
Peer reviewed version

Link to published version (if available):
10.1001/jama.2023.3651

Link to publication record in Explore Bristol Research
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via American Medical Association at https://jamanetwork.com/journals/jama/fullarticle/2802856. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/
Heterogeneous Treatment Effects of Therapeutic-dose Heparin in Patients Hospitalized for COVID-19: An Exploratory Analysis of a Multiplatform Randomized Clinical Trial

Ewan C. Goligher MD PhD*1,2,3,4
Patrick R. Lawler MD MPH*1,2,3,5
Thomas P. Jensen MS6
Victor Talisa PhD7
Lindsay R. Berry PhD6
Elizabeth Lorenzi PhD6
Bryan J. McVerry MD8
Chung-Chou Ho Chang PhD9
Eric Leifer PhD10
Charlotte Bradbury MD PhD11
Jeffrey Berger MD12
Beverly Hunt MD PhD13
Lana A. Castellucci MD14,15
Lucy Z. Kornblith MD16,17
Anthony Gordon MD18
Colin McArthur MD19,20,21
Steven Webb MD19,22
Judith Hochman MD12
Matthew Neal MD8,23
Ryan Zarychanski MD MSc**24,25
Scott Berry PhD**6
Derek C. Angus MD MPH**7,8

For the REMAP-CAP, ATTACC, and ACTIV-4a investigators

*Equally contributing first authors
**Equally contributing senior authors

Institutional Affiliations

1. Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, Canada
2. Department of Medicine, Division of Respirology, University Health Network, Toronto, Canada
3. Toronto General Hospital Research Institute, Toronto, Canada
4. Department of Physiology, University of Toronto, Toronto, Canada
5. Peter Munk Cardiac Centre, University Health Network, Toronto, Canada
6. Berry Consultants, Austin, TX
7. Department of Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA
8. University of Pittsburgh Medical Centre, Pittsburgh, PA
9. Department of Medicine, University of Pittsburgh, Pittsburgh, PA
10. National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD
11. University of Bristol and University Hospitals Bristol and Weston NHS Foundation Trust, Bristol
12. New York University (NYU) Grossman School of Medicine, New York, NY
13. Kings Healthcare Partners, London, United Kingdom
14. Ottawa Hospital Research Institute, Ottawa, Canada
15. University of Ottawa, Ottawa, Canada
16. Zuckerberg San Francisco General Hospital, San Francisco, CA
17. University of California San Francisco, San Francisco, CA
18. Imperial College Healthcare NHS Trust, St. Mary’s Hospital, London, United Kingdom
19. Australian and New Zealand Intensive Care Research Centre, Monash University, Melbourne, Australia
20. Auckland City Hospital, Auckland, New Zealand
21. Medical Research Institute of New Zealand, Wellington, New Zealand
22. St. John of God Subiaco Hospital, Subiaco, WA
23. Department of Surgery, University of Pittsburgh, Pittsburgh, PA
24. CancerCare Manitoba, Winnipeg, Manitoba, Canada
25. University of Manitoba, Winnipeg, Manitoba, Canada

Address for correspondence:
Dr Ewan Goligher
Toronto General Hospital
585 University Ave., 9-MaRS-9024
Toronto, ON, Canada
M5G 2N2
Email: ewan.goligher@utoronto.ca
Phone: +1 416 340-4800 ext. 6810

Word Count:
3,274

Registration:
Clinicaltrials.gov registration numbers NCT02735707, NCT04505774, NCT04359277, NCT04372589

Funding for original trial:
REMAP-CAP
Supported by the European Union — through FP7-HEALTH-2013-INNOVATION: the Platform for European Preparedness Against (Re-)emerging Epidemics (PREPARE) consortium (602525), and Horizon 2020 research and innovation program: the Rapid European Covid-19 Emergency Research response (RECOVER) consortium (101003589) — and by the Australian National Health and Medical Research Council (APP1101719), the Health Research Council of New Zealand (16/631), the Canadian Institutes of Health Research (Strategy for Patient-Oriented Research Innovative Clinical Trials Program Grant - 158584, and COVID-19 Rapid Research Operating Grant - 447335), the U.K. NIHR and the NIHR Imperial Biomedical Research Centre,
the Health Research Board of Ireland (CTN 2014-012), the UPMC Learning While Doing Program, the Translational Breast Cancer Research Consortium, the French Ministry of Health (PHRC-20-0147), the Minderoo Foundation, Amgen, Eisai, the Global Coalition for Adaptive Research, and the Wellcome Trust Innovations Project (215522). Anthony Gordon is funded by an NIHR Research Professorship (RP-2015-06-18), Manu Shankar-Hari by an NIHR Clinician Scientist Fellowship (CS-2016-16-011), Alexis Turgeon is funded by a Canada Research Chair – Tier 2. Ryan Zarychanski is the recipient of the Lyonel G Israels Research Chair in Hematology (University of Manitoba).

**ATTACC**
The ATTACC platform was supported by grants from the Canadian Institutes of Health Research, LifeArc, Thistledown Foundation, Research Manitoba, Cancercare Manitoba Foundation, Victoria General Hospital Foundation, Ontario Ministry of Health, and the Peter Munk Cardiac Centre.

**ACTIV-4a**
The ACTIV-4a platform was sponsored by the National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda and administered through OTA-20-011. The research was, in part, funded by the National Institutes of Health (NIH) Agreement 1OT2HL156812-01.
Key Points

Question:
What patient characteristics determine differences in the effect of therapeutic-dose heparin among patients hospitalized for moderate or severe COVID-19?

Findings:
Heterogeneity in treatment effect in the multi-platform randomized clinical trial of therapeutic-dose heparin for moderate or severe COVID-19 was evaluated using three different approaches: conventional subgroup analyses, risk-based analysis, and effect-based analysis. All three approaches found that heparin was more likely to be beneficial in those who were less severely ill at presentation or had lower BMI, and more likely to be harmful in sicker patients and those with higher BMI.

Meaning:
Analyzing heterogeneity of treatment effect may help to individualize treatment decision-making with heparin in patients hospitalized for COVID-19.
Abstract

Importance
Randomized trials of therapeutic-dose heparin in patients hospitalized with COVID-19 produced conflicting results, possibly due to heterogeneity of treatment effect (HTE) across individuals. Better understanding of HTE would facilitate individualized clinical decision-making.

Objective
To evaluate HTE of therapeutic-dose heparin for patients hospitalized for COVID-19.

Design, setting, and participants
Exploratory analysis of the multiplatform adaptive randomized trial of therapeutic-dose heparin vs usual care pharmacological thromboprophylaxis in 3,320 patients hospitalized for COVID-19 enrolled in North America, South America, Europe, Asia, and Australia between April 2020 and January 2021. HTE was assessed by conventional subgroup analyses of baseline characteristics; a multivariable outcome-prediction model (risk-based approach); and, a multivariable causal forest model (effect-based approach). Analyses primarily used Bayesian statistics, consistent with the original trial.

Exposures
Participants were randomized to therapeutic-dose heparin or usual care pharmacologic thromboprophylaxis.

Outcomes
Organ support-free days, assigning zero days to those who died in hospital and the number of days free of cardiovascular or respiratory organ support up to day 21 for those who survived to hospital discharge (where an odds ratio [OR]>1 = benefit and OR<1 = harm); and hospital survival.

Results
Age (median 60 years), sex (38% female), race (32% known non-white), and ethnicity (45% Hispanic) were similar between patients randomized to therapeutic-dose heparin or usual care. In conventional subgroup analyses, the effect of therapeutic-dose heparin on organ support-free days differed between patients requiring organ support at baseline or not (median ORs = 0.85 vs. 1.30; posterior probability of difference in OR, 99.8%), between females and males (median ORs = 0.87 vs. 1.16; posterior probability of difference in OR, 96.4%), and between patients with lower body mass index (BMI) <30 kg/m² compared to higher BMI groups (BMI ≥30 kg/m², posterior probability of difference in odds ratios >90% for all comparisons). In the risk-based analysis, patients at lowest risk of poor outcome had the highest propensity for benefit from heparin (lowest risk decile: posterior probability of odds ratio>1, 92%) while those at highest risk were most likely to be harmed (highest risk decile: posterior probability of OR<1, 87%). The effect-based analysis identified a subset of patients at high risk of harm (p=0.05 for difference in treatment effect) who tended to have high BMI and were more likely to require organ support at baseline.
Conclusions and Relevance
Among patients hospitalized for COVID-19, the effect of therapeutic-dose heparin was heterogeneous. Across all 3 approaches to assessing HTE, heparin was more likely to be beneficial in those who were less severely ill at presentation or had lower BMI, and more likely to be harmful in sicker patients and those with higher BMI.
Introduction

Thrombosis and inflammation contribute to critical illness or death in patients hospitalized for COVID-19. Multiple randomized clinical trials (RCTs) evaluated the benefit of therapeutic-dose heparin in these patients, with varying results. The multi-platform randomized clinical trial (mpRCT) and Therapeutic Anticoagulation versus Standard Care as a Rapid Response to the COVID-19 Pandemic (RAPID) RCT both reported clinical benefit from therapeutic-dose heparin in non-critically ill hospitalized COVID-19 patients; the mpRCT and other trials observed no benefit and probable harm from therapeutic or intermediate dose heparin in critically ill COVID-19 patients. Such divergent findings point to differences in heparin’s effect according to patient characteristics, also known as heterogeneity of treatment effect (HTE).

Estimates of between-group differences in RCT outcomes like those summarized above represent average treatment effects (ATE) across all study participants. Ideally, clinicians need estimates of treatment effect for individual patients (the individual treatment effect, ITE), but that would require knowledge of counterfactuals that cannot be observed (the patient’s outcome under treatment and without treatment). The ITE may vary widely within a population of patients because of differences among patients in relevant characteristics such as baseline health status, severity of infection, susceptibility to harm from anticoagulation, and multiple other factors. Although ITE is not estimable, treatment effects can be estimated for patient subgroups defined by a shared set of relevant characteristics (the conditional average treatment effect, cATE). Subgroup analysis is a conventional approach to estimating cATE but has important limitations, so methods have been developed for more valid estimation of treatment effect in subgroups. These include risk- and effect-based analysis of HTE,
data-driven modelling of CATEs using machine learning techniques.16 These newer methods improve upon conventional subgroup analysis8 but may yield dissimilar or even conflicting results, which can limit their utility in clinical decision-making.15,17

In this post-hoc exploratory study we applied these analytical approaches to pooled mpRCT data to empirically test the ability of these methods to identify heterogeneity in heparin’s treatment effects in patients with COVID-19, and to test for consistency between the results of these approaches as to which patients hospitalized for COVID-19 might benefit from the intervention.

Methods

Study Design and Inclusion Criteria

The mpRCT was a collaboration between the Antithrombotic Therapy to Ameliorate Complications of COVID-19 (ATTACC),18 Accelerating COVID-19 Therapeutic Interventions and Vaccines-4 Antithrombotics Inpatient platform trial (ACTIV-4a), and Randomized, Embedded, Multifactorial Adaptive Platform Trial for Community-Acquired Pneumonia (REMAP-CAP) trial platforms.19 Eligibility criteria, interventions, outcome measures, and data collection were prospectively aligned across the three platforms to evaluate the effect of pragmatically defined regimens of therapeutic-dose heparin administered prophylactically or usual care pharmacologic thromboprophylaxis on mortality and organ support-free days to day 21 (see Table E1 for details). Ethics and regulatory approval were obtained at each participating center. Race was ascertained by patient self-report based on selection from fixed categories with the option of “other”.

Commented [EG1]: For the sake of keeping the introduction short and focused, we deleted the following and also emphasized the point further in the Discussion:

“While mpRCT data are largely from 2020, we undertook the effort on the premise that the findings are still relevant to contemporary management of hospitalized COVID-19 patients in the absence of data demonstrating that mechanisms of disease driving outcomes in this population are fundamentally different in the current stage of the pandemic.”
The trial prospectively stratified the primary analysis based on illness severity, evaluating distinct treatment effects in hospitalized patients with laboratory-confirmed COVID-19 with severe illness (critically ill or requiring ICU-level care) or moderate illness (hospitalized non-critically ill patients not requiring ICU-level care). Patients with moderate illness were further analyzed according to baseline D-dimer level. ICU-level care was defined as the use of respiratory or cardiovascular organ support (oxygen delivered by high-flow nasal cannula, non-invasive or invasive ventilation, or the use of vasopressors or inotropes). The mpRCT employed Bayesian hierarchical modelling to separately estimate the posterior probability of treatment benefit in the primary endpoint in patients with severe COVID-19, moderate COVID-19 with elevated baseline D-dimer (more than two times the upper limit of normal), moderate COVID-19 without elevated baseline D-dimer, and moderate COVID-19 with missing baseline D-dimer. Adaptive analyses were planned to determine the final sample size based on accumulating information about the probability of superiority or futility. The mpRCT completed enrolment in severe COVID-19 patients in 7 of 2020, based on a statistical conclusion of futility, and subsequently completed enrolment in moderate COVID-19 patients in January of 2021, based on a statistical conclusion of superiority across all D-dimer groups. These mpRCT results were described in two separate reports \(^2\) \(^-\) \(^4\) including one-at-a-time subgroup analyses exploring HTE within moderate and severe COVID-19 separately.\(^2\) \(^-\) \(^4\) For the present study, we present analyses conducted across the entire population in the mpRCT with laboratory-confirmed COVID-19, including both moderately and severely ill patients.

\textit{Outcomes}
The primary outcome for the mpRCT was organ support–free days, evaluated on an ordinal scale that combined in-hospital death (assigned a value of −1) and the number of days free of cardiovascular or respiratory organ support up to day 21 among patients who survived to hospital discharge. Patients who never required organ support were assigned a value of 22. Patients who were discharged from the hospital before day 21 were assumed to be alive and free of organ support from discharge through day 21. Any death during the index hospitalization through 90 days was assigned the worst score on the outcome scale (−1). This end point reflects both survival to hospital discharge and the need for and duration of ICU-level interventions, with higher values indicating better outcomes.

A secondary outcome for this study was survival to hospital discharge, a key component of the primary outcome.

**Statistical Analysis**

HTE was evaluated using three strategies planned after publication of the original mpRCT results (Figure 1).

**Conventional Subgroup Analysis Approach**

Differences in treatment effect were explored within prespecified clinical subgroups defined by variables that were deemed to potentially moderate treatment effect (age, sex, race, ethnicity, body mass index (BMI), illness severity (moderate vs. severe), baseline respiratory support requirements, baseline d-dimer, pre-randomization thromboprophylaxis dosing, time from hospital admission), using conventional subgroup analysis. Here, the influence of each
potential effect modifier on treatment effect is evaluated one-at-a-time in separate models, in isolation from all other potential effect modifiers. Race and ethnicity were collected as these characteristics may influence both the risk of outcome and the magnitude of treatment effect. The treatment effect in each subgroup was estimated using a Bayesian cumulative logistic regression model with a statistical interaction (product term) with treatment effect to calculate the proportional odds ratio (OR) for organ support-free days, where an OR >1 indicated treatment benefit and an OR <1 indicated treatment harm.

Each subgroup analysis model was adjusted for illness severity, age, sex, baseline D-dimer, trial site/country, and enrolment time period (in 2-week intervals). Weakly informative Dirichlet prior probability distributions were specified for the baseline probabilities for each value of organ support-free days. Independent standard normal priors were specified on the treatment effects in each subgroup analysis. The model was fit using a Markov chain Monte Carlo algorithm with 100,000 samples from the joint posterior distribution, which allowed calculation of the posterior distributions for the odds ratios, including medians and 95% credible intervals (CrI), and the posterior probability of superiority of therapeutic-dose heparin (indicated by an OR >1). The magnitude of statistical evidence for differences in treatment effect was quantified in each subgroup analysis by computing the posterior probability that the OR was greater in one subgroup vs the others. Analyses were repeated for the secondary outcome of survival to discharge.

Risk-Based Approach
Differences in treatment effect were explored across levels of risk of poor outcome (effect modification by predicted risk) using previously described methods.\textsuperscript{13,14} First, risk was estimated for each patient, as follows. A cumulative logistic regression risk prediction model for organ support-free days was internally derived from the population using a pre-specified set of candidate risk predictors; final variables for inclusion were selected based on the model with lowest Bayesian information criterion among models with all possible combinations of the following: age, sex, race, BMI, cardiac history, diabetes, chronic kidney disease, chronic respiratory disease, chronic liver disease, immunosuppression, baseline vasopressor requirement, baseline respiratory support, D-dimer, neutrophil count, lymphocyte count, creatinine, and platelet count. Statistical interactions (product terms) between candidate risk predictors were not tested because it was not computationally possible to test all possible combinations. Missing data were multiply imputed using a multivariate imputation by chained equations (MICE) algorithm. The model was used to compute risk for each patient in terms of a risk score, computed as the linear combination of their covariate values and the final risk model coefficients (log odds ratio), multiplied by –1 to obtain a positive value. Patients were ranked by risk score and grouped by deciles of risk.

Then, risk score was evaluated as a potential effect modifier by computing cATE within each decile of risk, specifying a statistical interaction (product term) between risk decile and treatment assignment in a Bayesian cumulative logistic regression model (constructed as described above for the conventional subgroup analysis models). Smoothing of the treatment effects across quantiles of risk score was induced with a first-order normal dynamic linear
model prior distribution (dynamic borrowing). Additional analyses were performed with independent priors to estimate the treatment effects by decile without smoothing (borrowing).

**Effect-Based Approach**

Differences in treatment effect were evaluated across levels of predicted treatment effect (effect modification by predicted effect) using a non-parametric causal forest method (Figure E1). Because causal forest methods are most easily applied to binary outcomes, this analysis employed survival status at hospital discharge (alive or dead) as the outcome of interest. The treatment effect was quantified in terms of the absolute difference in survival rate. In this application, the causal forest method estimates cATE for each patient as a conditional absolute rate difference (cARD), the difference between weighted survival outcome averages among treated and control patients with similar values for potential effect modifiers (the more similar the values, the higher the weights).

Before applying the causal forest method and to reduce computational complexity for identifying effect modifiers in the face of a modest sample size, a prognostic model of the probability of survival at hospital discharge was computed by random forest modelling using the same set of candidate risk predictors used to construct the risk-based model. The causal forest model of treatment effect on hospital survival was constructed using the predicted probability of survival for each patient along with all candidate effect modifiers. The resulting model estimated cARD for each patient. Positive cARD values indicate a predicted improvement in survival due to therapeutic-dose heparin, while negative values indicate a predicted decrease in survival.
A specialized Monte-Carlo cross-validation procedure with 100 repetitions was used to estimate observed absolute difference in rate of hospital survival within deciles of predicted cARD, infer the 95% confidence limits, and test relevant null hypotheses. Specifically, we tested whether the observed absolute differences in the rate of survival increased monotonically across deciles of the cARD predicted for each patient; and whether the observed difference in rate of survival in the lowest decile of predicted cARD was significantly different from the rest (this latter test was defined post hoc).

Contributions of each candidate effect modifier to cARD prediction were quantified by Shapley Additive Explanations (SHAP) scores, which allocate credit for each cARD prediction among the variables. Specifically, a SHAP captures the difference in cARD attributable to the difference in candidate effect modifier value between patients. For example, a BMI SHAP of -0.06 for a one unit increase in BMI is associated with an absolute 6 percent decrease in cARD. See Supplement eMethods for more details. To further characterize patients in the lowest decile and compare them to others, we compared observed baseline characteristics between groups.

All statistical analyses are detailed in the Supplement eMethods and were performed using R v 4.1.3 and STAN v 2.21.0.

Results

The original primary mpRCT analyses included 1,098 patients hospitalized with severe COVID-19 and 2,219 with moderate COVID-19. Since publication, the primary endpoint was ascertained for 4 additional patients with severe COVID-19 and 1 patient with moderate COVID-19.
19 was removed from the analysis set as they were double-counted in two platforms. Thus, 3,320 patients were included for analysis in the present study. Baseline characteristics for the population are shown in Table 1, with missing data summarized in Table E2. In the overall population (n=3,320), median organ support-free days was 22 (interquartile range, 8 to 22); 17.3% (n=575) of patients died in hospital. In patients with moderate COVID-19 (n=2,218), median organ support-free days was 22 (interquartile range, 22 to 22) and hospital mortality was 7.8% (n=173). In patients with severe COVID-19 (n=1,102), median organ support-free days was 3 (interquartile range, −1 to 16) and hospital mortality was 36.5% (n=402).

In the overall mpRCT population, therapeutic-dose heparin was not associated with an increase in organ support-free days (median value for the posterior distribution of the OR: 1.05, 95% CrI 0.91, 1.22).

**Conventional subgroup analysis**

The effect of therapeutic-dose heparin on organ support-free days differed substantially between moderate and severe COVID-19 groups (median OR: 1.29 vs. 0.85; posterior probability of difference in OR, 99.8%), which the original trials defined as not requiring (moderate) vs requiring (severe) critical-care level organ support at baseline; between males and females (median OR: 1.16 vs. 0.86; posterior probability of difference in OR, 96.4%); and between the subgroup with BMI <30 kg/m² in comparison to BMI subgroups with BMI ≥30 kg/m² (posterior probability of difference in OR, >90% for all comparisons) [Figure 2, Panel A]. The remaining subgroup analyses are presented in Figure 2A and Table E3. Body mass index (BMI) was distributed similarly between patients with moderate COVID-19 (median 30.0,
interquartile range 26.5-34.8) and severe COVID-19 (median 30.4, interquartile range 26.6-35.4).

Similarly, the effect of therapeutic-dose heparin on hospital survival differed substantially between moderate and severe COVID-19 groups (median OR 1.22 vs. 0.82, posterior probability of difference in OR, 96.3%), between patients with BMI <30 kg/m² and those with higher BMI (posterior probability of difference in OR, >90% for all comparisons), and according to baseline respiratory support (posterior probability of difference in OR, >90% for all comparisons) (Figure 2, Panel A). Treatment effect on hospital survival was not meaningfully different between females and males (median OR, 0.85 vs. 1.02, posterior probability of difference in OR, 79%).

Risk-based Approach

The estimated odds ratios (and log odds ratios) for each variable included in the final derived model of cumulative odds of organ support-free days are shown in Table 2. The risk of having a low value for organ support-free days was distributed bimodally across the mpRCT population (Figure E2) and within each of the participating platforms (Figure E3). Patients were ranked by predicted risk and grouped by decile (Table E4). Increasing risk score was associated with progressively lower values for organ support-free days: in the lowest decile the risk of death or organ dysfunction was very low while the proportion of patients with death or prolonged organ failure was highest in the highest decile, with a consistent gradient across deciles (Figure E4).

The effect of therapeutic-dose heparin on organ support-free days and survival to hospital discharge for each risk score decile is shown in Figure 2B. The observed treatment effect in the 60% of the cohort at lowest risk (groups 1-6) (risk score ≤ 3.46) suggested probable benefit
from therapeutic dose-heparin (posterior probability of odds ratio >1, >80%). The observed treatment effect in the 30% at highest risk (groups 8-10) (risk score >5.14) suggested that harm from therapeutic-dose heparin was more probable than not (posterior probability of odds ratio >1, <50%). A similar pattern of results was obtained from estimates of cATE obtained without dynamic borrowing (Figure E5) and for survival to hospital discharge (Figure 2, Panel B). The implications of risk-based HTE in terms of absolute risks for poor clinical outcomes are shown for representative patients with varying risk profiles in Table E5: although patients at lowest risk exhibited the highest relative benefit (in terms of odds ratio, Figure 2, Panel B), the absolute risk difference is relatively constant for representative patients with low or moderate risk scores because treatment effect varies inversely with risk. In a representative patient with a high risk score, treatment increases the absolute risk of a poor outcome.

The level of respiratory support at baseline was a major determinant of risk group assignment. All 996 patients in risk groups 8-10 (Figure 2) were receiving either high flow nasal oxygen, non-invasive ventilation, or invasive mechanical ventilation at baseline (Figure E6), whereas very few patients in risk groups 1-6 were receiving these respiratory supports at baseline (2/1992, 0.1%). Most patients receiving high flow nasal oxygen were in risk groups 8-10 (292/425, 69%).

**Effect-based Approach**

The distribution of observed absolute differences in hospital survival obtained by cross-validation in each decile of predicted cARD is shown in Figure 2C. Observed hospital survival did not monotonically increase with deciles of predicted benefit (p=0.38). The point estimate for
absolute risk reduction in the lowest decile suggested a non-statistically significant association between therapeutic-dose heparin and possible harm (ARR –5.7%, 95% CI –17%, 6%), and the effect of therapeutic-dose heparin on mortality differed between patients in the lowest decile in comparison to all others (post hoc p-value for statistical interaction = 0.05). SHAP scores indicated that lower predicted risk of death and lower BMI exhibited the strongest associations with greater cARD (Figure E7). Patients in the lowest cARD decile group (in whom the treatment was associated with possible harm, Figure 2C) tended to have high BMI and were more likely to require ICU admission at baseline (Table E6).

Discussion

Using three analytic strategies, this exploratory analysis of the multiplatform trial of therapeutic-dose heparin for patients hospitalized for COVID-19 found that the treatment effect varied substantially according to the baseline risk of poor outcome, primarily reflected by the degree of organ support required at baseline. The threshold level of risk that defined whether heparin was more likely to be beneficial or harmful appears to coincide with the transition from moderate COVID-19 (no organ support at baseline) to severe COVID-19 (requirement for respiratory or cardiovascular organ support at baseline). Higher BMI was also associated with a higher risk of harm with therapeutic-dose heparin. Importantly, baseline severity of illness and BMI were associated with differential treatment effect irrespective of the analytical approach used to characterize HTE. This consistency is noteworthy in view of the previously reported differences in results obtained by different HTE modelling strategies17 and supports the relevance of these characteristics to inform decisions about treatment with
therapeutic-dose heparin in patients hospitalized for COVID-19, despite the exploratory nature of the analysis.

The multi-platform trial was designed to estimate treatment effect according to severity of illness and, among moderately ill patients, baseline D-dimer level. This design decision was motivated in part by anticipation of potential HTE based on these factors, and partly by the fact that therapeutic-dose heparin was conceptualized as a “treatment” for severe COVID-19 in patients with severe illness and as a means to prevent progression to severe illness in patients with moderate COVID-19. Nevertheless, the trial was predicated on anticipated potential benefit in both moderately and severely ill patients; if anything, prior observational evidence was suggestive of greater potential risk of thrombosis in critically ill patients, suggesting greater potential benefit. In retrospect, this trial design decision was fortuitous as the present analysis suggests that the distinction between moderate and severe COVID-19 appears to optimally differentiate patients who derive benefit or harm with therapeutic-dose heparin. If the trial had been primarily designed to enroll a broad cohort of all hospitalized patients, regardless of severity of illness, it would likely have reported a single average treatment effect indicative of neither benefit nor harm.

To appraise the credibility of observed HTE, it is important to consider both the consistency of the result across multiple analytical techniques and to establish whether there is a plausible mechanistic basis for HTE. HTE according to baseline severity of illness was observed across the three strategies deployed to evaluate HTE in this study. This suggests that therapeutic-dose heparin should be avoided in severely ill patients requiring organ support, and that its benefit is limited to hospitalized patients who do not require organ support at baseline. Of note, this
analysis cannot determine whether heparin should be continued or discontinued when patients transition from moderate to severe illness. Elevated BMI was also identified as an important predictor of harm from therapeutic-dose heparin across all three HTE strategies. Because BMI was unrelated to severity state, clinicians may consider shifting the risk-benefit decision-making for patients with body mass index > 30 kg/m^2 in patients with moderate COVID-19. Finally, subgroup analyses suggested that female sex was associated with a low probability of benefit from therapeutic-dose heparin and male sex was associated with a high probability of benefit. Sex was not an influential contributor to HTE in the risk-based and effect-based analyses.

The mechanisms accounting for HTE according to baseline risk and severity of illness, BMI, and sex are uncertain. It is possible that the putative effect of heparin on inflammation and thrombosis may only be relevant before organ injury and dysfunction are established. Obesity is generally associated with an increased risk of thrombosis and vascular inflammation, particularly in combination with COVID-19, so one might expect therapeutic-dose heparin to be beneficial rather than harmful in this population. However, dosing heparin to target therapeutic levels may be challenging in patients with high BMI. Observational data suggest that among patients hospitalized for COVID-19, males are at higher risk of thrombosis and death in comparison to females and the excess mortality in males was attributable to higher thrombotic risk. This observation might account for the observed differences in treatment effect by sex, although the underlying sex-specific mechanism remains uncertain.

Given trends in epidemiology, risk, treatment, and outcomes of COVID-19, the generalizability of mpRCT results generated during the first year of the pandemic to the contemporary management of COVID-19 warrants consideration. While rising population
immunity due to widespread vaccination and infection and the circulation of less virulent variants of SARS-CoV-2 have markedly reduced the risk of developing severe disease, the mechanisms (such as exaggerated host thrombotic and inflammatory responses) responsible for organ dysfunction and death among those patients who do develop severe disease are likely unchanged. The putative benefits of heparin therefore likely remain relevant for some patients, though the continued evolution of SaRS-CoV-2 bears consideration when generating inferences from trials conducted early in the pandemic. The findings of the present analysis highlight the importance of assessing risk in determining whether therapeutic-dose heparin should be prophylactically administered in patients with moderate COVID-19 without established thrombosis, and the risk score and effect-based predictors provide a basis for conducting this risk assessment.

The results of the multiplatform trial illustrate the importance of considering HTE in the design of trials and the utility of adaptive trial design methods to prospectively account for HTE. Under traditional approaches to trial design, the primary “cost” of designing a trial to estimate separate treatment effects in independent strata within the trial population is a marked increase in sample size requirement, an increase that may ultimately be unnecessary if the treatment effect is actually similar across strata. Such designs may be perceived as a “risk” in terms of increased costs and time to reach trial conclusions. Adaptive trial designs utilizing Bayesian hierarchical statistical models can mitigate this risk by dynamically borrowing information about treatment effect between subgroups. This entails that there is no major adverse impact on sample size requirement unless there is substantial HTE, in which case divergent treatments effect can be identified and reported, as in the multiplatform trial.
Because the sources and determinants of HTE are often unknown at the outset of clinical trials, it is challenging to prospectively account for HTE in trial design. Future innovations in adaptive trial design could utilize early trial phases to use data-driven techniques for risk-based and effect-based modelling on initial trial data to detect strong HTE signals. These signals could drive adaptations in the trial design to account for HTE discovered during the trial.

In conclusion, the effect of therapeutic-dose heparin administered prophylactically in patients hospitalized for COVID-19 appears to vary substantially according to the baseline risk of having a poor outcome, the severity of COVID-19 at baseline, and body mass index. Patients with moderate COVID-19 who require only low-flow supplemental oxygen (or none) are more likely to benefit from treatment; patients with severe COVID-19 who require organ support or ICU-level care do not benefit from therapeutic-dose heparin.
References


<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Therapeutic-dose Heparin (n=1707)</th>
<th>Usual care pharmacologic thromboprophylaxis (n=1613)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years) (median [IQR])</td>
<td>60 (50 - 69)</td>
<td>60 (51 - 69)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1094 (64%)</td>
<td>980 (61%)</td>
</tr>
<tr>
<td>Female</td>
<td>613 (36%)</td>
<td>633 (39%)</td>
</tr>
<tr>
<td>Racea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aboriginal</td>
<td>111 (7%)</td>
<td>80 (5%)</td>
</tr>
<tr>
<td>Asian</td>
<td>107 (6%)</td>
<td>108 (7%)</td>
</tr>
<tr>
<td>Black</td>
<td>224 (13%)</td>
<td>165 (10%)</td>
</tr>
<tr>
<td>Other</td>
<td>127 (8%)</td>
<td>145 (9%)</td>
</tr>
<tr>
<td>White</td>
<td>928 (54%)</td>
<td>892 (55%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>210 (12%)</td>
<td>223 (14%)</td>
</tr>
<tr>
<td>Hispanic ethnicity (n/N, %)</td>
<td>479/1073 (45%)</td>
<td>438/963 (45%)</td>
</tr>
<tr>
<td><strong>Baseline severity state</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate COVID-19</td>
<td>1171 (69%)</td>
<td>1047 (65%)</td>
</tr>
<tr>
<td>Severe COVID-19</td>
<td>536 (31%)</td>
<td>566 (35%)</td>
</tr>
<tr>
<td><strong>Region of enrolment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>873 (51%)</td>
<td>801 (50%)</td>
</tr>
<tr>
<td>Europe</td>
<td>585 (34%)</td>
<td>595 (37%)</td>
</tr>
<tr>
<td>South America</td>
<td>245 (14%)</td>
<td>213 (13%)</td>
</tr>
<tr>
<td>Australia/Asia</td>
<td>4 (0%)</td>
<td>4 (0%)</td>
</tr>
<tr>
<td><strong>Comorbidities</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body mass index (kg/m²) (median [IQR])</td>
<td>30.0 (26.5 - 35.2)</td>
<td>30.2 (26.6 - 34.9)</td>
</tr>
<tr>
<td>Cardiovascular diseasec</td>
<td>658 (19.8%)</td>
<td>581 (17.5%)</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>521 (15.7%)</td>
<td>501 (15.1%)</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>139 (4.2%)</td>
<td>112 (3.4%)</td>
</tr>
<tr>
<td>Chronic respiratory diseasef</td>
<td>359 (10.8%)</td>
<td>330 (9.9%)</td>
</tr>
<tr>
<td>Chronic liver disease</td>
<td>20 (0.6%)</td>
<td>14 (0.4%)</td>
</tr>
<tr>
<td>Immuno-suppressive diseaset</td>
<td>118 (3.6%)</td>
<td>119 (3.6%)</td>
</tr>
<tr>
<td><strong>Condition at enrolment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaspressors at baseline</td>
<td>95 (2.9%)</td>
<td>109 (3.3%)</td>
</tr>
<tr>
<td>Respiratory support at baseline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None SUPPLEMENTARY</td>
<td>1130 (66%)</td>
<td>998 (62%)</td>
</tr>
<tr>
<td>High flow nasal oxygen</td>
<td>202 (12%)</td>
<td>223 (14%)</td>
</tr>
<tr>
<td>Non-invasive ventilation</td>
<td>237 (14%)</td>
<td>225 (14%)</td>
</tr>
<tr>
<td>Invasive mechanical ventilation</td>
<td>138 (8%)</td>
<td>167 (10%)</td>
</tr>
<tr>
<td><strong>Pre-randomization thromboprophylaxis dosing</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low dose</td>
<td>522 (31%)</td>
<td>479 (30%)</td>
</tr>
<tr>
<td>Intermediate dose</td>
<td>138 (8%)</td>
<td>156 (10%)</td>
</tr>
<tr>
<td>Subtherapeutic dose</td>
<td>18 (1%)</td>
<td>8 (0%)</td>
</tr>
<tr>
<td>Therapeutic dose</td>
<td>51 (3%)</td>
<td>13 (1%)</td>
</tr>
<tr>
<td>Unknown dose</td>
<td>978 (57%)</td>
<td>957 (59%)</td>
</tr>
<tr>
<td>Laboratory values (median [IQR])</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>D-dimer (ratio, upper limit of normal at site)</td>
<td>1.67 (1.00 – 2.82)</td>
<td>1.63 (1.00 – 2.94)</td>
</tr>
<tr>
<td>Neutrophil count (10^9/L)</td>
<td>6.06 (4.00 – 8.90)</td>
<td>6.10 (4.05 – 9.00)</td>
</tr>
<tr>
<td>Lymphocyte count (10^9/L)</td>
<td>0.81 (0.60 – 1.20)</td>
<td>0.85 (0.60 – 1.25)</td>
</tr>
<tr>
<td>Creatinine (mg/dL)</td>
<td>0.89 (0.72 – 1.13)</td>
<td>0.86 (0.70 – 1.09)</td>
</tr>
<tr>
<td>Platelet count (10^9/L)</td>
<td>229 (176 – 300)</td>
<td>227 (175 – 298)</td>
</tr>
</tbody>
</table>

Values are n (%) unless otherwise noted

1 Participants in the multi-platform randomized clinical trial (mpRCT) of therapeutic-dose heparin for moderate and severe COVID-19 enrolled between April 2020 and January 2021, for whom primary outcome is known
2 Race was ascertained by patient self-report from a selection of fixed categories. When multiple races were selected, race is reported as “other”.
3 Severe COVID-19 defined as need ICU-level of care with organ support at baseline, including high flow nasal oxygen > 20 litres per minute, non-invasive ventilation, invasive mechanical ventilation, or vasopressors. Moderate COVID-19 was defined as hospitalization for COVID-19 without need for ICU-level of care with organ support.
4 Cardiovascular disease defined as baseline history of heart failure, myocardial infarction, coronary artery disease, peripheral arterial disease, or cerebrovascular disease (stroke or transient ischemic attack) in the ATTACC (Antithrombotic Therapy to Ameliorate Complications of COVID-19) and ACTIV-4a (A Multicenter, Adaptive, Randomized Controlled Platform Trial of the Safety and Efficacy of Antithrombotic Strategies in Hospitalized Adults with COVID-19) platforms and as a baseline history of New York Heart Association class IV symptoms in the REMAPCAP platform (Randomized, Embedded, Multifactorial Adaptive Platform Trial for Community-Acquired Pneumonia).
5 Chronic respiratory disease defined as a baseline history of asthma, chronic obstructive pulmonary disease, bronchiectasis, interstitial lung disease, primary lung cancer, pulmonary hypertension, active tuberculosis, or the receipt of home oxygen therapy.
6 Immunosuppressive disease or therapy defined as concurrently having any of the following conditions: HIV, leukemia, metastatic cancer, myeloma, lupus, multiple sclerosis, Rheumatoid arthritis, Psoriatic arthritis, Crohn’s disease, granulomatosis with polyangiitis, sarcoidosis, monoclonal gammopathy of unknown significance, ankylosing spondylitis, psoriasis, receiving chemotherapy or radiation, transplant recipient, receiving high-dose or long-term steroid treatment.
Table 2. Risk model for organ support-free days estimated in the full population

<table>
<thead>
<tr>
<th>Model variable</th>
<th>Risk model coefficients</th>
<th>Log odds ratio $^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (per 10 years)</td>
<td>0.73 (0.68, 0.77)</td>
<td>-0.320</td>
</tr>
<tr>
<td>Sex (female)</td>
<td>1.29 (1.10, 1.51)</td>
<td>0.253</td>
</tr>
<tr>
<td>Body mass index (per 5 kg/m$^2$)</td>
<td>0.92 (0.88, 0.97)</td>
<td>-0.078</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>0.78 (0.67, 0.91)</td>
<td>-0.251</td>
</tr>
<tr>
<td>Immuno-suppressive disease or therapy $^d$</td>
<td>0.58 (0.44, 0.77)</td>
<td>-0.539</td>
</tr>
<tr>
<td><strong>Baseline organ support</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vasopressor requirement</td>
<td>0.49 (0.35, 0.70)</td>
<td>-0.707</td>
</tr>
<tr>
<td>High flow nasal oxygen</td>
<td>0.07 (0.05, 0.08)</td>
<td>-2.702</td>
</tr>
<tr>
<td>Non-invasive ventilation</td>
<td>0.05 (0.04, 0.06)</td>
<td>-3.104</td>
</tr>
<tr>
<td>Invasive mechanical ventilation</td>
<td>0.04 (0.03, 0.05)</td>
<td>-3.289</td>
</tr>
<tr>
<td>Neutrophils (per $5 \times 10^9$/L)</td>
<td>0.74 (0.68, 0.81)</td>
<td>-0.300</td>
</tr>
<tr>
<td>Lymphocytes (per $5 \times 10^9$/L)</td>
<td>1.35 (1.10, 1.66)</td>
<td>0.300</td>
</tr>
<tr>
<td>Platelets (per $25 \times 10^9$/L)</td>
<td>1.07 (1.04, 1.09)</td>
<td>0.064</td>
</tr>
</tbody>
</table>

$^a$Unadjusted for treatment assignment and no statistical interactions between model variables were specified in the risk model.

$^b$Confidence intervals are reported because the internally derived risk scores were derived from a frequentist ordinal logistic model.

$^c$The log odds ratio is used to compute the risk score components.

$^d$Immunosuppressive disease or therapy defined as concurrently having any of the following conditions: HIV, leukemia, metastatic cancer, myeloma, lupus, multiple sclerosis, Rheumatoid arthritis, Psoriatic arthritis, Crohn’s disease, Wegener’s granulomatosis, sarcoidosis, monoclonal gammopathy of unknown significance, ankylosing spondylitis, psoriasis, receiving chemotherapy or radiation, transplant recipient, receiving high-dose or long-term steroid treatment.
Figure Legends

Figure 1. Three strategies to evaluate heterogeneity of treatment effect (HTE).

Caption:

Strategy 1: Conventional strategy tests for HTE using subgroup analyses. Differences in treatment effect among subgroups are evaluated in a regression model with an independent variable representing one of several potential effect modifiers (patient, disease, or management characteristics) that might modify treatment effect. A separate model is computed for each potential effect modifier.

Strategy 2: Risk-based strategy tests for HTE according to risk of the outcome estimated using a risk model derived and internally validated in the trial data. In this approach the patient and disease characteristics are handled as predictors of outcome (candidate risk predictors). These risk predictors are combined in a single risk model to compute a single candidate effect modifier, the predicted risk of outcome. See Text and eMethods for details.

Strategy 3: Effect-based strategy tests for HTE according to predicted treatment effect computed from a model combining multiple variables potentially associated with treatment effect (trained on part of the data, the training dataset) and comparing predicted vs observed treatment effect on the remaining data (test dataset). In this approach, patient and disease characteristics are used to compute a model of the difference in outcome with and without treatment; this model is used to compute a single candidate effect modifier, the predicted treatment effect. See Text and eMethods for details.
OSFDs = organ support-free days; cATE = conditional average treatment effect (the treatment effect within a subgroup); cARD = conditional absolute rate difference in hospital survival; ARD = observed absolute rate difference in hospital survival.

Figure 2. Heterogeneity of treatment effect (HTE).
Panel 1 and 2 dots represent median odds ratio value, whiskers represent 95% credible intervals. Panel 3 dots represent median value for the observed absolute risk reduction (ARR) in each repetition, whiskers represent 95% confidence intervals. See text and eMethods for full details.

TH = therapeutic-dose heparin; PTP = usual care pharmacological thromboprophylaxis.

Panel A. Conventional subgroup analyses.
Panel B. Risk-based HTE for organ support-free days shown by risk deciles (ranging from lowest, group 1, to highest, group 10). All patients in risk groups 8-10 required respiratory organ support at baseline vs 2/1992 (0.1%) of patients in risk groups 1-6 (see Table E2). Clinical benefit was deemed substantially more probable than not (posterior probability of odds ratio >1 above 80%) in risk groups 1-6. A similar pattern was observed for hospital survival, although the posterior probability of benefit from heparin was lower for hospital survival vs organ support-free days.

Panel C. Effect-based HTE for hospital survival shown by deciles of predicted conditional absolute rate difference (cARD) in hospital survival derived from repeated cross-validation using a causal machine learning algorithm (n=100 repetitions). See text and eMethods for full details.