# Development and validation of risk prediction model for venous thromboembolism in postpartum women: multinational cohort study

Alyshah Abdul Sultan,[1,2,3] Joe West,[2] Matthew J Grainge,[2] Richard D Riley,[1] Laila J Tata,[2] Olof Stephansson,[4,5] Kate M Fleming,[2,6] Catherine Nelson-Piercy,[7] Jonas F Ludvigsson[2,8]

[1]Research Institute of Primary Care and Health Sciences, Keele University, Keele ST5 5BG, UK

[2]Division of Epidemiology and Public Health, University of Nottingham, City Hospital, Nottingham NG5 1PB, UK

[3]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[4]Department of Medicine, Solna, Clinical Epidemiology Unit, Karolinska Institutet, Stockholm, Sweden

[5]Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

[6]Public Health Institute, Liverpool John Moores University, Liverpool L3 2ET, UK

[7]Women's Health Academic Centre, Guy's & St Thomas' Foundation Trust, St Thomas' Hospital, London SE1 7EH, UK

[8]Department of Paediatrics, Örebro University Hospital, Örebro, Sweden

Correspondence to: alyshah.sultan@hotmail.com

Additional material is published online only. To view please visit the journal online.

## ABSTRACT

### OBJECTIVE
To develop and validate a risk prediction model for venous thromboembolism in the first six weeks after delivery (early postpartum).

### DESIGN
Cohort study.

### SETTING
Records from England based Clinical Practice Research Datalink (CPRD) linked to Hospital Episode Statistics (HES) and data from Sweden based registry.

### PARTICIPANTS
All pregnant women registered with CPRD-HES linked data between 1997 and 2014 and Swedish medical birth registry between 2005 and 2011 with postpartum follow-up.

### MAIN OUTCOME MEASURE
Multivariable logistic regression analysis was used to develop a risk prediction model for postpartum venous thromboembolism based on the English data, which was externally validated in the Swedish data.

### RESULTS
433 353 deliveries were identified in the English cohort and 662 387 in the Swedish cohort. The absolute rate of venous thromboembolism was 7.2 per 10 000 deliveries in the English cohort and 7.9 per 10 000 in the Swedish cohort. Emergency caesarean delivery, stillbirth, varicose veins, pre-eclampsia/eclampsia, postpartum infection, and comorbidities were the strongest predictors of venous thromboembolism in the final multivariable model. Discrimination of the model was similar in both cohorts, with a C statistic above 0.70, with excellent calibration of observed and predicted risks. The model identified more venous thromboembolism events than the existing national English (sensitivity 68% *v* 63%) and Swedish guidelines (30% *v* 21%) at similar thresholds.

### CONCLUSION
A new prediction model that quantifies absolute risk of postpartum venous thromboembolism has been developed and externally validated. It is based on clinical variables that are available in many developed countries at the point of delivery and could serve as the basis for real time decisions on obstetric thromboprophylaxis.

## Introduction

The risk of venous thromboembolism increases significantly during pregnancy, peaks during the postpartum period, and is one of the leading causes of maternal mortality in developed countries.[1 2] In the United Kingdom, 50% of venous thromboembolism related maternal deaths occur during the postpartum period.[2] Targeted thromboprophylaxis can prevent postpartum venous thromboembolism with minimum harm; however, variations exist in the threshold (set on the basis of certain risk factors) at which intervention is recommended.[3-6] Although risk factors such as previous venous thromboembolism and certain thrombophilias are recognised as warranting intervention on their own, women with these risk factors represent a small proportion of all women giving birth. The vast majority of postpartum venous thromboembolisms occur in women without these specific risk factors.

Recommendations for thromboprophylaxis among a large proportion of postpartum women with only one or more "low to moderate" risk factors (such as obesity, caesarean section, and postpartum haemorrhage) are inconsistent across countries.[3-6] In the UK, women are categorised into low, intermediate, and high risk groups in the postpartum period with respect to advice on the duration of pharmacological thromboprophylaxis according to an additive ordinal point based scoring system,[5] which is not externally validated. A similar system is used in Sweden,[6] which is more conservative than those in the UK and Canada. Such categorisation, although visually pleasing and easily implemented in practice, may disguise the large variation in the actual risk of venous thromboembolism in those risk groups. Using a model to make predictions for individual women is thus more accurate and is preferred to the risk grouping approach.[7 8]

The aim of our study was to develop and externally validate a new prediction model that can generate absolute predicted risk of first venous thromboembolism within the first six weeks postpartum on the basis

## WHAT IS ALREADY KNOWN ON THIS TOPIC

The risk of venous thromboembolism (VTE) peaks during the first six weeks after delivery (postpartum) and is one of the leading causes of maternal mortality in developed countries

Targeted thromboprophylaxis can prevent VTE, but clinical risk prediction of VTE in postpartum women is rudimentary

## WHAT THIS STUDY ADDS

A risk prediction model has been developed and externally validated

This can be used as a tool to identify women at high risk of VTE, as it provides a woman's individual absolute predicted risk within the first six weeks postpartum.

The model is based on clinical variables that are likely to be available at the point of childbirth and performed better than the UK and Swedish thromboprophylaxis guidelines

of each woman's individualised clinical risk profile, and to compare this with the existing UK guideline on thromboprophylaxis from the Royal College of Obstetricians and Gynaecologists (RCOG) and the Swedish guideline.[5][6]

## Methods
### Data sources and study population
For this study, we used data from England to develop our postpartum venous thromboembolism risk score and data from Sweden to externally validate it. We have previously published the incidence of venous thromboembolism in both obstetric populations (England and Sweden) and found comparable estimates.[9][10]

### Derivation cohort (England)
The Clinical Practice Research Datalink (CPRD) is a large, longitudinal UK primary care database that covers 6% of the population.[11] Approximately 98% of the UK population are registered with general practitioners, who are responsible for almost the entirety of a patient's medical care.[12] All general practitioners participating in the CPRD are trained to record information by using the general practice Vision software. More than 50% of CPRD practices are linked to Hospital Episode Statistics (HES), which contains information on all hospital admissions in England. The anonymised patient identifiers from CPRD and HES have been linked by a trusted third party using NHS number, date of birth, postcode, and sex. Firstly, patients are matched on the basis of their NHS number (more than 90% of patients are linked in this way). The remaining patients are then linked probabilistically on postcode, date of birth, and sex. As HES covers only English hospitals, we excluded practices in Northern Ireland, Wales, and Scotland. To develop our risk prediction model, we used data on women (registered with a CPRD-HES linked practice) with no previous history of venous thromboembolism whose pregnancy ended in live birth or stillbirth between 1997 and 2014 and who had at least six weeks' postpartum follow-up.

### Validation cohort (Sweden)
The Swedish national inpatient register (IPR) was established in 1965 and has had complete national coverage since 1987.[13] More than 99% of all somatic and psychiatric hospital discharges from across Sweden are registered in the inpatient register. From 2001, IPR also includes hospital based outpatient consultations. Diagnoses in IPR are coded according to the Swedish international classification of disease system. It is mandatory for all physicians (private and publically funded) to deliver data to the IPR. Each hospital discharge and outpatient consultation is keyed to an individual's unique personal identity number,[14] which is issued to every individual in Sweden. The number is based on the combination of date of birth and a four digit number and is used by various private and public sectors to identify each individual. It is also used by the National Board of Health and Welfare to link data across various registers at the individual level. The Swedish Medical Birth Registry (SBR) contains information on more than 98% of all delivery records in Sweden since 1973. The SBR has been subjected to many quality checks, and the recorded data are of high quality and reliable.[15][16] For the purpose of validating a postpartum thrombosis risk score, we included information on pregnancies in women with no history of venous thromboembolism resulting in a live birth or stillbirth between 1 July 2005 and 31 December 2011. Using data from 2005 onwards allowed the acquisition of the national prescription data (National Prescribed Drug Register) for all of the Swedish study population.

### Defining venous thromboembolism
We defined our outcome as the occurrence of a first venous thromboembolism (deep venous thrombosis or pulmonary embolism) within the first six weeks after delivery. In both our derivation and validation cohorts, we defined venous thromboembolism by using an algorithm externally validated in the UK's primary care data with high accuracy (positive predictive value 84%).[17] Briefly, the algorithm considered a diagnosis of venous thromboembolism to be valid if it was accompanied by a prescription for an anticoagulant within 90 days of the event or if the patient died within 30 days of the event. This definition has also been shown to produce estimates of venous thromboembolism during the antepartum and postpartum periods in both English and Swedish data that are comparable to existing literature on the subject.[9][10]

### Candidate predictors, missing data, and power calculations
For each pregnant woman, we extracted information on sociodemographic and lifestyle characteristics, pre-existing comorbidities, and characteristics and complications of pregnancy and delivery from the medical record.[18] Definitions and the international classification of disease (ICD) codes used for each predictor are summarised in supplementary tables A and B. Primarily, we selected our candidate predictors from the most recent version of the RCOG's thromboprophylaxis guidelines.[5] These guidelines are already based on extensive literature review and expert consensus in opinion. We also added predictors not included in the guideline but identified in previous studies as important obstetric risk factors for venous thromboembolism that we were able to measure reliably in our data. These included diabetes, hypertension, and infant birth weight.[19-21] We measured antenatal parity that did not include current birth to avoid confusion about changing parity status during the course of pregnancy and allow for a standardised measure of parity during both antepartum and postpartum periods. For instance, a woman considered nulliparous during her first antenatal visit will remain in that category through the course of that pregnancy and the subsequent postpartum period and will be considered to have parity 1 for her subsequent pregnancy.

Our derivation cohort had missing information on pre-pregnancy body mass index (23%) and infant birth

weight (20%). We used multiple imputation to replace missing values by using a chained equation approach based on all candidate predictors. We created 10 imputed datasets for missing variables that were then combined across all datasets by using Rubin's rule to obtain final model estimates. Using the same method, we also imputed values for women with missing information on pre-pregnancy body mass index or their infant's birth weight (8.6%) in our Swedish validation cohort. On the basis of an estimated 300 venous thromboembolism events during the first six weeks postpartum and 22 candidate predictors in our derivation cohort, we had an effective sample size of 14 venous thromboembolism events per predictor, above the minimum requirement suggested by Peduzzi et al.[22]

### Statistical analysis for model development and validation

We treated occurrence of venous thromboembolism during the first six weeks postpartum as a binary outcome measure. For each of the 22 candidate predictors, we used a univariable logistic regression model to calculate the unadjusted odds ratio. For derivation of the risk prediction model, we initially included all candidate predictors in a multivariable logistic regression model. We fitted a clustering term to take account of consecutive pregnancies within women during the study period and used fractional polynomials to model potential non-linear relations between outcome and continuous predictors.

Through backwards elimination, we excluded (except for age at delivery, which was considered a prior predictor and retained in the model regardless of statistical significance[23]) candidate predictors from the

multivariable model that were not statistically significant (P>0.1 based on change in log likelihood). After elimination, we reinserted excluded predictors into the final model to further check whether they became statistically significant. We also rechecked fractional polynomial terms at this stage and re-estimated them if necessary. We formed the risk equation for predicting the log odds of venous thromboembolism by using the estimated β coefficients multiplied by the corresponding predictors included in our model together with the average intercept across patient clusters. This process ultimately led to an equation for the predicted absolute risk of venous thromboembolism: predicted risk=1/(1+e$^{-riskscore}$), where the "risk score" is the predicted log odds of venous thromboembolism from the developed model.

We assessed the performance of the model in terms of the C statistic and calibration slope (where 1.00 is ideal). The C statistic represents the probability that for any randomly selected pair of women with and without a venous thromboembolism, the women who had a venous thromboembolism had a higher predicted risk.[24] A value of 0.50 represents no discrimination and 1.00 represents perfect discrimination. We then did internal validation to correct measures of predictive performance for optimism (over-fitting) by bootstrapping 100 samples of the derivation data. We repeated the model development process in each bootstrap sample (as outlined above, including variable selection) to produce a model, applied the model to the same bootstrap sample to quantify apparent performance, and applied the model to the original dataset to test model performance (calibration slope and C statistic) and optimism (difference in test performance and apparent performance). We then estimated the overall optimism across all models (for example, derive shrinkage coefficient=average calibration slope from each of the bootstrap samples).[25] To account for over-fitting during the development process, we multiplied the original β coefficients by the uniform shrinkage factor in the final model. At this point, we re-estimated the intercept on the basis of the shrunken β coefficients to ensure that overall calibration was maintained, producing a final model.

We applied our risk prediction model to each woman in the external validation cohort on the basis of the presence of one or more risk factors (box 1). We examined the performance of this final model (in the original English data and then in the Swedish data) in terms of discrimination by calculating the C statistic. We examined calibration by plotting agreement between predicted and observed risks across tenths of predicted risk. For this external validation in Swedish data, we recalibrated the intercept on the basis of the incidence of venous thromboembolism and mean centring all predictors; however, we also assessed the performance of our model by applying the intercept from the English data because of the similar baseline venous thromboembolism risk in both populations. We applied the existing UK guidelines to the English data and the Swedish guidelines to the Swedish data to assess the number of women who qualified for pharmacological thromboprophylaxis.[5 6]

---

**Box 1: Risk prediction model**

Risk score from a logistic regression model to predict venous thromboembolism (VTE) in the first six weeks postpartum. Risk score=−9.103+0.94×(0.227*smoker*+1.221*varicose veins*+0.848*comorbidities (cardiac, renal, or inflammatory bowel disease)*+0.721*pre-eclampsia/eclampsia*+0.421*diabetes*+0.502*postpartum haemorrhage*+1.151*stillbirth*+1.097*postpartum infection*+(0.750*emergency section*/0.563*elective section*)+(0.165*parity of 1*/0.481*parity of 2*/0.566*parity of ≥3*)−0.0000798*age at delivery*$^3$+0.0000214(*age at delivery*$^3$ log (*age at delivery*))+0.00026641*BMI*$^3$−0.0000650(*BMI*$^3$ log (*BMI*))−22156315 *infant birth weight*$^{-2}$+3455223.4(*infant birth weight*$^{-2}$ log (*baby's birth weight*)))

All variables are coded as binary (0 or 1 for absence or presence of a risk factor), except for age, body mass index (BMI), and birth weight. These three variables were transformed on the basis of fractional polynomial regression (first degree) analysis. The value −9.103 is the intercept, and other numbers are the estimated regression coefficients for the predictors, which indicate their mutually adjusted relative contribution to the outcome risk. The regression coefficients represent the log odds ratio for a change of 1 unit in the corresponding predictor. The predicted risk of VTE=1/1+e$^{-riskscore}$.

*Example 1*—A 20 year old nulliparous woman who had an emergency caesarean section and has a BMI of 32. She had no history of comorbidities, developed no pregnancy related complications, and delivered a baby with birth weight of 3368 g. She has a predicted risk of 0.11% of developing a VTE within the first six weeks of delivery. *Interpretation*: if 1000 women with the same risk factors are followed, one will develop VTE within six weeks of childbirth.

*Example 2*—A 36 year old woman with a BMI of 45 who underwent an emergency caesarean section complicated with postpartum haemorrhage and infection. She had a history of cardiac disease and varicose veins. Her predicted VTE risk is 4.9% within the first six weeks of delivery. *Interpretation*: if 1000 women with the same risk factors are followed, 49 will develop VTE within six weeks of childbirth.

log=natural logarithm

---

This was based on the clinical risk factors we were able to measure reliably in our data. We then calculated the sensitivity, specificity, and positive predictive value for those women and compared them with the same proportion of women with the highest risk based on our prediction model. We repeated this analysis after excluding women already given a prescription for thromboprophylaxis. We also formally compared the number of venous thromboembolism events that were identified and missed, on the basis of our prediction model and existing guidelines, by using McNemar's test for discordant pairs, in both the English and Swedish data. Finally, we carried out a decision curve analysis to compare our prediction model with the existing thromboprophylaxis guidelines in the English and Swedish data.[24 26] This analysis assumes that the threshold probability of the disease at which a patient would opt for intervention is informative on how the patient weighs the relative harm of a false positive or a false negative prediction. This is then used to calculate the net benefit of the model across a wide range of threshold probabilities.[24] The most basic interpretation of a decision curve is that the model with the highest net benefit at a particular threshold has the highest clinical value.[26]

We used Stata version 13 for all statistical analyses. This study was conducted and reported in line with the Transparent Reporting of a multivariate prediction model for Individual Prediction or Diagnosis (TRIPOD) guidelines.[25]

### Patient involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for design or implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

## Results
### Study participants

In our derivation cohort from England, we analysed information on 321 415 women with 433 353 delivery episodes that resulted in live births or stillbirths with a complete six weeks of post-delivery follow-up. Our validated Swedish cohort had information on 498 918 women with 662 387 deliveries. Table 1 summarises the basic characteristics of the study population. Broadly, women in both cohorts had similar pre-pregnancy body mass index, delivery age, and prevalence of comorbidities (with the exception of varicose veins). Compared with England, women in Sweden were less likely to smoke and had fewer delivery related complications.

### Model development, performance measure, and validation

In the English development dataset, 312 venous thromboembolism events occurred during the first six weeks postpartum with an absolute rate of 7.2 per 10 000 deliveries. Univariable associations between postpartum venous thromboembolism and potential predictors are listed in supplementary table C. Of the 22 candidate predictors, 15 were statistically significantly associated with venous thromboembolism in our final multivariable model (table 2). Table 3 shows apparent and internal validation performance statistics of our risk prediction model. After adjustment for optimism, our final risk prediction model was able to discriminate postpartum women with and without venous thromboembolism with a C statistic of 0.70 (95% confidence interval 0.67 to 0.73). The agreement between the observed and predicted proportion of events showed excellent apparent calibration (fig 1, top), but a uniform shrinkage factor of 0.94 was needed to adjust predictor coefficients in the final model for optimism (table 3). Box 1 shows our final risk prediction model, which we integrated in a windows based calculator. Figure 2 shows a screen shot of our windows based risk calculator, which can be integrated into a designated website or a general practice/hospital computer system.

**Table 1 |** Basic characteristics of study populations. Values are numbers (percentages) unless stated otherwise

| Variable | Derivation cohort (England) (n=433 353) | Validation cohort (Sweden) (n=662 387) |
|---|---|---|
| Venous thromboembolism events | 315 (0.07) | 521 (0.08) |
| Social and demographic factors: | | |
| Mean (SD) age at delivery, years | 29.38 (5.90) | 30.32 (5.23) |
| Mean (SD) body mass index* | 24.05 (4.90) | 24.62 (4.57) |
| Smoker† (latest record before delivery) | 93 264 (21.52) | 32 502 (4.91) |
| Deliveries in 2004 or thereafter | 280 498 (64.73) | 662 387 (100) |
| Comorbidities‡: | | |
| Varicose veins | 10 935 (2.52) | 5156 (0.78) |
| Heart disease | 4431 (1.02) | 5072 (0.77) |
| Kidney disease | 4168 (0.96) | 6666 (1.01) |
| Inflammatory bowel disease | 2126 (0.49) | 5285 (0.80) |
| Pregnancy complications: | | |
| Pre-eclampsia/eclampsia | 9966 (2.30) | 24 013 (3.63) |
| Diabetes§ | 14 604 (3.37) | 14 948 (2.26) |
| Hypertension§ | 41 300 (9.53) | 7980 (1.20) |
| Parity: | | |
| Nulliparous¶ | 244 233 (56.36) | 293 176 (44.26) |
| 1 | 130 121 (30.03) | 242 341 (36.59) |
| 2 | 38 599 (8.91) | 88 803 (13.41) |
| ≥3 | 20 400 (4.71) | 38 067 (5.75) |
| Delivery characteristics/complications: | | |
| Preterm birth (<37 weeks) | 31 526 (7.27) | 31 728 (4.79) |
| Postpartum haemorrhage | 42 978 (9.92) | 48 383 (7.30) |
| Spontaneous/assisted vaginal delivery | 328 416 (75.78) | 547 654 (82.68) |
| Elective caesarean section | 44 143 (10.19) | 58 012 (8.76) |
| Emergency caesarean section | 60 794 (14.03) | 56 721 (8.56) |
| Multiple delivery (twins or more) | 6550 (1.51) | 9308 (1.41) |
| Stillbirth | 1972 (0.46) | 2286 (0.35) |
| Puerperal acute infection | 13 681 (3.16) | 48 383 (7.30) |
| Infant's mean (SD) birth weight**, g | 3368.35 (596.80) | 3519.80 (581.9) |
| Missing information: | | |
| Infant birth weight | 87 305 (20.14[20.15?]) | 930 (0.14) |
| Body mass index | 98 868 (22.81) | 57 173 (8.63) |

*Pregnancies with missing pre-pregnancy body mass index: 23% (England), 8.6% (Sweden).
†Latest smoking record before delivery.
‡Comorbidities recorded ever before delivery; for specific disease group for comorbidities, see supplementary table A.
§Including gestational and pre-existing.
¶Antenatal parity.
**Pregnancies with missing infant birthweight information: 20% (England), 0.1% (Sweden).

**Table 2 | Final multivariable analysis for venous thromboembolism risk within six weeks of delivery in derivation cohort**

| Variable | Model 1 (based on backwards elimination): odds ratio* (95% CI) | β coefficients |
|---|---|---|
| Smoker (latest record before delivery) | 1.25 (0.97 to 1.62) | 0.22684105 |
| Varicose veins | 3.39 (2.25 to 5.10) | 1.2210805 |
| Comorbidities (heart, kidney, or inflammatory bowel disease) | 2.33 (1.47 to 3.71) | 0.8476927 |
| Pre-eclampsia/eclampsia | 2.06 (1.32 to 3.20) | 0.72127433 |
| Diabetes | 1.52 (0.97 to 2.39) | 0.42119233 |
| Nulliparous (antenatal) | Reference | – |
| Parity 1 | 1.18 (0.91 to 1.53) | 0.16456948 |
| Parity 2 | 1.62 (1.13 to 2.33) | 0.48143018 |
| Parity 3 or more | 1.76 (1.56 to 2.68) | 0.5664196 |
| Postpartum haemorrhage | 1.65 (1.21 to 2.24) | 0.50183134 |
| Spontaneous/assisted vaginal delivery | Reference | – |
| Elective caesarean section | 1.76 (1.26 to 2.44) | 0.56321456 |
| Emergency caesarean section | 2.11 (1.60 to 2.80) | 0.75035197 |
| Stillbirth | 3.16 (1.29 to 7.73) | 1.1514008 |
| Postpartum infection | 2.99 (2.07 to 4.33) | 1.0969922 |
| **Fractional polynomial transformed** | | |
| Age$^3$ | – | −0.00007986 |
| Age$^3 \times$ln (age) | – | 0.00002147 |
| BMI$^3$ | – | 0.00026641 |
| BMI$^3 \times$ln (BMI) | – | −0.00006501 |
| Infant birth weight$^{-2}$ | – | −22156315 |
| Infant birth weight$^{-2} \times$ln (birth weight) | – | 3455223.4 |
| Constant† | – | −9.103121 |

Multiple imputation was carried out for missing body mass index (BMI) and birthweight data (10 imputations).
Age was included in model as a priori predictor. Predictors were retained in model at 10% level of significance.
*For binary variables, odds ratio is based on women without particular risk factor under study.
†Average across population. Constant was re-estimated after adjustment of model for optimism to ensure that overall calibration was maintained.

### External validation

In our Swedish cohort, 521 women had postpartum venous thromboembolism with an absolute rate of 7.9 per 10 000 deliveries. Applying our final risk prediction model (box 1) to the independent population after recalibration of the intercept gave a C statistic of 0.73 (0.71 to 0.75) and excellent calibration (fig 1, bottom; supplementary figure A), with the calibration slope only slightly above 1 (table 3). The mean predicted risk of venous thromboembolism based on our model was calculated to be 0.08% (min=7.73×10$^{-13}$, max=12.9%). Box 1 gives two clinical examples of the application of our risk prediction model.

### Comparing prediction models with existing guidelines

According to the UK's postpartum thromboprophylaxis guideline,[5] 35% of women in the English data qualified for pharmacological thromboprophylaxis for at least 10 days after delivery. The sensitivity and positive predictive value of the UK guideline based on our data were calculated to be 63% and 0.13% respectively. Applying our prediction model to identify this proportion of women (that is, 35% of women would qualify, which related to having a predicted absolute risk threshold of 6.3 per 10 000 deliveries) resulted in slightly higher sensitivity (68%) and positive predictive value (0.14%) (table 4). In the Swedish population, 6% of the postpartum women qualified for pharmacological thromboprophylaxis on the basis of national guidelines. The sensitivity and positive predictive value were calculated to be 21% and 0.26% respectively. Identifying this proportion of postpartum women on the basis of our model (that is, 6% qualifying, which related to an absolute risk threshold of 18 per 10 000 deliveries) resulted in a sensitivity of 30% and a positive predictive value of 0.38%. McNemar's test for discordant pairs comparing venous thromboembolisms identified using our prediction model and the existing guidelines was statistically significant in both the English (P=0.02) and Swedish (P<0.001) cohorts. These results remained virtually unchanged when we excluded 1% and 3% of women who were given pharmacological thromboprophylaxis in the English and Swedish data (supplementary table D). The performance of our model at various arbitrary thresholds is shown in supplementary table E. Finally, for each modelled threshold probability of venous thromboembolism, our model gave higher net benefit than the existing guidelines (supplementary figures B and C).

### Discussion

We have developed a new risk prediction model to calculate the absolute risk of venous thromboembolism during the first six weeks postpartum in a large representative sample of postpartum women in England. We then externally validated this model in a Swedish national cohort. Overall, our prediction model had excellent calibration and useful discrimination, with a C statistic of greater than 0.70 in both the English and Swedish data. Our risk prediction model performed better than the current UK and Swedish national thromboprophylaxis guidelines in terms of sensitivity and positive predictive value at similar thresholds on the basis of the risk factors that we were able to capture reliably in the women's medical/pregnancy records.

### Strength and limitations of study

Our risk prediction algorithm has several advantages over those in use in many developed countries.

**Table 3 | Model diagnostics (with 95% CI)**

| Measure | Apparent performance* | Test performance† | Average optimism‡ | Optimism corrected§ | External validation (Swedish data) |
|---|---|---|---|---|---|
| C statistic¶ | 0.72 (0.69 to 0.75) | 0.70 (0.70 to 0.71) | +0.020 | 0.70 (0.67 to 0.73) | 0.73 (0.71 to 0.75) |
| Calibration slope | 1.00 (0.88 to 1.11) | 0.94 (0.93 to 0.95) | +0.061 | 0.94 (0.81 to 1.04) | 1.11 (1.01 to 1.20) |

*Refers to performance estimated directly from dataset that was used to develop prediction model.
†Determined by developing model in each bootstrap sample (100 samples with replacement), calculating performance (bootstrap performance), and applying bootstrap model in original sample.
‡Average difference between model performance in bootstrap data and test performance in original dataset.
§Subtracting average optimism from apparent performance.
¶Probability that for any randomly selected pair of women with and without venous thromboembolism (VTE), women with VTE had higher predicted risk.[24] Value of 0.50 represents no discrimination and 1.00 represents perfect discrimination.
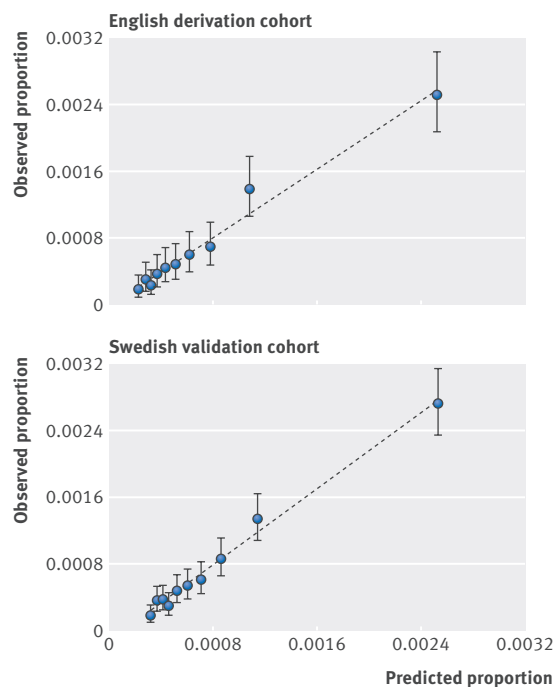
Fig 1 | Assessing calibration in English derivation cohort (top) and Swedish validation cohort (bottom). Intercept was re-estimated for Swedish population on basis of incidence of venous thromboembolism and mean centring all predictors

The model is based on absolute risks determined and validated in two very large and independent populations.[27] It is built from easily available clinical and demographic variables, implying that it can be straightforwardly applied in clinical practice and is readily amenable to further external validation in many countries that have routine data available for such a purpose. Although our model equation may seem complicated compared with the existing thromboprophylaxis guidelines, it can be easily integrated to a user friendly online calculator to be implemented in practice, being not dissimilar to those for QThrombosis.[28]

We found that less than 1% and 3% of women were given pharmacological thromboprophylaxis during the postpartum period in the English and Swedish data respectively. Although the performance of our model remained unchanged when we excluded women already given thromboprophylaxis, we have probably underestimated this proportion of women in the English data owing to unavailability of prescriptions emanating from secondary care. We believe, however, that thromboprophylaxis practices would be unlikely to have a huge effect on our estimates, as the risk of venous thromboembolism is significantly high well beyond the recommended intervention period of up to seven days[29] (guideline changed to 10 days in 2015[5]) after delivery for the vast majority of women at high risk.[30] Furthermore, evidence suggests inadequate use of thromboprophylaxis,[31 32] a belief supported by the finding of no significant change in the incidence of postpartum venous thromboembolism over time in our English data along with the recent increase in national

venous thromboembolism related deaths observed in the UK.[2 33]

Although our model has excellent calibration across the observed spectrum of absolute predicted risks, this spectrum mainly includes predicted risks that are small even for women with multiple risk factors. However, a large portion of these women qualify for pharmacological thromboprophylaxis on the basis of these small risks according to the current guidelines. This is because venous thromboembolism is the leading cause of direct maternity mortality in the UK, and venous thromboembolism related death may be prevented through targeted thromboprophylaxis. Our model enables prediction of postpartum venous thromboembolism at an individual level, which is an improvement on the currently used un-validated ordinal point based system that categorises women into low, intermediate, and high risk groups.

Although our model identifies more venous thromboembolism events than the current UK and Swedish guidelines and has been externally validated, it also missed 32% of postpartum venous thromboembolisms. This is not surprising given that a previous nationwide study from the UK reported that only 70% of women with antenatal pulmonary embolism had classic risk factors for venous thromboembolism. However, as a screening tool, our prediction model will have important implications for identifying those in whom thromboprophylaxis may be recommended.

We excluded women with a history of venous thromboembolism from our study as they represent a small proportion of women for whom the decision to give anticoagulants is less controversial. We also did not include information on a variety of risk factors (for example, protein C and S deficiency, factor V Leiden and prothrombin gene mutation, and family history of venous thromboembolism) in our model. However, by excluding women with a history of venous thromboembolism from our study, we may have limited its effect on our estimates. Furthermore, universal screening for some of those risk factors (such as factor V Leiden) is not routinely recommended in pregnant women, so pragmatically it cannot be used to predict the risk of first venous thromboembolism in the wider general population. The current RCOG thromboprophylaxis guideline recommends risk assessment for venous thromboembolism intrapartum or immediately after childbirth.[5] Thus our risk assessment tool can be used to quantify risk of venous thromboembolism during that period. However, our prediction model should not be used for women with one or more risk factors not measured in our model (for example, immobilisation due to fracture or in general) and should not be relied on to the exclusion of clinical judgment for prescribing thromboprophylaxis.

We acknowledge that our model does not take into account disease severity for specific comorbidities owing to data limitations and inadequate number of venous thromboembolism events leading to a lack of power to stratify on disease severity, if known. However, our prediction model is in line with the existing

Fig 2 | Screenshot of windows based risk prediction program. Based on hypothetical data and coefficients

guidelines that also do not take into account disease severity. Women who become pregnant are generally healthier and have lower prevalence of pre-existing comorbidities compared with the general population. In our study, the prevalence of heart, kidney, and inflammatory bowel disease was around 1% (even after use of our broad definitions). Thus the proportion of women experiencing severe heart and kidney disease during pregnancy will be even lower, and these women will probably be cared for very differently from those with well controlled comorbidities. We also believe that

whereas in theory our model could be improved by ascertaining more detail on the severity of risk factors, this would negatively affect the ability to apply this in practice where information on risk factors will need to be obtained accurately in a relatively short clinical assessment.

### Comparison with other studies

Although several studies have examined risk factors for venous thromboembolism during the postpartum period, studies specifically designed to develop and

**Table 4 | Comparing current guidelines with risk prediction model**

| Statistics | English data: total No postpartum women=433 353; total No VTE events=312 (imputed results) | | Swedish data: total No pregnancies=662 387; total No VTE events=521 (imputed results) | | |
| --- | --- | --- | --- | --- | --- |
| | Women given thromboprophylaxis based on RCOG postnatal thromboprophylaxis guidelines* | Risk prediction model (England): top 35% cut-off (threshold=6.3 per 10 000 deliveries) | Women given thromboprophylaxis based on Swedish national guidelines† | Risk prediction model (Sweden): top 6% cut-off (threshold=18 per 10 000 deliveries) | Risk prediction model (Sweden): top 35% cut-off (threshold=7.2 per 10 000 deliveries) |
| Total No (%) postpartum women warranting thromboprophylaxis | 149 402 (34.5) | 149 402 (34.5) | 41 254 (6.2) | 41 254 (6.2) | 231 835 (35) |
| Observed VTE events‡ | 197 | 212 | 109 | 158 | 355 |
| Mean predicted risk per 10 000 pregnancies | 12.3 | 13.0 | 25.8 | 31.6 | 14.2 |
| Sensitivity§, % (95% CI) | 63.1 (57.5 to 68.5) | 67.9 (62.5 to 73.1) | 20.9 (17.5 to 24.7) | 30.3 (26.4 to 34.5) | 68.1 (63.9 to 72.1) |
| Positive predictive value¶, % (95% CI) | 0.13 (0.11 to 0.15) | 0.14 (0.12 to 0.16) | 0.26 (0.21 to 0.31) | 0.38 (0.32 to 0.45) | 0.15 (0.13 to 0.17) |
| Specificity**, % (95% CI) | 65.6 (65.4 to 65.7) | 65.6 (65.4 to 65.7) | 93.8 (93.7 to 93.8) | 93.8 (93.7 to 93.9) | 65.1 (64.9 to 65.2) |

*Women with either two low risk factors (varicose veins, age >35 years, overweight, body mass index (BMI) 30-39, parity ≥3, smoker, puerperal infection, elective caesarean section, multiple delivery, preterm birth, stillbirth, pre-eclampsia/eclampsia, or postpartum haemorrhage) or one high risk factor (comorbidities (inflammatory bowel disease (IBD), heart disease, kidney disease), BMI ≥40, or emergency caesarean section)).
†Women with two clinical risk factors (elective caesarean section, age ≥40 years, BMI ≥30, or any comorbidities (heart disease, IBD, or kidney disease)).
‡In women warranting thromboprophylaxis.
§Percentage of true positive venous thromboembolism (VTE) cases correctly identified on basis of current thromboprophylaxis guidelines/risk prediction model.
¶Percentage of women without VTE diagnosis correctly identified on basis of thromboprophylaxis guidelines/risk prediction model.
**Likelihood that women above treatment threshold will develop VTE.

validate the risk prediction model are scarce. Previously, two risk models were constructed using Swedish data. One was based on a weighted risk score for exposures associated with at least a fivefold increase in risk of venous thromboembolism[6 34]; the other, an individualised risk assessment tool, was based on absolute risks of venous thromboembolism.[35] Both risk models failed to include some of the common clinical risk factors such as postpartum haemorrhage, type of caesarean section, and puerperal infection, which are important predictors of venous thromboembolism.[5] Furthermore, the first model (weighted risk score) was based on a small number of pregnant women (<2500) from a single centre that does not comprehensively inform the performance of the model. The UK, Sweden, and Canada use a points based algorithm to identify high risk postpartum women for pharmacological thromboprophylaxis.[4-6] The US guidelines are more conservative and advise against the routine use of pharmacological thromboprophylaxis.[3] These guidelines do not take into account the individualised absolute risk of venous thromboembolism based on women's complex clinical risk profile, and nor have they been externally validated, which is crucial to facilitate their translation into practice.

### Conclusion and policy implications
Our study has three important implications for clinical practice. Firstly, our prediction model can be used as a tool to identify women at high risk on the basis of their absolute predicted risk of venous thromboembolism within the first six weeks postpartum. The algorithm is based on standard clinical variables that are likely to be available at the point of childbirth and that could be readily integrated into secondary care computer systems or developed into an app for handheld devices for ease of use. Secondly, our prediction model performed better than the current UK and Swedish thromboprophylaxis guidelines (based on the risk factors recorded in data registries) in terms of identifying a higher proportion of venous thromboembolism events. Finally, our risk prediction model could be used to establish new treatment thresholds in clinical practice through consensus development of national guidelines. For example, the Swedish approach of targeting 6% of women may be a template for the UK, as the incidence of venous thromboembolism does not vary much between the two countries. On the other hand, the current Swedish threshold may be too conservative and may be leading to under-treatment. Naturally, such changes in guidelines will need to take into account the perspectives of the healthcare providers, practitioners, and women in each country and consider the potential benefits and harms of any threshold that is chosen, which is beyond the scope of this study.

Transparency: The lead author (the manuscript's guarantor) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Data sharing: No additional data available.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) 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 and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/.

1   Drife J. Thromboembolism. *Br Med Bull* 2003;67:177-90. doi:10.1093/bmb/ldg010.

2   Knight M, Tuffnell D, Kenyon S, Shakespeare J, Gray R, Kurinczuk JJ, eds. *Saving lives, improving mothers' care: surveillance of maternal deaths in the UK 2011-13 and lessons learned to inform maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2009-13.* National Perinatal Epidemiology Unit.University of Oxford, 2015.

3   Bates SM, Greer IA, Middeldorp S, et al. VTE, thrombophilia, antithrombotic therapy, and pregnancy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2012;141(2 Suppl):e691-736S.

4   Chan WS, Rey E, Kent NE, et al. VTE in Pregnancy Guideline Working Group Society of Obstetricians and Gynecologists of Canada. Venous thromboembolism and antithrombotic therapy in pregnancy. *J Obstet Gynaecol Can* 2014;36:527-53. doi:10.1016/S1701-2163(15)30569-7.

5   Royal College of Obstetricians and Gynaecologists. *Reducing the risk of venous thromboembolism during pregnancy and puerperium: Green-top guideline No. 37a.*RCOG Press, 2015.

6   Lindqvist PG, Hellgren M. Obstetric thromboprophylaxis: the Swedish guidelines. *Adv Hematol* 2011;2011:157483. doi:10.1155/2011/157483.

7   Steyerberg EW, Moons KG, van der Windt DA, et al. PROGRESS Group. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;10:e1001381. doi:10.1371/journal.pmed.1001381.

8   Hemingway H, Croft P, Perel P, et al. PROGRESS Group. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* 2013;346:e5595. doi:10.1136/bmj.e5595.

9   Abdul Sultan A, Tata LJ, Grainge MJ, West J. The incidence of first venous thromboembolism in and around pregnancy using linked primary and secondary care data: a population based cohort study from England and comparative meta-analysis. *PLoS One* 2013;8:e70310. doi:10.1371/journal.pone.0070310.

10  Abdul Sultan A, West J, Stephansson O, et al. Defining venous thromboembolism and measuring its incidence using Swedish health registries: a nationwide pregnancy cohort study. *BMJ Open* 2015;5:e008864. doi:10.1136/bmjopen-2015-008864.

11  Medicine and Healthcare Products Regulatory Agency. Clinical Practice Research Database. 2016. http://www.cprd.com/intro.asp.

12  Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015;44:827-36. doi:10.1093/ije/dyv098.

13  Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;11:450. doi:10.1186/1471-2458-11-450.

14  Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol* 2009;24:659-67. doi:10.1007/s10654-009-9350-y.

15  Centre for Epidemiology, National Board of Health and Welfare. The Swedish Medical Birth Registry - a summary of content and quality. 2003. http://www.socialstyrelsen.se/publikationer2003/2003-112-3.

16  Cnattingius S, Ericson A, Gunnarskog J, Källén B. A quality study of a medical birth registry. *Scand J Soc Med* 1990;18:143-8.

17  Lawrenson R, Todd JC, Leydon GM, Williams TJ, Farmer RD. Validation of the diagnosis of venous thromboembolism in general practice database studies. *Br J Clin Pharmacol* 2000;49:591-6. doi:10.1046/j.1365-2125.2000.00199.x.

18  Abdul Sultan A, Grainge MJ, West J, Fleming KM, Nelson-Piercy C, Tata LJ. Impact of risk factors on the timing of first postpartum venous thromboembolism: a population-based cohort study from England. *Blood* 2014;124:2872-80. doi:10.1182/blood-2014-05-572834.

19  Liu S, Rouleau J, Joseph KS, et al. Maternal Health Study Group of the Canadian Perinatal Surveillance System. Epidemiology of pregnancy-associated venous thromboembolism: a population-based study in Canada. *J Obstet Gynaecol Can* 2009;31:611-20. doi:10.1016/S1701-2163(16)34240-2.

20  James AH, Jamison MG, Brancazio LR, Myers ER. Venous thromboembolism during pregnancy and the postpartum period: incidence, risk factors, and mortality. *Am J Obstet Gynecol* 2006;194:1311-5. doi:10.1016/j.ajog.2005.11.008.

21  Blondon M, Quon BS, Harrington LB, Bounameaux H, Smith NL. Association between newborn birth weight and the risk of postpartum maternal venous thromboembolism: a population-based case-control study. *Circulation* 2015;131:1471-6, discussion 1476. doi:10.1161/CIRCULATIONAHA.114.012749.

22  Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373-9. doi:10.1016/S0895-4356(96)00236-3.

23  Sultan AA, Tata LJ, West J, et al. Risk factors for first venous thromboembolism around pregnancy: a population-based cohort study from the United Kingdom. *Blood* 2013;121:3953-61. doi:10.1182/blood-2012-11-469551.

24  Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565-74. doi:10.1177/0272989X06295361.

25  Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1-73. doi:10.7326/M14-0698.

26  Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016;352:i6. doi:10.1136/bmj.i6.

27  Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ* 2016;353:i3140. doi:10.1136/bmj.i3140.

28  Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithm (QThrombosis) to estimate future risk of venous thromboembolism: prospective cohort study. *BMJ* 2011;343:d4656. doi:10.1136/bmj.d4656.

29  Royal College of Obstetricians and Gynaecologists. *Reducing the risk of thrombosis and embolism during pregnancy and the puerperium: green-top Guideline No. 37a.*RCOG Press, 2009.

30  Kamel H, Navi BB, Sriram N, Hovsepian DA, Devereux RB, Elkind MS. Risk of a thrombotic event after the 6-week postpartum period. *N Engl J Med* 2014;370:1307-15. doi:10.1056/NEJMoa1311485.

31  Friedman AM, Ananth CV, Lu Y-S, D'Alton ME, Wright JD. Underuse of postcesarean thromboembolism prophylaxis. *Obstet Gynecol* 2013;122:1197-204. doi:10.1097/AOG.0000000000000007.

32  Knight M. UKOSS. Antenatal pulmonary embolism: risk factors, management and outcomes. *BJOG* 2008;115:453-61. doi:10.1111/j.1471-0528.2007.01622.x.

33  Knight M. *Kenyon S, Brocklehurst P, Neilson J, Shakespeare J, Kurinczuk JJ, eds. Saving lives, improving mothers' care: lessons learned to inform future maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2009-12.* National Perinatal Epidemiology Unit, University of Oxford, 2014.

34  Lindqvist PG, Torsson J, Almqvist A, Björgell O. Postpartum thromboembolism: severe events might be preventable using a new risk score model. *Vasc Health Risk Manag* 2008;4:1081-7. doi:10.2147/VHRM.S2831.

35  Lindqvist PG, Kublikas M, Dahlbäck B. Individual risk assessment of thrombosis in pregnancy. *Acta Obstet Gynecol Scand* 2002;81:412-6. doi:10.1034/j.1600-0412.2002.810507.x.

## Supplementary tables and figures