Biometrics 0, 1–7 DOI: 0000-0000-0000
January 2023

Discussion on "Bayesian meta-analysis of penetrance for cancer risk" by Thanthirige Ruberu, Danielle Lakshika Braun, Giovanni Parmigiani, and Swati Biswas

Gianluca Baio

Department of Statistical Science, University College London

**email:* g.baio@ucl.ac.uk

Summary:
Key words: Evidence synthesis,Meta-analysis,Population-adjustment.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

It is a pleasure to contribute to the discussion of the paper by Rubero, Braun, Parmigiani and Biswas (henceforth RBPB). Clearly, the authors have written a thought-provoking and potentially highly impactful piece of work, which I believe can improve the way in which different sources of information are suitably collected and pooled together and contribute to a wide range of applied disciplines.

I should start by saying that I am no expert in applications of statistical modelling to genomics, which is the motivating example for RBPB. However, my own research area (statistical modelling for health technology assessment, HTA) is, arguably, one where the use of evidence synthesis and meta-analytic tools is most prevalent. Consequently, in applications with which I have most familiarity, there has historically been interesting discussion on suitable methodologies to pool information from published studies, particularly from a Bayesian perspective (notably, among others Spiegelhalter et al., 2004; Welton et al., 2012; Dias et al., 2018).

Coming from this vantage point, I would like to discuss a few points that perhaps RBPB could fruitfully account for to improve upon their own framework.

2. Availability of input data in different formats

One of the main objectives for RBPB is to construct a model that can suitably pool together information that is originally presented in formats that are, at face value, not comparable. Their example concentrates on summary statistics from published literature reporting odds ratios rather than relative risks. Their modelling framework proceeds to define a composite likelihood function in which all the studies on the same summary statistics model a suitable transformation. For instance, the studies reporting relative risks y_s^{RR} are modelled using $\log y_s^{RR} \sim \text{Normal}(\ldots, \ldots)$.

This strategy may well be dictated by the data availability — it is possible that GWAS publications tend to simply report the crude summary statistics (e.g. "a RR of 2.1 with a confidence interval [1.73 − 2.41]") and of course, if that is the case, then the modelling strategy of RBPB is reasonable. However, in much of the literature in the HTA arena, it is not uncommon that data are presented in terms of the original counts. For instance, a study s may report the number of events y_s^t out of the n_s^t individuals who have been observed in the $t-$th exposure group. In this situation, it would be perhaps helpful to model directly this information, for instance with a structure similar to the following:

$$\begin{aligned} y_s^t &\sim \text{Binomial}(\pi_s^t, n_s^t) \\ \text{logit}(\pi_s^t) &= \beta_0 + \sum_{k=1}^{K} X_{ks}\beta_k \, [+\ldots], \end{aligned}$$

where the model parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_K)$ can be used to determine baseline and covariates effects and the notation $[+\ldots]$ indicates that the model may include more structure, e.g. in terms of random effects. I wonder whether inducing a posterior on the parameters and therefore on the relevant measure (e.g. OR or RR) can lead to more efficient use of the original data sources, while still allowing for the pooled estimates to be joined up, as in the model derived by RBPB. Of course, one obvious (positive) implication may be in terms of avoiding the necessity to approximate the variance of the estimates produced in each study s (e.g. using a fixed value, w in RBPB's notation).

3. Prior distributions

RBPB have a specific section in their paper dedicated to the definition of the prior distributions for the parameters of their modelling framework. I appreciate that the technical reader would have more details in the appendix, as well as the restrictions in the publication space.

However, I still wonder whether a more explicit discussion of the implications of these priors would facilitate uptake and help practitioners (who may be much less versed in the intricacies of Bayesian modelling) navigate the actual meaning of this modelling. In particular, I would suggest tools such as *prior predictive checks* (Vehtari et al., 2017; McElreath, 2018). In a nutshell, these procedures amount to generating out-of-sample data $\boldsymbol{y}^*$ (assumed exchangeable with the in-sample data $\boldsymbol{y}$) using the prior predictive distribution

$$p(\boldsymbol{y}^*) = \int p(\boldsymbol{y}^* \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$$

to evaluate the impact of the modelling assumptions in terms of the plausibility of the out-of-sample, in comparison to the actually observed data $\boldsymbol{y}$, thus allowing a deeper understanding of the practical meaning of the prior assumptions. Of course, this goes beyond the specifics of RBPB's model and it is more of a (relatively obvious) observation and suggestion of good practice from the Bayesian modelling.

I also think that exploration of alternative ways of including minimal prior information into the model, for instance through *Penalized Complexity* (PC) priors (Simpson et al., 2017), may be very helpful. PC priors are designed to regularize inference without forcing too strong information and tend to favor a "base" model (e.g. when a parameter is fixed to some given value), thus needing strong evidence from the data to depart from that basic assumption.

For example, consider the case of binary data y_s for a number of studies $s = 1, \ldots, S$, modelled as

$$y_s \sim \text{Binomial}(\pi_s, n_s)$$
$$\text{logit}(\pi_s) = \alpha + \gamma_s,$$

where $\alpha \sim \text{Normal}(0, v)$ is an overall baseline "prevalence" for the underlying outcome and $\gamma_s \sim \text{Normal}(0, \sigma_\gamma)$ is a study-specific random effect term, used to model correlation across the studies. We might consider a "base" model with $\sigma_\gamma = 0$ (which would then estimate the overall probability of occurrence using complete pooling across the studies) and a suitable PC prior is defined modelling $\sigma_\gamma \sim \text{Exponential}(\lambda)$, with the parameter $\lambda = -\dfrac{\log \alpha}{\sigma_0}$ fixed to impose that $\Pr(\sigma > \sigma_0) = \alpha$, for some arbitrarily defined constants σ_0 and α, used to control the level of information we are willing to include in the prior.

4. Comparability of studies

One of the key assumption in any meta-analytic modelling is about the (*alleged*) relative homogeneity of the underlying studies. Of course, this is pretty much a form of art as it is science, a cynic would perhaps say — in that, in my view, this is something about which we can often fail to be even remotely confident, but whose consequences may be dire, in terms of the model output. Once again, it is possible that in the area of application discussed by RBPB, this is much less of an issue and studies tend to be judged as reasonably "similar", which in part invalidates my argument and suggestion here.

In HTA, however, there is a growing body of literature exploring the issue of *population adjustment* (Signorovitch et al., 2012; Phillippo et al., 2018; Remiro-Azócar et al., 2021b, 2022). In the context of Network Meta-Analysis (NMA), we are often faced with a set of studies reporting data on a comparison of intervention A vs intervention C (e.g. the status quo), as well as a set of studies comparing a different active intervention B vs intervention C (the same reference).

Naïve comparisons that aim at making use of the common comparator C to obtain an indirect estimate of the "treatment" effect for A vs B may be extremely biased if the underlying populations observed in the different set of studies have broadly different characteristics; as an example, consider a study comparing a new drug A against aspirin C conducted in a very young and generally healthy population, while a separate study comparing an existing active treatment B against aspirin may select older and sicker patients to start with, leading to poor overlap in the baseline characteristics.

In HTA, it is now increasingly common to look at ways to "re-balance" and "transport" effects to map onto the target population of interest, e.g. using Bayesian hierarchical models

(Phillippo et al., 2018) or models based on parametric G-computation (Remiro-Azócar et al., 2022) to obtain an estimate of the relevant treatment effect (under different philosophies, leading to some discussion over the ideal setting, as discussed in Remiro-Azócar et al., 2021a).

Perhaps this aspect could also feed into a wider specification of RBPB's modelling structure, particularly if it can or should be expanded to questions best answered by network meta-analytic analyses.

5. Conclusions

I believe that RBPB's paper must be commended as a potentially very important contribution that can improve the way that studies on human genomic data are used for policy and decision-making, on the back of sound and extensive statistical analysis. As I have hopefully highlighted in my comments, there is perhaps a slightly different angle to consider, which embeds the results of the statistical modelling into a wider decision-analytic modelling structure, in which the pooled estimates are used to determine population dynamics and plan, say, for screening or vaccination programmes. Ultimately, I would very much welcome the combination of data and modelling in the field considered by RBPB with a more direct economic assessment, possibly extending all the way to considerations of the expected value of collecting additional information and research prioritisation (Jackson et al., 2022).

References

Dias, S., Ades, A. E., Welton, N. J., Jansen, J. P., and Sutton, A. J. (2018). *Network meta-analysis for decision-making.* John Wiley & Sons.

Jackson, C. H., Baio, G., Heath, A., Strong, M., Welton, N. J., and Wilson, E. C. (2022). Value of information analysis in models to inform health policy. *Annual review of statistics and its application* **9,** 95–118.

McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan.* Chapman and Hall/CRC.

Phillippo, D. M., Ades, A. E., Dias, S., Palmer, S., Abrams, K. R., and Welton, N. J. (2018). Methods for population-adjusted indirect comparisons in health technology appraisal. *Medical decision making* **38,** 200–211.

Remiro-Azócar, A., Heath, A., and Baio, G. (2021a). Conflating marginal and conditional treatment effects: Comments on "assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study". *Statistics in Medicine* **40,** 2753–2758.

Remiro-Azócar, A., Heath, A., and Baio, G. (2021b). Methods for population adjustment with limited access to individual patient data: a review and simulation study. *Research synthesis methods* **12,** 750–775.

Remiro-Azócar, A., Heath, A., and Baio, G. (2022). Parametric G-computation for compatible indirect treatment comparisons with limited individual patient data. *Research synthesis methods* **13,** 716–744.

Signorovitch, J. E., Sikirica, V., Erder, M. H., Xie, J., Lu, M., Hodgkins, P. S., Betts, K. A., and Wu, E. Q. (2012). Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value in Health* **15,** 940–947.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science* **32,** 1–28.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation*, volume 13. John Wiley & Sons.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing* **27,** 1413–1432.

Welton, N. J., Sutton, A. J., Cooper, N., Abrams, K. R., and Ades, A. (2012). *Evidence synthesis for decision making in healthcare*, volume 132. John Wiley & Sons.

Received Oct 2023