1. Introduction

Sensitivity analyses are used in clinical trials to test how robust our inferences are to data limitations, and to violations in the assumptions made for the specified statistical model. Per EMA recommendations (EMA, 2010), the impact of missing data should be investigated as part of the planned statistical analyses. Modelling techniques may often assume that data are missing at random, meaning that unobserved data are distributed in the same way as the observed data. This may lead to bias when estimating a treatment difference. For example, if subjects withdraw from the study due to deterioration then the unobserved data may be less favourable than the observed data. Tipping point analyses may be used to assess how violations in the missing at random assumption can affect the estimated treatment difference and how far we can deviate from this assumption before the trial results become non-significant. One possible tipping point method utilizes multiple imputation techniques to test a range of assumptions regarding the possible values of unobserved data following a subject’s withdrawal from a study. The methods described in this paper are particularly relevant for assessing the impact of missing data in a mixed effect model repeated measurement (MMRM) but these ideas can also be extended to event based data analysed using a negative binomial model.

2. Multiple Imputation

1. An appropriate (congenial) regression model is fitted to the data using an MCMC approach which samples from the joint posterior distribution for the estimated regression coefficients.
2. A number of independent samples are drawn so that we can repeat the analysis multiple times using different posterior estimates.
3. For each sample iteration, missing values can be calculated for each subject using the parameter estimates and a subject’s individual covariates (including randomised treatment arm) and their non-missing values.
4. We consider the treatment arm of interest and the comparator treatment which we will call the experimental and reference treatments respectively.
5. For both arms, data are imputed. In a tipping point analysis we use a missing at random (MAR) assumption, however for other multiple imputation techniques missing not at random assumptions such as jump to reference may be implemented.
6. MAR assumes that missing data are distributed in the same way as non-missing data.
7. This gives us an imputed dataset for each sample from the MCMC process.
8. The now complete data for the desired time point are analysed for each MCMC sample iteration using an ANCOVA model.
9. The results of each analysis are then combined using Rubin’s rule (Rubin, 1987) to provide a single estimate of the treatment difference between the reference and experimental arms.

3. Tipping Point Analysis

In this method of tipping point analysis, the data are imputed multiple times as described in Section 2. However, fixed values $\delta_E$ and $\delta_R$ are then added to the imputed values (at step 5) for the experimental and reference arms respectively before the imputed datasets are analysed. Deltas can be either additive for continuous endpoints or multiplicative if the endpoint is a rate. This process is then repeated multiple times, varying the values of $\delta_E$ and $\delta_R$ independently to provide a range of treatment estimates. It is then assessed for which values of $\delta_E$ and $\delta_R$ the treatment difference becomes non-significant (i.e. the ‘tipping point’).

**Example**

Consider a fictional two-arm trial where the primary endpoint $Y$ is continuous and measured across multiple timepoints. In this case, a greater value of $Y$ is desirable. We will consider two different scenarios to illustrate how to interpret the results of a tipping point analysis. Under both scenarios, the primary analysis was carried out using an MMRM model on the non-missing data which found the results to be statistically significant in favour of the experimental treatment. A tipping point analysis is performed, varying $\delta_E$ and $\delta_R$ between -15 and 15 units. The chosen significance level is 0.05 in both cases. In scenario 1 the results are robust to missing data, whereas in scenario 2 the results are sensitive to missing data.

4. Practical Considerations

**MCMC Convergence**

The imputation stage uses an MCMC procedure which needs to be run for sufficient iterations to converge to the joint posterior distribution of the model parameters. The posterior sample also needs to be large enough to minimize uncertainty around the estimates. In MCMC procedures, each draw is dependent on the value of the previous draw which will result in autocorrelation in the sample. For these reasons, it is important to choose a sufficiently large number of draws, level of thinning and number of burn-in iterations to ensure that we have simulated an independent random sample from the posterior distribution. A good way to check that the sample is converging is to run the whole analysis two or three times with a different random seed each time. If convergence is occurring then the results should be similar regardless of the random seed selected.

**Run Times**

This method of tipping point analysis is very computationally intensive which can lead to program run times being extremely long. For example an analysis with 10,000 subjects at 4 timepoints and 7 different values for each delta may take around 24 hours using an MCMC sample of 1,000 with 100 thinning. It is worth making sure that any programs used will work on a small number of MCMC iterations before attempting to produce final results. It may also be a good idea to subset the input dataset while developing code to cut down on run times. If you are running programs on a work server, it is advisable to run programs overnight or at weekends. This is because more memory may be available due to fewer people using the server, which may shorten run times and also reduce the impact on other users.

**Quality Control**

Due to the random element of the tipping point analysis it is unlikely that a QC statistician will get exactly the same results. However, the final estimates from QC analyses should agree within a certain degree of precision. For example if you would usually present your estimates to 3 decimal places then it may be acceptable for a difference of 0.001 between the production and QC results.

**Choice of Deltas**

Choosing values of $\delta_E$ and $\delta_R$ where we can see a tipping point may be challenging. A good strategy may be to start off with relatively large increments between the deltas to see approximately where a tipping point occurs before repeating the analysis with gradually smaller increments between the deltas so that we can see at precisely where the results become non-significant. It is also advisable to seek clinical guidance when choosing the deltas so that they are clinically feasible for the endpoint of interest.

5. Summary

Multiple imputation techniques allow us to test various assumptions about the missing data. Tipping point analysis is a good way to test how sensitive trial results are to missing data. This can be judged by testing at which values of $\delta_E$ and $\delta_R$ a tipping point occurs. A tipping point occurring when there is a small difference between $\delta_E$ and $\delta_R$ implies high sensitivity to missing data. Conversely, if a tipping point only occurs when there is a large difference between $\delta_E$ and $\delta_R$, then we can conclude that the primary analysis results are robust to the missing data in the trial.

Several practical issues should be considered when planning and implementing a tipping point analysis. Run times are likely to be long and the choice of deltas may need to be varied. The analysis may also need to be run several times to check MCMC convergence has occurred. This may result in the analysis being very time consuming from start to finish which should be taken into account when planning timelines for completion.

References

1. EMA, Guidance on Missing Data in Confirmatory Clinical Trials; 2010
2. Rubin DB. Multiple Imputation for Nonresponse in Surveys; 1987