Should RECOVERY have used Response Adaptive Randomization? Evidence from a simulation study

Tamir Sirkis (✉ tamir@live.co.uk)

University of Exeter Medical School <https://orcid.org/0000-0002-2631-6105>

Jack Bowden

Benjamin Jones

Methodology

Keywords: COVID-19, RECOVERY, response-adaptive randomization, REMAP-CAP, coronavirus, adaptive trial, platform trial, simulation

Posted Date: October 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-814621/v2>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Should RECOVERY have used Response Adaptive Randomization?

Evidence from a simulation study

Tamir Sirkis^{1*}, tamir@live.co.uk; Benjamin Jones, b.g.jones@exeter.ac.uk²; Jack Bowden,
j.bowden2@exeter.ac.uk¹

1. University of Exeter College of Medicine and Health, Exeter, U.K.

2. NIHR ARC South West Peninsula (PenARC), College of Medicine and Health, University of
Exeter

*corresponding author, tamir@live.co.uk

21

22 **Abstract**

23 **Background**

24 The Randomised Evaluation of COVID-19 Therapy (RECOVERY) trial is aimed at addressing the urgent
25 need to find effective treatments for patients hospitalised with suspected or confirmed COVID-19.

26 The trial has had many successes, including discovering that dexamethasone is effective at reducing
27 COVID-19 mortality, the first treatment to reach this milestone in a randomised controlled trial.

28 Despite this, it continues to use standard or 'fixed' randomization (FR) to allocate patients to
29 treatments. We assessed the impact of implementing response adaptive randomization (RAR) within
30 RECOVERY using an array of performance measures, to learn if it could be beneficial going forward.

31 This design feature has recently been implemented within the REMAP-CAP trial.

32 **Methods**

33 Trial data was simulated to closely match the data for patients allocated to either standard care or
34 dexamethasone in the RECOVERY trial from March-June 2020, representing two out of five arms
35 tested throughout this period. Two forms of FR and two forms of RAR were tested. Randomization
36 strategies were performed at the whole trial level as well as within three pre-specified patient
37 subgroups defined by patients' respiratory support level.

38 **Results**

39 RAR strategies led to more patients being given dexamethasone and a lower mortality rate in the
40 trial. Subgroup specific RAR reduced mortality rates even further. RAR did not induce any
41 meaningful bias in treatment effect estimates, but reduced statistical power compared to FR, with
42 subgroup level adaptive randomizations exhibiting the largest power reduction.

43 **Conclusions**

Using RAR within RECOVERY could have resulted in fewer deaths in the trial. However, a larger trial would have been needed to attain the same study power. This could potentially have prolonged the time to full approval of the drug, unless RAR itself led to an increased recruitment rate. Deciding how to balance the needs of patients within a trial and future patients who have yet to fall ill is an important ethical question of our time. RAR deserves to be considered as a design feature in future trials of COVID-19 and other diseases.

Keywords: COVID-19, RECOVERY, response-adaptive randomization, REMAP-CAP, coronavirus, adaptive trial, platform trial, simulation

Background

Coronavirus disease 2019 (COVID-19) is a condition caused by the severe acute respiratory syndrome coronavirus 2 (1). On March 11 2020, the global incidence and virulence of COVID-19 met the criteria for the World Health Organisation to declare it a pandemic (2). At the time of writing, the disease has caused 200 million cases and 4 million deaths worldwide (3). Furthermore, some patients who contracted COVID-19 report experiencing long covid, a condition consisting of many symptoms such as profound fatigue that can persist long after the initial infection has passed (4). It has been estimated there are around two million people in the UK who have suffered from long covid (5). The effects of the pandemic have been far reaching and extend beyond just those infected. For example, many patients with suspected cancer are not receiving appropriate early management, which experts believe will lead to increased mortality as their condition remains untreated (6). Likewise, the pandemic has been noted to increase and exacerbate mental health problems such as stress and anxiety (7). In terms of the wider societal impact, the pandemic has also led to a sharp increase in extreme poverty, 1.5 billion students to miss out on education, and increasing food insecurity (8).

When COVID-19 first emerged, very little was known about its pathophysiology and, as a result, clinicians were unsure which treatments would reduce COVID-19's associated morbidity and

mortality. To address this, large scale randomized clinical trials were quickly designed, authorised, and initiated in patients with severe COVID-19 symptoms in a bid to find effective treatments. One of the most high profile examples is the **Randomised Evaluation of COVID-19 ThERapY (RECOVERY)** trial (9). It commenced on March 19 2020, eight days after the pandemic announcement, with the aim of discovering new treatments that are effective in reducing 28-day mortality in patients hospitalised with confirmed or suspected COVID-19(9). To date, 15 treatments have been trialled, most consisting of repurposed drugs (9). Nine treatments have had results published, and three of these have been declared successful: dexamethasone (10), tocilizumab(11) and REGENERON (12). When analysing the dexamethasone data, researchers found a statistically significant effect in all patients, where fewer patients died on dexamethasone (22.9%) compared to standard care (25.7%) (10), but also performed a stratified analysis in three pre-specified patient subgroups who received different levels of respiratory support at the time of randomization: (i) no oxygen, (ii) oxygen only, and (iii) oxygen through invasive mechanical ventilation. The subgroup analysis uncovered important treatment effect heterogeneity: The most severely ill patients in group (iii) received a larger benefit from taking dexamethasone compared to group (ii), while relatively stable patients in group (i) appeared to fare better with standard care (although this was not statistically significant) (10). These results are shown in Table 1. The National Institute for Health and Care Excellence issued guidelines based on this research, stating dexamethasone should be given to hospitalised patients with COVID-19 if they require oxygen (13).

Patient subgroup	Number of patients	Proportion of patients	Number allocated standard care	Number allocated dexamethasone	Mortality on standard care	Mortality on dexamethasone
No oxygen	1535	23.9%	1034	501	14.0%	17.8%

Oxygen-only	3883	60.3%	2604	1279	26.2%	23.3%
Invasive ventilation	1007	15.7%	683	324	41.4%	29.3%

89

90 *Table 1: Mortality data for the RECOVERY trial stratified by patient group.*

91 RECOVERY is an excellent example of a modern adaptive platform trial (14). Unlike a traditional trial
92 where all design aspects (including the treatments to be compared and the sample size) must be
93 decided before the trial commences, adaptive platform trials have the freedom to continue
94 indefinitely. Whilst the trial is ongoing, new experimental treatments can be added and tested, old
95 experimental treatments showing little benefit can be dropped and effective experimental
96 treatments can be 'graduated' to become the *de-facto* standard of care. Many adaptive trial designs
97 exist, but their common aim is to be more flexible and resource efficient. Supporters of adaptive
98 designs assert that they are more ethical (15), although this is not universally accepted (16). One of
99 the most controversial features that can be incorporated into an adaptive trial is response-adaptive
100 randomization (RAR). RECOVERY did not incorporate this, preferring instead to use standard 'fixed'
101 randomization (FR) probabilities when allocating patients to experimental or standard care arms
102 during the trial. Under a RAR scheme, allocation generally starts in a FR state, randomising patients
103 to all trial arms with equal probability. RAR subsequently facilitates the adaptation of the allocation
104 ratio, as interim analyses begin to show that there is a genuine difference between outcomes in the
105 different arms, to favour treatments that have a higher estimated probability of a favourable
106 outcome. A recent example of a COVID-19 study that uses RAR is the **R**andomised, **E**mbodied,
107 **M**ulti-factorial, **A**ddaptive **P**latform Trial for **C**ommunity-**A**cquired **P**neumonia (REMAP-CAP) (17). This
108 is a trial that aims to evaluate multiple interventions simultaneously for community acquired
109 pneumonia but has a sub-platform, REMAP-COVID, created to assess COVID-19 treatments (18).

The aim of this paper is to investigate, by simulation, the possible benefit of applying RAR instead of FR to assign patients to either dexamethasone or standard of care in the RECOVERY trial. We hypothesised that applying RAR would reduce the number of deaths amongst trial participants by allocating more patients to their optimal treatment. To implement RAR, we use the REMAP-CAP algorithm as well as our own bespoke tuning algorithm. In different simulations we apply each RAR method across the whole patient cohort, and then separately within patient subgroups (i)-(iii). We apply FR using allocation ratios of 1:1 and 2:1 (with respect to standard care: dexamethasone), with the latter having been used in RECOVERY. To quantify benefit, we focus on the following four metrics:

- The proportion of patients allocated to each treatment;
- The expected number of deaths throughout trial;
- The statistical power to detect a treatment effect in all patients and in patient subgroups;
- The bias and mean squared error the treatment effect estimate.

Methods

Simulation set up

The parameters of the simulation were selected to closely resemble the observed results of the RECOVERY trial. The original trial followed up a total of 6425 patients across the dexamethasone (N = 2104) and standard of care (N = 4321) arms in 81 days between March 19 and June 8, 2020. At this point the trial was halted and results of its efficacy were published (10). To provide a means to implement RAR, patient outcome data was simulated for 100 days (or blocks) of 80 patients to give a total of 8000 patients. This sample size and trial duration intentionally exceeds that of RECOVERY, because it enables our study to evaluate the statistical properties of the design at both smaller and larger sample sizes than the actual trial. To match RECOVERY, 24%, 60% and 16% of each block were drawn from patient subgroups (i)-(iii), respectively. Patient outcomes (Y) representing the primary

outcome of mortality at 28 days were generated from a Bernoulli distribution, where for patient i randomized in block j in patient subgroup k , on treatment l :

$$Y_{i,j,k,l} \sim \text{bern}(P_{k,l}),$$

where $i = 1, \dots, 80$, $j = 1, \dots, 100$, $k = \text{i, ii, iii}$, $l = 0$ (standard of care) and 1 (standard of care + dexamethasone). The values of $P_{k,l}$ (the 28 day mortality rate) match the rates observed in RECOVERY (see Table 1).

Randomization allocation strategies

Six allocation strategies (two FR and four RAR) were investigated as part of this simulation. FR was investigated using both a 1:1 and a 2:1 standard care: dexamethasone ratio. Given the underlying outcome rates assumed in each trial arm are similar (Table 1), 1:1 allocation, or a 50% probability of receiving either, drug is near-optimal in terms of statistical power according to Neyman's rule (19) (the exact value being a 51%/49% split). For further details see Technical Appendix A. The latter 1:2 strategy was used in the RECOVERY trial. RAR was investigated using two randomization algorithms, our tuning algorithm (T) and the algorithm used in REMAP-CAP (RMC). We use T_f and RMC_f , to denote the RAR allocation procedures applied to trial patients across the full patient cohort. We use T_s and RMC_s , to denote RAR allocation procedures applied within each patient subgroup. Specifically, the probability of patients in block j and subgroup k being allocated to the dexamethasone group given treatment and outcome data on all preceding patients in blocks $1, \dots, j-28$ is denoted by $\alpha_{j,k}$, where:

154 $\alpha_{j,k}$

$$155 = \begin{cases} \frac{1}{2} : \text{for 1:1 Fixed equal Randomization (FeR)} \\ \frac{1}{3} : \text{for 1:2 Fixed unequal Randomization (FuR)} \\ \frac{\theta(1)^s}{\theta(1)^s + (1 - \theta(1))^s} : \text{Cohort Tuning algorithm (T}_f\text{)} \\ \sqrt{\frac{\theta(1)}{\sum_{j=1}^{J-28} n_{j,1} + 1}} / \left(\sqrt{\frac{\theta(1)}{\sum_{j=1}^{J-28} n_{j,1} + 1}} + \sqrt{\frac{\theta(0)}{\sum_{j=1}^{J-28} n_{j,0} + 1}} \right) : \text{Cohort REMAPCAP algortihm (RMC}_f\text{)} \\ \frac{\theta_k(1)^s}{\theta_k(1)^s + (1 - \theta_k(1))^s} : \text{Subgroup Tuning algorithm (T}_s\text{)} \\ \sqrt{\frac{\theta_k(1)}{\sum_{j=1}^{J-28} n_{j,k,1} + 1}} / \left(\sqrt{\frac{\theta_k(1)}{\sum_{j=1}^{J-28} n_{j,k,1} + 1}} + \sqrt{\frac{\theta_k(0)}{\sum_{j=1}^{J-28} n_{j,k,0} + 1}} \right) : \text{Subgroup REMAPCAP algortihm (RMC}_s\text{)} \end{cases}$$

156

157 Where $\theta(l)$ represents the posterior probability that treatment l is optimal based on patients who
 158 have been in the trial for at least 28 days, either for the full cohort or for subgroup k (further details
 159 supplied in Technical Appendix B); s is the proportion of trial stages completed at point of adaption;
 160 and $n_{j,k,l}$ represents number of patients in stage k in the cohort/sub-group that have been allocated
 161 treatment l .

162 In addition, allocation probabilities in RAR schemes T and RMC were constrained by a maximum and
 163 minimum value according to the following rule:

$$164 \alpha_{j,k} = \begin{cases} 0.9 \text{ if } \alpha_{j,k} > 0.9 \\ \alpha_{j,k} \text{ if } 0.1 \leq \alpha_{j,k} \leq 0.9 \\ 0.1 \text{ if } \alpha_{j,k} < 0.1 \end{cases}$$

165 Both the T and RMC RAR procedures used a “burn-in” period (a period where adaptive
 166 randomization was not applied) for the first 34 days of trial recruitment, meaning the first 2720
 167 patients or 34% of the trial were allocated in a fixed 1:1 ratio. Only after this point, new patients
 168 were allocated using RAR, with the $\alpha_{j,k}$ ratio being updated every 7 days. Each simulation was
 169 performed 1000 times.

Trial metrics

To assess the performance of the methods, the following summary measures were calculated across the 1000 simulated trials:

- The expected or average number of patients allocated to Dexamethasone $E[N_d]$
- The expected or average number of deaths $E[N_Y]$ at point where 6450 patients were allocated to a treatment, a number chosen due to its proximity to the 6425 patients recruited in RECOVERY.
- The expected probability (or statistical power) of rejecting the null hypothesis of a zero treatment effect at the end of each block $j = 1, \dots, 100$ in the trial, $E[Rej_j]$, calculated by $\sum_{t=1}^T \frac{\begin{cases} Rej_j = 1 & \text{if } p \leq 0.05 \\ Rej_j = 0 & \text{if } p > 0.05 \end{cases}}{T}$ for trials $t=1, \dots, 1000$. This was calculated using a logistic regression model.
- The relative bias in the treatment effect estimate $bias_k = E\left[\frac{(\hat{P}_{k,0} - \hat{P}_{k,1}) - (P_{k,0} - P_{k,1})}{(P_{k,0} - P_{k,1})}\right]$, where $P_{k,l}$ is the actual mortality rate and $\hat{P}_{k,l}$ is the estimated mortality rate in subgroup k ($1=(i)$, $2=(ii)$, $3=(iii)$) on treatment l (1 for dexamethasone, 0 for standard of care) from the trial. This bias is only calculated for RAR procedures, as only RAR induces bias in treatment effect estimates, as explained in technical appendix C.

Summary

To summarise, the simulation study investigated six treatment allocation methods:

- 1:1 FR across all patients (FeR)
- 1:2 FR across all patients (FuR)
- T algorithm across all patients (T_f)
- RMC algorithm across all patients (RMC_f)
- T algorithm within subgroups (i)-(iii) separately (T_s)

- RMC algorithm within subgroups (i)-(iii) separately (RMC_s)

In order to benchmark the performance of each strategy in terms of overall mortality, the expected number of deaths, $E[N_Y]$, is quantified under four theoretical strategies shown in Table 2. Strategies 1 and 2 describe allocation processes where the full cohort receives the same treatment, and therefore should be compared with T_f and RMC_f allocation strategies. In contrast, strategies 3 and 4 describe subgroup-specific treatment allocation and should be compared to T_s and RMC_s.

Strategy	Description	Expected mortality (%)
1. Dexamethasone	All patients given dexamethasone	1,477 (22.9%)
2. Standard Care	All patients given standard care	1,655 (25.7%)
3. Sub-group optimal	No oxygen group given standard care; other groups given dexamethasone	1,419 (22.0%)
4. Sub-group least optimal	No oxygen group given dexamethasone; other groups given standard care	1,713 (26.6%)

Table 2: Theoretical strategies leading to the best and worst mortality rates on patients given standard care and dexamethasone

Results

Allocation ratios

Both cohort RAR methods led to more patients receiving dexamethasone compared to either FR approach. T_f led to slightly more patients receiving dexamethasone than RMC_f. When considering subgroup-specific RAR, each subgroup has its own trend. In subgroup (i), both adaptive methods allocate slightly less patients to standard care compared to FuR, with T_s allocating standard care to

the most patients. In subgroups (ii) and (iii), the RAR algorithms allocate mainly to dexamethasone, with RMC_s allocating dexamethasone to the most patients. However, the treatment allocation disparity is much larger in the subgroup (iii). These differences are demonstrated in Figure 1.

Figure 1: Allocation proportions within the full cohort and within each patient subgroup for all FR and RAR procedures.

Horizontal blue line shows $n = 6400$, the sample size of RECOVERY

The RAR algorithms also differ in how they “ramp up” allocation to the optimal treatment as the trial progresses. T_f increases randomization to the optimal treatment at a steadier rate from the end of the burn-in period, whereas RMC_f increases randomization faster in the early trial stages but also plateaus earlier. This is consistent with the pattern in subgroup-level RAR methods, T_s and RMC_s, for subgroups (i) and (ii). However, in subgroup (iii), the RMC_s algorithm ramps up allocation to dexamethasone faster across the whole trial. This is demonstrated in Figures 2 and 3.

224

225

226 *Figure 2: The proportion of patients receiving dexamethasone across the trial for FR and cohort-level RAR frameworks.*

227 *Vertical line at 42.5% trial progress represents the end of the burn-in period.*

228

229

230

Figure 3: the allocation of patients to dexamethasone and no treatment in T_s and RMC_s . Vertical line at 42.5% trial progress represents the end of the burn-in period.

Mortality rates

FR led to the highest number of expected deaths (with FuR being worse than FeR). This was followed by cohort-level RAR, with the lowest mortality rates observed when using subgroup-specific RAR. There was little difference between the randomisation algorithms, with RMC_f and RMC_s attaining a marginally lower mortality rate than T_f and T_s respectively. The expected mortality figures are given in Table 3, expressed as deaths prevented relative to FuR .

Randomization strategy	Average no. of deaths at n = 6400 (percentage)	Deaths prevented compared to FuR at n = 6400
FuR	1583 (24.7%)	-
FeR	1556 (24.3%)	27
T_f	1528 (23.9%)	55
RMC_f	1527 (23.9%)	56

T_s	1516 (23.7%)	67
RMC_s	1514 (23.7%)	69

Table 3: Expected mortality rates in whole trial for each randomization method, given at point that RECOVERY trial was finished.

Statistical power

FeR led to the highest study power on a cohort level, as predicted by Neyman's rule. FuR, T_f and RMC_f all performed similarly, with FuR having slightly more power by the end of the simulation and at the point where the simulation has almost the same sample size as RECOVERY. Comparing the cohort RAR algorithms, RMC_f performs slightly better than T_f in most points of the trial. These results are shown in Figure 4.

Figure 4: Comparing fixed randomization and cohort-level adaptive randomization approaches in terms of study power across the length of the trial.

For subgroup-specific RAR procedures, the difference in power is minimal between randomization approaches; both RMC_s and T_s produce similar power at all stages. However, there is notable differences between subgroups. Subgroup (iii), for whom the treatment effect is largest, has the highest power and is the only subgroup where the trial reaches 90% power during the study. The other two groups don't reach 80% power by the end of the study. This is demonstrated in Figure 5.

Figure 5: Expected power at each trial stage using subgroup-specific adaptive randomization.

Bias in mortality rate estimation

The bias associated with using cohort RAR procedures is shown for all stages of the trial in Figure 6 and in sub-group specific and cohort RAR at the point where the trial reaches the sample size observed in RECOVERY in Figure 7. Both RMC_f and RMC_s are associated with higher bias than T_f and

T_s. Sub-group specific RAR results show bias is lowest in subgroup iii for both algorithms, with subgroup i having the highest bias for the T_s algorithm and subgroup ii having the highest bias for the RMC_s algorithm

Figure 6: Bias in calculating treatment effect across the trial using the cohort RAR approaches.

273

274

275 *Figure 7: Bias of treatment effect estimates in the full cohort and each subgroup using the four RAR schemes (T_f , RMC_f , T_s ,*
276 *RMC_s) at the sample size observed in RECOVERY (6425).*

277

278 Discussion

279 Understanding the results

280 It is important to consider why the different approaches led to different mortality rates. Full cohort
281 RAR approaches led to more patients receiving dexamethasone, which was shown to be superior to
282 standard care in most patients, and therefore led to fewer deaths overall in the trial simulation.
283 Subgroup RAR approaches improved mortality even further because they allowed more patients in
284 subgroup (i) to receive standard care and allowed dexamethasone allocation in subgroup (iii) to be
285 ramped up even faster.

Study power is important because it indicates how many patients are likely to be required to reach statistically significant results. Moreover, in the context of the pandemic, it would mean publishing positive results earlier which would lead to earlier use of the treatment in a real-world context. FeR showed the most study power. This is to be expected as Neyman's allocation formula indicated that the optimal allocation for study power is almost 1:1 (19). More generally, the power of all FR and cohort RAR strategies was seen to be inversely proportional to how much their average allocation ratio skewed away from 1:1. Although both FR approaches have more power than the cohort RAR approaches, it is important to mention FuR, the approach used in the RECOVERY trial, is only marginally more powerful despite leading to the most in-trial deaths. Nevertheless, the original RECOVERY trial still arguably benefitted from a larger control group because it had as many as five treatments trialled at any one time that would be compared to it.

Wider implications

Our results illustrate the trade-off in reducing patient deaths within the trial using RAR and getting statistically significant results earlier using FR. On one hand, publishing results earlier almost certainly led to lower overall deaths outside the trial. Patients in RECOVERY accounted for only 15% of all patients with COVID-19 in hospitals (20), and the subsequent use of dexamethasone has already been estimated to have saved a million lives globally (21). On the other hand, RAR designs have been shown to improve trial recruitment, precisely because patients understand they have a higher chance of receiving a well-performing treatment. (22). This could have arguably increased the sample size available to counteract any loss of power. Article 8 of the World Medical Association Declaration of Helsinki states that the goal of acquiring knowledge must not come before trial participants' interests (23), but a case can be made that FR procedures do just that.

There is no doubt that the implementation of RAR in a trial is more challenging than FR. This is especially true for subgroup specific RAR. While the subgroups used were declared before the start of the trial, it would be impossible to know that the treatment effect would be different across the

three groups, or whether other pre-specified subgroups should have been split for RAR instead. Using RAR at the subgroup level means splitting the sample, leading to smaller sample sizes in each group. This led to a low statistical power for subgroup (i) but a higher power for subgroup (iii) due to the large treatment effect. This meant that the subgroup where the treatment effect is greatest would benefit most from RAR.

Limitations

One important part of the model that remains unaccounted for is patient drift. Patient drift occurs when the trial cohort's characteristics, and therefore their likelihood of responding to a treatment, changes throughout the course of a trial (24). When using FR, the effect of patient drift will be minimised as any changes will be independent of treatment arm allocation (e.g. if patients present later in the trial had fewer comorbidities, both the standard care and dexamethasone groups would have exhibited lower mortality rate). This is not the case when using RAR because arms that performs better will receive a larger proportion of the patients as the trial progresses. Consequently, when calculating response rates, it must be noted that the characteristics of patients could potentially be unbalanced across arms. Throughout the pandemic, data shows the type of people susceptible to catching COVID-19 has changed dramatically. This occurred in terms of age-groups, ethnicities, socio-economic class, and geography (25). Likewise, the virus itself is likely to have changed, as mutations occur, and as different clones and variants become more common (26). Resources in treating the pandemic may also change, affecting how likely patients are to survive. For example, the typical care that COVID-19 patients get varies as clinical knowledge in treating the disease improves. Additionally, if more ventilators are procured or there are fewer COVID-19 patients in a hospital, a larger proportion of patients may be placed on ventilators by doctors. This has the dual effect of making the full cohort more likely to survive as more patients can receive adequate respiratory support and making the ventilator sub-group healthier as ventilators do not have to be reserved for only the most critical patients. Therefore, had RAR been applied, it is likely

the death rate would be skewed by the confounding effect of difference in patient, illness, and management characteristics.

Also notable is the simplicity of the simulation. RECOVERY has multiple trial arms, with arms being dropped and added dynamically. In contrast, we simulated it as a two-arm trial. Furthermore, for simplicity in the simulation, patient outcomes are generated in uniform batches. This contrasts with what happened in RECOVERY as, the number of hospitalised patients available to recruit varied significantly throughout the trial period, and this would have affected how RAR worked. There was a sharp decrease in hospitalised patients towards the end of the recruitment period (27). This would have likely meant that there would be more information attained at the start of the trial, and therefore the proportion of patients randomised to dexamethasone would have increased faster than it did in the simulation. Although our simulation study was simplistic, we believe the results paint a broadly accurate picture of how the operating characteristics of RECOVERY would differ using FR and RAR procedures.

Conclusion

Using RAR within RECOVERY could have resulted in more patients being given the optimal treatment, and therefore fewer deaths in the trial. The benefits of RAR were even more pronounced when used within pre-specified subgroups. However, more patients would have been required to attain the same study power under RAR, leading to a longer trial period (assuming the same recruitment rate). Bias in treatment effect estimation arises in RAR trials, but only to a negligible extent. The use of RAR deserves to be considered for use in future platform trials. The trade-off in balancing the needs of patients within and beyond the trial should be acknowledged by trialists more clearly, and patient groups themselves consulted before deciding what balance to strike.

List of abbreviations

RAR – Response Adaptive Randomization

360 FR – Fixed Randomization

361 COVID-19: coronavirus disease 2019

362 RECOVERY – Randomised Evaluation of COVID-19 Therapy

363 REMAP-CAP – Randomised, Embedded, Multi-factorial Adaptive Platform Trial for

364 Community Acquired Pneumonia

365 FeR – Fixed equal Randomization

366 FuR – Fixed unequal Randomization

367 T_f – Tuning protocol on the full patient cohort

368 RMC_f – REMAP-CAP randomization protocol on the full patient cohort

369 T_s – Tuning protocol on each subgroup individually

370 RMC_s – REMAP-CAP randomization protocol on each subgroup individually

371 **Declarations**

372 **Ethics approval and consent to participate**

373 Not applicable.

374 **Consent for publication**

375 Not applicable.

376 **Availability of data and materials**

377 Data sharing is not applicable to this article as no datasets were generated or analysed during the
378 current study.

379

380 Project name: RECOVERY RAR simulation

381 Project home page: https://github.com/ts482/RECOVERY_RAR

382 Operating system(s): Windows 10

383 Programming language: R statistical language

384 License: GNU GPL

385 **Competing interests**

386 The authors declare that they have no competing interests.

387 **Funding**

388 JB is funded by an Expanding Excellence in England (E3) research awarded to the University of
389 Exeter.

390 TS is funded by NHS Business Services Authority through the NHS Student Bursaries scheme.

391 BJ is supported by the National Institute for Health Research (NIHR) Applied Research Collaboration
392 (ARC) at the Royal Devon and Exeter NHS Foundation Trust. The views expressed in this publication
393 are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health
394 in England.

395

396 **Authors' contributions**

397 TS created the code simulation and wrote the first draft of the manuscript. JB contributed to the
398 conception and design of the work, code simulation and helped edit the manuscript. BJ made
399 substantial contributions to the conception and design of the work and helped edit the manuscript.

400 **Acknowledgements**

401 Not applicable.

Technical appendix

Technical Appendix A: Neyman rule for optimal allocation

Neyman allocation is a procedure that maximises study power for two arms given their estimated binary response rates (19). Rosenberger et al. (19) shows that optimal allocation ratio to treatment 1 for maximising power can be given by: $\frac{\sqrt{\hat{p}_1}}{\sqrt{\hat{p}_1} + \sqrt{\hat{p}_0}}$. In the case of the RECOVERY trial, $\hat{p}_1$ is the mortality rate on dexamethasone (22.9%) and $\hat{p}_0$ is the observed mortality rate on standard care (25.7%). Substituting these into the equation indicates that 49% of patients should receive dexamethasone to maximise the power of the study. Substituting the mortality rates on each treatment for the subgroups i, ii and iii (found in Table 1), the optimal allocation ratios to dexamethasone would be 53%, 49% and 46% respectively.

Technical Appendix B: Calculating the posterior outcome probability

To calculate the posterior probability that one treatment arm yields superior results to the other, a Bayesian beta-binomial model was assumed for each trial arm. The prior beta distribution can be given as $\pi(p|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1}$, where $B(\alpha, \beta) = \int_0^1 p^{\alpha-1} (1-p)^{\beta-1} dp$ and represents the Beta function. α and β are set to one. The binary outcome of each patient is mortality within 28 days after randomization and can be regarded as an independent Bernoulli trial. Therefore, a sample of n patients follows a binomial distribution, such that $f(y|p) = \binom{n}{y} p^y (1-p)^{n-y}$, where n is the total number of patients and y is the number of successes (or deaths within 28 days of randomization). Combining these distributions creates the posterior probability density function $\pi(p|y)$, given as $\frac{1}{B(y+\alpha, n-y+\beta)} p^{y+\alpha-1} (1-p)^{n-y+\beta-1}$, which is a beta($y+\alpha, n-y+\beta$) distribution.

The parameters indexing the posterior beta distribution are therefore $\alpha = y + \alpha$ and $\beta = n - y + \beta$. From this we can create a beta posterior distribution at each point of the interim analysis $i = 1, \dots, 100$ for each trial arm $j = 0$ (dexamethasone), 1 (standard care), and 100 points $k = 1, \dots, 100$. To estimate, for example, the posterior probability that the posterior mean for arm 0 is greater than the posterior mean of arm 1, we randomly generate 100 draws from each distribution and count the proportion of times that the arm 0 draw is the largest.

Technical Appendix C: Bowden and Trippa rule for bias following a RAR procedure

In general, FR schemes lead to unbiased estimates of the treatment effect. Conversely, RAR procedures induces bias into the treatment effect estimate, because the sample size of the trial arm becomes positively correlated with the estimate itself (28). Specifically, the bias of a proportion parameter is given by:

$$\frac{Cov[n_j, \hat{P}_j]}{E[n_j]} = P_j - E[\hat{P}_j] = -Bias(\hat{P}_j)$$

Where n_j is the sample size of treatment arm j , P_j is the true mortality rate in treatment arm j , and $\hat{P}_j$ is the estimated mortality rate in treatment arm j based on trial data.

References

1. CDC. Healthcare Workers [Internet]. Centers for Disease Control and Prevention. 2020 [cited 2021 Jun 24]. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/hcp/non-us-settings/overview/index.html>
2. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 [Internet]. [cited 2021 May 31]. Available from: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>
3. Coronavirus Pandemic (COVID-19) – the data - Statistics and Research [Internet]. Our World in Data. [cited 2021 Jun 1]. Available from: <https://ourworldindata.org/coronavirus-data>

- 450 4. Nabavi N. Long covid: How to define it and how to manage it. *BMJ*. 2020 Sep 7;370:m3489.
- 451 5. Whitaker M, Elliott J, Chadeau-Hyam M, Riley S, Darzi A, Cooke G, et al. Persistent symptoms
452 following SARS-CoV-2 infection in a random community sample of 508,707 people [Internet].
453 2021 Jun [cited 2021 Jun 27]. Available from:
454 <http://spiral.imperial.ac.uk/handle/10044/1/89844>
- 455 6. Oncology TL. COVID-19 and cancer: 1 year on. *Lancet Oncol*. 2021 Apr 1;22(4):411.
- 456 7. Torales J, O'Higgins M, Castaldelli-Maia JM, Ventriglio A. The outbreak of COVID-19 coronavirus
457 and its impact on global mental health. *Int J Soc Psychiatry*. 2020 Jun 1;66(4):317–20.
- 458 8. 2020 Year in Review: The impact of COVID-19 in 12 charts [Internet]. [cited 2021 Jun 27].
459 Available from: [https://blogs.worldbank.org/voices/2020-year-review-impact-covid-19-12-](https://blogs.worldbank.org/voices/2020-year-review-impact-covid-19-12-charts)
460 [charts](https://blogs.worldbank.org/voices/2020-year-review-impact-covid-19-12-charts)
- 461 9. University of Oxford. Randomised Evaluation of COVID-19 Therapy [Internet]. *clinicaltrials.gov*;
462 2021 Apr [cited 2021 May 27]. Report No.: NCT04381936. Available from:
463 <https://clinicaltrials.gov/ct2/show/NCT04381936>
- 464 10. Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med*. 2021 Feb 25;384(8):693–
465 704.
- 466 11. Abani O, Abbas A, Abbas F, Abbas M, Abbasi S, Abbass H, et al. Tocilizumab in patients
467 admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label,
468 platform trial. *The Lancet*. 2021 May 1;397(10285):1637–45.
- 469 12. Group RC, Horby PW, Mafham M, Peto L, Campbell M, Pessoa-Amorim G, et al. Casirivimab and
470 imdevimab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised,
471 controlled, open-label, platform trial. *medRxiv*. 2021 Jun 16;2021.06.15.21258542.
- 472 13. Recommendations | COVID-19 rapid guideline: managing COVID-19 | Guidance | NICE
473 [Internet]. NICE; [cited 2021 Jul 15]. Available from:
474 <https://www.nice.org.uk/guidance/ng191/chapter/Recommendations>
- 475 14. Angus DC, Alexander BM, Berry S, Buxton M, Lewis R, Paoloni M, et al. Adaptive platform trials:
476 definition, design, conduct and reporting considerations. *Nat Rev Drug Discov*. 2019
477 Oct;18(10):797–807.
- 478 15. Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson LV, et al. Adaptive
479 designs in clinical trials: why use them, and how to run and report them. *BMC Med*. 2018 Feb
480 28;16(1):29.
- 481 16. Proschan M, Evans S. Resist the Temptation of Response-Adaptive Randomization. *Clin Infect*
482 *Dis*. 2020 Dec 1;71(11):3002–4.
- 483 17. Interleukin-6 Receptor Antagonists in Critically Ill Patients with Covid-19. *N Engl J Med*. 2021
484 Apr 22;384(16):1491–502.
- 485 18. Bonten MJM. Randomized, Embedded, Multifactorial Adaptive Platform Trial for Community-
486 Acquired Pneumonia [Internet]. *clinicaltrials.gov*; 2020 Oct [cited 2021 May 26]. Report No.:
487 NCT02735707. Available from: <https://clinicaltrials.gov/ct2/show/NCT02735707>

- 488 19. Rosenberger WF, Stallard N, Ivanova A, Harper CN, Ricks ML. Optimal Adaptive Designs for
489 Binary Response Trials. *Biometrics*. 2001;57(3):909–13.
- 490 20. Wise J, Coombes R. Covid-19: The inside story of the RECOVERY trial. *BMJ*. 2020 Jul
491 8;370:m2670.
- 492 21. Robinson J. Steroid has saved the lives of one million COVID-19 patients worldwide, figures
493 show [Internet]. *The Pharmaceutical Journal*. [cited 2021 Aug 3]. Available from:
494 [https://pharmaceutical-journal.com/article/news/steroid-has-saved-the-lives-of-one-million-](https://pharmaceutical-journal.com/article/news/steroid-has-saved-the-lives-of-one-million-covid-19-patients-worldwide-figures-show)
495 [covid-19-patients-worldwide-figures-show](https://pharmaceutical-journal.com/article/news/steroid-has-saved-the-lives-of-one-million-covid-19-patients-worldwide-figures-show)
- 496 22. Tehranisa JS, Meurer WJ. Can response-adaptive randomization increase participation in acute
497 stroke trials? *Stroke*. 2014 Jul;45(7):2131–3.
- 498 23. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research
499 Involving Human Subjects. *JAMA*. 2013 Nov 27;310(20):2191.
- 500 24. Villar SS, Bowden J, Wason J. Response-adaptive designs for binary responses: How to offer
501 patient benefit while being robust to time trends? *Pharm Stat*. 2018;17(2):182–97.
- 502 25. Venkatesan P. The changing demographics of COVID-19. *Lancet Respir Med*. 2020 Dec
503 1;8(12):e95.
- 504 26. Luring AS, Hodcroft EB. Genetic Variants of SARS-CoV-2—What Do They Mean? *JAMA*. 2021
505 Feb 9;325(6):529.
- 506 27. UK: Daily new hospital admissions for COVID-19 [Internet]. *Our World in Data*. [cited 2021 Jul
507 27]. Available from: <https://ourworldindata.org/grapher/uk-daily-covid-admissions>
- 508 28. Bowden J, Trippa L. Unbiased estimation for response adaptive clinical trials. *Stat Methods*
509 *Med Res*. 2017 Oct;26(5):2376–88.

510