Best practice in modelling and simulation: initiatives from PSI and EFPIA

Michael O'Kelly
Quintiles Advisory Services Analytics

Summary of presentation

- Work on Best Practice since 2011.
- European Federation of Pharmaceutical Industries and Associations (EFPIA) working group
  - > Model Informed Drug Discovery and Development (MID3).
  - > MID3 publication on Good Practice;
  - > Comprehensive and agreed across a large number of pharma companies.
- Statisticians in the Pharmaceutical Industry (PSI) Special Interest Group (SIG)
  - > Modelling and Simulation SIG Best Practice initiative –
  - > includes template for specification of modelling and simulation;
  - > emphasis on flexibility of requirements for Best Practice;
  - > can be used on its own or as a tool for implementing MID3 Good Practice
- People and Best Practice

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.
- 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.
- Rob Hemmings of EMA called for a Best Practice document.

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.
- 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.
- Rob Hemmings of EMA called for a Best Practice document.

EMA: “Best practice” depends on importance of project

EUROPEAN MEDICINES AGENCY
SCIENCE · MEDICINES · HEALTH

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

Hemmings R. M&S good practices and next steps. EMA-EFPIA Modelling and Simulation Workshop, 30 Nov – 1 Dec 2011, London. Available at https://www.youtube.com/watch?v=BSsaUmMuUAE&index=12&list=PL7K5dNgKnawY96v4FlgjTwUFAgUCXrRmw. Accessed 31 March 2016.

EMA: “Best practice” depends on importance of project

EUROPEAN MEDICINES AGENCY
SCIENCE · MEDICINES · HEALTH

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

QUINTILES

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

Best Practice, background

- 2012: EFPIA started working group for “model-informed drug discovery and development” (MID3).
  - Represented: Pfizer, Bayer, Roche, AZ, GSK, J&J, Merck, BI, Novartis, Novo Nordisk.
- Jan 2016: MID3 publish Good Practice in *CPT: Pharmacometrics and Systems Pharmacology*.
  - plus supplement listing 103 papers/slidesets: projects using modelling and simulation.

- Meanwhile…
- 2014: Michael O’K submitted draft Best Practice to SIG for review.
- 2015 February: PSI SIG Hackathon to try out the SIG Best Practice proposal.
- 2015 May: SIG Best Practice document presented at PSI annual conference.
- 2015 November: Best Practice document adopted by PSI Board pending acceptance by *Pharmaceutical Statistics.*

MID3 paper on Good Practices

- MID3 paper “Good Practices in Model-Informed Drug Discovery and Development (MID3): practice, application and documentation.
  - Wide industry representation: Pfizer, Bayer, Roche, AstraZeneca, GSK, J&J, Merck, BI, Novartis, Novo Nordisk.
  - Aim of MID3 working group:
    - “to assemble a collection of “good practice” recommendations in order to address the heterogeneity in both the quality and content of MID3 documentation submitted to European Regulatory agencies”.
  - Aim then expanded to include
    - foster integration of MID3 into broader research and development (R&D) environment;
    - illustrate use of MID3 applications;
    - evaluate the impact of MID3 on R&D efficiency

MID3 group definition of model informed drug discovery and development

- ‘‘A quantitative framework for prediction and extrapolation, centered on knowledge and inference generated from integrated models of compound, mechanism and disease level data and aimed at improving the quality, efficiency and cost effectiveness of decision making’’.
- The concept that R&D decisions are ‘‘informed’’ rather than ‘‘based’’ on model derived outputs is a central tenet.

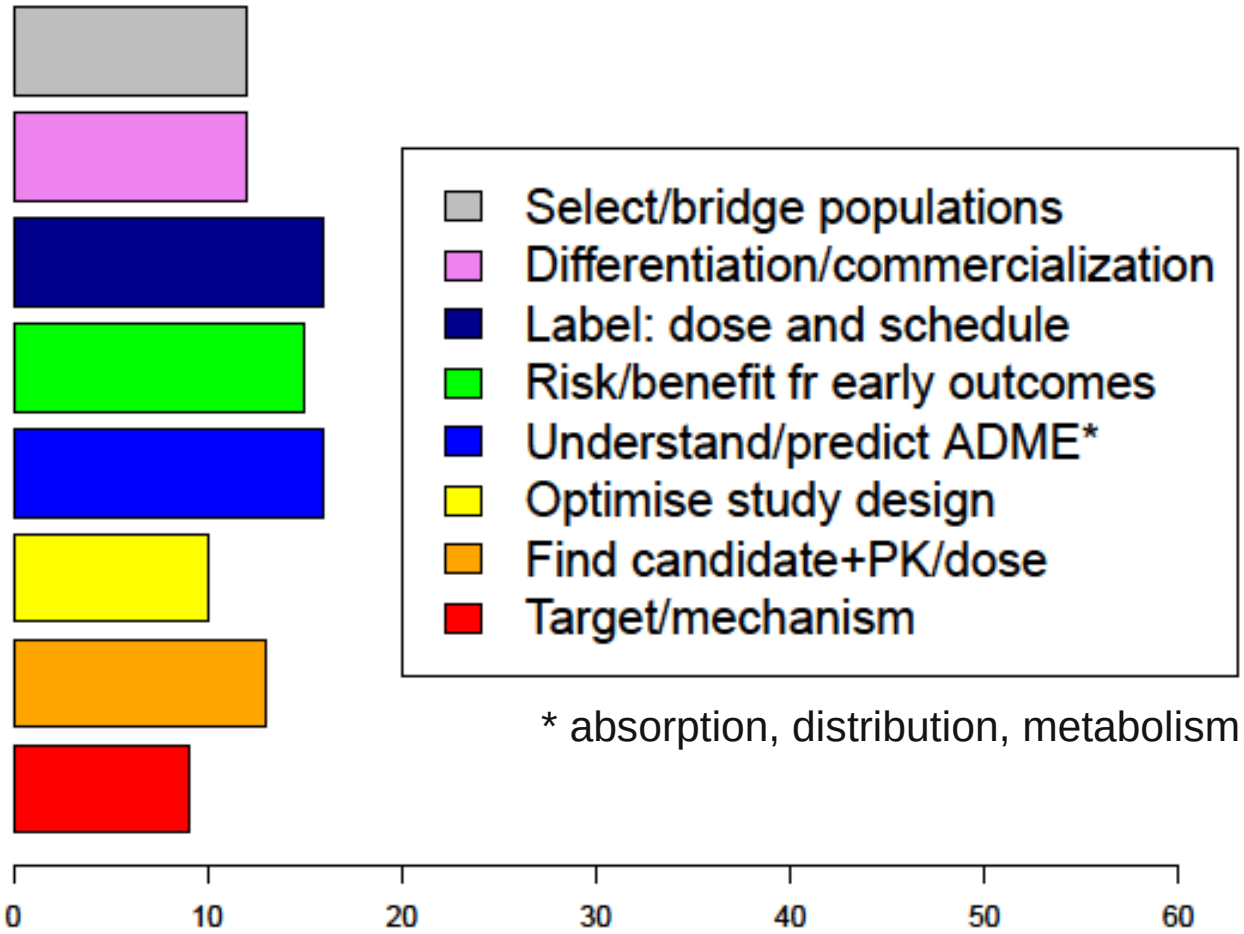
Summary of MID3 paper

- Why MID3 is important.
- What MID3 is; its challenges and applications.
- How to do MID3: the core “Good Practice” part of the paper.

Summary of MID3 paper

- Why MID3 is important.
- What MID3 is; its challenges and applications.
- How to do MID3: the core “Good Practice” part of the paper.

Summary of MID3 paper – survey of 103 publications "illustrative rather than exhaustive"



* absorption, distribution, metabolism, and excretion

Summary of MID3 paper

- Why MID3 is important.
- What MID3 is; its challenges and applications.
- How to do MID3: the core “Good Practice” part of the paper.

Summary of MID3 paper

- Challenges and opportunities
  - How to get MID3 accepted and used throughout.
  - How to argue systematically and scientifically for value of MID3
    - use “High, Medium, Low” impact categories for MID3 projects.

Summary of MID3 paper

- Why MID3 is important.
- What MID3 is; its challenges and applications.
- How to do MID3: the core “Good Practice” part of the paper.

Summary of MID3 paper

- Why MID3 is important.
- What MID3 is; its challenges and applications.
- How to do MID3: the core “Good Practice” part of the paper.

Summary of MID3 paper: Good Practice

- Recommended documentation of planning, conduct, and reporting of MID3 analyses.

MID3 headings for Good Practice

Components of Good Practice plans		
Analysis plan	Simulation plan	Report

• Introduction
• Objectives
• Data plan
• Data exploration
• Methods
  • Model building
  • Selection+evaluation
  • Qualification
• Assumptions
• Results
	• Introduction
	• Objectives
	• Additional data
	• Methods
	  • Identify model
	  • Limitations
	  • Qualification
	• Assumptions
	• Results
		• Synopsis
		• Introduction
		• Objectives
		• Data
		• Methods
		  • Identify model
		  • Limitations
		  • Qualification
		• Assumptions
		• Results
		• Applications/simulations
		• Discussion
		• Conclusion
		• Appendices

MID3 headings for Good Practice

Components of Good Practice plans		
Analysis plan	Simulation plan	Report

• Introduction
• Objectives
• Data plan
• Data exploration
• Methods
  • Model building
  • Selection+evaluation
  • Qualification
• Assumptions
• Results
	• Introduction
	• Objectives
	• Additional data
	• Methods
	  • Identify model
	  • Limitations
	  • Qualification
	• Assumptions
	• Results
		• Synopsis
		• Introduction
		• Objectives
		• Data
		• Methods
		  • Identify model
		  • Limitations
		  • Qualification
		• Assumptions
		• Results
		• Applications/simulations
		• Discussion
		• Conclusion
		• Appendices

MID3 headings for Good Practice – includes recommendations for each heading

Components of Good Practice plans		
Analysis plan	Simulation plan	Report

• Introduction
• Objectives
• Data plan
• Data exploration
• Methods
  • Model building
  • Selection+evaluation
  • Qualification
• Assumptions
• Results
	• Introduction
	• Objectives
	• Additional data
	• Methods
	  • Identify model
	  • Limitations
	  • Qualification
	• Assumptions
	• Results
		• Synopsis
		• Introduction
		• Objectives
		• Data
		• Methods
		  • Identify model
		  • Limitations
		  • Qualification
		• Assumptions
		• Results
		• Applications/simulations
		• Discussion
		• Conclusion
		• Appendices

Best Practice in modelling and simulation

MID3

- When to use simulation
- Key elements for a good plan
- Quality control
- Iterative nature of the MID3 process

PSI

- When to use simulation
- Key elements for a good plan
- Quality control
- Iterative nature of modelling and simulation

Best Practice in modelling and simulation

MID3

- Agreed across 10 companies.
- Emphasis on integrating MID3 into the general pharmaceutical development process
- Three planning documents.
- Lists key recommended elements.

PSI

- Authored by SIG, to be adopted by PSI.
- Emphasis on providing a tool for Best Practice for the working statistician
- One specification for a project.
- Emphasis on flexibility – specification should include key elements or justify their absence.

Best Practice in modelling and simulation

MID3

- Emphasis on hypothesis **generation** rather than hypothesis **testing**.
- Tends not to go into detail on technical requirements.

PSI

- Allows for possibility of hypothesis testing.
- Suggests including “less likely” scenarios in simulations.
- Considers technical detail, e.g., operating characteristics; use of confidence intervals; measure of stochastic variability in simulations; randomisation seed; software version.
- Includes guidance on when to include simulation.

Best Practice in modelling and simulation

MID3

- Emphasis on hypothesis **generation** rather than hypothesis **testing**.
- Tends not to go into detail on technical requirements.

PSI

- Allows for possibility of hypothesis testing.
- Suggests including “less likely” scenarios in simulations.
- Considers technical detail, e.g., operating characteristics; use of confidence intervals; measure of stochastic variability in simulations; randomisation seed; software version.
- Includes guidance on when to include simulation.

PSI Best Practice document can be used as a tool or template to implement Best Practice as described by MID3 and/or PSI. MID3 and PSI SIG share vision of good practice harmonised across the uses of modelling and simulation.

PSI Modelling and simulation SIG Best Practice document

Short: 11 pages

Contains sections as follows

- Definitions
- Scope of the document
- Objective
- Introduction
- The project specification
  - Quality control
  - Presentation of results
- Changes to the specification

PSI Modelling and simulation SIG Best Practice document

Contains sections as follows

- **Definitions**
- Scope of the document
- Objective
- Introduction
- The project specification
  - Quality control
  - Presentation of results
- Changes to the specification

Definitions of modelling and simulation

- “Project to answer specified scientific and business questions using model(s) with estimates that:
  - have a stochastic element, and/or
  - have combinations of attributes that are not present together in a single dataset or in the individual assumptions used by the model, and/or
  - extrapolate beyond the directly specified assumptions and/or beyond the support provided by the data used by the model.”
- vs. MID3 definition above:
  - “A quantitative framework for prediction and extrapolation, centered on knowledge and inference generated from integrated models of compound, mechanism and disease level data and aimed at improving the quality, efficiency and cost effectiveness of decision making”

PSI Modelling and simulation SIG Best Practice document

Contains sections as follows

- Definitions
- Scope of the document
- Objective
- Introduction
- The project specification
  - Quality control
  - Presentation of results
- Changes to the specification

PSI Modelling and simulation SIG Best Practice document

Contains sections as follows

- Definitions
- Scope of the document
- Objective
- **Introduction**
- The project specification
  - Quality control
  - Presentation of results
- Changes to the specification

Introduction: when to include simulation in a project

- Closed-form approaches vs. simulation.

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.

PSI Modelling and simulation SIG Best Practice document

Contains sections as follows

- Definitions
- Scope of the document
- Objective
- Introduction
- The project specification
  - Quality control
  - Presentation of results
- Changes to the specification

PSI Modelling and simulation SIG Best Practice document

Contains sections as follows

- Definitions
- Scope of the document
- Objective
- Introduction
- **The project specification**
  - Quality control
  - Presentation of results
- Changes to the specification

Checklist for a best-practice specification – include these, or say why not

- Summary – optional
- Introduction, including
  - clinical/statistical background,
  - objectives,
  - metrics/criteria for conclusions.
- Simulation and analysis/design
  - Scenarios assumed and assumptions made, including sensitivity analyses.
  - State and justify assumptions (including some less likely ones).
  - Data sets
  - Analysis
  - Operating characteristics
  - Logistics – the execution environment, location of the programming code.

PSI Modelling and simulation SIG Best Practice document

Contains sections as follows

- Definitions
- Scope of the document
- Objective
- Introduction
- The project specification
  - Quality control
  - Presentation of results
- Changes to the specification

A selection of key elements for the specification

- 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

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, coverage of CIs

PSI Modelling and simulation SIG Best Practice document

Contains sections as follows

- Definitions
- Scope of the document
- Objective
- Introduction
- The project specification
  - **Quality control**
  - **Presentation of results**
- Changes to the specification

Best practice – during, after, and next time around…

- Quality control – level of QC should be appropriate
  - from review of specification (low-impact project)
  - to independent programming of project (some high-impact projects)
- Presentation of results
  - may vary depending on audience – plan in advance outputs for each audience
  - use of confidence intervals
- Changes to specification
  - Specification should be auditable, e.g.,
    - revision history
    - formal amendment (as in protocol amendment)
    - include old versions as appendices

Principle: do what is necessary for Best Practice, but not more

- 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.

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 mvtBinaryEP 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.

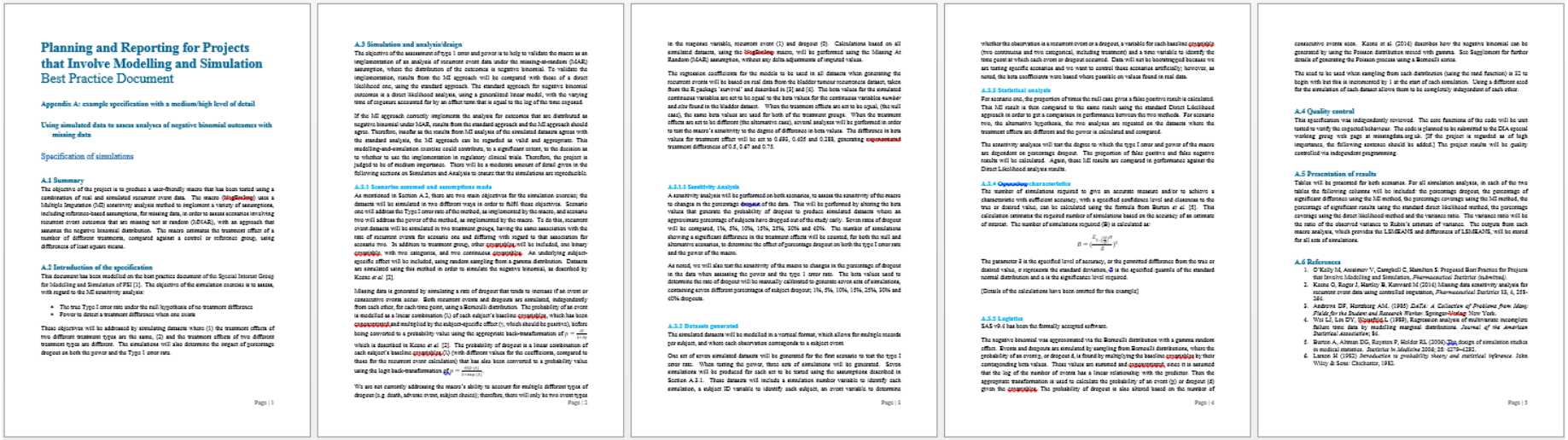
B.3 Quality control
Given that this modelling and simulation task is of low importance and will not of itself lead to a decision, the specification will be submitted to the requestor of the calculation, but no further QC of the production of results is planned. The output will be checked against the 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.

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 macro that has been tested using a combination of real and simulated recurrent event data. The macro (SingleSimp) uses a Multiple Imputation (MI) sensitivity analysis method to implement a variety of assumptions, including reference-based assumptions, for missing data, in order to assess scenarios involving recurrent event outcomes that are missing not at random (MNAR), with an approach that assumes the negative binomial distribution. The macro estimates the treatment effect of a number of different treatments, compared against a control or reference group, using differences of least square means.

A.2 Introduction of the specification

This document has been modelled on the best practice document of the Special Interest Group for Modelling and Simulation of PSI [1]. The objective of the simulation exercises is to assess, with regard to the MI sensitivity analysis:

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

These objectives will be addressed by simulating datasets 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 simulations 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 1 error and power is to help to validate the macro 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 time of exposure 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 outcomes that are distributed as negative binomial under MAR, results from the standard approach and the MI approach should agree. Therefore, results in the results from MI analysis of the simulated datasets agrees with the standard analysis, the MI approach can be regarded as valid and appropriate. This modelling-and-simulation exercise could contribute, to a significant extent, to the decision as to whether to use the implementation in regulatory clinical trials. Therefore, the project is judged to be of medium importance. There will be a moderate amount of detail given in the following sections on Simulation and Analysis to ensure that the simulations are reproducible.

A.3.1 Scenarios assessed and assumptions made

As mentioned in Section A.2, there are two main objectives for the simulation exercises; the datasets will be simulated in two different ways in order to fulfil these objectives. Scenario one will address the Type I error rate of the method, as implemented by the macro, and scenario two will address the power of the method, as implemented by the macro. To do this, recurrent event datasets will be simulated in two treatment groups, having the same association with the rate of recurrent events for scenario one and differing with regard to that association for scenario two. In addition to treatment group, other covariates will be included, one binary covariate, with two categories, and two continuous covariates. An underlying subject-specific effect will be included, using random sampling from a gamma distribution. Datasets are simulated using this method in order to simulate the negative binomial, as described by Keene *et al.* [2].

Missing data is generated by simulating a rate of dropout that tends to increase if an event or consecutive events occur. 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 (λ) of each subject's baseline covariates, which has been exponentiated and multiplied by the subject-specific offset (t), which should be positive), before being converted to a probability value using the appropriate back-transformation $p(\lambda) = \frac{\exp(\lambda)}{1+\exp(\lambda)}$ which is described in Keene *et al.* [2]. The probability of dropout is a linear combination of each subject's baseline covariates (λ) (with different values for the coefficients, compared to those for the recurrent event calculation) that has also been converted to a probability value using the logit back-transformation $p(\lambda) = \frac{\exp(\lambda)}{1+\exp(\lambda)}$.

We are not currently addressing the macro's ability to account for multiple different types of dropout (e.g. death, adverse event, subject choice); 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 SingleSimp macro, will be performed using the Missing At Random (MAR) assumption, without any delta adjustments of imputed 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 [3] and [4]. The beta values for the simulated continuous variables are set to be equal to the beta values for the continuous variables *number* and *size* found in the bladder dataset. When the treatment effects are set to be equal, (the null case), the same beta values are used for both of the treatment groups. When the treatment effects are set to be different (the alternative case), several analyses will be performed in order to test the macro's sensitivity to the degree of difference in beta values. The difference in beta values for treatment effect will be set to 0.492, 0.405 and 0.288, generating exponentiated treatment differences of 0.5, 0.67 and 0.75.

A.3.1.1 Sensitivity Analysis

A sensitivity analysis will be performed on both scenarios, to assess the sensitivity of the macro to changes in the percentage dropout of the data. This will be performed by altering the beta 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%, 25%, 30% and 40%. The number of simulations showing 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 macro.

As noted, we will also test the sensitivity of the macro to changes in the percentage of dropout in the data when assessing the power and the type 1 error rate. The beta values used to determine the rate of dropout will be manually calibrated to generate seven sets of simulations, containing seven different percentages of subject dropout; 1%, 5%, 10%, 15%, 25%, 30% and 40% dropout.

A.3.2 Datasets generated

The simulated datasets will be modelled in a vertical format, which allows for multiple records per subject, and where each observation corresponds to a subject event.

One set of seven simulated datasets will be generated for the first scenario to test the type I error rate. When testing the power, three sets of simulations will be generated. Seven simulations will be produced for each set to be tested using the assumptions described in Section A.2.1. These datasets will include a simulation number variable to identify each simulation, a subject ID variable to identify each subject, an event variable to determine

Page | 3

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 locf/nocb/ipolated because we are testing specific scenarios and we want to control these scenarios artificially; however, as noted, the beta coefficients were based where possible on values found in real data.

A.3.3 Statistical analysis

For scenario one, the proportion of times the null case gives a false positive result is calculated. This MI 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 repeated as 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 macro are dependent on percentage dropout. The proportion of false positive and false negative results will be calculated. Again, these MI results are compared in performance against the Direct Likelihood analysis results.

A.3.4 Operating characteristics

The number of simulations required to give an accurate measure and/or to achieve a characteristic with sufficient accuracy, with a specified confidence level and closeness 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 an estimate of interest. The number of simulations required (R) is calculated as:

$$R = \left(\frac{z_{1-\frac{\alpha}{2}}\sigma}{\delta}\right)^2$$

The parameter δ is the specified level of accuracy, or the guaranteed difference from the true or desired value, σ represents the standard deviation, α is the specified quantile of the standard normal distribution and α is the significance level required.

[Details of the calculations have been omitted for this example]

A.3.5 Logistics

SAS v9.4 has been the formally accepted software.

The negative binomial was approximated via the Bernoulli distribution with a gamma random effect. Events and dropouts are simulated by sampling from Bernoulli distributions, where the probability of an event, or dropout d, is found by multiplying the baseline covariates by their corresponding beta values. These values are summed and exponentiated, since it is assumed that the log of the number of events has a linear relationship with the predictor. Then the appropriate transformation is used to calculate the probability of an event (η) or dropout (d) given the covariates. The probability of dropout is also altered based on the number of

Page | 4

consecutive events seen. Keene et al. (2014) describes 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 Bernoulli series.

The seed to be used when sampling from each distribution (using the rand function) is 52 to begin with but this is incremented by 1 at the start of each simulation. Using a different seed for the simulation of each dataset allows them to be completely independent of each other.

A.4 Quality control

This specification was independently reviewed. The core functions of the code will be unit tested to verify the expected behaviour. The code is planned to be submitted to the DIA special working group web page at missingdata.org.uk. [If the project is regarded as of high importance, the following sentence should be added:] The project results will be quality controlled via independent programming.

A.5 Presentation of results

Tables will be presented 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 differences 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 estimate of variance. The output from each macro analysis, which provides the LSMEANS and differences of LSMEANS, will be stored for all sets of simulations.

A.6 References

1. O'Kelly M, Antonov V, Campbell C, Hamilton S. Proposed Best Practice for Projects that Involve Modelling and Simulation. *Pharmaceutical Statistics* (Submitted).
2. Keene O, Roger J, Hartley R, Kenward M (2014) Missing data sensitivity analysis for recurrent event data using controlled imputation, *Pharmaceutical Statistics* 13, 6, 258-264
3. Andrews DF, Herzberg AM. (1985) *DATA: A Collection of Problems from Many Fields for the Student and Research Worker*. Springer-Verlag New York.
4. Wei LJ, Lin DY, Weissfeld L. (1989), Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *Journal of the American Statistical Association*; 84.
5. Burton A, Altman DG, Royston P, Holder RL (2006) The design of simulation studies in medical statistics. *Statistics in Medicine* 2006; 25: 4279–4292.
6. Larson H (1982) *Introduction to probability theory and statistical inference*. John Wiley & Sons: Chichester, 1982.

Page | 5

A.7 Supplement

Justification 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 λ. Here we describe how this data can be approximated by using discrete event simulation with Bernoulli trials. Note that a sufficiently short interval, Δt, can with a probability close to one only contain either 0 or 1 events. Indeed for small Δt the probability of one event occurring in interval $[t, t+\Delta t]$ is proportional to the length of the interval, $\lambda \Delta t$.
Using the properties of a Poisson process we get that the probability to have more than one event is $o(\Delta t)$. Thus, the probability of no event is approximately $q = 1 - \lambda \Delta t$, or $1 - p$.
Then the events in the non-overlapping intervals of this length, Δt, can be considered to be independent Bernoulli trials.

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

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k}$$
$$= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

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

That,

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

for each $k = 0, 1, 2, \ldots$ This proves that when events generated that are consistent with the above assumptions, the number of events, N, in an interval of fixed length t, has a probability function of:

$$P_n(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, k = 0, 1, 2, \ldots$$

where N is the Poisson random variable with parameter λt. This shows how the Bernoulli trials can be used to generate approximately a Poisson process in continuous time.

For recurrent events the negative binomial distribution can be derived as a mixture of Poisson distributions where the mixing distribution acts multiplicatively on the intensity of the Poisson process. The mixing distribution is a gamma distribution with mean λ. The negative binomial distribution is also called a gamma-Poisson Distribution for this reason. The gamma-Poisson

Page | 6

model is a special case of generalized linear mixed models with a Poisson distribution and log link. In this model, observations Y_{it} are made on the ith subject in interval t:

$$Y_{it} | \eta_i \sim Poisson(U_{it})$$

and

$$\ln(U_{it}) = X_{it}\beta + \eta_i$$

where η_i is a subject specific effect with some distribution on the real line. The design matrix X_{it} accounts for both the treatment type and the baseline covariates. Set $\psi_i = \exp(\beta_i)$ and $\psi_{it} = \exp(X_{it}\beta)$, then:

$$Y_{it} | \psi_i \sim Poisson(\psi_i \psi_{it})$$

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

In the planned simulations, the ψ_i is generated from a gamma distribution and multiplied by the event rate, ψ_i, which is the exponentiated linear predictor described in Section A.2.1.

Page | 7

PSI proposed Best Practice document

- PSI SIG Best Practice document
  - Authored by volunteers from the SIG
    - SIG members from variety of pharmaceutical companies.
    - All authors of the Best Practice document from contract research organizations.
    - Agreed to be adopted by Board of PSI.
  - Aim of document
    - Define Modelling and Simulation;
    - when to use modelling and simulation vs. closed form solutions;
    - elements of a specification for a modelling and simulation project;
    - managing changes to the specification.

MID3 paper on Good Practices

- MID3 paper “Good Practices in Model-Informed Drug Discovery and Development (MID3): practice, application and documentation.
  - Wide industry representation: Pfizer, Bayer, Roche, AstraZeneca, GSK, J&J, Merck, BI, Novartis, Novo Nordisk.
  - Aim of MID3 working group:
    - “to assemble a collection of “good practice” recommendations in order to address the heterogeneity in both the quality and content of MID3 documentation submitted to European Regulatory agencies”.
  - Aim then expanded to include
    - foster integration of MID3 into broader research and development (R&D) environment;
    - illustrate use of MID3 applications;
    - evaluate the impact of MID3 on R&D efficiency

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

- Agreement all aspects of the process
- Some differences in emphasis
- The two groups are working together to promote good practice
- The two groups will participate at session on Best Practice at 2016 annual conference of the statisticians professional body PSI, London, UK

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

- Agreement all aspects of the process
- Some differences in emphasis
- The two groups are working together to promote good practice
- The two groups will participate at session on Best Practice at 2016 annual conference of the statisticians professional body PSI, London, UK

MID3 and PSI SIG share vision of good practice harmonised across the uses of modelling and simulation.

Essential to a successful modelling and simulation project

- 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, Lowis 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* (submitted).

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.