

BMJ Open Contribution of the elevated thrombosis risk of males to the excess male mortality observed in COVID-19: an observational study

Kenneth Roy Cohen , David Anderson, Sheng Ren, David J Cook

To cite: Cohen KR, Anderson D, Ren S, *et al.* Contribution of the elevated thrombosis risk of males to the excess male mortality observed in COVID-19: an observational study. *BMJ Open* 2022;**12**:e051624. doi:10.1136/bmjopen-2021-051624

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-051624>).

Received 26 March 2021
Accepted 07 February 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

OptumLabs, United Health Group, Minnetonka, Minnesota, USA

Correspondence to

Dr Kenneth Roy Cohen;
ken.cohen@optum.com

ABSTRACT

Background The mortality rate of COVID-19 is elevated in males compared with females.

Objective Determine the extent that the elevated thrombotic risk in males relative to females contributes to excess COVID-19 mortality in males.

Design Observational study.

Setting Data sourced from electronic medical records from over 200 US hospital systems.

Participants 60 877 patients aged 18 years and older hospitalised with COVID-19.

Exposure Exposure variable: biological sex; key variable of interest: thrombosis.

Primary outcome measures Primary outcome was COVID-19 mortality. We measured: (1) mortality rate of males relative to females, (2) rate of thrombotic diagnoses occurring during hospitalisation for COVID-19 in both sexes and (3) mortality rate when evidence of thrombosis was present.

Results The COVID-19 mortality rate of males was 29.9% higher than that of females. Males had a 35.8% higher rate of receiving a thrombotic diagnosis compared with females. The mortality rate of all patients with a thrombotic diagnosis was 40.0%—over twice that of patients with COVID-19 without a thrombotic diagnosis (adjusted OR 2.50 (2.37 to 2.64), $p < 0.001$). When defining thrombosis as either a documented thrombotic diagnosis or a D-dimer level ≥ 3.0 $\mu\text{g/mL}$, 16.4% of the excess mortality in male patients could be explained by increased thrombotic risk.

Conclusions Our findings suggest the higher COVID-19 mortality rate in males may be significantly accounted for by the elevated risk of thrombosis among males. Understanding the mechanisms that underlie increased male thrombotic risk may allow for the advancement of effective anticoagulation strategies that reduce COVID-19 mortality in males.

INTRODUCTION

Male sex has emerged as a risk factor for increased COVID-19 morbidity and mortality, with the risk of mortality in males being 30%–40% higher than in females.^{1 2} This heightened mortality risk for males is independent of advancing age, and cannot be fully explained by higher rates among males of other comorbidities known to increase

Strengths and limitations of this study

- Data derived from 60 877 hospitalised patients with COVID-19 comprised of blended claims and Electronic Health Record (EHR) data.
- Natural language processing (NLP) engine allowing for detailed interrogation of EHR data.
- Independent manual verification of NLP engine accuracy.
- Data elements could be missed if not expressed in a standard fashion that can be recognised by the NLP engine.
- True magnitude of underlying thrombosis in this population could be underestimated by both claims analysis and NLP interrogation of EHR data.

mortality in COVID-19, such as obesity, diabetes mellitus, hypertension or underlying cardiopulmonary disease.^{3 4} Therefore, the aetiology of this sex difference in COVID-19 mortality is largely unexplained.

Thrombosis appears to play an important role in the morbidity and mortality of COVID-19, with patients at increased risk of both microvascular and macrovascular thrombosis. In a meta-analysis of 42 studies representing data from 8271 patients with COVID-19, the authors documented an overall venous thromboembolism rate of 21% and a pulmonary embolism (PE) rate of 13%. For patients admitted to the Intensive Care Unit (ICU), the rate of thromboembolism was even higher, at 31%. Notably, the pooled odds of mortality were 74% higher among patients with COVID-19 who developed thromboembolism, compared with those who did not.⁵

Additionally, thromboses may go undiagnosed during a hospital stay for COVID-19. Markedly elevated D-dimer levels are highly correlated with underlying thrombosis, confirmed in two recent studies. In the first study, a level of 3.0 $\mu\text{g/mL}$ had a sensitivity of 70.0%, specificity of 96.7% with a positive

predictive value of 87.5% for underlying thrombosis.⁶ In the second study, a level >2.5 $\mu\text{g/mL}$ had a sensitivity of 63% and a specificity of 85% for the diagnosis of PE.⁷

The elevated thrombotic risk of males relative to females was well established in the medical literature long before the emergence of COVID-19. In multiple studies of unprovoked deep vein thrombosis (DVT) and PE where hereditary thrombophilia was excluded, the rate of recurrent thrombosis in males has been reported to be anywhere from two to more than three times higher than females.^{8–12} Despite extensive research on the topic, the aetiology of this sex-based difference in the thrombotic risk of males remains unknown.¹³

We therefore understand that males are at risk of higher COVID-19 mortality, that COVID-19 mortality risk is increased when patients have underlying thrombosis, and that unrelated to COVID-19, males are more prone to thrombosis relative to females. It is therefore plausible that the elevated thrombotic risk of males relative to females may partially explain the excess mortality rate observed in males with COVID-19. In this study, we sought to understand whether males with COVID-19 have a higher rate of thrombosis compared with females with COVID-19, and whether a higher rate of thrombosis contributes to the excess mortality seen in males.

METHODS

Data source

To explore a potential connection between thrombosis, sex and mortality risk, we analysed hospitalisation data from over 200 geographically dispersed hospital systems. The data for this study were sourced from electronic medical records (EMR) data and post-EMR coding. These EMRs were processed with a natural language processing (NLP) engine that produces a homogenised set of codified and non-codified information, including lab results, medications, symptoms and various observational extracts of text from the EMR. The at-scale extraction of this detailed information allows insight into the clinical manifestations of the COVID-19 population. We examined the subsets of patients who survived COVID-19 compared with those who expired. For each of the groups, we compared the male and female incidence of receiving a thrombosis diagnosis code while hospitalised for COVID-19.

Inclusion criteria

From these data, we selected patients aged 18 years or older who were hospitalised for COVID-19 between 3 March 2020 and 11 June 2021. There were no exclusions for other comorbidities or underlying diseases. We limited our analysis to those patients who survived and were discharged from the hospital, those discharged to hospice care, those who died in the hospital and those who died after being discharged to home. Patients who died soon after discharge to home or who were discharged to hospice care were treated as deceased cases due to

COVID-19. COVID-19 readmission cases were excluded from our study.

Outcomes and study variables

The primary outcome of interest was mortality from COVID-19. The exposure variable was biological sex, and the key variable of interest was thrombosis. We classified thrombotic diagnosis codes into four major conditions: (1) myocardial infarction (MI), (2) DVT/PE, (3) stroke and (4) peripheral arterial occlusion. A full list of ICD-10 codes used to identify thrombosis can be found in online supplemental table 1.

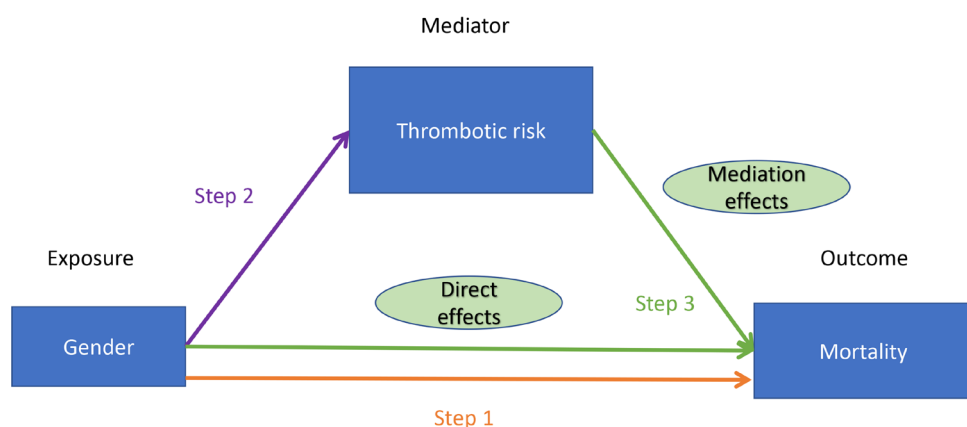
Because thrombotic diagnoses are under-reported in inpatients with COVID-19,¹⁴ we also examined the peak D-dimer value of the subset of hospitalised patients in whom this was measured. A D-dimer level ≥ 3.0 $\mu\text{g/mL}$ was used as a surrogate for underlying thrombosis when a thrombotic diagnosis code was absent.

Statistical methods

To determine whether the elevated thrombotic risk in males is associated with higher COVID-19 mortality, we used mediation analysis,¹⁵ where elevated thrombotic risk was the mediator between the outcome (mortality) and the exposure variable (gender). Mediation analysis consists of three steps, which are illustrated in figure 1. First, we needed to confirm that male patients had higher mortality rate than female patients in our study population. Second, we needed to confirm that males had a higher incidence of a thrombosis diagnosis relative to females. Third, we needed to demonstrate that both males and females with thrombosis diagnoses had a higher mortality rate than those who did not. If all three of these conditions are supported by statistically significant differences between groups, we can combine regression models from the first and third hypotheses to estimate the proportion of sex effect on mortality that can be explained by the elevated thrombotic risk. Besides estimating this proportion, we also examined the effect of a thrombosis diagnosis together with evidence of a D-dimer level ≥ 3.0 $\mu\text{g/mL}$, recognising that such levels are also likely indicative of underlying thrombosis.

Logistic regression models were used in each of the three steps. The response variables and covariates of interest are illustrated in figure 1. We used the Wald test for regression coefficients in logistic regression models to assess statistical significance in all steps. All hypothesis tests were one-sided at the 0.05 significance level. We report 95% CIs and one-tailed p values. We choose the more powerful one-sided test because all three associations described in figure 1/table 1 (male and higher risk of mortality, male and higher risk of thrombosis, and thrombosis and higher risk of mortality) are well documented in literature as we discussed in the Introduction section, and we only wanted to test if these hypotheses were true with respect to directions supported by literature. All regression models were adjusted for age (binary variable of age >65 years or not) and comorbidities

Mediation Analysis Flowchart



Each arrow (step) is a regression model. It starts from a covariate of interest and points to the response variable.

Figure 1 Mediation analysis flow chart.

Table 1 Full mediation analysis results of thrombosis diagnosis

Mediation analysis steps

Step 1:		Female	Male
Documentation of the higher mortality rate in males	Number of patients	29 680	31 197
	Survived	24 438	24 009
	Deceased	5242	7188
	Sex distribution among deceased patients	42.2%	57.8%
	Rate of deceased patients by sex	17.7%	23.0%
	Unadjusted OR	1.40 (1.34 to 1.45) p<0.001	
	Adjusted OR	1.41 (1.35 to 1.48) p<0.001	
Step 2:		Female	Male
Documentation of the higher rate of thrombosis in males	Number of patients	29 680	31 197
	No thrombosis diagnosis	25 625	25 407
	Had thrombosis diagnosis	4055	5790
	Sex distribution among patients with thrombosis	41.2%	58.8%
	Thrombosis prevalence by sex	13.7%	18.6%
	Unadjusted OR	1.44 (1.38 to 1.50) p<0.001	
	Adjusted OR	1.34 (1.28 to 1.40) p<0.001	
Step 3:		No thrombosis	Had thrombosis
Documentation of the higher mortality rate associated with a thrombosis diagnosis	Number of patients	51 032	9845
	Survived	42 537	5910
	Deceased	8495	3935
	Thrombosis distribution among deceased patients	68.3%	31.7%
	Mortality rates by thrombosis	16.7%	40.0%
	Unadjusted OR	3.33 (3.18 to 3.49) p<0.001	
	Adjusted OR	2.50 (2.37 to 2.64) p<0.001	
	Adjusted OR for sex	1.38 (1.31 to 1.44) p<0.001	

derived from the Charlson Index.¹⁶ These comorbidities can be found in online supplemental table 2. Note that two comorbidities—MI and cerebrovascular diseases—are not adjusted in our regression models since they overlap with our thrombosis definitions shown in online supplemental table 1.

The proportion of excess male mortality explained by the elevated thrombotic risk is defined by the ratio of the mediation effect to the total effect (mediation+direct effect) of gender on mortality. This proportion was estimated using the methods explained in this paper,¹⁷ which was implemented by an R package ‘mediation’.¹⁸ Our R codes for mediation analysis using this package can be found in online supplemental materials III. Because thrombotic diagnoses are under-reported in inpatients as discussed above, we estimated the proportion based on the presence of a thrombosis diagnosis only as well as a definition of elevated thrombotic risk that included D-dimer values >3.0 $\mu\text{g/mL}$ ¹⁹ or presence of a thrombotic diagnosis. We also reported the proportion explained by thrombosis diagnoses only on the subset whose D-dimer was measured to make sure the proportion explained by elevated thrombotic risk was not subject to selection bias. The upper bound of the D-dimer normal range (0.5 $\mu\text{g/mL}$) was used to normalise D-dimer values.

Patient and public involvement

No patients or members of the public were involved in the design, conduct, reporting or dissemination plans of our research.

RESULTS

Our study population was comprised of 60 877 patients with COVID-19 with an average age of 62 years. There were slightly more male patients (51.3%) than female patients. Among study patients, 16.2% had a reported thrombotic diagnosis during their hospital stay and 20.4% died or were discharged to hospice care. We also identified a subset of 31 817 patients (52.3%) out of the study population (with or without a thrombotic diagnosis) who had D-dimer values measured during their hospital stay. Normalised mean peak D-dimer during a hospital stay was 4.35 (435% of the normal range upper bound). Characteristics of the full study population and this patient subset are shown in table 2. Percentages of comorbid conditions stratified by gender and thrombosis diagnosis are shown in online supplemental tables 2 and 3 for these two populations.

We found that the mortality rate in males with COVID-19 was higher by 29.9% compared with females with COVID-19 (an absolute rate difference of 5.3%; adjusted OR=1.41 (1.35 to 1.48), $p<0.001$). Compared with females with COVID-19, males with COVID-19 had a rate of receiving a thrombotic diagnosis during their hospital stay that was 35.8% higher (an absolute difference of 4.9%, OR=1.34 (1.28 to 1.40), $p<0.001$), confirming the higher rate of thrombotic diagnoses in

Table 2 Characteristics of the full hospitalised population and the D-dimer subset

Study sample characteristics	Summary statistics*	
	Full population dataset	D-dimer analysis subset
Sample size	60 877	31 817
Deceased	12 430 (20.42)	6633 (20.85)
Thrombophilia	9845 (16.17)	5160 (16.22)
Thrombosis history	4085 (6.71)	2147 (6.75)
Mechanical ventilation	6803 (11.17)	3941 (12.39)
Deceased or intubated	13 960 (22.93)	7478 (23.50)
MI	3625 (5.96)	1732 (5.44)
DVT/PE	5150 (8.46)	3000 (9.43)
Stroke	1891 (3.11)	854 (2.68)
PAO	290 (0.48)	100 (0.31)
Age average (SD)	61.70 (18.20)	62.84 (16.54)
Sex: male	31 197 (51.25)	16 969 (53.33)
D-dimer average	—	435.39%
D-dimer median	—	184.40%

*With the exception of D-dimer statistics, sample size and age, data are otherwise reported as number (percentage) of patients within each dataset. D-dimer statistics are based on the upper bound of the normal range for each measure. DVT, deep vein thrombosis; MI, myocardial infarction; PAO, peripheral arterial occlusion; PE, pulmonary embolism.

males. Additionally, we found an over twofold difference in mortality between patients with and without a thrombotic diagnosis (40.0% vs 16.7%; adjusted OR=2.50 (2.37 to 2.64), $p<0.001$). Having verified that thrombosis is more prevalent in males (OR=1.34) and that thrombosis is a strong risk factor for mortality (OR=2.50), we conclude that the elevated thrombotic risk accounts for a portion of the excess mortality in males. All results of this three-step mediation analysis are shown in table 1.

Based on these findings, we then sought to determine what proportion of the excess male mortality in patients with COVID-19 might be related to the elevated thrombotic risk. When using only a documented diagnosis of thrombosis, the proportion of the mortality effect explained by thrombosis was significant at 12.1% ($p<0.001$), with a 95% CI (9.44% to 15.3%) (table 3). Because of under-reporting of thrombotic diagnoses in hospitalised patients with COVID-19, we also considered a D-dimer level ≥ 3.0 $\mu\text{g/mL}$ as a surrogate of underlying thrombosis. We therefore combined the proportion of increased mortality in males as predicted by a thrombotic diagnosis with the proportion explained by a D-dimer level >3.0 $\mu\text{g/mL}$. In doing so, the total proportion of excess male mortality potentially explained by the elevated thrombosis risk in males increased to 16.4% with a 95% CI (11.4% to 22.6%) when the thrombosis definition was expanded to include D-dimer levels >3.0 $\mu\text{g/mL}$ (table 3). Note that in this subset of patients who

Table 3 Proportion of excess mortality explained by elevated thrombotic risk

Mediators	Proportion mediated (explained)
Thrombosis diagnosis on whole dataset	12.1% (9.44% to 15.3%) p<0.001
Thrombosis diagnosis on D-dimer subset	10.6% (7.01% to 15.2%) p<0.001
Thrombosis diagnosis codes or D-dimer >3.0 µg/mL	16.4% (11.4% to 22.6%) p<0.001

had D-dimer measured (52.3% of the study cohort), the proportion of COVID-19 cases with a documented thrombotic diagnosis alone was 10.6% (table 3), which was very close to the 12.1% of cases with a documented thrombotic diagnosis in the study cohort. This indicates that an additional 5.8% of cases can be explained by elevated D-dimer levels, and that there was almost no bias in the selection of the subcohort that had D-dimer measured.

DISCUSSION

The data extracted from our large population of hospitalised patients with COVID-19 confirm a higher mortality rate among those patients with thrombosis. It also demonstrates that the incidence of thrombosis is higher in males relative to females and may explain up to 16.4% of the excess mortality seen in males with COVID-19.

Because the known genetic risks for thrombophilia (factor V Leiden, prothrombin G20210A, protein C and S abnormalities, etc) are not sex-linked chromosomal mutations, they cannot explain this excess male risk. There are, however, genetic mutations found on the X and Y chromosomes which might explain a portion of this increased thrombophilia in males.¹² Additionally, severe COVID-19 infection is itself associated with a prothrombotic state. Although the prothrombotic mechanisms are not fully understood, several have been postulated. Activation of the coagulation cascade through direct viral invasion of the vascular endothelium may cause severe endothelial injury with disruption of fibrinolytic activity and release of von-Willebrand factor.²⁰ Thrombosis may be triggered by the generalised cytokine activation seen in severe COVID-19 with resultant activation of platelets and the complement system.²¹ Lastly, COVID-19 has been associated with the development of procoagulant autoantibodies, including antiphospholipid antibodies. Diabetes has also been shown to independently increase the risk of thrombosis in hospitalised patients with COVID-19.²²

To our knowledge, this is the first study to specifically address the significantly elevated risk of thrombosis and its associated excess mortality in males with COVID-19 relative to females. Although mortality as a function of sex was not specifically addressed in the meta-analysis of the 42 studies noted in the introduction,⁵ that study also documented that thrombosis contributes to excess

mortality in COVID-19. Moreover, among those 42 studies, 29 documented the sex of the patients. Of those 29 studies, 27 (93%) documented a higher percentage of thromboembolism in males relative to females. Of all cases of thromboembolism across those 29 studies, 70% occurred in males and 30% occurred in females.

Studies of thromboprophylaxis among hospitalised patients with COVID-19 have documented improved outcomes. A systematic review and pooled analysis of 35 studies looked at anticoagulation strategies in 4685 hospitalised patients with COVID-19.²³ This review suggested that standard prophylactic doses of anticoagulation were associated with significant reductions in venous thromboembolism and arterial thrombosis events, with intermediate and therapeutic doses of anticoagulation providing no additional benefit. A more recent Randomized Controlled Trial (RCT) documented improved outcomes with full dose anticoagulation compared with thromboprophylaxis in a non-critically ill population hospitalised with COVID-19. The benefit was most apparent in the subset of patients with the highest D-dimer levels.²⁴ A second RCT found that compared with standard heparin thromboprophylaxis, therapeutic doses of low molecular weight heparin reduced major thromboembolism and death in a population of high-risk hospitalised patients with COVID-19 and very elevated D-dimer levels.²⁵ These studies did not report stratification of the anticoagulation benefit by sex of the patients.

Our study has limitations. NLP extraction of EMR data may miss certain data elements if they are not expressed in a standard fashion that can be recognised by the NLP engine. The accuracy of the output of the NLP engine was confirmed in a subset of patients by manual chart review. Similarly, analysis of claims data might miss patients with underlying thrombotic diagnoses if these diagnoses were assumed to be part of the COVID-19 clinical syndrome and therefore not independently submitted. Lastly, when interpreting the results from the D-dimer analysis, it cannot be stated with certainty that a markedly elevated D-dimer level is related to underlying thrombosis.

In conclusion, our data support the concept that a significant portion of the excess male mortality in COVID-19 is related to the elevated risk of thrombosis in males relative to females. Understanding the mechanisms that drive the elevated thrombotic risk in males, as well as those that drive thrombosis in COVID-19, may allow for the development of more effective anticoagulation strategies that reduce the mortality risk for males diagnosed with COVID-19. To help develop these strategies, reanalysis of outcome by sex in recently published trials and new randomised clinical trials is suggested to examine whether more intense anticoagulation regimens in males hospitalised with COVID-19 will reduce thromboses and improve clinical outcomes including mortality.

Contributors KRC, DA, SR and DJC had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design—all authors. Acquisition, analysis or interpretation

of data—all authors. Drafting of the manuscript—all authors. Critical revision of the manuscript for important intellectual content—all authors and Amy Okaya, OptumLabs. Editing of the manuscript—Amy Okaya. Statistical analysis—SR. Supervision—KRC. Guarantor—KRC.

Funding The study was funded by OptumLabs, the research and development arm of United Health Group, and the authors KRC, DJC, SR and DA are full-time employees of United Health Group.

Disclaimer The authors played an active role in all aspects of study development, including the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval As this study represents an exploratory secondary analysis of existing de-identified patient data, no IRB review and approval were required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The data are not available for public use but, under certain conditions, may be made available to editors and their approved reviewers under a data use agreement to confirm the findings of the current study.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Kenneth Roy Cohen <http://orcid.org/0000-0002-6141-9142>

REFERENCES

- Martins-Filho PR, Tavares CSS, Santos VS. Factors associated with mortality in patients with COVID-19. A quantitative evidence synthesis of clinical and laboratory data. *Eur J Intern Med* 2020;76:97–9.
- The Sex, Gender, and COVID-19 Project [Internet]. Men, sex, gender and COVID-19. Available: <https://globalhealth5050.org/the-sex-sex-and-covid-19-project/men-sex-sex-and-covid-19/> [Accessed 28 December 2020].
- Docherty AB, Harrison EM, Green CA, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* 2020;369:m1985.
- Peckham H, de Grijter NM, Raine C, et al. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ICU admission. *Nat Commun* 2020;11:6317.
- Malas MB, Naazie IN, EClinicalMedicine. Thromboembolism risk of COVID-19 is high and associated with a higher risk of mortality: a systematic review and meta-analysis. Available: <https://pubmed.ncbi.nlm.nih.gov/33251499/> [Accessed Nov 2020].
- Cui S, Chen S, Li X, et al. Prevalence of venous thromboembolism in patients with severe novel coronavirus pneumonia. *J Thromb Haemost* 2020;18:1421–4.
- Logothetis CN, Weppelmann TA, Jordan A, et al. D-Dimer testing for the exclusion of pulmonary embolism among hospitalized patients with COVID-19. *JAMA Netw Open* 2021;4:e2128802.
- Christiansen SC, Lijfering WM, Helmerhorst FM, et al. Sex difference in risk of recurrent venous thrombosis and the risk profile for a second event. *J Thromb Haemost* 2010;8:2159–68.
- Baglin T, Luddington R, Brown K, et al. High risk of recurrent venous thromboembolism in men. *J Thromb Haemost* 2004;2:2152–5.
- Rodger MA, Kahn SR, Wells PS, et al. Identifying unprovoked thromboembolism patients at low risk for recurrence who can discontinue anticoagulant therapy. *CMAJ* 2008;179:417–26.
- McRae S, Tran H, Schulman S, et al. Effect of patient's sex on risk of recurrent venous thromboembolism: a meta-analysis. *Lancet* 2006;368:371–8.
- Douketis J, Tosetto A, Marcucci M, et al. Risk of recurrence after venous thromboembolism in men and women: patient level meta-analysis. *BMJ* 2011;342:d813.
- Roach REJ, Cannegieter SC, Lijfering WM. Differential risks in men and women for first and recurrent venous thrombosis: the role of genes and environment. *J Thromb Haemost* 2014;12:1593–600.
- Piazza G, Morrow DA, Diagnosis MDA. Diagnosis, management, and pathophysiology of arterial and venous thrombosis in COVID-19. *JAMA* 2020;324:2548–9.
- Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986;51:1173–82.
- Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005;43:1130–9.
- Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods* 2010;15:309–34.
- Tingley D, Yamamoto T, Hirose K. Mediation: R package for causal mediation analysis. *J Stat Soft [Internet]* 2014;59:1–38 www.jstatsoft.org/index.php/jss/article/view/v059i05
- Zhang L, Yan X, Fan Q, et al. D-Dimer levels on admission to predict in-hospital mortality in patients with Covid-19. *J Thromb Haemost* 2020;18:1324–9.
- Ackermann M, Verleden SE, Kuehnel M, et al. Pulmonary vascular Endothelialitis, thrombosis, and angiogenesis in Covid-19. *N Engl J Med* 2020;383:120–8.
- Magro C, Mulvey JJ, Berlin D, et al. Complement associated microvascular injury and thrombosis in the pathogenesis of severe COVID-19 infection: a report of five cases. *Transl Res* 2020;220:1–13.
- Calvisi SL, Ramirez GA, Scavini M, et al. Thromboembolism risk among patients with diabetes/stress hyperglycemia and COVID-19. *Metabolism* 2021;123:154845.
- Patell R, Chiasakul T, Bauer E, et al. Pharmacologic thromboprophylaxis and thrombosis in hospitalized patients with COVID-19: a pooled analysis. *Thromb Haemost* 2021;121:076–85.
- Investigators A. ACTIV-4a Investigators, REMAP-CAP Investigators, et al. therapeutic anticoagulation with heparin in Noncritically ill patients with Covid-19. *N Engl J Med* 2021;385:790–802.
- Spyropoulos AC, Goldin M, Giannis D. Efficacy and safety of Therapeutic-Dose heparin vs standard prophylactic or Intermediate-Dose heparins for thromboprophylaxis in high-risk hospitalized patients with COVID-19: the HEP-COVID randomized clinical trial. *JAMA Intern Med. Published online* 2021.

Does the elevated thrombosis risk of males relative to females help account for the excess male mortality observed in Covid-19? An observational study
Supplementary Materials

I. ICD-10 Codes Used to Define Thrombosis Conditions During Hospital Stay

Table e1. All ICD-10 Codes Used for Thrombosis Definition (limited to first 5 digits)

II. Comorbid Condition Summary

Table e2 Prevalence of Comorbid Conditions in the Full Study Population

Table e3. Prevalence of Comorbid Conditions in the Subpopulation with D-dimer and CRP Values

III. R codes for mediation analysis

I. ICD-10 Codes Used to Define Thrombosis Conditions During Hospital Stay**Table e1. All ICD-10 Codes Used for Thrombosis Definition (limited to first 5 digits)**

Thrombosis groups	ICD-10 codes
Myocardial infarction	I2101, I2102, I2109, I2111, I2119, I2121, I2129, I213, I214, I219, I21A1, I21A9, I220, I221, I222, I228, I229
Peripheral arterial occlusions	E0852, E0952, E1052, E1152, E1352, I7401, I7409, I7410, I7411, I7419, I742, I743, I744, I745, I748, I749, I7501, I7502, I7581, I7589, I96
Stroke	I6300, I6301, I6302, I6303, I6309, I6310, I6311, I6312, I6313, I6319, I6320, I6321, I6322, I6323, I6329, I6330, I6331, I6332, I6333, I6334, I6339, I6340, I6341, I6342, I6343, I6344, I6349, I6350, I6351, I6352, I6353, I6354, I6359, I636, I638, I6381, I6389, I639, I601, I6502, I6503, I6509, I651, I6521, I6522, I6523, I6529, I658, I659, I6601, I6602, I6603, I6609, I6611, I6612, I6613, I6619, I6621, I6622, I6623, I6629, I663, I668, I669
Deep venous thrombosis and pulmonary embolus	I2602, I2609, I2692, I2693, I2694, I2699, I8000, I8001, I8002, I8003, I8010, I8011, I8012, I8013, I8020, I8021, I8022, I8023, I8024, I8025, I8029, I803, I808, I809, I81, I820, I821, I8221, I8222, I8229, I823, I8240, I8241, I8242, I8243, I8244, I8245, I8246, I8249, I824Y, I824Z, I825Y, I825Z, I8260, I8261, I8262, I8271, I8281, I8289, I8290, I82A1, I82B1, I82C1

II. Comorbid Condition Summary**Table e2 Prevalence of Comorbid Conditions in the Full Study Population**

Sex	Female	Female	Male	Male
Thrombotic condition	No	Yes	No	Yes
Myocardial infarction	3.67%	39.63%	6.95%	42.69%
Congestive heart failure	14.28%	33.27%	17.04%	31.09%
Peripheral vascular disease	4.44%	10.83%	6.76%	14.82%
Cerebrovascular disease	2.47%	23.65%	2.8%	23.02%
Dementia	8.68%	16.79%	7.43%	10.54%
Chronic pulmonary disease	28.34%	32.03%	22.95%	25.99%
Rheumatic disease	4.64%	6.34%	1.82%	1.64%
Peptic ulcer disease	0.64%	1.55%	0.82%	1.62%
Mild liver disease	5.81%	7.37%	7.29%	8.38%
Diabetes without complication	29.93%	37.09%	34.1%	38.13%
Diabetes with complication	14.62%	25.57%	17.9%	25.92%
Hemiplegia or paraplegia	0.54%	3.7%	0.86%	4.44%
Renal disease	16.93%	30.26%	23.47%	33.45%
Malignancy	5.67%	8.38%	6.82%	9.67%
Moderate severe liver disease	0.71%	1.5%	1.13%	1.88%
Metastatic solid tumor	1.21%	2.07%	1.41%	2.09%
Acquired immunodeficiency syndrome	1%	4.07%	1.28%	4.23%

Table e3. Prevalence of Comorbid Conditions in the Subpopulation with D-dimer and CRP Values

Sex	Female	Female	Male	Male
Thrombotic Condition	No	Yes	No	Yes
Myocardial infarction	3.88%	36%	6.87%	39.61%

Congestive heart failure	14.37%	32.72%	16.29%	29.21%
Peripheral vascular disease	4.56%	9.65%	6.25%	14.2%
Cerebrovascular disease	2.24%	20.83%	2.39%	19.82%
Dementia	8.21%	14.41%	6.62%	9.78%
Chronic pulmonary disease	30.59%	33.29%	23.34%	27.12%
Rheumatic disease	5.3%	6.37%	1.75%	1.54%
Peptic ulcer disease	0.63%	1.38%	0.65%	1.54%
Mild liver disease	6.61%	7.04%	7.11%	8.21%
Diabetes without complication	34.19%	38.71%	36.25%	38.83%
Diabetes with complication	15.28%	25.01%	17.16%	24.3%
Hemiplegia or paraplegia	0.45%	3.09%	0.66%	3.96%
Renal disease	17.32%	29.15%	22.21%	32.97%
Malignancy	5.66%	8.13%	6.26%	8.73%
Moderate severe liver disease	0.57%	1.09%	0.84%	1.47%
Metastatic solid tumor	1%	1.76%	1.11%	1.73%
Acquired immunodeficiency syndrome	1.19%	4.47%	1.34%	4.19%

III. R codes for mediation analysis

Create functions for mediation analysis

```
library(dplyr); library(tidy); library(mediation)

#### Mediation Analysis Pipeline ####
### Step 1: outcome and gender ###
med_step.1 <- function(cohort, outcome, adj_vars, tab_row_names){
  unadj_formula = as.formula(paste(outcome, '~', 'gender'))
  adj_formula = as.formula(paste(outcome, '~', paste0(c('gender', adj_vars), c
collapse= ' + ')))

  outcome_gender_tab = table(cohort[, outcome], cohort$gender) %>% t()
  gender.prop_within_died = paste(round(100*(prop.table(outcome_gender_tab,2)
[,2]),2), "%", sep = "")
  died.prop_within_gender = paste(round(100*(prop.table(outcome_gender_tab,1)
[,2]),2), "%", sep = "")
  N_gender_overall = table(cohort$gender)

  outcome_main_unadj = glm(unadj_formula, data = cohort, family = binomial)
  outcome_gender_unadj = summary(outcome_main_unadj)$coefficient["gendermale"
,c(1,4)]
  outcome_gender_unadj_ci = confint(outcome_main_unadj)["gendermale",] %>% exp()
p()

  outcome_main = glm(adj_formula, data = cohort, family = binomial)
  outcome_gender_adj = summary(outcome_main)$coefficient["gendermale",c(1,4)]
  outcome_gender_adj_ci = confint(outcome_main)["gendermale",] %>% exp()

  main_step1_num = cbind(N_gender_overall, outcome_gender_tab, gender.prop_wi
thin_died, died.prop_within_gender) %>% t() %>% as.data.frame()
```

```

    rownames(main_step1_num) = tab_row_names
    main_step1_unadj = paste(round(exp(outcome_gender_unadj[1]), 2), " [", round(
d(outcome_gender_unadj_ci[1],2), " ",
        round(outcome_gender_unadj_ci[2],2), "] p-value =", signif
, signif(outcome_gender_unadj[2], 3), sep = "")
    main_step1_adj = paste(round(exp(outcome_gender_adj[1]), 2), " [", round(ou
tcome_gender_adj_ci[1],2), " ",
        round(outcome_gender_adj_ci[2],2), "] p-value =", si
gnif(outcome_gender_adj[2], 3), sep = "")
    return(rslt.list = list(tab = main_step1_num, unadj = main_step1_unadj, adj
= main_step1_adj))
}

### Step 2: mediator and gender ###
med_step.2 <- function(cohort, coag_def, adj_vars, tab_row_names){
  unadj_formula = as.formula(paste(coag_def, '~', 'gender'))
  adj_formula = as.formula(paste(coag_def, '~', paste0(c('gender', adj_vars),
collapse=' + ')))

  hyper_coag_gender_tab = table(cohort[,coag_def], cohort$gender) %>% t()
  gender.prop_within_coag = paste(round(100*(prop.table(hyper_coag_gender_tab
,2)[,2]),2), "%", sep = "")
  coag.prop_within_gender = paste(round(100*(prop.table(hyper_coag_gender_tab
,1)[,2]),2), "%", sep = "")
  N_gender = table(cohort$gender)

  mediator_main_unadj = glm(unadj_formula, data = cohort, family = binomial)
  gender_unadj = summary(mediator_main_unadj)$coefficient["gendermale",c(1,4)]
}
gender_unadj_ci = confint(mediator_main_unadj)["gendermale",] %>% exp()

mediator_main = glm(adj_formula, data = cohort, family = binomial)
gender_adj = summary(mediator_main)$coefficient["gendermale",c(1,4)]
gender_adj_ci = confint(mediator_main)["gendermale",] %>% exp()

main_step2_num = cbind(N_gender, hyper_coag_gender_tab, gender.prop_within_
coag, coag.prop_within_gender) %>% t() %>% as.data.frame()
  rownames(main_step2_num) = tab_row_names
  main_step2_unadj = paste(round(exp(gender_unadj[1]), 2), " [", round(gender
_unadj_ci[1],2), " ",
      round(gender_unadj_ci[2],2), "] p-value =", signif
(gender_unadj[2], 3), sep = "")
  main_step2_adj = paste(round(exp(gender_adj[1]), 2), " [", round(gender_adj
_ci[1],2), " ",
      round(gender_adj_ci[2],2), "] p-value =", signif(gen
der_adj[2], 3), sep = "")
  return(rslt.list = list(tab = main_step2_num, unadj = main_step2_unadj, adj
= main_step2_adj, mediator_model = mediator_main))
}

```

```

med_step.2.cont <- function(cohort, mediator, adj_vars){
  unadj_formula = as.formula(paste(mediator, '~', 'gender'))
  adj_formula = as.formula(paste(mediator, '~', paste0(c('gender', adj_vars),
collapse=' + ')))

  mediator_main_unadj = glm(unadj_formula, data = cohort, family = gaussian(link = "log"))
  gender_unadj = summary(mediator_main_unadj)$coefficient["gendermale",c(1,4)]
}
gender_unadj_ci = confint(mediator_main_unadj)["gendermale",]

mediator_main = glm(adj_formula, data = cohort, family = gaussian(link = "log"))
gender_adj = summary(mediator_main)$coefficient["gendermale",c(1,4)]
gender_adj_ci = confint(mediator_main)["gendermale",]

main_step2_unadj = paste(round(gender_unadj[1], 2), " [", round(gender_unadj_ci[1],2), ", ",
round(gender_unadj_ci[2],2), "] p-value =", signif(gender_unadj[2], 3), sep = "")
main_step2_adj = paste(round(gender_adj[1], 2), " [", round(gender_adj_ci[1],2), ", ",
round(gender_adj_ci[2],2), "] p-value =", signif(gender_adj[2], 3), sep = "")
return(rslt.list = list(unadj = main_step2_unadj, adj = main_step2_adj, mediator_model = mediator_main))
}

### Step 3.1: outcome and mediator + gender ###
med_step.3 <- function(cohort, outcome, coag_def, adj_vars, tab_row_names){
  unadj_formula = as.formula(paste(outcome, '~', coag_def))
  adj_formula = as.formula(paste(outcome, '~', paste0(c('gender', coag_def, adj_vars), collapse=' + ')))

  outcome_hcoag_tab = table(cohort[, outcome], cohort[,coag_def]) %>% t()
  hyper_coag.prop_within_outcome = paste(round(100*(prop.table(outcome_hcoag_tab,2)[,2]),2), "%", sep = "")
  outcome.prop_within_hyper_coag = paste(round(100*(prop.table(outcome_hcoag_tab,1)[,2]),2), "%", sep = "")
  N_hyper_coag = table(cohort[,coag_def])
  # N_outcome = table(cohort$died_or_vent)

  response_main_unadj = glm(unadj_formula, data = cohort, family = binomial)
  hyper_coag_unadj = summary(response_main_unadj)$coefficient[coag_def,c(1,4)]
}
hyper_coag_unadj_ci = confint(response_main_unadj)[coag_def,] %>% exp()

response_main = glm(adj_formula, data = cohort, family = binomial)
hyper_coag_adj = summary(response_main)$coefficient[coag_def,c(1,4)]
hyper_coag_adj_ci = confint(response_main)[coag_def,] %>% exp()

```

```

response_gender_adj = summary(response_main)$coefficient["gendermale",c(1,4
)]
response_gender_adj_ci = confint(response_main)["gendermale",] %>% exp()

main_step3_num = cbind(N_hyper_coag, outcome_hcoag_tab, hyper_coag.prop_wit
hin_outcome, outcome.prop_within_hyper_coag) %>% t() %>% as.data.frame()
colnames(main_step3_num) = c("No hypercoagulation", "Had hypercoagulation")
rownames(main_step3_num) = tab_row_names
main_step3_unadj = paste(round(exp(hyper_coag_unadj[1]), 2), " [", round(hy
per_coag_unadj_ci[1],2), ", ",
round(hyper_coag_unadj_ci[2],2), "] p-value =", si
gnif(hyper_coag_unadj[2], 3), sep = "")
main_step3_adj = paste(round(exp(hyper_coag_adj[1]), 2), " [", round(hyper
coag_adj_ci[1],2), ", ",
round(hyper_coag_adj_ci[2],2), "] p-value =", signif
(hyper_coag_adj[2], 3), sep = "")
main_step3_gender_adj = paste(round(exp(response_gender_adj[1]), 2), " [",
round(response_gender_adj_ci[1],2), ", ",
round(response_gender_adj_ci[2],2), "] p-valu
e =", signif(response_gender_adj[2], 3), sep = "")
return(rslt.list = list(tab = main_step3_num, unadj = main_step3_unadj, adj
= main_step3_adj,
gender_adj = main_step3_gender_adj, response_model
= response_main))
}

### Step 3.2: Proportion Mediated ####
mediation.pipeline <- function(mediator_model, response_model, mediator_name)
{
  set.seed(12345)
  med.fit = mediate(mediator_model, response_model, treat = "gender", mediato
r = mediator_name,
control.value = "female", treat.value = "male", robustSE
= TRUE, sims = 1000) # , covariates = list(age_gt_65 = 1)
prop.est = as.character(signif(c(c(summary(med.fit)$n.avg, summary(med.fit)
$n.avg.ci)*100, summary(med.fit)$n.avg.p), 3))
prop.est.fmt = paste(prop.est[1], " [", prop.est[2], ", ", prop.est[3], "]
p-value =", prop.est[4], sep = "")
return(rslt.list = list(fit = med.fit, proportion = prop.est.fmt))
}

```

Implement Mediation Analysis

```

## df_main_cohort_comorb is the analytic dataset we created, which contains a
ge, gender, deceased or not, hyper coagulation/thrombosis, and individual cha
rlson comorbidities.
### Step1: outcome and gender ####
main.step1.rslt_d = med_step.1(df_main_cohort_comorb, "deceased", c("age_gt_6
5", colnames(df_comorb_analytic)[c(3,4,6:18)]), tab_row_names)

```



```
### Step2: hcoag and gender ####
main.step2.rslt = med_step.2(df_main_cohort_comorb, "hyper_coag", c("age_gt_65", colnames(df_comorb_analytic)[c(3,4,6:18)]), tab_row_names)

### Step3.1: outcome and hcoag + gender ####
main.step3.rslt_d = med_step.3(df_main_cohort_comorb, "deceased", "hyper_coag", c("age_gt_65", colnames(df_comorb_analytic)[c(3,4,6:18)]), tab_row_names)

### Step3.2: Proportion Mediated ####
mediator_name = "hyper_coag"
main.med.rslt_d = mediation.pipeline(main.step2.rslt$mediator_model, main.step3.rslt_d$response_model, mediator_name)
main.med.rslt_d$proportion

## [1] "12.1 [9.44, 15.3] p-value =0"
```