Statistical review of
Effect of Dexamethasone in Hospitalized Patients with COVID-19 – Preliminary Report

July 2nd, 2020

Dr Tim Morris, PhD - Statistician, MRC Clinical Trials Unit at University College London.

Dr Darren Dahly, PhD - Principal Statistician, HRB Clinical Research Facility Cork. Senior Lecturer in Patient-Focused Research Methods, School of Public Health, University College Cork.

Professor Kerry Hood, PhD - Professor of Trials, Centre for Trials Research, Cardiff University.

Professor Simon Gates, PhD - Professor of Biostatistics and Clinical Trials, Cancer Research UK Clinical Trials Unit, Institute of Cancer and Genomic Sciences, University of Birmingham, UK

The following review has been prepared in collaboration with members of the MRC-NIHR Trials Methodology Research Partnership[1]. The reviewers named above, and other, unnamed discussants of the paper, are all qualified statisticians with experience in clinical trials. Our objective is to provide a rapid review of publications, preprints and protocols from clinical trials of COVID-19 treatments, independent of journal specific review processes. We aim to provide timely, constructive, focused, clear advice aimed at improving both the research outputs under review, as well as future studies. Given our collective expertise (clinical trial statistics) our reviews focus on the designs of the trials and other statistical content (methods, presentation and accuracy of results, inferences). This review reflects the expert opinions of the named authors, and does not imply endorsement by the MRC-NIHR Trials Methodology Research Partnership, its wider membership, or any other organization.

Here we review *Effect of Dexamethasone in Hospitalized Patients with COVID-19 – Preliminary Report* by Horby *et al*[2], published as a preprint on medRxiv on the 22nd of June, 2020.

Overall, this was a well-conducted, well-designed, practice-changing clinical trial providing the most convincing evidence to date of an efficacious treatment in hospitalised COVID-19 patients. However, there are a number of issues, primarily around the reporting of the trial, that should ideally be resolved prior to final publication. We also urge caution against over-interpreting the estimated subgroup-specific effects based on level of respiratory support, especially the suggestion of harm in the group not on oxygen or mechanical ventilation.

Study Summary

NOTE: RECOVERY is an adaptive platform trial aimed at identifying beneficial treatments for patients hospitalized with COVID-19. The fact that it's a *platform* trial means that multiple treatments are being simultaneously tested in different study arms; whereas the term *adaptive* indicates that there is scope in the study protocol to change the available treatments, and apparently the targeted sample size, as more evidence comes to light. However, there weren't other adaptive elements in the trial such as patient enrichment or response adaptive randomisation; and the design and subsequent analysis of study data were done under a frequentist statistical framework. Finally, the preprint under review was solely focused on dexamethasone, one of the treatments under consideration. Following from these points, for the purposes of this review, we can think of RECOVERY as a two arm parallel trial aimed at evaluating the efficacy of dexamethasone in hospitalized patients with confirmed or suspected COVID-19.

The overall trial recruited 11,320 patients between March 19th and June 8th, 2020, from 176 National Health Service (NHS) hospital organisations in the United Kingdom. They were allocated 2:1 into usual-care, or usual-care plus one of the available treatments being tested. Thus 4321 were allocated to usual-care, and 2104 were allocated to usual-care plus open-label dexamethasone (6 mg once daily for up to 10 days), with the remainder allocated to other treatment options (which were not further considered in the preprint). The primary outcome was all-cause mortality within 28 days of randomisation. Other key outcomes (all within 28 days of randomisation) included time to discharge from the hospital; and receipt of invasive mechanical ventilation (including ECMO) or death in those patients not receiving invasive mechanical ventilation at randomisation. All analyses were done on an intention-to-treat basis.

The main finding of the trial was that 28-day mortality in the dexamethasone arm was 21.6% (454/2104) compared to 24.6% (1065/4321) in the usual-care arm, with an age-adjusted hazard ratio of 0.83 (95% CI 0.74 to 0.92; $p < 0.001$) favoring the dexamethasone arm. Further, based on a prespecified subgroup analysis, the study demonstrated heterogeneity of the treatment effect by level of respiratory support. For patients who were not yet on any support, the estimated HR was 1.22 (95% CI 0.93 to 1.61; $p = 0.14$), whereas for patients on oxygen only, or invasive mechanical ventilation, the HRs were 0.80 (0.70 to 0.92; $p = 0.002$) and 0.65 (0.51 to 0.82; $p < 0.001$), respectively.

Patients allocated to dexamethasone also experienced shorter hospitalisation times (median 12 days vs. 13 days) were more likely to be discharged within 28 days (64.6% vs 61.1%; HR 1.11, 95% CI 1.04 to 1.19; $p = 0.002$), and that this effect was greater in patients who were on mechanical ventilation at randomisation. Finally, among the subset of patients who were not on mechanical ventilation at randomisation, the probability of progressing to mechanical ventilation,

2

death, or either of these, was lower in patients allocated to dexamethasone (respectively, RR 0.76 [95%CI 0.61 to 0.96; p = 0.021]; 0.91 [0.82 to 1.01, p = 0.07]; 0.91 [0.82 to 1.00, p = 0.049]).

Following from these results, the authors concluded that, "The RECOVERY trial provides clear evidence that treatment with dexamethasone 6 mg once daily for up to 10 days reduces 28-day mortality in patients with COVID-19 who are receiving respiratory support" and that, "Dexamethasone provides an effective treatment for the sickest patients with COVID-19 and, given its low cost, well understood safety profile, and widespread availability, is one that can be used worldwide."

We sincerely thank the authors for their contribution to our collective understanding of COVID-19, for their commitment to the timely dissemination of research results.

Major comments

The overall reporting of the trial was suboptimal, which was exacerbated by a lack of clarity and consistency in other study documentation such as the protocol and statistical analysis plan.

Unfortunately, and quite surprisingly given the quality of the study overall and the experience of the investigators, we found the reporting to be lacking on multiple fronts. Further, while clear and comprehensive reporting is paramount for interpreting for even relatively simple trials, its value is amplified for out-of-the-ordinary designs such as RECOVERY. Below we highlight a few key areas for improvement:

- While adaptive designs can offer many important advantages of traditional fixed trial designs, it is crucial that any processes that will be used to alter trial parameters are clearly described in the trial documentation. Further, it is important to demonstrate how the type 1 error rate will be maintained, given that this study was conducted under a frequentist framework for statistical inference (notably, this is common even for adaptive trials featuring Bayesian methods, usually through comprehensive simulations). For the trial at hand, it was apparently stopped based on a sample size calculation made during the trial, but the sample size calculation is only vaguely described (and based on a risk difference when the trial report focused on hazard ratios) and there are no details given on the process that led to that calculation. The paper refers back to the protocol, but the protocol lacks information on exactly how the sample size would be updated and who would be responsible for making the resulting decisions. To the degree that the protocol does outline how adaptive decisions would be made, it is largely limited to noting that the DMC would view interim data on a bi-weekly basis to evaluate treatment efficacy and harms, but details on what criteria would lead to a recommendation or decision to stop early are

not provided. This implies that the RECOVERY trial data were observed multiple times during the course of the trial, but again there is no information on how this was incorporated into the eventual inferences being made based on the trial results. It could be possible that we simply missed this information, but it wasn't for lack of effort. If this information does exist it should be made more readily accessible. Notably, and as the RECOVERY authors will certainly be aware, there is a CONSORT extension for adaptive studies that would be useful to consider[3].

- Overall, the statistical methods described in the preprint were not well described and were not fully consistent with the protocol or the statistical analysis plan. Importantly, there was no description of how decisions would be made to use adjusted estimates following from tests of imbalance at baseline (e.g. "table 1 tests"), and there were limited details provided in the protocol or SAP on how the prespecified subgroup analyses would be conducted. In the preprint itself (like the SAP), it only notes that chi-squared trend-tests would be used (presumably based on the binary outcome at 28 days?), but this doesn't conform to the Cox regression models that were used to estimate treatment effects, which could have included the relevant interaction term (which raises the question of whether the subgroup specific estimates and CIs are the product of such a model, or from separate, subgroup-specific models).

- Finally, we felt there was a lack of clarity around recruitment and randomisation. In line with CONSORT, though a minor detail, the method of generating the random sequence should be provided; and the patient flow diagram (S1) should be included in the main body of the text. Importantly, it would be useful to see the potential eligible number of patients that presented but that weren't randomized. This might be especially useful in light of the paradoxical observation that those on the mechanical ventilation at baseline were substantially younger than the enrolled patients that weren't, which likely merits some explanation in a trial that intended to recruit a broadly defined patient sample where we might reasonably expect age and level of support at baseline to be positively correlated.

Recommendations:
For future studies
- Please consult CONSORT and any relevant extensions when reporting your trial. Take special care to ensure that what was done and reported is consistent with other trial documentation (which themselves should be as detailed as possible) and where there are differences, clearly note and justify them.

For this study
- Revisit the trial report in light of CONSORT and the extension for adaptive trials.

The prominently featured subgroup analysis needs to be cautiously interpreted.

The paper focused on the results of a prespecified subgroup analysis, for which dexamethasone appeared to result in a reduction in mortality in those receiving mechanical ventilation or oxygen at randomisation, but not in those receiving no respiratory support. However, this was one of five pre-planned subgroup analyses (age, sex, level of respiratory support, days since symptom onset, and predicted 28-day mortality risk), whose results could have been similarly emphasised depending on the result. Given that there were no apparent efforts to control the type 1 error rate, it seems plausible this particular subgroup effect was a chance finding, and thus we feel that it should be more cautiously interpreted. Importantly, if it is indeed a chance finding, and dexamethasone works similarly regardless of ventilation, it may lead to patients who are not on oxygen or mechanical ventilation missing out on an efficacious treatment. Further, because the subgroup is defined by a modifiable treatment factor (the decision to provide respiratory support), it doesn't seem out of the realm of possibility that decisions to provide respiratory support could be influenced by this framing of the result.

Compounding this issue is the suggestion made in the paper that dexamethasone might actually be harmful, an inference that seems to be based on the observation that the bulk of the frequentist 95% confidence interval for the subgroup specific effect falls above the null. From our perspective this seems like too strong of an inference based on a frequentist 95% CI, especially in light of the observation that patients not on oxygen or mechanical ventilation at randomisation were less likely to later need those supports if they were in the dexamethasone arm.

Overall, we would recommend that these results be interpreted cautiously until they can be confirmed in subsequent trials, though it is not clear how equipoise has been affected by the published RECOVERY results, which in turn could wind up making such confirmation difficult, if not impossible, at this point.

Recommendations:
For this study
- We would suggest some consideration of the multiplicity in the subgroup analyses, adding more details about how the subgroup specific effects were estimated, and urge caution against interpreting the point-estimates as "the effect" particularly in light of the frequentist methods used.

For the reader
- In our opinion, the subgroup specific results should be interpreted cautiously, particularly the suggestion of harm in the group not receiving respiratory support, awaiting confirmation of the result.

Minor points

- By "rate ratio", the authors mean hazard ratio and at one point say "rate ratio and risk ratio, both denoted by RR". This may be confusing for readers. If using a Cox model, why not use "HR"? Further, in our opinion, the hazard ratio estimand is not most relevant for a mortality outcome when death occurs in such a short time-frame[4]. The patient flow diagram indicates that there will be continued data collection so that mortality can be assessed at day 28 for all patients, and it would be useful to see the updated analysis, particularly the risk difference for 28-day all cause mortality, once those data are available.

- The models used to estimate treatment effects were adjusted for age. This was a post-hoc decision based on the observation of a "statistically significant" difference between study arms in the mean age. Though the adjustment did not seemingly affect the inferences drawn from the study (i.e. the unadjusted results were essentially the same), the appropriateness of this choice is still debatable. Ideally, investigators should identify strong prognostic factors (e.g. age, clinical status at baseline) and make note in the statistical analysis plan that these will be adjusted for in the eventual analysis. This is good practice even if the factors aren't "imbalanced" at baseline[5], and by preregistering the choice (before any data are collected) it protects the investigators from any accusation of p-hacking[6], though to be clear this isn't a concern in this case. Finally, it is worth noting that adjustment changes the estimand when using an HR - conditional and marginal HRs are not strictly comparable since they have different true values[7].

- Of the three trial registrations, only the EUDRACT registration was prospective (dated March 17). The ISRCTN registration reports that the primary outcome was death in hospitalised patients by 28 days post randomisation, and that this was changed during the study to all-cause mortality at 28 days. What they have actually done in the analysis is closer to the former definition as they have made the assumption that anyone discharged from hospital was then alive up until 28 days.

- The outcomes 'need for renal dialysis or haemofiltration', 'major cardiac arrhythmia' and 'duration of ventilation' are described in the methods section but never analysed (though analyses of two of these are said to be in preparation)

- It would be ideal to also present confidence intervals for the respective risk differences (along with the reported intervals for the estimated rate ratios).

- According to the paper (though not detailed in the protocol or SAP) the reported analysis was intended to provide 90% power to detect the effect of interest with a type 1 error rate of 1%. Consequently, it would be more consistent to feature the 99% CIs in the text, tables and figures, rather than the 95% CIs that are there now (with the 99% CIs reported as footnotes).

- The trial used a 2:1 allocation ratio, and while this is common in multi-arm designs because it reduces the correlation of each comparison vs. control[8], it would be good to explicitly justify this choice in the paper.

- There was reportedly a subgroup analysis based on patients' predicted 28 day mortality risk. However, no details are provided on the specific model used to predict this risk (including in the protocol). While not pertinent to the trial at hand, a validation of this model in the RECOVERY data would be very welcome.

- From the trial registration it seems that pregnant women were given a different corticosteroid (prednisolone), it would be good to add how many there were to table 1.

- Some of the ambiguities could be cleared up by sharing the analysis scripts. It's hard to envision a reason not to do this, other than historical precedent of researchers not doing it.

- It would be worth noting in the statistical analysis plan who was responsible for the analysis and what quality measures (e.g. SOPs, testing of analysis scripts, replication by a second analyst, etc) were in place to ensure it was done correctly.

- In the description of the sample size, we suggest explicitly noting that the 28-day mortality of 20% specifically refers to the usual-care arm, just to avoid any confusion that it might refer to a value averaged over all trial arms.

Open Data

No.

Open Analysis Code

No.

Pre-registered study design

Yes.

PubPeer

There may be comments on the PubPeer page for the published version of this paper.
https://pubpeer.com/publications/0F03B831DA92455784F3C696F006F7

References

1. MRC-NIHR Trials Methodology Research Partnership. https://www.methodologyhubs.mrc.ac.uk/about/tmrp/.

2. Horby, P. *et al. Effect of Dexamethasone in Hospitalized Patients with COVID-19: Preliminary Report*. http://medrxiv.org/lookup/doi/10.1101/2020.06.22.20137273 (2020) doi:10.1101/2020.06.22.20137273.

3. Dimairo, M. *et al.* The Adaptive designs CONSORT Extension (ACE) statement: a checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. *BMJ* m115 (2020) doi:10.1136/bmj.m115.

4. Kahan, Brennan. Treatment estimands in clinical trials of patients hospitalised for COVID-19: ensuring trials ask the right questions. (2020) doi:10.17605/OSF.IO/7WXK9.

5. Kahan, B. C., Jairath, V., Doré, C. J. & Morris, T. P. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials* **15**, 139 (2014).

6. Hawkes, N. Sixty seconds on . . . P-hacking. *BMJ* k4039 (2018) doi:10.1136/bmj.k4039.

7. Hauck, W. W., Anderson, S. & Marcus, S. M. Should We Adjust for Covariates in Nonlinear Regression Analyses of Randomized Trials? *Control. Clin. Trials* **19**, 249–256 (1998).

8. Dunnett, C. W. A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *J. Am. Stat. Assoc.* **50**, 1096–1121 (1955).

9. Schulz, K. F., Altman, D. G., Moher, D. & for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* **340**, c332–c332 (2010).

CONSORT CHECKLIST

To support the review, we completed the CONSORT checklist[9] below. Material taken from the paper (or other documents, which will be labelled) is in *italics.* Our comments are in **bold**.

Title and abstract

1a Identification as a randomised trial in the title

No.

1b Structured summary of trial design, methods, results, and conclusions.

Title: Identification of the study as randomised	No
Authors: Contact details for the corresponding author	Yes
Trial design: Description of the trial design (eg, parallel, cluster, non-inferiority)	Yes
Methods	
Participants: Eligibility criteria for participants and the settings where the data were collected	No
Interventions: Interventions intended for each group	Yes
Objective: Specific objective or hypothesis	Yes
Outcome: Clearly defined primary outcome for this report	Yes
Randomisation: How participants were allocated to interventions	Yes
Blinding (masking): Whether or not participants, care givers, and those assessing the outcomes were blinded to group assignment	Yes
Results	
Numbers randomised: Number of participants randomised to each group	Yes
Recruitment: Trial status	No
Numbers analysed: Number of participants analysed in each group	Yes
Outcome: For the primary outcome, a result for each group and the estimated effect size and its precision	Yes
Harms: Important adverse events or side-effects	No
Conclusions: General interpretation of the results	Yes
Trial registration: Registration number and name of trial register	Yes
Funding: Source of funding	Yes

Introduction

Background and objectives

2a Scientific background and explanation of rationale

Yes

2b Specific objectives or hypotheses

In the absence of reliable evidence from large-scale randomized clinical trials, there is great uncertainty about the effectiveness of corticosteroids in COVID-19. Prior to RECOVERY, many COVID-19 treatment guidelines stated that corticosteroids were either 'contraindicated' or 'not recommended'19 although in China, corticosteroids are recommended for severe cases.20 Practice has varied widely across the world: in some series, as many as 50% of cases were treated with corticosteroids.21,22 Here we report the results of a randomized controlled trial of dexamethasone in patients hospitalized with COVID-19.

Methods

Trial design

3a Description of trial design (such as parallel, factorial) including allocation ratio

The Randomised Evaluation of COVID-19 therapy (RECOVERY) trial is a randomized, controlled, open-label, adaptive, platform trial comparing a range of possible treatments with usual care in patients hospitalized with COVID-19. [**preprint**]

From version 6.0 of the protocol, a factorial design will be used such that eligible and consenting participants may be randomised to one of the treatment arms in Randomisation Aand, simultaneously, to one of the treatment arms in Randomisation B. [**protocol V6.0 2020-05-14**]

3b Important changes to methods after trial commencement (such as eligibility criteria), with reasons

Through the play of chance in the unstratified randomization, mean age was 1.1 years higher in those allocated dexamethasone than those allocated usual care (Table 1). To account for this imbalance in an important prognostic factor, the estimates of rate ratios and risk ratios (both hereon denoted RR) were adjusted for baseline age. This adjustment was not specified in the first version of the statistical analysis plan, but was added once the imbalance in age (a key prognostic factor) became apparent. Results with and without age-adjustment are provided and show that it does not alter the conclusions materially. [**preprint**]

RECOVERY is a randomized trial among patients hospitalized for COVID-19. All eligible patients receive usual standard of care in the participating hospital and are randomly allocated between no additional treatment and one of several active treatment arms. Over time, additional treatment arms have been added (see Table). In version 4.0 of the protocol, a second randomization was introduced for those trial participants with hypoxia (oxygen saturation <92% on air or receiving oxygen) and inflammation(C-reactive protein ≥75 mg/dL), comparing the addition of tocilizumab vs. control on top of the treatment assigned in the first randomization. In version 6.0, a factorial design was introduced to the first randomization such that participants were also randomized to convalescent plasma vs. no additional treatment.As outlined in the protocol, if one or more of the active treatments was not available at the hospital or is believed, by the attending clinician, to be contraindicated (or definitely indicated) for the specific patient, then random allocation was between the remaining treatment arms.The original and final protocol are included in the supplementary material to this publication, together with summaries of the changes made [**preprint supplementary info**]

Participants

4a Eligibility criteria for participants

Hospitalized patients were eligible for the trial if they had clinically suspected or laboratory confirmed SARS-CoV-2 infection and no medical history that might, in the opinion of the attending clinician, put the patient at significant risk if they were to participate in the trial. Initially, recruitment was limited to patients aged at least 18 years but the age limit was removed from 9 May 2020. Pregnant or breast-feeding women were eligible

4b Settings and locations where the data were collected

The trial was conducted at 176 National Health Service (NHS) hospital organizations in the United Kingdom (see Supplementary Appendix), supported by the National Institute for Health Research Clinical Research Network.

Interventions

5 The interventions for each group with sufficient details to allow replication, including how and when they were actually administered

Eligible and consenting patients were assigned in a ratio of 2:1 to either usual standard of care or to usual standard of care plus dexamethasone 6 mg once daily (oral or intravenous) for up to 10 days (or until discharge if sooner) or to one of the other suitable and available treatment arms (see Supplementary Appendix) using web-based simple randomization with allocation concealment

Outcomes

6a Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed

The primary outcome was all-cause mortality within 28 days of randomization. Secondary outcomes were time to discharge from hospital, and among patients not receiving invasive mechanical ventilation at randomization, subsequent receipt of invasive mechanical ventilation(including extra-corporeal membrane oxygenation) or death. Subsidiary clinical outcomes included cause-specific mortality, receipt of renal hemodialysis or hemofiltration, major cardiac arrhythmia (recorded in a subset), and receipt and duration of ventilation [**preprint**]

Primary Outcome Measures :
All-cause mortality [Time Frame: Within 28 days after randomisation]
Secondary Outcome Measures :
Duration of hospital stay [Time Frame: Within 28 days and up to 6 months after the main randomisation]
Need for (and duration of) ventilation [Time Frame: Within 28 days and up to 6 months after the main randomisation]
Composite endpoint of death or need for mechanical ventilation or ECMO [Time Frame: Within 28 days and up to 6 months after the main randomisation] [**registry**]

6b Any changes to trial outcomes after the trial commenced, with reasons

Sample size

7a How sample size was determined

As stated in the protocol, appropriate sample sizes could not be estimated when the trial was being planned at the start of the COVID-19 pandemic. As the trial progressed, the trial Steering Committee, blinded to the results of the study treatment comparisons, formed the view that, if 28-day mortality was 20% then a comparison of at least 2000 patients allocated to active drug and 4000 to usual care alone would yield at least 90% power at two-sided P=0.01 to detect a clinically relevant absolute difference of 4 percentage points between the two groups (a proportional reduction of one-fifth). [**preprint**]

The larger the number randomised,the more accurate the results will be, but the numbers that can be randomised will depend critically on how large the epidemic becomes. If substantial numbers are hospitalised in the participating centres then it may be possible to randomise

several thousand with moderate disease and a few thousand with severe disease.Some indicative sample sizes and projected recruitment will be estimated using emerging data for several different scenarios. Sample size and recruitment will be monitored by the Steering Committee (SC) throughout the trial. [**SAP v1.0 09_06_20**]

No additional information provided in protocol v6.0 (May 14)

7b When applicable, explanation of any interim analyses and stopping guidelines

The interim trial results will be monitored by an independentData Monitoring Committee(DMC). The most important task for the DMC will be to assess whether the randomised comparisons in the study have provided evidence on mortality that is strong enough (with a range of uncertainty around the results that is narrow enough) to affect national and global treatment strategies. In such a circumstance, the DMC will inform the Trial Steering Committee who will make the results available to the public and amend the trial arms accordingly. New trial arms can be added as evidence emerges that other candidate therapeutics should be evaluated. [**protocol V6.0 2020-05-14**]

Randomisation

Sequence generation

8a Method used to generate the random allocation sequence

using web based simple randomization

8b Type of randomisation; details of any restriction (such as blocking and block size)

Baseline data collected using a web-based case report form included demographics, level of respiratory support, major comorbidities, suitability of the study treatment for a particular patient and treatment availability at the study site. Eligible and consenting patients were assigned in a ratio of 2:1 to either usual standard of care or to usual standard of care plus dexamethasone 6 mg once daily (oral or intravenous) for up to 10 days (or until discharge if sooner) or to one of the other suitable and available treatment arms (see Supplementary Appendix) using web-based simple randomization with allocation concealment.

Allocation concealment mechanism

9 Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned

13

...using web-based simple randomization with allocation concealment. [**preprint**]

Eligible patients will be randomised using a 24/7 secure central web-based randomisation system, developed and hosted within NDPH, University of Oxford. Users of the system will have no insight into the next allocation, given that simple randomisation is being used. In the event that a patient is randomised inadvertently more than once during the same hospital admission, the first allocation will be used. The implementation of the randomisation procedure will be monitored by the Senior Trials Programmer, and the SC notified if an error in the randomisation process is identified [**SAP v1.0 09_06_20**]

Implementation

10 Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions

The details of the allocated study treatments will be displayed on the screen and can be printed or downloaded. The hospital clinicians are responsible for administration of the allocated treatments. [**protocol V6.0 2020-05-14**]

Blinding

11a If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how.

Open label.

11b If relevant, description of the similarity of interventions

NA

Statistical methods

12a Statistical methods used to compare groups for primary and secondary outcomes

For the primary outcome of 28-day mortality, the hazard ratio from Cox regression was used to estimate the mortality rate ratio. The few patients (4.8%) who had not been followed for 28 days by the time of the data cut (10 June 2020) were either censored on 8 June 2020 or, if they had already been discharged alive, were right-censored at day 29 (that is, in the absence of any information to the contrary they were assumed to have survived 28 days). Kaplan-Meier survival curves were constructed to display cumulative mortality over the 28-day period. Cox regression was used to analyze the secondary outcome of hospital discharge within 28 days, with patients who died in hospital right-censored on day 29. For the pre-specified composite secondary outcome of invasive mechanical ventilation or death within 28 days (among those not receiving

invasive mechanical ventilation at randomization), the precise date of invasive mechanical ventilation was not available and so a log-binomial regression model was used to estimate the risk ratio.

12b Methods for additional analyses, such as subgroup analyses and adjusted analyses

Pre-specified analyses of the primary outcome were performed in five subgroups defined by characteristics at randomization: age, sex, level of respiratory support, days since symptom onset, and predicted 28-day mortality risk. Observed effects within subgroup categories were compared using a chi-square test for trend.

Results

Participant flow (a diagram is strongly recommended)

13a For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome

13b For each group, losses and exclusions after randomisation, together with reasons

Figure S1: Trial profile - Flow of participants through the RECOVERY trial
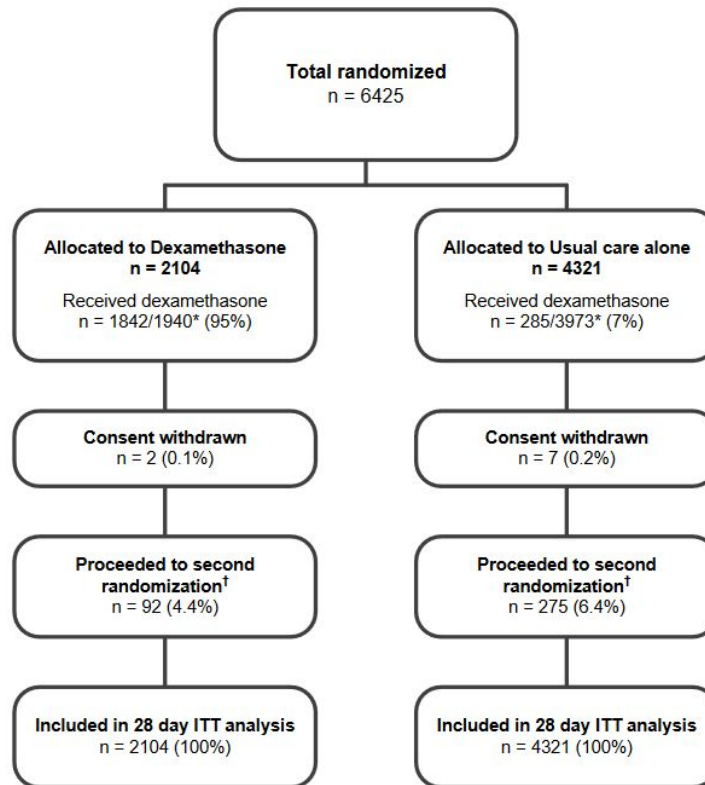
ITT – intention to treat
* 1940/2104 (92.1%) and 3973/4321 (91.9%) patients have a completed follow-up form at time of analysis.
† Randomization to Tocilizumab vs. No additional treatment in accordance with protocol version 4.0 or later. In addition, 14 patients were additionally randomized to convalescent plasma vs control (5 [0.2%] patients allocated to dexamethasone arm vs 9 [0.2%] patients allocated to usual care) in accordance with protocol V6.0.

Recruitment

14a Dates defining the periods of recruitment and follow-up

Of the 11,320 patients randomized between 19 March and 8 June...

14b Why the trial ended or was stopped

As the trial progressed, the trial Steering Committee, blinded to the results of the study treatment comparisons, formed the view that, if 28-day mortality was 20% then a comparison of at least 2000 patients allocated to active drug and 4000 to usual care alone would yield at least 90% power at two-sided P=0.01 to detect a clinically relevant absolute difference of 4 percentage points between the two groups (a proportional reduction of one-fifth). Consequently,

on 8 June 2020, the Steering Committee closed recruitment to the dexamethasone arm since enrolment exceeded 2000 patients.

Baseline data

15 A table showing baseline demographic and clinical characteristics for each group

Of the 11,320 patients randomized between 19 March and 8 June, 9355 (83%) were eligible to be randomized to dexamethasone (that is dexamethasone was available in the hospital at the time and the patient had no known indication for or contraindication to dexamethasone). Of these, 2104 were randomized to dexamethasone and 4321 were randomized to usual care (Figure S1), with the remainder being randomized to one of the other treatment arms.

Table 1: Baseline characteristics by randomized allocation and level of respiratory support received

	Treatment allocation		Respiratory support received at randomization		
	Dexamethasone (n=2104)	Usual care (n=4321)	No oxygen received (n=1535)	Oxygen only (n=3883)	Invasive mechanical ventilation (n=1007)
Age, years	66.9 (15.4)	65.8 (15.8)	69.3 (17.6)	66.7 (15.3)	59.0 (11.5)
<70	1142 (54%)	2506 (58%)	660 (43%)	2149 (55%)	839 (83%)
≥70 to <80	467 (22%)	860 (20%)	338 (22%)	837 (22%)	152 (15%)
≥80	495 (24%)	955 (22%)	537 (35%)	897 (23%)	16 (2%)
Sex					
Male	1338 (64%)	2750 (64%)	892 (58%)	2462 (63%)	734 (73%)
Female*	766 (36%)	1571 (36%)	643 (42%)	1421 (37%)	273 (27%)
Number of days since symptom onset	8 (5-13)	9 (5-13)	6 (3-10)	9 (5-12)	13 (8-18)
Respiratory support received					
No oxygen received	501 (24%)	1034 (24%)	1535 (100%)	0 (0%)	0 (0%)
Oxygen only	1279 (61%)	2604 (60%)	0 (0%)	3883 (100%)	0 (0%)
Invasive mechanical ventilation	324 (15%)	683 (16%)	0 (0%)	0 (0%)	1007 (100%)
Previous diseases					
Diabetes	521 (25%)	1025 (24%)	342 (22%)	950 (24%)	254 (25%)
Heart disease	586 (28%)	1171 (27%)	519 (34%)	1074 (28%)	164 (16%)
Chronic lung disease	415 (20%)	931 (22%)	351 (23%)	883 (23%)	112 (11%)
Tuberculosis	6 (<0.5%)	19 (<0.5%)	8 (1%)	11 (<0.5%)	6 (1%)
HIV	12 (1%)	20 (<0.5%)	5 (<0.5%)	21 (1%)	6 (1%)
Severe liver disease	37 (2%)	82 (2%)	32 (2%)	72 (2%)	15 (1%)
Severe kidney impairment	167 (8%)	358 (8%)	120 (8%)	253 (7%)	152 (15%)
Any of the above	1174 (56%)	2417 (56%)	911 (59%)	2175 (56%)	505 (50%)
SARS-Cov-2 test result					
Positive	1702 (81%)	3553 (82%)	1198 (78%)	3144 (81%)	913 (91%)
Negative	213 (10%)	397 (9%)	182 (12%)	398 (10%)	30 (3%)
Test result not yet known†	189 (9%)	371 (9%)	155 (10%)	341 (9%)	64 (6%)

Results are count (%), mean ± standard deviation, or median (inter-quartile range). * Includes 6 pregnant women. † SARS-Cov-2 test results are captured on the follow-up form, so are currently unknown for some. There was a significant (2p=0.008) difference in mean age between those allocated dexamethasone and those allocated usual care, but no significant differences between these groups in any other baseline characteristic. The 'oxygen only' group includes non-invasive ventilation.

Numbers analysed

16 For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups

Outcomes and estimation

17a For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)

17b For binary outcomes, presentation of both absolute and relative effect sizes is recommended

Significantly fewer patients allocated to dexamethasone met the primary outcome of 28-day mortality than in the usual care group (454 of 2104 patients [21.6%] allocated dexamethasone vs. 1065 of 4321 patients [24.6%] allocated usual care; rate ratio, 0.83; 95% confidence interval [CI], 0.74 to 0.92; P<0.001) (Figure 1A).

Allocation to dexamethasone was associated with a shorter duration of hospitalization than usual care (median 12 days vs. 13 days) and a greater probability of discharge within 28 days (rate ratio 1.11 [95% CI 1.04 to 1.19]; p=0.002) (Table 2) with the greatest effect seen among those receiving invasive mechanical ventilation at baseline (test for trend p=0.002) (Figure S3a).

Table 2: Effect of allocation to dexamethasone on main study outcomes

| | Treatment allocation | | | |
	Dexamethasone (n=2104)	Usual care (n=4321)	RR (95% CI)	p-value
Primary outcome:				
28-day mortality	454 (21.6%)	1065 (24.6%)	0.83 (0.74-0.92)	<0.001
Secondary outcomes:				
Discharged from hospital within 28 days	1360 (64.6%)	2639 (61.1%)	1.11 (1.04-1.19)	0.002
Receipt of invasive mechanical ventilation or death*	425/1780 (23.9%)	939/3638 (25.8%)	0.91 (0.82-1.00)	0.049
Invasive mechanical ventilation	92/1780 (5.2%)	258/3638 (7.1%)	0.76 (0.61-0.96)	0.021
Death	360/1780 (20.2%)	787/3638 (21.6%)	0.91 (0.82-1.01)	0.07

RR=Rate Ratio for the outcomes of 28-day mortality and hospital discharge, and risk ratio for the outcome of receipt of invasive mechanical ventilation or death (and its subcomponents). Estimates of the RR and its 95% confidence interval are adjusted for age in three categories (<70 years , 70-79 years, and 80 years or older). * Analyses exclude those on invasive mechanical ventilation at randomization.

Ancillary analyses

18 Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory

In a pre-specified subgroup analysis by level of respiratory support received at randomization, there was a significant trend showing the greatest absolute and proportional benefit among those patients receiving invasive mechanical ventilation at randomization (test for trend

p<0.001) (Figure 2). Dexamethasone reduced 28-day mortality by 35% in patients receiving invasive mechanical ventilation (rate ratio 0.65 [95 % CI 0.51 to 0.82]; p<0.001) and by 20% in patients receiving oxygen without invasive mechanical ventilation (rate ratio 0.80 [95 % CI 0.70 to 0.92]; p=0.002) (Figure 1B-C). However, there was no evidence of benefit among those patients who were not receiving respiratory support (rate ratio 1.22 [95 % CI 0.93 to 1.61]; p=0.14) (Figure 1D). Sensitivity analyses without age-adjustment produced similar findings (Table S2).

Patients with longer duration of symptoms (who were more likely to be on invasive mechanical ventilation at randomization) had a greater mortality benefit, such that dexamethasone was associated with a reduction in 28-day mortality among those with symptoms for more than 7 days but not among those with more recent symptom onset (test for trend p<0.001) (Figure S2)
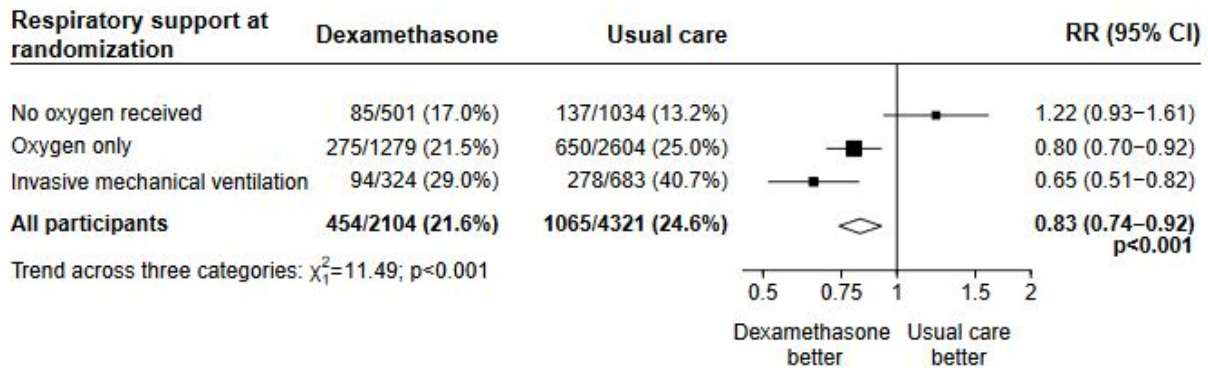
Among those not on invasive mechanical ventilation at baseline, the number of patients progressing to the pre-specified composite secondary outcome of invasive mechanical ventilation or death was lower among those allocated to dexamethasone (risk ratio 0.91 [95% CI a 0.82 to 1.00]; p=0.049) (Table 2) but with significantly greater effects among patients receiving oxygen at randomization (test for trend p=0.008) (Figure S3b). Subsidiary clinical outcomesThe risk of progression to invasive mechanical ventilation was lower among those allocated dexamethasone vs. usual care (risk ratio 0.76 [95% CI 0.61 to 0.96]; p=0.021) (Table 2).Preliminary analyses indicate no excess risk of any particular cause of death (in particular there was no excess of deaths due to non-COVID infection).

Figure 1: 28−day mortality in all patients (panel a) and separately according to level of respiratory support received at randomization (panels b−d)
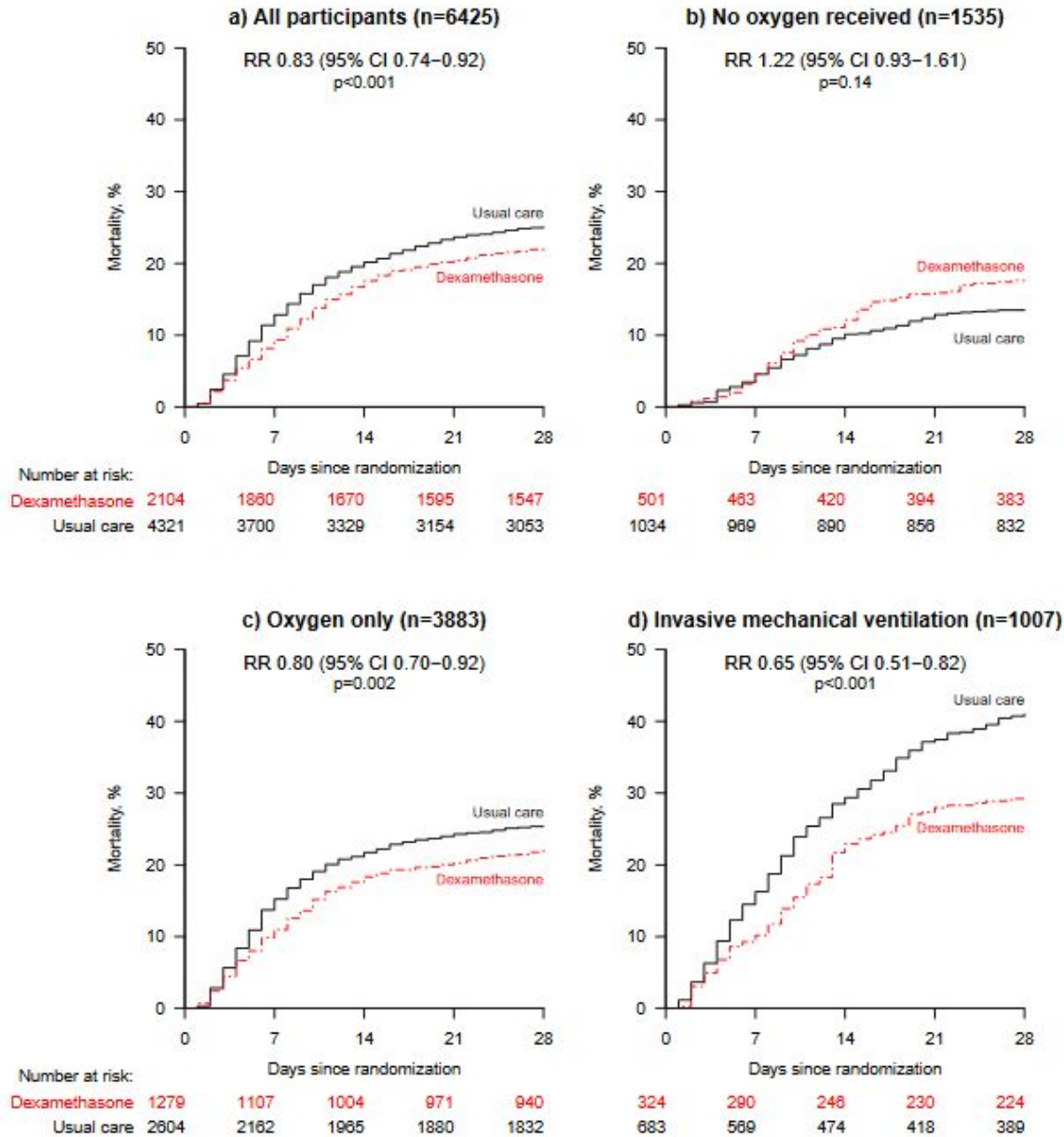
RR=age−adjusted rate ratio. CI=confidence interval. The 'oxygen only' group includes non−invasive ventilation. Note: in the RECOVERY trial press release of 16 June 2020, effects in subgroups of level of respiratory support received were shown with 99% CIs, not 95% CIs as inadvertently stated. The age−adjusted rate ratio and 99% confidence intervals remain unchanged in this analysis: no oxygen required, RR 1.22 (99% CI 0.86−1.75); oxygen only, RR 0.80 (99% CI 0.67−0.96); invasive mechanical ventilation, RR 0.65 (99% CI 0.48−0.88).

Figure 2: Effect of allocation to dexamethasone on 28-day mortality by level of respiratory support received at randomization

Respiratory support at randomization	Dexamethasone	Usual care		RR (95% CI)
No oxygen received	85/501 (17.0%)	137/1034 (13.2%)		1.22 (0.93–1.61)
Oxygen only	275/1279 (21.5%)	650/2604 (25.0%)		0.80 (0.70–0.92)
Invasive mechanical ventilation	94/324 (29.0%)	278/683 (40.7%)		0.65 (0.51–0.82)
All participants	454/2104 (21.6%)	1065/4321 (24.6%)		0.83 (0.74–0.92) p<0.001

Trend across three categories: χ^2_1=11.49; p<0.001

0.5 0.75 1 1.5 2
Dexamethasone better   Usual care better

RR=age-adjusted rate ratio. CI=confidence interval. Subgroup-specific RR estimates are represented by squares (with areas of the squares proportional to the amount of statistical information) and the lines through them correspond to the 95% confidence intervals. The 'oxygen only' group includes non-invasive ventilation. Note: in the RECOVERY trial press release of 16 June 2020, effects in subgroups of level of respiratory support received were shown with 99% CIs, not 95% CIs as inadvertently stated. The age-adjusted rate ratio and 99% confidence intervals remain unchanged in this analysis: no oxygen required, RR 1.22 (99% CI 0.86–1.75); oxygen only, RR 0.80 (99% CI 0.67–0.96); invasive mechanical ventilation, RR 0.65 (99% CI 0.48–0.88).

Harms

19 All important harms or unintended effects in each group (for specific guidance see CONSORT for harms42)

Not reported.

Discussion

Limitations

20 Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses

Not discussed.

Generalisability

21 Generalisability (external validity, applicability) of the trial findings

Prior to the completion of this trial, many COVID-19 treatment guidelines have stated that corticosteroids are either 'contraindicated' or 'not recommended' in COVID-19.19 These should now be updated, as has already happened within the UK.27 Dexamethasone provides an effective treatment for the sickest patients with COVID-19 and, given its low cost, well understood safety profile, and widespread availability, is one that can be used worldwide.

Interpretation

22 Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence

Prior to the completion of this trial, many COVID-19 treatment guidelines have stated that corticosteroids are either 'contraindicated' or 'not recommended' in COVID-19.19 These should now be updated, as has already happened within the UK.27 Dexamethasone provides an effective treatment for the sickest patients with COVID-19 and, given its low cost, well understood safety profile, and widespread availability, is one that can be used worldwide.

Other information

Registration

23 Registration number and name of trial registry

The RECOVERY trial is registered with ISRCTN (50189673) and clinicaltrials.gov (NCT04381936).

Protocol

24 Where the full trial protocol can be accessed, if available

Protocol V6.0: https://www.recoverytrial.net/files/recovery-protocol-v6-0-2020-05-14.pdf
SAP v1.0: https://www.recoverytrial.net/files/recovery-sap-v1-0-09_06_20.pdf

Funding

25 Sources of funding and other support (such as supply of drugs), role of funders

Medical Research Council and National Institute for Health Research (Grant ref: MC_PC_19056).