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

```r
library(dplyr); library(tidyr); 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), 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()

  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_within_died, died.prop_within_gender) %>% t() %>% as.data.frame()
```

3

```r
  rownames(main_step1_num) = tab_row_names
  main_step1_unadj = paste(round(exp(outcome_gender_unadj[1]), 2), " [", round(outcome_gender_unadj_ci[1],2), ", ",
                           round(outcome_gender_unadj_ci[2],2), "] p-value =", signif(outcome_gender_unadj[2], 3), sep = "")
  main_step1_adj = paste(round(exp(outcome_gender_adj[1]), 2), " [", round(outcome_gender_adj_ci[1],2), ", ",
                         round(outcome_gender_adj_ci[2],2), "] p-value =", signif(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(gender_adj[2], 3), sep = "")
  return(rslt.list = list(tab = main_step2_num, unadj = main_step2_unadj, adj = main_step2_adj, mediator_model = mediator_main))
}
```

4

```r
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(l
ink = "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 = "l
og"))
  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_unad
j_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(gen
der_adj[2], 3), sep = "")
  return(rslt.list = list(unadj = main_step2_unadj, adj = main_step2_adj, med
iator_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, ad
j_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()
```

5

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance
placed on this supplemental material which has been supplied by the author(s)

BMJ Open

```r
  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

```r
## 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)
```

6

```r
### Step2: hcoag and gender ####
main.step2.rslt = med_step.2(df_main_cohort_comorb, "hyper_coag", c("age_gt_6
5", 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$mediator_model, main.ste
p3.rslt_d$response_model, mediator_name)
main.med.rslt_d$proportion

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