Supplementary Material

Arthur M. Albuquerque, Lucas Tramujas, Lorenzo R. Sewanan, Donald R. Williams, James M. Brophy

Table of Contents

# eMethods

## Bayesian meta-analysis

Our meta-regression random-effect model is defined as:

where is the observed mean log odds ratio of tocilizumab versus control and is the known sampling variance in study . is the intercept, which represents the overall effect of tocilizumab in patients on **invasive mechanical ventilation**. is the difference between patients on simple oxygen only (SOO) and patients on invasive mechanical ventilation (). is the difference between patients on noninvasive ventilation (NIV) and patients on invasive mechanical ventilation (). Both coefficients are multiplied by , which is dummy-coded. Lastly, represents the between-study heterogeneity.

In this case, we are able to assess tocilizumab’s effect in each subgroup while assuming a common between-study heterogeneity:

* SOO =
* NIV =
* IMV =

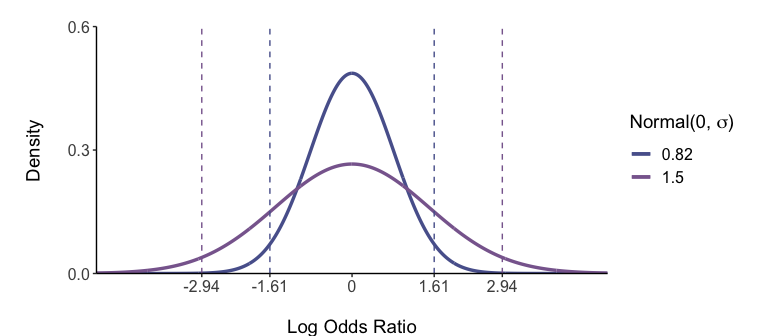
### Weakly informative priors

Because we applied the Bayesian framework, we assigned a prior distribution for each parameter. In our main model, we implemented priors that cover plausible values for all parameters, assigning limited density to impossible values, and thus employed little influence in the results (hereafter, known as weakly informative priors). These are our weakly informative priors:

Now, we will explain the rationale underlying these distributions.

We find highly unlikely that a pharmacological treatment, such as tocilizumab, will yield a 80% odds reduction in 28-days all-cause mortality regardless of the subgroup of patients, as suggested by empirical evidence. Thus, for (Intercept), we set a prior distribution of in the log odds ratio scale.

Moreover, we set very weakly informative priors for and (Coefficients), because we did not expect large differences between the invasive mechanical subgroup () and the other two respiratory support subgroups.

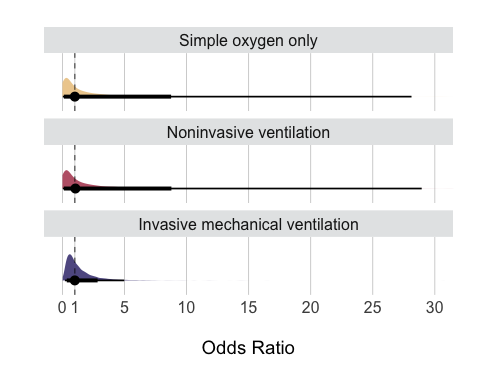


To facilitate the understanding of these distributions, here are the corresponding 95% quantile intervals in the linear scale.

|  | Log scale | | Linear scale | |
| --- | --- | --- | --- | --- |
| Prior | Mean | SD | Mean | 95% CI |
| Intercept | 0 | 0.82 | 1 | [0.2, 5] |
| Coefficients | 0 | 1.50 | 1 | [0.1, 18.9] |

Of note, while the “Intercept” can be interpreted as tocilizumab’s effect (odds ratio) in the linear scale, the coefficients pose a different interpretation. In this model, the coefficients correspond to the difference between a subgroup (SOO or NIV) in comparison to the intercept (IMV subgroup) in the log scale. While each coefficient corresponds to subtraction operation per se, the interpretation changes in the linear scale. When exponentiated, a subtraction in the log scale yields a division operation. Thus, the coefficient parameters should be interpreted as the ratio of odds ratios between the IMV and the other corresponding subgroup (SOO or NIV).

Another way to assess the plausibility of the aforementioned priors is to perform a prior predictive check, which can be visualized below:



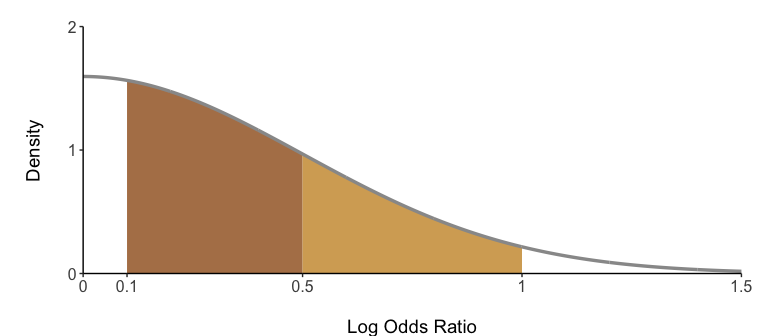
Point estimates depict the median and interval bars depict the 80% and 95% quantile intervals.

As expected, the distribution for IMV approximately ranges from 0.2 to 5.0. The intervals for SOO and NIV are wider given the addition of another prior distribution for each subgroup in comparison to IMV. Nevertheless, the prior predictive check above confirms the weakly informative nature of the aforementioned prior distributions.

Lastly, we will now discuss the weakly informative prior distribution for . Because we wanted to perform unconditional inferences beyond the included studies, we fitted a random-effect meta-analysis. In this model, one assumes there is within-study heterogeneity (represented by , the known sampling variance in study ) and the between-study heterogeneity (represented by ).

Although the definition of small or large between-study heterogeneity is arbitrary, previous work suggests cutoff values (“reasonable” heterogeneity between 0.1 and 0.5, “fairly high” between 0.5 and 1.0, and “fairly extreme” for values larger than 1.0 log odds ratio).[7,13] We added a category for low heterogeneity (between 0 and 0.1).

The distribution yields plausible probabilities in each of these ranges.



Here are the corresponding probabilities within each of the aforementioned heterogeneity ranges:

| Heterogeneity Range | | | |
| --- | --- | --- | --- |
| Low | Reasonable | Fairly High | Fairly Extreme |
| 16% | 52% | 27% | 5% |

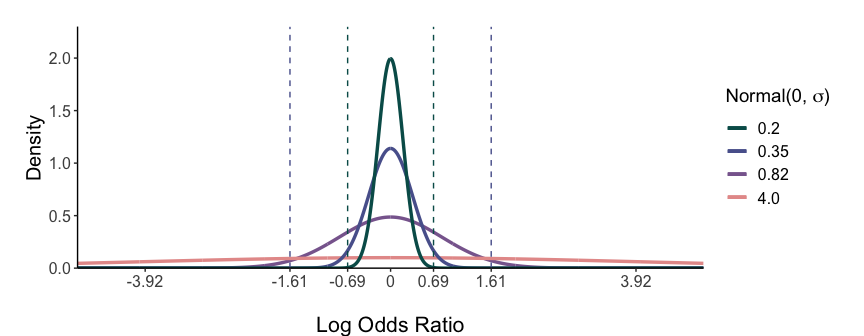
### Alternative priors

To check whether the choice of weakly informative priors meaningfully impacted our results or our conclusions, we also fitted models using vague or informative priors.

Vague priors:

Informative priors:

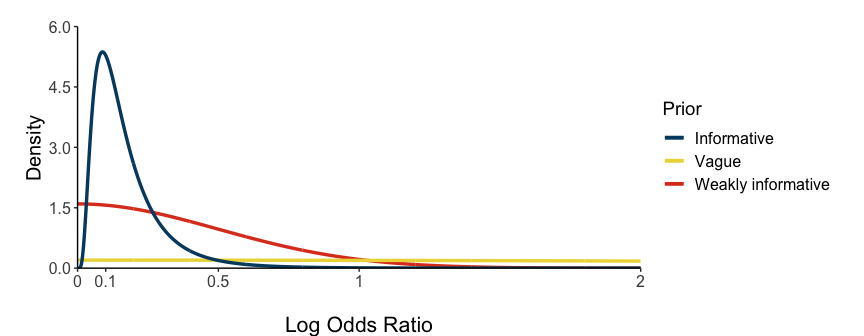
Here are graphical representations of these normal distributions (along with the weakly informative mentioned before):



| Log scale | | Linear scale | |
| --- | --- | --- | --- |
| Mean | SD | Mean | 95% CI |
| 0 | 0.20 | 1 | [0.7, 1.5] |
| 0 | 0.35 | 1 | [0.5, 2] |
| 0 | 0.82 | 1 | [0.2, 5] |
| 0 | 4.00 | 1 | [0, 2540.2] |

Again, we note that the interpretation in the linear scale is different between the Intercept (tocilizumab’s effect as odds ratio) and Coefficients (relative difference between subgroups as ratio of odds ratios).

Here are graphical representations of distributions for the between-study standard deviation () (along with the weakly informative mentioned before):



|  | | Heterogeneity Range | | | |
| --- | --- | --- | --- | --- | --- |
| Prior | Distribution | Low | Reasonable | Fairly High | Fairly Extreme |
| Weakly informative | Half-Normal(0.5) | 16% | 52% | 27% | 5% |
| Vague | Half-Normal(4.0) | 2% | 8% | 10% | 80% |
| Informative | Log-Normal(-1.975, 0.67) | 31% | 66% | 3% | 0% |

## Simulation analysis

Bayesian methods allows one to incorporate external evidence in the analysis through the prior distribution. Here, we simulated six randomized controlled trials (RCT) of different sample sizes comparing tocilizumab and control treatment in patients on invasive mechanical ventilation. We then incorporated these RCTs in our analysis by defining the prior distribution for this subgroup based on these simulations.

All RCTs were set to find an effect size of log odds ratio of -0.26. This value is the equal to log(0.77), which was chosen based on WHO’s meta-analysis ([page 14 in Supplement 2](doi:10.1001/jama.2021.11330)). This is the mean odds ratio of tocilizumab vs. control in patients using corticosteroids (overall results). We decided to use this value to reflect an skeptical view to heterogeneity of treatment effect across subgroups, and thus the “real” effect in this subgroup would be equal to the largest body of evidence for tocilizumab in all hospitalized COVID-19 patients on corticosteroids.

As mentioned above, we incorporated these RCTs through the prior distribution. To this end, we re-fitted our primary model six different times to incorporate these RCTs separately. More specifically, we did not sequentially incorporate these RCTs to our model. Instead, we fitted six separate models.

Here is the model in which we incorporated these simulations:

This is identical to our primary model in which we used weakly informative priors. However, now the prior for the intercept () - which corresponds to tocilizumab’s effect in the invasive mechanical ventilation subgroup - was set to incorporate the RCTs mentioned above.

A normally distributed prior is defined by the mean and standard deviation (SD). We set the mean to be equal to -0.26, as explained before. We will now explain how we defined the SD of each corresponding prior distribution.

Given that all six simulated RCTs were set to find the same effect size, the only difference between them was the total number of included patients: 200, 500, 1000, 1500, 2000, or 4000. To calculate the standard deviation of each corresponding prior based on the number of total patients included, one must also assume the proportion of patients included in each treatment arm and the mortality risk in the control arm:

1. We assumed equal allocation in both treatment arms
2. Adapting from the suggestions in the [GRADE guidelines](http://dx.doi.org/10.1016/j.jclinepi.2012.01.012), we found a striking discrepancy between the control mortality risk in this data (52%) in comparison to another [previously published meta-analysis](https://jamanetwork.com/journals/jama/fullarticle/2770279?utm_campaign=articlePDF&utm_medium=articlePDFlink&utm_source=articlePDF&utm_content=jama.2020.17023) (34% in patients on IVM and using corticosteroids). Thus, we have decided to use 43% (arithmetic mean between 34 and 52) as our reference risk in the IVM subgroup
3. The mortality risk in the tocilizumab was calculated using the following formula (Box 1 in [Doi et al., 2020](https://doi.org/10.1016/j.jclinepi.2020.08.019)):

where is the mortality risk in the tocilizumab group, is the mortality risk in the control group and is the odds ratio mentioned above. Thus, the tocilizumab risk is equal to 0.37.

In summary, we are simulating RCTs with mean OR equal to 0.77, control risk mortality of 43, and tocilizumab risk of 37.

Based on these values, we can estimate the standard deviation (SD) with the following [formula](https://doi.org/10.1016/j.jclinepi.2008.07.006):

where , , and are number of events and follow this 2x2 table:

| Event | Tocilizumab | Control |
| --- | --- | --- |
| Death | a | c |
| No death | b | d |

As similarly shown in the supplementary material of [Higgins and Spiegelhalter, 2002](https://academic.oup.com/ije/article/31/1/96/655931?login=true), we can estimate these values as:

where and are the sample sizes in the tocilizumab and control arms, respectively. As mentioned above, we assume equal allocation in both treatment arms, thus .

Finally, we can estimate the based on the 6 different sample sizes mentioned above:

| Sample size in each treatment arm | Total sample size | SD |
| --- | --- | --- |
| 100 | 200 | 0.29 |
| 250 | 500 | 0.18 |
| 500 | 1,000 | 0.13 |
| 750 | 1,500 | 0.11 |
| 1,000 | 2,000 | 0.09 |
| 2,000 | 4,000 | 0.06 |

In summary, there are 4 parameters in these models, which are , , , and . The priors for the latter three parameters will be the same in every model:

On the other hand, the priors for are described as:

where is equal to -0.26 in every model, and the ranges from 0.29 to 0.06, as described in the table above.

## Deriving risk difference from odds ratio

We used the odds ratio as our primary estimand, following the suggestions by [Doi et al., 2020](https://doi.org/10.1016/j.jclinepi.2020.08.019) and [Doi et al., 2021](https://doi.org/10.1016/j.jclinepi.2021.08.003).

Moreover, we derived the risk in the tocilizumab group using the following formula (Box 1 in [Doi et al., 2020](https://doi.org/10.1016/j.jclinepi.2020.08.019)):

where is the mortality risk in the tocilizumab group, is the mortality risk in the control group and is the odds ratio.

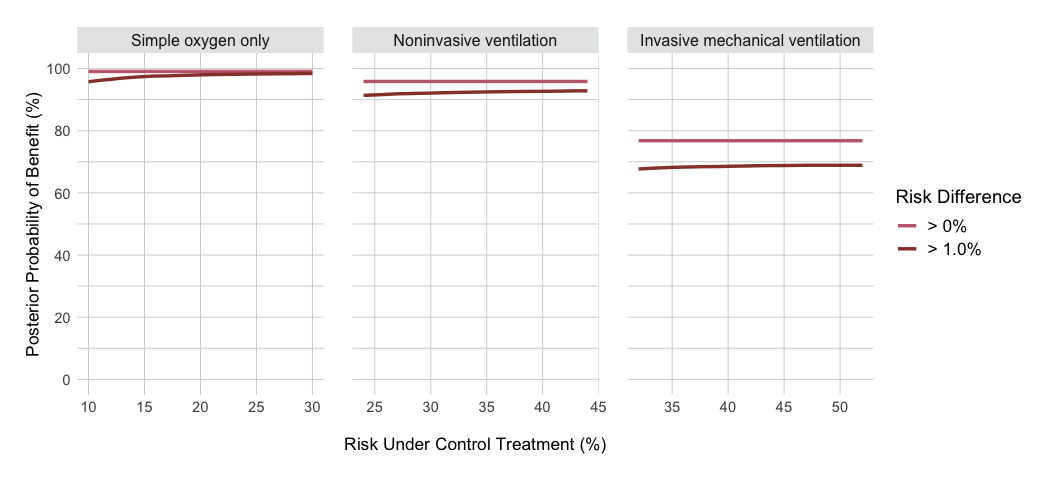
We then calculated the risk difference (RD) with the following formula, which was also the procedure suggested by [Doi et al., 2020](https://doi.org/10.1016/j.jclinepi.2020.08.019) (page 4):

Adapting from the suggestions in the [GRADE guidelines](http://dx.doi.org/10.1016/j.jclinepi.2012.01.012), we assumed different mortality risks in each subgroup. For the simple oxygen only and noninvasive ventilation subgroups, we used the average mortality risk in each subgroup based on the data included in this reanalysis of a [previously published meta-analysis](10.1001/jama.2021.11330). In contrast, regarding the invasive mechanical ventilation (IVM) subgroup, we found a striking discrepancy between the control mortality risk in this data (52%) in comparison to another [previously published meta-analysis](https://jamanetwork.com/journals/jama/fullarticle/2770279?utm_campaign=articlePDF&utm_medium=articlePDFlink&utm_source=articlePDF&utm_content=jama.2020.17023) (34% in patients on IVM and using corticosteroids). Thus, we have decided to use 43% (arithmetic mean between 34 and 52) as our reference risk in the IVM subgroup. Recognizing the potential variability of the subgroup baseline risks, we estimated the risk differences with twenty different plausible baseline risks for each subgroup (spanning +- 10% change from the reference risks mentioned above).

| Subgroup | Control Risk |
| --- | --- |
| Simple oxygen only | 20 +- 10% |
| Noninvasive ventilation | 34 +- 10% |
| Invasive mechanical ventilation | 42 +- 10% |

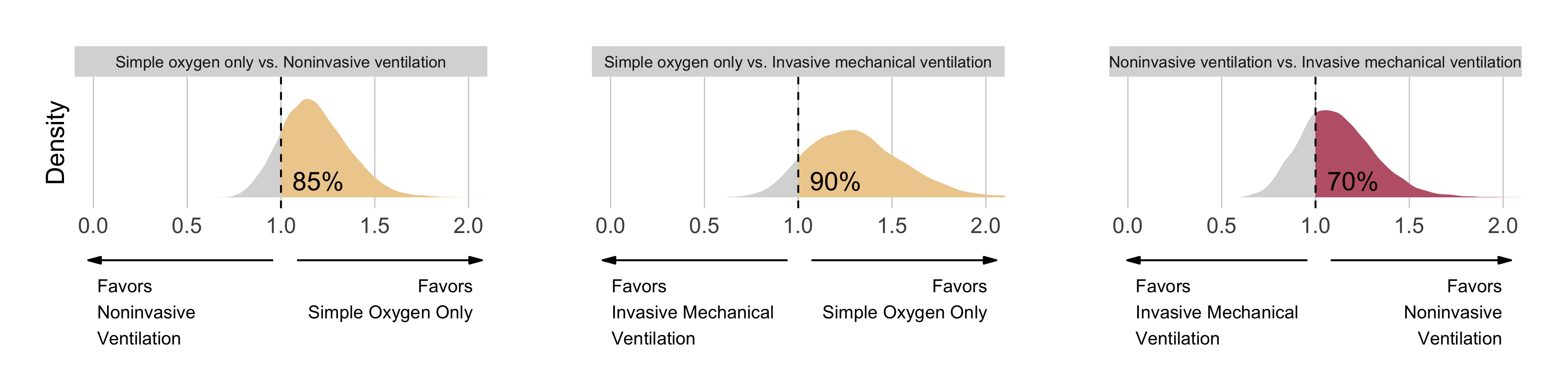
# eFigures and eTables

## eFigure 1



Posterior probabilities of benefit per subgroup in the risk difference scale assuming weakly informative priors. Each line represents the posterior probability of benefit for a specific cutoff, such as risk difference greater than 0% or 1%, across plausible ranges of mortality risk under control treatment. Underlying weakly informative priors are N(0, 0.82) for the mean effect, N(0, 1.5) for the coefficients, and HN(0.5) for the between-study standard deviation. N(mu, sigma) = Normal(mean, standard deviation); HN(sigma) = Half-Normal(standard deviation)

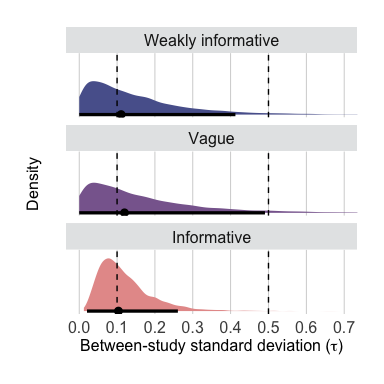
## eFigure 2



Posterior distributions for comparisons of effect sizes between subgroups while assuming weakly informative priors. Each distribution represents the ratio of odds ratios of two subgroups. On top of each distribution, there is a percentage representing the posterior probability of a ratio of odds ratios greater than 1.0. Arrows on the bottom represent - in that comparison - which subgroup benefited to a greater extent tocilizumab’s effect on mortality reduction. For example, assuming our model, there was a 85% probability that tocilizumab reduces mortality to a greater extent in the simple oxygen subgroup in comparison to noninvasive ventilation. Underlying weakly informative priors are N(0, 0.82) for the mean effect, N(0, 1.5) for the coefficients, and HN(0.5) for the between-study standard deviation. N(mu, sigma) = Normal(mean, standard deviation); HN(sigma) = Half-Normal(standard deviation).

## eTable 1

## eFigure 3



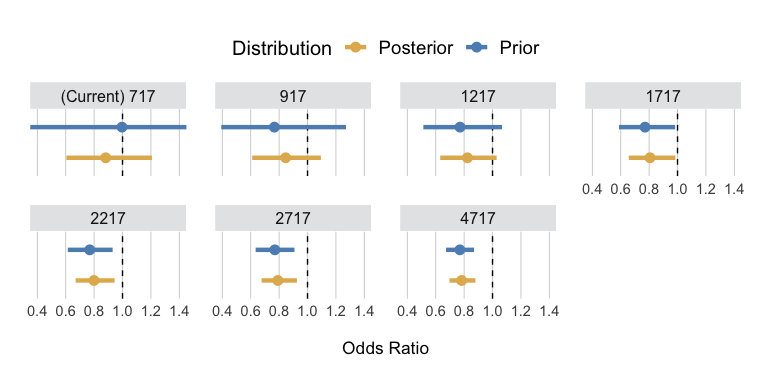
Posterior distributions (log scale) of the between study standard deviation (tau) upon different underlying prior distributions (weakly informative, vague or informative). Tau is a proxy for the between-study heterogeneity in random-effect meta-analyses. Weakly informative priors: Intercept N(0, 0.82); Coefficients N(0, 1.5); Between-study standard deviation HN(0.5) / Vague priors: Intercept N(0, 4); Coefficients N(0, 4); Between-study standard deviation HN(4) / Informative priors: Intercept N(0, 0.35); Coefficients N(0, 0.2); Between-study standard deviation LN(-1.975, 0.67). N(mu, sigma) = Normal(mean, standard deviation); HN(sigma) = Half-Normal(standard deviation); LN(mu, sigma) = Log-Normal(mean, standard deviation).

## eTable 2

## eTable 3

## eTable 4

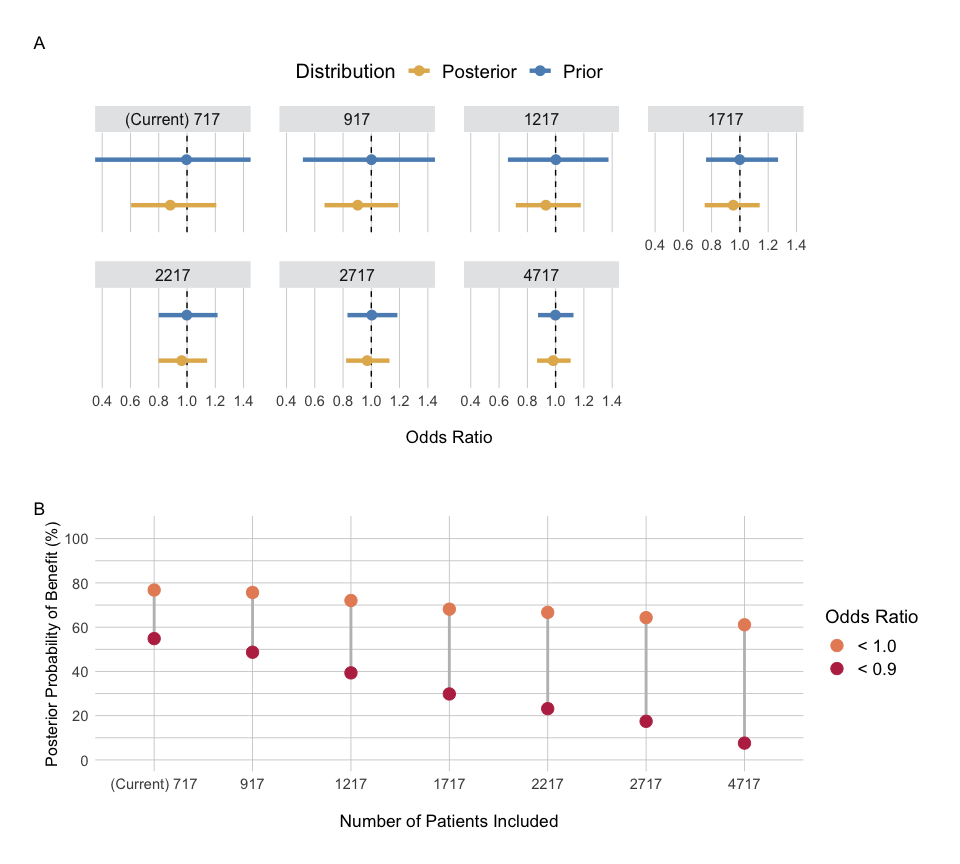
## eFigure 4



Results for the invasive mechanical ventilation subgroup from meta-analyses using an informative prior based on simulated randomized controlled trials (eTable 4). Each panel represents a different model, in which the prior distribution in centered at 0.77 odds ratio. The label on top of each panel depict the number of total patients on invasive mechanical ventilation included in each respective model (current plus simulated patients). Point estimates depict the median and interval bars represent the 95% compatibility (highest density) intervals for both prior and posterior distributions. The posterior distribution for ‘(Current) 717 patients’ depicts the results previously shown in Figures 1A and 1B for this subgroup. Underlying weakly informative priors are N(0, 1.5) for the coefficients, and HN(0.5) for the between-study standard deviation. N(mu, sigma) = Normal(mean, standard deviation); HN(sigma) = Half-Normal(standard deviation).

## eTable 5

## eFigure 5



Results for the invasive mechanical ventilation subgroup from meta-analyses using an informative prior based on simulated randomized controlled trials (eTable 5). In contrast to the results shown in eFigure4, the prior distributions in these analyses are centered at 1.0 odds ratio. Panel A: Each panel represents a different model. The label on top of each panel depict the number of total patients on invasive mechanical ventilation included in each respective model (current plus simulated patients). Point estimates depict the median and interval bars represent the 95% compatibility (highest density) intervals for both prior and posterior distributions. The posterior distribution for ‘(Current) 717 patients’ depicts the results previously shown in Figures 1A and 1B for this subgroup. Panel B shows the posterior probability of benefit for different thresholds (OR < 1.0 and < 0.9). Underlying weakly informative priors are N(0, 1.5) for the coefficients, and HN(0.5) for the between-study standard deviation. N(mu, sigma) = Normal(mean, standard deviation); HN(sigma) = Half-Normal(standard deviation).