A Practical Guide to Inconsistency Checks in Bayesian Network Meta-Analysis

Caitlin Daly[1], Beatrice C Downing[1], Nicky J Welton[1]

[1] NICE Guidelines Technical Support Unit, Population Health Sciences, Bristol Medical School, University of Bristol

Introduction

Network meta-analysis (NMA) [1] combines trial evidence to estimate the relative differences between several interventions within a connected network. This is achieved by making the consistency assumption: that the relative treatment effect $d_{c,k}$ between any pair of treatments k relative to c is the difference between the effect of treatments k and c relative to a reference treatment 1:

$$d_{c,k} = d_{1,k} - d_{1,c} \qquad (1)$$

By combining trial evidence, we make the assertion that the trial populations are similar enough to be combined – that any participant could receive any of the interventions and that study designs and participants do not differ in factors (effect modifiers) that might change the relative treatment effects [1]. Differences in effect modifiers between trials that make the *same* treatment comparison can lead to heterogeneity, which is a measure of how much the estimate of a particular treatment effect differs between different studies and can be captured by the between-study variance in a random-effects model. Inconsistency in NMA arises when there are differences in effect modifiers between trial evidence on *different* treatment comparisons in the network. For example, suppose we have *direct* estimates from trials of AvB, AvC, and BvC. Then we can form an *indirect* estimate of $d_{BC}^{Indirect}$ using the evidence on d_{AB}^{Direct} and d_{AC}^{Direct} via equation (1). If there are differences in effect modifiers between the AvB, AvC, and BvC evidence then the direct estimate d_{BC}^{Direct} may differ from the indirect estimate $d_{BC}^{Indirect}$ causing inconsistency. Inconsistency can only be assessed where there are "loops" of evidence in the network informed by separate, independent trials, so that both direct and indirect estimates are available. When there are differences between direct and indirect estimates the underlying consistency assumption (equation (1)) is violated and the NMA results may be unreliable. It is therefore essential to check for inconsistency when possible.

If the only loops are made entirely by a single multi-armed trial (>2 arms), or if there are no loops, we cannot assess inconsistency.

This guide aims to provide a practical step-by-step approach to conducting inconsistency checks for Bayesian network meta-analyses. Bayesian network meta-analyses can be conducted in WinBUGS or OpenBUGS using code provided by the NICE Decision Support Unit (DSU) Technical Support Document 2 (TSD2) [2] for fixed and random effect models that assume consistency, for a range of different outcomes. Where networks contain loops informed by different trials, the **unrelated mean**

effects (UME) model can be fitted to provide a global assessment of inconsistency using code provided in the NICE DSU Technical Support Document 4 (TSD4) [3]. In an unrelated mean effects model, the relative treatment effect for each pair of treatments with direct data is estimated separately, rather than from the consistency assumption (equation (1)), so that the direct and indirect effects are no longer constrained to be consistent with each other.

To explore whether the direct and indirect evidence for particular treatment comparisons is inconsistent, the contribution to the posterior mean residual deviance can be plotted for the UME (inconsistency) model against the NMA (consistency) model in a deviance contribution plot (dev-dev plot) [1, 3]. Further exploration via node-splitting can be used to obtain the direct and indirect estimates and perform a local test for inconsistency using code provided in TSD 4 [3].

Whilst the NICE DSU TSDs [2, 3] provide code to conduct inconsistency checks, in our experience these analyses can be complex, require some adaptation to the code, and several complications arise in practical applications in guidelines. In this guide we have summarised some practical tips and give a step-by-step guide to conducting a Bayesian network meta-analysis in WinBUGS, including checks to assess heterogeneity and inconsistency at a global and local level. We begin with some general guidance on conducting the analyses, and then give a step-by-step guide together with a flow-chart.

General guidance on running inconsistency checks

- Standard code for the unrelated mean effects (UME) model provided in TSD4 [3] requires the treatments to be listed in increasing order within studies. Whilst it is good practice to arrange the data in this way, a simple adjustment to the code can be made to accommodate unordered treatments within studies. To achieve this the following code may be used, which should replace the priors for the treatment effects in the standard code:

  ```
  # vague priors for treatment effects
  for(c in 1:nt){ d[c,c]<-0 }
  for(c in 1:(nt-1)){
          for(k in (c+1):nt){
                  d[c,k]~dnorm(0,0.0001)
                  d[k,c] <- -d[c,k]
                  }
          }
  ```

  The accompanying files with the prefix Ch7 in the supplementary folder provide UME codes developed by Dias et al. [1], adjusted for unordered treatments within studies.

- To enable a fair comparison of the model fit between the NMA model and the UME model, both models must be fit **_on the exact same data_**.
  - It is important, particularly for the deviance contribution plot, to have the studies and treatments within the studies to be arranged in the same order in the NMA and UME model datasets.
  - If a continuity correction has been applied to the data to stabilise the results for the NMA model, then the inconsistency models should be run on the continuity corrected data as well (and vice versa).

- - - Example of continuity correction: In studies with zero cells, add 0.5 to numerators, add 1 to denominators. However, note that studies with zero cells for all treatment arms should be excluded from the analysis.
- To enable a fair comparison of the model fit between the NMA and UME models, the same assumptions for priors, likelihood, and link (scale of the analysis) should be modelled, so that the only difference is whether the consistency assumption is applied or not.
  - If a fixed effect (FE) model is selected for the NMA, then a FE UME model should be fitted.
  - If a random effects (RE) model is selected for the NMA, then a RE UME model should be fitted.
    - The same prior (e.g., vague or informative) on the between-study heterogeneity parameter (e.g., standard deviation or variance) should be applied.
  - The same assumptions should also be imposed when running the node-split models.

- When running the models, it is important to first check for convergence.
  - Run at least two chains.
  - To save computer memory, first monitor `d`, `sd` (if RE model), and `totresdev`.
  - Inspect the history and bgr plots to assess whether convergence is satisfactory.
    - When fitting RE models, inspect density plot of `sd` to see if it has been sufficiently estimated by the data. If there is not enough evidence to estimate the between-study SD, consider an informative prior [4, 5]. If an informative prior is preferred, it must be used for both the NMA and UME models for the purposes of inconsistency checking.
  - Once convergence is deemed satisfactory, monitor any additional parameters of interest (e.g., `DIC`, `dev`).

Steps for carrying out NMA with inconsistency checking:
Here we set out the steps to perform an NMA with inconsistency checks. A summary of the steps is also provided as a flowchart in the appendix.

1. First fit fixed effect (FE) and random effects (RE) NMA models to the data.
    - Ensure convergence is satisfactory for both models and there is enough evidence to estimate the between-study standard deviation in the RE model.
        o If there are convergence issues and there are zero cells within the data, consider applying a continuity correction.
        o If there is not enough evidence to estimate the between-study SD, consider an informative prior [4, 5].
    - Complete the following model fit statistics table for both NMA models.

Table X: Model fit statistics for [outcome]

Model[a]	Posterior median of between-study standard deviation (95% credible interval)	Posterior mean of total residual deviance[b]	DIC[c]
NMA: Fixed effect	---		
NMA: Random effects			
UME: X			

[a] Results based on X samples after a burn-in of X iterations
[b] Total residual deviance compared to X data points
[c] Deviance information criteria (DIC) – lower values preferred.

- Compare the posterior mean of the total residual deviance as well as the DIC between both models. Are there any meaningful differences of at least 3 or 5 points between the models? If so, the model with the lower model fit statistics is preferred. If there are no meaningful differences, the FE model should be selected, but ensure it fits the data reasonably well by comparing the total residual deviance to the number of data points (the total number of study arms) – they should be similar. The selected base-case model (FE or RE) should be noted as the model to proceed the inconsistency checks with.

2. On the **exact same data**, fit a FE or RE UME model (corresponding to the base-case NMA model).
    - Specifying the initial values for the treatment effect parameters ($d[c,k]$) for the UME model is complicated. We have provided R code to generate these initial values in the accompanying file "R code to generate initial values for ds_treatment order does not matter.R".
    - Check if convergence of the UME model is satisfactory, and if a RE model is fitted ensure there is enough evidence to estimate the between-study standard deviation.
        o If there are convergence issues for the UME model and there are zero cells within the data, consider applying a continuity correction. If this helps stabilise the results, then the NMA (consistency) model needs to be re-run on the continuity corrected data **for the purpose of checking for inconsistency only**. The final NMA results however should be presented based on the NMA model fit to the raw data (unless continuity corrections were needed for the NMA model too).
        o Similarly, if there is not enough evidence to estimate the between-study SD in the UME model, consider using an informative prior [1, 4, 5]. If an informative prior for the between studies SD is used in the UME model, then the NMA model would needs to be re-run with the same informative prior for SD **for the purpose of checking inconsistency only**. The final NMA results however should

be presented based on the NMA model with a vague prior (unless an informative prior was also needed for the NMA model).
- Monitor the treatment effects (`d`), between-study SD (`sd`) if applicable, the deviance values (`dev`), total residual deviance (`totresdev`), and `DIC`.

3. Complete the model fit statistics table for the UME model. In an external program such as Excel or R, create a deviance contribution plot, where the `dev`'s from the base-case NMA model are plotted along the x-axis, and the `dev`'s from the UME model are plotted along the y-axis (e.g., Figure 1).
   - Resources are available to create these plots in the download folder that accompanies this report. To produce plots in Excel, see "Excel template for deviance plots.xlsx"; to produce plots in R, see "devdev_function.R".
   - Note that the deviance values are outputted by WinBUGS as `dev[i,k]`, where `k` corresponds to the kth arm in the ith study. As such, the data informing the deviance contribution plots should be set up in a format similar to the table presented below, so that the deviance values from the NMA and UME models correspond to each other. Use the indices of the studies (`i`) to identify any points notably below the line of equality (for these points, the `dev[i,k]` value for the UME model will be notably smaller than that of the NMA model).

Node	NMA	UME
`dev[1,1]`		
`dev[1,2]`		
`dev[2,1]`		
…		

4. Check for evidence of inconsistency by comparing the model fit statistics between the base-case NMA model and the UME (inconsistency) model, and by inspecting the deviance contribution plot.
   - Is the total residual deviance or DIC of the UME model at least 3 or 5 points less than that of the NMA model?
   - In the case of RE models, is the between-study SD of the UME meaningfully smaller than that of the NMA model? Note that this can be the case even if the total residual deviance or DIC is not very different between the NMA and UME models. A reduction in between-study SD still indicates inconsistency.
   - In the deviance contribution plot, are there any points notably below the line of equality? For example, in Figure 1A, no points fell below the line of equality. However, in Figure 1B, multiple points fall below the line of equality. The dashed lines in Figure 1B each mark an improvement in residual deviance by multiples of 0.5. There are obvious improvements in the fit of at least two data points. The inconsistency model better predicted the data point in arm 2 of the study in row 4 in the data (an improvement in residual deviance between 1 and 1.5), as well as the data point in arm 2 of the study in row 2 in the data (an improvement in residual deviance of at least 2). Improvements in residual deviance of at least 0.5 should be flagged as well.

   If the answer to any of the above questions is 'yes', then proceed to the steps outlined below. Otherwise, there is no evidence of inconsistency and no further steps are required.
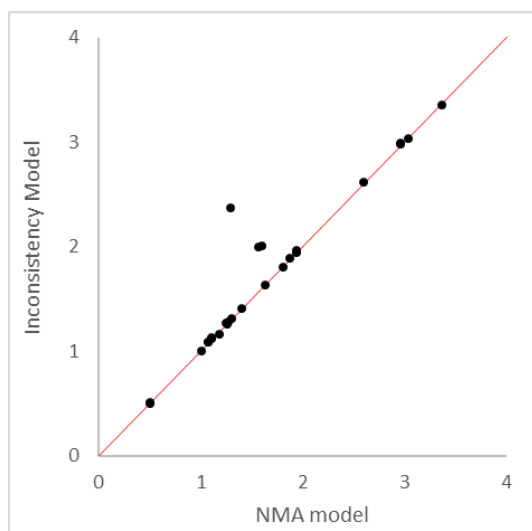
Figure 1A: Example of deviance contribution plot with no points notably below line of equality.
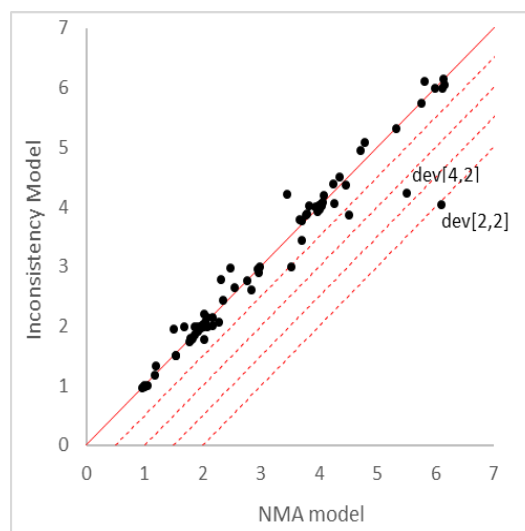
Figure 1B: Example of deviance contribution plot with some points notably below line of equality.

5. Check the data for potential data extraction errors that may be contributing to potential inconsistency. Running node-split models is time intensive, so to minimise risk of having to run them more than once, it might save time to check the data at this point.
    *   For any studies notably below line of equality in the deviance contribution plot, identify the loops of evidence they contribute to.
        o   Double check the data in all studies within these identified evidence loops.
    *   If any data extraction errors are identified, then Steps 1-5 will have to be repeated (including re-running the NMA) with the corrected data.

6. If there is evidence of inconsistency identified from the in the deviance contribution plot, this can be explored further using node-splitting. Fit a FE or RE node-split model (corresponding to the base-case NMA model).
    *   If standard TSD2 [2] NMA code has been used, then the `gemtc` package [6, 7] in R may be used. See R code provided in the accompanying files "R code for node spliting_logit FE.R" and "R code for node splitting_logit RE .R".
        o   The R code will produce plots of direct, indirect, and NMA estimates of treatment effects for the treatment comparisons that have both direct & indirect evidence.
        o   It will also provide the direct and indirect estimates for these treatment comparisons (note these will be on the log scale if modelling ratios).
            ▪   Use these to produce a table of the direct, indirect, and NMA estimates.
            ▪   Note that comparisons with only direct evidence (i.e., a spur) or only indirect evidence (i.e., not compared in a trial) will not be included in the R output. In these cases, use estimates from NMA model to populate the table; the direct or indirect estimates will equal the NMA estimates. Where there is no direct or indirect evidence, leave the cell blank or note there is no evidence (e.g., N/A).
        o   In addition, the R code will output the total residual deviance, DIC, and between study-SD for each model. These should be compared with that of the NMA model outputted by `gemtc` – note `gemtc` calculates the DIC differently from WinBUGS.

- o Be careful when interpreting the p-values, as there are multiple comparisons and the false positive rate for detecting a difference between the direct and indirect estimates increases with these comparisons.
    - o Note that the `gemtc` package does not accept non-integer values (e.g., 0.5) in the data for dichotomous outcomes.
        - ▪ If a continuity correction is required to stabilise the results, there are two options:
        1) Manually calculate the relative effects in each study (e.g., log-ORs, log-HRs) and input this as contrast data into R.
        2) Run the node-split models in WinBUGS using the code from TSD4[3].
    - o It might take a long time to produce convergence plots – hit Esc button to stop R from producing them.
- If a more complex NMA model has been fitted, seek assistance from the TSU, as this will require the node-split model to be written and run in WinBUGS.

7. Use the results of the node-splitting analyses in combination with the deviance contribution plot to identify loops in the network with potential inconsistency.
    - In all studies contributing direct evidence to these loops with potential inconsistency:
    i) Double check the data for any data extraction errors. If data extraction errors have been identified, correct the data and repeat Steps 1-7 again on the updated data.
    ii) Review the inclusion/exclusion criteria of these studies for potential treatment effect modifiers.

8. Attempt to explain reasons for potential inconsistencies. This may formally be done through meta-regression, where a treatment effect modifier is included as a covariate [1, 8].

9. Check to see if presence of inconsistency will impact the treatment recommendations by conducting a threshold analysis [9, 10]. The TSU can assist you with this.

Summary

- This step-by-step guide is designed to help guideline developers to conduct inconsistency checks in guidelines that use a Bayesian NMA. Whilst we have attempted to cover many of the practical issues that arise in guidelines, each dataset has its own unique features. The NICE Guidelines Technical Support Unit is available to support NICE guideline developers to conduct NMAs and associated inconsistency checks. Please do not hesitate to contact the TSU with any queries or requests for assistance.

Appendix: Flowchart summarising inconsistency checks process
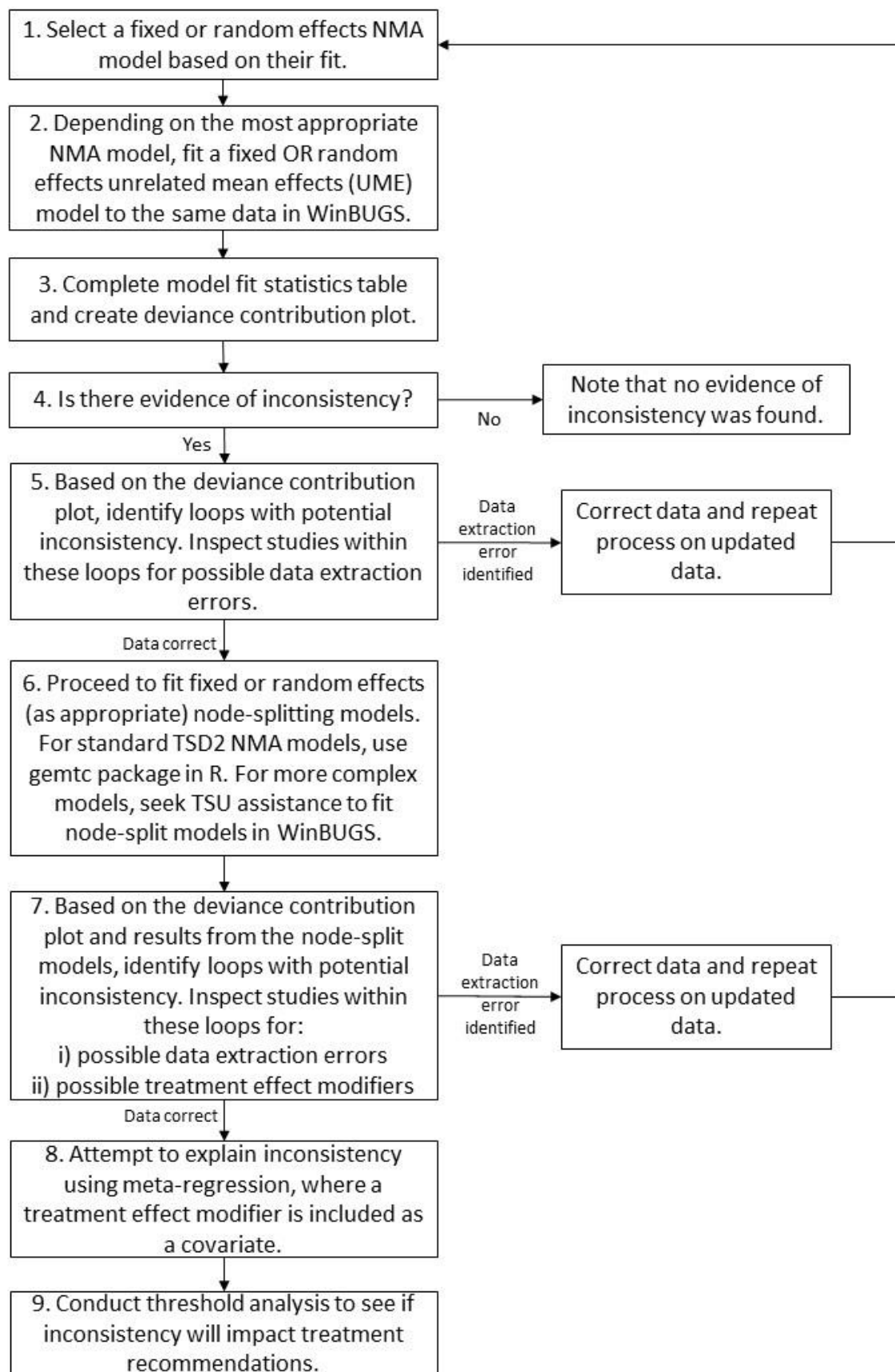
References

1.   Dias, S., et al., *Network meta-analysis for decision making*. Statistics in Practice. 2018: Wiley. 488.

2.   Dias, S., et al., *NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis of randomised controlled trials*, in *Technical Support Document*. 2011.

3.   Dias, S., et al., *NICE DSU Technical Support Document 4: Inconsistency in networks of evidence based on randomised controlled trials*, in *Technical Support Document*. 2011.

4.   Turner, R.M., et al., *Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis.* Statistics in Medicine, 2015. **34**(6): p. 984-998.

5.   Rhodes, K.M., R.M. Turner, and J.P.T. Higgins *Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data.* Journal of Clinical Epidemiology, 2015. **68**: p. 52-60.

6.   van Valkenhoef, G. and J. Kuiper, *gemtc: Network Meta-Analysis Using Bayesian Methods. R package.* 2016, CRAN.

7.   van Valkenhoef, G., et al., *Automated generation of node-splitting models for assessment of inconsistency in network meta-analysis.* Research Synthesis Methods, 2016. **7**: p. 80-93.

8.   Dias, S., et al., *NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment*, in *Technical Support Document*. 2011.

9.   Phillippo, D.M., et al., *Sensitivity of treatment recommendations to bias in network meta-analysis.* Journal of the Royal Statistical Society: Series A (Statistics in Society), 2018. **181**(3): p. 843-867.

10. Phillippo, D.M., et al., *Threshold Analysis as an Alternative to GRADE for Assessing Confidence in Guideline Recommendations Based on Network Meta-analyses.* Annals of Internal Medicine, 2019. **170**(8): p. 538-546.