

# Best Practice for Modelling and Simulation: EFSPI Special Interest Group proposal

Michael O'Kelly, member SIG for Modelling and Simulation.



# Acknowledgements

- SIG members and PSI members who reviewed draft Best Practice proposal.
- SIG Chair Chris Campbell and Professor Chris Jennison provided the 2015 “Hackathon” that tested the Best Practice proposal
  - › and the participants in the “Hackathon”, who provided fruitful input on the SIG Best Practice proposal, resulting in many improvements to the Best Practice document.
- Co-authors of Best Practice document Vlad Anisimov, Chris Campbell, Sinéad Hamilton.

O’Kelly M, Anisimov V, Campbell C, Hamilton S. Proposed Best Practice for Projects that Involve Modelling and Simulation. *Pharmaceutical Statistics* (accepted).

# Best Practice, background

- November, 2011: European Federation of Pharmaceutical Industry Associations (EFPIA) and European Medicines Agency (EMA) organized workshop on Modelling and Simulation (M&S).
  - › FDA attendee.
- Rob Hemmings of EMA called for a Best Practice document.
- EMA stressed that levels of pre-specification and justification of assumptions etc. for modelling and simulation would and should vary depending on “importance” of the project and its outcome in the process of approval.

# Best Practice, background

- November, 2011: European Federation of Pharmaceutical Industry Associations (EFPIA) and European Medicines Agency (EMA) organized workshop on Modelling and Simulation (M&S).
  - › FDA attendee.
- Rob Hemmings of EMA called for a Best Practice document.
- EMA stressed that levels of pre-specification and justification of assumptions etc. for modelling and simulation would and should vary depending on “importance” of the project and its outcome in the process of approval.

# EMA: “Best practice” depends on importance of project



EMA-EFPIA Modelling  
and Simulation Workshop

Good practices and next steps

Robert Hemmings, EMA

## M&S good practices

- Different standards for different exercises (L,M,H)
- Standard should be high!
  - Assumptions (not only mathematical)
  - Model building rationale
  - Model testing
  - Inference
  - Sensitivity analyses / Challenge assumptions
  - Reporting
- Detail of regulatory response might be vary according to impact

# EMA: “Best practice” depends on importance of project



EMA-EFPIA Modelling  
and Simulation Workshop

Good practices and next steps

Robert Hemmings, EMA

## M&S good practices

- Different standards for different exercises (L,M,H)
- Standard should be high!
  - Assumptions (not only mathematical)
  - Model building rationale
  - Model testing
  - Inference
  - Sensitivity analyses / Challenge assumptions
  - Reporting
- Detail of regulatory response might be vary according to impact

# PSI early discussions with EMA on Best Practice

- EMA suggestion: should cover
  1. Pre-specification
  2. Analysis
  3. Check of assumptions
  4. Presentation of results
  5. Sensitivity analysis
- Conclusions of simulations should be robust to missing data

# EFSPI proposed Best Practice document

- EFSPI SIG Best Practice document
  - › Authored by volunteers from the SIG
    - » SIG members from variety of pharmaceutical companies.
    - » All authors of the Best Practice document were from contract research organizations.
    - » Agreed to be adopted by Board of PSI.

# EFSPI Modelling and simulation SIG Best Practice document

- Defines modelling and simulation.
  - › Tricky – e.g., what’s the difference between prediction and simulation?
- Declares its scope to be **all** uses of modelling and simulation.
  - › Controversial: can same Best Practice cover use of modelling and simulation in pharmacometrics and in simulating the running of a clinical trial?
- When should we use modelling and simulation (vs., say, employing a closed-form solution)?
- Heart of Best Practice: the specification of the project.
  - › EFSPI proposal names key elements of a specification.
  - › Important: allows flexibility in what elements may be included and in level of detail provided.
  - › Best Practice requires the plan to justify omission of elements, justify level of detail.
- Quality control.
- Changes to the modelling and simulation plan.

# Introduction: role of the specification in Best Practice

- Specification enables
  - › users to answer in advance the question “will this modelling and simulation project answer my research question”?
  - › team members to act consistently to achieve the planned outputs;
  - › sponsor to assess whether the project achieved its goals.
- SIG document allows the flexibility necessary for Best Practice in this area where the regulatory and scientific importance of the projects varies widely.

# Example best-practice specification for low-impact work

## Planning and Reporting for Projects that Involve Modelling and Simulation Best Practice Document

### Appendix B: example specification with a low level of detail

#### Using simulated data to verify an estimate of probability of success

##### Specification of simulations

###### B.1 Introduction

Given five treatment development programs with known probability of success, it is desired to know the probability of zero successes and of four and five successes. These probabilities have been calculated analytically. It is requested that a simulation be run to verify that the analysis is correct.

Since this is a one-off query on whose evidence alone no decision will be made, this is judged a project of low importance. Therefore the clinical background is not described; nor are metrics and criteria for decisions appropriate.

###### B.2 Simulation and analysis/design

As noted, this project is of low importance and no decision will be made by it alone. Therefore the description of the elements of the simulation and analysis will not be detailed and some elements are not applicable.

###### B.2.1 Scenarios assumed and assumptions made

Probabilities of 0.1, 0.2, 0.2, 0.05 and 0.4 were given for programs 1-5, respectively. Since the objective was simple verification of an existing calculation, no justification is given here of these probabilities. Since the question answered is theoretical, just one given scenario is used.

###### B.2.1.1 Sensitivity analyses

This project is not required to assess assumptions, so sensitivity to assumptions is not planned to be analysed via sensitivity analyses

###### B.2.2 Data sets generated

Temporary sets of binary outcomes will be generated. Data will not be bootstrapped because a simple verification is sufficient. Three million binary outcomes are simulated for each program.

Page 1

###### B.2.3 Statistical analysis

The number of instances of zero, four and five successes was calculated for each of the 3 million simulations, and the probability of zero, four and five successes in a simulated instance was calculated and plotted.

###### B.2.4 Operating characteristics

Given that this modelling and simulation task is to be a sanity check, the number of simulations required to achieve a given accuracy with 5% confidence will be approximated. The probability of five successes is small ( $<1/100$ ) so a precision of 0.001 is desired. Using the formula of Burton *et al.* (2006), with  $\alpha=0.05$ , and approximating the variance of the probabilities as  $5 * p(1-p)$  where  $p=0.2$ , 3 million simulations will provide precision of approximately 0.001.

###### B.2.5 Logistics

The R language package `myBinaryEP` will be used to simulate the binary outcomes. The package allows for correlations between the outcomes, but this was not required for the primary objective. R version 3.0.1 will be used. See Appendix for the R code used. The seed used was 1.

###### B.3 Quality control

Given that this modelling and simulation task is of low importance and will not of itself lead to a decision, the specification will be submitted to the requestor of the calculation, but no further QC of the production of results is planned. The output will be checked against the requestor's calculations.

###### B.4 Presentation of results

A table will be presented of the probability of zero, four and five successes among the five programs, calculated as the proportion of instances of zero, four and five successes in 3 million simulations of the five programs. The number of instances will also be plotted in a histogram with one stack for each level of successes, a stack for zero successes, 1 success, ... and so on up to five successes. These outputs are judged sufficient to act as a check of the analytic estimates, which is the objective of this project.

The precision of the result (standard error) will be presented in a footnote to the plot. Given the inclusion of precision, no confidence intervals will be presented. Given the theoretical nature of the problem and the corresponding simplicity of the simulation, no bias is to be expected in the simulation-based estimates.

The results of the modelling and simulation will not be stored. The R code will be stored in [location]. A note of the contents of the table output will be included as a comment in the R code.

Page 2

# Example best-practice specification, high-impact work

## Planning and Reporting for Projects that Involve Modelling and Simulation Best Practice Document

Appendix A: example specification with a medium-high level of detail

Using simulated data to assess analyses of negative binomial outcomes with missing data

Specification of simulations

### A.1 Summary

The objective of the project is to produce a user-friendly means that has been tested using a combination of real and simulated recurrent event data. The means **designing** will use a Multiple Imputation (MI) scenario analysis method to impute a variety of appropriate, including reference-based assumptions, for missing data, in order to assess accurate surviving recurrent event outcomes that are missing at random (MAR), with an approach that accounts the negative binomial distribution. The means estimate the treatment effect of a number of different treatments, compared against a control or reference group, using difference of mean square means.

### A.2 Introduction of the specification

This document has been modified on the best practice document of the Special Interest Group for Modelling and Simulation of RCTs. The objective of the simulation exercise is to assess, with respect to the MI scenario analysis:

- The true Type I error under the null hypothesis of no treatment difference
- Power to detect a treatment difference when one occurs

These objectives will be addressed by simulating scenarios where (1) the treatment effects of two different treatment types are the same, (2) and the treatment effects of two different treatment types are different. The simulation will also determine the impact of percentage dropout on both the power and the Type I error rate.

Page 1

### A.3 Simulation and analysis design

The objective of the assessment of type I error and power is to help to validate the means as an implementation of an analysis of recurrent event data under the missing-at-random (MAR) assumption, where the distribution of the outcome is negative binomial. To validate the implementation, results from the MI approach will be compared with those of a direct likelihood one, using the standard approach. The standard approach for negative binomial outcomes is a direct likelihood analysis, using a generalized linear model, with the varying rates of response accounted for by an offset term that is equal to the log of the time exposed.

If the MI approach correctly implements the analysis for scenarios that are distributed as negative binomial under MAR, results from the standard approach and the MI approach should agree. Therefore, results of the results from MI analysis of the simulated datasets agree with the standard analysis, the MI approach can be regarded as valid and appropriate. This multiple-imputation scenario based approach, to a significant extent, is to determine as to whether to set the implementation to negative binomial data. Therefore, the project is judged to be of medium importance. There will be a moderate amount of work given in the following sections on Simulation and Analysis to ensure that no simulation or analysis is overlooked.

#### A.3.1 Scenario assessed and assumptions made

As mentioned in Section A.2, there are two main objectives for the simulation exercise; the datasets will be simulated in two different trials in order to fulfil these objectives. Scenarios will address the Type I error rate of the method, as implemented by the means, and scenario two will address the power of the method, as implemented by the means. To do this, recurrent event datasets will be simulated in two treatment groups having the same relationship with the rate of recurrent events for scenario one and differing with respect to that association for scenario two. In addition to treatment group, other covariates will be included, such as binary covariate, with two categories, and no continuous covariate. An underlying random-effects effect will be included, using random sampling from a gamma distribution. Datasets are simulated using the method in order to estimate the negative binomial, as described by Kuznetsov et al. [2].

Missing data is generated by simulating a rate of dropout that tends to increase as the event or continuous event occurs. Both recurrent events and dropouts are simulated, independently from each other, for each time point, using a Bernoulli distribution. The probability of an event is modelled as a linear combination (1) of each subject's baseline covariates, which has been  $\log$ -transformed and multiplied by the subject-specific effect, (2) the subject's previous values being converted to a probability value using the appropriate back-transformation of  $\mu = \frac{\exp(\eta)}{1 + \exp(\eta)}$ , which is described in Kuznetsov et al. [2]. The probability of dropout is a linear combination of each subject's baseline covariates (1) with different values than Bernoulli distribution, as compared to those that in Kuznetsov et al. [2], which has also been converted to a probability value using the log back-transformation ( $\mu = \frac{\exp(\eta)}{1 + \exp(\eta)}$ ).

We are currently addressing the mean's ability to account for multiple different types of dropout (e.g. death, adverse event, subject dropout); therefore, there will only be two event types

Page 2

in the response variable, recurrent event (1) and dropout (2). Calculations based on all simulated datasets, using the likelihood method, will be performed using the fitting An R Package (SAR) (SAR); however, without any data requirements of input values.

The regression coefficients for the models to be used in all datasets when generating the recurrent events will be based on real data from the bladder tumour recurrence dataset, taken from the R package 'survival' and described in [1] and [3]. The true values for the simulated continuous variables are set to equal to the true values for the continuous variable number and/or fixed on the bladder dataset. When the treatment effects are set to be equal, (the null case), the same true values are used for both of the treatment groups. When the treatment effect is not to be different (the alternative case), event analysis will be performed in order to test the mean's sensitivity to the degree of difference in true values. The difference in true values for treatment effect will be set to 0.02, 0.05 and 0.125, generating **coloured** treatment differences of 0.1, 0.47 and 0.7.

#### A.3.1.1 Sensitivity Analysis

A sensitivity analysis will be performed on both scenarios, to assess the sensitivity of the means to change in the percentage **degree** of missing data. This will be performed by altering the true values that generate the probability of dropout to produce simulated datasets where an approximate percentage of subjects have dropped out of the study early. Seven rates of dropout will be compared, 1%, 5%, 10%, 15%, 20%, 25% and 40%. The number of simulations during a significant difference in the treatment effects will be counted, for both the null and alternative scenarios, to determine the effect of percentage dropout on both the type I error rate and the power of the means.

As noted, we will also test the sensitivity of the means to changes in the percentage of dropout in the data when assessing the power and the type I error rate. The true values used to determine the rate of dropout will be manually calibrated to generate seven sets of estimates, containing seven different percentage of subject dropout: 1%, 5%, 10%, 15%, 20%, 25% and 40% dropouts.

#### A.3.1.2 Dataset generated

The simulated datasets will be modelled in a normal form, which allows for multiple records per subject, and whose each observation corresponds to a subject over

one set of seven simulated datasets will be generated for the three scenarios to test the type I error rate. When testing the power, three sets of simulations will be generated. Seven estimates will be generated for each such to be tested using the maximum likelihood method in Section A.1. These datasets will include a simulation variable variable to identify each scenario, a subject ID variable to identify each subject, an event variable to determine

whether the observation is a recurrent event or a dropout, a variable for each baseline covariate (two continuous and two categorical, including treatment) and a time variable to identify the time point at which each event or dropout occurred. Data will not be benchmarked because we are using specific scenarios and we want to control these scenarios artificially; however, as noted, the true coefficients have been varied according to values fixed to real data.

#### A.3.1.3 Statistical analysis

For scenario one, the proportion of times the null case gives a false positive result is calculated. The 10% result is then compared to the same result using the standard Direct Likelihood approach in order to get a comparison in performance between the two methods. For scenario two, the alternative hypothesis, the two analyses are requested on the datasets where the treatment effects are different and the power is calculated and compared.

The sensitivity analysis will test the degree to which the type I error and power of the means are dependent on percentage dropout. The proportion of false positive and false negative results will be calculated. Again, these 10% results are compared in performance against the Direct Likelihood analysis results.

#### A.3.1.4 Coverage characteristics

The number of simulations required to give an accurate measure (and/or to achieve a comparison with sufficient accuracy, with a specified confidence level and clearance to the true or desired value, can be calculated using the formula from Burton et al. [5]. This calculation estimates the required number of simulations based on the accuracy of the number of success. The number of simulations required (N) is calculated as:

$$N = \left( \frac{z_{1-\alpha/2}}{d} \right)^2$$

The parameter  $z_{1-\alpha/2}$  is the specified level of accuracy, or the permitted difference from the true or desired value, or represents the standard deviation,  $d$  is the specified percent of the standard normal distribution and  $\alpha$  is the significance level required.

(Details of the calculation have been omitted for this example)

#### A.3.1.5 Software

SAS v9 has been the primary software adopted.

The negative binomial was approximated by the Bernoulli distribution with a gamma random effect. Events and dropouts are simulated by sampling from Bernoulli distributions, where the probability of the event or dropout is, in theory, by employing the baseline covariates by their corresponding true values. These values are simulated and approximated, since it is assumed that the log of the number of events has a linear relationship with the predictor. Thus the appropriate transformation is used to calculate the probability of an event (1) or dropout (2) given the appropriate. The probability of dropout is also defined based on the number of

continuous events seen. Kuznetsov et al. (2014) describe how the negative binomial can be generated by using the Poisson distribution mixed with gamma. See Supplement for further details of generating the Poisson process using a Weibull series.

The need to use fixed sampling from each distribution (using the rand function) is 22 to begin with but this is recommended by 1 at the start of each simulation. Using a different seed for the emulators of each dataset allows them to be completely independent of each other.

#### A.4 Quality control

The specification was independently reviewed. The core functions of the code will be unit tested to verify the computerisation. The code is planned to be submitted to the CEA equal working group with input in meetings on 14th. Of the project is required as of high importance, the delivery timeline should be added. The project means will be quality controlled via independent programming.

#### A.5 Presentation of results

Tables will be generated for both scenarios. For all simulation analyses, in each of the two tables the following columns will be included: the percentage dropout, the percentage of significant difference using the MI method, the percentage coverage using the MI method, the percentage of significant results using the standard direct likelihood method, the percentage coverage using the direct likelihood method and the variance ratio. The variance ratio will be the ratio of the observed variance to Rubin's variance of variance. The output from each means analysis, which provides the LOG-RATE and difference of LOG-RATE, will be noted for all sets of simulations.

#### A.6 References

1. O'Keefe M, Anonim V, Campbell C, Sheehan S. Present Best Practice for Projects that Involve Modelling and Simulation. *Pharmaceutical Statistics* (submitted).
2. Kuznetsov A, Roper J, Mandy B, Kuznetsov M (2014) Missing Data Scenario Analysis for Recurrent Event Data using Generalized Imputation. *Pharmaceutical Statistics* 13, 4, 289-304.
3. Anderson DR, Hartzberg ADL (1985) *A Collection of Problems from Wiley's Probability and Statistics*. John Wiley and Sons: New York.
4. Wu Y, Lin DY, Shen D (1995) Empirical Analysis of Recurrent Intergroup Relapse Time Data by Modeling Marginal Densities. *Journal of the American Statistical Association*, 94.
5. Warner A, Arora DC, Rayson P, Bolder RL (2010) *Using designs of simulation studies in medical research*. *Statistics in Medicine* 29(24): 2517-2525.
6. Lerner H (1982) *Stochastic processes in probability theory and statistical inference*. John Wiley & Sons: Chichester, 1982.

Page 1

model is a special case of generalized linear mixed models with a Poisson distribution and log link. In this model, observations  $y_i$  are made on the  $i$ th subject in scenario 1

$$y_i | x_i \sim \text{Poisson}(\mu_i)$$

$$\ln(\mu_i) = \gamma_0 + \beta + \eta_i$$

where  $\eta_i$  is a subject specific effect with same distribution on the real line. The design matrix  $X_i$  accounts for both the treatment type and the baseline covariates. Set  $\eta_i = \exp(\gamma_i) \mu_i$ ,  $\eta_i = \exp(\gamma_i) \mu_i$ , then

$$y_i | x_i \sim \text{Poisson}(\mu_i \eta_i)$$

The negative binomial model is a special case of this generalized linear mixed model with a Poisson distribution and log link, where  $\eta_i$  has a gamma distribution.

In the closed simulation,  $\eta_i$  is generated from a gamma distribution and multiplied by the prior mean,  $\mu_i$ , which is the exponential mean prior distributed in Section A.1.

### A.7 Supplement

#### A.7.1 Application of the use of Bernoulli and Gamma distributions when generating Poisson event data and the Negative Binomial distribution

Assume that the recurrent event data in time interval  $T$  are generated by a Poisson process with parameter  $\lambda$ . Then we describe how this data can be approximated by using discrete event simulation with Bernoulli trials. Consider a sufficiently small time interval  $\Delta t$  with a probability close to unity, then only comes either 0 or 1 event. Instead for each  $\Delta t$ , the probability of one event occurring is termed  $\lambda \Delta t$  is proportional to the length of the interval  $\Delta t$ . Using the properties of a Poisson process we get that the probability to have more than one event within  $\Delta t$ . Thus, the probability of one event is  $\lambda \Delta t$  and  $1 - \lambda \Delta t$ . Thus the events in the non-overlapping intervals of this length,  $\Delta t$ , can be considered to be independent Bernoulli trials.

Now we can address our study time,  $T$ , into  $n = T/\Delta t$  independent intervals of equal length. Thus independent Bernoulli trials each have a probability of an event  $p = \lambda \Delta t$ . Calculating Bernoulli trials with probability  $p$  we get that the number of events in the study of length  $n$  is binomial with parameters  $(n, p)$ , where

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \binom{n}{k} p^k (1-p)^{n-k}$$

It is well-known [6, 7] that for a large  $n$  and small probability  $p$  such that  $np \rightarrow \lambda$ , the binomial distribution is approximated by a Poisson distribution with parameter  $\lambda$ .

Thus,

$$P_n(k) \approx \frac{\lambda^k e^{-\lambda}}{k!}$$

So, which is 0.1, 1. This proves that when events generated that are consistent with the above assumptions, the number of events,  $X$ , in an interval of fixed length,  $T$ , has a probability function of

$$P_n(k) \approx \frac{\lambda^k e^{-\lambda}}{k!}, \lambda = 0.3, 2, \dots$$

where  $X$  is the Poisson random variable with parameter  $\lambda$ . This shows how the Bernoulli trials can be used to generate approximately a Poisson process in simulation time.

For recurrent events the negative binomial distribution can be derived as a mixture of Poisson distributions where the mixing distribution is exponentially distributed as the mixture of Poisson process. This mixing distribution is a gamma distribution with mean  $\lambda$ . The negative binomial distribution is also called a gamma-Poisson distribution for the same reason. The gamma-Poisson

Page 4

Page 17

# Best Practice for modelling and simulation, the work of the two groups, MID3 and EFSPI

- Agreement all aspects of the process
- Some differences in emphasis
- The two groups are working together to promote good practice
- The two groups are participating at session on Best Practice at 2016 annual American Statistical Association Biopharmaceutical Workshop with FDA speakers.

# Informal survey of Best Practice: background.

- MID3 group
  - › MID3 paper on Good Practice, supplementary material
  - › Lists 103 publications describing projects involving modelling and simulation.
- These slides make use the MID3 supplementary material.
- These slides do not represent the views of the MID3 group or the authors of the MID3 paper.

Marshall S, Burghaus R, Cosson V, Cheung S, Chenel M, DellaPasqua O, Frey N, Hamren B, Harnisch L, Ivanow F, Kerbusch T, Lippert J, Milligan P, Rohou S, Staab A, Steimer J, Tornøe C, Visser S. (2016) Good Practices in Model-Informed Drug Discovery and Development (MID3): practice, application and documentation. *CPT: Pharmacometrics & Systems Pharmacology*; (2016) 5, 93–122, available at <http://onlinelibrary.wiley.com/doi/10.1002/psp4.12049/epdf>, accessed 22Apr2016

# Survey

- Random sample of 48 of the 103 publications listed by the MID3 paper.
  - › 12, 17 and 6 of low, medium and high regulatory importance.
    - » Regulatory importance assessed by the authors.
  - › Publications freely downloadable for 36.
  - › Of these 36, 29 included an element of simulation.
  - › Survey includes six slidesets, 23 journal papers.
- Limitations
  - › Many journal papers are not written with regulatory requirements in mind.
  - › Survey includes six slidesets - less space to describe elements of best practice.
    - » Included only very slightly fewer of the elements of best practice surveyed.
  - › Assessments in this survey not independently reviewed.
- Average “Best Practice” items present, out of a possible 21 items
  - › 14.3, 12.3 & 8.3 for papers of low, medium and high regulatory importance, respectively.

# Attributes assessed in survey – planning & presentation

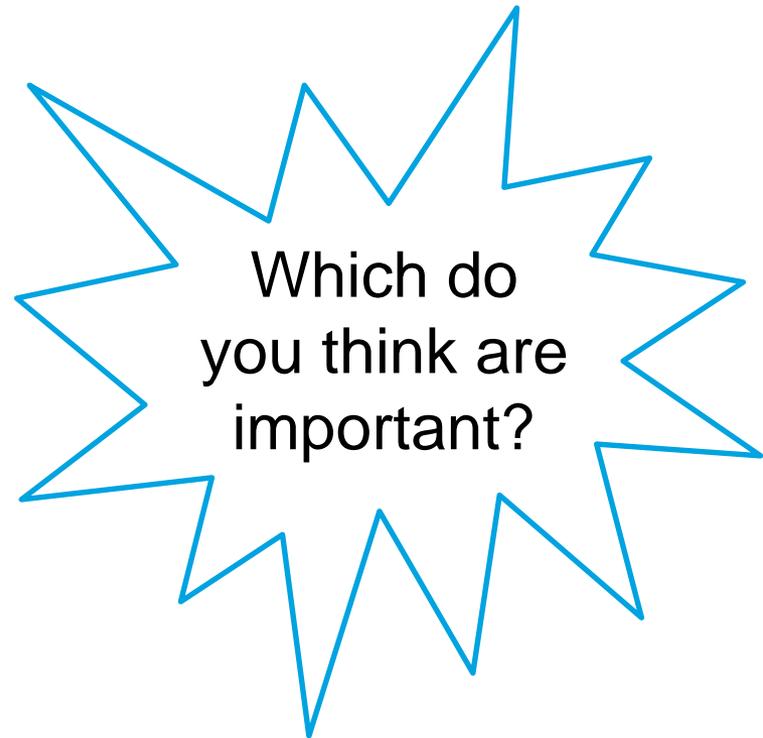
- Statement of objective
- Clinical background
- Criteria for conclusion
- Input data described
- Assumptions for scenarios described
- Assumptions for scenarios justified
- Model assumptions stated
- Model assumptions justified
- Model assumptions checked
- Sensitivity analyses presented
- Any sensitivity results unfavourable to thesis?
- Limitations described
- Analysis described



13 items

# Attributes assessed in survey – planning & presentation

- Statement of objective
- Clinical background
- Criteria for conclusion
- Input data described
- Assumptions for scenarios described
- Assumptions for scenarios justified
- Model assumptions stated
- Model assumptions justified
- Model assumptions checked
- Sensitivity analyses presented
- Any sensitivity results unfavourable to thesis?
- Limitations described
- Analysis described



# Attributes assessed in survey – planning & presentation

- Statement of objective
- Clinical background
- Criteria for conclusion
- Input data described
- Assumptions for scenarios described
- Assumptions for scenarios justified
- Model assumptions stated
- Model assumptions justified
- Model assumptions checked
- Sensitivity analyses presented
- Any sensitivity results unfavourable to thesis?
- Limitations described
- Analysis described

# Attributes assessed in survey – planning & presentation

• Statement of objective	26 (90%)
• Clinical background	26 (90%)
• Criteria for conclusion	2 (7%)
• Input data described	26 (90%)
• Assumptions for scenarios described	18 (62%)
• Assumptions for scenarios justified	16 (55%)
• Model assumptions stated	26 (90%)
• Model assumptions justified	21 (72%)
• Model assumptions checked	21 (72%)
• Sensitivity analyses presented	8 (31%)
• Any sensitivity results unfavourable to thesis?	0 (0%)
• Limitations described	14 (48%)
• Analysis described	14 (48%)

# More technical attributes assessed in survey

- Software stated
- Software version given
- Seed(s) stated
- Programming code available
- QC process described
- Measure of simulation uncertainty
- Results include confidence intervals (CIs)
- Any operating characteristics assessed
  - › e.g. type I error, power, bias, coverage of CIs



8 items

# More technical attributes assessed in survey

- Software stated
- Software version given
- Seed(s) stated
- Programming code available
- QC process described
- Measure of simulation uncertainty
- Results include confidence intervals (CIs)
- Any operating characteristics assessed
  - › e.g. type I error, power, bias, coverage of CIs

Which do  
you think are  
important?

# More technical attributes assessed in survey

- Software stated
- Software version given
- Seed(s) stated
- Programming code available
- QC process described
- Measure of simulation uncertainty
- Results include confidence intervals (CIs)\*
- Any operating characteristics assessed
  - › e.g. type I error, power, bias, coverage of CIs

# More technical attributes assessed in survey

- Software stated 18 (62%)
- Software version given 14 (48%)
- Seed(s) stated 2 (7%)
- Programming code available 4 (14%)
- QC process described 1 (3%)
- Measure of simulation uncertainty 5 (17%)
- Results include confidence intervals (CIs)\* 17 (59%)
- Any operating characteristics assessed 11 (38%)
  - › e.g. type I error, power, bias, coverage of CIs

# Essential to a successful program simulation

- Integrate the quantitative experts, the clinical experts and the other decision-makers
- The whole team is needed, to decide on
  - > what aspects of development program to assess
  - > assumptions
  - > scenarios
  - > how to interpret results
  - > what new alternatives to assess

# References

Burman C-F, Wiklund S. (2011) Modelling and simulation in the pharmaceutical industry—some reflections. *Pharmaceutical Statistics*; 10: 508-516.

Burton A, Altman D, Royston P, Holder R (2006) The design of simulation studies in medical statistics. *Statistics in medicine*; 25: 4279-4292.

Bonate P, Gillespie W, Ludden T, Rubin DB, Stanski D (1999) *Simulation in drug development: good practices*. In Draft publication of the Centre for Drug Development Science (CDDS), Holford NHG, Hale M, Ko H, Steimer J-L, Sheiner LB, Peck CC (eds). University of California: San Francisco;; pp. 1–18.

Gaydos B, Anderson K, Berry D, Burnham N, Chuang-Stein C, Dudinak J, Fardipoor P, Gallo P, Givens S, Lewis R, Maca J, Pinheiro J, Pritchett Y, Krams M (2009) Good practices for adaptive clinical trials. *Drug Information Journal*; 43: 539–556.

Marshall S, Burghaus R, Cosson V, Cheung S, Chenel M, DellaPasqua O, Frey N, Hamren B, Harnisch L, Ivanow F, Kerbusch T, Lippert J, Milligan P, Rohou S, Staab A, Steimer J, Tornøe C, Visser S. (2016) Good Practices in Model-Informed Drug Discovery and Development (MID3): practice, application and documentation. *CPT: Pharmacometrics & Systems Pharmacology*; (2016) 5, 93–122, available at <http://onlinelibrary.wiley.com/doi/10.1002/psp4.12049/epdf>, accessed 22Apr2016

O’Kelly M, Anisimov V, Campbell C, Hamilton S (2016) Proposed Best Practice for Projects that Involve Modelling and Simulation. *Pharmaceutical Statistics* (accepted).

Smith M, Marshall A (2010) Importance of protocols for simulation studies in clinical drug development. *Statistical Methods in Medical Research*; 20: 613-622.

US Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. (1999) *Guidance for industry: population pharmacokinetics*. US Food and Drug Administration: Rockville, MD.