Dexamethasone 12 mg versus 6 mg for Patients with COVID-19 and Severe Hypoxaemia: a Pre-Planned, Secondary Bayesian Analysis of the COVID STEROID 2 Trial

Electronic Supplementary Material

Table of contents

Enrolment criteria

Detailed enrolment criteria adapted from the trial protocol and primary trial publication [1,2]:

Inclusion criteria

Hospitalised adult patient (≥18 years) with documented coronavirus disease 2019 (COVID-19) receiving at least 10 L/min of supplemental oxygen (regardless of delivery system) or requiring mechanical ventilation (continuous use of continuous positive airway pressure, non-invasive ventilation, or invasive mechanical ventilation).

Exclusion criteria

- Use of systemic corticosteroids for other indications than COVID-19 in daily doses higher than 6 mg dexamethasone equivalents
- Use of systemic corticosteroids for COVID-19 for ≥5 consecutive days
- Invasive fungal infection
- Active tuberculosis
- Fertile woman (<60 years of age) with positive urine or plasma human chorionic gonadotropin
- Known hypersensitivity to dexamethasone
- Previously randomised in the COVID STEROID 2 trial
- Informed consent not obtainable

Detailed outcome definitions

Detailed outcome definitions adapted from the trial protocol and primary trial publication [1,2]:

1. Days alive without life support at day 28: calculated as the total number of days alive without any of the three types of life support listed below within 28 days of randomisation:
    a. Invasive mechanical ventilation: use of mechanical ventilation via a cuffed endotracheal tube.
    b. Circulatory support: continuous infusion of any vasopressor or inotropic agent for ≥1 hour.
    c. Kidney replacement therapy (KRT): any form of acute or chronic intermittent or continuous KRT, including days in between intermittent KRT. Periods with up to 3 days between intermittent KRT were counted as days with KRT [2].
2. Serious adverse reactions: ≥1 of the following from randomisation to day 28 (or for available days, for patients where consent for further data registration was withdrawn):
    a. New episodes of septic shock
    b. Invasive fungal infection
    c. Clinically important gastrointestinal bleeding
    d. Anaphylactic reaction to intravenous dexamethasone
3. All-cause mortality at day 28: death from any cause within 28 days of randomisation
4. All-cause mortality at day 90: death from any cause within 90 days of randomisation
5. Days alive without life support at day 90: defined as above, except follow-up continued until day 90.
6. Days alive and out of hospital at da 90: calculated as the total number of days that the patient was alive and out of hospital within 90 days of randomisation.

For the three count outcomes, death was not penalised (i.e., non-survivors did not receive the lowest possible value of 0 days), as has previously been done [3-5], and consequently, *post hoc* analyses assigning non-survivors 0 days was conducted to ease comparison with other trials. For serious adverse reactions, additional details on the individual components are presented elsewhere [2]. For all outcomes, additional detailed results for their individual components have been reported in the primary trial report [2].

Additional methodological details

This section outlines the exact model specifications, adapted from the protocol and statistical analysis plan for the Bayesian analysis of the COVID STEROID 2 trial. Additional details and rationale are provided in the published protocol and statistical analysis plan [6].

Model specifications

The count outcomes (days alive without life support at day 28 and 90, days alive and out of hospital at day 90) were analysed using hurdle-negative binomial models, with both parts of the models adjusted for the stratification variables (site, age below 70 years and use of invasive mechanical ventilation at baseline).

The binary secondary outcomes (mortality at day 28 and 90 and serious adverse reactions) were analysed using logistic regression models adjusted for the stratification variables.

For all analyses, non-centred parameterisations of the intercepts (i.e., conventional intercepts) were used.

In all analyses, sites with ≤12 patients were merged with other small sites within the same country, to ease estimation of small site effects (nuisance parameters) and to avoid problems with the model diagnostics outlined below (the smallest site only included 1 patient, which would otherwise cause problems for the approximate cross-validation approach).

Priors

We used weakly informative priors centred on neutral effects and including all plausible effect sizes for the primary analysis and all *post hoc* analyses. Sensitivity analyses were conducted using sceptic priors for the treatment effects, and the same weakly informative priors for all other parameters. Further details and rationale are provided in the protocol [6].

Primary outcome

We used a *normal(mean -0.85, SD 1.5)* prior for the logistic regression intercept (corresponding to a probability distribution for the control group baseline risk of having 0 days alive without life support centred on 30% and with 95% central probability mass between 2 and 89%), a *normal(2.8, 2.25)* prior for the zero-truncated negative-binomial model intercept (corresponding to a probability distribution for the control group centred on 16.4 and with 95% central probability mass between 0.2 and 1352 days alive without life support), a *gamma(0.01, 0.01)* prior for the zero-truncated negative-binomial model shape parameter, *normal(0, 1)* priors for all other variables including the treatment effect in both parts of the model (corresponding to probability distributions for the odds ratios/incidence rate ratios centred on 1.00 and with 95% central probability mass between 0.14 and 7.10), and a sceptic *normal(0, 0.15)* prior for the treatment

effect in both parts of the model in the sensitivity analysis (corresponding to probability distributions for the odds ratios/incidence rate ratios centred on 1.00 and with 95% central probability mass between 0.75 and 1.34).

Secondary outcomes

For the three binary secondary outcomes, we used the same priors as outlined for the logistic regression part of the hurdle-negative binomial model used for analysing the primary outcome. For the secondary count outcomes, we used the same priors as for the primary outcome.

Technical details and model diagnostics

We used Stan's [7] default dynamic Hamiltonian Monte Carlo sampler with 4 chains with 5,000 warm-up iterations and 15,000 post-warm-up iterations each (60,000 posterior samples in totals), and required bulk/tail effective sample sizes (ESS) of 10,000 for the parameter of primary interest (the treatment effect). We assessed convergence using overlain density and trace plots and using the updated *Rhat* statistic, which we required to be ≤1.01 for all parameters [8,9]. We assessed model fit using graphical posterior predictive checks (of entire predicted distributions and predicted mean values, both in the full sample and in the two treatment groups separately) [10] and Pareto-smoothed importance sampling leave-one-out cross-validation (primarily focused on the effective number of parameters compared to the actual number of parameters in each models) [11].

All model diagnostics were generally adequate; bulk/tail ESS were >10,000 for the treatment effect in all models, and adequately large (>7,000) for all additional parameters. For one sensitivity analysis (days alive without life support at day 90 with non-survivors assigned zero days), the model was refit twice during the leave-one-out cross-validation procedure due to two influential observations, which were thus adequately handled. The resulting effective number of parameters was slightly larger than the actual number of parameters, but this was accepted for this sensitivity analysis as all other checks were adequate and as results were in general concordance with the other analyses. Graphical posterior predictive checks revealed that the hurdle-negative binomials models adequately captured the expected/mean values but were not able to adequately generate similarly distributed data. This was primarily due to the unexpected high proportion of patients with the maximum values for days alive without life support at day 28 and day 90 (as discussed in the primary text) [2]. While these models were not able to adequately generate similarly distributed data, the models were considered adequate as the means and associated uncertainty (which was the measures used in all subsequent calculations) were adequately captured. Additional *post hoc* models used to challenge the results to the unexpected distributions provided similar results (described below). Graphical posterior predictive checks for the logistic regression models indicated no issues.

Calculation of effect estimates

As stated in the protocol and discussed in the primary text, we planned to primarily present model results by calculating conditional adjusted estimates for a reference patient in each group with all adjustment variables to their most common value [6]. Conditional effects have the benefit of being directly interpretable for similar patients as compared to average treatment effects that reflect an average across the entire trial population and may thus be an average of different treatment effects in different subpopulations within a study. Thus, different conditional effects (i.e., effects for different reference patients) in a trial may vary, and in some cases, they may provide results that are do not appear representative of the full trial. This was unexpectedly the case in this trial, especially for serious adverse reactions, likely due to a relatively overall event rate, and possibly due to differences in event rates between sites. For the planned reference patients, the adjusted risks of serious adverse reactions in each group were very low and not representative of the full trial. While the relative effects were comparable to the full trial population, the absolute effects on serious adverse reactions were thus very low (leading to small probabilities of clinically important effects in either direction), and likely substantially underestimated the risk of serious adverse reactions in both treatment groups.

Consequently, we deviated from the planned approach to presentation and primarily present *average treatment effects*, estimated by using the joint posteriors to calculate the predicted estimates for all patients included in the analyses *assuming* that they had received either dexamethasone 12 mg or 6 mg (i.e., predicting the actual and counterfactual outcomes for all patients with either treatment). Posterior distributions of predicted *mean* values for all patients in the trial assuming assignment to either treatment were then presented and used to calculate relative and absolute differences and probabilities of different effect sizes. Of note, this approach is similar to the G-computation [12] approach used in some of the primary analyses of the trial [2], with the joint posterior distributions used in place of bootstrap resampling.

In addition to the average treatment effects, supplementary results for three representative reference patients are presented in this supplement:

1) Reference A: a patient included in the site including most patients in the trial (an Indian site), with age below 70 years and not on invasive mechanical ventilation at baseline (corresponding to the planned reference patient, i.e., all adjustment variables set to their most common values).
2) Reference B: a patient included in the site including second most patients in the trial (a Danish non-intensive care unit (ICU) site), with age below 70 years and not on invasive mechanical ventilation at baseline.
3) Reference C: a patient included in the site ranking #5 in terms of recruitment (a Danish site strictly including ICU patients with the majority being on invasive mechanical ventilation at randomisation; the top #3 site did not solely include ICU patients and the top #4 site included only ICU patients, but with the majority not on invasive mechanical ventilation at

baseline, and both only included slightly more patients), with age below 70 years and receiving invasive mechanical ventilation at baseline.

For all models and reference patients, predicted values for the count outcomes outside the actual possible ranges were possible as all models were by definition not restricted to any maximum values and as the additional, *post hoc* sensitivity analysis models were also not bounded to be strictly positive. The small proportion of values outside the valid ranges were truncated before summarisation and calculation of the presented results. Of note, as the relative treatment effect is modelled on the logarithmic scale in the hurdle-negative binomial models primarily used to analyse the primary outcome, the absolute differences for different reference patients will by definition be estimated as larger when the control group number of days alive without life support is larger, as the relative effect is assumed by the model to be consistent across reference patients. Additional models allowing different relative effects for different patients were not planned *a priori* and were consequently not used.

Additional *post hoc* models used

Similar to the primary frequentist analyses [2], we *post hoc* supplemented the planned model used for assessing the two days alive without life support outcomes with additional modelling strategies used to assess the influence of the unexpected distribution.

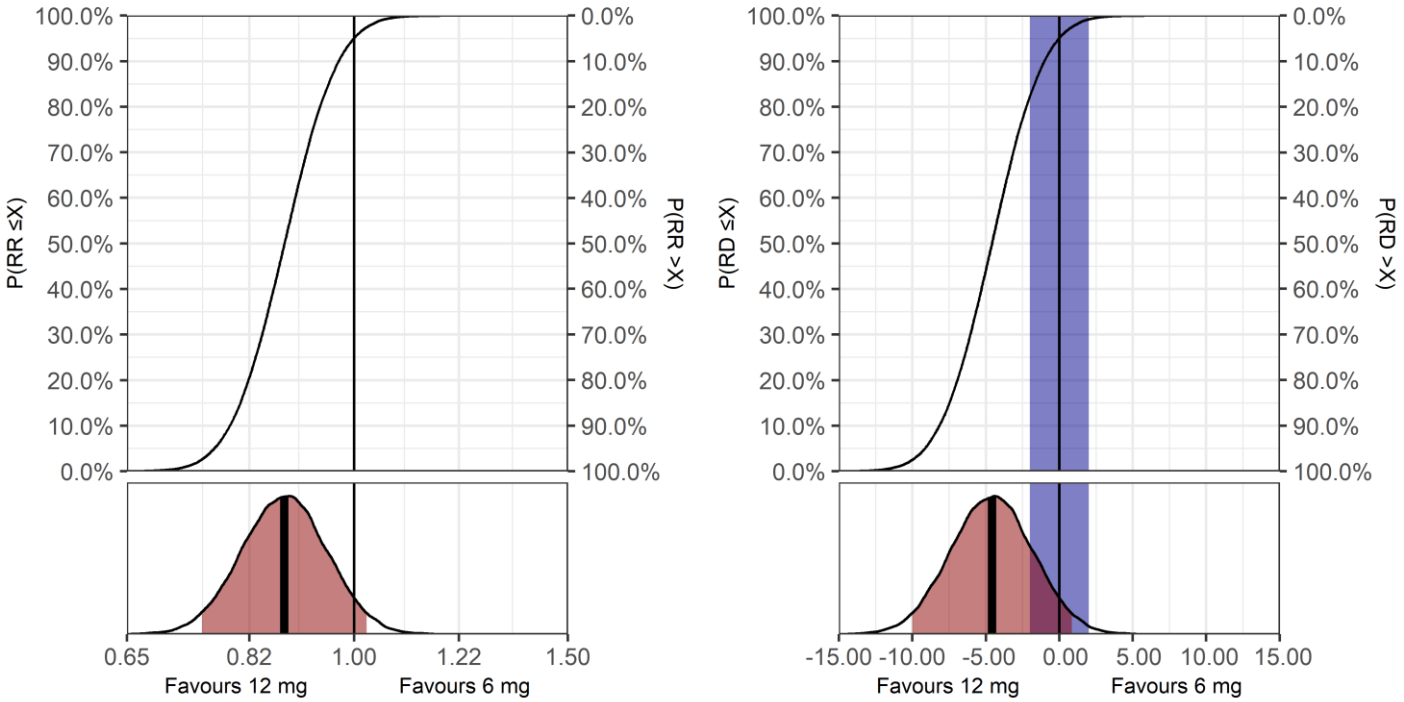
First, we used a Bayesian bootstrapping procedure, using a conventional linear regression model (which can be interpreted as using a flat, non-informative prior on all parameters) adjusted for stratification variables to estimate predicted values and differences in each group. We used 50,000 bootstrap samples with 10,000 weighted samples from the trial population using a flat Dirichlet prior in each bootstrap sample. Results were summarised and interpreted as for the other Bayesian models.

Second, we used Bayesian linear regression models adjusted for stratification variables and with weakly informative, neutral priors. For days alive without life support at day 28, a *normal(15, 15)* prior for the intercept [corresponding to a probability distribution for the control group centred on 15 and with 95% central probability mass between -14.4 and 44.4 days alive without life support] and *normal(0, 5)* priors for the treatment effect and all adjustment variables [corresponding to probability distributions centred on 0 and with 95% central probability mass between -9.8 and 9.8 days alive without life support] were used. For days alive without life support at day 90, a *normal(45, 45)* prior for the intercept [corresponding to a probability distribution for the control group centred on 45 and with 95% central probability mass between -43.2 and 133.2 days alive without life support] and *normal(0, 15)* priors for the treatment effect and all adjustment variables [corresponding to probability distributions centred on 0 and with 95% central probability mass between -29.4 and 29.4 days alive without life support] were used.

Posterior predictive checks revealed that this model, too, adequately captured the mean values with appropriate uncertainty, but was not able to generatively reproduce the sample distribution either. Of note, using the full posterior distribution for the mean value when using weakly informative priors overwhelmed by the data can be compared to the similar strategy of using conventional bootstrap resampling as used *post hoc* in the primary trial report [2].

Additional results

Figure S1. Mortality at day 90



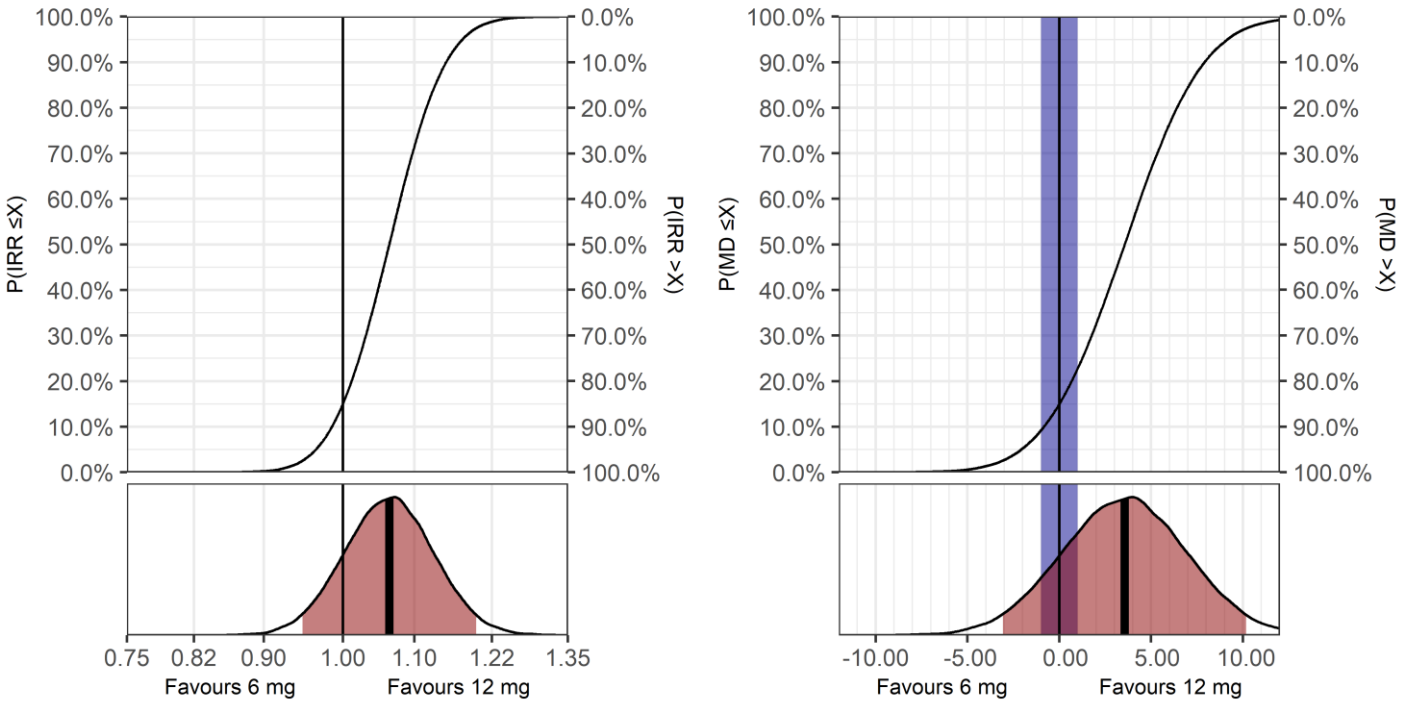
Full posterior probability distributions for the effect of the treatment on 90-day all-cause mortality (primary analysis using weakly informative priors). Left plot displays the relative difference (relative risk, RR), while the right plot displays the absolute difference (risk difference, RD) in percentage points. These results are adjusted or all stratification variables and calculated as average treatment effects, as outlined in the methods section in the main text. An RR <1 or RD <0 favours 12 mg dexamethasone; an RR >1 or RD >0 favours 6 mg dexamethasone.

The upper subplots display the cumulative posterior distributions, corresponding to the probabilities of effect sizes (X-axis) ≤ the values on the left Y-axis and > the values on the right Y-axis.

The lower subplots display the entire posterior distributions, with the bold, vertical line indicating the median value (used as the point estimate) and the area highlighted in red indicating the percentile-based 95% credible interval.

The vertical black lines represents *exactly* no difference, and the area highlighted in blue in the absolute effects plots represent effect sizes smaller than the pre-defined minimally clinically important difference of 2 percentage points in either direction [2].

Figure S2. Days alive without life support at day 90



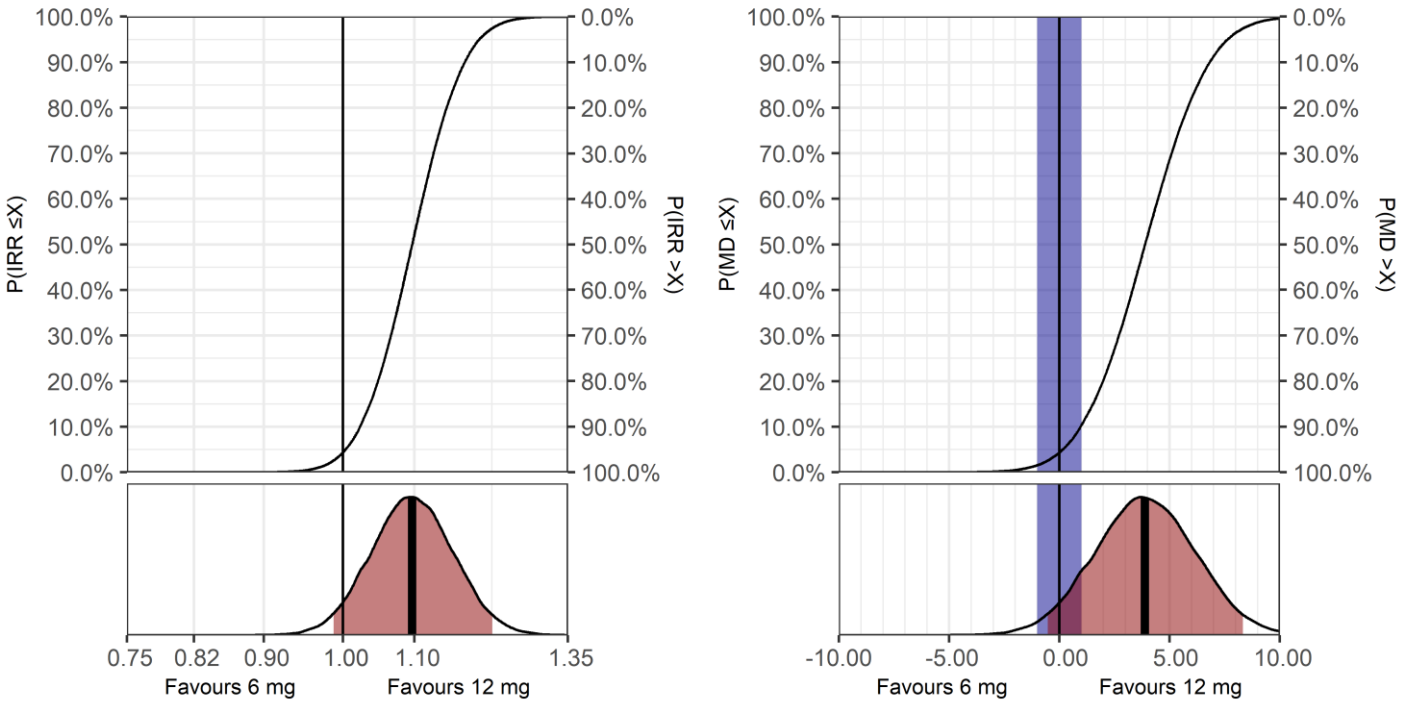
Full posterior probability distributions for the effect of the treatment on days alive without life support at day 90 (primary analysis using weakly informative priors). Left plot displays the relative difference (incidence rate ratio, IRR), while the right plot displays the absolute difference (mean difference, MD) in days. These results are adjusted for all stratification variables and calculated as average treatment effects, as outlined in the methods section in the main text. An IRR >1 or MD >0 favours 12 mg dexamethasone; an IRR <1 or MD <0 favours 6 mg dexamethasone.

The upper subplots display the cumulative posterior distributions, corresponding to the probabilities of effect sizes (X-axis) ≤ the values on the left Y-axis and > the values on the right Y-axis.

The lower subplots display the entire posterior distributions, with the bold, vertical line indicating the median value (used as the point estimate) and the area highlighted in red indicating the percentile-based 95% credible interval.

The vertical black lines represents *exactly* no difference, and the area highlighted in blue in the absolute effects plots represent effect sizes smaller than the pre-defined minimally clinically important difference of 1 day in either direction [2].

Figure S3. Days alive and out of hospital at day 90



Full posterior probability distributions for the effect of the treatment on days alive and out of hospital at day 90 (primary analysis using weakly informative priors). Left plot displays the relative difference (incidence rate ratio, IRR), while the right plot displays the absolute difference (mean difference, MD) in days. These results are adjusted for all stratification variables and calculated as average treatment effects, as outlined in the methods section in the main text. An IRR >1 or MD >0 favours 12 mg dexamethasone; an IRR <1 or MD <0 favours 6 mg dexamethasone.

The upper subplots display the cumulative posterior distributions, corresponding to the probabilities of effect sizes (X-axis) ≤ the values on the left Y-axis and > the values on the right Y-axis.

The lower subplots display the entire posterior distributions, with the bold, vertical line indicating the median value (used as the point estimate) and the area highlighted in red indicating the percentile-based 95% credible interval.

The vertical black lines represents *exactly* no difference, and the area highlighted in blue in the absolute effects plots represent effect sizes smaller than the pre-defined minimally clinically important difference of 1 day in either direction [2].

Table S1. Additional descriptive outcome data

Variable	Dexamethasone 12 mg (n=497)	Dexamethasone 6 mg (n=485)
Days alive without life support at day 28 (non-survivors assigned 0 days)[1]	22.0 (0.0 to 28.0) days	20.0 (0.0 to 28.0) days
Days alive without life support at day 28 (best/worst-case)[2]	22.0 (6.0 to 28.0) days	20.0 (4.0 to 28.0) days
Days alive without life support at day 28 (worst/best-case)[3]	22.0 (5.0 to 28.0) days	21.0 (4.0 to 28.0) days
Days alive without life support at day 90 (non-survivors assigned 0 days)[4]	84.0 (0.0 to 90.0) days	80.0 (0.0 to 90.0) days
Days alive and out of hospital at day 90 (non-survivors assigned 0 days)[5]	60.5 (0.0 to 78.0) days	48.0 (0.0 to 76.0) days

[1] Data for days alive without life support at day 28 were missing in 11 patients (6 patients in the dexamethasone 12 mg group and 5 patients in the dexamethasone 6 mg group).

[2] Assuming that patients with missing data were alive without life support in the 12 mg group and not alive without life support in the 6 mg group on all days not accounted for.

[3] Assuming that patients with missing data were not alive without life support in the 12 mg group and alive without life support in the 6 mg group on all days not accounted for.

[4] Data for days alive without life support at day 90 were missing in 15 patients (7 patients in the dexamethasone 12 mg group and 8 patients in the dexamethasone 6 mg group).

[5] Data for days alive and out of hospital at day 90 were missing in 14 patients (7 in each group).

Table S2. *Post hoc* analyses of days alive without life support at day 28

Analysis	Effect estimates				Probability of effects with 12 mg dexamethasone				
	Dexamethasone 12 mg	Dexamethasone 6 mg	Relative difference	Absolute difference	Any benefit	Any harm	Clinically important benefit	Clinically important harm	No clinically important difference
Non-survivors assigned 0 days	Mean: 16.5 (15.4 to 17.7) days	Mean: 15.0 (13.9 to 16.1) days	IRR: 1.11 (1.00 to 1.22)	MD: 1.6 (0.0 to 3.1) days	97.8%	2.2%	77.4%	0.1%	22.6%
Best/worst-case analysis	Mean: 17.9 (16.8 to 19.1) days	Mean: 16.4 (15.3 to 17.5) days	IRR: 1.09 (1.00 to 1.20)	MD: 1.5 (-0.1 to 3.1) days	97.1%	2.9%	74.9%	0.1%	25.0%
Worst/best-case analysis	Mean: 17.6 (16.5 to 18.8) days	Mean: 16.6 (15.5 to 17.7) days	IRR: 1.06 (0.97 to 1.17)	MD: 1.1 (-0.5 to 2.7) days	90.6%	9.4%	53.3%	0.6%	46.2%
Bayesian bootstrap	Mean: 17.7 (16.7 to 18.6) days	Mean: 16.4 (15.4 to 17.4) days	IRR: 1.08 (1.00 to 1.16)	MD: 1.3 (-0.1 to 2.6) days	96.9%	3.1%	64.8%	0.0%	35.2%
Linear model	Mean: 17.6 (16.7 to 18.5) days	Mean: 16.4 (15.5 to 17.3) days	IRR: 1.08 (1.00 to 1.16)	MD: 1.2 (-0.1 to 2.5) days	96.9%	3.1%	63.5%	0.0%	36.5%

Post hoc analyses using different outcome definitions (non-survivors assigned 0 days), best/worst- and worst/best- case analysis and different modelling strategies for the primary outcome. All analyses were conducted in the 491 (12 mg group) versus 480 (6 mg group) patients with available data for the primary outcome, except the best/worst- and worst/best-case analyses where all 497 versus 485 patients were included (after imputing missing days as alive without life support in the 12 mg group and not alive without life support in the 6 mg group, and vice versa). All analyses were adjusted for the stratification variables, and effect sizes are presented as average treatment effects as outlined in the methods section (main text), summarised using median posterior values as point estimates and percentile-based 95% credible intervals (CrIs). Results estimated for reference patients are presented in **Table S3**.

Any benefit is the probability of a MD >0 days (IRR >1); any harm is the probability of a MD <0 days (IRR <1); no clinically important difference is the probability of an absolute MD <1 days; clinically important benefit/harm are probabilities of effect sizes larger than no clinically important difference in either direction. All definitions of clinically important effect sizes were pre-specified in the protocol [2].

Abbreviations: IRR: incidence rate ratio (>1 favours 12 mg); MD: mean difference (>0 favours 12 mg).

Table S3. Effect estimates for days alive without life support at day 28 in different analyses for reference patients

Analysis	Effect estimates				Probability of effects with 12 mg dexamethasone				
	Dexamethasone 12 mg	**Dexamethasone 6 mg**	**Relative difference**	**Absolute difference**	**Any benefit**	**Any harm**	**Clinically important benefit**	**Clinically important harm**	**No clinically important difference**
Reference patient A (largest Indian site, age below 70 years, not on invasive mechanical ventilation at baseline)									
Primary analysis	Mean: 20.9 (18.1 to 24.3) days	Mean: 19.4 (16.7 to 22.5) days	IRR: 1.08 (0.98 to 1.19)	MD: 1.6 (-0.3 to 3.5) days	94.9%	5.1%	72.6%	0.4%	27.0%
Sceptic prior	Mean: 20.9 (18.0 to 24.2) days	Mean: 19.4 (16.8 to 22.5) days	IRR: 1.08 (0.99 to 1.17)	MD: 1.5 (-0.3 to 3.3) days	95.1%	4.9%	70.6%	0.3%	29.1%
Non-survivors assigned 0 days	Mean: 20.2 (17.1 to 23.4) days	Mean: 18.1 (15.1 to 21.3) days	IRR: 1.11 (1.00 to 1.24)	MD: 2.0 (0.0 to 4.1) days	97.7%	2.3%	84.7%	0.1%	15.1%
Best/worst-case analysis	Mean: 21.0 (18.1 to 24.4) days	Mean: 19.1 (16.5 to 22.2) days	IRR: 1.10 (1.00 to 1.21)	MD: 1.9 (0.0 to 3.8) days	97.4%	2.6%	81.5%	0.1%	18.3%
Worst/best-case analysis	Mean: 20.8 (18.0 to 24.2) days	Mean: 19.5 (16.8 to 22.6) days	IRR: 1.07 (0.97 to 1.17)	MD: 1.3 (-0.6 to 3.3) days	91.7%	8.3%	63.2%	0.8%	36.0%
Bayesian bootstrap	Mean: 20.7 (18.5 to 22.7) days	Mean: 19.4 (17.2 to 21.5) days	IRR: 1.06 (1.00 to 1.14)	MD: 1.3 (-0.1 to 2.6) days	96.9%	3.1%	64.9%	0.0%	35.0%
Linear model	Mean: 20.6 (18.6 to 22.7) days	Mean: 19.4 (17.4 to 21.4) days	IRR: 1.06 (1.00 to 1.14)	MD: 1.2 (-0.1 to 2.5) days	96.9%	3.1%	63.7%	0.0%	36.2%

Reference patient B (largest Danish non-ICU site, age below 70 years, not on invasive mechanical ventilation at baseline)									
Primary analysis	Mean: 23.6 (20.0 to 28.0) days	Mean: 21.8 (18.4 to 25.9) days	IRR: 1.08 (0.99 to 1.19)	MD: 1.8 (-0.2 to 3.9) days	95.7%	4.0%	78.1%	0.3%	21.6%
Sceptic prior	Mean: 23.5 (19.9 to 27.9) days	Mean: 21.8 (18.5 to 25.9) days	IRR: 1.08 (0.99 to 1.17)	MD: 1.7 (-0.3 to 3.6) days	95.2%	4.5%	75.0%	0.4%	24.6%
Non-survivors assigned 0 days	Mean: 23.3 (20.2 to 26.7) days	Mean: 21.6 (18.5 to 24.9) days	IRR: 1.08 (1.00 to 1.17)	MD: 1.7 (0.0 to 3.5) days	97.7%	2.3%	80.0%	0.1%	19.9%
Best/worst-case analysis	Mean: 23.5 (20.0 to 27.9) days	Mean: 21.4 (18.2 to 25.4) days	IRR: 1.10 (1.00 to 1.20)	MD: 2.1 (0.1 to 4.2) days	97.8%	2.1%	85.3%	0.1%	14.6%
Worst/best-case analysis	Mean: 23.3 (19.7 to 27.6) days	Mean: 21.7 (18.4 to 25.7) days	IRR: 1.07 (0.98 to 1.17)	MD: 1.5 (-0.5 to 3.6) days	93.3%	6.5%	69.7%	0.7%	29.6%
Bayesian bootstrap	Mean: 23.4 (20.7 to 25.8) days	Mean: 22.1 (19.4 to 24.6) days	IRR: 1.06 (1.00 to 1.12)	MD: 1.3 (-0.1 to 2.6) days	96.8%	3.2%	64.8%	0.0%	35.1%
Linear model	Mean: 23.2 (20.7 to 25.7) days	Mean: 22.0 (19.5 to 24.5) days	IRR: 1.06 (1.00 to 1.12)	MD: 1.2 (-0.1 to 2.5) days	96.9%	3.1%	63.7%	0.0%	36.2%
Reference patient C (large Danish ICU-only site, age below 70 years, on invasive mechanical ventilation at baseline)									
Primary analysis	Mean: 12.2 (9.0 to 16.0) days	Mean: 11.4 (8.5 to 15.0) days	IRR: 1.07 (0.91 to 1.25)	MD: 0.8 (-1.0 to 2.6) days	80.8%	19.2%	41.4%	2.7%	55.9%
Sceptic prior	Mean: 12.2 (9.2 to 16.0) days	Mean: 11.4 (8.5 to 14.9) days	IRR: 1.07 (0.96 to 1.20)	MD: 0.8 (-0.5 to 2.2) days	89.3%	10.7%	39.4%	0.3%	60.3%
Non-survivors assigned 0 days	Mean: 12.4 (9.9 to 15.0) days	Mean: 11.2 (8.7 to 13.8) days	IRR: 1.11 (1.00 to 1.24)	MD: 1.2 (0.0 to 2.5) days	97.8%	2.2%	64.1%	0.0%	35.9%

Best/worst-case analysis	Mean: 12.5 (9.3 to 16.3) days	Mean: 11.4 (8.5 to 14.9) days	IRR: 1.09 (0.93 to 1.28)	MD: 1.1 (-0.8 to 3.0) days	87.6%	12.4%	53.5%	1.4%	45.0%
Worst/best-case analysis	Mean: 12.0 (8.9 to 15.7) days	Mean: 11.5 (8.5 to 15.0) days	IRR: 1.05 (0.90 to 1.23)	MD: 0.6 (-1.3 to 2.4) days	73.8%	26.2%	31.6%	4.5%	63.9%
Bayesian bootstrap	Mean: 12.1 (9.3 to 14.8) days	Mean: 10.8 (8.0 to 13.5) days	IRR: 1.12 (1.00 to 1.27)	MD: 1.3 (-0.1 to 2.6) days	97.0%	3.0%	64.6%	0.0%	35.3%
Linear model	Mean: 12.2 (9.5 to 14.9) days	Mean: 11.0 (8.2 to 13.7) days	IRR: 1.11 (0.99 to 1.26)	MD: 1.2 (-0.1 to 2.5) days	96.9%	3.1%	63.7%	0.0%	36.2%

Results from the analyses primarily presented in **Table 2** (main text) and **Table S2** (average treatment effects for the entire trial population) with adjusted (for stratification variables) estimated presented in this table for three reference patients (described in detail above); of note, the underlying analyses are the same, but conditional effects for different reference patients are presented in this table. The probabilities of any benefit and any harm do not add to exactly 100% for all analyses in all reference patients, due to truncation of a small proportion of predicted values outside the valid range; where values were truncated for both treatment arms, the estimated effect is exactly neutral. For additional details regarding the number of patients in each analysis and description of the analyses, see the rest of the text and the footnotes to **Table 2** (main text) and **Table S2**.

Results are summarised using median posterior values as point estimates and percentile-based 95% credible intervals (CrIs).

Any benefit is the probability of a MD >0 days (IRR >1); any harm is the probability of a MD <0 days (IRR <1); no clinically important difference is the probability of an absolute MD <1 days; clinically important benefit/harm are probabilities of effect sizes larger than no clinically important difference in either direction. All definitions of clinically important effect sizes were pre-specified in the protocol [2].

Abbreviations: ICU: intensive care unit; IRR: incidence rate ratio (>1 favours 12 mg); MD: mean difference (>0 favours 12 mg).

Table S4. Effect estimates for the binary secondary outcomes in different analyses for reference patients

Analysis	Effect estimates				Probability of effects with 12 mg dexamethasone				
	Dexamethasone 12 mg	**Dexamethasone 6 mg**	**Relative difference**	**Absolute difference**	**Any benefit**	**Any harm**	**Clinically important benefit**	**Clinically important harm**	**No clinically important difference**
Reference patient A (largest Indian site, age below 70 years, not on invasive mechanical ventilation at baseline)									
Serious adverse reactions at day 28 – primary analysis	Prob.: 0.9% (0.2% to 2.6%)	Prob.: 1.1% (0.3% to 3.1%)	RR: 0.82 (0.55 to 1.21)	RD: -0.2% (-0.9% to 0.2%)	84.1%	15.9%	0.1%	0.0%	99.9%
Serious adverse reactions at day 28 – sceptic prior	Prob.: 1.0% (0.3% to 2.7%)	Prob.: 1.0% (0.3% to 2.9%)	RR: 0.93 (0.74 to 1.18)	RD: -0.1% (-0.5% to 0.2%)	72.3%	27.7%	0.0%	0.0%	100.0%
Mortality at day 28 – primary analysis	Prob.: 26.3% (18.6% to 35.6%)	Prob.: 31.4% (22.7% to 41.4%)	RR: 0.84 (0.67 to 1.04)	RD: -5.0% (-11.3% to 1.0%)	94.8%	5.2%	83.7%	1.2%	15.2%
Mortality at day 28 – sceptic prior	Prob.: 27.6% (19.8% to 36.8%)	Prob.: 30.1% (21.8% to 39.6%)	RR: 0.92 (0.79 to 1.06)	RD: -2.5% (-6.9% to 1.8%)	87.2%	12.8%	58.7%	1.9%	39.4%
Mortality at day 90 – primary analysis	Prob.: 26.9% (19.2% to 36.1%)	Prob.: 31.9% (23.3% to 41.8%)	RR: 0.84 (0.69 to 1.03)	RD: -4.9% (-11.0% to 0.9%)	95.1%	4.9%	83.5%	0.9%	15.6%
Mortality at day 90 – sceptic prior	Prob.: 28.1% (20.3% to 37.3%)	Prob.: 30.7% (22.4% to 40.2%)	RR: 0.92 (0.79 to 1.06)	RD: -2.6% (-6.9% to 1.7%)	88.3%	11.7%	60.1%	1.7%	38.2%
Reference patient B (largest Danish non-ICU site, age below 70 years, not on invasive mechanical ventilation at baseline)									
Serious adverse reactions at day 28 – primary analysis	Prob.: 5.7% (2.5% to 11.4%)	Prob.: 6.9% (3.1% to 13.5%)	RR: 0.83 (0.57 to 1.20)	RD: -1.1% (-4.3% to 1.2%)	84.1%	15.9%	24.9%	0.7%	74.4%

Serious adverse reactions at day 28 – sceptic prior	Prob.: 6.1% (2.7% to 12.0%)	Prob.: 6.5% (2.9% to 12.7%)	RR: 0.94 (0.75 to 1.17)	RD: -0.4% (-2.2% to 1.1%)	72.3%	27.7%	3.5%	0.2%	96.3%
Mortality at day 28 – primary analysis	Prob.: 12.0% (6.8% to 19.7%)	Prob.: 14.8% (8.6% to 23.7%)	RR: 0.81 (0.62 to 1.04)	RD: -2.8% (-7.0% to 0.6%)	94.8%	5.2%	67.1%	0.3%	32.5%
Mortality at day 28 – sceptic prior	Prob.: 12.7% (7.3% to 20.4%)	Prob.: 14.1% (8.2% to 22.5%)	RR: 0.90 (0.75 to 1.08)	RD: -1.4% (-4.2% to 1.0%)	87.2%	12.8%	31.4%	0.4%	68.2%
Mortality at day 90 – primary analysis	Prob.: 13.9% (8.1% to 22.1%)	Prob.: 17.1% (10.2% to 26.6%)	RR: 0.82 (0.64 to 1.04)	RD: -3.1% (-7.5% to 0.6%)	95.1%	4.9%	71.2%	0.3%	28.4%
Mortality at day 90 – sceptic prior	Prob.: 14.7% (8.8% to 23.0%)	Prob.: 16.3% (9.8% to 25.3%)	RR: 0.90 (0.76 to 1.07)	RD: -1.6% (-4.7% to 1.1%)	88.3%	11.7%	38.1%	0.5%	61.4%
Reference patient C (large Danish ICU-only site, age below 70 years, on invasive mechanical ventilation at baseline)									
Serious adverse reactions at day 28 – primary analysis	Prob.: 8.9% (3.9% to 17.3%)	Prob.: 10.6% (4.7% to 20.5%)	RR: 0.83 (0.58 to 1.19)	RD: -1.7% (-6.2% to 1.8%)	84.1%	15.9%	42.4%	1.9%	55.6%
Serious adverse reactions at day 28 – sceptic prior	Prob.: 9.4% (4.2% to 18.1%)	Prob.: 10.1% (4.5% to 19.1%)	RR: 0.94 (0.76 to 1.16)	RD: -0.6% (-3.1% to 1.5%)	72.3%	27.7%	11.6%	1.1%	87.3%
Mortality at day 28 – primary analysis	Prob.: 14.6% (7.8% to 24.7%)	Prob.: 18.0% (9.7% to 29.7%)	RR: 0.81 (0.63 to 1.04)	RD: -3.3% (-8.3% to 0.7%)	94.8%	5.2%	72.6%	0.5%	26.9%
Mortality at day 28 – sceptic prior	Prob.: 15.4% (8.2% to 25.8%)	Prob.: 17.0% (9.2% to 28.2%)	RR: 0.90 (0.76 to 1.08)	RD: -1.6% (-5.0% to 1.2%)	87.2%	12.8%	39.2%	0.7%	60.1%
Mortality at day 90 – primary analysis	Prob.: 21.9% (13.0% to 33.7%)	Prob.: 26.3% (15.8% to 39.5%)	RR: 0.83 (0.67 to 1.03)	RD: -4.3% (-10.2% to 0.8%)	95.1%	4.9%	80.7%	0.7%	18.6%

Mortality at day 90 – sceptic prior	Prob.: 22.9% (13.6% to 34.9%)	Prob.: 25.2% (15.2% to 37.8%)	RR: 0.91 (0.78 to 1.06)	RD: -2.2% (-6.2% to 1.5%)	88.3%	11.7%	54.2%	1.2%	44.5%

Results from the analyses primarily presented in **Table 2** (main text; average treatment effects for the entire trial population, main text) with adjusted (for stratification variables) estimated presented in this table for three reference patients (described in detail above); of note, the underlying analyses are the same, but conditional effects for different reference patients are presented in this table. Analyses were conducted in all patients with available data (see **Table 1**, main text). Results are summarised using median posterior values as point estimates and percentile-based 95% credible intervals (CrIs).

Any benefit is the probability of a RD <0 percentage points (RR <1); any harm is the probability of a RD >0 percentage points (RR >1); no clinically important difference is the probability of an absolute RD <2 percentage points; clinically important benefit/harm are probabilities of effect sizes larger than no clinically important difference in either direction. All definitions of clinically important effect sizes were pre-specified in the protocol [2].

Abbreviations: ICU: intensive care unit; prob.: probability; RD: risk difference in percentage points (<0 favours 12 mg); RR: relative risk (<1 favours 12 mg).

Table S5. *Post hoc* analyses of days alive without life support and days alive and out of hospital at day 90

Analysis	Effect estimates				Probability of effects with 12 mg dexamethasone				
	Dexamethasone 12 mg	Dexamethasone 6 mg	Relative difference	Absolute difference	Any benefit	Any harm	Clinically important benefit	Clinically important harm	No clinically important difference
Days alive without life support at day 90									
Non-survivors assigned 0 days	Mean: 57.0 (53.6 to 60.3) days	Mean: 52.2 (48.8 to 55.7) days	IRR: 1.09 (1.00 to 1.19)	MD: 4.8 (0.0 to 9.5) days	97.5%	2.5%	94.0%	0.9%	5.1%
Bayesian bootstrap	Mean: 58.9 (55.8 to 61.9) days	Mean: 54.6 (51.2 to 58.0) days	IRR: 1.08 (0.99 to 1.17)	MD: 4.2 (-0.3 to 8.9) days	96.5%	3.5%	91.8%	1.3%	6.9%
Linear model	Mean: 58.9 (55.8 to 62.0) days	Mean: 54.7 (51.5 to 57.9) days	IRR: 1.08 (1.00 to 1.17)	MD: 4.2 (-0.3 to 8.6) days	96.8%	3.2%	92.2%	1.1%	6.7%
Days alive and out of hospital at day 90									
Non-survivors assigned 0 days	Mean: 43.7 (40.6 to 46.9) days	Mean: 40.0 (37.0 to 43.2) days	IRR: 1.09 (0.98 to 1.21)	MD: 3.7 (-0.7 to 8.0) days	95.1%	4.9%	88.5%	1.8%	9.7%

Post hoc analyses using different outcome definitions (non-survivors assigned 0 days) and different modelling strategies for days alive without life support and days alive and out of hospital at day 90. All analyses were conducted in all patients with available data (**Table 1**, main text). All analyses were adjusted for the stratification variables, and effect sizes are presented as average treatment effects as outlined in the methods section (main text), summarised using median posterior values as point estimates and percentile-based 95% credible intervals (CrIs). Results estimated for reference patients are presented in **Tables S6 and S7**.

Any benefit is the probability of a MD >0 days (IRR >1); any harm is the probability of a MD <0 days (IRR <1); no clinically important difference is the probability of an absolute MD <1 days; clinically important benefit/harm are probabilities of effect sizes larger than no clinically important difference in either direction. All definitions of clinically important effect sizes were pre-specified in the protocol [2].

Abbreviations: IRR: incidence rate ratio (>1 favours 12 mg); MD: mean difference (>0 favours 12 mg).

Table S6. Effect estimates for days alive without life support at day 90 in different analyses for reference patients

Analysis	Effect estimates				Probability of effects with 12 mg dexamethasone				
	Dexamethasone 12 mg	**Dexamethasone 6 mg**	**Relative difference**	**Absolute difference**	**Any benefit**	**Any harm**	**Clinically important benefit**	**Clinically important harm**	**No clinically important difference**
Reference patient A (largest Indian site, age below 70 years, not on invasive mechanical ventilation at baseline)									
Primary analysis	Mean: 65.1 (53.9 to 79.4) days	Mean: 60.9 (50.3 to 74.4) days	IRR: 1.07 (0.95 to 1.21)	MD: 4.2 (-3.6 to 12.2) days	85.5%	14.5%	78.9%	9.5%	11.5%
Sceptic prior	Mean: 64.9 (53.7 to 78.9) days	Mean: 61.3 (50.7 to 74.7) days	IRR: 1.06 (0.95 to 1.18)	MD: 3.5 (-3.5 to 10.9) days	83.8%	16.2%	76.2%	10.4%	13.4%
Non-survivors assigned 0 days	Mean: 65.3 (56.9 to 72.9) days	Mean: 59.9 (50.9 to 68.1) days	IRR: 1.09 (1.00 to 1.20)	MD: 5.5 (0.0 to 11.1) days	97.5%	2.5%	94.6%	1.1%	4.4%
Bayesian bootstrap	Mean: 64.7 (56.9 to 72.1) days	Mean: 60.4 (52.4 to 68.0) days	IRR: 1.07 (0.99 to 1.16)	MD: 4.3 (-0.3 to 9.0) days	96.6%	3.4%	91.9%	1.2%	6.9%
Linear model	Mean: 64.5 (57.5 to 71.5) days	Mean: 60.2 (53.2 to 67.3) days	IRR: 1.07 (1.00 to 1.15)	MD: 4.2 (-0.3 to 8.7) days	96.8%	3.2%	92.2%	1.1%	6.7%
Reference patient B (largest Danish non-ICU site, age below 70 years, not on invasive mechanical ventilation at baseline)									
Primary analysis	Mean: 77.5 (62.4 to 90.0) days	Mean: 72.5 (58.1 to 90.0) days	IRR: 1.06 (0.95 to 1.20)	MD: 4.6 (-3.9 to 13.5) days	83.7%	13.3%	78.1%	9.1%	12.8%
Sceptic prior	Mean: 77.3 (62.2 to 90.0) days	Mean: 73.0 (58.7 to 90.0) days	IRR: 1.05 (0.95 to 1.18)	MD: 3.9 (-3.9 to 12.3) days	81.6%	15.2%	75.1%	10.4%	14.5%

Non-survivors assigned 0 days	Mean: 75.7 (68.0 to 82.2) days	Mean: 71.9 (63.3 to 79.0) days	IRR: 1.05 (1.00 to 1.12)	MD: 3.8 (0.1 to 8.1) days	97.7%	2.3%	92.9%	0.7%	6.5%
Bayesian bootstrap	Mean: 77.9 (69.4 to 85.5) days	Mean: 73.5 (64.7 to 81.6) days	IRR: 1.06 (1.00 to 1.13)	MD: 4.3 (-0.3 to 8.9) days	96.6%	3.4%	91.9%	1.2%	6.9%
Linear model	Mean: 77.0 (68.4 to 85.5) days	Mean: 72.8 (64.1 to 81.4) days	IRR: 1.06 (1.00 to 1.13)	MD: 4.2 (-0.3 to 8.7) days	96.8%	3.2%	92.2%	1.1%	6.7%
Reference patient C (large Danish ICU-only site, age below 70 years, on invasive mechanical ventilation at baseline)									
Primary analysis	Mean: 62.3 (45.3 to 85.1) days	Mean: 58.2 (42.1 to 80.0) days	IRR: 1.07 (0.91 to 1.26)	MD: 4.0 (-5.6 to 14.0) days	79.5%	20.2%	73.3%	14.9%	11.7%
Sceptic prior	Mean: 61.9 (45.4 to 84.6) days	Mean: 58.5 (42.7 to 80.0) days	IRR: 1.06 (0.93 to 1.20)	MD: 3.4 (-4.1 to 11.4) days	81.0%	18.7%	73.2%	12.5%	14.3%
Non-survivors assigned 0 days	Mean: 59.6 (50.0 to 67.4) days	Mean: 55.3 (44.9 to 64.1) days	IRR: 1.08 (1.00 to 1.18)	MD: 4.2 (0.0 to 8.8) days	97.6%	2.4%	93.4%	0.8%	5.8%
Bayesian bootstrap	Mean: 59.1 (48.4 to 68.4) days	Mean: 54.8 (43.9 to 64.3) days	IRR: 1.08 (0.99 to 1.18)	MD: 4.3 (-0.3 to 8.9) days	96.6%	3.4%	91.9%	1.2%	6.9%
Linear model	Mean: 58.5 (49.1 to 67.9) days	Mean: 54.3 (44.8 to 63.8) days	IRR: 1.08 (1.00 to 1.17)	MD: 4.2 (-0.3 to 8.7) days	96.8%	3.2%	92.2%	1.1%	6.7%

Results from the analyses primarily presented in **Table 2** (main text) and **Table S5** (average treatment effects for the entire trial population) with adjusted (for stratification variables) estimated presented in this table for three reference patients (described in detail above); of note, the underlying analyses are the same, but conditional effects for different reference patients are presented in this table. The probabilities of any benefit and any harm do not add to exactly 100% for all analyses in all reference patients, due to truncation of a small proportion of predicted values outside the valid range; where values were truncated for both treatment arms, the estimated effect is exactly neutral. For additional details regarding the number of patients in each analysis and description of the analyses, see the rest of the text and the footnotes to **Table 2** (main text) and **Table S5**.

Results are summarised using median posterior values as point estimates and percentile-based 95% credible intervals (CrIs).

Any benefit is the probability of a MD >0 days (IRR >1); any harm is the probability of a MD <0 days (IRR <1); no clinically important difference is the probability of an absolute MD <1 days; clinically important benefit/harm are probabilities of effect sizes larger than no clinically important difference in either direction. All definitions of clinically important effect sizes were pre-specified in the protocol [2].

Abbreviations: ICU: intensive care unit; IRR: incidence rate ratio (>1 favours 12 mg); MD: mean difference (>0 favours 12 mg).

Table S7. Effect estimates for days alive and out of hospital at day 90 in different analyses for reference patients

Analysis	Effect estimates				Probability of effects with 12 mg dexamethasone				
	Dexamethasone 12 mg	**Dexamethasone 6 mg**	**Relative difference**	**Absolute difference**	**Any benefit**	**Any harm**	**Clinically important benefit**	**Clinically important harm**	**No clinically important difference**
Reference patient A (largest Indian site, age below 70 years, not on invasive mechanical ventilation at baseline)									
Primary analysis	Mean: 56.1 (47.8 to 64.6) days	Mean: 51.2 (42.5 to 60.2) days	IRR: 1.09 (0.99 to 1.22)	MD: 4.8 (-0.8 to 10.5) days	95.5%	4.5%	90.8%	2.0%	7.1%
Sceptic prior	Mean: 54.9 (46.7 to 63.4) days	Mean: 52.4 (44.0 to 61.1) days	IRR: 1.05 (0.96 to 1.14)	MD: 2.5 (-2.1 to 7.1) days	85.8%	14.2%	74.0%	6.7%	19.3%
Reference patient B (largest Danish non-ICU site, age below 70 years, not on invasive mechanical ventilation at baseline)									
Primary analysis	Mean: 66.1 (57.3 to 75.4) days	Mean: 63.0 (53.6 to 72.9) days	IRR: 1.05 (0.97 to 1.14)	MD: 3.1 (-1.9 to 8.4) days	88.5%	11.5%	79.2%	5.5%	15.4%
Sceptic prior	Mean: 65.4 (56.6 to 74.8) days	Mean: 63.8 (54.6 to 73.4) days	IRR: 1.03 (0.96 to 1.10)	MD: 1.6 (-2.9 to 6.1) days	76.2%	23.8%	60.8%	12.5%	26.7%
Reference patient C (large Danish ICU-only site, age below 70 years, on invasive mechanical ventilation at baseline)									
Primary analysis	Mean: 48.1 (38.9 to 57.6) days	Mean: 44.3 (34.5 to 54.4) days	IRR: 1.09 (0.98 to 1.22)	MD: 3.8 (-0.8 to 8.6) days	94.6%	5.4%	88.0%	2.0%	10.0%
Sceptic prior	Mean: 47.3 (38.0 to 56.9) days	Mean: 45.3 (35.7 to 55.2) days	IRR: 1.04 (0.96 to 1.14)	MD: 2.0 (-1.8 to 5.8) days	84.4%	15.6%	69.2%	6.3%	24.5%

Results from the analyses primarily presented in **Table 2** (main text) and **Table S5** (average treatment effects for the entire trial population) with adjusted (for stratification variables) estimated presented in this table for three reference patients (described in detail above); of note, the underlying analyses are the same, but conditional effects for different reference patients are presented in this table. The probabilities of any benefit and any harm do not add to exactly 100% for all

analyses in all reference patients, due to truncation of a small proportion of predicted values outside the valid range; where values were truncated for both treatment arms, the estimated effect is exactly neutral. For additional details regarding the number of patients in each analysis and description of the analyses, see the rest of the text and the footnotes to **Table 2** (main text) and **Table S5**.

Results are summarised using median posterior values as point estimates and percentile-based 95% credible intervals (CrIs).

Any benefit is the probability of a MD >0 days (IRR >1); any harm is the probability of a MD <0 days (IRR <1); no clinically important difference is the probability of an absolute MD <1 days; clinically important benefit/harm are probabilities of effect sizes larger than no clinically important difference in either direction. All definitions of clinically important effect sizes were pre-specified in the protocol [2].

Abbreviations: ICU: intensive care unit; IRR: incidence rate ratio (>1 favours 12 mg); MD: mean difference (>0 favours 12 mg).

STROBE checklist

Completed Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement checklist [13] for this manuscript. Additional details can be found in the primary trial protocol, the primary trial report and in the protocol and statistical analysis plan for the secondary Bayesian analyses [1,2,6]. To avoid unnecessary duplication of items, this manuscript is reported according to STROBE despite the randomised design, as STROBE covers all applicable items not specified in the primary trial report adhering to the Consolidated Standards of Reporting Trials (CONSORT) statement [2,14].

Abbreviations: ESM: electronic supplementary material (this file).

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract – 1,4
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found – 4
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported – 5
Objectives	3	State specific objectives, including any prespecified hypotheses – 5
Methods		
Study design	4	Present key elements of study design early in the paper – 6-7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection – 6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up – 6, ESM
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed – not applicable
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable – 6-7, ESM
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group – 6, ESM, primary manuscript/protocols
Bias	9	Describe any efforts to address potential sources of bias – 6-7
Study size	10	Explain how the study size was arrived at – primary manuscript/protocol
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why – 6-8, primary manuscript/protocol
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding – 7-8, ESM, primary manuscript/protocols
		(*b*) Describe any methods used to examine subgroups and interactions – not applicable/primary manuscript/protocols
		(*c*) Explain how missing data were addressed – 8

		(*d*) If applicable, explain how loss to follow-up was addressed – 8
		(*e*) Describe any sensitivity analyses – 7-8, ESM
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed – 9, primary manuscript
		(b) Give reasons for non-participation at each stage – primary manuscript
		(c) Consider use of a flow diagram – primary manuscript
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders – Table 1, ESM, primary manuscript
		(b) Indicate number of participants with missing data for each variable of interest – Table 1, ESM
		(c) Summarise follow-up time (eg, average and total amount) – 9
Outcome data	15*	Report numbers of outcome events or summary measures over time – Table 1, ESM
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included – 9-10, ESM
		(*b*) Report category boundaries when continuous variables were categorized – 6
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period – 9-10, Table 2, ESM
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses – 9-10, ESM
Discussion		
Key results	18	Summarise key results with reference to study objectives – 11-12
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias – 11-12
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence – 11-12
Generalisability	21	Discuss the generalisability (external validity) of the study results – 11-12
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based – 13

*Give information separately for exposed and unexposed groups.

References

[1] Munch MW, Granholm A, Myatra SN, et al (2021) Higher vs. Lower Doses of Dexamethasone in Patients with COVID-19 and Severe Hypoxia (COVID STEROID 2) trial: protocol and statistical analysis plan. Acta Anaesthesiol Scand 65:834-845.

[2] The COVID STEROID 2 Trial Group (2021) Effect of 12 mg vs 6 mg of Dexamethasone on the Number of Days Alive Without Life Support in Adults With COVID-19 and Severe Hypoxemia: The COVID STEROID 2 Randomized Trial. JAMA. https://doi.org/10.1001/jama.2021.18295.

[3] Yehya N, Harhay MO, Curley MAQ, Schoenfeld DA, Reeder RW (2019) Reappraisal of ventilator-free days in critical care research. Am J Respir Crit Care Med 200:828–836.

[4] Contentin L, Ehrmann S, Giraudeau B (2014) Heterogeneity in the Definition of Mechanical Ventilation Duration and Ventilator-Free Days. Am J Respir Crit Care Med 189:998-1002.

[5] Writing Committee for the REMAP-CAP Investigators (2020) Effect of Hydrocortisone on Mortality and Organ Support in Patients With Severe COVID-19: The REMAP-CAP COVID-19 Corticosteroid Domain Randomized Clinical Trial. JAMA 324:1317-1329.

[6] Granholm A, Munch MW, Myatra SN, et al (2021) Higher vs. Lower Doses of Dexamethasone in Patients with COVID-19 and Severe Hypoxia (COVID STEROID 2) trial: protocol for a secondary Bayesian analysis. Acta Anaesthesiol Scand 65:702-710.

[7] Carpenter B, Guo J, Hoffman MD, et al (2017) Stan: A Probabilistic Programming Language. J Stat Softw. https://doi.org/10.18637/jss.v076.i01.

[8] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2014) Bayesian Data Analysis. 3rd ed. CRC Press.

[9] Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner PC (2020) Rank-Normalization, Folding, and Localization: An Improved Rhat for Assessing Convergence of MCMC. Bayesian Anal 16:667-718.

[10] Gabry J, Simpson D, Betancourt M (2019) Visualization in Bayesian workflow. J R Stat Soc A 182:389-402.

[11] Vehtari A, Gelman A, Gabry J (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput 27:1413-1432.

[12] Chatton A, Le Borgne F, Leyrat C, et al (2020) G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: a comparative simulation study. Sci Rep 10:9219.

[13] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP (2008) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. J Clin Epidemiol 61:344-349.

[14] Schulz KF, Altman DG, Moher D, CONSORT Group (2010) CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 340:c332.