Elicitation of priors for intervention effects in educational trial data

Qing Zhang[1], Germaine Uwimpuhwe[1], Dimitris Vallis[12], Akansha Singh[1], Tahani Coolen-Maturi[1], Jochen Einbeck[1]

[1] Durham University, Durham, UK,
qing.zhang@durham.ac.uk, germaine.uwimpuhwe@durham.ac.uk,
akansha.singh@durham.ac.uk, tahani.maturi@durham.ac.uk,
jochen.einbeck@durham.ac.uk
[2] The Policy Institute, King's College London, UK
dimitris.vallis@kcl.ac.uk

Abstract. Effect sizes for educational interventions are commonly small, and hence decisions to re-grant efficacy trials (small trials with homogeneous populations under idealized conditions) as effectiveness trials (larger trials with heterogeneous populations) are often based on limited evidence from the efficacy trial itself. However, supplementary evidence may be available on how (past) effectiveness trials with similar outcomes tend to perform. This work proposes a Bayesian approach of making use of such evidence for re-granting decisions.

Keywords: Randomized controlled trial; Multilevel model; Meta-analysis

1 Introduction

We introduce a Bayesian approach to support re-granting decisions for efficacy trials. By harvesting prior information from past effectiveness trials, and feeding this information in form of a prior distribution on the intervention parameter into a multilevel model, we aim to gain principled evidence on the effect size which would likely have been observed if the efficacy trial had been be run under the less idealized conditions of an effectiveness trial. The Education Endowment Foundation (EEF) established in 2011 have commissioned more than 200 educational efficacy and effectiveness trials in England. Two case studies funded by the EEF have been considered, involving re-granting decisions for the 'Lexia® Core5® Reading' and 'Maths Count' efficacy trials, to which we refer in brief as "Lexia trial" and "Maths trial" henceforth. For each case study, a series of 'similar' EEF effectiveness trials has been identified, their effect sizes obtained, and combined through a fixed-effect meta-analysis. The combined effect size was then used to inform the prior distribution in a Bayesian multilevel model for the estimation of the posterior. These analyses demonstrate how incorporating prior information can lead to more precise estimates of intervention effects, which, de-

spite often resulting in smaller posterior effect sizes, lead to narrower confidence intervals[1] and a greater likelihood of identifying significant intervention effects.

2 Methodology

In accordance with the multi-site design of the efficacy trial to be re-granted (here, Lexia or Maths), a multilevel model can be defined following Singh et al. [2] as

$$Y_{ij} = \beta_0 + \beta_1 P_{ij} + \beta_2 T_{ij} + b_{0j} + b_{2j} T_{ij} + \epsilon_{ij}. \quad (1)$$

Here, Y_{ij} and P_{ij} represent the post-test and pre-test score for pupil i in school j, respectively, b_{0j}, b_{2j} school-specific (random) intercept and slope terms, and β_2 the intervention effect with prior distribution

$$\pi(\beta_2) = N(\mu, \sigma^2), \quad (2)$$

where typically μ and σ^2 need to be informed by a collection of K past effectiveness trials. A complication is that the outcome measures of the past effectiveness trials, and the efficacy trial at hand, may be different, so that the absolute scale of β_2 would be meaningless. This requires the outcome measure of the efficacy trial to be standardized (mean zero, unit variance). We consider two approaches to adjust the results from the K trials to this standardized scale. In the first approach, we re-fit a Bayesian multilevel model of type (1) with standardized outcomes to each trial $k = 1, \ldots, K$, yielding full posterior treatment distributions with mean μ_k and variance σ_k^2, which are then meta-analyzed to obtain $\mu = \frac{\sum_k w_k \mu_k}{\sum_k w_k}$, where $w_k = 1/\sigma_k^2$. In the second approach, we manually extract the intervention effects, in form of effect sizes, from the evaluation reports of the effectiveness trials, and then meta-analyze these effect sizes. The rationale for this approach is provided in Section 3.

3 Manual elicitation of priors from evaluation reports

For simplicity of presentation, we consider only a single past effectiveness trial, with estimated intervention effect $\hat{\beta}$ (obtained from a non-standardised analysis of the data) and standard error se($\hat{\beta}$). Denote by $\hat{\sigma}_T^2$ the estimated total variance (usually, unconditional) from that trial. This defines the effect size $ES = \hat{\beta}/\hat{\sigma}_T$, with standard error se(ES) (Hedges [1]). Effect size is a common measure used to estimate intervention effects in educational trials. These values will usually be provided in trial evaluation reports. Recall that the objective is to find a prior for the intervention parameter of the efficacy trial to be re-granted, denoted β_2 in (1). We denote this parameter now by β^* and assume that the analysis of

[1] This is technically a 'credible interval' but for ease of presentation we will use the term 'confidence interval' irrespective of whether it has been obtained in a Bayesian or frequentist way.

this efficacy trial will be carried out using standardised data. Accordingly, we denote by $\hat{\beta}^*$, ${\hat{\sigma}_T^*}^2$ and ES^* an estimate of the intervention effect, the total unconditional variance, and the effect size, for the efficacy trial. The task is to elicit suitable values of μ^* and σ^* which can be used to specify a prior distribution $\pi(\beta^*) = N(\mu^*, {\sigma^*}^2)$. The key insights are as follows: If the outcome variable is standardised, then the true total variance exactly equals 1 so that one will have ${\hat{\sigma}_T^*}^2 \approx 1$. Since effect size is a dimensionless concept it is reasonable to assert that $ES = ES^*$. Combining these, one has $\hat{\beta}^* \approx ES^* = ES$. Hence, the effect size estimate from the evaluation report can be used as the prior mean for the standardised Bayesian analysis of the efficacy trial. In order to specify the prior variance, we note that

$$\text{se}(ES^*) = \text{se}(\hat{\beta}^*/\widehat{\sigma}_T^*) = \sqrt{\text{var}(\hat{\beta}^*/\widehat{\sigma}_T^*)} \gtrsim \sqrt{\text{var}(\hat{\beta}^*)} = \text{se}(\hat{\beta}^*), \tag{3}$$

where the property at $\gtrsim$ follows from results in Goodman [3]. That is, the variance of $\hat{\beta}^*/\widehat{\sigma}_T^*$ will tend to be a bit larger than that of $\hat{\beta}^*$ (since the denominator, while approximately equal to 1, will still induce additional variation). Derivation (3) implies that $\text{se}(\hat{\beta}^*) \lesssim \text{se}(ES^*)$ and hence it is reasonable to use $\text{se}(ES^*)$ as an upper bound for $\text{se}(\hat{\beta}^*)$, to represent our prior uncertainty about $\hat{\beta}^*$. Finally, we follow analogous reasoning as above to assert that $\text{se}(ES) = \text{se}(ES^*)$. In summary, following the extraction of an effect size ES and its standard error, $\text{se}(ES)$, from an evaluation report, one has the remarkably simple result

$$\pi(\beta^*) = N(ES, \text{se}(ES)^2) \tag{4}$$

as the prior for our efficacy trial. While this will remain valid irrespectively of whether the past effectiveness trial has been using standardised or non-standardised data, it is important to recall that the analysis of the efficacy trial always needs to use standardised data. It also needs noting that $\text{se}(ES)$ will often not be given in the evaluator's report, but will instead need to be inferred from a reported 95% confidence interval for the effect size, $CI_{ES} = (\underline{ES}, \overline{ES})$, via $\text{se}(ES) = (\overline{ES} - \underline{ES})/(1.96 \times 2)$.

In some special situations one may not be able to use (4), e.g. because the evaluation reports do not provide effect sizes and standard errors. Apart from such rare cases, Equation (4) will generally be usable, and can equally be used for a meta-analyzed effect size arising from a collection of trials.

4 Results

Figure 1 displays the fixed-effect meta-analyses that were performed on all collected priors using inverse-variance weighting to generate the combined prior for both the manual and Bayesian methodologies within the Lexia and Maths efficacy trials. The priors are represented by squares, and the horizontal lines correspond to the 95% confidence intervals (CI). Larger squares indicate a greater weight in the meta-analysis. For the Lexia trial, the manual priors ranged from

-0.06 to 0.13 with a combined prior of 0.061. The heterogeneity among the studies was low ($I^2 = 2\%$), indicating a high level of consistency across studies. The Bayesian priors showed a slightly reduced combined prior of 0.031, with no observed heterogeneity ($I^2 = 0\%$). For the Maths trial, the manual priors varied more significantly, ranging from -0.01 to 0.27 with a combined prior of -0.000. Heterogeneity was substantial ($I^2 = 64\%$), reflecting variability in the study outcomes. Bayesian priors showed a combined prior of 0.036 with moderate heterogeneity ($I^2 = 56\%$).

Application of the methodologies to the two efficacy trials is provided in Tables 1 and 2. We see that the priors based on the two elicitation approaches are similar. Employing the priors obtained from re-fitted Bayesian model, one finds that despite reducing posterior means, the posterior precision has increased, indicating, in the case of Lexia, increased support towards a positive re-granting decision.

The application of meta-analyzed priors resulted in considerably shifted posterior estimates compared to those obtained using non-informative priors. This discrepancy elucidates the substantial influence that the choice of prior can exert on the posterior treatment effect within a Bayesian analytical framework. The employment of a non-informative prior, often representing a state of relative agnosticism regarding prior knowledge, yields a posterior distribution that is more influenced by the data from the current analysis. Conversely, informative priors, particularly those derived from meta-analytical methods, can markedly temper the posterior estimates, thus manifesting a more moderated effect size that is informed by a broader base of pre-existing evidence. Such moderation is observable in the decreased mean and standard deviation of the posterior estimates, suggesting that the informative priors contribute to a more nuanced and potentially more reliable estimation of treatment effects.

Table 1. Meta-analyzed priors from manual computation and resulting from refitted Bayesian models.

Trial	Priors	Meta-analysed ES	Prior	
		CI_{ES}	Mean	S. Dev.
Lexia	Manual	0.061	0.061	0.032
		$(-0.001, 0.124)$		
	Bayesian	0.031	0.031	0.030
		$(-0.028, 0.089)$		
Maths	Manual	-0.000	-0.000	0.010
		$(-0.019, 0.019)$		
	Bayesian	0.036	0.036	0.031
		$(-0.025, 0.097)$		

Table 2. Prior and posterior intervention means and standard deviations for the intervention parameter β_2, using non-informative priors, and priors resulting from re-fitted and meta-analyzed Bayesian model estimates.

Trial	Priors source	Prior		Posterior	
		Mean	S. Dev.	Mean	S. Dev.
Lexia	Non-Informative	0.000	100.000	0.078	0.066
	Meta-analysis	0.031	0.030	0.038	0.027
Maths	Non-Informative	0.000	100.000	0.203	0.130
	Meta-analysis	0.036	0.031	0.046	0.030

Acknowledgments: Special thanks to the EEF for funding this research. We would like to thank FFT for enabling access to EEF's data archive. This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

References

1. Hedges, L. V. (2007). Effect Sizes in Cluster-Randomized Designs. Journal of Educational and Behavioural Statistics, 32, 341–370.
2. Singh, A., Uwimpuhwe, G., Li, M., Einbeck, J., Higgins, S. and Kasim, A. (2021). Multi-site educational trials: estimating the effect size and its confidence intervals. International Journal of Research and Method in Education, 45, 18–38.
3. Goodman, L. A. (1962). The Variance of the Product of K Random Variables. Journal of the American Statistical Association, 57, 55–60.

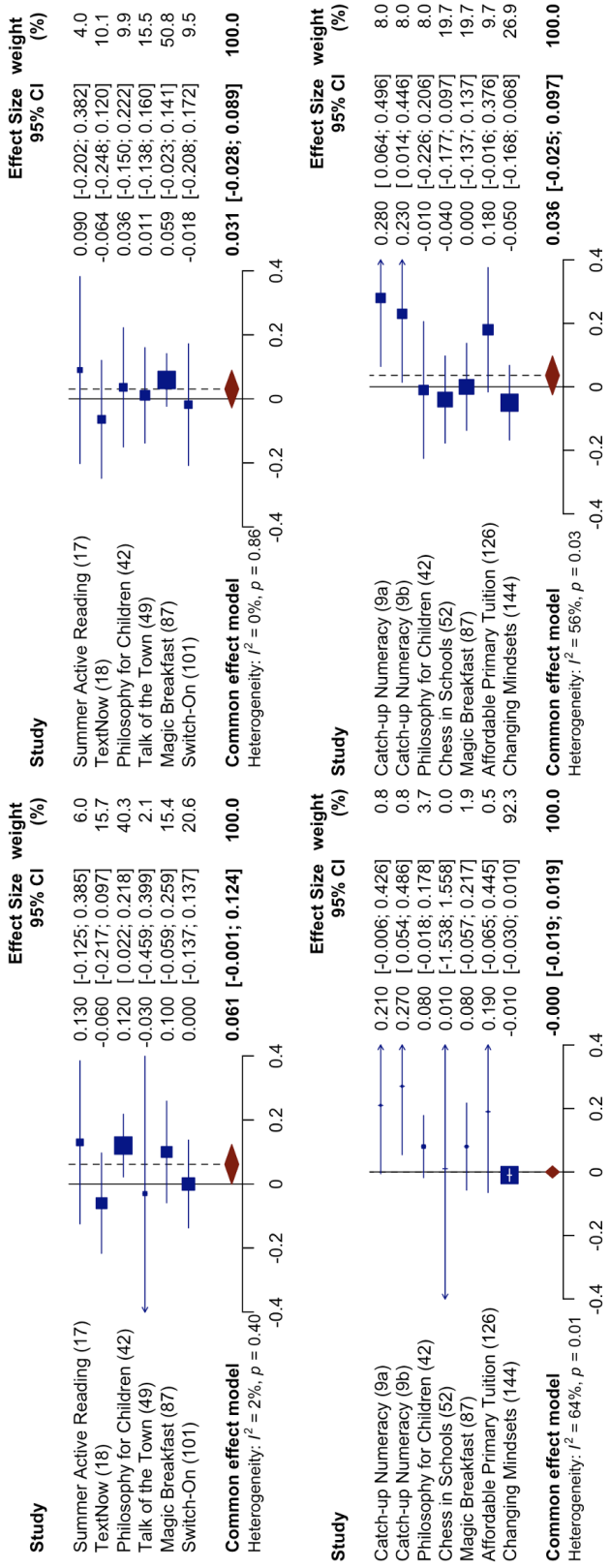

Fig. 1. Meta-analyses of manual versus Bayesian priors across two trials (Upper panel: Lexia trial; lower panel: Maths trial; left column: manual priors; right column: Bayesian priors)