Is poor adherence to good modelling practices making us worry too much about effect modifiers? A case study in moderate to severe plaque psoriasis.

Tim Disher, PhD, RN | Sandpiper Analytics

CO199

Background

Uncritical use “off the shelf” network meta-analysis models without data exploration and consideration of modifications may overestimate heterogeneity and lead to inappropriate conclusions or use of more complex population adjustment methods. This research explores the impact of poor modelling practices in a case study in plaque psoriasis.

Methods

We use an example of a large network in moderate to severe plaque psoriasis to examine how assumptions of proportional odds across thresholds in the baseline arm leads to borrowing of information across studies, and compare this to an a model where each study’s baseline treatment has individual intercepts estimated. We further examine the influence of logit compared to probit link. Models are compared in terms of absolute (total residual deviance) and relative fit (DIC), in addition to the magnitude of the baseline risk adjustment beta and between trial standard deviation. Base models are based those described in NICE TSD 2[1] and TSD 3[2] and modifications are subsequently made for link and study intercept models.

Results

Logit link consistently led to large improvements in DIC compared to probit link models. Between trial SD was smaller for probit models, but treatment effects were also smaller and SD as a percentage of average basic parameter was comparable. Models that fit distinct intercepts for baseline arm in each trial let to large improvements in total residual deviance (443 vs 500-688 on 370 data points), and DIC. Between trial standard deviation in distinct intercept models included trivial heterogeneity in credible intervals (0.02). Betas estimated in models with study specific intercepts are one fifth the magnitude of those from the default TSD model. Depending on DIC threshold used, a fixed effect model could be justified, leading to additional 4-6 significant comparisons.

Model	Coefficient for BLR Meta-regression	Total residual Deviance	Between Trial SD	Number of significant comparisons	DIC difference vs simplest model
Logit FE – Study Z		443		40	0
Logit RE – Study Z		422	0.16 (0.02 to 0.29)	36	-4
Logit RE BLR – Study Z	-0.19 (-0.29 to -0.1)	415	0.13 (0.03 to 0.25)	34	-1
Logit FE BLR – Study Z	-0.18 (-0.29 to -0.1)	429		37	2
Probit RE BLR - Study Z	-0.09 (-0.14 to -0.05)	433	0.07 (0.01 to 0.14)	34	14
Probit FE BLR – Study Z	-0.1 (-0.14 to -0.06)	445		36	17
Probit RE – Study Z		436	0.11 (0.05 to 0.18)	37	17
Logit RE – TSD Z		534	0.19 (0.07 to 0.31)	37	19
Probit FE – Study Z		473		39	30
Logit FE – TSD Z		568		41	31
Logit RE BLR – TSD Z	-0.98 (-1.07 to -0.86)	516	0.26 (0.2 to 0.33)	34	31
Logit FE BLR – TSD Z	-0.52 (-0.64 to -0.38)	597		39	62
Probit RE – TSD Z		599	0.12 (0.06 to 0.19)	37	89
Probit FE – TSD Z		644		40	108
probit-ord-fe-blr	-0.44 (-0.57 to -0.27)	669		39	134
probit-ord-re-blr	-0.99 (-1.06 to -0.89)	577	0.15 (0.12 to 0.19)	34	

Green highlight indicates best balance of parsimony and model fit; Orange highlight indicates most commonly used model in recent NMAs.
Definitions: Study Z – individual baseline intercepts for each threshold in each study; TSD Z – individual reference threshold for each study with distance between subsequent thresholds shared across all treatments
Abbreviations: FE – fixed effect; RE – random effects; BLR – Baseline risk; SD – standard deviation; DIC – deviance information criteria

Discussion

We find strong evidence in favour of logit link in these data. Since logit and probit links only differ in the tails this difference may be driven by a combination of large treatment effects and rare events. Further, we find that the default model approach of assuming all baseline therapies share the same distance between intercepts leads to very bad absolute model fit and is a key driver of heterogeneity and preference for adjusted models. Authors who have used “REZ” models to relax the proportional odds assumption[3], also relax the shared intercepts assumption, suggesting that preference for these models may not necessarily indicate violation of the proportional odds assumption

Conclusions

Modelling decisions should be based on exploration and consideration of key assumptions. Standard model code should be seen as a starting point, but may be improved upon and lead to important differences. Failure to follow best practices may unnecessarily overstate heterogeneity in NMAs.

References

1. Dias, S., Welton, N. J., Sutton, A. J., & Ades, A. E. (2011). NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials.
2. Dias, S., Sutton, A. J., Welton, N. J., & Ades, A. E. (2011). NICE DSU technical support document 3: heterogeneity: subgroups, meta-regression, bias and bias-adjustment.
3. Fahrbach, K., Sarri, G., Phillippo, D. M., Neupane, B., Martel, S. E., Kiri, S., & Reich, K. (2021). Short-term efficacy of biologic therapies in moderate-to-severe plaque psoriasis: a systematic literature review and an enhanced multinomial network meta-analysis. *Dermatology and Therapy*, *11*, 1965-1998.

SANDPIPER ANALYTICS

sandpiper-analytics.com

tim.disher@sandpiper-analytics.com

linkedin.com/in/tim-disher/

ISPOR 2025 – Montreal, QC