Report of the first seven agents in the I-SPY COVID trial: a phase 2, open label, adaptive platform randomised controlled trial

The I-SPY COVID Consortium[a]

Summary

Background An urgent need exists to rapidly screen potential therapeutics for severe COVID-19 or other emerging pathogens associated with high morbidity and mortality.

Methods Using an adaptive platform design created to rapidly evaluate investigational agents, hospitalised patients with severe COVID-19 requiring ≥6 L/min oxygen were randomised to either a backbone regimen of dexamethasone and remdesivir alone (controls) or backbone plus one open-label investigational agent. Patients were enrolled to the arms described between July 30, 2020 and June 11, 2021 in 20 medical centres in the United States. The platform contained up to four potentially available investigational agents and controls available for randomisation during a single time-period. The two primary endpoints were time-to-recovery (<6 L/min oxygen for two consecutive days) and mortality. Data were evaluated biweekly in comparison to pre-specified criteria for graduation (i.e., likely efficacy), futility, and safety, with an adaptive sample size of 40–125 individuals per agent and a Bayesian analytical approach. Criteria were designed to achieve rapid screening of agents and to identify large benefit signals. Concurrently enrolled controls were used for all analyses. https://clinicaltrials.gov/ct2/show/NCT04488081.

Findings The first 7 agents evaluated were cenicriviroc (CCR2/5 antagonist; n = 92), icatibant (bradykinin antagonist; n = 96), apremilast (PDE4 inhibitor; n = 67), celecoxib/famotidine (COX2/histamine blockade; n = 30), IC14 (anti-CD14; n = 67), dornase alfa (inhaled DNase; n = 39) and razuprotafib (Tie2 agonist; n = 22). Razuprotafib was dropped from the trial due to feasibility issues. In the modified intention-to-treat analyses, no agent met pre-specified efficacy/graduation endpoints with posterior probabilities for the hazard ratios [HRs] for recovery ≤1.5 between 0.99 and 1.00. The data monitoring committee stopped Celecoxib/Famotidine for potential harm (median posterior HR for recovery 0.5, 95% credible interval [CrI] 0.28–0.90; median posterior HR for death 1.67, 95% CrI 0.79–3.58).

Interpretation None of the first 7 agents to enter the trial met the prespecified criteria for a large efficacy signal. Celecoxib/Famotidine was stopped early for potential harm. Adaptive platform trials may provide a useful approach to rapidly screen multiple agents during a pandemic.

Funding Quantum Leap Healthcare Collaborative is the trial sponsor. Funding for this trial has come from: the COVID R&D Consortium, Allergan, Amgen Inc., Takeda Pharmaceutical Company, Implicit Bioscience, Johnson & Johnson, Pfizer Inc., Roche/Genentech, Apotex Inc., FAST Grant from Emergent Venture George Mason University, The DoD Defense Threat Reduction Agency (DTRA), The Department of Health and Human Services Biomedical Advanced Research and Development Authority (BARDA), and The Grove Foundation. Effort sponsored by the U.S. Government under Other Transaction number W15QKN-16-9-1002 between the MCDC, and the Government.

Introduction

At the dawn of the SARS-CoV-2 pandemic in 2020, understanding of the pathogenesis of respiratory failure due to this virus was limited, and differences compared to other viral pneumonias and other causes of acute respiratory distress syndrome were unclear. In response, a wide variety of potential therapeutic approaches were proposed, with explosive growth in the number of new clinical trials. Innovative large-scale Phase 3 platform trials, such as RECOVERY and

---

[a]List provided in Appendix.
Corresponding author. D. Clark Files, Atrium Health Wake Forest Baptist, Medical Centre Boulevard, Winston-Salem, NC 27104, USA.
Email address: clark.files@wakehealth.edu (C. Files).

Research in context

Evidence before this study

At the outset of the COVID-19 pandemic, there was a major unmet need for a trial mechanism to rapidly screen potentially useful therapeutic agents but there was limited knowledge about the optimal approach to such an effort. To date, most Phase 2 clinical trials in severe COVID-19 have been stand-alone trials of a single agent compared to placebo. A Pubmed search using the terms "clinical trials" "COVID-19" "severe" and "therapeutics" were performed from dates ranging from January 2020 to December 2022 to review the available evidence. Adaptive platform trials offer the ability to adapt to emerging science and evaluate multiple agents under a single protocol for further testing in definitive trials.

Added value of this study

The use of an adaptive platform trial mechanism with a Bayesian statistical design with mortality and rate of recovery

as the primary endpoints successfully tested several new therapies among a geographically diverse patient population of hospitalised severe COVID-19 patients in both academic and community hospitals. None of the first seven agents tested on this platform met the prespecified graduation/efficacy endpoints and one investigational agent was terminated early for possible harm.

Implications of all the available evidence

Lessons learned from the ISPY COVID experience will provide useful for evaluating novel therapeutics in the future; this design may have value for evaluating therapeutics in critical illnesses including sepsis, acute hypoxemic respiratory failure, and acute respiratory distress syndrome.

REMAP-CAP, had a major impact on clinical care by identifying significant benefits of glucocorticoids and other investigational agents in severe COVID-19.[1,2] Though early in the pandemic, a phase 2 platform to screen biologically plausible effective agents that could markedly reduce severe COVID-19 morbidity and mortality for subsequent Phase 3 studies was urgently needed. At the time, most Phase 2 studies for severe COVID-19 remained single agent trials, each requiring substantial time to identify sites and assemble trial machinery before launching, only to disband after investigation of a single agent. Moreover, most patients in the United States have not had access to clinical trials for novel therapies for severe COVID-19 because enrollment is frequently restricted to high-volume academic centres.[3]

To address these challenges, we developed and implemented an investigator-initiated Phase 2 adaptive platform trial to test potential therapies for severe COVID-19, conducted at geographically dispersed sites across the United States that enrolled a diverse population and included centres that may not traditionally participate in clinical trials (e.g., community hospitals, integrated delivery networks). Our objective, similar to the AGILE-ACCORD platform trial[4] in less severe patients, was to identify agents with large signals of benefit for further testing in Phase 3 trials. This report summarises key results from the first seven agents evaluated in the trial, which has adapted over time and continues to enroll participants in subsequent study arms.

Methods

Trial design and oversight

The I-SPY COVID trial is a multicentre phase 2 adaptive platform trial designed to rapidly screen potential

therapeutics for severe COVID-19 (NCT04488081). A detailed description of the study rationale and protocol have been previously published.[5,6] The study is conducted under a central IRB mechanism (Wake Forest University Health Sciences IRB00066805). Briefly, the study consists of a randomised cohort consisting of eligible patients who consent to participation and an observational cohort in which no treatment is assigned but clinical data are collected through medical records. The observational cohort will be reported in a subsequent manuscript. Participants in the randomised cohort were randomised to one of up to five separate treatment arms, including a control arm and up to four separate open-label investigational agent arms (Supplemental Fig. S1). An individual participant was never randomised to receive more than one investigational agent and each investigational agent arm was compared to participants randomised to the control arm during the same time epoch. The number of active agents on the platform ranged from three to four during the time-period reported here. All participants received a backbone treatment regimen of corticosteroids (dexamethasone 6 mg recommended) and remdesivir, based on the results of the ACTT-1 Study[7,8] and the RECOVERY Collaborative Group.[1] A master protocol governs trial conduct and operations and permits investigational regimens to enter and leave the study using a protocol amendment without interrupting enrollment.[9] The full study protocol active during the time of conduct reported here is available in the Supplement.

The trial was designed and is governed by an independent group of investigators whose members are blinded to information about outcomes until enrollment and 28-day outcomes in each arm are complete; an independent data monitoring committee (DMC) reviewed unblinded data on a bi-weekly basis with unblinded

study statisticians (see Supplement for details). The study protocol and associated amendments were approved by a central institutional review board (Wake Forest University; Winston-Salem, North Carolina), with reliance agreements from participating sites. The trial is conducted in accordance with Good Clinical Practice guidelines and the Helsinki Declaration, and all participants, or their surrogates, provide informed consent prior to enrollment in the interventional arm of the study.

Participants

During this period which includes patients randomised between 30 July 2020 and 11 June 2021, participants were enrolled at 20 medical centres across the United States, representing a mix of academic and community hospitals, and integrated delivery networks (Supplemental Fig. S2). Hospitalised patients aged ≥18 years with confirmed SARS-CoV-2 infection and a requirement for at least 6 L/min nasal cannula oxygen were eligible. Major exclusion criteria included duration of high-flow oxygen requirement or mechanical ventilation >72 h, severe liver disease, need for renal replacement therapy, co-enrollment in another clinical trial for a novel therapeutic agent requiring an IND, or estimated six-month mortality of >50% from underlying chronic conditions. Additional exclusion criteria existed for specific agents where appropriate. If a patient met master protocol inclusion/exclusion but subsequently randomised to an investigational agent where they met an agent specific exclusion, that participant was moved to the control arm of the study, but not analysed as a concurrent control for study outcomes. Using this method, the concurrent control arm for each individual agent only consists of control patients that were eligible for that investigational agent. A detailed list of master protocol inclusion and exclusion criteria and drug specific appendices are available in the study protocol in the Supplemental Data.

Agent selection

Eleven candidate investigative agents were initially identified through a partnership with the COVID R&D Consortium, an industry-led organisation formed to accelerate new COVID-19 therapies and vaccines. Of these, four were selected for trial inclusion by the I-SPY COVID Agent Committee (apremilast, icatibant, cenicriviroc, razuprotafib) based on potential mechanistic benefit, risk profile, scalability, and availability; subsequent agents were proposed by investigators, the Defense Threat Reduction Agency (DTRA; one of the study funders), and the Biomedical Advanced Research and Development Authority (BARDA; another study funder). Celecoxib/famotidine, dornase alfa, and IC14 were the next agents that were ready for testing and selected by the Agents Committee, as previously described.[6] The doses and duration of the agents were those proposed by industry partners and investigators and generally drawn from studies in other populations (see Supplemental Table S2). Given the urgent nature of the pandemic, dose finding studies were not a part of this trial. This manuscript describes the results of these first seven agents that entered the trial. Subsequent agents on the ISPY COVID platform will be reported later.

Randomisation and masking

Study staff and investigators contacted eligible patients, or their surrogates, to assess interest in study participation. Interested patients were randomised (stratified by site and requirement for mechanical ventilation) to either the control arm (backbone alone) or backbone plus one of up to 4 investigational agents active in the trial at any one time. Multiple parallel active agent arms with a common control arm can increase the efficiency of clinical trial. A higher allocation to the control arm was used to reduce stochastic noise within that arm, which affects all comparisons. A ratio of 2 (control):1:1:1:1 initially, which was changed to 1.4 (control):1:1:1:1 in January 2021 to enhance enrollment in the interventional arms. The effect of this change on trial operating characteristics was assessed using simulations (see Statistical analysis section and Supplement for details). During study design, patient advocates advised that detailed discussion by study teams of all five potential treatment arms would be too onerous and confusing for potential participants.[6] After randomisation, study teams re-approached participants or surrogates for informed consent, discussing the individual agent's rationale and potential risks and benefits in the setting of severe COVID-19. Investigational agents were dispensed by local research pharmacies in an open-label design.

Procedures

Sites provided non-study related treatment in accordance with local standards of care; all sites were strongly encouraged to adhere to best practices for management of severe COVID-19 and ARDS, including low tidal volume ventilation for mechanically ventilated patients. At enrollment, demographics, severity of illness, comorbidities, and concomitant medical therapies were collected. Follow up information was collected daily for clinical status and COVID-19 severity using a modified WHO ordinal scale (Supplemental Table S1),[10] vital signs and key laboratory values, adverse events, and subsequent treatments administered.

Outcomes

The initial primary outcome for the trial was time to durable recovery, defined as requiring less than 6 L/min oxygen for at least 2 consecutive days. In May 2021, after 1254 patients had been enrolled (677 in the randomised cohort and 577 in the observational cohort), in-hospital

mortality was added as a primary endpoint to form a family of two primary endpoints. This change was prompted by study leadership and the DMC based on emerging evidence from other trials on how "recovery" endpoints operated during severe COVID-19 and the likely value of including mortality as a primary endpoint.[11] Pre-specified secondary outcomes included progression to mechanical ventilation or death (for patients who were not mechanically ventilated at enrollment) and Common Terminology Criteria for Adverse Events (CTCAE) Grade 3 or higher adverse events.

Statistical analysis

The primary analysis was conducted on the modified intention-to-treat (mITT) population, consisting of randomised patients who provided informed consent. In addition, as pre-specified in the statistical analysis plan (SAP; Supplement), we performed analyses on the intention-to-treat (ITT) population, which in this design included patients who were randomised but did not sign consent to proceed to treatment with the agent assigned. The outcomes of randomised patients who did not consent to treatment with the assigned agent were tracked in the observational cohort and could thus be included in the ITT analysis. Further information about these analyses is included in the Supplementary Methods. Of note, in the statistical analysis plan included in the data supplement, the mITT is referred to as the ITT group, while the ITT group is referred to as the "super ITT" group. For all analyses, *only* concurrent and eligibility matched control patients were included in the analysis (i.e., patients who met the agent-specific eligibility criteria and were enrolled contemporaneously to that agent being active on the platform).

Bayesian proportional-hazard Weibull regression was used to model the cause-specific hazard function for recovery as a function of study arm, adjusting for baseline COVID severity. Death prior to recovery was treated as a competing event. Patients were censored at 60 days if they were still alive and not recovered. Similarly, Bayesian proportional-hazard Weibull regression was used to model the hazard functions for all-cause mortality, and for time to mechanical ventilation or death for patients who were not mechanically ventilated at baseline. In these analyses, follow-up times were censored at 60 days if patients were still at risk. Weakly informative priors were used for both models. Models were fit using Markov chain Monte Carlo methods with 4000 draws from the joint posterior distribution; we report medians of the posterior hazard ratio (HR) distributions for investigational agents compared with concurrently randomised controls, and 95% two-sided quantile-based credible intervals. Detailed descriptions of the models are found in the SAP.

During the trial, HRs for recovery and mortality, as well as the cumulative incidence and mortality functions, were updated every two weeks for each arm and reviewed by the DMC. Enrollment in an arm was planned to be stopped if the HRs met predefined criteria for efficacy (>97.5% posterior probability of HR for recovery >1 [i.e., higher recovery rate] OR >90% posterior probability of HR for mortality <1 [i.e., reduced mortality], in which case an agent would 'graduate'), or futility (>90% posterior probability of HR for recovery <1.5 AND <50% posterior probability for HR of mortality < 1) in the mITT population. The criteria based on mortality were added after 677 randomised patients (see 'Outcomes'); however, prior to this time point, the DMC was considering mortality as a key secondary endpoint when evaluating continuation or termination of specific arms. The criteria were designed to achieve rapid screening of agents and to identify large signals for benefit, acknowledging potential type II errors for smaller treatment effects. In the context of this trial, graduation would indicate that an agent should be prioritised for further testing in a larger Phase 3 trial. At least 40 enrolled patients were required to drop an arm for futility, and at least 50 were required for graduation; the maximum allowed number of patients in an arm was 125. The DMC met every 2 weeks and could stop an agent for safety concerns at any time. See the DMC Charter for additional details.

Simulations were used to estimate trial operating characteristics under a variety of possible scenarios ranging from pessimistic, in which the null hypothesis of no benefit holds for every treatment, to optimistic cases in which several of the regimens are truly effective. The bi-weekly interim analyses were accounted for in the simulations, as were different allocation ratios to the control arm. The power ranged from high ($\geq$85%) for scenarios with big effect sizes on recovery ($HR_r \geq 1.7$) or mortality rates ($HR_m \leq 0.5$) to low (around 20%) for small effect sizes ($HR_r = 1.2$ and $HR_m = 0.8$, respectively). The type 1 error rate varied between 4% and 17%, depending on scenario, which was considered acceptable for a phase 2 signal seeking trial. Multiplicity were not accounted for in the analyses of the secondary outcomes. See the Statistical Analysis Plan for additional details.

Role of the funding source

The funders had no role in data collection or analysis or preparation of this manuscript; some funders reviewed this manuscript for accuracy. All authors had access to the dataset and take responsibility for the content and submission of this manuscript.

Results

Patients were enrolled to the arms described in this manuscript between July 30, 2020 and June 11, 2021 (Supplemental Tables S2 and S3). Of the 1101 patients who were eligible and offered participation in the trial during this period, 814 were randomised, and 642
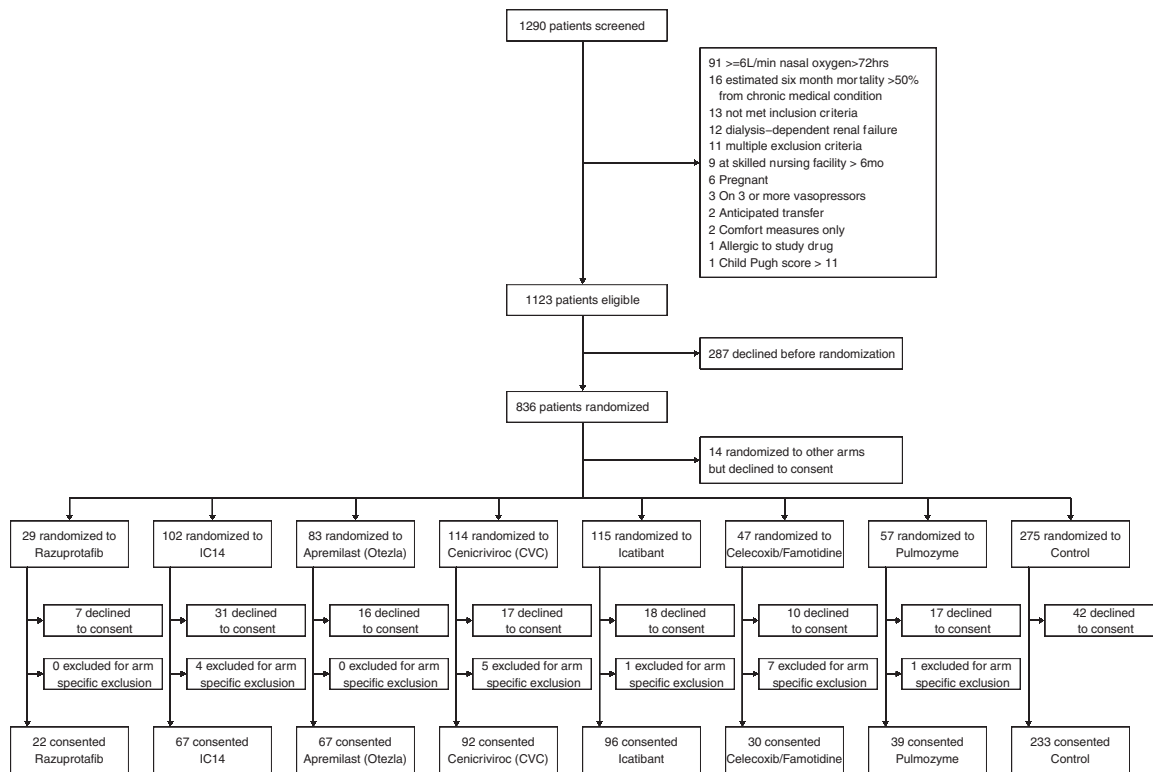
Fig. 1: Consort flow chart of the ISPY COVID trial. Among the 172 participants who declined consent, the number and percentage of individuals who declined to sign consent after being randomised to a particular agent were: 7 (24%) for Razuprotafib, 16 (19.3%) for Apremilast, 10 (21.3%) for FamCox, 17 (14.9%) for CVC, 31 (30.4%) for IC14, 18 (15.7%) for Icatibant, 17 (29.8%) for Pulmozyme, and 42 (15.3%) for Control. The remaining 14 individuals who declined consent were randomised to other investigational arms that were started before the discontinuation of agents reported in this manuscript. Individuals that consented to Control arm were "shared" among investigational agent arms if they were randomised in a time epoch that included the active investigational agent.

provided informed consent and were enrolled in the randomised cohort to one of the six treatment or control arms (Fig. 1). Among the 172 individuals who declined to sign consent for an investigational agent arm or control, the number and percentage of individuals that declined to participate in each randomised arm include: 7 (24%) for razuprotafib, 16 (19.3%) for apremilast, 10 (21.3%) for famotidine/celecoxib, 17 (14.9%) for CVC, 31 (30.4%) for IC14, 18 (15.7%) for icatibant, 17 (29.8%) for pulmozyme, and 42 (15.3%) for control. The remaining 14 individuals who declined consent were randomised to other investigational arms that were started before the discontinuation of agents reported in this manuscript.

The mean age of enrolled patients ranged from 59 to 67, and the majority of participants were male (Table 1 and Supplemental Table S4). Enrollment reflected high rates of underrepresented minorities in the United States, with 16–42% of participants identifying as LatinX and 12–23% identifying as Black. The majority of participants were overweight or obese (77–91%, depending on study arm), and most participants had at least one major comorbidity, with hypertension, diabetes, and chronic lung or kidney disease being the most common. The average duration of COVID-19 symptoms prior to enrollment was 9–10 days; the average duration of hospitalisation prior to enrollment was 2 days. At baseline, most participants (83–94%, across study arms) were not mechanically ventilated. On study day 1 (day after enrollment), relatively few required vasopressors (6–15%, across study arms), but the majority were on treatment-dose antimicrobials across all arms.

The four initial study agents were apremilast (phosphodiesterase-4 inhibitor), cenicriviroc (CCR2/5 antagonist), icatibant (bradykinin antagonist), and razuprotafib (Tie2 agonist). The next three agents to enter the trial were celecoxib/famotidine (cyclooxygenase-2 and histamine blockers); IC-14 (anti-CD14 monoclonal antibody), and inhaled dornase alfa (recombinant DNAse) (Supplemental Fig. S3 and Supplemental Table S2). The number of patients enrolled and evaluated in the mITT analyses for each arm was as follows: razuprotafib n = 22; apremilast n = 67; cenicriviroc n = 92; icatibant n = 96; celecoxib/famotidine n = 30; dornase n = 39; IC14 n = 67; controls n = 233.

Articles

Characteristics	Razuprotafib	Razuprotafib controls	Apremilast	Apremilast controls	Cenicriviroc	Cenicriviroc controls	Icatibant	Icatibant controls	Celecoxib/ famotidine	Celecoxib/ famotidine controls	Dornase	Dornase controls	IC14	IC14 controls
	(n = 22)	(n = 67)	(n = 67)	(n = 143)	(n = 92)	(n = 169)	(n = 96)	(n = 183)	(n = 30)	(n = 37)	(n = 39)	(n = 88)	(n = 67)	(n = 76)
Baseline data														
Demographics														
Age (years, mean ± SD)	67 ± 10	66 ± 13	66 ± 14	67 ± 14	67 ± 13	67 ± 15	64 ± 14	67 ± 14	65 ± 12	56 ± 17	63 ± 14	61 ± 17	59 ± 16	60 ± 17
Female sex (n (%))	7 (32%)	32 (48%)	22 (33%)	56 (39%)	31 (34%)	63 (37%)	31 (32%)	71 (39%)	8 (27%)	11 (30%)	16 (41%)	31 (35%)	25 (37%)	27 (36%)
Hispanic or Latino	4 (18%)	11 (16%)	13 (19%)	39 (27%)	28 (30%)	52 (31%)	31 (32%)	56 (31%)	8 (27%)	18 (49%)	13 (33%)	41 (47%)	28 (42%)	36 (47%)
Race (n (%))														
American Indian or Alaska Native	0 (0%)	2 (3%)	0 (0%)	2 (1%)	1 (1%)	2 (1%)	1 (1%)	2 (1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Asian	0 (0%)	0 (0%)	3 (4%)	2 (1%)	5 (5%)	3 (2%)	3 (3%)	3 (2%)	1 (3%)	1 (3%)	2 (5%)	2 (2%)	4 (6%)	1 (1%)
Black or African American	3 (14%)	13 (19%)	14 (21%)	27 (19%)	11 (12%)	29 (17%)	15 (16%)	34 (19%)	9 (30%)	7 (19%)	9 (23%)	15 (17%)	10 (15%)	14 (18%)
Native Hawaiian or other Pacific Islander	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)	0 (0%)	1 (3%)	1 (3%)	1 (1%)	0 (0%)	1 (1%)
White or Caucasian	15 (68%)	36 (54%)	37 (55%)	72 (50%)	44 (48%)	81 (48%)	49 (51%)	86 (47%)	12 (40%)	15 (41%)	20 (51%)	37 (42%)	31 (46%)	33 (43%)
Two or more races not described above	2 (9%)	3 (4%)	3 (4%)	10 (7%)	12 (13%)	17 (10%)	13 (14%)	18 (10%)	1 (3%)	1 (3%)	0 (0%)	10 (11%)	0 (0%)	8 (11%)
Unknown or not reported	2 (9%)	14 (21%)	10 (15%)	30 (21%)	19 (21%)	38 (22%)	14 (15%)	41 (22%)	7 (23%)	12 (32%)	7 (18%)	23 (26%)	22 (33%)	19 (25%)
COVID severity at baseline, (n (%))														
WHO Level 5 (≥6 L supplemental oxygen)	19 (86%)	52 (78%)	52 (78%)	105 (73%)	66 (72%)	126 (75%)	65 (68%)	133 (73%)	22 (73%)	20 (54%)	31 (79%)	65 (74%)	61 (91%)	56 (74%)
WHO Level 5 (non-invasive ventilation or HFNO)	3 (14%)	2 (3%)	8 (12%)	15 (10%)	13 (14%)	16 (9%)	15 (16%)	17 (9%)	5 (17%)	1 (3%)	4 (10%)	6 (7%)	2 (3%)	3 (4%)
WHO Level 6 or 7 (mechanically ventilated)	0 (0%)	13 (19%)	7 (10%)	23 (16%)	13 (14%)	27 (16%)	16 (17%)	33 (18%)	3 (10%)	4 (11%)	4 (10%)	17 (19%)	4 (6%)	17 (22%)
Comorbidities at baseline (n (%))														
Any of listed conditions	22 (100%)	66 (96%)	65 (97%)	138 (97%)	87 (95%)	161 (95%)	94 (98%)	175 (96%)	22 (73%)	23 (62%)	31 (79%)	67 (76%)	49 (73%)	54 (71%)
Congestive heart failure	1 (5%)	4 (6%)	8 (12%)	8 (6%)	6 (7%)	12 (7%)	8 (8%)	11 (6%)	3 (10%)	2 (5%)	0 (0%)	6 (7%)	3 (4%)	5 (7%)
Diabetes	7 (32%)	22 (33%)	21 (31%)	49 (34%)	38 (41%)	57 (34%)	39 (41%)	63 (34%)	14 (47%)	13 (34%)	10 (26%)	31 (35%)	19 (28%)	26 (34%)
Kidney disease	4 (18%)	9 (13%)	7 (10%)	24 (17%)	13 (14%)	24 (14%)	10 (10%)	28 (15%)	3 (10%)	3 (8%)	2 (5%)	9 (10%)	3 (4%)	7 (9%)
Hypertension	13 (59%)	39 (58%)	40 (60%)	90 (63%)	58 (63%)	105 (62%)	64 (67%)	113 (62%)	18 (60%)	15 (41%)	23 (59%)	45 (51%)	37 (55%)	36 (47%)
Pulmonary disease	2 (9%)	14 (21%)	15 (22%)	28 (20%)	21 (23%)	35 (21%)	26 (27%)	38 (21%)	5 (17%)	1 (3%)	7 (18%)	12 (14%)	10 (15%)	12 (16%)
Solid organ transplant	1 (5%)	4 (6%)	7 (10%)	7 (5%)	4 (4%)	7 (4%)	7 (7%)	10 (5%)	0 (0%)	3 (8%)	1 (3%)	7 (8%)	0 (0%)	4 (5%)
Medication use at enrollment: steroids (n (%))														
Yes	22 (100%)	62 (93%)	65 (97%)	134 (94%)	86 (93%)	155 (92%)	89 (93%)	169 (92%)	26 (86%)	33 (89%)	35 (90%)	80 (91%)	56 (84%)	67 (88%)
No	0 (0%)	4 (6%)	1 (1.5%)	7 (5%)	6 (7%)	9 (5%)	5 (5%)	9 (5%)	2 (7%)	1 (3%)	4 (10%)	2 (2%)	6 (9%)	2 (3%)
Unknown	0 (0%)	1 (1%)	1 (1.5%)	2 (1%)	0 (0%)	5 (3%)	2 (2%)	5 (3%)	2 (7%)	3 (8%)	0 (0%)	6 (7%)	5 (7%)	7 (9%)
Medication use at enrollment: remdesivir (n (%))														
Yes	20 (91%)	53 (79%)	62 (93%)	106 (74%)	79 (86%)	123 (73%)	79 (82%)	135 (74%)	23 (77%)	28 (75%)	32 (82%)	67 (76%)	57 (85%)	57 (75%)
No	2 (9%)	12 (27%)	4 (6%)	34 (24%)	12 (13%)	41 (24%)	16 (17%)	43 (23%)	4 (13%)	8 (22%)	7 (18%)	18 (21%)	8 (12%)	16 (21%)
Unknown	0 (0%)	2 (3%)	1 (1%)	3 (2%)	1 (1%)	5 (3%)	1 (1%)	5 (3%)	3 (10%)	1 (3%)	0 (0%)	3 (3%)	2 (3%)	3 (4%)
Cointerventions on day 1 (n (%))														
Shock on vasopressors	0 (0%)	10 (15%)	8 (12%)	22 (15%)	8 (9%)	24 (14%)	11 (11%)	29 (16%)	2 (7%)	2 (5%)	6 (15%)	11 (13%)	4 (6%)	9 (12%)
Treatment dose antimicrobials	12 (55%)	37 (55%)	41 (61%)	81 (57%)	53 (58%)	94 (56%)	67 (70%)	104 (57%)	17 (57%)	15 (41%)	20 (51%)	48 (55%)	36 (54%)	41 (54%)

(Table 1 continues on next page)

Characteristics	Razuprotafib (n = 22)	Razuprotafib controls (n = 67)	Apremilast (n = 67)	Apremilast controls (n = 143)	Cenicriviroc (n = 92)	Cenicriviroc controls (n = 169)	Icatibant (n = 96)	Icatibant controls (n = 183)	Celecoxib/famotidine (n = 30)	Celecoxib/famotidine controls (n = 37)	Dornase (n = 39)	Dornase controls (n = 88)	IC14 (n = 67)	IC14 controls (n = 76)
(Continued from previous page)														
On RRT (renal replacement therapy) on day 1 (n (%))														
Yes	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)	2 (2%)	0 (0%)	0 (0%)	0 (0%)	2 (5%)	0 (0%)	0 (0%)	0 (0%)
Creatinine, mg/dL on day 1														
Mean ± SD	1.1 ± 0.5	1.1 ± 0.7	1.2 ± 0.6	1.2 ± 0.8	1.2 ± 0.7	1.2 ± 0.8	1.1 ± 0.8	1.2 ± 0.8	0.9 ± 0.3	1.0 ± 0.6	1.0 ± 0.7	1.2 ± 1.0	0.9 ± 0.4	1.2 ± 0.8
Platelets, thousands on day 1														
Mean ± SD	269 ± 95	275 ± 109	266 ± 121	270 ± 101	255 ± 112	267 ± 101	263 ± 90	264 ± 98	254 ± 90	286 ± 135	268 ± 94	247 ± 83	292 ± 115	250 ± 81
Total bilirubin (mg/dL) on day 1														
Mean ± SD	0.8 ± 0.5	0.6 ± 0.3	0.6 ± 0.4	0.6 ± 0.5	0.6 ± 0.7	0.6 ± 0.5	0.5 ± 0.3	0.6 ± 0.5	0.6 ± 0.2	0.7 ± 0.5	0.7 ± 0.6	0.6 ± 0.6	0.7 ± 0.4	0.5 ± 0.3

Patients randomised to the Control arm were "shared" among investigational agent if they were randomised during the same time epoch that the investigational agent was on the Platform.

***Table 1*: Selected baseline and day 1 characteristics by enrollment arm.**

The razuprotafib arm was terminated early for logistical reasons: specifically, because concerns for treatment-related decreases in blood pressure, and monitoring requirements made it challenging to administer this agent during a pandemic when sites were experiencing considerable strain. These concerns prompted a pre-planned pause in enrolment to the razuprotafib arm after enrollment of 14 patients to the arm (safety cohort) before continuing to enroll in total 22 patients, after which the agent was discontinued. At the time of discontinuation, the posterior probability of HR >1 for recovery and HR <1 for mortality were 0.43 and 0.32, respectively, in the mITT population. Detailed results for this agent will be reported separately. The remaining six investigational agents described here met pre-specified criteria for futility for the primary endpoint of time-to-recovery (Table 2), with posterior probabilities of 0.99–1.00 in the mITT population. Enrollment in each arm was terminated upon meeting this criterion, except for celecoxib/famotidine, which was terminated early (n = 30) by the DMC for potential harm. At the time of agent termination, posterior probabilities of HR for recovery >1, which were used to assess agents' graduation, ranged from 0.01 for celecoxib/famotidine to 0.23 for cenicriviroc. Posterior probabilities of HR <1 for mortality ranged from 0.09 for celecoxib/famotidine to 0.72 for IC14. Fig. 2 depicts the time-to-recovery for each agent compared to concurrent controls. Fig. 3 depicts survival over 60 days for each agent compared to concurrent controls. The observed proportions of patients who recovered, or died by 60 days of follow-up for each agent, and concurrently randomised controls are presented in Table 3 and Supplemental Table S5.

The median posterior hazard ratios (HR) for recovery and death in the mITT population are provided in Table 4, and the model-based cumulative probabilities are shown in Supplemental Figs. S4 and S5. The recovery rate was worse in the celecoxib/famotidine and IC14 arms compared to concurrent controls (median posterior HRs 0.50, 95% CrI 0.28–0.90; and 0.63; 95% CrI 0.40–0.96, respectively; HR < 1 indicates lower recovery rate). For the other 4 agents, the posterior cause-specific hazard for recovery was not different than controls, though major improvements in recovery were excluded (upper bound of 95% CrI's ≤ 1.2). Credible intervals for the endpoint of mortality were wider (Table 4). Patterns observed in mITT analysis were mirrored in ITT analyses (Supplemental Tables S6 and S7; Supplemental Figs. S6 and S7). Of note, a small number of patients were discharged after only one day of requiring <6 L oxygen and therefore did not meet the primary endpoint of durable recovery (48 h). These patients were censored for the primary analyses, and a sensitivity analysis treating these patients as recovered did not substantively impact the results (Supplemental Table S8).

Outcome	Action	Criterion	Arm	Posterior probability[a]
Recovery (HR > 1 is better)	Graduation	Pr(HRr > 1.0 | data) ≥ 0.975	Apremilast	0.101
			Cenicriviroc	0.225
			Icatibant	0.163
			IC14	0.017
			Celecoxib/famotidine	0.011
			Dornase	0.125
	Futility	Pr(HRr < 1.5 | data) ≥ 0.900	Apremilast	**0.999**
			Cenicriviroc	**0.998**
			Icatibant	**1.000**
			IC14	**1.000**
			Celecoxib/famotidine	**1.000**
			Dornase	**0.998**
Death (HR < 1 is better)	Graduation	Pr(HRm < 1.0 | data) ≥ 0.900	Apremilast	0.428
			Cenicriviroc	0.138
			Icatibant	0.389
			IC14	0.716
			Celecoxib/famotidine	0.089
			Dornase	0.383
	Futility	Pr(HRm < 1.0 | data) ≤ 0.500	Apremilast	**0.428**
			Cenicriviroc	**0.138**
			Icatibant	**0.389**
			IC14	0.716
			Celecoxib/famotidine	**0.089**
			Dornase	**0.383**

Bold posterior probability indicates criterion was met. *Abbreviations*: Hazard ratios (HR) as follows: HRr (*recovery*, in which values >1 indicate better outcomes, i.e., higher instantaneous recovery rate), HRm (mortality, in which values <1 indicate better outcomes, i.e., longer time to death), Pr (probability). [a]The posterior probability column shows the probabilities Pr(HRr > 1.0 \| data), Pr(HRr < 1.5 \| data), and Pr(HRm < 1.0 \| data) (i.e., the corresponding probabilities after having observed the data in the trial and given the priors and the model, as specified in the SAP). Since the same posterior probability, Pr(HRm < 1.0 \| data), is used in both the graduation and the futility criteria for death, its numerical value is the same for all agents in the rows corresponding to the graduation and the futility criteria (e.g., 0.716 for IC14). This probability is however compared to different thresholds in the graduation and futility criteria (≥0.9 for graduation and ≤0.5 for futility).

***Table 2*: Posterior probabilities for stopping criteria by agent, action, and outcome: modified intention-to-treat population.**
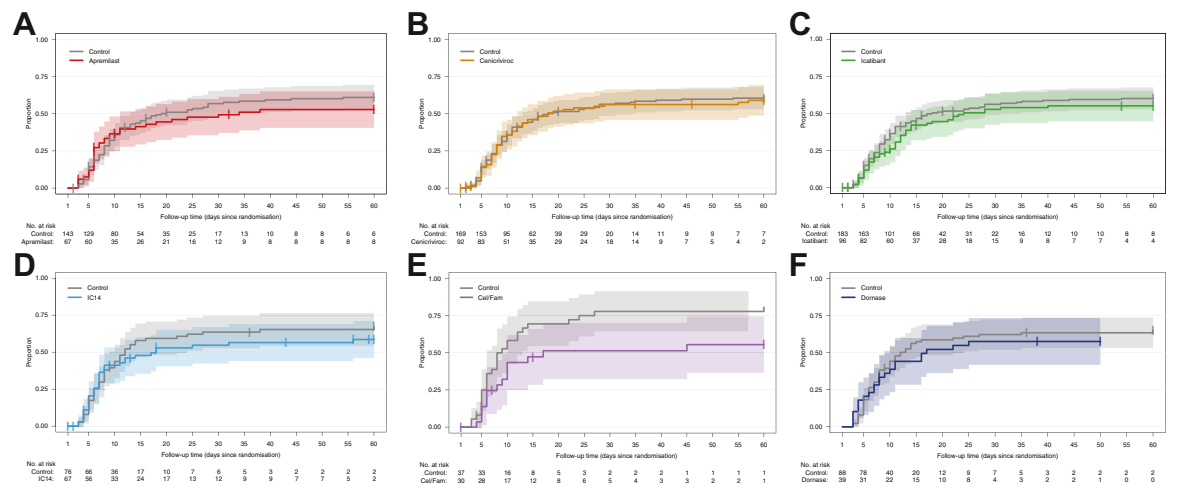
***Fig. 2*: Time-to-Recovery, Modified Intention-To-Treat Population.** Aalen-Johansen cumulative incidence curves showing the proportion of recovered patients in the mITT population up to 60 days of follow-up post randomisation in the presence of death prior to recovery as a competing risk. Numbers below each panel show patients at risk across time. Shaded areas show 95% pointwise confidence intervals. A. Apremilast arm, B. Cenicriviroc arm, C. Icatibant arm, D. IC14 arm, E. Celecoxib/Famotidine arm, F. Dornase arm.
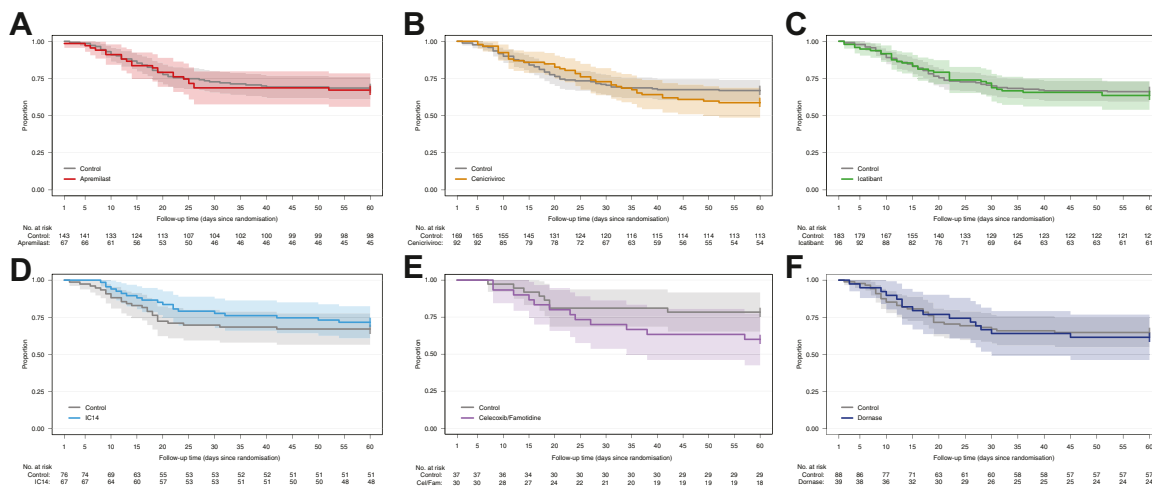
Fig. 3: Survival Over Time, Modified Intention-To-Treat Population. Kaplan–Meier curves for overall survival in the mITT population according to treatment arm. The x-axis shows time in days. The numbers below each panel show patients at risk over time. Shaded areas show 95% pointwise confidence intervals. A. Apremilast arm, B. Cenicriviroc arm, C. Icatibant arm, D. IC14 arm, E. Celecoxib/Famotidine arm, F. Dornase arm.

As most participants were not mechanically ventilated at enrollment, progression to mechanical ventilation or death was pre-specified as an important secondary outcome and is presented in Fig. 4. No differences were observed for this endpoint for any agent, compared to concurrent controls (Supplemental Table S9).

No statistically significant differences in proportions of systematically captured clinically important events were detected in any arms compared to concurrent controls (p > 0.05 for all comparisons) (Supplemental Fig. S8), except that the celecoxib/famotidine arm had a lower proportion of participants requiring renal replacement therapy compared to concurrent controls

(p = 0.04, Fisher's Exact). Three adverse events were adjudicated by the Safety Working Group as probably related to cenicriviroc. One participant reported nausea (1/92, 1%), resulting in discontinuation of study drug, and two participants experienced grade 1 and grade 2 elevations in liver transaminases (2/92, 2%) (Supplemental Table S10); no other adverse events were deemed probably or definitely related to study drug for the other study arms by the Safety Working Group.

Discussion

The first seven agents to be fully studied in I-SPY COVID did not provide a large notable benefit in time-to-recovery or mortality for patients hospitalised with severe COVID-19. All six agents met the futility stopping criterion, and one combination agent (celecoxib/famotidine) was stopped early due to potential harm. Similarly, no improvement in the secondary endpoint of progression to mechanical ventilation or death was identified for patients who were not mechanically ventilated at enrollment (>80% of patients in all arms).

When the I-SPY COVID trial was designed early in the pandemic, high volumes of patients with critical illness strained hospital resources, and knowledge about the pathogenesis of SARS-CoV-2-related lung injury was limited. The innovative approach for this trial design was to rapidly screen agents to identify marked benefit on time-to-recovery, prioritising high agent throughput versus the risk of discarding treatments with potentially smaller benefits. While none of these first agents displayed large efficacy signals, and therefore none graduated to further testing in a Phase 3 trial, these results do not exclude the possibility of these agents having smaller beneficial effects, i.e., a Type II error is possible.

Agent	Recovered	Died
Apremilast (n = 67)	34 (51%)	22 (33%)
Apremilast Controls (n = 143)	80 (56%)	45 (31%)
Cenicriviroc (n = 92)	52 (57%)	38 (41%)
Cenicriviroc controls (n = 169)	95 (56%)	56 (33%)
Icatibant (n = 96)	49 (51%)	35 (36%)
Icatibant controls (n = 183)	102 (56%)	62 (34%)
IC14 (n = 67)	36 (54%)	19 (28%)
IC14 controls (n = 76)	48 (63%)	25 (33%)
Celecoxib/famotidine (n = 30)	15 (50%)	12 (40%)
Celecoxib/famotidine controls (n = 37)	29 (78%)	8 (22%)
Dornase (n = 39)	22 (56%)	15 (38%)
Dornase controls (n = 88)	55 (62%)	31 (35%)

Note that controls may appear in more than one row if they were concurrent to multiple agents.

Table 3: Observed recovery and all-cause mortality proportions at 60 days, by agent and concurrent controls, modified intention-to-treat population.

Outcome	Arm	Median posterior hazard ratio (95% CrI)
Recovery (HR > 1 is better)	Apremilast	0.78 (0.52, 1.14)
	Cenicriviroc	0.88 (0.63, 1.23)
	Icatibant	0.85 (0.60, 1.17)
	IC14	0.63 (0.40, 0.96)
	Celecoxib/famotidine	0.50 (0.28, 0.90)
	Dornase alfa	0.76 (0.48, 1.19)
Death (HR < 1 is better)	Apremilast	1.05 (0.66, 1.71)
	Cenicriviroc	1.24 (0.84, 1.83)
	Icatibant	1.06 (0.71, 1.59)
	IC14	0.86 (0.48, 1.51)
	Celecoxib/famotidine	1.67 (0.79, 3.58)
	Dornase alfa	1.09 (0.61, 1.92)

The posterior distributions of the cause-specific hazard ratios for recovery were computed using Bayesian proportional-hazard Weibull regressions. The cause-specific hazards were modeled as a function of study arm, adjusting for baseline COVID severity. Similarly, Bayesian proportional-hazard Weibull regressions were used to model the hazard functions for all-cause mortality. In these analyses, follow-up times were censored at 60 days if patients were still at risk for either outcome. Weakly informative priors were used for all models' parameters; see SAP for details. The table shows the medians of the posterior hazard ratio distributions for investigational agents compared with concurrently randomised controls, and 95% quantile-based credible intervals.

Table 4: Median posterior hazard ratio by agent, modified intention-to-treat.

a pulmonary-specific outcome (since COVID-19 was thought to primarily cause respiratory failure at the time) relevant for hospitals during surge conditions. In addition, this endpoint afforded greater statistical power compared to mortality, particularly in a population with varying levels of respiratory support at enrollment. However, there are limitations to this endpoint. It is unclear if a low level of oxygen use for 2 consecutive days accurately reflects clinical recovery, and the association of this endpoint with longer-term patient-centred outcomes remains to be defined. Lastly, mortality and recovery should both be measured as key or primary endpoints, as other COVID-19 studies have found that mortality and recovery endpoints may be disparate.[12] A theoretical drug that reduces mortality in a severely ill cohort may actually prolong recovery. In this trial our cohort had prolonged recovery—we found that ~12% of enrolled patients were alive but not recovered by day 60.

One strength of this trial design is that control groups for each agent were composed of patients enrolled contemporaneously, who would have been eligible for treatment with that agent. This approach has not been uniformly employed in platform trials but accounts for changing therapies and outcomes over time, introduction of vaccines, and the changing pattern of SARS-CoV-2 variants that are cardinal features of this pandemic. Other strengths of the trial design include: i) engagement at a wide range of sites including both academic and community hospitals, which has enabled enrollment of a diverse population; ii) rapid data collection to enable timely analysis of patient outcomes, essential for platform designs; and iii) the platform design, which enables efficient testing of multiple agents using the same infrastructure.

For example, although it met criteria for futility for time to recovery, IC14's final posterior HR for death [0.86 (95% CrI 0.48–1.51)] (HR < 1 favors decreased mortality) and posterior probability of graduation (0.72) indicate that a smaller beneficial effect remains possible.

Endpoint selection has been a persistent challenge for Phase 2 studies in critical illness with a lack of validated, proximate, patient-centred surrogate endpoints. Time-to-recovery, defined as requiring <6 L/min oxygen for ≥2 days, was selected as the primary endpoint for I-SPY COVID, with the goal of focusing on

Fig. 4: Progression to Mechanical Ventilation or Death in Patients Not Mechanically Ventilated at Baseline, By Agent, mITT Population. Kaplan–Meier curves for progression to mechanical ventilation or death in patients not mechanically ventilated at baseline in the mITT population according to treatment arm. The x-axis shows time in days. The numbers below each panel show patients at risk over time. Shaded areas show 95% pointwise confidence intervals.

Among the study limitations is the potential for Type 1 or Type 2 error that comes with the strategic choice to rapidly screen many agents for large signals of benefit with strong DMC oversight to stop agents at any time when potential harm signals emerge. It is not possible to have both small sample size and great precision. The open label design that facilitates rapid cycling of agents and enables a mix of different study drug routes (e.g., IV, SQ, inhaled) may introduce bias. Additionally, the inability to perform dose finding studies due to the urgency of the pandemic may have led to ineffective dosing. While objective endpoints (e.g., mortality and time-to-recovery) and a large pool of clinical sites that each enroll relatively few patients per arm mitigates some challenges of an open label approach, investigator bias cannot be completely removed. To compensate for this limitation, registration-focused blinded Phase 3 trials of graduating agents is the intended next step for any graduating agent. Based on the level of respiratory support required—ranging from 6 L/min nasal cannula (<10% of patients) to invasive mechanical ventilation, with most patients on high flow oxygen—lung injury severity for enrolled patients was probably heterogenous. However, the inclusion criteria as designed allowed the trial to consistently enroll patients with early lung injury even as treatment practices around endotracheal intubation shifted over the course of the pandemic.[13] The process of randomisation followed by consent may have led to bias and imbalances in arm accrual; however, the ITT analyses, which included all randomised patients, had similar outcomes to the mITT analyses. Finally, the biological heterogeneity within severe COVID-19 has been more substantial than might have been initially anticipated for lung injury emerging from a single etiologic agent[14] and that heterogeneity has not yet been incorporated into these analyses. To address this issue, biospecimens have been collected from participants, and analyses of these specimens are ongoing. We feel that the results of this trial for some agents should not limit future consideration of testing in larger trials of patients with COVID or non-COVID severe acute respiratory failure.

As an ongoing platform trial, I-SPY COVID continues to evaluate additional agents and is evolving with the experience of the investigative team. We now collect data on vaccination (including type and number of doses) and the use of novel therapies not widely used or not available at the time the trial began (e.g., monoclonal antibodies, additional COVID-directed immunosuppressive therapies, and other anti-viral agents), along with longer-term patient-reported outcomes post-discharge, among others. We have reduced the number of agents being studied concurrently to two (in addition to the control arm) to facilitate informed consent *prior* to randomisation, and we are reevaluating the target sample size to increase the statistical power of the trial. Our ongoing efforts to refine the trial design are being pursued with hopes that similar approaches may be useful for non-COVID-ARDS, sepsis, acute hypoxemic respiratory failure, or other forms of critical illness. More detailed reports including biological results will be submitted in the near future for each of the agents discussed in this report.

In conclusion, the I-SPY COVID trial rapidly evaluated potential therapeutic agents for COVID-19 using an adaptive platform design in a diverse set of clinical centres. Although none of these agents produced large signals for efficacy, the trial demonstrated that it is feasible to evaluate multiple agents in parallel arms in search of large treatment effects in a time of great clinical need. Adaptive platform trials can provide a flexible infrastructure for enabling more rapid integration of emerging knowledge during an ongoing pandemic and continuously improving trial design. The experience with this trial and other innovative trial designs during the COVID-19 pandemic can help accelerate discovery in critical care and ultimately bring effective therapies to patients with severe illness.

References

1 Group RC, Horby P, Lim WS, et al. Dexamethasone in hospitalized patients with Covid-19. *N Engl J Med*. 2021;384(8):693–704.
2 Angus DC, Derde L, Al-Beidh F, et al. 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*. 2020;324(13):1317–1329.
3 Woodcock J, Araojo R, Thompson T, Puckrein GA. Integrating research into community practice - toward increased diversity in clinical trials. *N Engl J Med*. 2021;385(15):1351–1353.
4 Griffiths G, Fitzgerald R, Jaki T, et al. AGILE-ACCORD: a randomized, multicentre, seamless, adaptive phase I/II platform study to determine the optimal dose, safety and efficacy of multiple candidate agents for the treatment of COVID-19: a structured summary of a study protocol for a randomised platform trial. *Trials*. 2020;21(1):544.
5 Calfee C. Clinical trial design during and beyond the pandemic: the I-SPY COVID trial. *Nat Med*. 2022;28(1):9–11.
6 Files DC, Matthay MA, Calfee CS, et al. I-SPY COVID adaptive platform trial for COVID-19 acute respiratory failure: rationale, design and operations. *BMJ Open*. 2022;12(6):e060664.
7 Beigel JH, Tomashek KM, Dodd LE, et al. Remdesivir for the treatment of Covid-19 - final report. *N Engl J Med*. 2020;383(19):1813–1826.
8 Wang Y, Zhang D, Du G, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multi-centre trial. *Lancet*. 2020;395(10236):1569–1578.
9 Woodcock J, LaVange LM. Master protocols to study multiple therapies, multiple diseases, or both. *N Engl J Med*. 2017;377(1):62–70.
10 Organization WH. *WHO R&D blueprint: COVID-19 therapeutic trial synopsis*. 2020.
11 Douin DJ, Siegel L, Grandits G, et al. Evaluating primary endpoints for COVID-19 therapeutic trials to assess recovery. *Am J Respir Crit Care Med*. 2022;206(6):730–739.
12 ACTIV-3–Therapeutics for Inpatients with COVID-19 (TICO) Study Group. Tixagevimab-cilgavimab for treatment of patients hospitalised with COVID-19: a randomised, double-blind, phase 3 trial. *Lancet Respir Med*. 2022;10(10):972–984.
13 Matthay MA, Thompson BT, Ware LB. The Berlin definition of acute respiratory distress syndrome: should patients receiving high-flow nasal oxygen be included? *Lancet Respir Med*. 2021;9(8):933–936.
14 Sinha P, Furfaro D, Cummings MJ, et al. Latent class analysis reveals COVID-19-related acute respiratory distress syndrome subgroups with differential responses to corticosteroids. *Am J Respir Crit Care Med*. 2021;204(11):1274–1285.