**Summary of results for priors on beta**

Which priors result in the lowest/highest bias?

* Choice of prior not important if there is enough information in the categories of the outcome (i.e. symmetric)
* When skewed: all priors perform similarly but prior on the R-square seems to correct the bias.

How does bias change with different:

1. Number of categories: not much difference for symmetric; more important with skewed outcome probabilities (higher underestimation of treatment with higher categories regardless of effect size)
2. Sample size: smaller sample there is greater bias (results being driven more by the priors given lack of data/sparsity in some outcome categories)
3. Effect size: similar results across scenarios
4. Distribution of outcome probabilities: more bias for skewed (data sparsity)
5. Adaptive design: greater bias but likely due to the smaller sample size (more being driven by the prior) and the nature of adaptive trials with an early stopping rule. Symmetric probabilities tend to over estimate the treatment effect when there is one (though not necessarily an issue particularly if they want to stop early if there is a treatment effect and show there really is a beneficial effect)

Which priors result in the lowest/highest coverage?

Choice of prior not important based on coverage; though with 30 categories with skewed probabilities perhaps an R-squared prior would be best (under-coverage otherwise). Under-coverage for the adaptive design as well.

How does coverage change with different:

1. Number of categories:
2. Sample size:
3. Effect size:
4. Distribution of outcome probabilities:
5. Adaptive design:

Which priors result in the lowest/highest MSE?

Really not much difference with a larger sample size – MSE likely driven by variability in estimates given the small sample size. MSE smaller for symmetric but this was also due to the bias observed.

Which priors result in the lowest/highest mean posterior probabilities?

How do posterior probabilities change with different:

With smaller effect size, smaller posterior probability with higher number of categories Higher posterior probability for higher sample size. For skewed probabilities with small small sample size, best to use R-square prior otherwise no difference.

1. Number of categories:
2. Sample size:
3. Effect size:
4. Distribution of outcome probabilities:
5. Adaptive design:

Which priors result in the lowest/highest proportion declaring superiority?

How does this change with different:

For skewed probabilities with small and moderate effect sizes, may be better to use prior on R-square (with increasing number of categories especially). For symmetric probabilities the choice doesn’t appear to matter, though slightly higher proportion declaring superiority when the R-square is not used.

1. Number of categories:
2. Sample size:
3. Effect size:
4. Distribution of outcome probabilities:
5. Adaptive design:

Which priors result in the lowest/highest proportion of trials stopping early?

How does this change with different:

Again, R-square prior has higher proportion that stop early for skewed probabilities but best not to use this for symmetric. Those with large SD seem to have highest proportion of trials that stop early with large effect size. This proportion of trials stopping early decreases with increasing number of categories.

1. Number of categories:
2. Sample size:
3. Effect size:
4. Distribution of outcome probabilities:

**Summary of results for priors on alpha**

Which priors result in the lowest/highest bias?

* When the value of the hyperparameter decreases, bias goes towards zero for skewed data regardless of effect size, sample size. Normal distribution for the cut-points also results in small bias, except for 30 categories and small sample size where there is a slight over estimation.
* For symmetric: little difference between the choice in priors, though hyperparameters close to zero tend to underestimate with 30 categories.
* Adaptive design: over estimates treatment effect for U shape (not surprising): this is larger for increasing number of categories and effect size for U shape. Choice does not differ for U shape for larger sample sizes, though for small sample sizes perhaps hyperparameter closer to zero is more appropriate. For skewed, choosing a hyperparameter below one (but not too close to zero) brings the bias closer to zero across categories.

Which priors result in the lowest/highest coverage?

* Majority of priors are close to nominal coverage level, apart from 1/J and 0 hyperparameters, particularly for when there are 30 categories. Nominal coverage decreases

Which priors result in the lowest/highest proportion declaring superiority?

For skewed: Dirichlet with 1 or 0.5 less likely to declare superiority compared to 1/J or zero (that perform similarly). Normal performs better than 1/0.5. For U-shape, results are quite similar for a fixed design: for an adaptive design, using hyperparameters closer to zero for larger number of categories (30) would be sufficient.

Which priors result in the lowest/highest proportion of trials stopping early?

Similar to above: for skewed, priors with hyperparameters closer to zero more effective for stopping the trial early when there is an effect (though be cautious as more likely to stop when there is no effect too, even with small sample sizes). Same for U-shaped except for large effect/large sample size.

How does this change with different:

Again, R-square prior has higher proportion that stop early for skewed probabilities but best not to use this for symmetric. Those with large SD seem to have highest proportion of trials that stop early with large effect size. This proportion of trials stopping early decreases with increasing number of categories.

1. Number of categories:
2. Sample size:
3. Effect size:
4. Distribution of outcome probabilities: