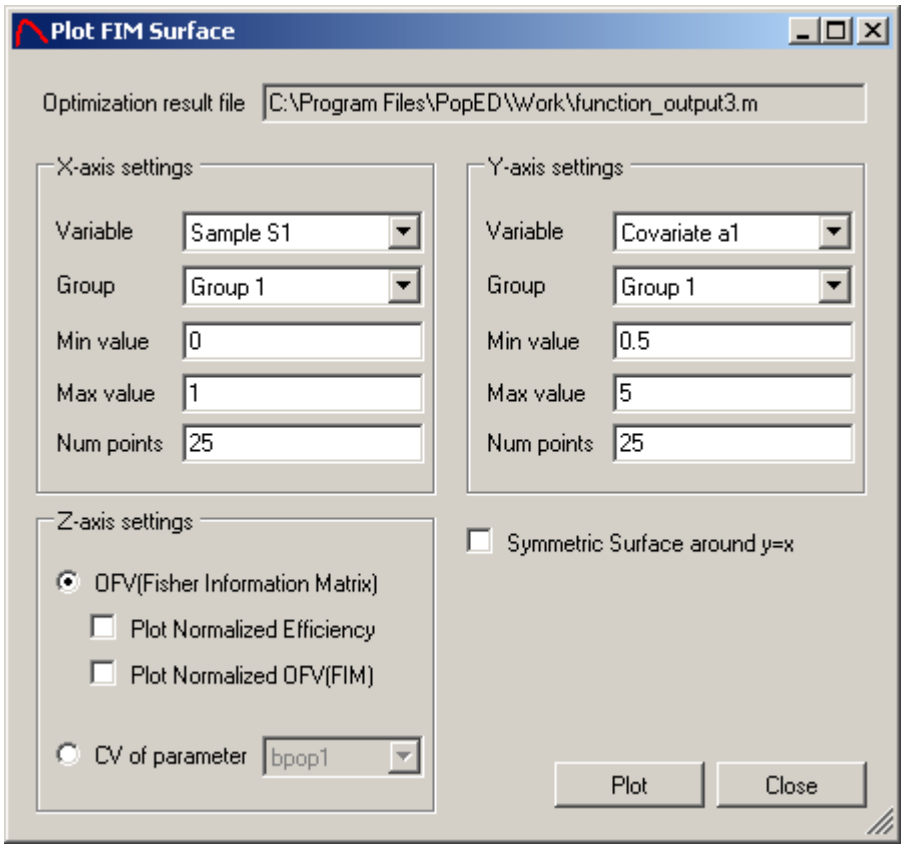
opt_value, *opt_time*, *opt_group* and *opt_model* are output vectors. *opt_value* contains the dependent variable value, opt_time the time point that the *opt_value* is calculated at, *opt_group* in which group the DV are calculated and *opt_model* has information about which sub-model the sample belong to.

Example:

```
plot_model_results([ 1 1],[ 1000 1000],[ 0 0],
[1 1],function_input(),1,0,[1 1 1; 1 1 1],function_output1(),0)
```

Plot Fisher Information Matrix Surface

The surface of the OFV can be plotted against two of the design variables. This plot is available in the GUI from the Optimal Design menu. To get access to this plot a valid output function file needs to be specified.



The x and y axis settings are similar and all the variables in the current design are available in the Variable combo box. Choose a group for that design variable and a min/max value of the axis. Num of points to split the axis in and to calculate the OFV should also be specified. The surface will calculate the OFV number of points in the x-axis multiplied by the number of points in the y-axis, i.e. 25x25 = 625 in the example above. The surfaces that can be plotted are the OFV or the normalized OFV, i.e. scaled to the power of one divided by the number of unfixed parameters in the model (FIM^(1/p)). The third alternative is to plot the normalized efficiency, i.e. the value of OFV divided by the optimal OFV scaled with the number of parameters in the FIM. One other option is to plot the SE (%) of an unfixed parameter in the model. The CV will be presented as a percentage value in the plot. If the surface is known to be symmetric, e.g. plotting sample1 and sample2 within the same group the *Symmetric Surface around y=x* can be checked, this will half the run-time. This option could only be used when the num points are the same for the two dimensions. If discrete variables are plotted the 3d plot will disregard the num of points and instead use the number of discrete values of the discrete variable that are within the min and max value. The grouping of the sample times, the covariates and the discrete variables will be considered in the plot.

This diagnostic must be considered very carefully even though it could be a powerful tool to visualize the surface of the FIM. In most cases the dimensionality of the FIM is greater than 3 and to look only of 2 of these dimensions and draw conclusions could be misleading. There is also possible to change the min and max values to higher or lower values than the bounds used in the optimization. In this case e.g. the efficiency calculations might be false because the optimal OFV might be in the wider region. Also be very careful to use variables that are not optimized for in the surface, e.g.

only optimization on sampling schedule but plotting a sample versus the dose. Again the optimal OFV might be wrong. This option is still available because sequential optimization might have been performed before the last optimization.

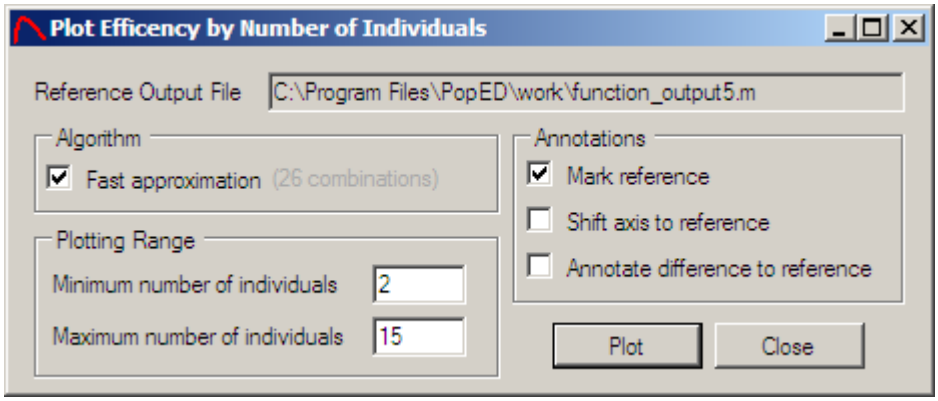
If the plot is not visible the render method might be wrong, this could be the case when the OFV is very large. Change the render from OpenGL to z-buffer in the Matlab figure properties. Another possibility is that the FIM calculation is nearly singular or singular. Try to change the max and min value to be close to the optimal design or use the script version of the surface plot to catch warning messages.

The script version of the FIM surface plot is available from the Matlab command window and the function is named:

```
plot3d_fim
```

The script function has a lot of options and will not be described here but there is a describing help text in the function. Type *plot3d_fim* in the Matlab command window and help text will be printed. For example there is an option to save the plot to a figure file without viewing it that could be useful for cluster/script users.

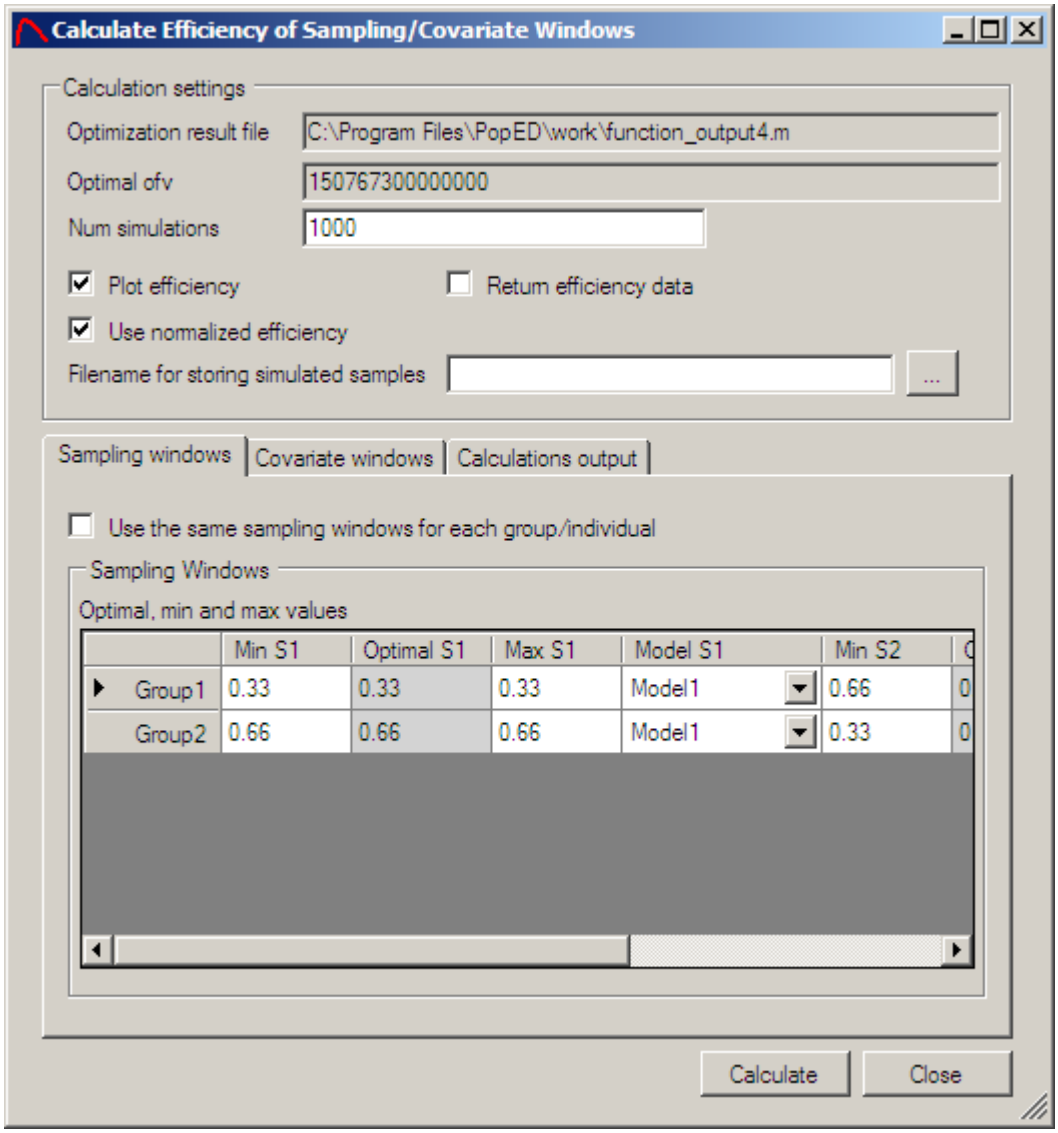
Relate efficiency to individuals



Reporting designs or the relationship between e.g. two designs is often done by reporting efficiency. In general, the efficiency between two designs might not be intuitive to interpret and furthermore the efficiency is related to the size of the model and also calculated differently for each criterion, e.g. D-optimal, A-optimal etc. Instead; a more interesting and intuitive way of comparing designs might be to relate the efficiency to the number of more/less individuals needed to get the same information as the reference design. This is in fact a true optimization problem because the information one gain when adding an extra individual depends on in which design group the individual will be added. This plot enables the user to see how adding or removing individuals from the study will relate to certain efficiencies. The best efficiency is shown by adding the individuals in the optimal groups and the worst efficiency is shown by placing the individuals in the group with least information.

Fast approximation is an option that allows the calculation of the optimal number of individuals be calculated faster. This option is likely to be exact in most cases but no mathematical proof of that has been provided. A good practice is to always use the *fast approximation* while exploring designs. The range of the number of individuals to show in the plot could be changed and the “maximum” number of individuals in the plot is the *total max groupsize* and the minimum number of individuals is the *total min groupsize*. If the *mark reference* option is set the reference design in the efficiency calculation is indicated by a dashed line. The *shift axis to reference* option allows the user to make the number of individuals axis related to the reference design in the efficiency calculation. Finally; the *annotate difference to reference* option mark the efficiency difference and number of individual difference to the reference design. This mark is only meaningful if the reference design has not been optimized for number of individuals but otherwise the option will show the maximum gain in information by optimizing on group assignment.

Calculate Design Windows



Often the optimal sampling times, optimal covariates values or optimal discrete variables values are not feasible in practice. Therefore borders around each optimal value could be calculated that will produce a feasible design instead of an optimal design, this is called sampling windows or covariate windows. This option is available in the GUI from the Optimal Design menu and it's also available as a script function.

Depending if covariate and/or discrete variables are in the design, tab pages for covariates windows and discrete variable windows will be enabled. Sampling windows is always available because samples need to be taken in an informative design. In the sampling windows the optimal sampling time is written and the window is defined by changing the min and max values. The optimal value and the model could not be changed. The sampling window couldn't be outside of the borders defined before the optimization. If discrete variables is used the value for each discrete variable could be changed from the optimal value to another available discrete value for that particular discrete variable. Grouping will be considered for the samples. The *Filename for storing simulated samples* enabled the user to store the samples for the sampling/covariate windows as well as the expected CVs and the efficiency for each simulation. The results will be stored in a *.csv file with the number of rows equal to the number of simulations. The columns of the file will start with the sample times for group 1 then the sample times for group2 etc. After the sample times the covariates for group1, group2 etc will be entered followed by the discrete variables for group1, group2 etc. The last columns will contain the CVs (in fractions) for each parameter that is not fixed (in the same order as the output file) and lastly the efficiency (in fractions) is outputted in the last column.

The tool will calculate the efficiency of a sample within a window by randomly (uniformly) take sample times from the window and calculate the OFV and compare the value to the optimal OFV. This could be done by using the normalized OFV (scaled to the number of parameters) or the un-scaled efficiency. The number of simulations indicates how many samples that will be drawn from the windows. If grouping is used for the sampling times, covariates or the discrete variables, the samples will be grouped when calculating the sampling windows. The first sample in the grouping will define the borders for the sampling window, e.g. if sample 1 and sample 3 is grouped, the sampling window for sample1 will be the window for sample 3 regardless of the values for sample 3 and they will of course have the same sampling time sampled from the sampling window. There is also a possibility to plot the efficiency in a Matlab figure window by checking the *Plot efficiency* check box.

The windows function will return the efficiency in a vector and it's visible in the calculation output tab after the calculations if the *Return efficiency data* is checked.

The script version of the sampling windows calculation function is:

```
[eff min_eff mean_eff max_eff] =
efficiency_in_windows(iNumSimulations,optdetfim,xt_windows,a_windows,x_wind
ows,bNormalizedEfficiency,bPlot,popedInput,strFileName)
```

iNumSimulations is a scalar that specifies how many samples that will be picked from the sampling windows.

optdetfim is a scalar with the optimal OFV from the optimization.

xt_windows is a matrix with the sampling windows for all groups. For each row in the matrix the min and max value of the window for each sample should be entered.

a_windows is a matrix with the covariate windows for all groups. For each row in the matrix the min and max value of the window for each covariate should be entered.

x_windows is a matrix with the discrete variable windows for all groups. For each row in the matrix the min and max value of the window for each discrete variable should be entered. From the GUI, the min and max value for each discrete variable will be the same.

bNormalizedEfficiency if the efficiency should be scaled to the number of unfixed parameters in the model or not. 1 if normalized, 0 otherwise.

bPlot if the efficiency should be plotted, true or false.

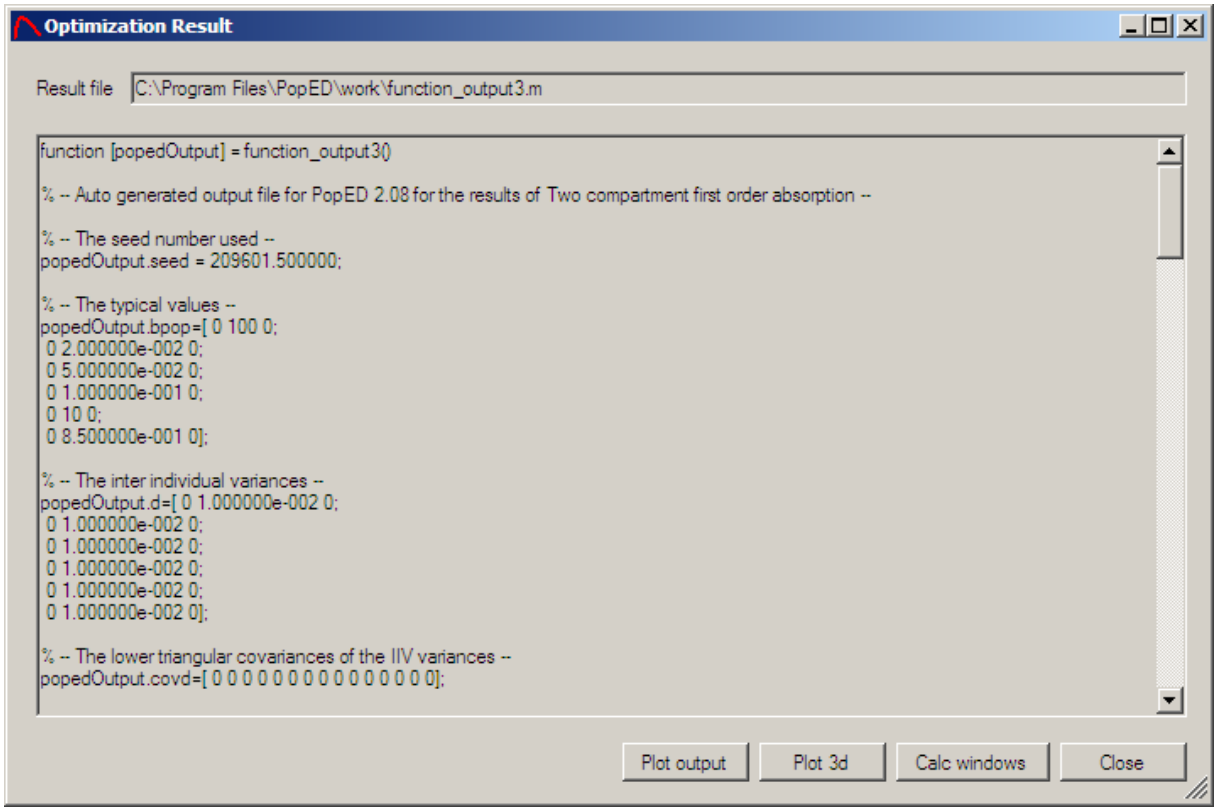
popedInput is the PopED settings function created from the xml file (see above). The function is often accessible by typing *function_input()*.

strFileName is the filename and path to the *.csv file where the results should be stored. If *strFileName* is an empty string ‘’, no output file will be created.

Example:

```
efficiency_in_windows(1000,1.154813e020,[0 1.2 0 0.5 1 2],[50
150],[],0,1,function_input(),’’)
```

View Optimization Output



The optimization result is a Matlab function file that contains a lot of information about the optimization result. The variables in the output function file are.

- seed – The seed number used in the optimization.
- bpop – The typical values, the distributions and variability.
- d – The IIV distribution, the distributions and variability.
- covd – The lower triangular (matrix) vector of covariance for the IIV variances.
- docc – The inter occasion variances, the distributions and variability.
- sigma – The Residual variability.
- NumOcc – The number of occasions.
- xt – The final sampling times.
- x – The final discrete values of the discrete variables.
- a – The final covariates.
- gni – Number of samples per subject.

- groupsize – Group size for each group.
- modelswitch – The vector defining which samples that belong to which sub-model.
- iFIMCalculationType – The full or reduced FIM.
- ApproximationMethod – FO, FOCE or FOCEI.
- FOCENumInd – The number of individual samples if a conditional approximation method was used.
- fmf – The Fisher Information Matrix.
- imf – The inverse of Fisher Information Matrix.
- ofvmf – The objective function value.
- d_switch – If optimized using D or ED-optimal design.
- ofv_calc_type – The criterion used to calculate the OFV. E.g. det(FIM) = 1 and ln(det(FIM)) = 4.
- param_var – A vector with the unfixed parameter variances, the vector start with the unfixed bpop, then the unfixed d, the unfixed docc and last is the unfixed sigma.
- param_cv – A vector with the unfixed parameter CV in fractions, the vector start with the unfixed bpop, then the unfixed d, the unfixed docc and last is the unfixed sigma.
- notfixed_bpop – A vector with a 1 if a bpop is not fixed and a 0 if the bpop is fixed.
- notfixed_d – A vector with a 1 if a d is not fixed and a 0 if the d is fixed.
- notfixed_sigma– A vector with a 1 if a sigma is not fixed and a 0 if the sigma is fixed.
- notfixed_docc – A vector with a 1 if a docc is not fixed and a 0 if the docc is fixed.

- filename – A string containing the filename and path of the output file, used by the GUI if the COM server is not used..

The function output could also be called within a script by typing function_outputi() where i is the function output number.

Update initial design from output

The initial design (sample times, covariates, discrete variables, number of samples per group and the number of individuals per group) could be updated after an optimization. This is done by clicking *Update initial design from output* under the Optimal Design menu. The output file must be consistent with the xml file (i.e. similar number of groups, sample points etc.)

Convert output structure to xml

The output structure, e.g. *function_output.m,* could be converted to an xml file. This feature is used by the GUI but might also be used by third party software that wants to communicate with PopED. To convert an output file; use the following syntax:

```
convert_output_to_xml(function_output10(),strXmlFileName)
```

where *strXMLFileName* is set to the name that the xml file will have. This could also be an empty string and then the file name of the xml file will automatically be the filename and path of the function output but with the extension .xml instead of .m.

Model Templates

A number of model templates are distributed with the PopED program to illustrate how different models can be implemented. The model templates can be accessed by the Wizard; new option, under the File menu in the GUI. The model templates directory could be extended by own user defined templates by saving a PopED GUI settings file (*.xml) in the template directory. A recommendation is to write user defined templates as general as possible and not to add specific design of models that are already in the template directory. The templates available in the PopED version 2.10 are:

- One compartment IV Bolus dose. Parameterized by CL and V.
- One compartment zero order infusion. Parameterized by ke and V.
- One compartment 1st order absorption. Parameterized by ka, ke, F and CL.
- One compartment IV Bolus dose with Michaelis Menten elimination. Parameterized by Vmax, Km and V. Explicit solution.
- One compartment IV Bolus dose with Michaelis Menten elimination. Parameterized by Vmax, Km and V. Differential equation.

- One compartment with transit compartment absorption. Parameterized by ka, CL, V, MTT and n.
- One compartment IV bolus with a direct Emax effect. Parameterized by CL, V, E0, Emax and EC50.
- One compartment 1st order absorption with repeated dosing. Parameterized by ka, ke, V and tau.
- One compartment 1st order absorption with repeated dosing implemented as differential equation. Parameterized by ka, ke, V and tau.
- Two compartment IV Bolus dose. Parameterized by V, k12, k21 and ke.
- Two compartment zero order infusion. Parameterized by V, k12, k21 and ke.
- Two compartment 1st order absorption. Parameterized by ka, k12, k21, V, F and ke.
- Two compartment 1st order absorption with a lag time. Parameterized by ka, CL, Vc, Vp, Q, F and tlag.
- Two compartment 1st order absorption, multiple dosing. Parameterized by ka, k12, k21, V and F.
- Three compartment IV Bolus Dose, Parameterized by V, k12, k21, k13, k31 and ke.
- Three compartment zero order infusion. Parameterized by V, k12, k21, k13, k31 and ke.
- Three compartment 1st order absorption. Parameterized by ka, k12, k21, k13, k31. V, F and ke.
- Biological rhythm (3 cos rhythms). Each rhythm parameterized by Peak time, Amplitude and the period.
- Linear disease progression model with symptomatic effect. Possibility to optimize on start and stop time of study.
- Linear disease progression model with protective effect. Possibility to optimize on start and stop time of study.
- Linear disease progression model with symptomatic and protective effect. Possibility to optimize on start and stop time of study.

Most of the model templates are analytical solutions of the different compartment models; therefore it could be good to use the templates because the analytic

solutions are much faster than the differential equations. Fill free to distribute and share user written templates with other users of PopED.

Examples

The examples are available in the <*PopED Install dir*>*PopED Examples* and will cover more of the script specific options of the PopED program. The examples can be run from the GUI (except example 2) and most of the design options can be changed and viewed by the GUI version of PopED. The examples are not always optimized for speed, instead they are written to be easy to understand and to follow.

To run an example, open the PopED GUI settings file (*.xml) for the example and change the path to the model file. The model file for each example will be in the same directory as the PopED GUI settings file. For example 2, after changing the model file path, open the Matlab version of PopED (see Running PopED) and type *run* in the Matlab command prompt.

Example 1, Multiple Drug optimization and grouping of sample times

In this example optimization will be done in three drugs concurrently. PopED will find an optimal design for parameter estimation of the three drugs at the same time. The drug models are written with analytic solutions but an identical differential equation example is also available by setting the sixth covariate to a value (0=ode, 1=analytic, 2=linear ode solver). The linear solver uses both the inhomogeneous solver as well as the homogeneous solver. Look in the model file for more information. All drugs have different profiles; one is a one-compartment drug with a zero order infusion, the second drug is a one-compartment drug with an IV bolus dose and finally the third drug is a one-compartment drug with 1st order absorption. The design contains four different groups with 10 individuals in each group, all with four sample times. The sampling schedule is coded as 12 different sample times per group but the sample times are grouped in 3, i.e. to cover the measurements of the three drugs at the same time. There is also grouping of the sample times between groups, i.e. the first sample of the first and third group are grouped together meaning that they will be taken at the same time. There is also a grouping of the last sample time between the second and the 4th group. Use the model validation tools from the PopED GUI (plot model and simulate) to look at the different drug profiles.

The grouping forces some of the sample times to be the same within and between groups. The grouping assumes that the residual variability of simultaneous samples of the three drugs is independent. This is probably not a correct assumption but the residual variability correlation in the samples only depends on the difference in the analysis method of the samples.

Before running the experiment, transform the PopED GUI settings xml file to a Matlab function (see Convert PopED Settings to script version).

Example 2, User defined distributions on model parameters

In this example a user defined distribution (UDD) is assumed on the typical value parameter for clearance. For a parameter like CL a normal distribution for the typical value or a normal/transformed IIV distribution probably would be better but this example is only an illustration how a user defined distribution can be implemented. The user defined distribution can be discrete or stochastic and in this example a discrete distribution of 25 values for CL is used. In ED-optimal design the expectation value can be approximated by a Laplace integral approximation but this is not used for the user defined distributions. For a user defined distribution that is not stochastic, the ED-sample size variable should be set to the length of the distribution and the stop criteria for ED-optimal design should be small (see ED-optimal design settings). A user defined distribution function should have this syntax:

```
function ret = user_distribution(ret,t,sample_number,globalStructure)
%A user defined distribution function.
%The value from the user defined distribution (UDD) should be returned in
%ret.
%Input:
```

```
%t is a vector that contains all the distributions of the
%parameters in the model. t(i) == 3 is a user defined distribution,
% 0 is point distribution, 1 a normal and 2 a uniform distribution. Ret is
%also an input variable with the values for the other distributions, if
%t(i)~=3 the value ret(i) should not be change, but the user distribution
%value can depend on the other parameters current values. The sample number
%is the ED-sample number; this shouldn't be larger than the distribution
length for a discrete distribution.
typCL = [1 2 4 3 3.2];
IIVCL = [0.01 0.02 0.012 0.011 0.022];
j=1;
for i=1:length(t)
    if (t(i)==3) %This is an user defined distribution parameter
        if (j==1) %If it's the first parameter with a UDD
            ret(i) = typCL(mod(sample_number-1,length(typCL))+1);
        else
            ret(i) = IIVCL(mod(sample_number-1,length(IIVCL))+1);
        end
        j=j+1;
    end
end

end
```

In the example the distribution is defined by a definition outside the distribution function, i.e. in the run.m file, and then passed to the distribution function via the *globalStructure* (*user_data*). This is to illustrate that if a large discrete distribution is used it might be computationally run-time efficient to declare the distribution outside the distribution function. Look in the run.m file and the user_distribution.m file to see how the UDD is defined. The model is a one-compartment model with an IV bolus dose. The path to the UDD might need to be changed in the run.m file.

Before running the experiment, transform the PopED GUI settings xml file to a Matlab function (see Convert PopED Settings to script version).

Example 3, Analytic Fisher Information Matrix of PK-PD model

In this example, analytic derivatives are used/calculated instead of numerical approximations of the FIM. There are two sub-models; a one-compartment IV bolus dose and a direct e-max effect model. Run the example from the GUI. If the run produced an error, it might be that the analytical derivative file has not been flushed to the operating file system before the execution of the FIM. Retry the run or run it from the Matlab command line in the example directory by typing:

```
poped2_0('analytic_derivative.xml');
```

Example 4, Discrete list of sample times

In this example, there is a possibility to optimize the best set of a finite discrete number of samples. The example can be run within the GUI and uses the discrete variables (*x*) as the sample times. This is assigned in the model, settings the sampling time vector (*xt*) to the discrete variables. The number of samples must still be set and entered in order to produce the right dimensions in the Fisher Information Matrix. But the actual sample time (border, initial) doesn't matter; it is the discrete variable border/initial that is used. It is also possible to use the model in simulations and typical value plot because of the if-statement in the model. The model is a one-compartment 1st order absorption model.

Abbreviations and Variables

FIM – Fisher Information Matrix ~ Inverse of the expected Covariance-Variance matrix of the unknown parameter.

Criterion Function – The function that is optimized in the optimization procedure.

OFV – Objective Function Value – See criterion function.

RS – Random Search – A random search algorithm.

SG – Stochastic Gradient – A stochastic gradient search algorithm.

LS – Line Search – A line search algorithm.

GUI – Graphical User Interface – A graphical tool that communicates with the user.

bpop – Population mean variable.

b – Inter Individual variable.

d - Inter individual distribution variance.

docc – Occasion variability variance.

a – Covariate variable.

x – Discrete variable.

g – Parameter definition vector.

xt – Sample time vector.

epsi – Residual variable for all sample and individuals.

sigma – Residual distribution variance.

IIV – Inter individual Variability – The variability between individuals.

BSV – Between Subject Variability See IIV.

NONMEM – NON linear Mixed Effects Modeling.

UDD – User Defined Distributions.