Hypothesis-free analysis of deep vein thrombosis aetiology: a Mendelian randomization study

Andrei-Emil Constantinescu[1], Caroline J Bull[1,2], Jie Zheng[2], Benjamin Elsworth[2], Ingeborg Hers[3], Nicholas J Timpson[2], Samantha F Moore[3] and Emma E Vincent[1,2*]

[1]School of Cellular and Molecular Medicine, University of Bristol, Bristol, United Kingdom

[2]Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

[3]School of Physiology, Pharmacology and Neuroscience, University of Bristol, Bristol, United Kingdom

*Corresponding author:

Emma Vincent

Biomedical Sciences Building

Room G53

Bristol, UK

BS8 1TW

emma.vincent@bristol.ac.uk

Summary

Background

Deep vein thrombosis (DVT) is the formation of a thrombus/clot in the deep veins, when part of this clot breaks off it can travel to the lungs, resulting in pulmonary embolism. These two conditions together are known as venous thromboembolism (VTE), a leading cause of death and disability worldwide. Despite the prevalence of VTE, we do not fully understand what causes it and it is often overlooked as a major public health problem. Confirming and identifying risk factors associated with DVT is likely to lead to a reduction in the incidence, morbidity and mortality of VTE especially where these risk factors are modifiable. We can do this, by exploiting the availability of summary genetic data from genome-wide association studies (GWAS) of numerous phenotypes, including DVT, which permits hypothesis-free causal inference.

Objectives

To identify novel risk factors for DVT and to assess the causality of factors previously shown to be associated with DVT.

Methods

Two-sample Mendelian randomization (MR) was performed using summarised genetic data. Inverse variance weighted (IVW) estimates were calculated and validated by additional methods more robust to horizontal pleiotropy (MR Egger, simple mode, weighted mode, and weighted median). Bidirectional and heterogeneity sensitivity analyses were performed to further evaluate our findings.

Results

Forty-seven exposures passed an exposure-exposure correlation-adjusted Bonferroni P-value threshold (5.43E-05). These included previously hypothesised risk factors for DVT (e.g. body mass index, varicose veins, height, hyperthyroidism) and novel associations (e.g. prospective memory, basal metabolic rate).

Conclusion

Our analyses confirmed causal associations of risk factors previously associated with DVT and highlighted several novel risk factors for the disease. Our study demonstrates the utility of using a hypothesis free Mendelian randomization approach for the identification of novel disease risk factors.

Introduction

Under normal physiological circumstances, platelets and fibrin form clots to prevent blood loss at sites of vessel injury (1). Thrombosis is characterised as the abnormal formation of a clot within a blood vessel, which leads to reduced blood flow through the circulatory system (2, 3). When these abnormal clots occur in the deep veins of the legs, groin or arms this is known as deep vein thrombosis (DVT), and when part of one of these clots breaks away and becomes lodged in the lungs this is known as pulmonary embolism. These two conditions are together termed venous thromboembolism (VTE), a leading cause of death and disability worldwide.

Common symptoms of DVT include pain, swelling and tenderness of the effected limb and redness/warmth at the site of the clots, however about half of those suffering DVT will have no symptoms. Undiagnosed/untreated DVT can lead to serious health problems with 1 in 10 of those untreated developing pulmonary embolism (PE), which can then lead to heart failure and in severe cases, death (4). The symptoms of DVT alone are often not specific or sufficient to make a diagnosis, but when considered in conjunction with known risk factors, can help determine the likelihood of DVT. Furthermore, evaluating these risk factors can be used to determine whether thromboprophylaxis should be administered to prevent DVT in high risk patients.

Patients with DVT are currently treated with anticoagulants. Intravenous heparin and oral warfarin (a vitamin K antagonist) have been used in combination to treat DVT for over 50 years. Whilst new treatments (dabigatran, rivaroxaban) have shown increased anticoagulative activity relative to the traditional treatment, they do not target the main cause(s) of DVT (2, 3). Hence, the identification of

novel *causal* risk factors for DVT is desired, as this could aid in the development of an efficient prophylactic drug (3). Current research has established several risk factors for DVT: these are genetic, such as deficiencies in anticoagulants antithrombin, protein C, and protein S or acquired, such as age, obesity, cancer, pregnancy, trauma, or smoking (2, 5, 6).

Whilst most associations with a large effect size identified by observational epidemiology have been found to be causal (e.g. a positive association between smoking and cancer), observational epidemiology is limited, especially in the case of weak associations (7). This may be due to confounding, reverse causation, or bias. Mendelian randomization (MR) applies the concept of randomized controlled trials (RCTs) to genetic epidemiology, bypassing the high costs and ethical issues associated with RCTs. In an MR framework, genetic variants are used as instrumental variables (IVs) to infer the causality of potential risk factors. MR is seeing increasing application in the field of epidemiology and has been proven to give reliable effect estimates, if certain assumptions are satisfied (Box 1) (8-14). Unlike one-sample MR, which requires individual-level data, two-sample MR can be conducted using summary-level data from published genome-wide association studies (GWAS) (15). Summary data from GWAS are often publicly available, and as genetic data for exposures and outcomes can be obtained from independent datasets, this makes two-sample MR a flexible, well-powered and cost-effective method to investigate causal associations (16).

Hypothesis-driven approaches based on previous research can be subject to publication bias, which may prevent the identification of novel risk factors (10). Here, we conducted a hypothesis-free two-sample MR analysis of 973 exposures on DVT. These exposures are those curated in an online repository of genetic data; MR-Base (9). Our aim was to identify both novel risk factors for DVT through a hypothesis-free analysis, and to test the causality of traits previously shown to be associated with the disease.

Methods

Data preparation

GWAS data for exposures

Hypothesis-free two-sample MR was conducted using the TwoSampleMR R package (9). Genetic data on exposures were obtained from the MR-Base platform of harmonised GWAS summary data. MR-Base is a database containing summary data from many GWAS with an in-built analytical platform capable of performing Mendelian randomization (9). The platform allows a hypothesis-free analysis of all the exposures in MR-Base to DVT to be conducted. The exposure data encompassed lifestyle (e.g. BMI and education), disease (e.g. ulcerative colitis and squamous cell carcinoma) and biological (e.g. bone density and oestrogen levels) traits. Prior to the MR analysis, we prepared the summary data from the GWAS available in MR-Base (https://mrcieu.github.io/TwoSampleMR).

A list of studies with available GWAS summary statistics was obtained through the MR-Base API in R Studio. Non-European (N=88) and duplicate (N=138) studies were automatically removed using the dplyr R package (https://github.com/tidyverse/dplyr). In the case of duplicate studies, those with the highest sample size were kept. VTE (DVT and PE) and VTE-related (e.g. phlebitis and thrombophlebitis) traits were removed (N=9). The genetic instruments used for the analysis were single-nucleotide polymorphisms (SNPs). Genetic confounding may bias MR estimates if SNPs are correlated (17), therefore linkage disequilibrium (LD) PLINK clumping (radius = 10,000kb; r2 = 0.001) was conducted to ensure the SNPs used to instrument exposures were independent. Depending on the nature of the exposure, the reported effect size for a given SNP was expressed along with the standard error (SE) as follows: as a one standard deviation (SD) increase in the level of the risk factor per risk allele for a continuous exposure or as an odds ratio (OR) for a binary exposure.

Deep vein thrombosis data

GWAS data for European DVT cases were obtained from the Neale Lab analysis of UK Biobank data (https://github.com/Nealelab/UK_Biobank_GWAS). During a 5-year period starting from

2005, ~500,000 participants aged 45 to 69 were employed to take part in UK Biobank. DVT data was collected through an online questionnaire, while diagnosis was confirmed by verbal interview with a trained nurse at one of the Biobank Assessment Centres in the UK. Samples were originally genotyped using a custom UK Biobank Affymetrix Axiom array (18). The latest Neale Lab data set version contains auto-curated phenotypes using PHESANT, followed by genotypic data selected through SNP quality control (QC). The GWAS data from the Neale Lab consortium was divided into multiple datasets, ordered by trait. Our outcome of interest (DVT) was presented in MR-Base as "Non-cancer illness code self-reported: deep venous thrombosis (dvt)"; these summary results relate to a GWAS of 6,767 cases and 330,392 controls.

Data harmonisation

For exposure and outcome data harmonisation, incorrect but unambiguous alleles were corrected, while ambiguous alleles were removed. In the case of palindromic SNPs (A/T or C/G), allele frequencies were used to solve ambiguities. Traits that did not have genetic variants in the DVT GWAS were excluded (N = 483), resulting in a final list of 973 exposure phenotypes with which to perform the MR analysis **(Supplementary Table 1**).

Mendelian Randomization Analyses

Hypothesis-Free MR Analysis of human traits on DVT

Hypothesis-free two-sample MR was conducted using the TwoSampleMR R package (9). The causal effect of a given trait on DVT was estimated using the inverse-variance weighted (IVW) method for traits with more than one SNP. Wald ratios were derived for traits with a single SNP. Additional MR methods were also performed as sensitivity analyses where genetic instruments were comprised of more than 3 SNPs (MR Egger, simple mode, weighted mode, and weighted median) (19).

Multiple testing correction

As our analysis required a large number of phenotypes to be studied for their association with DVT, we expected that some of these traits might be highly correlated with each other. Therefore, we used PhenoSpD (20) to estimate the number of independent variables present in a correlation structure comprised of the particular traits of interest in order to correct for multiple testing. We used metaCCA (21) to create a phenotypic correlation matrix by Pearson correlation between each phenotype, with the aid of GWAS summary data. This correlation matrix was used as an input for PhenoSpD to assess the independent phenotypes through matrix spectral decomposition (22, 23). Since PhenoSpD treats exposures from separate studies as independent (e.g. BMI from study A can't be found to correlate with hip circumference from study B, even though this is most likely the case), the number of total variables before the Bonferroni correction is more stringent.

Beta coefficient transformation

Historically, studies have used logistic regression to investigate the association between a trait and disease. However, an issue with logistic regression is that effect estimates might not be representative of the whole population. Therefore, linear mixed model (LMM) methodology has gained popularity in genetic epidemiology due to its ability to control for population structure. Unlike in logistic regression, odds ratios (ORs) and risk ratios (RRs) cannot be calculated directly, but rather approximated, as the outcome from an LMM applied on a binary trait ranges on a scale of 0-1. In our study, we converted the beta coefficients from our MR analysis to RRs, using previously described methodology (24).

MR Sensitivity Analyses

Horizontal Pleiotropy Analysis

Horizontal pleiotropy occurs when a SNP affects the outcome through a separate biological pathway than the exposure of interest (e.g. a genetic variant for cholesterol affects DVT not through cholesterol, but through another biological pathway). This can bias estimation of the causal effect of an exposure and subsequently leads to type I statistical errors, thus violating a key assumption of MR in a

similar way to genetic confounding. Therefore, MR-Egger regression was performed where exposures had more than 3 SNPs to test for this type of pleiotropy (27).

Heterogeneity Analysis

As part of the MR analysis, the causal effect of the genetic variants estimating for a single trait is assumed to be the same (homogenous). However, an increase in the number of instruments for an exposure can lead to heterogeneity, especially when there are multiple mechanisms through which the exposure might affect the outcome (e.g. variants associated with BMI may be associated with DVT via a number of alterations to the circulating metabolome) (28). To test for genetic heterogeneity, we used the maximum likelihood estimator and MR-Egger for the results which passed multiple testing correction.

Bidirectional MR

MR analysis of DVT on human traits

We performed a bidirectional MR analysis to assess the direction of the causal association between our exposures and DVT (i.e. to confirm that exposures alter risk of DVT and not vice-versa). As such, we performed an additional MR analysis, with DVT as the exposure and the traits causally associated with DVT (as evidenced by our primary analysis) as outcomes. This was conducted to identify potential pathways of reverse causation, which would invalidate MR assumptions (25, 26).

Results

Of the 973 phenotypes investigated, 945 were identified as independent by PhenoSpD, setting the Bonferroni P-value threshold for our MR analysis at 5.43E-5. Forty-seven phenotypes were found to be significantly associated with DVT at this threshold (**Figure 1**, **Table 1**). The results of the MR analyses for all exposures are shown in **Supplementary Table 2**. We were able to confirm the

association of traits related to adiposity, an established risk factor for DVT, such as "Body Mass Index" (Log RR: 0.40, 95% CI: 0.32 to 0.47; P = 1.60E-22), "Waist circumference" (Log RR: 0.50, 95% CI: 0.40 to 0.59; P = 1.74E-22) and "Hip circumference" (Log RR: 0.36, 95% CI: 0.28 to 0.45; P = 2.22E-13). Other risk factors previously found to be associated with DVT were "Comparative height size at age 10" (Log RR: 0.30, 95% CI: 0.20 to 0.40; P = 1.93E-06) and "Hyperthyroidism/thyrotoxicosis" (Log RR: 2.39, 95% CI: 1.88 to 2.90; P = 8.69E-18).

We found several novel associations, such as "Varicose veins" (Log RR: 1.90, 95% CI: 1.30 to 2.50; P = 2.36E-07) and "Varicose veins of the lower extremities" (Log RR: 3.40, 95% CI: 2.31 to 4.49; P = 5.13E-07), "Basal metabolic rate" (Log RR: 0.45, 95% CI: 0.36 to 0.54; P = 2.62E-20), "Treatment/medication code: warfarin" (Log RR: 4.29, 95% CI: 3.09 to 5.49; P = 1.40E-09), "Treatment/medication code: carbimazole" (Log RR: 3.60, 95% CI: 2.70 to 4.50; P = 2.41E-12) and "Prospective memory result" (Log RR: 1.46, 95% CI: 1.02 to 1.90; 5.33E-08).

Over 50% of the exposures which passed our P-value threshold were found to be heterogenous (N=27) using the maximum likelihood method. Of these, most (N=24) were traits related to body size (mass and adiposity). The remaining heterogenous traits were basal metabolic rate (PHet: 3.71E-03), warfarin (PHet: 5.66E-40), and comparative height size at age 10 (PHet: 1.56E-05). These findings coincided with our IVW and MR-Egger heterogeneity analyses.

Through our MR-Egger analysis, we found strong evidence of horizontal pleiotropy for one trait ("Qualifications: None of the above") (intercept = -5.69E-04, P = 3.35E-02). We were unable to assess whether the "Prospective memory result" trait was pleiotropic, as this exposure was instrumented using only 2 SNPs. The bidirectional MR analysis found that DVT is causally associated with "Treatment/medication code: warfarin" (P = 1.79E-30), representing a bidirectional causal effect, thus invalidating MR assumptions (**Table 2**).

Finally, "Eicosapentaenoate (EPA; 20:5n3)" (Log RR: 1.1, 95% CI: 0.75 to 1.45; P = 3.14E-07), "Stearidonate (18:4n3)" (Log RR: 1.09, 95% CI: 0.73 to 1.45; P = 1.22E-06), "Arachidonate (20:4n6)" (Log RR: 0.913, 95% CI: 0.61 to 1.22; P = 2.08E-06), and "Mania/bipolar disorder/manic

depression" (Log RR: 3.95, 95% CI: 2.60 to 5.30; P = 5.18E-06) were found to be associated with DVT. As instruments for these exposures were comprised of one SNP, we were unable to test for heterogeneity, horizontal pleiotropy or appraise the directionality of the association.

Discussion

We performed a hypothesis-free MR analysis of 973 exposures to DVT, of which 47 were found to pass a conservative P-value threshold for evidence of causality. We have confirmed the causal association of several previously established risk factors for DVT and have identified several novel associations.

One of the most well-known risk factors for DVT is adiposity and adiposity-related traits. As such, the association between these traits with DVT most likely represents a true causal relationship. Previous studies have confirmed that obesity leads to an increased incidence of DVT, and the estimate we report here for BMI largely coincides with that of a previous MR study (6). A hypothesised mechanism is that altered metabolism in people with higher adiposity levels leads to a hypercoagulable state, and due to an impaired venous return, increases the chance of thrombi formation (29, 30).

We also found that an increase in the fat-free mass of the body leads to an increased risk of DVT, which reinforces the findings from previous studies which attest that the physical increase in body measurements leads to an increase in DVT (30). As our waist circumference Log RR (0.5) was higher compared to that of BMI (0.4) and hip circumference (0.36), this suggests that the distribution of adiposity could be an important factor for DVT progression.

Height is also a well-documented risk factor for DVT and in support of this we show that "comparative height at age 10" was positively associated with DVT. Increased height leads to a greater volume of blood needed to be pumped throughout the body, which can increase the stress on the blood vessels, disrupting haemostasis. Height is also associated with an increase in body size, which might

have a standalone effect greater than that of metabolic changes due to obesity (29, 41). As expected, many body size related traits demonstrated heterogeneity.

Another risk factor that has been previously shown to be associated with DVT (by observational analysis) is hyperthyroidism (31, 32). Here we report a positive association between hyperthyroidism/thyrotoxicosis and DVT. Thyroid hormones (THs) regulate the metabolic processes in our body. An overabundance of these hormones leads to hyperthyroidism/thyrotoxicosis, leading to a hypercoagulable state and to changes in the basal metabolic rate and thermogenesis, both which affect body weight. Moreover, TH induces alterations in factor VIII synthesis and secretion, which in turn leads to an increase in thrombi formation (31, 32).

Hyperthyroidism influences basal metabolic rate and this in turn has a large impact on body weight. Increased basal metabolic rate may lead individuals to consume a larger amount of food compared to an average person. Moreover, the basal metabolic rate is regulated by thyroid hormones, and this makes sense considering that hyperthyroidism leads to an increased risk of DVT (31). Here, we found that an increase in basal metabolic rate is associated with DVT. Although hyperthyroidism and basal metabolic rate traits are clearly linked biologically, we did not find evidence of heterogeneity.

In MR-base there are genetic instruments that proxy for an increased likelihood of being prescribed a particular drug. Here, we found that the genetic instrument that proxies for an increased likelihood are being prescribed carbimazole is associated with increased risk of DVT. Carbimazole is a thionamide drug which has been used to treat thyrotoxicosis for over 60 years. It reduces the levels of circulating thyroid hormones (THs) by binding to thyroid peroxidase, the enzyme required for TH production. As hyperthyroidism/thyrotoxicosis is positively associated with DVT, we would expect that carbimazole to have a negative effect on the disease. Our MR analysis has shown that this is not the case, with the most probable explanation being that patients who take this type of thionamide drug are more likely to have been diagnosed with DVT.

We also found that the genetic instrument that proxies for an increased likelihood of being prescribed warfarin is associated with an increased risk of DVT. Warfarin is an anticoagulant used to

treat DVT which acts as a vitamin K antagonist, reducing the production of vitamin K-dependent proteins involved in coagulation (FVIIa, FIXa, FXa, and thrombin). However, initial warfarin dosage may result in skin necrosis and a hypercoagulable state due to reductions in protein C and protein S levels, paradoxically increasing the risk of DVT. Moreover, our sensitivity analysis identified a bidirectional causal effect between warfarin treatment and DVT. This would make sense, as individuals who are prescribed warfarin are more likely to already suffer from a form of VTE (2, 33).

There is some evidence that varicose veins may increase the risk of DVT (34) and here we demonstrate there is indeed a causal association. Varicose veins are characterised by their enlarged and twisted appearance. A common occurrence in varicose veins is the impaired action of leaflet valves, which prevent the blood from falling backwards. This results in the inability of the blood to fully return to the heart, leading to the enlargement of the veins, and in time, potentially an increased risk of DVT (34).

Venous blood stasis caused by immobility is also a known risk factor for DVT. Here we report a positive association between long standing illness, disability or infirmity with DVT. This most likely causes stasis of the blood flow in the veins and can be either due to a particular neurological condition or due to the paralysis of the lower limbs. Moreover, immobility may also arise due to hospitalisation and surgery or a prolonged work-, air travel-, computer-related immobility (35, 36).

In addition, current research suggests that comorbidities lead to a higher incidence of DVT. It can therefore be assumed that an individual taking prescription medication suffers from an ongoing medical condition (thus having a lower health rating), which in turn increases the risk of developing DVT. This depends on the comorbidity, as patients suffering from long-lasting conditions, such as cancer or chronic affections are more predisposed to developing DVT than other patients (37-40). This is consistent with our finding that “taking other prescription medications and overall health rating” is associated with an increased risk of DVT.

Four of the traits which passed the P-value threshold were associated with only one SNP: “Eicosapentaenoate (EPA; 20:5n3)”, “Stearidonate (18:4n3)”, “Arachidonate (20:4n6)”, and

"Mania/bipolar disorder/manic depression". This limited our capacity to discuss in further details the mechanistic insights and pathways through which these exposures might act, as we are unable to confirm the direction of a causal effect on DVT or conduct any additional sensitivity analysis. Unless additional instruments are found for these traits, a further colocalization analysis is required to assess the direction of their causal effects (42). We also report an association between prospective memory and DVT, however, we found no evidence from the literature to support this. As this trait was instrumented by 2 SNPs, we were unable to perform a horizontal pleiotropy analysis, and thus could not confirm that the genetic variants for prospective memory act only through this trait alone.

Finally, we found an association between low qualification and DVT. As this association was found to be due to horizontal pleiotropy, the genetic variants associated with this trait most likely do not act through the exposure, but rather through a different pathway, thus invalidating one of the MR assumptions. However, a case can still be made for the relationship between education and DVT. Previous research has highlighted that a lower socioeconomic status is associated with a decrease in school performance (43), and that this in turn is associated an increased incidence of VTE (44).

Strengths

Using MR, a genetic epidemiological method which utilises the availability of summary-level GWAS results, we were able to test the association between a number of exposures to a type of cardiovascular disease (DVT) for which the causes are still largely unknown. This makes using two-sample MR in a hypothesis-free manner an attractive approach, as this ensures that novel risk factor identification is not hindered by publication bias. Unlike observational epidemiology, which necessitates the collection of primary data, two-sample MR can use genetic data compiled from previous studies to appraise the association between an exposure and an outcome.

Our hypothesis-free approach has highlighted several exposures (prospective memory, basal metabolic rate) that have not been found through conventional methods, while confirming that adiposity, a previously-known risk factor, plays a large role in DVT aetiology. Detection of these

established associations increases the validity of our finding using two-sample MR. These findings may now be applied to bring additional insight into hypothesis-driven observational or laboratory studies. As genetic epidemiology is aided by the publication of new GWAS relating to the mediators of the human proteome and transcriptome, a more detailed analysis, outlying possible pathways through which an exposure might impact DVT will be possible (45-47).

Limitations

Although the number of available traits in MR-Base has risen significantly during the previous year, some traits are still in the process of being curated and introduced into the database. Moreover, it is possible that some risk factors might not possess any genetic instruments. As such, we were unable identify the association of some exposures (e.g. proteins which are associated with the disease – von Willebrand Factor or P-selectin) which were found to be potential risk factors in previous studies.

A limitation in the conversion of beta coefficients to RRs in the case of all-or-none outcome traits is represented by the ratio of the number of cases to the number of controls. When this ratio is very small, the RR cannot be calculated in the case of those traits with negative betas under a certain value. As such, one trait which we found to be significant ("Qualifications: College or University degree") in the initial stage of the analysis was left out.

As PhenoSpD is not able to assess the correlation between traits which come from different studies, the number of independent variables resulting from the PhenoSpD analysis was higher, resulting in a more stringent P-value threshold following Bonferroni correction. This might have elevated the type 2 error rate, where traits which have a true causal effect on the disease were not found to be significant as they did not pass our threshold. We have included a supplementary table with those traits that, although did not pass our P-value threshold, did show evidence for an association (**Supplementary Table 3**). Another cause of false-negative findings arises from the limited power of some instruments. This discrepancy in power leads to a variation in significance of traits which are most likely correlated. For example, although we found many traits related to adiposity to be associated with

DVT (e.g. BMI, weight, body fat percentage), exposures such as "Obesity Class 1" and "Body fat" were not.

The limitations of two-sample MR outlined above reflect that there is potential for better quality control of MR analyses, such as using only those instruments which pass a particular statistical power threshold, restricting the analysis to traits which possess more than one SNP, or using studies with a larger ratio of cases to controls in the case of those analyses where the outcome is a binary trait.

Conclusion

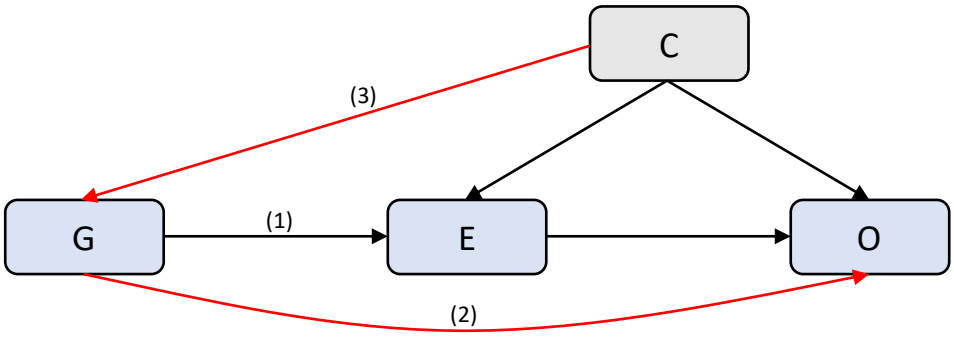
As previous studies on DVT using conventional approaches have not yielded conclusive results, here we used MR to investigate DVT aetiology. Genetic epidemiology has been gaining in popularity during the last decade, partly due to the decreasing cost of genome sequencing and as well as limitations of observational epidemiological methods in causal inference. Through a hypothesis-free approach we were able to confirm the association of previously identified risk factors for DVT (e.g. adiposity-related) and identify novel causal associations (e.g. hyperthyroidism, basal metabolic rate) with the disease. Further research is required to inform mechanistic understanding of how these exposures alter DVT risk. This could be achieved by incorporating gene expression (the human transcriptome and proteome) and pQTL (protein quantitative trait locus) data into the further MR analyses or by future hypothesis-driven studies, in an observational or laboratory setting.

References:

1. Baaten CCFMJ, ten Cate H, van der Meijden PEJ, Heemskerk JWM. Platelet populations and priming in hematological diseases. Blood Reviews. 2017;31(6):389-99.
2. Mackman N. New insights into the mechanisms of venous thrombosis. The Journal of Clinical Investigation. 2012;122(7):2331-6.
3. Stone J, Hangge P, Albadawi H, Wallace A, Shamoun F, Knuttien MG, et al. Deep vein thrombosis: pathogenesis, diagnosis, and medical management. Cardiovascular Diagnosis and Therapy. 2017;7(Suppl 3):S276-S84.
4. Zhu R, Hu Y, Tang L. Reduced cardiac function and risk of venous thromboembolism in Asian countries. Thrombosis Journal. 2017;15(1):12.
5. Samuelson Bannow BT, Konkle BA. Laboratory biomarkers for venous thromboembolism risk in patients with hematologic malignancies: A review. Thrombosis Research. 2018;163:138-45.
6. Klovaite J, Benn M, Nordestgaard BG. Obesity as a causal risk factor for deep venous thrombosis: a Mendelian randomization study. Journal of Internal Medicine. 2014;277(5):573-84.
7. Grimes DA, Schulz KF. False Alarms and Pseudo-Epidemics: The Limitations of Observational Epidemiology. Obstetrics & Gynecology. 2012;120(4).
8. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*. International Journal of Epidemiology. 2003;32(1):1-22.
9. Hemani G, Zheng J, Wade KH, Laurin C, Elsworth B, Burgess S, et al. MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. bioRxiv. 2016.
10. Evans DM, Davey Smith G. Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality. Annual Review of Genomics and Human Genetics. 2015;16(1):327-50.
11. Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. International Journal of Epidemiology. 2004;33(1):30-42.
12. Lewis SJ, Davey Smith G. Alcohol, ALDH2, and Esophageal Cancer: A Meta-analysis Which Illustrates the Potentials and Limitations of a Mendelian Randomization Approach. Cancer Epidemiology Biomarkers & Prevention. 2005;14(8):1967.
13. Lawlor Debbie A, Harbord Roger M, Sterne Jonathan AC, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine. 2008;27(8):1133-63.
14. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Human Molecular Genetics. 2014;23(R1):R89-R98.
15. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. International Journal of Epidemiology. 2017;46(6):1734-9.
16. Lawlor DA. Commentary: Two-sample Mendelian randomization: opportunities and challenges. International Journal of Epidemiology. 2016;45(3):908-15.
17. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. The American Journal of Clinical Nutrition. 2016;103(4):965-78.
18. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine. 2015;12(3):e1001779-e.
19. Hemani G, Bowden J, Haycock PC, Zheng J, Davis O, Flach P, et al. Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. bioRxiv. 2017.
20. Zheng J, Richardson T, Millard L, Hemani G, Raistrick C, Vilhjalmsson B, et al. PhenoSpD: an atlas of phenotypic correlations and a multiple testing correction for the human phenome. bioRxiv. 2017.

21. Cichonska A, Rousu J, Marttinen P, Kangas AJ, Soininen P, Lehtimäki T, et al. metaCCA: summary statistics-based multivariate meta-analysis of genome-wide association studies using canonical correlation analysis. Bioinformatics. 2016;32(13):1981-9.
22. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet. 2004;74(4):765-9.
23. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb). 2005;95(3):221-7.
24. Lloyd-Jones LR, Robinson MR, Yang J, Visscher PM. Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. Genetics. 2018;208(4):1397.
25. Zheng J, Baird D, Borges M-C, Bowden J, Hemani G, Haycock P, et al. Recent Developments in Mendelian Randomization Studies. Current Epidemiology Reports. 2017;4(4):330-45.
26. Lyngdoh T, Vuistiner P, Marques-Vidal P, Rousson V, Waeber G, Vollenweider P, et al. Serum Uric Acid and Adiposity: Deciphering Causality Using a Bidirectional Mendelian Randomization Approach. PLOS ONE. 2012;7(6):e39321.
27. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International Journal of Epidemiology. 2015;44(2):512-25.
28. Bowden J, Spiller W, Del Greco M F, Sheehan N, Thompson J, Minelli C, et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. International Journal of Epidemiology. 2018;47(4):1264-78.
29. Cushman M, O'Meara ES, Heckbert SR, Zakai NA, Rosamond W, Folsom AR. Body size measures, hemostatic and inflammatory markers and risk of venous thrombosis: The Longitudinal Investigation of Thromboembolism Etiology. Thrombosis research. 2016;144:127-32.
30. Kaze AD, Bigna JJ, Nansseu JR, Noubiap JJ. Body size measures and risk of venous thromboembolism: protocol for a systematic review and meta-analysis. BMJ open. 2018;8(3):e018958-e.
31. Mullur R, Liu Y-Y, Brent GA. Thyroid hormone regulation of metabolism. Physiological reviews. 2014;94(2):355-82.
32. van Zaane B, Squizzato A, Huijgen R, van Zanten AP, Fliers E, Cannegieter SC, et al. Increasing levels of free thyroxine as a risk factor for a first venous thrombosis: a case-control study. Blood. 2010;115(22):4344-9.
33. Binymin KA, Nasher M, Patel D. Warfarin-induced deep vein thrombosis. International medical case reports journal. 2014;7:123-5.
34. Müller B, Leutgeb, Engeser, Achankeng N, Szecsenyi, Laux. Varicose veins are a risk factor for deep venous thrombosis in general practice patients. Vasa. 2012;41(5):360-5.
35. Cushman M. Epidemiology and risk factors for venous thrombosis. Seminars in hematology. 2007;44(2):62-9.
36. Healy B, Levin E, Perrin K, Weatherall M, Beasley R. Prolonged work- and computer-related seated immobility and risk of venous thromboembolism. Journal of the Royal Society of Medicine. 2010;103(11):447-54.
37. Chung WS, Lin CL. Comorbid risks of deep vein thrombosis and pulmonary thromboembolism in patients with chronic pancreatitis: a nationwide cohort study. Journal of Thrombosis and Haemostasis. 2015;14(1):98-104.
38. Corraini P, Ording AG, Henderson VW, Szépligeti S, Horváth-Puhó E, Sørensen HT. Cancer, other comorbidity, and risk of venous thromboembolism after stroke: a population-based cohort study. Thrombosis Research. 2016;147:88-93.
39. Ording AG, Horváth-Puhó E, Garne JP, Nyström PW, Vyberg M, Sørensen HT, et al. Impact of comorbidity on risk of venous thromboembolism in patients with breast cancer: a Danish population-based cohort study. BMJ Open. 2014;4(6).
40. Smith AB, Horvath-Puhó E, Nielsen ME, Lash TL, Baron JA, Sørensen HT. Effect of comorbidity on risk of venous thromboembolism in patients with renal cell carcinoma. Urologic oncology. 2014;32(4):466-72.

41. Brækkan SK, Borch KH, Mathiesen EB, Njølstad I, Wilsgaard T, Hansen J-B. Body Height and Risk of Venous ThromboembolismThe Tromsø Study. American Journal of Epidemiology. 2010;171(10):1109-15.
42. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLOS Genetics. 2014;10(5):e1004383.
43. Thomson S. Achievement at school and socioeconomic background—an educational perspective. npj Science of Learning. 2018;3(1):5.
44. Kort D, Rein N, Meer FJM, Vermaas HW, Wiersma N, Cannegieter SC, et al. Relationship between neighborhood socioeconomic status and venous thromboembolism: results from a population-based study. Journal of Thrombosis and Haemostasis. 2017;15(12):2352-60.
45. Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA, et al. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Briefings in Bioinformatics. 2018;19(2):286-302.
46. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. Nature Communications. 2017;8:14357.
47. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558(7708):73-9.



Box 1. MR works in a similar way to randomized controlled trials (since alleles are randomly allocated at conception). It uses a genetic variant (G) as an instrument/proxy to determine whether an exposure (E) is causally associated with a disease outcome (O). During Mendelian Randomization, three conditions must be met to assure the validity of the analysis: 1) the instrument (G) is certainly associated with the exposure (E); 2) the association between the genetic instrument and the outcome (DVT) happens solely though the exposure; 3) the instrument is not associated with any confounder (8). These are invalidated by the presence of horizontal pleiotropy, where a genetic variant affects the outcome not through the studied exposure, but through a different pathway (27).

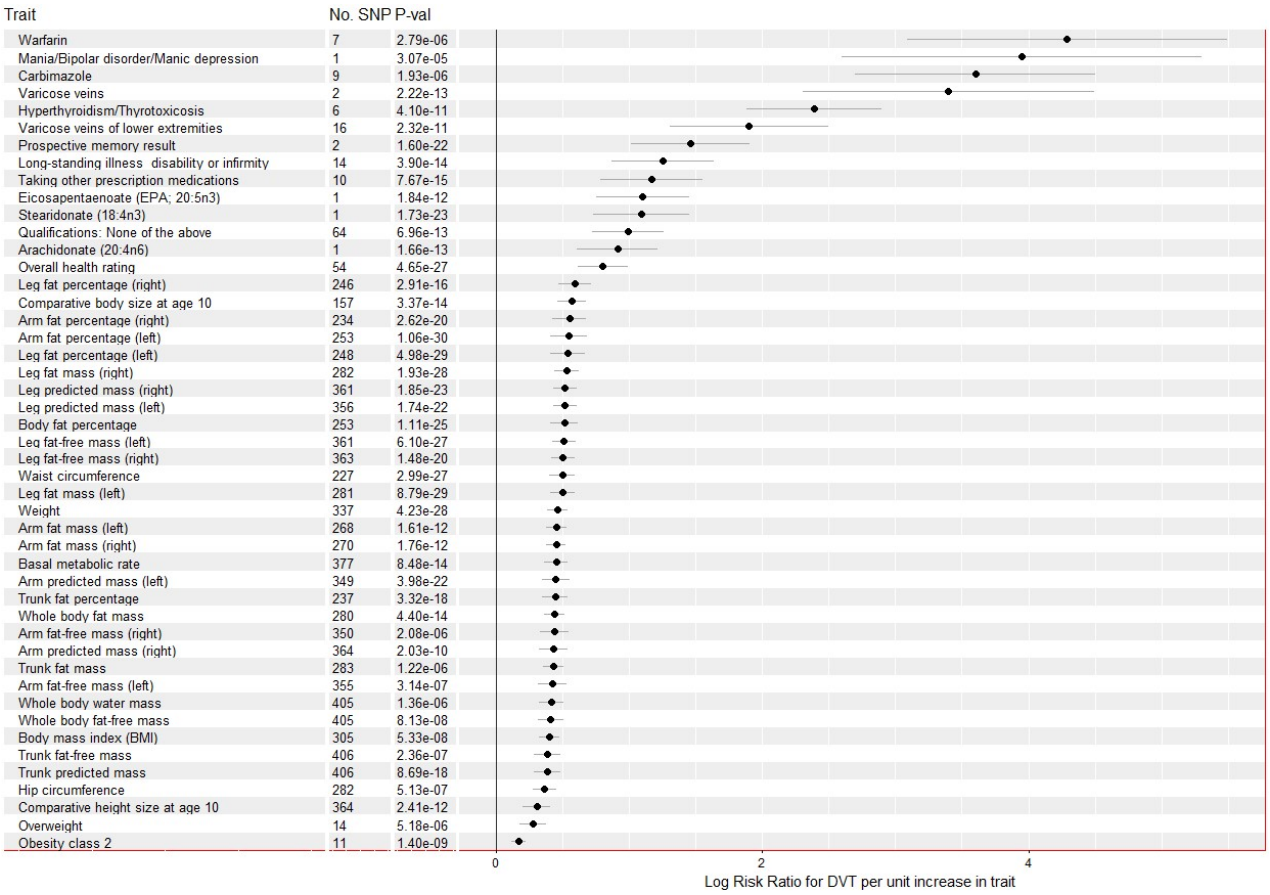


Figure 1. 1-to-many-forest plot of the exposures which passed the P-value threshold after multiple testing correction. Each trait is accompanied by two additional descriptive columns (No. SNP and Bonferroni-corrected P-value), while the Log Risk Ratio (RR) is displayed on the right, alongside with the standard error (SE). MR methods: Inverse Variance Weighted (SNP > 1) and Wald Ratio (SNP = 1).

Table 1

Main MR analysis. Methods: Inverse Variance Weighted (SNP > 1) and Wald Ratio (SNP = 1).

Exposure	Log Risk Ratio	SE	No. SNP	P-value Bonferroni	$P_{Het\ (ML)}$	P_{Plt}	CI (95%)
Arm fat mass (right)	0.45	0.04	270	1.06E-30	3.60E-01	2.82E-01	0.38 0.52
Arm fat mass (left)	0.45	0.04	268	4.98E-29	1.93E-01	1.35E-01	0.38 0.53
Leg predicted mass (right)	0.52	0.04	361	8.79E-29	1.34E-02	6.65E-01	0.43 0.60
Weight	0.46	0.04	337	1.93E-28	1.33E-03	8.57E-01	0.38 0.54
Leg fat mass (right)	0.53	0.05	282	4.23E-28	9.07E-03	4.98E-01	0.44 0.62
Leg predicted mass (left)	0.52	0.05	356	2.99E-27	5.18E-03	8.05E-01	0.43 0.60
Whole body fat mass	0.44	0.04	280	4.65E-27	1.75E-01	1.77E-01	0.36 0.51
Leg fat-free mass (left)	0.51	0.05	361	6.10E-27	4.73E-03	8.07E-01	0.42 0.60
Leg fat-free mass (right)	0.50	0.05	363	1.11E-25	5.05E-03	5.56E-01	0.41 0.59
Trunk fat mass	0.43	0.04	283	1.73E-23	2.90E-03	6.36E-01	0.35 0.51
Leg fat mass (left)	0.50	0.05	281	1.85E-23	3.71E-02	5.53E-01	0.40 0.59
Body mass index (BMI)	0.40	0.04	305	1.60E-22	6.81E-02	5.29E-01	0.32 0.47
Waist circumference	0.50	0.05	227	1.74E-22	1.65E-02	5.22E-01	0.40 0.59
Comparative body size at age 10	0.57	0.06	157	3.98E-22	5.18E-01	1.95E-01	0.46 0.68
Body fat percentage	0.51	0.05	253	1.48E-20	4.79E-02	6.35E-01	0.41 0.61
Basal metabolic rate	0.45	0.05	377	2.62E-20	3.71E-03	7.06E-01	0.36 0.54
Leg fat percentage (right)	0.59	0.06	246	3.32E-18	2.87E-03	2.40E-01	0.47 0.71
Non-cancer illness code self-reported: hyperthyroidism/thyrotoxicosis	2.39	0.26	6	8.69E-18	6.69E-01	3.87E-01	1.88 2.90
Trunk fat percentage	0.44	0.05	237	2.91E-16	2.43E-03	6.18E-01	0.35 0.54
Whole body water mass	0.42	0.05	405	7.67E-15	1.32E-04	3.44E-01	0.32 0.51
Arm predicted mass (left)	0.45	0.05	349	3.37E-14	1.53E-05	2.58E-01	0.34 0.55
Whole body fat-free mass	0.41	0.05	405	3.90E-14	2.06E-04	3.42E-01	0.31 0.50
Overall health rating	0.80	0.10	54	4.40E-14	5.14E-01	6.40E-01	0.61 0.99
Arm fat percentage (right)	0.55	0.07	234	8.48E-14	8.47E-17	6.94E-01	0.42 0.68
Arm fat-free mass (right)	0.44	0.05	350	1.66E-13	2.95E-04	2.18E-01	0.33 0.54
Hip circumference	0.36	0.04	282	2.22E-13	2.92E-04	8.76E-02	0.28 0.45
Arm predicted mass (right)	0.43	0.05	364	6.96E-13	9.35E-05	2.66E-01	0.32 0.54
Arm fat percentage (left)	0.55	0.07	253	1.61E-12	1.32E-24	6.98E-01	0.41 0.68
Leg fat percentage (left)	0.54	0.07	248	1.76E-12	7.00E-04	7.26E-01	0.40 0.67
Arm fat-free mass (left)	0.42	0.05	355	1.84E-12	3.14E-05	1.92E-01	0.32 0.53
Treatment/medication code: carbimazole	3.60	0.46	9	2.41E-12	5.21E-01	1.05E-01	2.70 4.50
Trunk fat-free mass	0.39	0.05	406	2.32E-11	2.46E-06	5.75E-02	0.29 0.48
Trunk predicted mass	0.38	0.05	406	4.10E-11	9.09E-06	5.13E-02	0.28 0.48
Qualifications: None of the above	0.99	0.14	64	2.03E-10	6.18E-01	3.35E-02	0.72 1.26
Treatment/medication code: warfarin	4.29	0.61	7	1.40E-09	5.66E-40	4.26E-01	3.09 5.49
Prospective memory result	1.46	0.23	2	5.33E-08	4.61E-01	NA	1.02 1.90
Long-standing illness disability or infirmity	1.25	0.20	14	8.13E-08	2.17E-01	4.46E-01	0.87 1.63
Diagnoses - main ICD10: I83 Varicose veins of lower extremities	1.90	0.31	16	2.36E-07	1.91E-01	5.04E-01	1.30 2.50
Eicosapentaenoate (EPA; 20:5n3)	1.10	0.18	1	3.14E-07	NA	NA	0.75 1.45
Non-cancer illness code self-reported: varicose veins	3.40	0.56	2	5.13E-07	4.42E-01	NA	2.31 4.49
Stearidonate (18:4n3)	1.09	0.18	1	1.22E-06	NA	NA	0.73 1.45
Taking other prescription medications	1.17	0.20	10	1.36E-06	4.83E-01	4.40E-01	0.79 1.55
Comparative height size at age 10	0.30	0.05	364	1.93E-06	1.56E-05	1.08E-01	0.20 0.40
Arachidonate (20:4n6)	0.91	0.16	1	2.08E-06	NA	NA	0.61 1.22
Obesity class 2	0.17	0.03	11	2.79E-06	5.45E-01	6.86E-01	0.11 0.22
Non-cancer illness code self-reported: mania/bipolar disorder/manic depression	3.95	0.69	1	5.18E-06	NA	NA	2.60 5.30
Overweight	0.28	0.05	14	3.07E-05	3.44E-01	1.71E-01	0.18 0.38

Table 2

Bidirectional MR of DVT on traits founds significant in our main MR analysis. Method: Inverse variance weighted (IVW).

Outcome	No. SNP	Beta	SE	P-value	P-value Bonferroni
Treatment/medication code: warfarin	9	0.29	0.02	3.81E-32	1.79E-30
Stearidonate (18:4n3)	5	1.35	0.50	6.78E-03	3.19E-01
Prospective memory result	9	0.40	0.17	1.64E-02	7.71E-01
Arachidonate (20:4n6)	5	0.24	0.30	2.73E-02	1.00E+00
Eicosapentaenoate (EPA; 20:5n3)	5	0.58	0.44	2.86E-02	1.00E+00
Obesity class 2	5	1.20	2.53	4.12E-02	1.00E+00
Overweight	5	-0.46	1.17	4.69E-02	1.00E+00
Non-cancer illness code self-reported: varicose veins	9	0.02	0.01	4.71E-02	1.00E+00
Treatment/medication code: carbimazole	9	0.00	0.01	8.84E-02	1.00E+00
Body mass index (BMI)	9	0.03	0.40	9.53E-02	1.00E+00
Weight	9	0.22	0.25	9.89E-02	1.00E+00
Overall health rating	9	0.27	0.16	1.87E-01	1.00E+00
Long-standing illness disability or infirmity	9	0.19	0.10	1.91E-01	1.00E+00
Body fat percentage	9	0.04	0.37	2.06E-01	1.00E+00
Whole body fat mass	9	0.13	0.36	2.38E-01	1.00E+00
Whole body fat-free mass	9	0.20	0.32	3.62E-01	1.00E+00
Whole body water mass	9	0.20	0.32	3.80E-01	1.00E+00
Basal metabolic rate	9	0.21	0.28	3.89E-01	1.00E+00
Leg fat percentage (right)	9	-0.25	0.28	3.96E-01	1.00E+00
Leg fat mass (right)	9	-0.04	0.29	4.15E-01	1.00E+00