

SIMULATION STUDY

Evaluating the effect of different non-informative prior specifications on the Bayesian proportional odds model in randomised controlled trials: a simulation study

Chris J. Selman^{1,2*}, Katherine J. Lee^{1,2}, Michael Dymock^{3,4}, Ian C. Marschner⁵, Steven Y.C. Tong^{6,7}, Mark Jones^{3,8} and Robert K. Mahar^{1,9,10}

*Correspondence:
chris.selman@unimelb.edu.au

¹Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Parkville, VIC, 3052, Australia

²Department of Paediatrics, University of Melbourne, Parkville, VIC, 3052, Australia

Full list of author information is available at the end of the article

Abstract

Background: Ordinal outcomes combine multiple distinct ordered patient states into a single endpoint and are commonly analysed using proportional odds (PO) models in clinical trials. When using a Bayesian approach, it is not obvious what the influence of a 'non-informative' prior is in the analysis of a fixed design or on early stopping decisions in adaptive designs.

Methods: This study compares different non-informative prior specifications for the Bayesian PO model in the context of both a two-arm trial with a fixed design and an adaptive design with an early stopping rule. We conducted an extensive simulation study, varying the effect size, sample size, number of categories and distribution of the control arm probabilities. The effect of the prior specifications is also compared using data from the Australian COVID-19 Trial.

Results: Our findings indicate that the choice of prior specification can introduce bias in the estimation of the treatment effect, particularly when control arm probabilities are right-skewed. The R-square prior specification resulted in the smallest bias and increased the likelihood of appropriately stopping early when there was a treatment effect. However, this specification exhibited larger biases when control arm probabilities were U-shaped when there was an early stopping rule. Dirichlet priors with concentration parameters close to zero resulted in the smallest bias when probabilities were right-skewed in the control arm, but were more likely to inappropriately stop early for superiority when there was no treatment effect and an early stopping rule.

Conclusions: The specification of non-informative priors in Bayesian adaptive trials with ordinal outcomes has implications for treatment effect estimation and early stopping decisions. We recommend the careful selection of priors that consider the possible distribution of control arm probabilities and that sensitivity analyses to the prior be conducted.

Keywords: ordinal outcome; proportional odds model; adaptive trial; non-informative priors; randomised controlled trials; simulation study; Bayesian

Introduction

Randomised controlled trials (RCTs) aim to answer causal questions by comparing outcomes from participants randomised between a control treatment and at least one intervention. One type of outcome that is becoming increasingly popular in trials is an ordinal outcome [1]. Ordinal outcomes combine multiple distinct ordered patient states into a single endpoint. The categories that make up the ordinal outcome must be monotonically ordered but there may not necessarily be an equal 'distance' between proximate categories. Although ordinal outcomes may be less familiar to researchers than other outcome types, they are commonly used in many medical settings. For example, the modified-Rankin scale [2] is typically used to measure outcomes in patients who suffer from stroke or traumatic brain injury with categories ranging from 0 (no symptoms) and 6 (dead), with increasing severity of disability in between. Ordinal outcomes became particularly prominent during the initial stages of the COVID-19 pandemic, when the World Health Organization developed the Clinical Progression Scale to measure distinguished health states for patients with COVID-19 [3] that was anchored at asymptomatic and uninfected, and death. Despite being relatively common, there is still a lack of understanding of the interpretation and appropriate analyses of ordinal outcomes in RCTs [1].

There are many approaches available to analyse ordinal outcomes in RCTs. The most common is a proportional odds (PO) model, which estimates a proportional (or common) odds ratio (OR) across each binary split (or 'cut-point') of the ordinal scale [1, 4–6]. The proportional OR represents the size of the shift in the distribution of the ordinal scale towards a favourable (or unfavourable) outcome for an intervention relative to the comparator, which can be used to summarise the efficacy of an intervention.

Alongside the increased use of ordinal outcomes, the COVID-19 pandemic also necessitated efficient trial designs that could expedite treatment decisions [7–9]. This included the use of adaptive designs that allow prespecified adaptations or changes to the design based on the accruing data from the trial. Some examples of adaptations that can be incorporated include early stopping, sample size re-estimation, response adaptive randomisation, subgroup selection, or some combination thereof [10].

Early stopping rules, particularly for superiority, are commonly used in trials that use ordinal outcomes [1]. Early stopping for superiority allows a trial to be stopped following a pre-specified interim analysis, if there is sufficient evidence that an intervention is superior to the comparator (or sometimes, other interventions), avoiding future patients being allocated to inferior treatments. A trial with such stopping rules can reduce the anticipated duration and sample size compared to a trial with a fixed design [11].

Bayesian methods are commonly used for analysing adaptive trials that allow for early stopping since they are naturally suited to multiple looks at the data, complicated models can be fit fairly easily, and prior distributions provide a principled way of updating inferences at each interim analysis [12–14]. In Bayesian inference,

the data are combined with prior distributions or beliefs about the parameters of interest to obtain a posterior distribution for the parameters [15]. The choice of prior distribution has a direct influence on the posterior distribution, particularly if an informative prior is chosen or the sample size is small, which can in turn affect the final estimate of the treatment effect. In an adaptive trial, the decision to stop a trial early for superiority is typically guided by the posterior probability that the intervention is superior to one or more comparators which can again be affected by the choice of prior. For example, if an informative prior strongly favours the intervention over the comparator, early stopping decisions may be (possibly inappropriately) accelerated which would increase the risk of stopping for success even if the intervention is not superior.

Non-informative priors are often used in the analysis of trials to reflect clinical equipoise and to allow the inference to be guided by the data. There are, however, numerous non-informative priors that can be used for the treatment effect and any other parameters in the model, and it is often unclear how ‘non-informative’ these priors may be for a complex analysis model, particularly in the analysis of ordinal outcomes. For example, although a diffuse prior centered around no treatment effect, such as a $\text{Normal}(0, \sigma)$ prior distribution with large σ , may be a natural choice for the treatment effect parameter, the influence on the posterior distribution is not well understood in the context of a PO model. For large σ , the inverse-logit transformation of the prior distribution pushes the mass towards the lower and upper bounds, creating a ‘bathtub’ like distribution. Instead, specifying a smaller σ may be more appropriate. Furthermore, PO models for ordinal outcomes use ordered cut-point parameters to distinguish between consecutive (outcome) categories that represent cumulative log-odds for the control group which also require prior distributions. There has been some methodological research surrounding the choice of priors however, this research has primarily focused on the choice for the prior when the outcome is continuous. In the context of PO models, it has been suggested that for the model cut-points a prior should be chosen that preserves the parameter ordering, such as a Dirichlet distribution [16, 17] or Normal distribution. However the impact that different non-informative priors have on the estimation of the treatment effect has not been explored in trial settings. Given the increased use of ordinal outcomes in trials that use Bayesian PO models and the numerous possible choices of prior distributions available, it is imperative to understand the impact that non-informative priors have on the posterior distribution for the treatment effect and whether they are truly ‘non-informative’, particularly in trials with interim analyses where the sample size may be small.

In this paper, we aim to evaluate the effect of different non-informative prior specifications on the estimation of the treatment effect from a Bayesian PO model. We focus on the context of a two-arm trial with 1:1 ratio of allocation that uses an ordinal outcome under 1) a fixed design that has a single analysis at the end of the study and, 2) an adaptive design with an early stopping rule for superiority at a single interim analysis. This is investigated using a simulation study and a case study of the Australasian COVID-19 Trial (ASCOT) trial data. As part of the simulation study we assess the impact of the following factors on the estimation of

the treatment effect and posterior probability of intervention superiority where the target estimand is a proportional OR:

- 1 Magnitude and direction of the proportional OR.
- 2 Trial sample size.
- 3 Number of categories of the ordinal outcome.
- 4 Distribution of the control arm outcome probabilities.

This paper proceeds as follows: we first define key notation, provide an overview of the PO model and the estimand of interest, before outlining candidate non-informative prior distributions. Next we describe and present the results of the simulation study followed by the results of the ASCOT case study [18]. We conclude by summarising the key results and provide guidance for specifying priors for Bayesian PO models.

Methods

Notation

The target estimand is the proportional OR that represents the relative increase in odds of a more favourable outcome for an intervention compared to a control arm. We start by outlining our notation:

- J is the number of categories in the ordinal outcome;
- β is the proportional log-OR;
- $b \in \{c, t\}$ represents the control arm c and intervention arm t ;
- π_b is a vector of ordinal outcome category probabilities for participants receiving arm b . That is:

$$\pi_b = (\pi_{b1}, \pi_{b2}, \dots, \pi_{bJ})^T \quad (1)$$

- $\alpha = (\alpha_2, \dots, \alpha_J)$ is a vector of model intercepts representing the cumulative log-odds for the control group;
- $a = 1, 2$ indicate the interim and final analysis times;
- \mathbf{Y}_{ab} is the vector of observed ordinal outcome for participants $n = 1, 2, \dots, n_a$ allocated to arm b at analysis a .
- $x_i \in \{0, 1\}$ is an indicator denoting assignment to intervention, where 0 = control and 1 = intervention.

Proportional odds model

The PO model can be used to estimate a treatment effect for an ordinal outcome with more than two categories. The model estimates a proportional log-OR (β) that is assumed to be constant across all binary splits of the ordinal scale, where for $j = 2, 3, \dots, J$:

$$\log \frac{P(Y_{ab} \geq j)}{P(Y_{ab} < j)} = \text{logit}[P(Y_{ab} \geq j)] = \alpha_j + \beta \mathbf{x} \quad (2)$$

When using this model for Bayesian inference, a prior must be specified for the baseline cumulative log-odds α and for the proportional log-OR β . We outline some options of possible priors for these parameters in a PO model in the following section.

Specification of non-informative priors for β

A non-informative prior on the treatment effect parameter β is usually specified to be symmetric around no effect (i.e., zero) with sufficient variability to encapsulate the appropriate range of potential effects if a non-informative prior is desired [19–21]. One common choice is a Normal distribution with a large variance parameter [1]. Although this prior appears to be a reasonable choice on the log-odds scale, when transforming the distribution using the inverse-logit transformation, the mass is pushed towards the bounds at zero and one, creating a ‘bathtub’ like distribution. Instead, using a smaller variance parameter, although appearing more informative, may still be rather non-informative. Other symmetric and potentially non-informative prior distributions that could be specified for the treatment effect include the Student- t distribution [21] (which approaches the shape of a Normal distribution when the degrees of freedom becomes large but has fatter tails with smaller degrees of freedom to reflect a wider range of prior beliefs), a Cauchy distribution (a t -distribution with one degree of freedom), or a Laplace distribution. The Laplace distribution is also known as the double-exponential distribution and is symmetric with a sharp peak at its centre and can be motivated as a scale mixture of Normal distributions. This distribution has fairly flat tails to mimic a non-informative prior when an appropriate value is specified for the standard deviation [22].

An alternative prior specification for the treatment effect parameter β is to put a prior on the beliefs about the location of the R^2 [23] (the proportion of variance in the ordinal outcome that is explained by the intervention) assuming a continuous latent variable specification for the ordinal outcome. This involves putting a prior on the location of R^2 , which can be achieved using a Beta distribution where the first shape hyperparameter is set to equal to half the number of predictors in the model [23] and the second shape hyperparameter is greater than zero. When both shape hyperparameters are equal (in this case, also 0.5), the prior mean and median of the R^2 is 0.5 and the distribution is symmetric. Like the normal prior with a large variance, this specification pushes most of the probability density toward zero and one. Choosing a value of the second shape hyperparameter that is not close to zero or one would be a reasonable diffuse prior since the prior belief does not hinge on whether the intervention has a strong effect on the outcome (if close to one), and whether there is high certainty that the intervention has no effect on the outcome (if close to zero). It can be unclear how to specify a prior value for the R^2 in a PO model as R^2 refers to the proportion of variance in the underlying continuous latent variable that is attributable to the predictors under a linear model, and typically, the R^2 is lower in a PO model compared to a linear model with a continuous outcome. However, the smaller the R^2 , the smaller the prior correlations among the ordinal outcome and intervention variable(s) are, and the more concentrated near zero the prior distribution is for the regression coefficient (i.e. the proportional log-OR).

In the current study, we compare the estimation performance of the PO model using the following priors on β :

- 1 $\beta \sim N(\mu = 0, \sigma^2 = 100^2)$
- 2 $\beta \sim N(\mu = 0, \sigma^2 = 2.5^2)$
- 3 $\beta \sim t(df = 1)$ (equivalent to a Cauchy distribution)
- 4 $\beta \sim \text{Laplace}(\mu = 0)$, with SD = 100
- 5 $\beta \sim \text{Laplace}(\mu = 0)$, with SD = 2.5
- 6 Mean $R^2 = 0.5$, equivalent to a Beta(0.5, 0.5) distribution

Specification of non-informative priors for α

Although we would expect the prior used for the treatment effect to have the largest effect on the posterior distribution for the treatment effect, the prior used for the cut-points may also be important. The Dirichlet distribution is a natural choice for the (implicit) prior on the model cut-points because the concentration parameters can be interpreted as prior counts, i.e. prior probabilities for each of the J categories at the covariate means [19]. For instance, if $\alpha_j = 1$ for all $j \in \{1, 2, \dots, J\}$, the Dirichlet prior is uniformly distributed across the categories. This implies a prior belief of one observation in each of the J ordinal categories when the predictors are at their sample means, indicating very weak prior knowledge that no category has probability zero. The Dirichlet prior distribution may be appropriate when modeling ordinal outcomes given its multivariate generalisation of the Beta distribution and that it is the conjugate prior of the categorical and multinomial distribution [24]. An alternative choice is Jeffrey's fixed reference prior where $\alpha_j = 0.5$ for all $j \in \{1, 2, \dots, J\}$ [17]. This prior maximises the 'distance' (based on the Kullback–Leibler divergence, a measure of how one probability distribution is different to another probability distribution) between the prior and the posterior, implying that the data has the largest possible effect on the posterior (and that the prior has the least possible effect on the posterior). An alternative is to specify a normal distribution for each cut-point that preserves the ordering in the model.

In this paper we investigate the following priors for the intercept terms:

- 1 Uniform Dirichlet: $\alpha \sim Dir(\mathbf{1})$ (sets of baseline risk are equally likely for each category).
- 2 Multivariate Jeffrey's reference prior for the cut-points: $\alpha \sim Dir(\mathbf{0.5})$.
- 3 $\alpha \sim Dir(\mathbf{0.001})$.
- 4 $\alpha \sim Dir(\frac{1}{J})$ (objective reciprocal prior specification)
- 5 $\alpha_j \sim N(0, 100^2)$

Simulation study

We conducted a simulation study to evaluate the impact that different non-informative priors for α and β have on the estimation of the treatment effect under a range of distributional, sample size, trial design, and effect size scenarios. In this

section, we outline the data generating mechanisms, analysis methods, and performance measures used in the simulation study.

Data-generating mechanisms

The data generation process assumes a true underlying proportional OR in each trial. We set up a hypothetical scenario where we emulate a two-arm trial with a fixed design, and another with a single interim analysis once half of the outcome data is available with the potential to stop early if the intervention is declared superior. Intervention assignment was generated using $X_i \sim \text{Bern}(0.5)$ - representing simple randomisation with an equal allocation ratio. The ordinal outcome data was then simulated by taking a random sample from a multinomial distribution for the control and intervention group respectively given the respective probabilities as described below. We restrict our focus to ordinal outcomes with $J \in \{4, 10, 30\}$ categories to reflect the wide range of ordinal outcomes that are used in practice.

We varied the distribution of the probabilities in each outcome category in the control group to reflect real-world scenarios. First, we set the probabilities in each outcome category to follow a skewed relationship as observed in the distribution of two ordinal outcomes that were both days alive and free of hospitalisation/ventilation 28-days post randomisation respectively in the ASCOT trial. Here, the probabilities were primarily concentrated to categories at the beginning of the outcome scale, exhibiting positive skewness across the distribution of the categories. Second, we defined probabilities in each outcome category to follow a U-shape distribution which is commonly observed in stroke trials that use the modified-Rankin scale where patients tends to fall in either the first two or three categories and the last two categories (depending upon the population), as was observed in the ESCAPE-NA1 trial [25]. We set probabilities in each category of the ordinal outcome in the control group by partitioning the sample space of a $\theta \sim \text{Beta}(\alpha, \beta)$ distribution into J equal partitions where the control probabilities for each category were set to equal the cumulative probability in each partition. Finally, we considered a scenario where the probabilities in each outcome category were equal.

Next we generated the cumulative log-odds for the intervention group assuming proportional odds, i.e. $\text{logit}(P(Y \geq j)) = \alpha_j + \beta$ where we explored an effect size of $\log(\text{OR}) = \log(1)$ (no effect), $\log(1.10)$ (small effect) and $\log(1.50)$ (moderate effect size) to reflect similar proportional ORs to that observed in the ASCOT trial. The intervention outcome probabilities π_t were then calculated using the inverse logit function. These probabilities along with the control group probabilities were used to generate the ordinal outcome.

Data was simulated for a total sample size of $n_{obs} \in \{100, 500\}$ records for each scenario under consideration. For simplicity, we assumed that the outcome is observed immediately. An interim analysis occurred at an information fraction of one-half, i.e. at a sample size of 50 or 250. The stopping rule was guided by the Bayesian posterior probability that the intervention was superior to the control arm. Specifically, superiority of an intervention was assessed using the posterior probability that

the proportional log-OR is greater than zero exceeds 0.99 (i.e. $P(\beta > 0) > 0.99$) at the interim analysis to align with more stringent stopping rules at an interim, then superiority was triggered and the trial was stopped early at the interim analysis. If the trial continued to the maximum sample size, superiority was declared if $P(\beta > 0) > 0.95$. For the fixed design, superiority was declared if $P(\beta > 0) > 0.95$. We note that the two designs are not comparable due to different Type I error rates, although the objective of this study is not to compare the designs, rather the impact of using different priors on the treatment effect given the study design. In total, 108 scenarios were considered in this simulation study.

Analysis methods

We estimated the parameters of the PO model (Equation 2) using Hamiltonian Monte Carlo (HMC) sampling, a Markov Chain Monte Carlo (MCMC)-sampling-based algorithm, to obtain the posterior distribution for the treatment effect using the different prior specifications for α and β outlined above. We initially assess the impact of changing the prior for the treatment effect β for a fixed prior for α ($\alpha \sim Dir(1)$), and then assess the impact of changing the prior for the cut-points but holding the prior for β fixed $\beta \sim N(0, 100^2)$. The median of the posterior distribution for β was used to determine performance metrics. The HMC method was implemented in the R programming environment [26] and Stan [27], the latter of which is a probabilistic programming language that specifies statistical models. The R packages *posterior* [28] and *rstanarm* [29] were used to summarise posterior distributions.

An iterative process was used to determine the number of simulations, n_{sim} for each scenario, to ensure that the required maximum tolerable upper bound of the Monte Carlo standard errors (MCSEs) for each performance measure was less than 0.05 [30]. We calculated the jackknife-after-bootstrap MCSE for each parameter using the method described in Koehler et al [31]. All scenarios achieved the upper bound when $n_{sim} = 1000$.

A burn-in of 5,000 iterations and a chain length of 10,000 samples per chain were used across four chains, with no thinning, resulting in a total of 20,000 post-warmup samples from the posterior distribution for each scenario. To evaluate the reliability of the posterior distributions, convergence diagnostics were recorded, including the effective sample size (ESS), the potential-scale reduction factor (\hat{R}), and the number of divergent transitions, which signal difficulties in properly exploring the parameter space. Divergent transitions, which occur when the HMC sampler encounters extreme curvature in the target distribution, were minimized or eliminated by adjusting the average target acceptance probability rate and increasing the maximum tree depth. If divergent transitions were present, a sensitivity analysis that removed the analysis method where such transitions occurred was performed to assess the robustness of the conclusions.

Performance measures

The performance measures of interest were the overall bias, relative bias, coverage and mean-squared error of the proportional log-OR, relative to the value used in the data generation [32]. Relative bias was calculated on the OR scale. For each scenario, we also estimated the posterior probability that the intervention was superior (whether this be at the interim or maximum sample size), the proportion of trials that declared superiority, and among the trials that incorporated an early stopping rule, the proportion of trials that stopped early.

Simulation study results

In this section, we summarise the results from the simulation study. There was a similar pattern of results for the fixed and adaptive designs so we focus on the latter and present the results for the fixed design in the supplementary material, unless where patterns were different.

Specification of priors for β

There was substantial bias in the estimated treatment effect when the outcome distribution in the control arm was right-skewed for all prior distributions apart from a prior specification on the R-squared, with the relative bias increasing for increasing number of categories (see Figure 1 and Figure S2 in Supplementary Material 1). However, the bias in the treatment effect was generally minimal irrespective of the prior specification when the control probabilities were U-shaped or uniform for the fixed design. The exception was when there was moderate effect size and sample size was small for the Laplace (small SD), Cauchy and R-square prior specifications when there was a positive bias, particularly for the adaptive designs. Coverage showed a similar pattern being close to nominal coverage level (95%) when the control probabilities were U-shaped or uniform, but with under-coverage when they were skewed and there were 30 categories. This was the case for both fixed and the adaptive design. Looking at the average posterior probability that the intervention was superior (see Figures S9 and S10 in Supplementary Material 1), it appears that with smaller effect size, the posterior probability decreased with higher number of categories in the ordinal outcome for all priors. In particular, when there were skewed probabilities in the control arm and a small sample size, it appears that the posterior probability was slightly higher when a prior on the R-squared was used compared with the alternative priors, which was the case for both the fixed and adaptive designs. When the control arm probabilities were U-shaped, the proportion of trials that declared superiority did not appear to differ substantially with the choice of prior, although there was a slightly higher proportion of trials declaring superiority with the Normal, Laplace and Cauchy prior compared to when a prior specified using the R-squared was used. Similar results were observed for the fixed design.

A higher proportion of trials with an adaptive design declared superiority when the R-squared prior was used with skewed control arm probabilities when there was a

small and moderate effect size, particularly with increasing number of categories in the ordinal outcome (Figure 2c). When there were skewed control arm probabilities, the R-squared prior results in the highest proportion of trials that would have stopped early, but when the control arm probabilities were U-shaped the R-squared prior resulted in the fewest trials stopping early. In the latter context, specifying Normal or Laplace priors with large standard deviations had the highest proportion of trials that stopped early particularly with the larger effect sizes. It appears that trials are less likely to stop early when the ordinal outcome is right-skewed in the control arm and has a larger number of categories, regardless of the prior specified for the treatment effect, with the exception of when the prior is specified on the R-squared.

Of note, the posterior distributions for all target parameters of interest across all scenarios appeared to converge to their stationary distribution, with $\hat{R} < 1.01$ for all priors and all scenarios. The bulk and tail effective sample sizes were all above the recommendation of 100 per chain [33, 34], indicating good mixing and low autocorrelation. The MCSE of all performance measures were less than 0.05 and are reported in the supplementary material for each prior specification and each scenario (Supplementary Material 2). There were no analyses where there were divergent transitions.

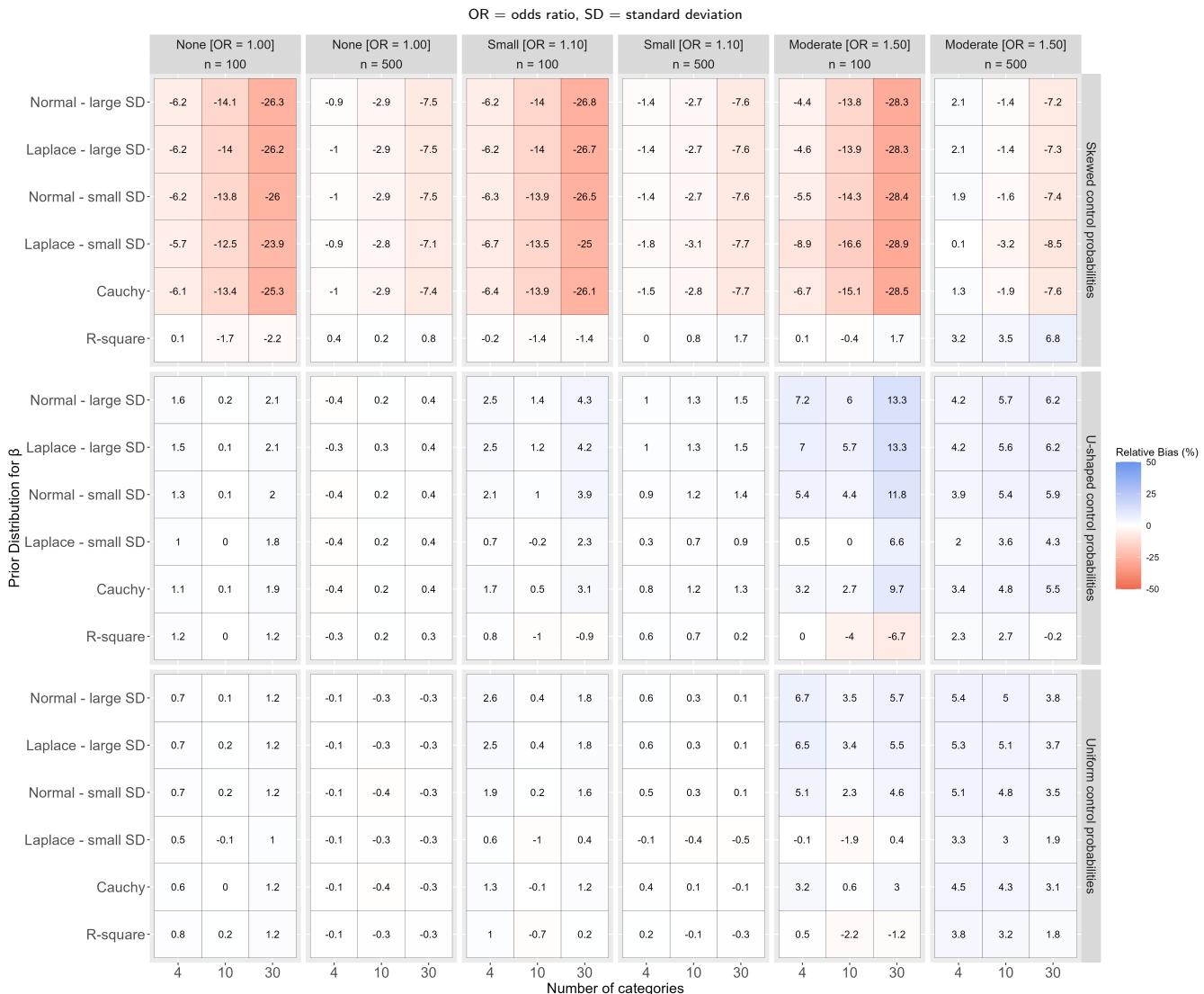
Figure 1a Relative bias in the odds ratio for the various scenarios with an adaptive design when varying the prior for the treatment effect

Figure 1b Coverage of the 95% credible interval for the log-odds ratio for the various scenarios with an adaptive design when varying the prior for the treatment effect

OR = odds ratio, SD = standard deviation

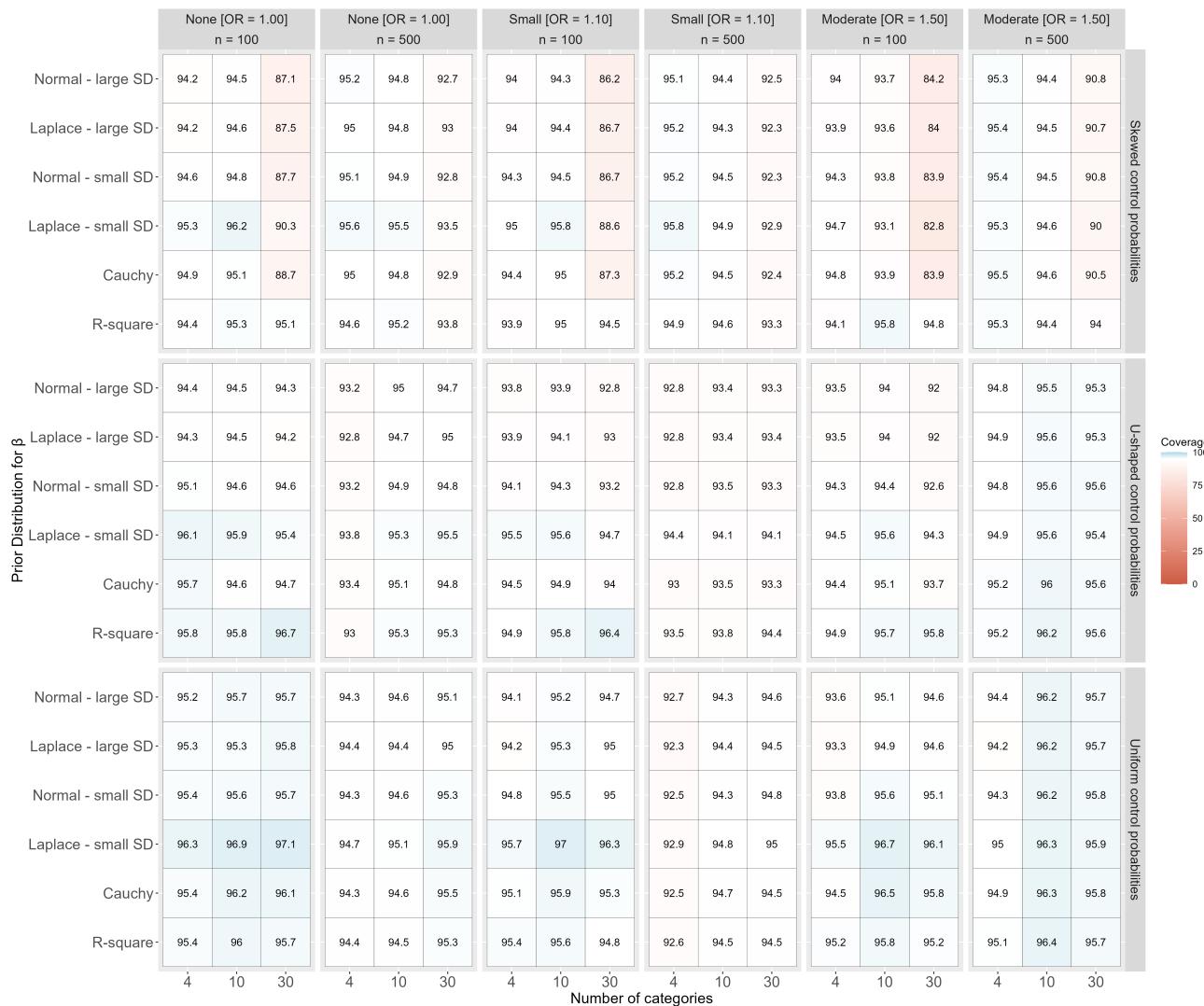
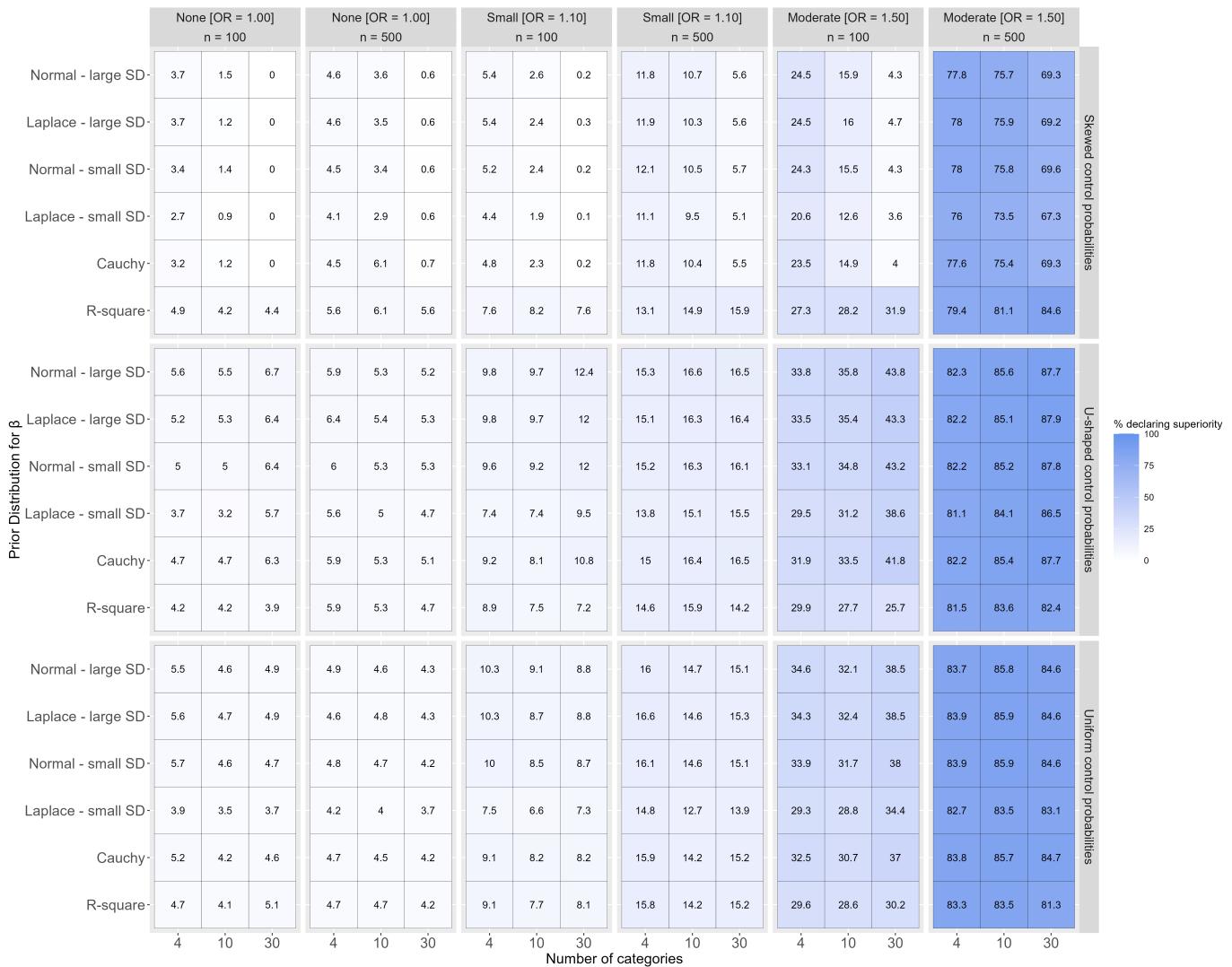


Figure 1c Proportion of trials declaring superiority in an adaptive design when varying the prior for the treatment effect

OR = odds ratio, SD = standard deviation



Specification of implicit priors for α

Keeping the prior for the treatment effect to be fixed at $\beta \sim N(0, 100)$, as the value of the concentration parameters of the Dirichlet distribution decreased, the relative bias approached zero when the control arm probabilities were right-skewed, regardless of the effect or sample size, for a fixed design (Supplementary Material - Figure S16). In contrast, there was minimal relative bias when the control arm probabilities were right-skewed when a Normal distribution was specified for the cut-points, except when there was a small sample size and the ordinal outcome had 30 categories where there was an overestimation of the treatment effect (relative bias 9% and 11% for skewed and U-shaped probabilities respectively). When the control arm probabilities were U-shaped or uniform, there was minimal bias in the treatment effect for all of the prior specifications except when there were 30 categories and the sample size was small.

When the design was adaptive, there was a tendency to overestimate the treatment effect with all priors for all scenarios (Figure 3a). This relative bias increased with the number of categories and effect size. The results were similar irrespective of the choice of concentration parameter with larger sample sizes, however, for smaller sample sizes, the priors with a concentration parameter closer to zero resulted in smaller bias for the treatment effect. When the control arm probabilities were right skewed, specifying a Dirichlet using $\alpha = 0.5$ or $\alpha = \frac{1}{J}$ resulted in lower bias than the other priors regardless of the number of categories in the outcome. Coverage was close to the nominal level of 95% for the majority of scenarios (Figure 3b), with the exception of Dirichlet priors where the concentration parameters were closest to zero which resulted in under-coverage.

When the control arm probabilities were skewed, adaptive designs were less likely to declare superiority and therefore less likely to stop early (Figure 3c) when a Dirichlet prior with $\alpha = 1$ or $\alpha = 0.5$ was used compared to other priors. Importantly, trials were more likely to stop early even when there was no treatment effect when using the Dirichlet prior with a small concentration parameter, with a probability of up to a 18% chance of stopping early in some scenarios.

Figure 2a Relative bias in the odds ratio for various scenarios with an adaptive design when varying the implicit prior on the cut-points

OR = odds ratio, SD = standard deviation

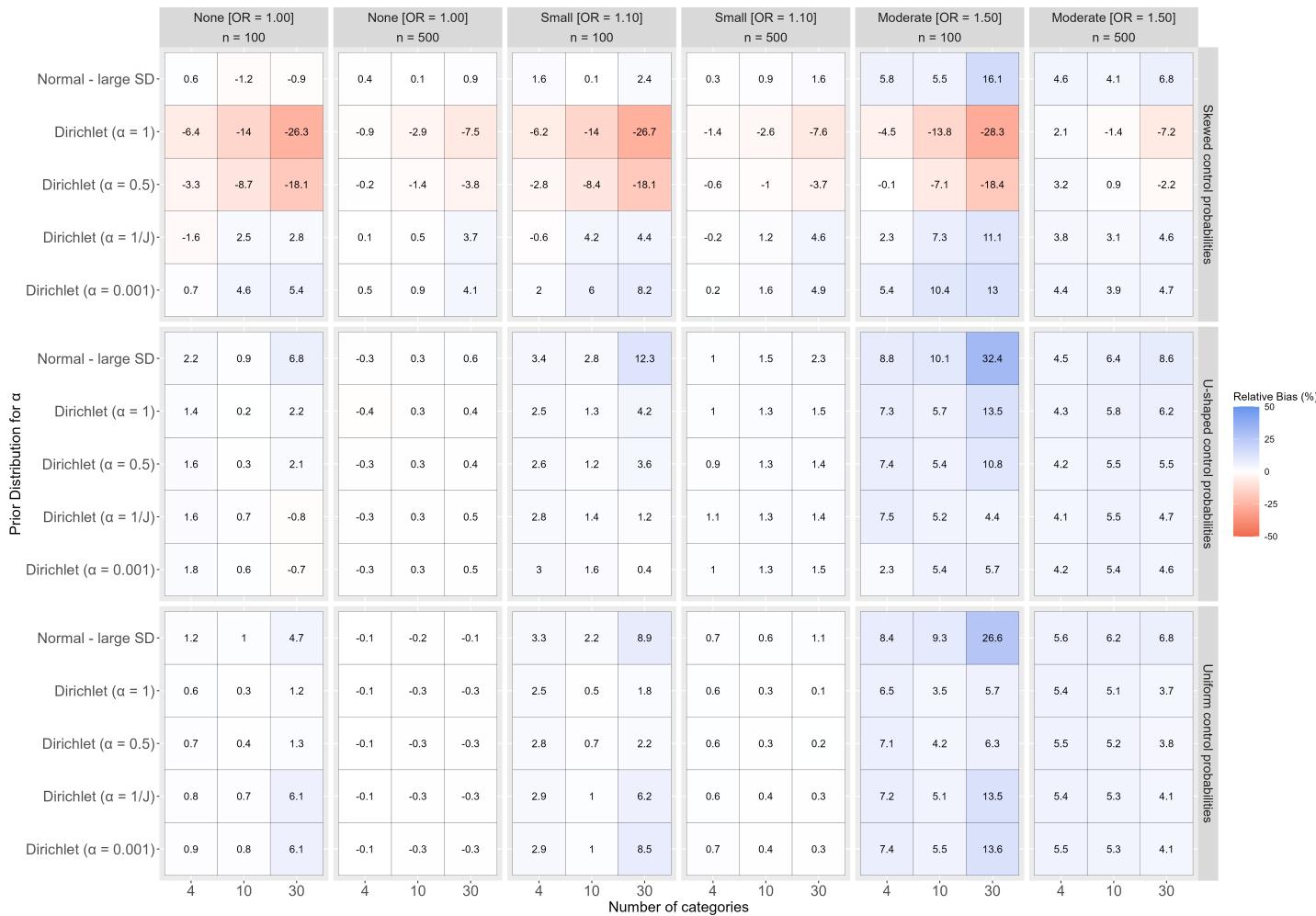


Figure 2b Coverage in the log-odds ratio for various scenarios with an adaptive design when varying the implicit prior on the cut-points

OR = odds ratio, SD = standard deviation

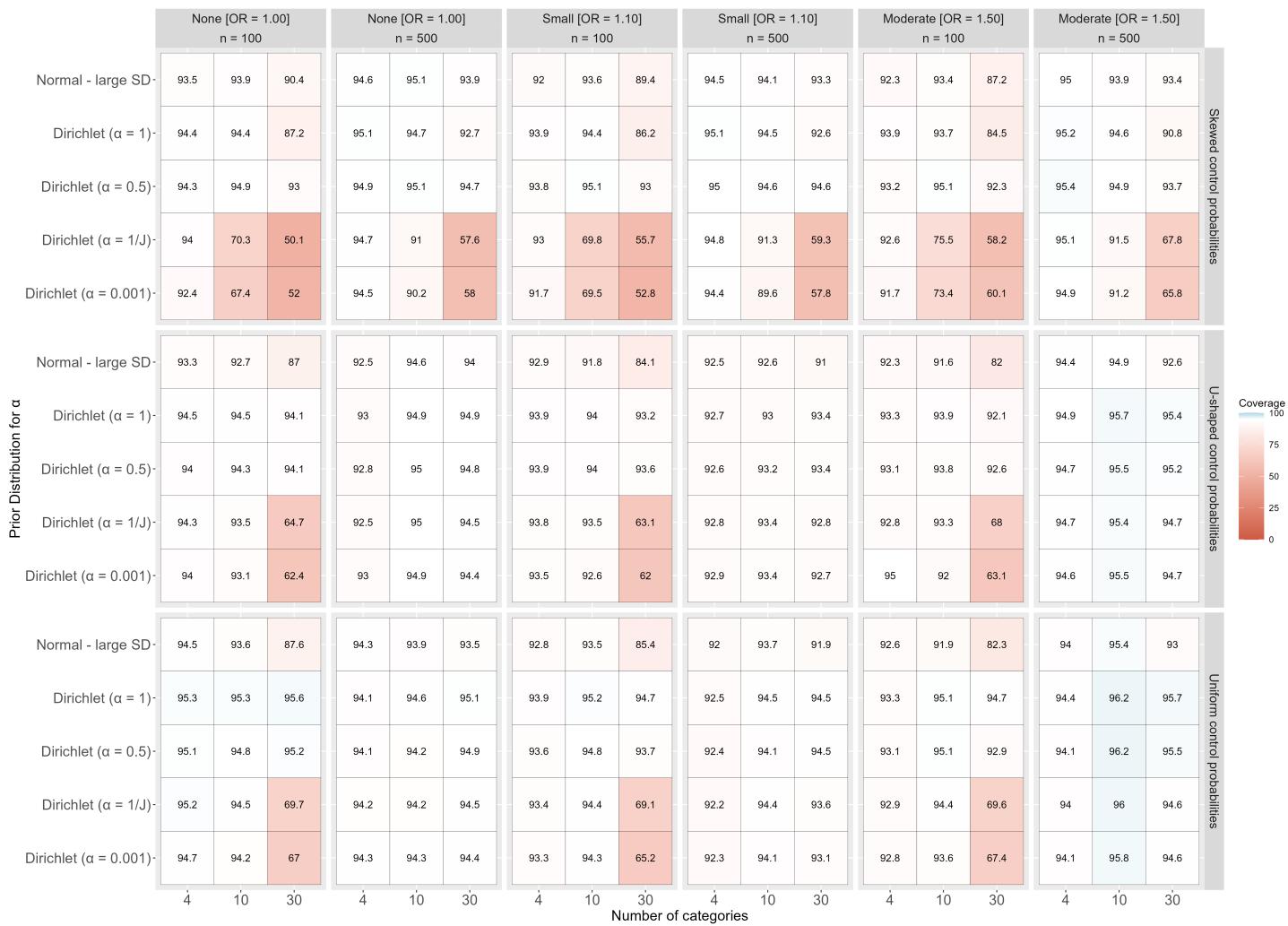
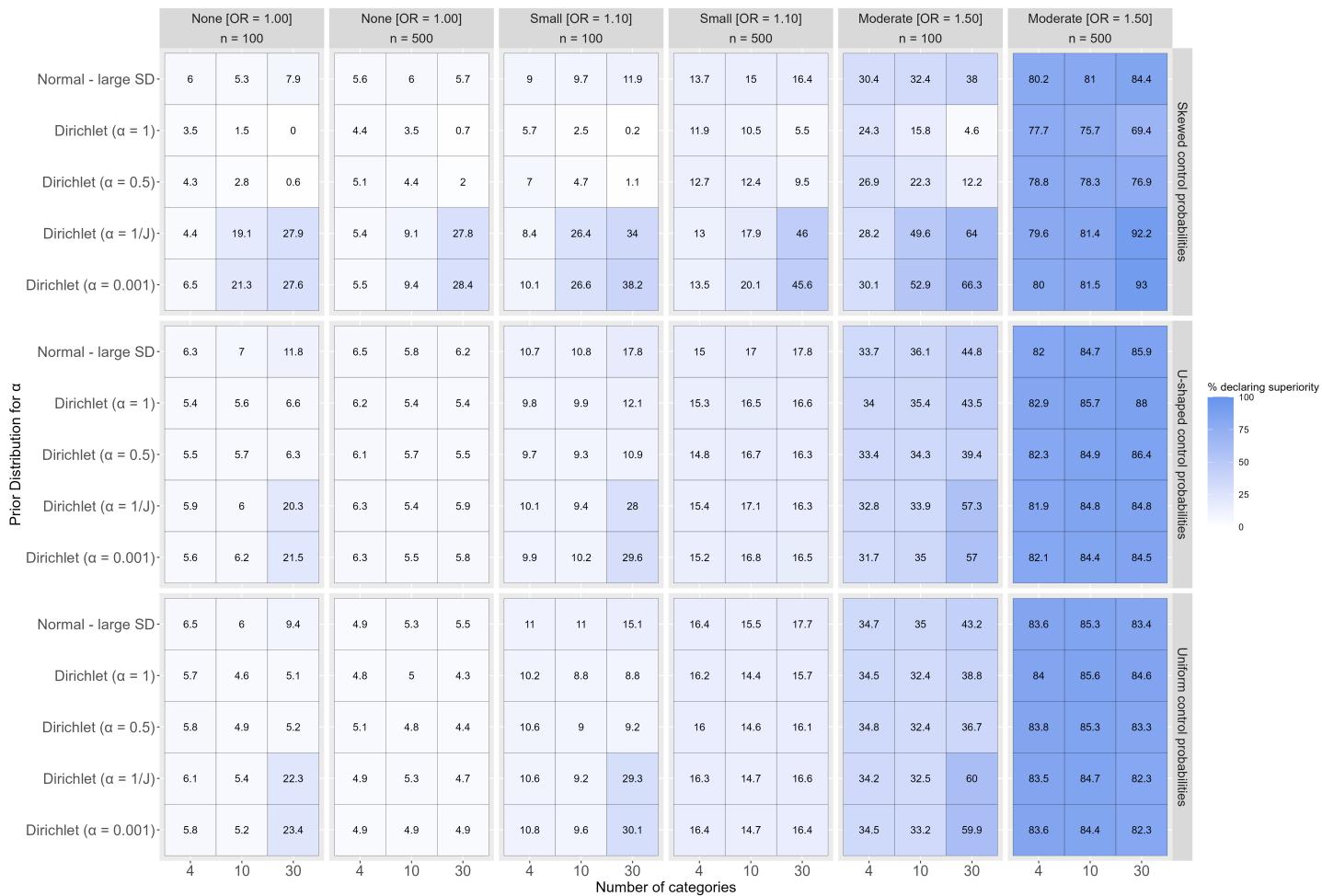


Figure 2c Proportion of trials declaring superiority in an adaptive design when varying the implicit prior on the cut-points

OR = odds ratio, SD = standard deviation



When altering the prior on the cut-points, divergent transitions sometimes occurred, despite increasing the target acceptance probability rate and maximum tree-depth to reduce the number that occurred. Divergent transitions were particularly common when the Dirichlet prior was used with small concentration parameters when the sample size was small and the outcome had 30 categories. We present the results with and without divergent transitions given some results had high \hat{R} and very small effective sample sizes (see Supplementary Material 3). When the simulated datasets resulting in divergent transitions were removed, all estimates of the \hat{R} were < 1.01 , effective sample sizes were sufficiently large, and the MCSE for the performance measures were less than 0.05.

The ASCOT case study

ASCOT was an adaptive platform trial that evaluated the efficacy of multiple interventions for hospitalised, non-critically ill COVID-19 patients. It evaluated interventions across three intervention domains: anticoagulation, antiviral, and therapeutic antibody domains, each comparing a different set of interventions to standard of care within that domain. The primary outcome was death or the need for organ support by day 28 (a binary outcome), but the trial had four secondary ordinal endpoints (described below). We focus on the anticoagulation domain, which trialed low-dose thromboprophylaxis with low-molecular-weight heparin (LMWH), intermediate-dose LMWH, and a low-dose LMWH with once-daily aspirin. The aspirin arm was discontinued during the trial due to external evidence, and a new arm with therapeutic-dose anticoagulation was introduced. The trial ended early because of funding limitations, slowing participant recruitment, and a recommendation from the data safety and monitoring committee to stop therapeutic anticoagulation due to lack of effectiveness.

ASCOT had four ordinal secondary endpoints:

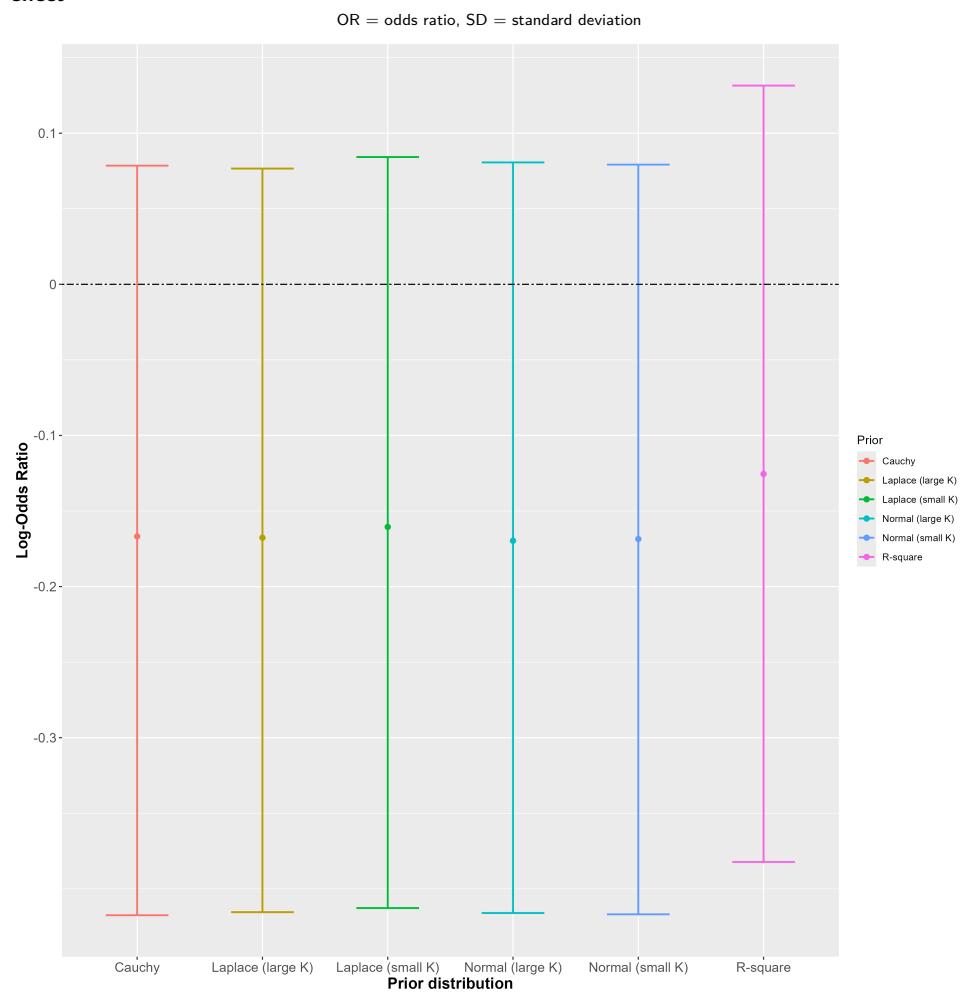
- 1 WHO 8-point ordinal outcome scale at day 28 post randomisation: (1) not hospitalised and no limitations on activities, (2) not hospitalised, limitations on activities, (3) hospitalised, not requiring supplemental oxygen and no longer requiring ongoing medical care, (4) hospitalised, not requiring supplemental oxygen but requiring ongoing medical care, (5) hospitalised, requiring supplemental oxygen, (6) hospitalised, on non-invasive ventilation or high flow oxygen devices, (7) hospitalised, on invasive mechanical ventilation or ECMO, and (8) death.
- 2 Modified Medical Research Council (mMRC) breathlessness scale (only asked among those diagnosed with COVID-19), a 5-point ordinal scale: (1) 'I only get breathless with strenuous exercise', (2) 'I get short of breath when hurrying on level ground or walking up a slight hill', (3) 'On level ground, I walk slower than people of the same age because of breathlessness, or I have to stop for breath when walking at my own pace on the level', (4) 'I stop for breath after walking about 100 yards or after a few minutes on level ground', and (5) 'I am too breathless to leave the house or I am breathless when dressing or undressing'.

- 3 Days alive and free of hospital by 28 days post randomisation: a 29-point ordinal outcome calculated as 28 minus the number of days spent in hospital, where patients dying within 28 days were assigned zero free days.
- 4 Days alive and free of ventilation by 28 days post randomisation: a 29-point ordinal outcome calculated as 28 minus the number of days free of invasive or non-invasive ventilation, where patients dying within 28 days were assigned zero free days.

The same prior distributions used in the simulation study were applied in the analysis of the anticoagulation domain of ASCOT using complete cases only [18]. For simplicity we focused on comparing only the low vs intermediate dose with LMWH and present the results for each of the four secondary ordinal endpoints. Supplementary Material 4 provides an overview of the distribution across categories by intervention group for each ordinal endpoint, which were largely skewed.

When the six priors for the treatment effect were applied to the analysis of the WHO scale, all methods estimated the proportional log-OR of the ordinal scale to be smaller than zero indicating the low-dose provided better odds of a more favourable outcome (Figure 4). All approaches resulted in a similar point estimate and large uncertainty irrespective of the prior (Figure 3). Similar estimated treatment effects were observed when the implicit prior on the cut-points is varied for all of these outcomes. Supplementary Material 4 indicates similar results across all other outcomes.

Figure 3 Estimated treatment effect of proportional log-ORs for the WHO eight-point scale in the Australasian COVID-19 Trial (ASCOT) obtained under the six prior specifications for treatment effect



Discussion

This study examined the impact of different non-informative prior distributions on parameters of a Bayesian PO model have on estimation and operating characteristics in the context of a fixed and an adaptive trial with an early stopping rule. When varying the prior for the treatment effect, results indicate that bias was substantial when the control arm probabilities are right-skewed for all priors apart from the R-squared prior specification, especially with increasing categories in the ordinal outcome and smaller sample size. If it is expected that participants will fall at both ends of the ordinal scale, then specifying a Normal, Cauchy or Laplace prior distribution for the treatment effect would be appropriate. There was not much variation in the results when a Laplace or Normal prior distribution was used for the treatment effect or with different values for the standard deviation for these priors. This implies that in fact, specifying a reasonably small or large variance term will still result in a rather ‘non-informative’ prior distribution that does not impact the treatment effect in the scenarios considered in this study. This study also found that when altering the implicit prior on the model cut-points while keeping the prior for the treatment effect fixed, decreasing the concentration parameters for a Dirichlet prior can reduce bias in treatment effect estimation for right-skewed control arm probabilities, with minimal bias when using a Normal distribution for cut-points (except for small sample sizes and 30 categories), and results in a higher proportion of trials to stop early.

The results from our study also show that it is more likely that divergent transitions will occur when a Dirichlet prior with concentration parameters close to zero is used for the control group probabilities which, although can attempt to be mitigated by increasing the target average acceptance probability rate, may still exist in some circumstances. Divergent transitions implies that there were issues in the sampler exploring the posterior distribution of the target parameter, which can lead to biased inference. Divergent transitions were more likely to occur when the sample size is small and with a larger number of categories. The implications of this is that careful consideration should be made for selecting an appropriate Dirichlet prior, particularly when it is known that the interim analysis will occur at a small sample size or the ordinal outcome has a large number of categories.

Symmetric priors like the Normal, Laplace or Cauchy distributions assume that the treatment effect could be in either direction. This may be unjustified if a large proportion of participants are expected to be at one end of the scale (and therefore there is data sparsity in the remaining categories) which may explain the bias with these approaches when the controlled probabilities are skewed. In contrast, the R-squared prior allows for the fact that the variation in the outcome might not be due to the treatment effect alone, but could also be explained by other factors such as the distribution of outcomes in the control arm. This could be more consistent with the skewed distribution of outcomes and hence may explain the minimal bias with this approach. When varying the prior for the control arm probabilities, the uniform Dirichlet prior (which reflects that there is equal probability in each category of the control arm) may lead to poor estimation of the control arm’s distribution in the

PO model, particularly when the control probabilities are skewed, resulting in an underestimation of the treatment effect. Trials that incorporated early stopping rules exhibited larger bias in the treatment effect for some of the priors than the fixed design, particularly for U-shaped control probabilities. This aligns with the expectations that trials that stop early are more likely due to random highs [35]. Adaptive designs that used the R-squared prior specification had the highest early stopping rates when the control data were skewed, particularly when the ordinal outcome had many categories.

We also found that decreasing the concentration parameters for a Dirichlet prior can reduce bias in treatment effect estimation for right-skewed control arm probabilities. Reducing the concentration parameters allow the Dirichlet prior to become more diffuse, meaning it gives less weight to a prior belief that the categories should be equally distributed and uniform. This therefore gives the model more freedom to adjust based on the actual data. When the control arm probabilities were U-shaped, the choice of prior for the control group probabilities had little impact on the estimation of the treatment effect, likely because the prior distributions are symmetric about the middle category and there is reasonable probability of falling in each category. We found that using a Dirichlet prior for the control group probabilities with a concentration parameter closer to zero resulted in a trial that was more likely to stop early and/or declare superiority, even when there was no association between intervention and outcome. This is likely because the lower value of the concentration parameter leads to more uncertainty in the prior belief about the outcome. This flexibility makes the model more sensitive to random fluctuations in the data, increasing the chances of detecting a treatment effect even in the absence of a true effect, leading to premature stopping of the trial. Another possible explanation is that the priors for the model parameters for the treatment effect and model ‘intercepts’ have been specified independently from each other *a priori* (and are therefore approximately independent in the posterior). However, in our study, we used standard treatment coding (assigning 0 and 1 for the treatment indicator) which does not ensure that these parameters are uncorrelated in the model. An alternative would be to center the treatment contrast specification, that is, to code the treatment variable as $\pm \frac{1}{2}$ [36], to ensure the treatment coefficient is approximately uncorrelated with the model ‘intercepts’ which may alleviate the observed bias. Although not the focus of this study, future research could explore centering of treatment covariates in more detail.

The results of this study has implications for practice. It is common for trials (whether fixed or adaptive) to specify a Normal prior distribution for the treatment effect and an implicit Dirichlet prior on the model cut-points in a Bayesian PO model [1] using $\alpha = \mathbf{1}$. However, the results from this study suggest that the most appropriate priors may depend on the distribution of participants in the control arm. Use of pilot data or similar studies would therefore be useful prior to data collection in determining the distribution of patients across the categories in order to guide the selection of priors. For example, if it is expected that control arm patients are more likely to fall in the first few categories of the ordinal outcome,

then an R-squared prior specification on the treatment effect might be more appropriate, particularly if the ordinal outcome has a large number of categories. If little is known about the distribution of the outcome in advance, then pre-specified sensitivity analyses could be conducted using various non-informative priors to assess the robustness of the results to the priors used at the final analysis. For example, if a diffuse Normal prior is used for the treatment effect, which is common in practice, then decreasing the value of the concentration parameters of the Dirichlet prior closer to zero or using a Normal prior for the cut-points may reduce bias and increase the likelihood of stopping the trial early when the control arm probabilities are skewed. We acknowledge that an exploration of the different priors would not be as useful for an interim analysis if one prior resulted in a threshold being reached to stop but another prior did not.

Our study has several strengths. The simulation study explores a number of realistic scenarios (108 in total). We have considered both a fixed and an adaptive design, and considered a range of prior distributions that could be used. We have also applied the same approaches to a case study with four different ordinal outcomes. However, this study does have its limitations. First, there are many other possible control arm distributions and adaptive designs that could be explored where the different non-informative priors could lead to different results. Second, we did not explore the impact that the specification of the priors may have in a multi-arm trial or in an analysis that adjusts for covariates in the model, such as randomisation strata, although we do not expect the latter to affect the results. Third, we only considered a limited set of commonly used priors and acknowledge that there are other alternatives that may warrant investigation. Finally, we did not consider the impact that departure from PO has on the results as this may not be a realistic assumption to make in practice. This should be investigated in future studies.

In summary, there are numerous possible non-informative prior distributions that can be specified for the Bayesian PO model, and careful specification of these should be considered when conducting such an analysis. Importantly, the most appropriate prior appears to depend on the distribution of the control arm probabilities. We recommend the use of the R-squared prior specification when the control arm is skewed, and a Dirichlet prior with concentration parameters close to zero in conjunction with a Normal prior for the treatment effect if the analysis occurs at a reasonable sample size (e.g. $n = 500$), otherwise concentration parameters set to one would be appropriate to avoid divergent transitions. Choosing an appropriate non-informative prior will result in better inference and therefore trial conclusions and adaptive decision-making.

Declarations

Acknowledgements

The authors would like to thank the Australian COVID-19 (ASCOT) trial team for providing access and use of the ASCOT trial data to illustrate the methods discussed in this paper, and to ASCOT investigator and statistical team member Tom Snelling for providing useful insights when shaping the manuscript.

Funding

This work forms part of Chris Selman's PhD, which is supported by the Research Training Program Scholarship, administered by the Australian Commonwealth Government and The University of Melbourne, Australia. Chris Selman's PhD was also supported by a Centre of Research Excellence grant from the National Health and Medical Research Council of Australia ID 1171422, to the AusTriM Research Network. Research at the Murdoch Children's Research Institute is supported by the Victorian Government's Operational Infrastructure Support Program. This work was supported by the Australian National Health and Medical Research Council (NHMRC) Centre for Research Excellence grants to the Victorian Centre for Biostatistics (ID 1035261) and the Australian Trials Methodology Research Network (ID 1171422), including through seed funding awarded to Robert Mahar. Katherine Lee is funded by an NHMRC investigator grant (ID 2017498). Michael Dymock is supported by a NHMRC Postgraduate Research Award (APP2022557). The funding bodies played no role in the study conception, design, data collection, data analysis, data interpretation or writing of the report.

Consent for Publication

Not applicable.

Availability of data and materials

The code and datasets generated for this simulation study are available on Github [37]:
https://github.com/chrisselman/noninfo_pomodel.

Data used for the illustrative example within this paper are available upon reasonable request from the ASCOT study team.

Abbreviations

ASCOT: Australasian COVID-19 Trial, ESS: Effective Sample Size, HMC: Hamiltonian Monte Carlo, LMWH: Low-Molecular-Weight Heparin, RCT: randomised controlled trial, OR: odds ratio, PO: proportional odds, MCSE: Monte Carlo standard error, mMRC: Modified Medical Research Council, WHO: World Health Organisation.

Ethics approval and consent to participate

Use of the ASCOT data was approved by both the ASCOT Global Trial Steering Committee and the MCRI Change Advisory Board.

Registration details

The ASCOT trial is registered on the Australian New Zealand Clinical Trials Registry (ANZCTR) with registration number ACTRN12620000445976 on the 6 April 2020.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CJS, KJL, RKM conceived the study. CJS conducted the simulation study and data analysis and wrote the first draft of the manuscript, with regular input from KJL and RKM. All authors contributed to the design of the study, critical revision of the manuscript, and take responsibility for its content.

Author details

¹Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Parkville, VIC, 3052, Australia. ²Department of Paediatrics, University of Melbourne, Parkville, VIC, 3052, Australia. ³Wesfarmers Centre of Vaccines and Infectious Diseases, The Kids Research Institute Australia, Nedlands, WA, 6009, Australia. ⁴School of Population and Global Health, The University of Western Australia, Nedlands, WA, 6009, Australia. ⁵NHMRC Clinical Trials Centre, University of Sydney, Camperdown, NSW, 2050, Australia. ⁶Department of Infectious Diseases, University of Melbourne, at the Peter Doherty Institute for Infection and Immunity, Parkville, VIC, 3052, Australia. ⁷Victorian Infectious Diseases Service, The Royal Melbourne Hospital, at the Peter Doherty Institute for Infection and Immunity, Parkville, VIC, 3052, Australia. ⁸Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, NSW, 2006, Australia. ⁹Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Parkville, VIC, 3052, Australia. ¹⁰Methods and Implementation Support for Clinical and Health Research Hub, University of Melbourne, Parkville, VIC, 3052, Australia.

References

1. Selman, C.J., Lee, K.J., Ferguson, K.N., Whitehead, C.L., Manley, B.J., Mahar, R.K.: Statistical analyses of ordinal outcomes in randomised controlled trials: a scoping review. *Trials* **25**(1), 1–18 (2024)
2. Bonita, R., Beaglehole, R.: Recovery of motor function after stroke. *Stroke* **19**(12), 1497–1500 (1988)
3. Mathioudakis, A.G., Fally, M., Hashad, R., Kouta, A., Hadi, A.S., Knight, S.B., Bakerly, N.D., Singh, D., Williamson, P.R., Felton, T., et al.: Outcomes evaluated in controlled clinical trials on the management of COVID-19: a methodological systematic review. *Life* **10**(12), 350 (2020)
4. Walker, S.H., Duncan, D.B.: Estimation of the probability of an event as a function of several independent variables. *Biometrika* **54**(1-2), 167–179 (1967)
5. McCullagh, P.: Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)* **42**(2), 109–127 (1980)
6. Harrell, F.E., et al.: *Regression Modeling Strategies: with Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis* vol. 3. Springer, ??? (2015)
7. Stallard, N., Hampson, L., Benda, N., Brannath, W., Burnett, T., Friede, T., Kimani, P.K., Koenig, F., Krisam, J., Mozgunov, P., et al.: Efficient adaptive designs for clinical trials of interventions for covid-19. *Statistics in Biopharmaceutical Research* **12**(4), 483–497 (2020)
8. Perkins, G.D., Couper, K., Connolly, B., Baillie, J.K., Bradley, J.M., Dark, P., De Soyza, A., Gorman, E., Gray, A., Hamilton, L., et al.: Recovery-respiratory support: respiratory strategies for patients with suspected or proven covid-19 respiratory failure; continuous positive airway pressure, high-flow nasal oxygen, and standard care: a structured summary of a study protocol for a randomised controlled trial. *Trials* **21**(1), 1–3 (2020)
9. Florescu, S., Stanciu, D., Zaharia, M., Kosa, A., Codreanu, D., Kidwai, A., Masood, S., Kaye, C., Coutts, A., MacKay, L., et al.: Effect of antiplatelet therapy on survival and organ support-free days in critically ill patients with covid-19: a randomized clinical trial. *Jama* **327**(13), 1247–1259 (2022)
10. Kairalla, J.A., Coffey, C.S., Thomann, M.A., Muller, K.E.: Adaptive trial designs: a review of barriers and opportunities. *Trials* **13**, 1–9 (2012)
11. McMurray, J.J., Packer, M., Desai, A.S., Gong, J., Lefkowitz, M.P., Rizkala, A.R., Rouleau, J.L., Shi, V.C., Solomon, S.D., Swedberg, K., et al.: Angiotensin-neprilysin inhibition versus enalapril in heart failure. *N Engl J Med* **371**, 993–1004 (2014)
12. Berry, D.A.: Bayesian clinical trials. *Nature reviews Drug discovery* **5**(1), 27–36 (2006)
13. Bast Jr, R.C., Holland, J.F.: *Holland-Frei Cancer Medicine* 8. PMPH-USA, ??? (2010)
14. Berry, D.A.: Introduction to Bayesian methods III: use and interpretation of Bayesian tools in design and analysis. Sage Publications Sage CA: Thousand Oaks, CA (2005)
15. Gelman, A., Shalizi, C.R.: Philosophy and the practice of bayesian statistics. *British Journal of Mathematical and Statistical Psychology* **66**(1), 8–38 (2013)
16. McKinley, T.J., Morters, M., Wood, J.L.: Bayesian model choice in cumulative link ordinal regression models (2015)
17. James, N.T., Harrell Jr, F.E., Shepherd, B.E.: Bayesian cumulative probability models for continuous and mixed outcomes. *arXiv preprint arXiv:2102.00330* (2021)
18. McQuilten, Z.K., Venkatesh, B., Jha, V., Roberts, J., Morpeth, S.C., Totterdell, J.A., McPhee, G.M., Abraham, J., Bam, N., Bandara, M., et al.: Anticoagulation strategies in non-critically ill patients with covid-19. *NEJM Evidence* **2**(2), 2200293 (2023)
19. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: *Bayesian Data Analysis*. Chapman and Hall/CRC, ??? (2014)
20. Gelman, A., Jakulin, A., Pittau, M.G., Su, Y.-S.: A weakly informative default prior distribution for logistic and other regression models (2008)
21. Sarma, A., Kay, M.: Prior setting in practice: Strategies and rationales used in choosing prior distributions for bayesian analysis. In: *Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems*, pp. 1–12 (2020)
22. Rochford, A.: Prior Distributions for Bayesian Regression Using PyMC. <https://austinrochford.com/posts/2013-09-02-prior-distributions-for-bayesian-regression-using-pymc.html>. Accessed: 2025-01-20 (2013)
23. Gabry, J., Goodrich, B.: Estimating Regularized Linear Models with rstanarm. <https://cran.r-project.org/web/packages/rstanarm/vignettes/lm.html#priors>. Accessed: 2025-01-20 (2024)

24. Tu, S.: The dirichlet-multinomial and dirichlet-categorical models for bayesian inference. Computer Science Division, UC Berkeley 2 (2014)
25. Hill, M.D., Goyal, M., Menon, B.K., Nogueira, R.G., McTaggart, R.A., Demchuk, A.M., Poppe, A.Y., Buck, B.H., Field, T.S., Dowlatshahi, D., et al.: Efficacy and safety of nerinetide for the treatment of acute ischaemic stroke (escape-na1): a multicentre, double-blind, randomised controlled trial. *The Lancet* **395**(10227), 878–887 (2020)
26. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2021). R Foundation for Statistical Computing. <https://www.R-project.org/>
27. Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A.: Stan: A probabilistic programming language. *Journal of statistical software* **76**(1) (2017)
28. Bürkner, P.-C., Gabry, J., Kay, M., Vehtari, A.: posterior: Tools for Working with Posterior Distributions. R package version 1.6.0 (2024). <https://mc-stan.org/posterior/>
29. Goodrich, B., Gabry, J., Ali, I., Brilleman, S.: rstanarm: Bayesian applied regression modeling via Stan (2020). <https://mc-stan.org/rstanarm>
30. Kelter, R.: The bayesian simulation study (basis) framework for simulation studies in statistical and methodological research. *Biometrical Journal*, 2200095 (2023)
31. Koehler, E., Brown, E., Haneuse, S.J.-P.: On the assessment of monte carlo error in simulation-based statistical analyses. *The American Statistician* **63**(2), 155–162 (2009)
32. Morris, T.P., White, I.R., Crowther, M.J.: Using simulation studies to evaluate statistical methods. *Statistics in medicine* **38**(11), 2074–2102 (2019)
33. Goodrich, B., Gabry, J., Ali, I., Brilleman, S.: rstanarm: Bayesian applied regression modeling via stan. R package version 2(1) (2020)
34. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., Bürkner, P.-C.: Rank-normalization, folding, and localization: An improved r \hat{r} for assessing convergence of mcmc (with discussion). *Bayesian analysis* **16**(2), 667–718 (2021)
35. Walter, S., Guyatt, G., Bassler, D., Briel, M., Ramsay, T., Han, H.: Randomised trials with provision for early stopping for benefit (or harm): the impact on the estimated treatment effect. *Statistics in medicine* **38**(14), 2524–2543 (2019)
36. Kraemer, H.C., Blasey, C.M.: Centring in regression analyses: a strategy to prevent errors in statistical inference. *International journal of methods in psychiatric research* **13**(3), 141–151 (2004)
37. Selman, C.J.: Simulation Study Code. GitHub (2025)