External validation of a model to predict women most at risk of postpartum venous thromboembolism: Maternity clot risk

Lu Ban[a,b,1] (lu.ban@hotmail.com), Alyshah Abdul Sultan[c, 2] (alyshah.abdulsultan@astrazeneca.com), Joe West[a,d] (joe.west@nottingham.ac.uk), Laila J Tata[d] (laila.tata@nottingham.ac.uk), Richard D Riley[c] (r.riley@keele.ac.uk) Catherine Nelson-Piercy[e] (Catherine.Nelson-Piercy@gstt.nhs.uk), Matthew J Grainge[d,*] (matthew.grainge@nottingham.ac.uk)

a NIHR Nottingham Biomedical Research Centre (BRC), Nottingham University Hospitals NHS Trust and the University of Nottingham. C-floor, South Block, Queen's Medical Centre, Derby Road, Nottingham UK, NG72UH.

b Nottingham Digestive Diseases Biomedical Research Centre, School of Medicine, University of Nottingham. E-floor, West Block, Queen's Medical Centre, Derby Road, Nottingham UK, NG72UH.

c Centre for Prognosis, School of Primary, Community and Social Care, Keele University. David Weatherall Building, Keele, Staffordshire, UK, ST5 5BG

d Division of Epidemiology and Public Health, School of Medicine, University of Nottingham. Clinical Sciences Building Phase 2, City Hospital, Nottingham, UK, NG5 1PB

e Women's Health Academic Centre, Guy's and St Thomas' NHS Foundation Trust, St Thomas Hospital. Westminster Bridge Rd, London, UK, SE1 7EH.

1 Present address: Evidera PPD, Beijing, China.

2 Present address: CVRM Epidemiology, BioPharmaceuticals Medical, AstraZeneca, Cambridge, UK, SG8 6EE

*Correspondence to: Dr Matthew J. Grainge

Email: matthew.grainge@nottingham.ac.uk,

Telephone: +44(0)115 8230456,

Orcid: https://orcid.org/0000-0001-7181-4042

Word count: 4,362

ABSTRACT

Introduction: Venous thromboembolism (VTE) is the leading cause of direct maternal mortality in high-income countries. We previously developed a risk prediction score for postpartum venous thromboembolism (VTE) in women without a previous VTE. In this paper, we provide further external validation and assess its performance across various groups of postpartum women from England.

Materials and Methods: Cohort study using primary and secondary care data covering England. We used data from QResearch comprising women with pregnancies ending in live birth or stillbirth recoded in Hospital Episodes Statistics between 2004 and 2015. Outcome was VTE in the 6 weeks postpartum. Our predictor variables included sociodemographic and lifestyle characteristics, pre-existing comorbidities, and pregnancy and delivery characteristics.

Results: Among 535,583 women with 700,185 deliveries, 549 VTE events were recorded (absolute risk of 7.8 VTE events per 10,000 deliveries). When we compared predicted probabilities of VTE for each woman from the original model with actual VTE events, we obtained a C-statistic of 0.67 (95% CI 0.65 to 0.70). However, our model slightly over-predicted VTE risk for the higher risk women (calibration slope=0.84; 95% CI 0.74 to 0.94). Performance was similar across groups defined by calendar time, socioeconomic status, age group and geographical area. The score performed comparably with the existing algorithm used by the UK Royal College of Obstetrician and Gynaecologists.

Conclusions: Our model enables flexibility in setting new treatment thresholds. Adopting it in clinical practice may help optimise use of low-molecular-weight heparin postpartum to maximise health gain by better targeting of high-risk groups.

Keywords: Electronic health data, venous thromboembolism, childbirth, risk prediction score, thromboprophylaxis, cohort study.

INTRODUCTION

Venous thromboembolism (VTE) is the leading direct cause of maternal mortality in high income countries and is associated with considerable preventable morbidity.[1,2] The absolute risk of VTE peaks in the six weeks following childbirth.[3–5] In 2016, we developed and published a risk prediction score which estimates the risk of VTE during the first six weeks after childbirth based on commonly recorded risk factors at the point of delivery.[6] This score was subsequently named the "Maternity Clot Risk" and is available from www.maternity-clot-risk.co.uk. The Maternity Clot Risk not only performed better than the current UK Royal College of Obstetrician and Gynaecologist (RCOG)[7] and Swedish postpartum thromboprophylaxis guidelines,[8] it also generates a predicted risk for each women which can be used in conjunction with pre-set thresholds for initiation of thromboprophylaxis. The score was originally developed using UK primary care data linked to secondary care data (Clinical Practice Research Datalink, CPRD) and was externally validated in an independent Swedish database[9] where it performed as well as in the original dataset.

The value of a risk prediction score to be used in clinical practice depends on how well it performs when it is applied in populations that are different from the population in which it was developed.[10,11] Furthermore, multiple external validation studies would be needed to fully realise the generalisability of a prediction model. Our original model was developed in UK practices that use a particular clinical computer system[12] (Vision, currently used by 9% of all practices in the UK[13]) and contribute data to the CPRD. In this paper we further validate the Maternity Clot Risk using the QResearch database, which records data from general practices that use another more commonly used system (used by 56% of all UK practices[13]) called

Egton Medical Information System (EMIS). The data have been recently linked to secondary care hospital data.

The predictive performance of a model tends to vary across settings, populations and time periods.[14] The aim of this study was therefore to perform an independent external validation of the predictive performance of the Maternity Clot Risk and assess its performance based on calendar time, age group, socioeconomic status and geographic region. External validation of prediction models is a necessary precursor to the important step change of clinicians being able to use the model in everyday clinical practice.

METHODS

A description of the initial study proposal can be found at

https://www.qresearch.org/research/approved-research-programs-and-projects/validating-a-postpartum-venous-thromboembolism-risk-prediction-model-using-qresearch/

Data source and study population

QResearch is a UK primary care database containing routinely collected healthcare data of anonymised patients from over 1,000 English general practices (https://www.qresearch.org/). QResearch has been recently linked to Hospital Episode Statistics (HES), a secondary care administrative database containing all inpatient admissions wholly or partially funded by the National Health Service in England. QResearch has been used for a wide range of clinical research, including the development and validation of various risk prediction models.[15–17] A cohort of women aged 12-59 years old with at least one delivery ending in live birth or stillbirth recorded in HES between January 2004 and December 2015 was extracted using version 41 of the linked QResearch database as the basis for the study population. HES maternity includes all births occurring in English NHS hospitals where over 97% of live births occur in England.[18] Some women had multiple deliveries included. Those with a history of VTE before the index delivery were excluded from the study.

Definition of outcome

VTE (deep vein thrombosis or pulmonary embolism) was defined based on the first ever recording of the event within the first six weeks postpartum using relevant diagnostic codes. A VTE was defined using a combination of VTE diagnoses in both primary and secondary care data and anticoagulant prescriptions as established

previously.[19] In brief, a diagnosis of VTE, in either the primary or secondary care section of the data, was considered to be confirmed if it was accompanied by a prescription for an anticoagulant in primary care within 90 days of the event or if the woman died within 30 days of the event.

Definition of predictors and subgroup variables

In line with our previous CPRD study,[6] we extracted information on sociodemographic and lifestyle characteristics, pre-existing comorbidities, and pregnancy and delivery characteristics and complications from both primary and secondary records. Methods used to define predictors in QResearch are described in supplementary table S1. We defined pre-existing medical conditions as varicose veins, cardiac disease (Ischemic heart disease, congenital heart disease, cardiac failure, cardiac arrhythmias or cardiomyopathy), renal disease (Glomerular disease, renal tubulointerstitial disease or renal failure) and inflammatory bowel disease (ulcerative colitis, Crohn's disease or non-specific IBD). Infection following delivery included infections of the respiratory system and urinary tract but not other puerperal infections.

Socioeconomic status was determined from the Townsend deprivation score, grouped into quintiles with 1 the least deprived and 5 the most deprived. Calendar time was also considered in order to provide illustration of model performance during periods when different proportions of women would have been receiving LMWH. We were unable to identify these women individually from both the development and validation datasets. Participants were grouped into one of ten geographic regions, which were based on former strategic health authorities in the UK. For the purpose

of subgroup analyses, age was grouped into three categories (<25 years, 25-34 years, ≥35 years).

Statistical analysis

As in the model development study,[6] we treated the occurrence of postpartum VTE as a binary outcome measure (occurrence in the first six weeks postpartum: yes or no). Continuous variables (i.e. age, pre-pregnancy body mass index (BMI) and baby's birth weight) were transformed in line with the Maternity Clot Risk equation (Box 1). To account for missing data, we used multiple imputation by chained equations to create five imputed datasets where any missing values for the BMI and the baby's birth weight were estimated based on other covariates and postpartum VTE. Multiple pregnancies in the same woman were accounted for by use of a clustering term.

To each imputed dataset, we applied the Maternity Clot Risk (Box 1) to provide a predicted VTE risk for each postpartum woman. The following methods were used to evaluate the extent to which our model correctly predicts which women developed postpartum VTE.

i. Discrimination - The ability of the score to differentiate between women who did and did not develop a first postpartum VTE event

ii. Calibration - Refers to how closely the predicted first postpartum VTE risk agrees with the observed risk. This differs from discrimination as it enables detection of whether the model over or under-estimates VTE risk, either universally or at specific risk levels.

iii. Subgroup analyses – Based on age, region, socioeconomic status and calendar time were presented to explore potential heterogeneity in model performance

between different clinically important demographic subgroups or over time.

iv. Sensitivity and positive predictive value of the model in predicting postpartum VTE compared with the algorithm currently used by the Royal College of Obstetricians and Gynaecologists (RCOG).

v. Decision curve analysis – Highlights the range of thresholds for intervention based on underlying risk of VTE where the model outperforms alternative strategies for intervention.

In a sensitivity analysis, we explored whether a re-calibrated model offered improved prognostic performance. The above methods along with justification for their use are explained fully in supplementary Appendix A.

Previous research suggested that at least 100 cases and 100 non-cases would be needed for validation studies, and our sample size far exceeds this.[20] All data management and analysis were conducted using Stata 15, and the findings reported according to the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) guidance.[21]

Box 1 Maternity Clot Risk

Risk score developed from a logistic regression model in the model development study to predict the first ever venous thromboembolism in the first six weeks postpartum.

Risk score=-9.103121 + 0.94×(0.22684105×smoker + 1.2210805×varicose veins + 0.8476927×comorbidities (cardiac, renal or inflammatory bowel disease) + 0.72127433×pre-eclampsia/eclampsia + 0.42119233×diabetes + 0.50183134×postpartum haemorrhage + 1.1514008×stillbirth + 1.0969922×postpartum infection + 0.56321456×elective section + 0.75035197×emergency section + 0.16456948×parity of 1 + 0.48143018×parity of 2 + 0.5664196×parity of ≥3 - 0.00007986×age at delivery^{3}+0.00002147×(age at delivery^{3}ln(age at delivery)) + 0.00026641×BMI^{3} - 0.00006501×(BMI^{3}ln(BMI)) -22156315×infant birth weight(g)$^{-2}$ + 3455223.4×(infant birth weight^{-2}ln(infant birth weight)))

Ethical Approval

This project was approved by the QResearch Advisory Group, Project reference ID R82.

RESULTS

Baseline characteristics

We included 535,583 women with 700,185 deliveries resulting in either a live birth or stillbirth with a complete six weeks of post-delivery follow-up. There were 549 first VTE events in the first six weeks postpartum corresponding to an absolute risk of 7.8 per 10,000 deliveries (95% CI 7.2 to 8.5). Table 1 shows the basic characteristics of the study population. Broadly, compared to women in CPRD, women in QResearch had similar age at delivery and prevalence of comorbidities, slightly higher mean BMI, were less likely to be nulliparous and smoke and had slightly fewer pregnancy and delivery related complications (Table 1). There were 17.8% with missing infant birth weight and 17.3% with missing BMI in QResearch, which was lower than in CPRD.

Prediction of VTE risk

Using the maternity clot risk formula, predicted risks of VTE were calculated for each woman in the cohort. The predicted risks ranged from 0 to 745 per 10,000 deliveries (maximum equivalent to 7.5% risk of VTE); median predicted risk =5.3 per 10,000, 10^{th} percentile 3.0 per 10,000, 90^{th} percentile 14.2 per 10,000. A total of 5.0% of women had a predicted risk of VTE of 0.2% or more (or 20 per 10,000 deliveries, n=34,832), 0.6% of women had a predicted risk of 0.5% or more (n=4,242) and 0.1% had a predicted risk of 1% or more (n=756). These numbers were taken from the first multiple imputed dataset, but all the key figures (predicted risks per 10,000 and percentages) were the same to at least 2 significant figures in the other imputed datasets.

Overall model performance

After obtaining predicted probabilities of VTE from the Maternity Clot Risk score, the overall C-statistic pooled over the imputed datasets was 0.67 (95% CI 0.65 to 0.70). Calibration slope was 0.84 (0.74-0.94) and calibration-in-the-large was 0.02 (-0.06 to 0.10). Similar results were observed in each imputed dataset (Supplementary Table S2). The plotted agreement between predicted and observed risks across tenths of predicted risks is shown in supplementary figure 1. Due to the small range of predicted risks, the figures show the predicted risks up to 30 per 10,000 deliveries only.

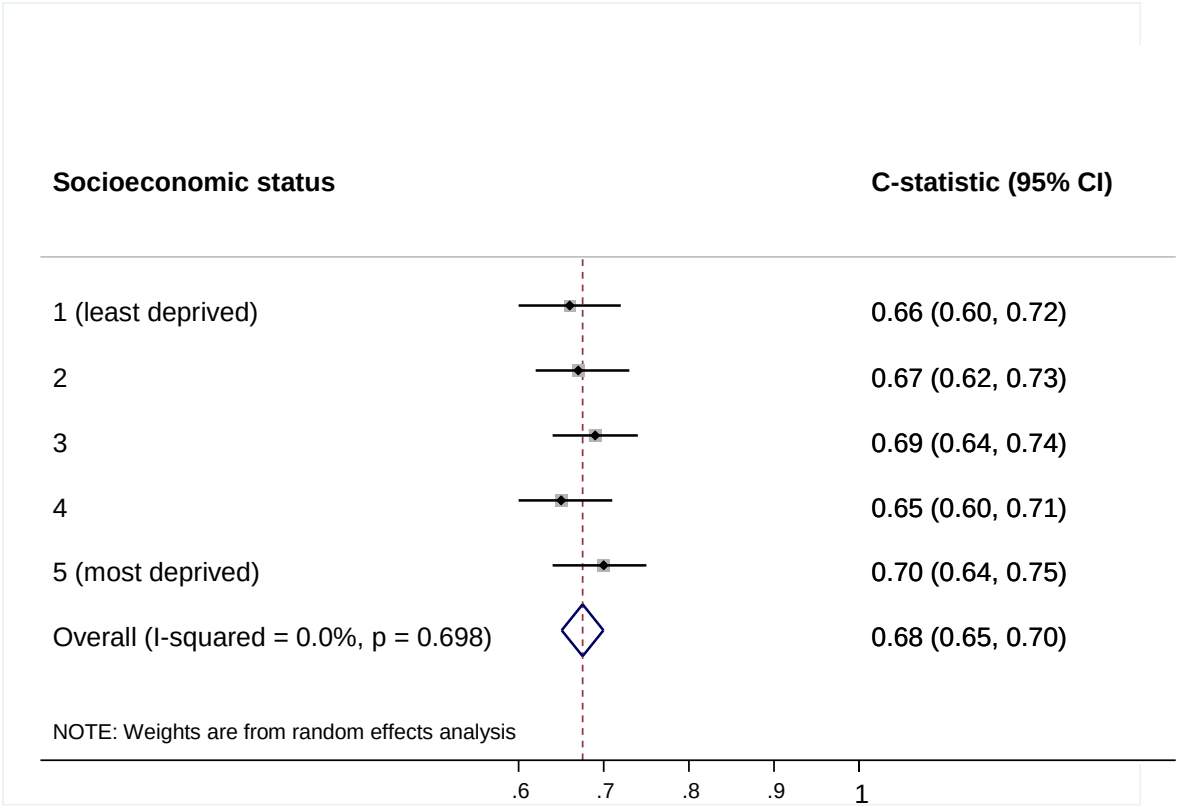
Variables	QResearch cohort (n=700,185)	CPRD cohort (n=433,353)
	n (%, if not otherwise specified)	n (%, if not otherwise specified)
VTE	549 (0.08)	315 (0.07)
Social and demographic factors:		
Mean (SD) age at delivery, years	29.85 (5.91)	29.38 (5.90)
Mean (SD) body mass index	25.06 (5.55)	24.05 (4.90)
Normal	315,624 (45.08)	-
Underweight	28,269 (4.04)	-
Overweight	141,313 (20.18)	-
Obese	94,063 (13.43)	-
Missing	120,916 (17.27)	-
Smoker (latest record before delivery)	128,029 (18.29)	93,264 (21.52)
Socioeconomic deprivation		
1 (least deprived)	130,173 (18.59)	-
2	140,584 (20.08)	-
3	151,064 (21.57)	-
4	144,769 (20.68)	-
5 (most deprived)	130,665 (18.66)	-
Missing	2,930 (0.42)	-
Comorbidities ever before delivery:		
Varicose veins	16,962 (2.42)	10,935 (2.52)
Heart disease	7,525 (1.07)	4,431 (1.02)
Kidney disease	5,314 (0.76)	4,168 (0.96)
Inflammatory bowel disease	3,756 (0.54)	2,126 (0.49)
Pregnancy complications:		
Pre-eclampsia/eclampsia	12,291 (1.76)	9,966 (2.30)
Diabetes	37,699 (5.38)	14,604 (3.37)
Hypertension	46,158 (6.59)	41,300 (9.53)
Antenatal parity		
Nulliparous	341,625 (48.79)	244,233 (56.36)
1	259,841 (37.11)	130,121 (30.03)
2	67,955 (9.71)	38,599 (8.91)

≥3	30,764 (4.39)	20,400 (4.71)
Delivery characteristics/complications:		
Preterm birth (<37 weeks)	49,610 (7.09)	31,526 (7.27)
Postpartum haemorrhage	62,244 (8.89)	42,978 (9.92)
Spontaneous/assisted vaginal delivery	523,360 (74.75)	328,416 (75.78)
Elective caesarean section	75,640 (10.80)	44,143 (10.19)
Emergency caesarean section	101,185 (14.45)	60,794 (14.03)
Multiple delivery (twins or more)	10,772 (1.54)	6,550 (1.51)
Stillbirth	3,312 (0.47)	1,972 (0.46)
Puerperal acute infection	14,043 (2.01)	13,681 (3.16)
Infant's mean (SD) birth weight, g	3356.17 (584.57)	3368.35 (596.80)
Missing information:		
Infant birth weight	124,299 (17.75)	87,305 (20.15)
Body mass index	120,916 (17.27)	98,868 (22.81)

Table 1 Characteristics of study population (number of deliveries=700,185, number of women=535,583) from QResearch cohort and CPRD cohort [6]

Performance by subgroup

Results from the analysis by different groups showed that the Maternity Clot Risk performed similarly in women of different socioeconomic groups (Figure 1), geographic regions (Figure 2), in women giving birth in different calendar periods (Figure 3) and women in different age groups (Figure 4), with minimal heterogeneity in all instances for both the C-statistic and calibration slope.



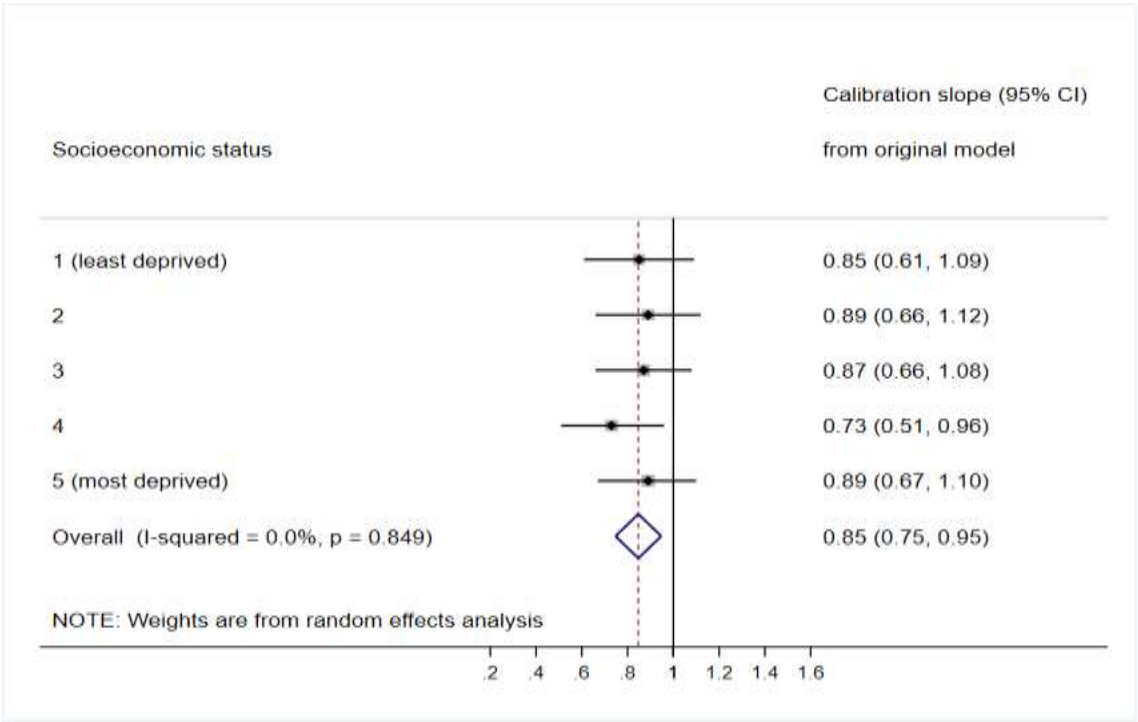
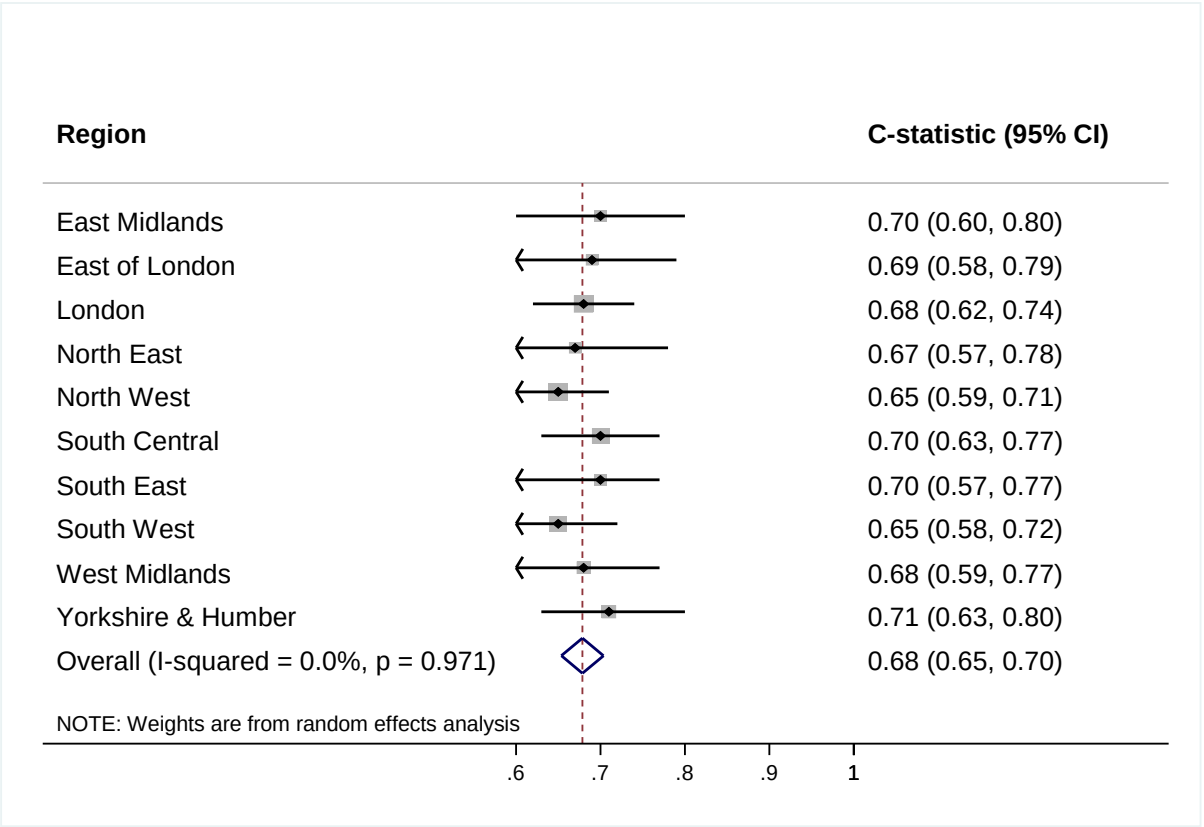


Figure 1 Model diagnostics by socioeconomic status; a) c-statistic, b) calibration slope (before re-calibration)



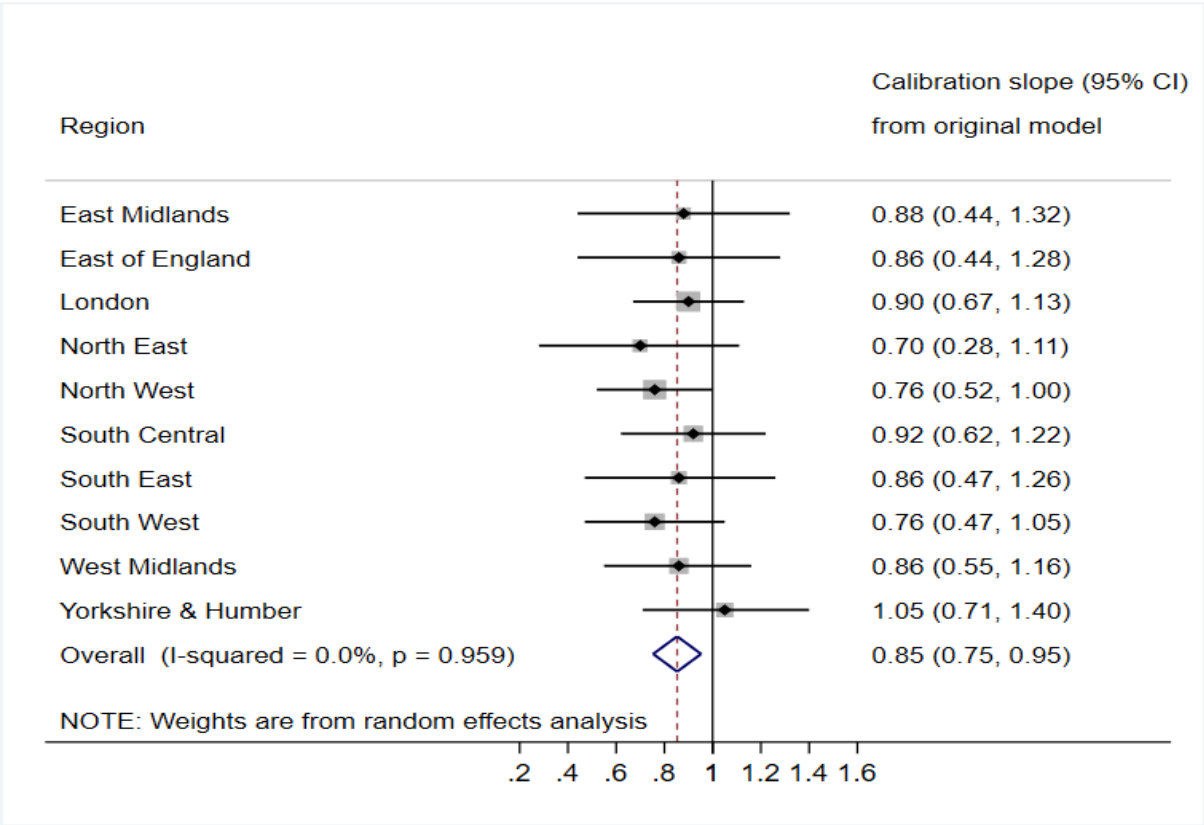
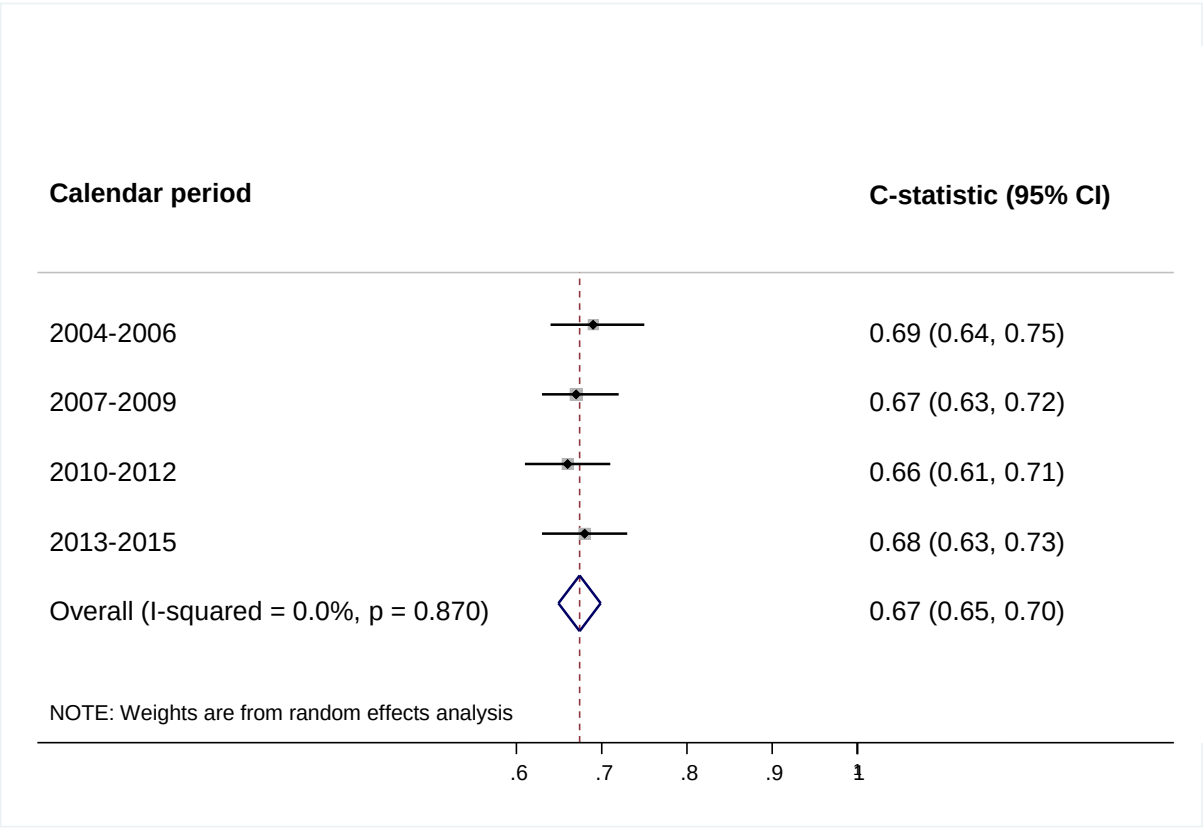


Figure 2 Model diagnostics by region; a) c-statistic, b) calibration slope (before re-calibration)



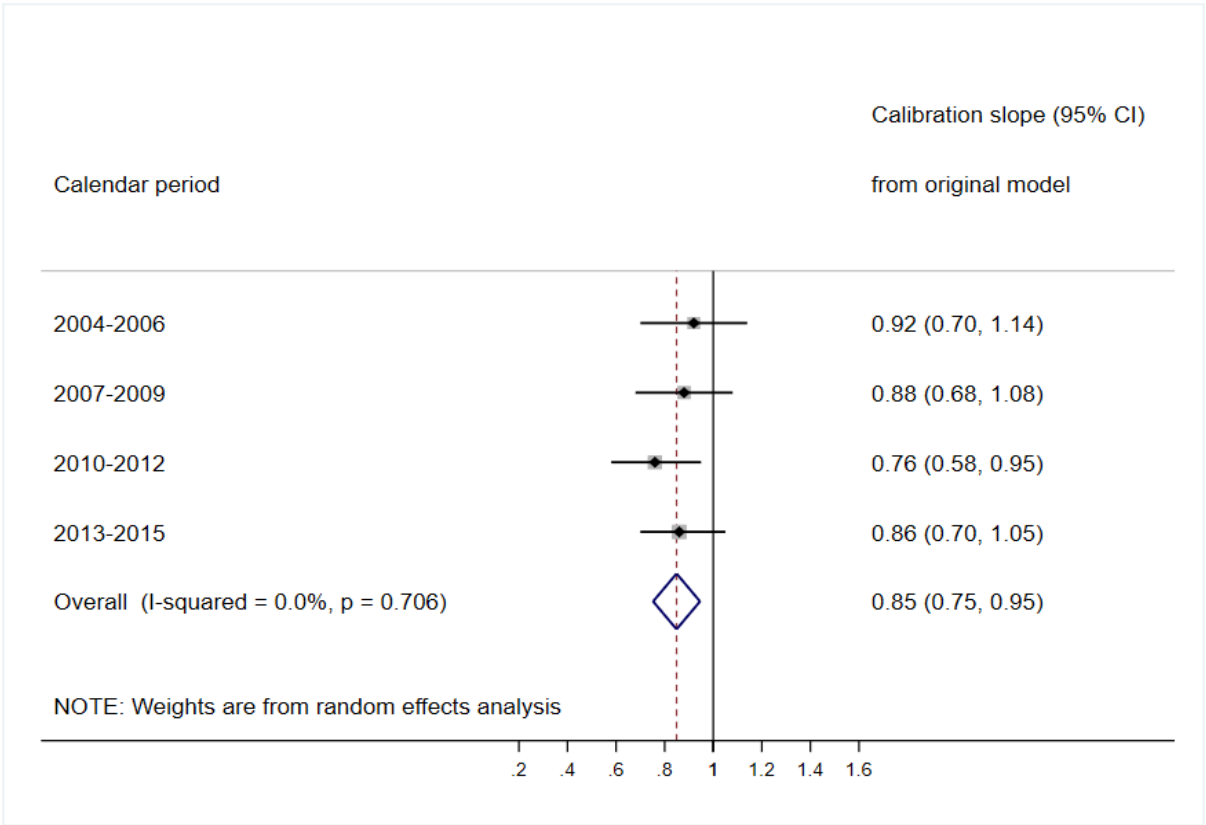
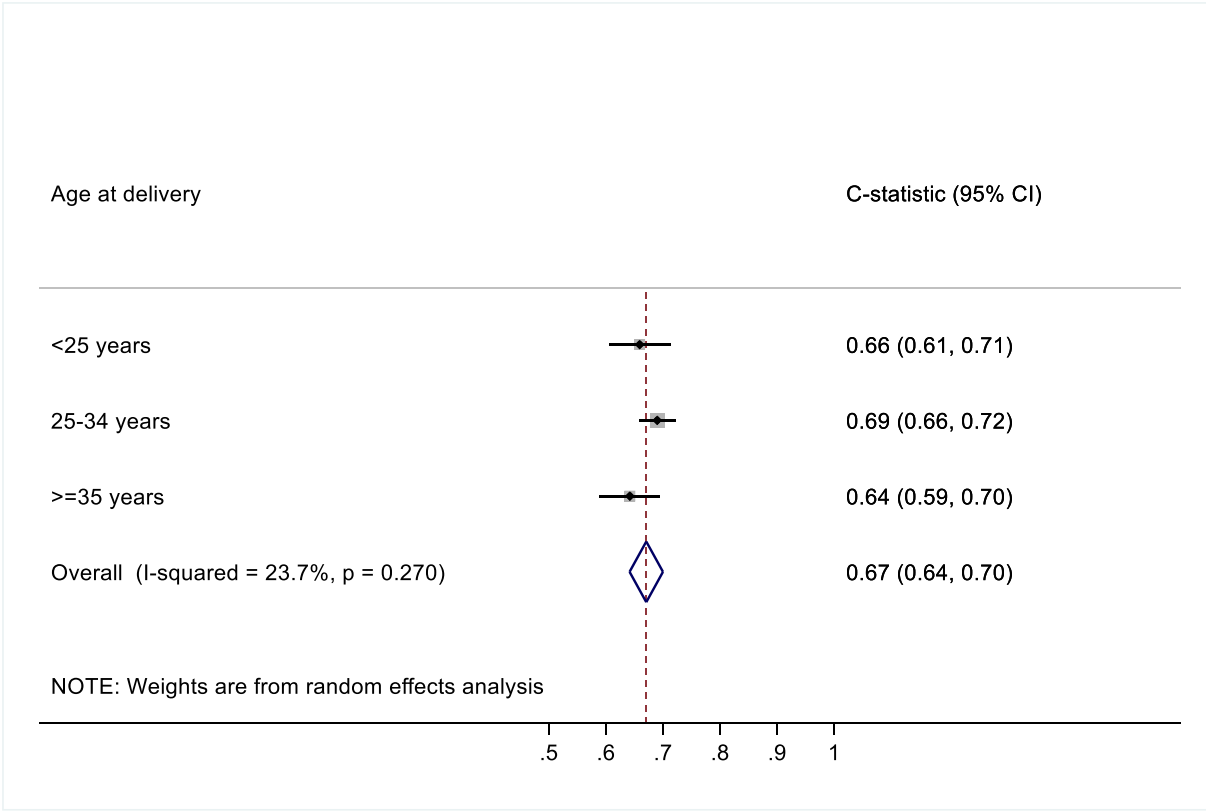


Figure 3 Model diagnostics by calendar time; a) c-statistic, b) calibration slope (before re-calibration)



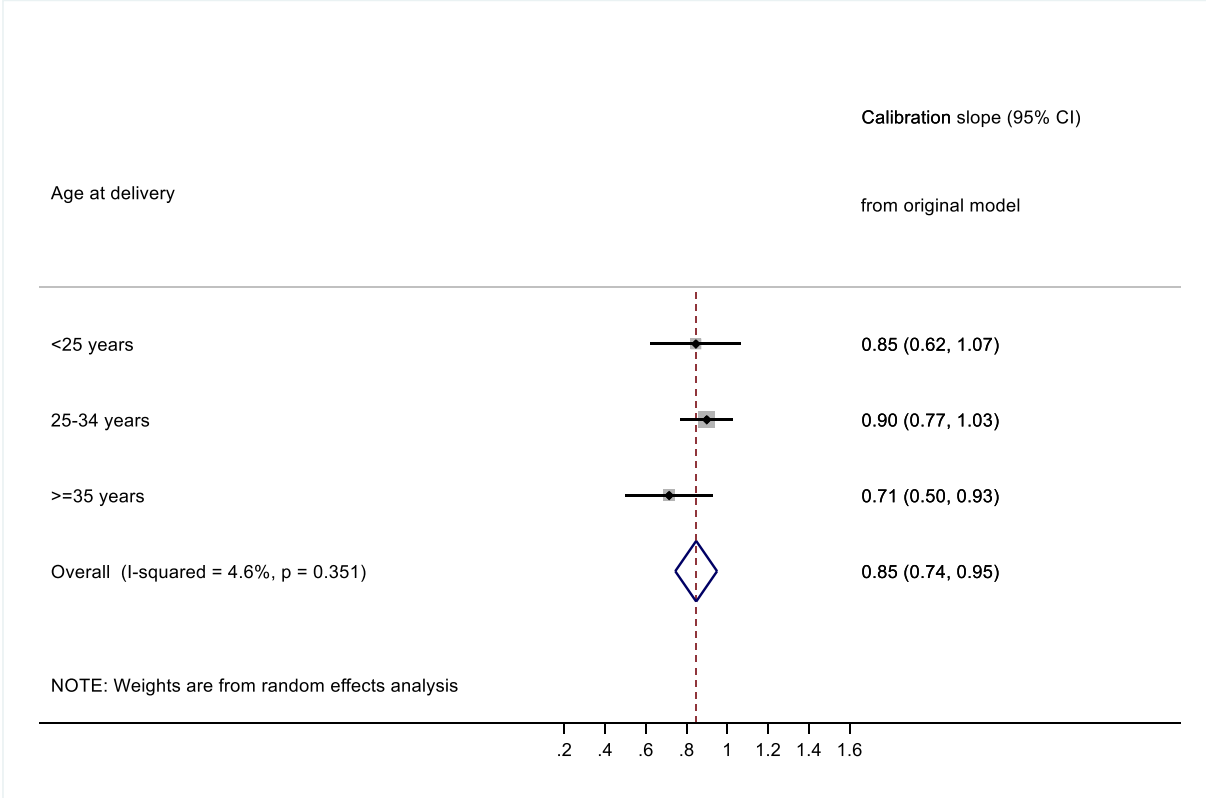


Figure 4 Model diagnostics by age at delivery; a) c-statistic, b) calibration slope (before re-calibration)

Comparison with the existing RCOG guideline

According to the current RCOG postpartum thromboprophylaxis guideline, 35.6% of women in the study population qualified for pharmacological thromboprophylaxis for at least 10 days after delivery. The results from the decision curve analysis (Supplementary Figure S2) show, although the net benefit was small, the Maternity Clot Risk was better than a treat-all or treat-none strategy between risk thresholds of 10 and 30 per 10,000 deliveries. It had higher net benefit than the current RCOG guideline between risk thresholds of 5 and 30 per 10,000 deliveries. Using the Maternity Clot Risk to identify the same proportion of women based on their predicted risks (i.e. risk threshold 6.77 per 10,000 deliveries) resulted in a slightly higher observed sensitivity (59.2, 95%CI 55.0 to 63.3 vs. than using the RCOG guideline (56.8, 95%CI 52.6 to 61.0)(Table 2), although the difference was not statistically significant.

Statistics	Based on RCOG guideline	Based on Maternity Clot Risk*
Total No(%) postpartum women warranting thromboprophylaxis	248,983 (35.6)	249,265 (35.6)
Observed VTE events	312	325
Mean predicted risk per 10,000 deliveries	13.2	14.0
Sensitivity (%)	56.8 (52.6-61.0)	59.2 (55.0-63.3)
Positive predictive value (%)	0.13 (0.11-0.14)	0.13 (0.12-0.15)
Specificity (%)	64.5 (64.3-64.6)	64.4 (64.3-64.5)

*women with a risk of VTE of 6.77 per 10,000 deliveries or above would be eligible for pharmacological thromboprophylaxis

Table 2 Comparing the Maternity Clot Risk with the existing RCOG thromboprophylaxis guideline from the original model (in imputed dataset 1, number of deliveries=700,185, number of VTE events=549)

Re-calibration results

After shrinking the original predictor coefficients by 0.79 (0.84*0.94) and re-estimating the intercept, the calibration slope was 1.00 (0.88 to 1.12). Results from other analyses remained largely unchanged (Supplementary Table S2, and Supplementary Figures S1 to S6).

DISCUSSION

Main findings

We have carried out an external validation of the Maternity Clot Risk in the largest available UK primary care dataset. It was conducted in an independent sample of women derived from UK general practice using a different clinical computer system to the CPRD. Applying the Maternity Clot Risk to the QResearch cohort resulted in an overall C-statistic of 0.67 (95% CI 0.65-0.70) and a calibration slope of 0.84 (0.74-0.94). The predictive performance was similar across time periods, socioeconomic and age groups and geographical regions. The Maternity Clot Risk had a slightly higher net benefit than the existing RCOG postpartum thromboprophylaxis guideline and the treat none strategy between risk thresholds of 10 and 30 VTE events per 10,000 deliveries. Our model has the potential to be used in maternity units if suitable thresholds for intervention could be established, although results should be interpreted in light of limitations.

Strengths and limitations

We have conducted an external validation of the Maternity Clot Risk in the UK population. It was conducted in the UK's largest primary care data with linkages to secondary care hospital data, with 549 cases. Data management and analysis were conducted by a researcher not involved in the original model development process (LB) but using the original Maternity Clot Risk and statistical methods, which further ensures robustness of our external validation. As primary care practices contributing data to QResearch use a different computer system, the women included in this study were different from those used to develop the original score. Moreover, computer systems used by CPRD and QResearch cover 67% of English practices making our findings generalisable to all women giving birth in the UK. The ethnic

diversity in England has been increasing over the last two decades and 86% of the population in England and Wales are white according to the 2011 UK census data.[22] Finally, our large sample size gave us the opportunity to assess model performance in various subgroups and assess heterogeneity based on these factors.

Limitations of this study surround the use of electronic health data for the development and validation of risk prediction models some of which have been previously highlighted.[6] We were unable to individually validate VTE events which occurred in our study due to the terms of the QResearch licence which protects the anonymity of practices which contribute to the QResearch data and individual patients within these practices. Whilst the validation of the algorithm we used to define VTE events excluded pregnant women, we have ourselves conducted methodological work on classification of pregnancy-associated VTE events using electronic sources such as CPRD and QResearch, and found rates of VTE in and around pregnancy that were comparable with existing values obtained from a systematic review on this topic [23]. Nonetheless, we must consider the impact of any misclassification in our outcome event. Misclassification of VTE events both in the development and validation data would attenuate the effect of the predictor variables on VTE risk (assuming that misclassification was unrelated to the predictor variables) and thus bias conclusions towards claiming the maternity clot risk calculator has a weaker prognostic performance than it actually does.

A specific limitation of the present study, which in part affects our VTE algorithm, is that we did not have information on prescriptions emanating from secondary care and were unable to separate prophylactic from therapeutic doses of LMWH in primary care data. The former meant we were unable to account for women already

on thromboprophylaxis during and after childbirth. This is an acknowledged limitation, in general, of developing prognostic models using real world data to identify individuals who should receive a medical intervention. Whilst our subgroup analysis showed that the model performance did not differ noticeably between different time periods, QResearch covered more recent data (2004 onwards) compared to the data used for model development (1997 onwards) and may have downplayed the impact on some of the well-established risk factors due to better awareness of VTE risks. The inability to separate prophylactic from therapeutic doses was due to a combination of incomplete dose data and overlap in therapeutic and prophylactic doses of LMWH preparations as dose is determined from body weight. Therefore, we cannot rule out that primary care prescriptions picked up by our algorithm were for women receiving 6 weeks prophylaxis due to a previous VTE (incorrect inclusion in our study cohort). However, this would only have resulted if the VTE code for their previous event was not recorded in our data. Alternatively, some women were receiving VTE prophylaxis for other reasons during pregnancy and according to RCOG guidelines the same women would receive prophylaxis for 6 weeks post-delivery. If an unconfirmed VTE code was included post-delivery in this instance, this would be picked up by our algorithm (false positive VTE). Less than 1% of women in our cohort received any anticoagulation during the pregnancy itself, so the potential impact of this on our findings is likely to be minimal. Whilst all these specific limitations could be overcome through further validation in a prospective study which formally adjudicates VTE events, such studies are liable to be smaller and less representative of a maternal population than those which make use of administrative health data.

Further limitations include that more than 17% of women had missing values on their pre-pregnancy BMI and their baby's birth weight. This is an improvement from our previous CPRD study and we used multiple imputation technique to minimise the risk of bias associated with missing data. Second, both CPRD and QResearch use a similar coding system (Read code version 2). There is another computer system used in England to record patient consultations (SystmOne) that uses a slightly different coding system (clinical terminology version 3). It is possible that VTE events around pregnancy may be coded differently in practices using SystmOne and so our model performance may not generalise to these practices. However, these practices only represent a small proportion of all practices in England at the present time. Third, there was some miscalibration when applying the Maternity Clot Risk in the QResearch population; indicating some overestimation of risk for women with high-predicted values. However, in the sensitivity analyses the re-calibrated score produced very similar results to the main analyses, indicating that the potential miscalibration had very little impact on the overall predictive performance of the score. Finally, the score was developed for women without history of VTE therefore cannot be applied to women with a previous VTE or with a known high-risk hereditary thrombophilia. Any woman who has had a VTE previously would be considered high risk by the RCOG and receive thromboprophylaxis for at least 6 weeks, regardless of other risk factors. Routine testing for thrombophilia is not commonplace in the UK with many women being diagnosed after a blood clot has occurred. Therefore, whilst we acknowledge the inability of our model to make predictions based on this, we believe it has less relevance in the identification of intermediate risk women who would receive thromboprophylaxis for 10 days based on RCOG guidelines.

Cohort comparison

In this validation we were able to test our model using data from a higher number of deliveries using QResearch than those originally used to develop the model from the CPRD.[6] Whilst most of the baseline characteristics were broadly similar across both databases, some differences were observed. In particular, women in QResearch had lower incidence of pre-eclampsia/eclampsia, postpartum haemorrhage and higher mean BMI. Similarly, the overall rate of VTE during the first six weeks after childbirth was also slightly higher than in the CPRD cohort despite applying the same algorithm. These differences may reflect some variations in the study population between CPRD and QResearch. For example, there is evidence that practices contributing to CPRD are slightly more affluent and have lower all-cause mortality compared to the general population.[24] In contrast, due to a wider coverage, the QResearch population could better reflect the English population demographics. Alternatively, it may reflect variations in the recording of medical events across various regions. In addition, applying the current RCOG postpartum thromboprophylaxis guideline in QResearch identified fewer VTE events compared to it applied in CPRD. This may be due to the difference in the observation time period as QResearch used more recent data. Nevertheless, both QResearch and CPRD cohorts showed that in the UK more than 35% of women qualify for short to long term postpartum pharmacological thromboprophylaxis.

Conclusion and policy implications

We have carried out a second external validation of the Maternity Clot Risk. Overall, its predictive performance is consistent with its performance in the development CPRD population and is similar across subgroups relating to age, socioeconomic status, region, and calendar period. In addition, re-calibration of the score did not

improve its performance considerably. Therefore we recommend using the original score (Box 1).

The two algorithms (RCOG and Maternity Clot Risk) correctly predicted a similar number of VTE events, there was a slightly higher sensitivity with our risk score, which was not statistically different. However, the Maternity Clot Risk allows the flexibility of setting new treatment thresholds based on absolute predicted risks of VTE. If adopted it may help optimise use of LMWH to maximise health gain by better targeting of high-risk groups. In the UK, over 35% of women qualify for pharmacological thromboprophylaxis (based on the current RCOG postpartum thromboprophylaxis guideline) with a corresponding mean VTE risk of 1 in 769 (0.13%) postpartum women (based on Maternity Clot Risk applied in QResearch data). Assuming that low-molecular-weight heparin (LMWH) reduces the risk of VTE by at least 50% (based on trial data in ambulatory patients with cancer[25]), 1,538 postpartum women would require LMWH to prevent one VTE event. Increasing this risk threshold would result in a lower number needed to treat and would potentially be more acceptable to the women themselves.[26] For instance, targeting the highest 15% of the population (with a corresponding absolute VTE risk of 1 in 476 (0.21%)) would reduce the number of postpartum women requiring treatment to 952. Of course, any such recommendation will need to carefully take into account the perspective of the health care providers, practitioners and women and consider the potential benefits and harms of any threshold for which further research is urgently needed. Further validation of the model, especially in populations more ethnically diverse than those previously used to develop and externally validate the model, should be taken into consideration for use of the model in maternity settings worldwide. Finally, whilst we restricted our model to 6 weeks post-delivery as this is

the interval over which most postpartum VTE events occur, future work could consider which factors predict later maternal VTE events (beyond 6 weeks).

Acknowledgements

QResearch is a database of primary care records from practices, which use the Egton Medical Information System (EMIS) software. We acknowledge the role of EMIS in this study and all practices and patients which contribute to this. We also acknowledge that HES data used in this analysis are re-used with permission from NHS Digital, who retain the copyright. We thank Ruth Jack (University Of Nottingham) for helping us with access to data on the QResearch server.

Funding

This work was supported by a Senior Clinical Research Fellowship awarded to JW from the University of Nottingham and Nottingham University Hospitals NHS Trust. This was used to fund the cost of £13,500 for extracting and providing the data under the terms of the QResearch data licence. This specific funding source had no role in study design, data collection, data analysis, data interpretation, writing of the report, or in the decision to submit the article for publication.

Competing Interests

C N-P reports personal fees from Sanofi, other from Leo-Pharma, outside the submitted work; and is the lead developer of the RCOG Green Top Guideline on thromboprophylaxis in pregnancy (37a). AAS is currently an employee at Astra Zeneca. This study was conducted before he commenced his employment at Astra Zeneca and this research does not impact in any way on the role he currently fulfils. None of the other authors have interests which could inappropriately influence the work.

Author Contributions

AAS, JW, LJT and MJG were responsible for the conception and design of the study.

LB carried out the data management and modelling with guidance from AAS, JW, LJT, RDR and MJG. C N-P provided clinical (obstetric medicine) input to the study and ensured relevance of the project to the RCOG thromboprophylaxis guidelines. LB and AAS produced the first draft of the manuscript. All authors were responsible for critical evaluation of the manuscript and contributed to subsequent drafts. All authors accept responsibility for the paper as published.

REFERENCES

[1] Knight M, Bunch K, Tuffnell D, Jayakody H, Shakespeare J, Kotnis R, Kenyon S, Kurinczuk JJ (Eds.), on behalf of MBBRACE-UK. Saving Lives, Improving Mothers' Care - Lessons learned to inform maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2014-16, Oxford, 2018.

[2] I.A. Greer, Pregnancy complicated by venous thrombosis, N. Engl. J. Med. 373 (2015) 540–547. https://doi.org/10.1056/NEJMcp1407434.

[3] A.A. Sultan, J. West, L.J. Tata, K.M. Fleming, C. Nelson-Piercy, M.J. Grainge, Risk of first venous thromboembolism in and around pregnancy: A population-based cohort study, Br. J. Haematol. 156 (2012) 366–373. https://doi.org/10.1111/j.1365-2141.2011.08956.x.

[4] A.A. Sultan, M.J. Grainge, J. West, K.M. Fleming, C. Nelson-Piercy, L.J. Tata, Impact of risk factors on the timing of first postpartum venous thromboembolism: a population-based cohort study from England., Blood. 124 (2014) 2872–2880. https://doi.org/10.1182/blood-2014-05-572834.

[5] H. Kamel, B.B. Navi, N. Sriram, D.A. Hovsepian, R.B. Devereux, M.S.V. Elkind, Risk of a Thrombotic Event after the 6-Week Postpartum Period, N. Engl. J. Med. 370 (2014) 1307–1315. https://doi.org/10.1056/NEJMoa1311485.

[6] A.A. Sultan, J. West, M.J. Grainge, R.D. Riley, L.J. Tata, O. Stephansson, K.M. Fleming, C. Nelson-Piercy, J.F. Ludvigsson, Development and validation of risk prediction model for venous thromboembolism in postpartum women: Multinational cohort study, BMJ. 355 (2016) i6253.

https://doi.org/10.1136/bmj.i6253.

[7] Royal College of Obstetricians and Gynecologists, Reducing the Risk of Venous Thromboembolism during Pregnancy and the Puerperium (Green-top Guideline No. 37a), London, 2015. https://www.rcog.org.uk/globalassets/documents/guidelines/gtg-37a.pdf.

[8] P.G. Lindqvist, M. Hellgren, Obstetric thromboprophylaxis: The Swedish guidelines, Adv. Hematol. 2011 (2011) Article ID 157483. https://doi.org/10.1155/2011/157483.

[9] J.F. Ludvigsson, E. Andersson, A. Ekbom, M. Feychting, J.L. Kim, C. Reuterwall, M. Heurgren, P.O. Olausson, External review and validation of the Swedish national inpatient register, BMC Public Health. 11 (2011) 450. https://doi.org/10.1186/1471-2458-11-450.

[10] R. Riley, D. van der Windt, P. Croft, K.G.M. Moons, Prognosis Research in Health Care: Concepts, Methods, and Impact, Oxford University Press, Oxford, 2019.

[11] T.P.A. Debray, Y. Vergouwe, H. Koffijberg, D. Nieboer, E.W. Steyerberg, K.G.M. Moons, A new framework to enhance the interpretation of external validation studies of clinical prediction models, J. Clin. Epidemiol. 68 (2015) 279–289. https://doi.org/10.1016/j.jclinepi.2014.06.018.

[12] E. Herrett, A.M. Gallagher, K. Bhaskaran, H. Forbes, R. Mathur, T. van Staa, L. Smeeth, Data Resource Profile: Clinical Practice Research Datalink (CPRD), Int. J. Epidemiol. 44 (2015) 827–836. https://doi.org/10.1093/ije/dyv098.

[13] E. Kontopantelis, R.J. Stevens, P.J. Helms, D. Edwards, T. Doran, D.M.

Ashcroft, Spatial distribution of clinical computer systems in primary care in England in 2016 and implications for primary care electronic medical record databases: A cross-sectional population study, BMJ Open. 8 (2018) 1–7. https://doi.org/10.1136/bmjopen-2017-020738.

[14] R.D. Riley, J. Ensor, K.I.E. Snell, T.P.A. Debray, D.G. Altman, K.G.M. Moons, G.S. Collins, External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: Opportunities and challenges, BMJ. 353 (2016) 27–30. https://doi.org/10.1136/bmj.i3140.

[15] J. Hippisley-Cox, C. Coupland, Development and validation of risk prediction algorithm (QThrombosis) to estimate future risk of venous thromboembolism: Prospective cohort study, BMJ. 343 (2011) 1–12. https://doi.org/10.1136/bmj.d4656.

[16] J. Hippisley-Cox, C. Coupland, Development and validation of risk prediction equations to estimate survival in patients with colorectal cancer: Cohort study, BMJ. 357 (2017) 1–25. https://doi.org/10.1136/bmj.j2497.

[17] J. Hippisley-Cox, C. Coupland, Development and validation of QMortality risk prediction algorithm to estimate short term risk of death and assess frailty: cohort study, BMJ. 358 (2017) j4208. https://doi.org/10.1136/bmj.j4208.

[18] Office for National Statistics, Birth characteristics in England and Wales: 2019, Stat. Bull. 2021 (2019). https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthcharacteristicsinenglandandwales/2019 (accessed February 27, 2021).

[19] R. Lawrenson, J.C. Todd, G.M. Leydon, T.J. Williams, R.D. Farmer, Validation

of the diagnosis of venous thromboembolism in general practice database studies., Br. J. Clin. Pharmacol. 49 (2000) 591–6. https://doi.org/10.1046/j.1365-2125.2000.00199.x.

[20] G.S. Collins, E.O. Ogundimu, D.G. Altman, Sample size considerations for the external validation of a multivariable prognostic model: A resampling study, Stat. Med. 35 (2016) 214–226. https://doi.org/10.1002/sim.6787.

[21] K.G.M. Moons, D.G. Altman, J.B. Reitsma, J.P.A. Ioannidis, P. Macaskill, E.W. Steyerberg, A.J. Vickers, D.F. Ransohoff, G.S. Collins, Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration., Ann. Intern. Med. 162 (2015) W1-73. https://doi.org/10.7326/M14-0698.

[22] Office for National Statistics, Ethnicity and National Identity in England and Wales: 2011, (2011). https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/ethnicityandnationalidentityinenglandandwales/2012-12-11#ethnicity-in-england-and-wales (accessed November 20, 2020).

[23] A. Abdul Sultan, L.J. Tata, M.J. Grainge, J. West, 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. 8 (2013). https://doi.org/10.1371/journal.pone.0070310.

[24] A. Maguire, B.T. Blak, M. Thompson, The importance of defining periods of complete mortality reporting for research using automated data from primary care., Pharmacoepidemiol. Drug Saf. 18 (2009) 76–83.

https://doi.org/10.1002/pds.1688.

[25] E.A. Akl, L.A. Kahale, R.A. Ballout, M. Barba, V.E.D. Yosuico, F.F. van Doormaal, S. Middeldorp, A. Bryant, H. Schünemann, Parenteral anticoagulation in ambulatory patients with cancer., Cochrane Database Syst. Rev. (2014) CD006652. https://doi.org/10.1002/14651858.CD006652.pub4.

[26] N. Gassmann, M. Viviano, M. Righini, P. Fontana, B. Martinez de Tejada, M. Blondon, Estimating the risk thresholds used by guidelines to recommend postpartum thromboprophylaxis., J. Thromb. Haemost. 19 (2021) 452–459. https://doi.org/10.1111/jth.15166.

External validation of a model to predict women most at risk of postpartum venous thromboembolism: Maternity clot risk

Supplementary material

Appendix A: Description of statistical methods for model evaluation

Discrimination

Discrimination is the ability of the score to differentiate between women who did and did not develop a first postpartum VTE event, and was examined by calculating the C-statistic (where 0.50 represents no discrimination and 1.00 represents perfect discrimination).[1] It can be interpreted as the probability that the score assigns a higher predicted risk to a randomly selected woman with VTE than it does to a randomly selected woman without a VTE. For example, the C-statistic 0.7 means that if we randomly select two women, 1 with a VTE and 1 without a VTE, there is a 70% probability that the model assigns a higher predicted risk to the former. We used Rubin's rule to combine the multiple imputed estimates to get the overall C-statistic.[2]

Calibration

Calibration refers to how closely the predicted postpartum VTE risk agrees with the observed risk. Calibration was examined by plotting agreement between predicted and observed risks across the entire spectrum of predicted risks using loess smooth curves, and also within tenths of predicted risk. Calibration was also measured through calculation of calibration slope and calibration-in-the-large.[1] The calibration slope (ideal value of 1) gives an indication of the degree of overfitting, and is obtained by regressing the outcome on the predicted probabilities for each woman. Calibration-in-the-large (ideal value of 0), compares the mean observed risk with the mean of the predicted risks, thus estimating the extent to which a model systematically over or under-estimates VTE risk. Again, Rubin's rule was used to obtain an overall estimate of the calibration slope across the imputed datasets.

Subgroup analyses

The predictive performance was evaluated for various subgroups in terms of the C-statistic and calibration slope. Forest plots were presented to display point estimates with 95% confidence intervals (95% CI) from a random effects meta-analysis. I^2 was used to quantify the amount of variability that is due to between-group heterogeneity rather than sampling error. Pooled values of the C-statistic and calibration slope across all imputed datasets within each subgroup category were calculated using the same method described above.

Comparison with RCOG algorithm

To compare the performance of Maternity Clot Risk in QResearch versus the existing RCOG postpartum thromboprophylaxis guideline, we applied the RCOG guideline to our cohort to assess the number of women who qualified for pharmacological thromboprophylaxis based on prescribed risk factors recorded in the data. We used the Maternity Clot Risk to identify the same proportion of women who should receive VTE prophylaxis as the RCOG guideline and calculated sensitivity (percent of women with VTE above the risk threshold), specificity (percent of women without VTE below the risk threshold), and positive predictive value (the percent of women above the risk threshold who develop VTE). This analysis assumes that the threshold is set so that same percentage of women receive prophylaxis as under current RCOG guidance. The subsequent decision curve analysis explores the effect of varying this threshold. Analyses were performed separately for each imputed dataset with no attempt to pool results (which was also the case for the decision curve analysis).

Decision curve analysis

Decision curves assess the consequences of applying a test or treatment in practice based on a risk threshold.[3] Intervention takes place when the probability of an event from a prediction model exceeds the risk threshold. The net benefit is calculated from the sensitivity and specificity of the model at each threshold probability, and calculated from the formula

net benefit = sensitivity × prevalence – (1 – specificity) × (1 – prevalence) × odds at threshold probability

where odds = threshold/(1-threshold) and prevalence is the probability of a VTE before the prediction model is applied.

Our decision curve analysis compares use of the maternity clot risk calculator against three alternative strategies, i) treat all, ii) treat no one, iii) treat according to RCOG algorithm; it will provide a visual display of the range of threshold probabilities for which each strategy is superior to all others (highest net benefit). The optimum threshold is not determined from the data but on how healthcare users and providers weigh up the benefits of intervention (averted VTE) against the costs of intervention (tolerability of LMWH and financial cost). A positive net benefit indicates that use of the strategy is preferred to the baseline scenario of treating no one.

Model re-calibration

Re-calibration uses the original model coefficients but applies a shrinkage factor to account for any over optimism whilst developing the model (so that the calibration slope is forced to be 1). In our case, the original predictor coefficients estimated in the development study [4] were shrunk by the calibration slope estimated in the current study multiplied by the original shrinking factor (0.94). We also re-estimated the intercept which enables more accurate estimation of predicted probabilities in

populations with a different baseline risk. In a sensitivity analysis, we compared the performance of the re-calibrated model with the original score.

References (Appendix A)

1. K.G.M. Moons, D.G. Altman, J.B. Reitsma, J.P.A. Ioannidis, P. Macaskill, E.W. Steyerberg, A.J. Vickers, D.F. Ransohoff, G.S. Collins, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration, Ann. Intern. Med. 162 (2015) W1–W73. https://doi.org/10.7326/M14-0698.
2. D. Rubin, Multiple imputation for nonresponse in surveys., John Wiley & Sons, Inc., Hoboken, NJ, 1987.
3. A.J. Vickers, B. Van Calster, E.W. Steyerberg, Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests, BMJ. 352 (2016). https://doi.org/10.1136/bmj.i6.
4. A.A. Sultan, J. West, M.J. Grainge, R.D. Riley, L.J. Tata, O. Stephansson, K.M. Fleming, C. Nelson-Piercy, J.F. Ludvigsson, Development and validation of risk prediction model for venous thromboembolism in postpartum women: Multinational cohort study, BMJ. 355 (2016) i6253. https://doi.org/10.1136/bmj.i6253.

Table S1 Algorithms used to define predictors in QResearch and CPRD

Variables	QResearch	CPRD
Social and demographic factors:		
Age	Age at delivery	Age at delivery
Body mass index (BMI), kg/m^2	Latest BMI before pregnancy	Latest BMI before pregnancy
Smoker	Latest record before delivery	Latest record before delivery
Comorbidities ever before delivery:		
Varicose veins	Diagnosis ever before delivery, identified from either primary or secondary care data	Diagnosis ever before delivery, identified from either primary or secondary care data
Heart disease	Diagnosis ever before delivery, identified from either primary or secondary care data	Diagnosis ever before delivery, identified from either primary or secondary care data
Kidney disease	Diagnosis ever before delivery, identified from either primary or secondary care data	Diagnosis ever before delivery, identified from either primary or secondary care data
IBD	Diagnosis ever before delivery, identified from primary care data	Diagnosis ever before delivery, identified from primary care data
Pregnancy complications:		
Pre-eclampsia/eclampsia	Diagnosis between 30 days after conception and 30 days after delivery, identified from secondary care data	Diagnosis between 30 days after conception and 30 days after delivery, identified from secondary care data
Diabetes (including both pre-existing and gestational diabetes)	Either a prescription of anti-diabetic drugs or a primary or secondary care medical diagnosis	Either a prescription of anti-diabetic drugs or a primary or secondary care medical diagnosis

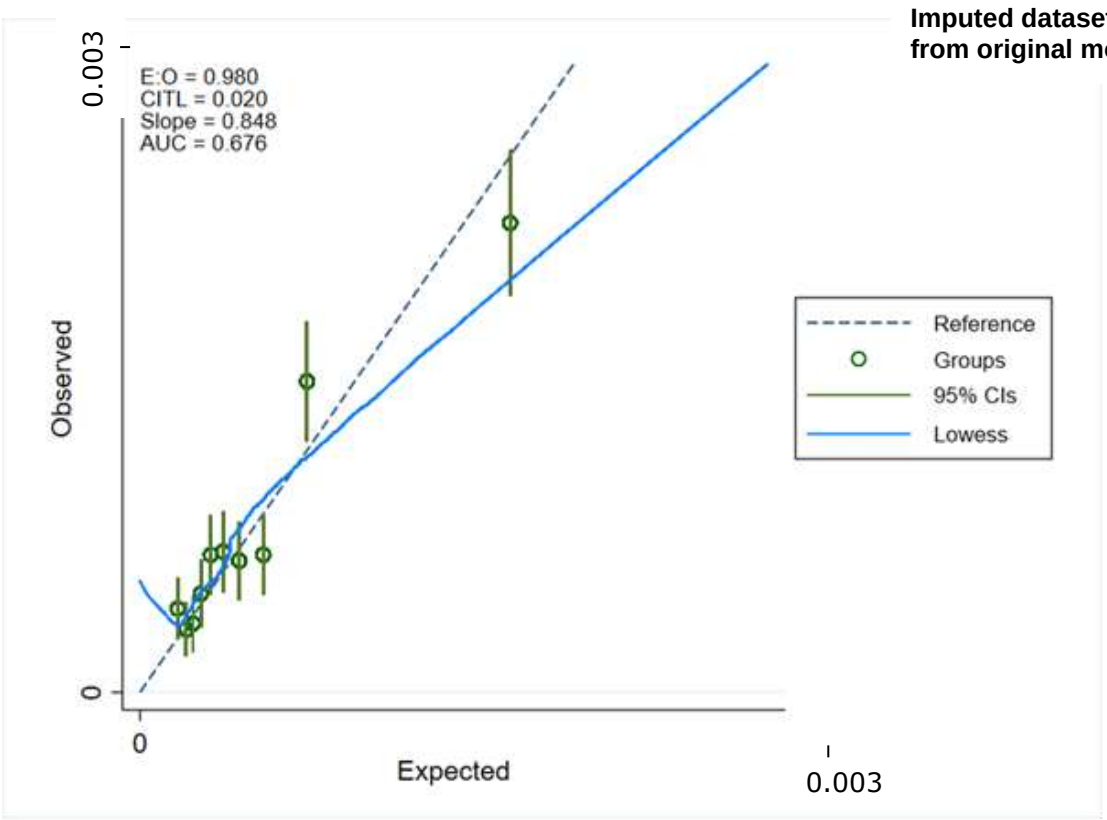
	code ever before delivery.	code diagnosis ever before delivery.
Hypertension (including both pre-existing and gestational hypertension)	Prescription of antihypertensive drugs or diagnosis ever before delivery, identified from both primary and secondary care data	Prescription of antihypertensive drugs or diagnosis ever before delivery, identified from both primary and secondary care data
Antenatal parity	Number of previous deliveries (not including current pregnancy), identified from HES maternity data	Number of previous deliveries (not including current pregnancy), identified from both HES maternity and mother-baby link data
Delivery characteristics/complications:		
Preterm birth	Birth before 37 weeks gestational age, identified from HES maternity data	Birth before 37 weeks gestational age, identified from HES maternity data
Postpartum haemorrhage	Diagnosis from secondary care data	Diagnosis from secondary care data
Delivery method	Identified using procedure codes from secondary care	Identified using procedure codes from secondary care
Multiple delivery (twins or more)	Identified using both procedure codes from secondary care and HES maternity data	Identified using procedure codes from secondary care and HES maternity data
Stillbirth	Diagnosis from secondary care data	Diagnosis from secondary care data
Puerperal acute infection	Diagnosis of respiratory or urinary tract infections, identified from both primary and secondary care data	Diagnosis of respiratory or urinary tract infections, identified from both primary and secondary care data

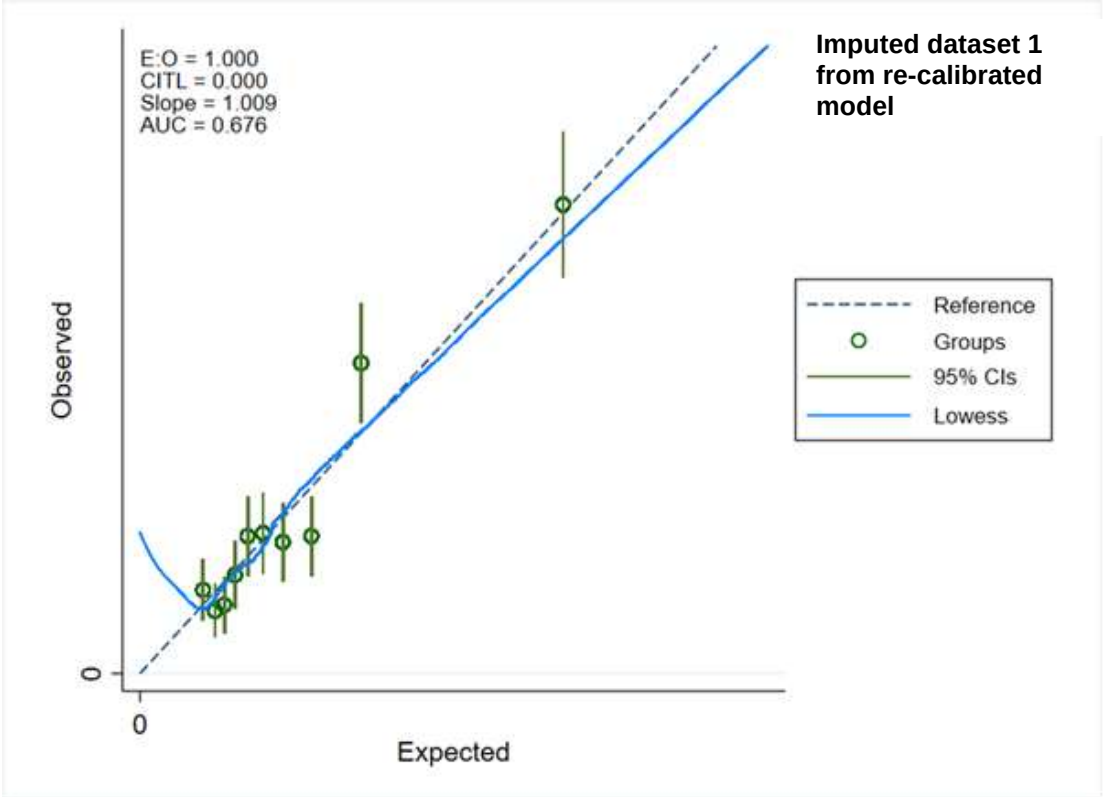
Infant’s birth weight, g	Identified from HES maternity data (the smallest birth weight used for multiple deliveries)	Identified from HES maternity data (the smallest birth weight used for multiple deliveries)

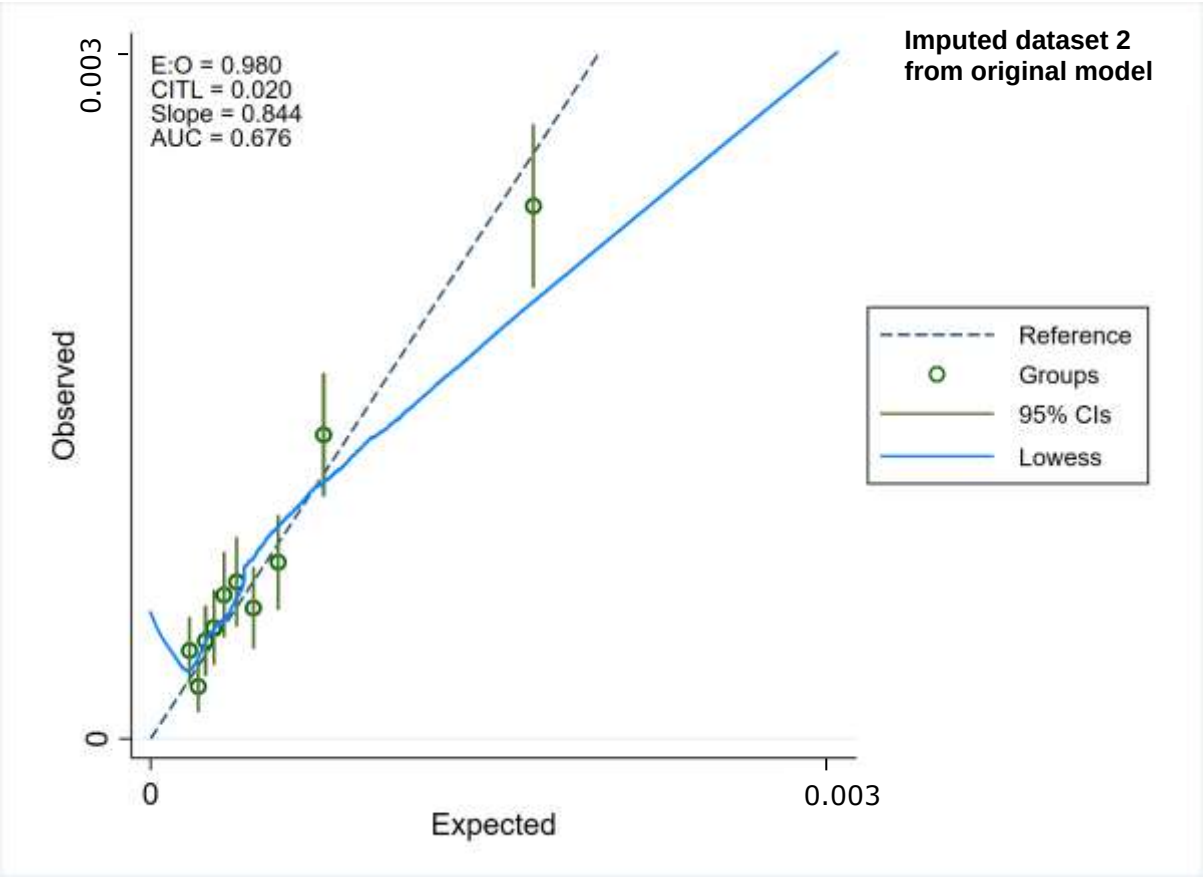
Table S2 Model performance (with 95% confidence interval) in each imputed dataset

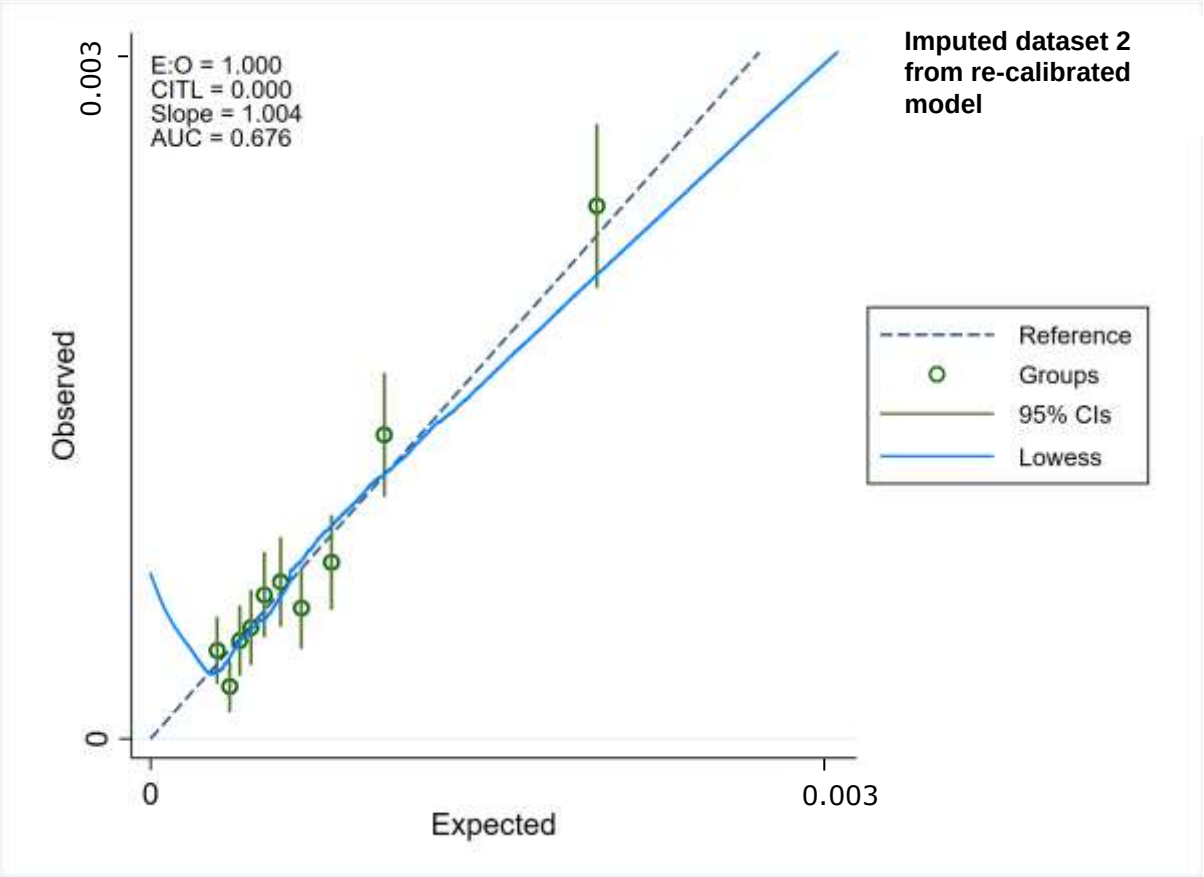
Imputed dataset 1	
C statistic	0.68 (0.65-0.70)
Calibration slope	0.85 (0.75-0.94)
After re-calibration:	
Calibration slope	1.01 (0.89-1.12)
Imputed dataset 2	
C statistic	0.68 (0.65-0.70)
Calibration slope	0.84 (0.75-0.94)
After re-calibration:	
Calibration slope	1.00 (0.89-1.12)
Imputed dataset 3	
C statistic	0.67 (0.65-0.70)
Calibration slope	0.83 (0.73-0.93)
After re-calibration:	
Calibration slope	0.99 (0.87-1.11)
Imputed dataset 4	
C statistic	0.68 (0.65-0.70)
Calibration slope	0.84 (0.74-0.94)
After re-calibration:	
Calibration slope	1.00 (0.89-1.12)
Imputed dataset 5	
C statistic	0.67 (0.65-0.70)
Calibration slope	0.84 (0.74-0.93)
After re-calibration:	
Calibration slope	1.00 (0.88-1.11)

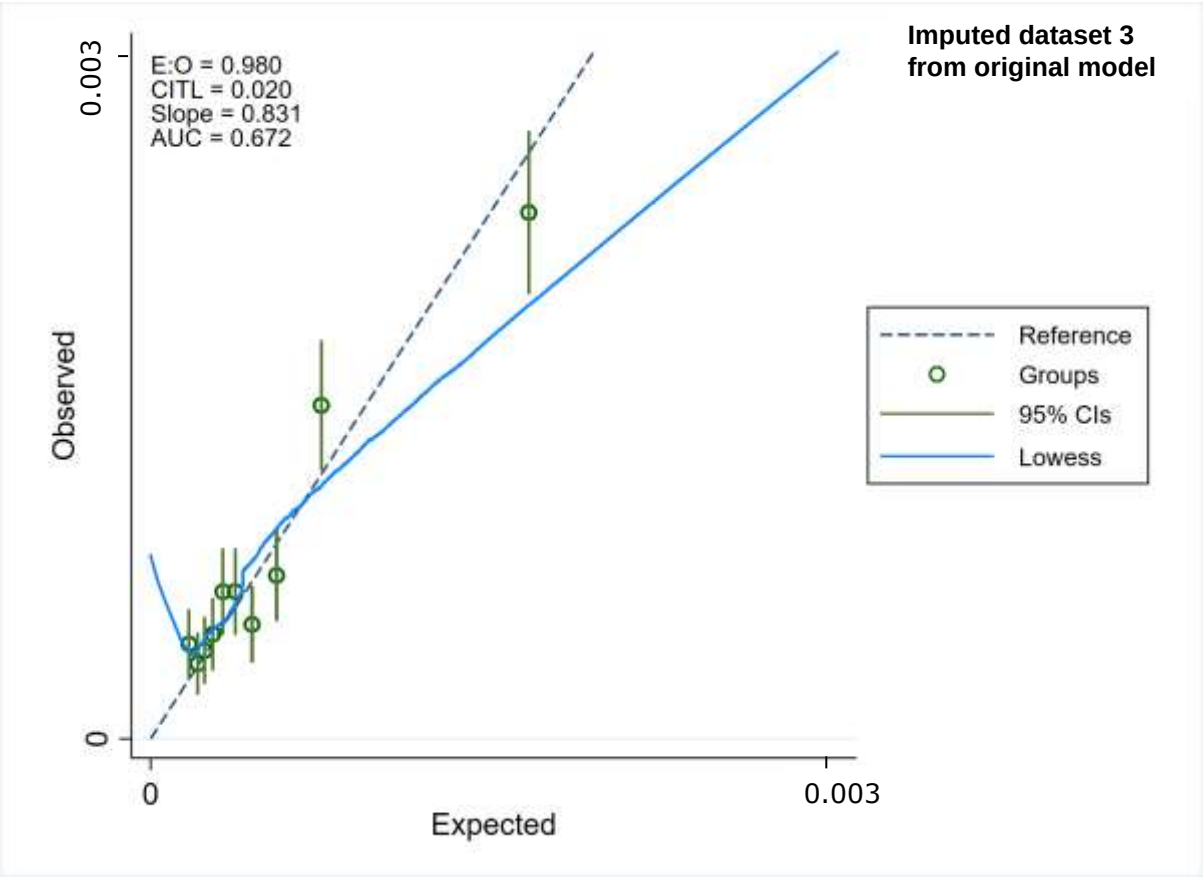
Figure S1 Calibration of the expected (predicted) risks from the original model and the observed risks across tenths of predicted risk with 95% CI and Lowess smoothing for each imputed dataset

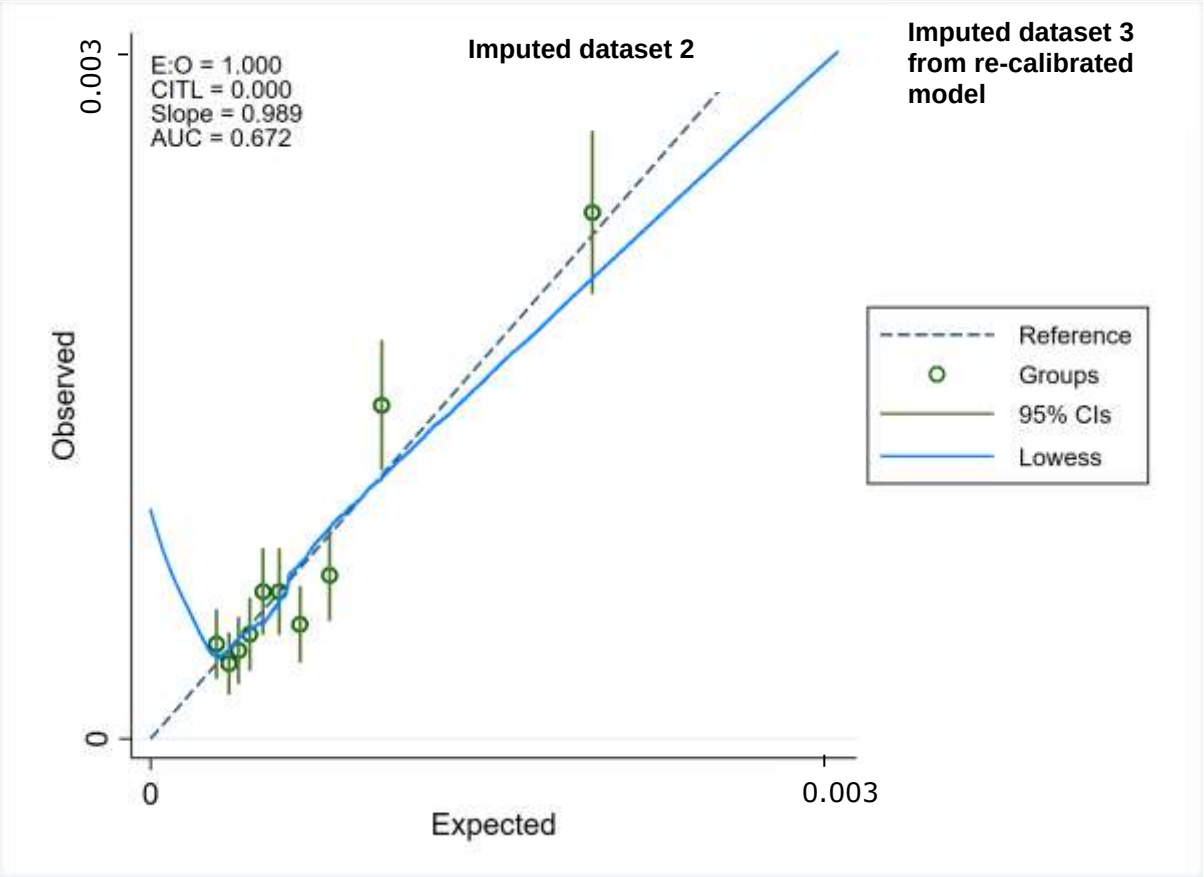


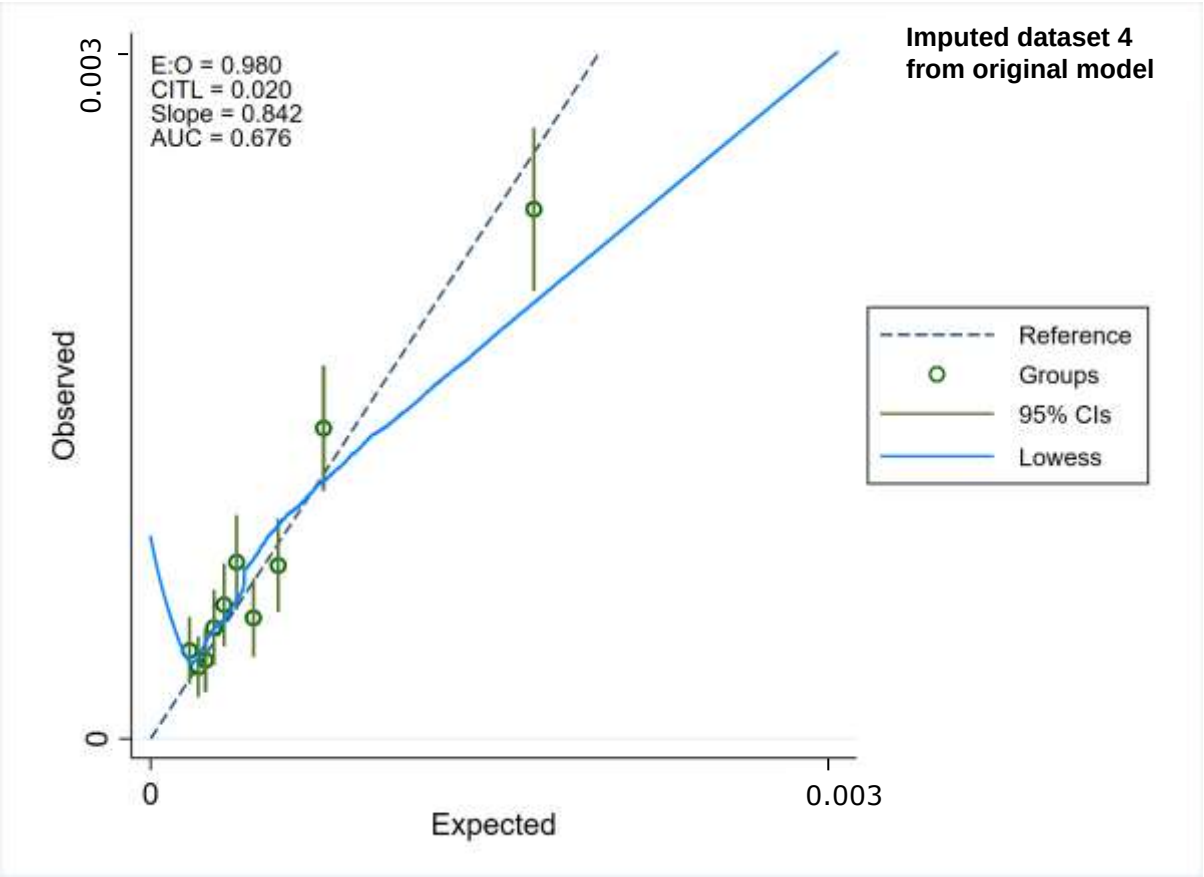



Imputed dataset 2
from original model
E:O = 0.980
CITL = 0.020
Slope = 0.844
AUC = 0.676
0.003
0
Observed
0
0.003
Expected
Reference
Groups
95% CIs
Lowess

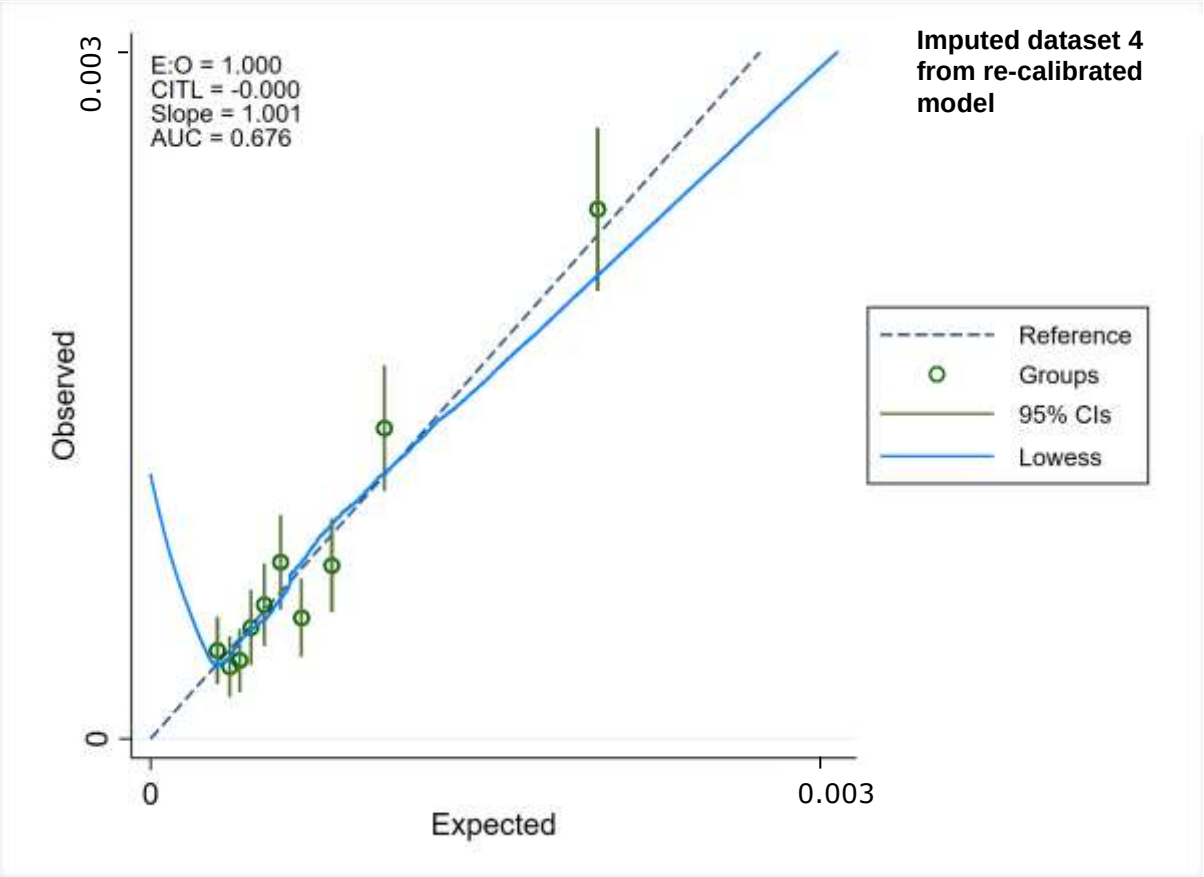

Imputed dataset 2
from re-calibrated
model
E:O = 1.000
CITL = 0.000
Slope = 1.004
AUC = 0.676
0.003
0
Observed
0
0.003
Expected
Reference
Groups
95% CIs
Lowess


Imputed dataset 3
from original model
E:O = 0.980
CITL = 0.020
Slope = 0.831
AUC = 0.672
0.003
0
Observed
0
0.003
Expected
Reference
Groups
95% CIs
Lowess

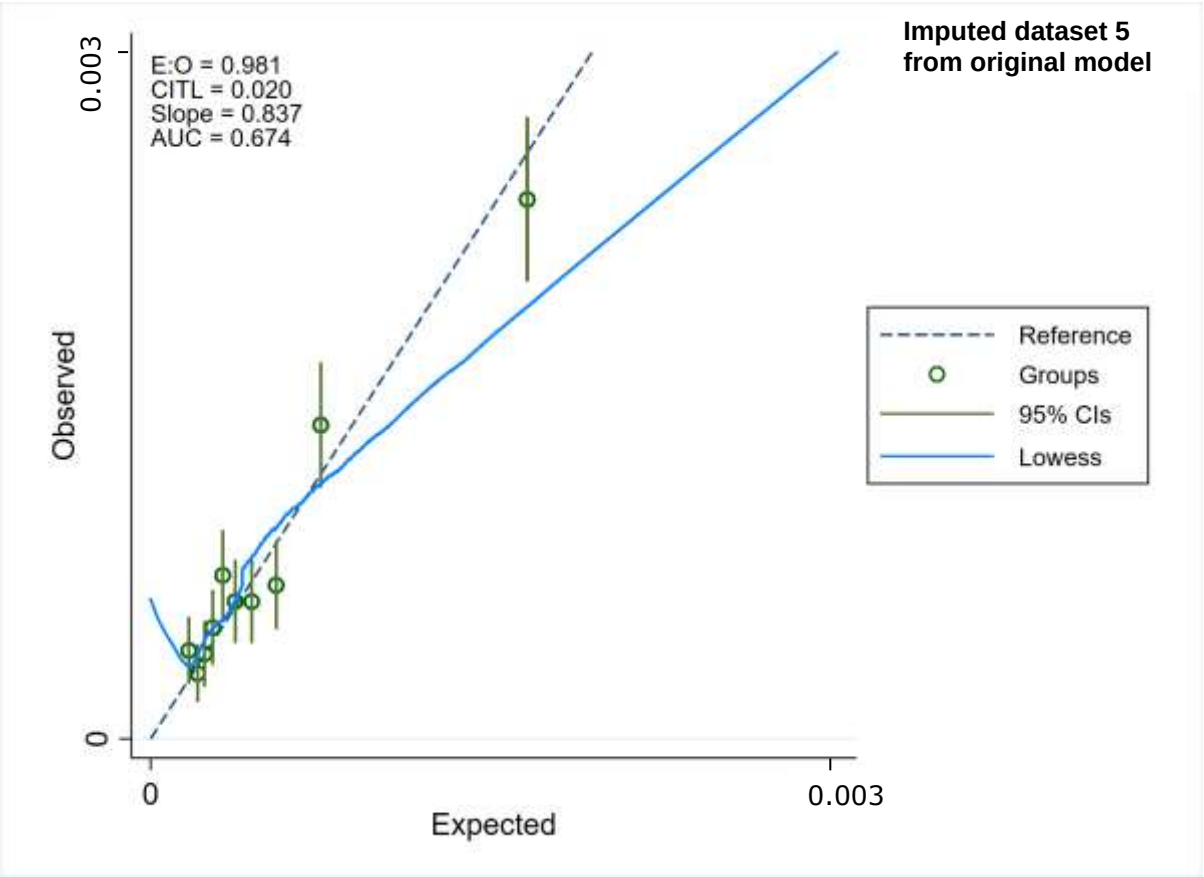

Imputed dataset 2
Imputed dataset 3
from re-calibrated
model
E:O = 1.000
CITL = 0.000
Slope = 0.989
AUC = 0.672
0.003
0
Observed
0
0.003
Expected
Reference
Groups
95% CIs
Lowess



Imputed dataset 4 from original model



Imputed dataset 4 from re-calibrated model


Imputed dataset 5
from original model
E:O = 0.981
CITL = 0.020
Slope = 0.837
AUC = 0.674
0.003
0
Observed
0
0.003
Expected
Reference
Groups
95% CIs
Lowess

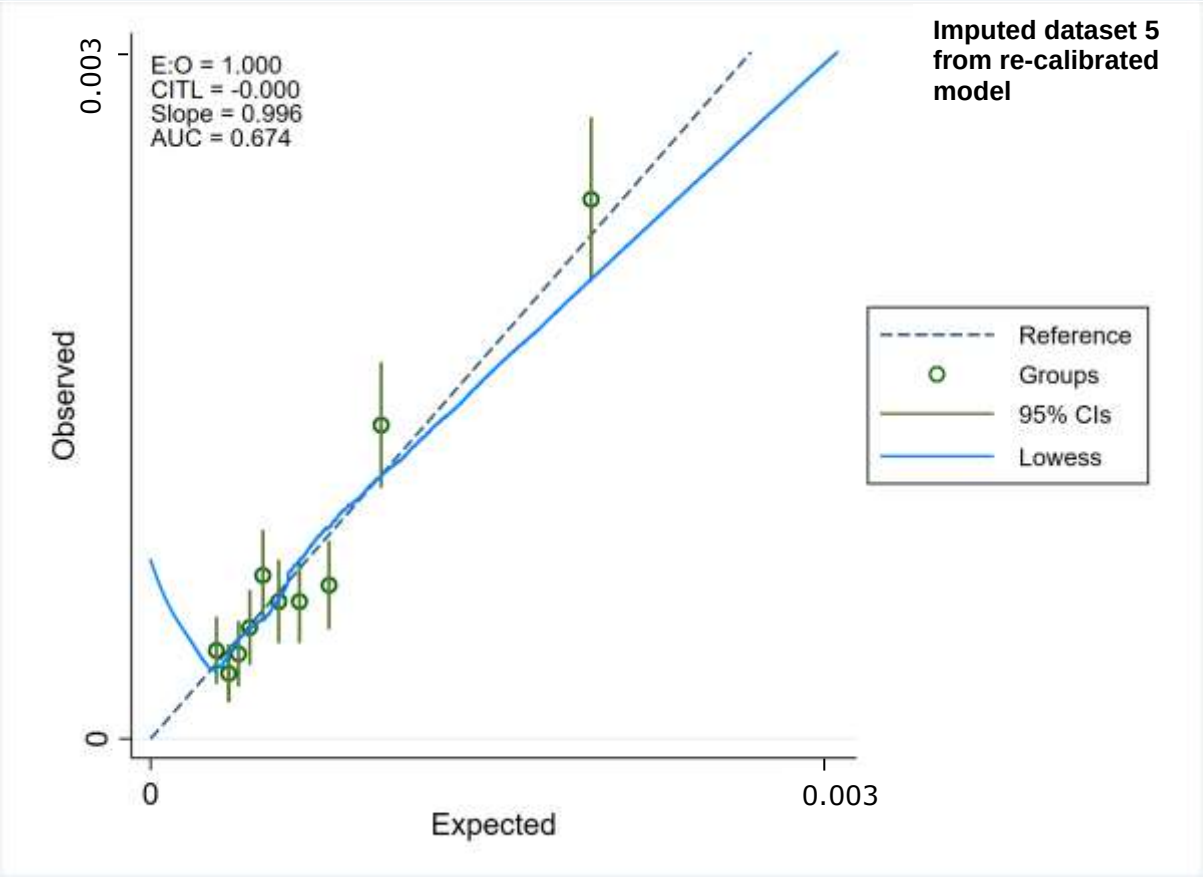
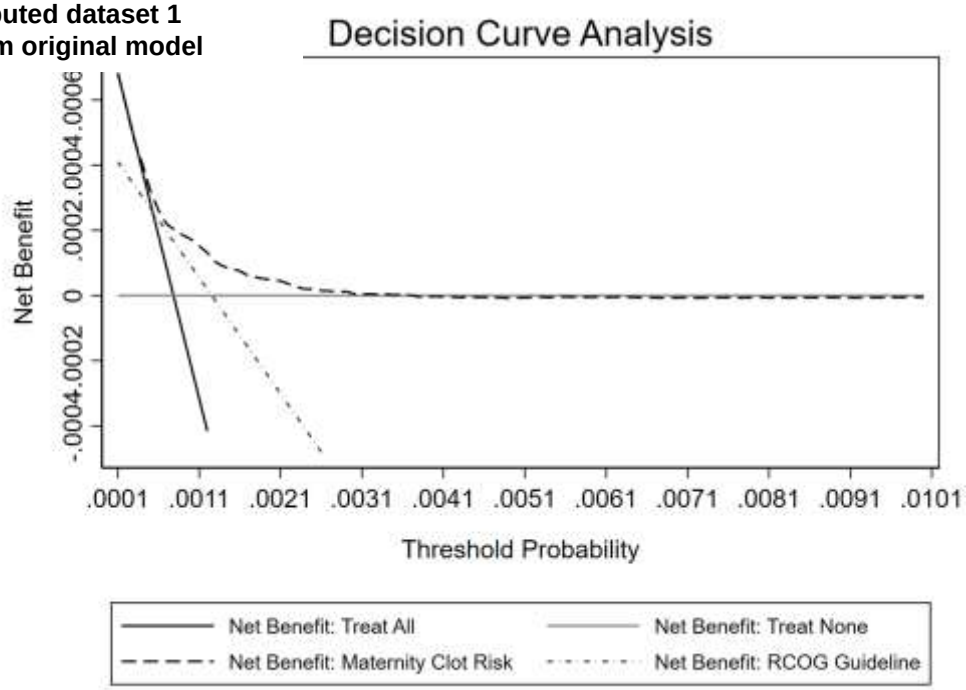
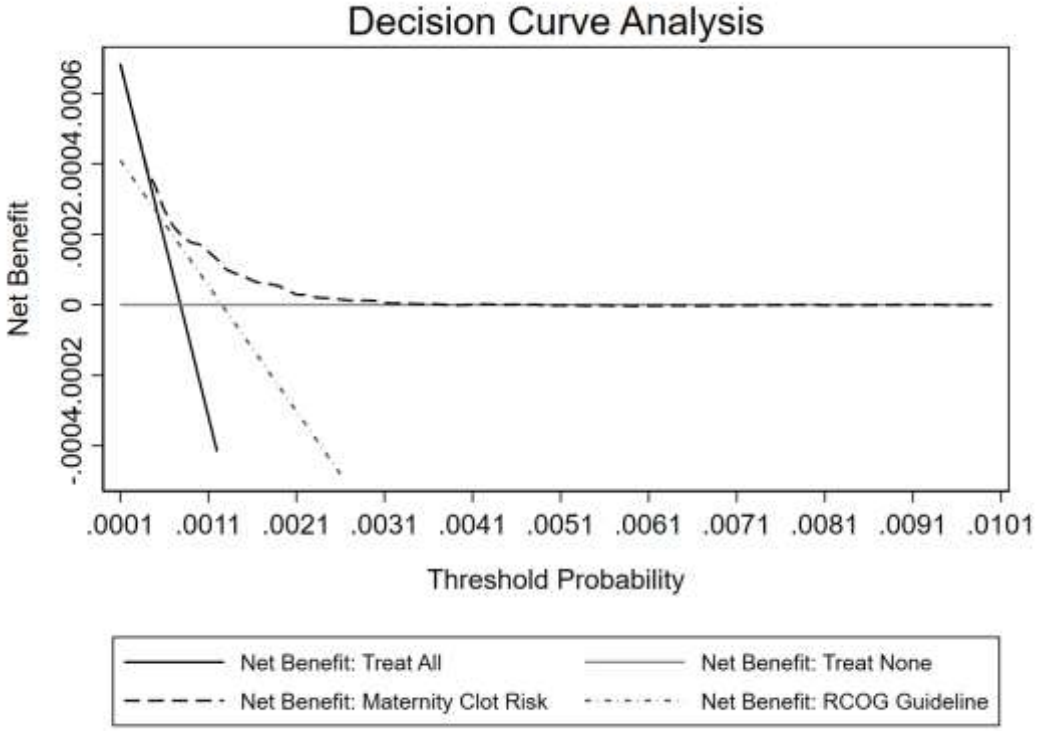

Imputed dataset 5
from re-calibrated
model
E:O = 1.000
CITL = -0.000
Slope = 0.996
AUC = 0.674
0.003
0
Observed
0
0.003
Expected
Reference
Groups
95% CIs
Lowess

Figure S2 Decision curve analysis using the postpartum VTE risk score. The analysis was repeated within each imputed dataset both from the original model and after re-calibration.

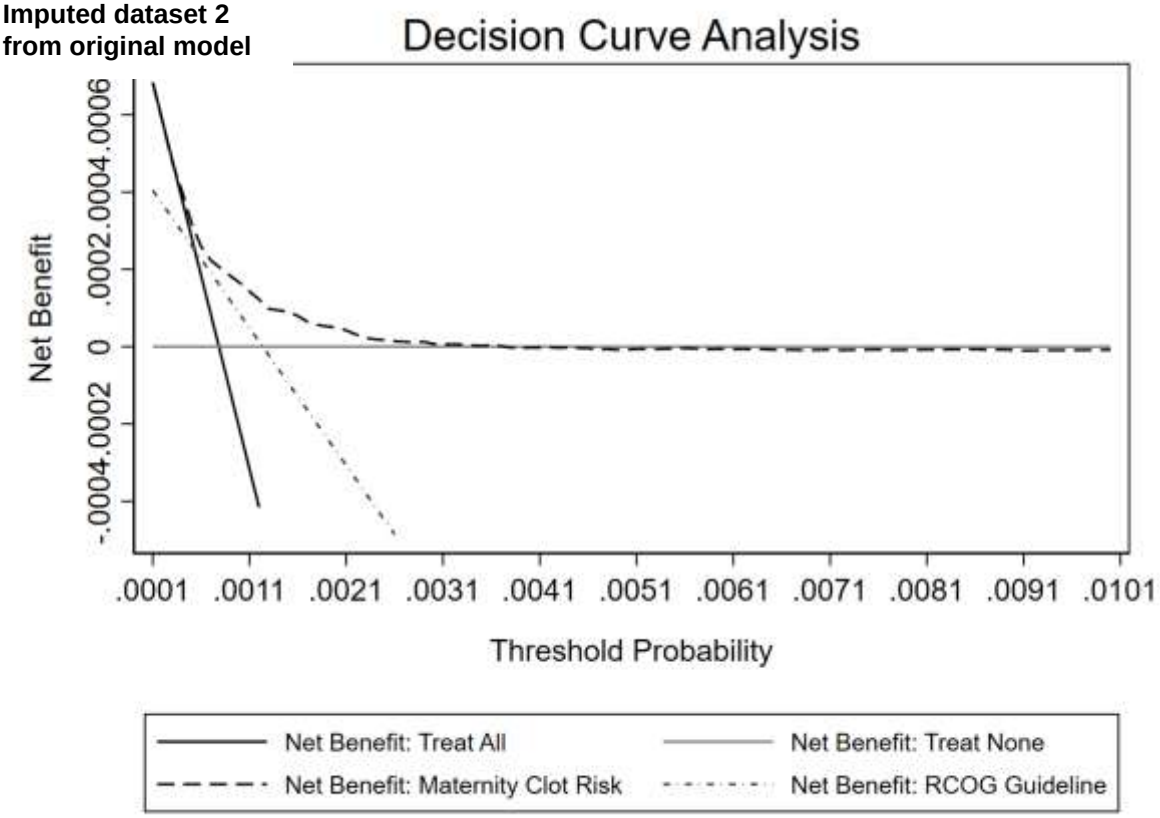
Imputed dataset 1 from original model



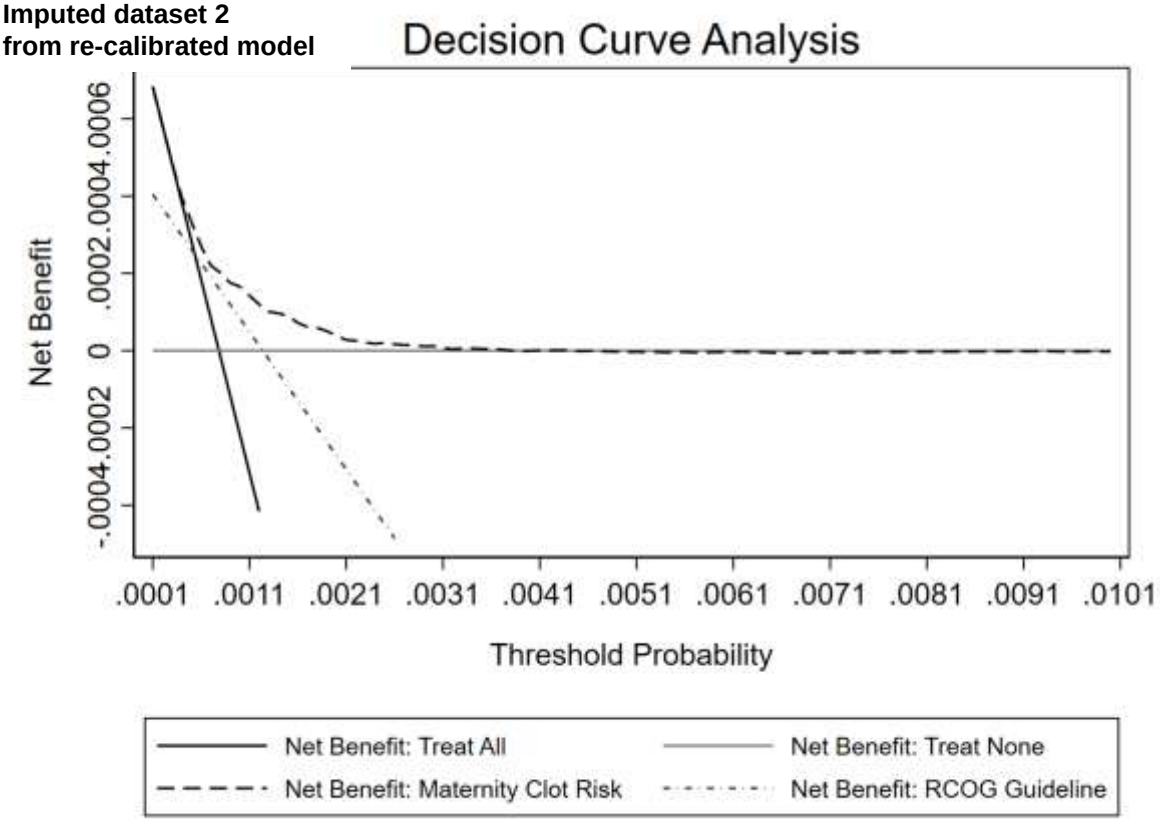
Imputed dataset 1 from re-calibrated model



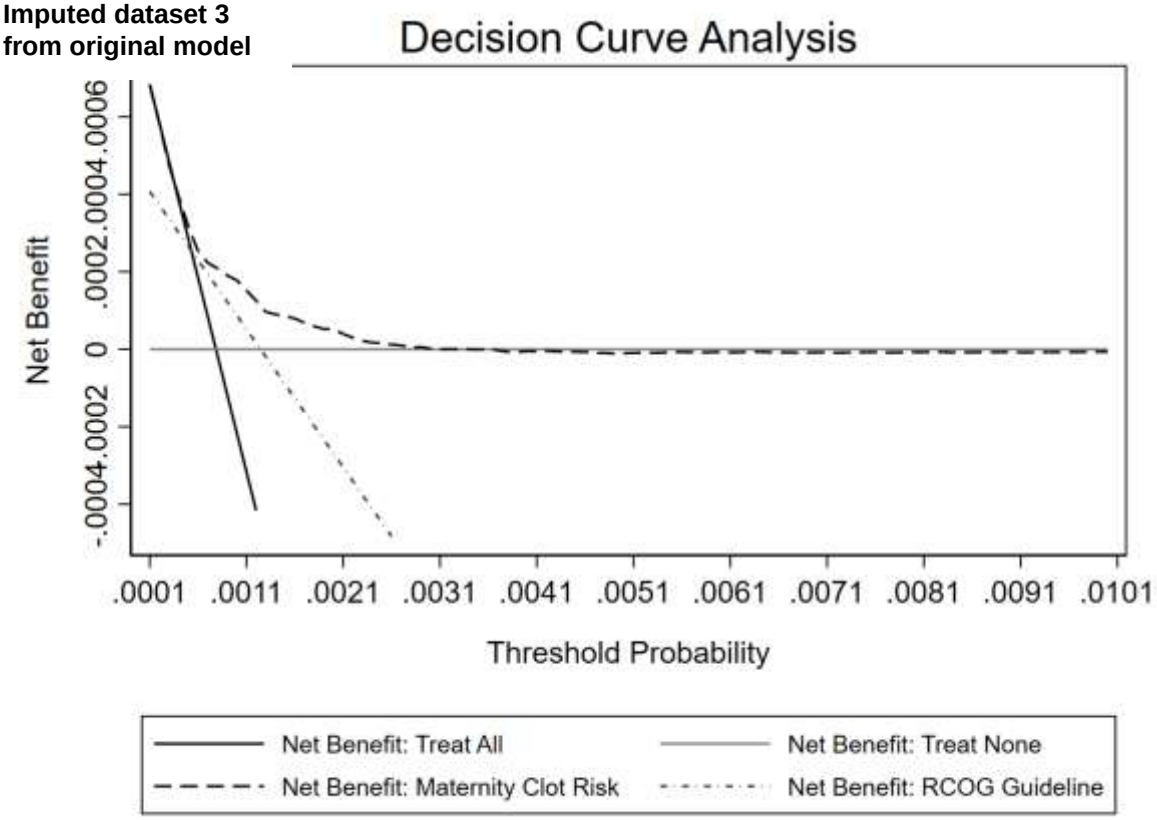
Imputed dataset 2 from original model



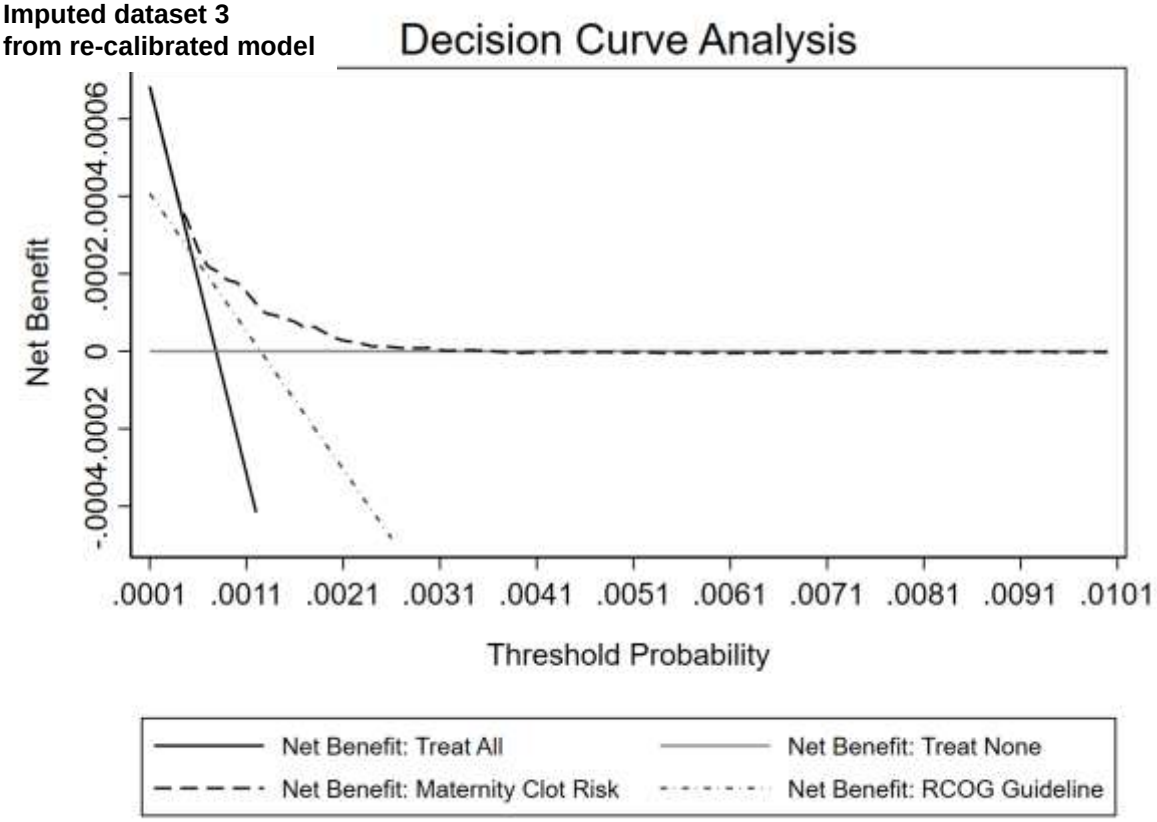
Imputed dataset 2 from re-calibrated model



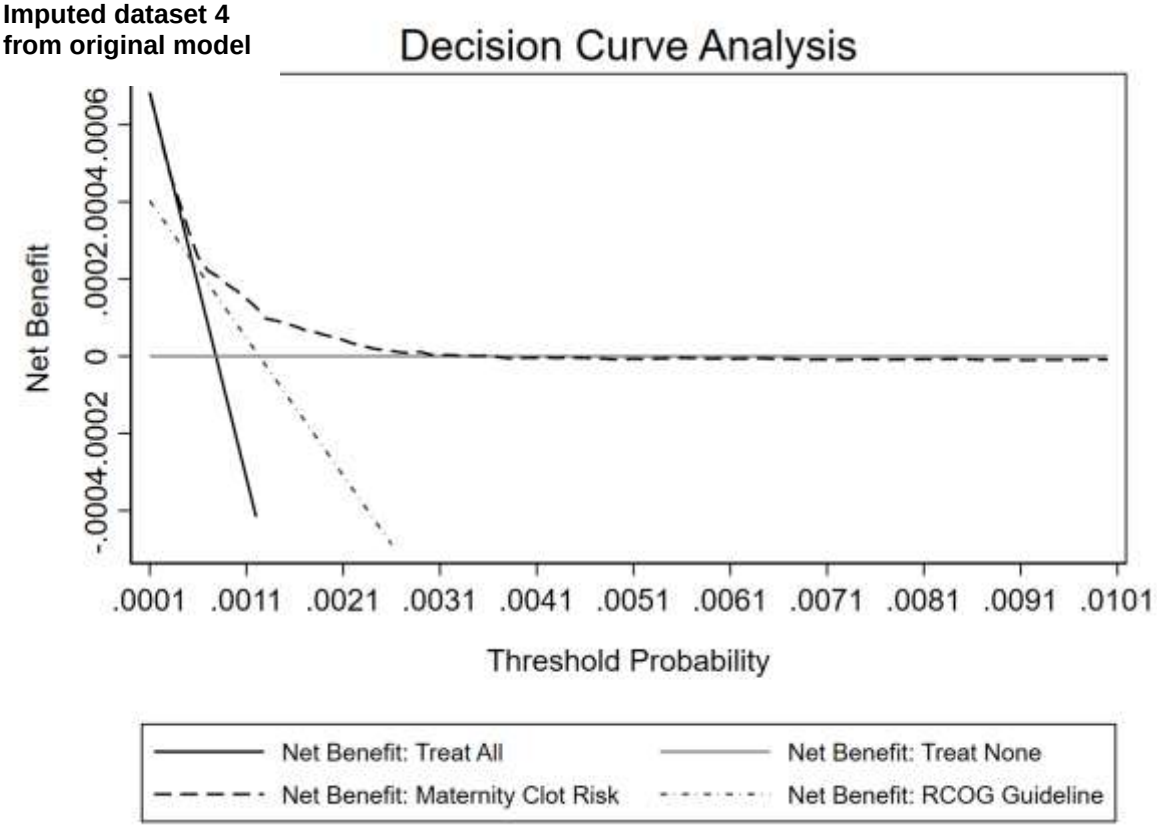
Imputed dataset 3 from original model



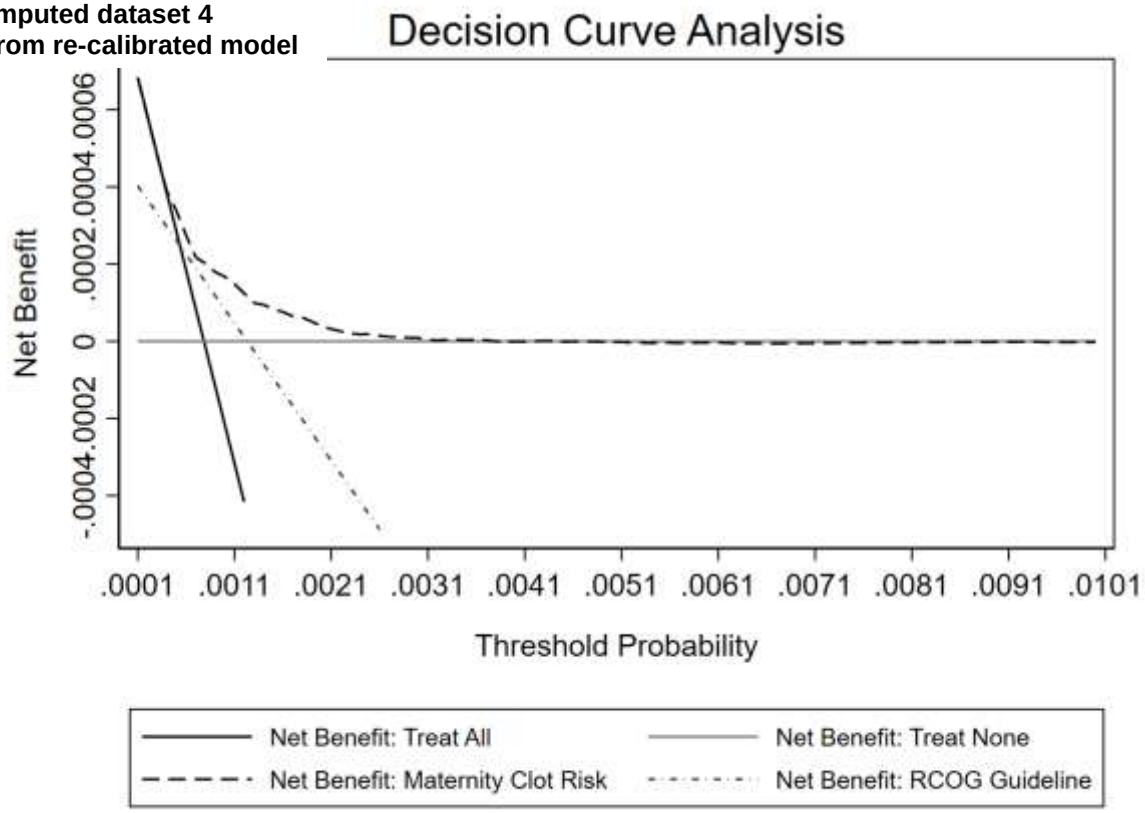
Imputed dataset 3 from re-calibrated model



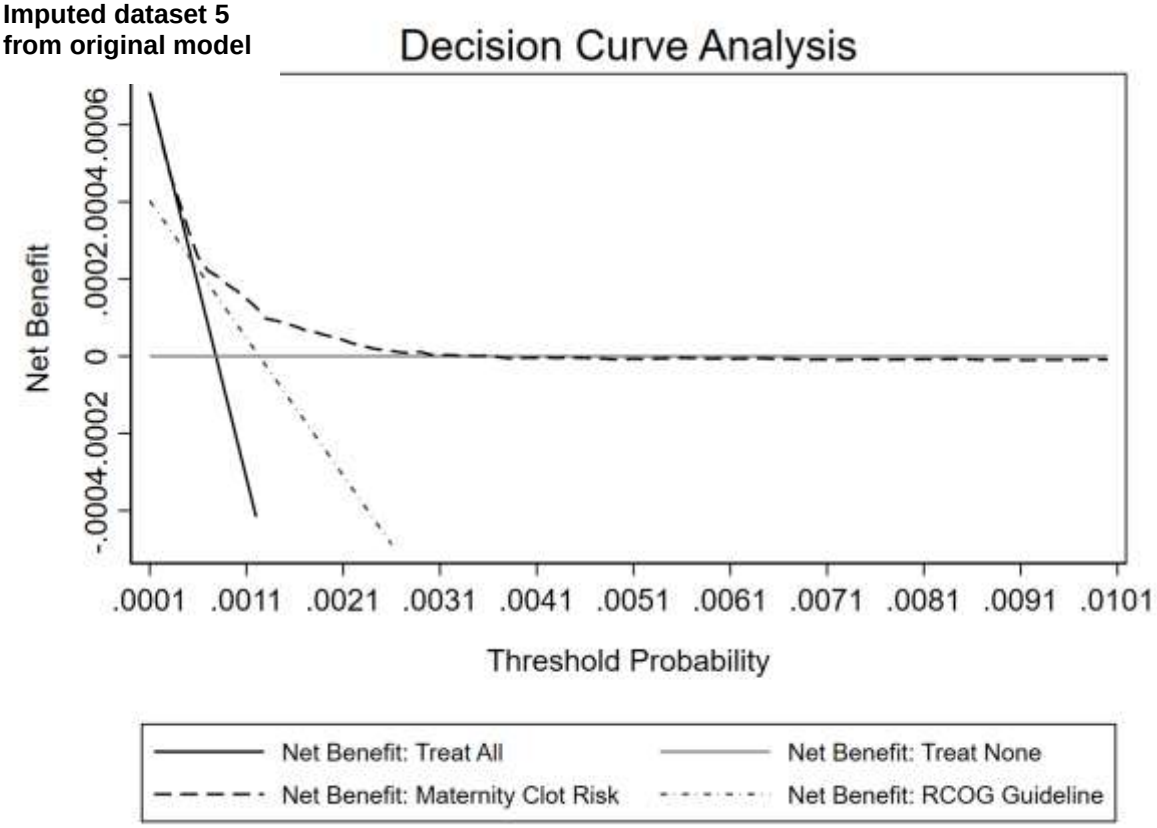
Imputed dataset 4 from original model



Imputed dataset 4 from re-calibrated model



Imputed dataset 5 from original model



Imputed dataset 5 from re-calibrated model

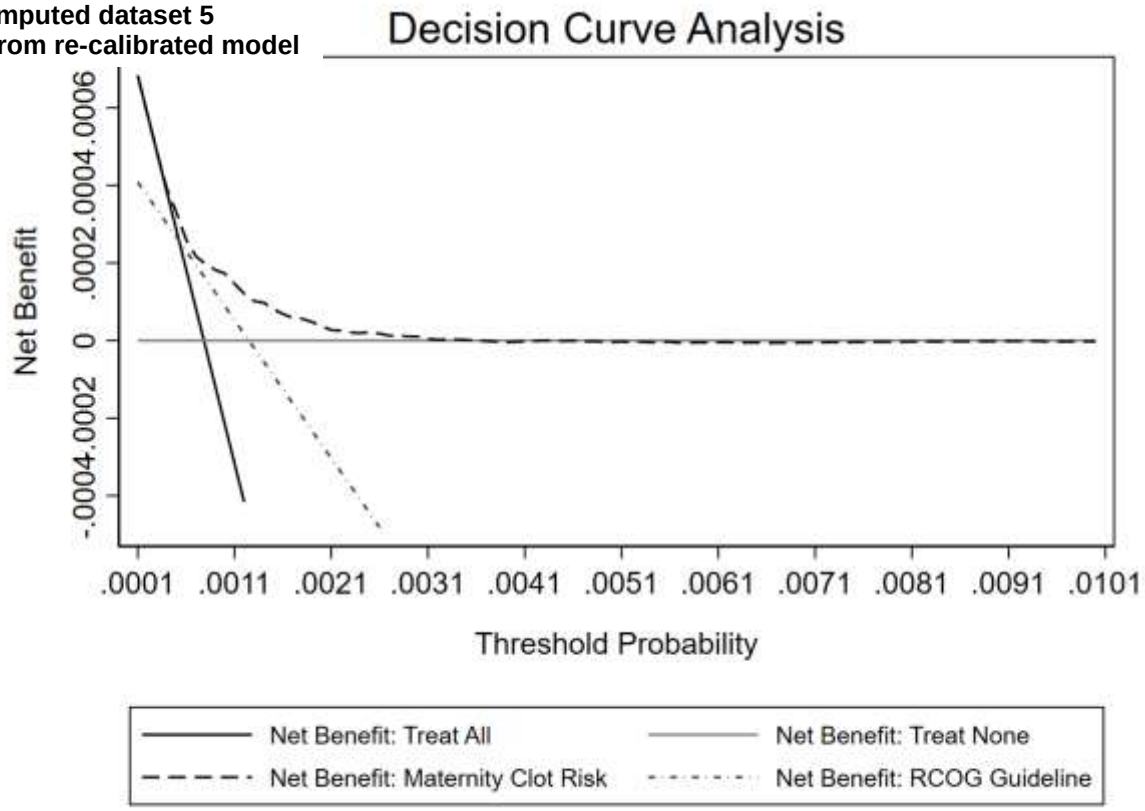


Figure S3 Calibration slope status after re-calibration by socioeconomic status

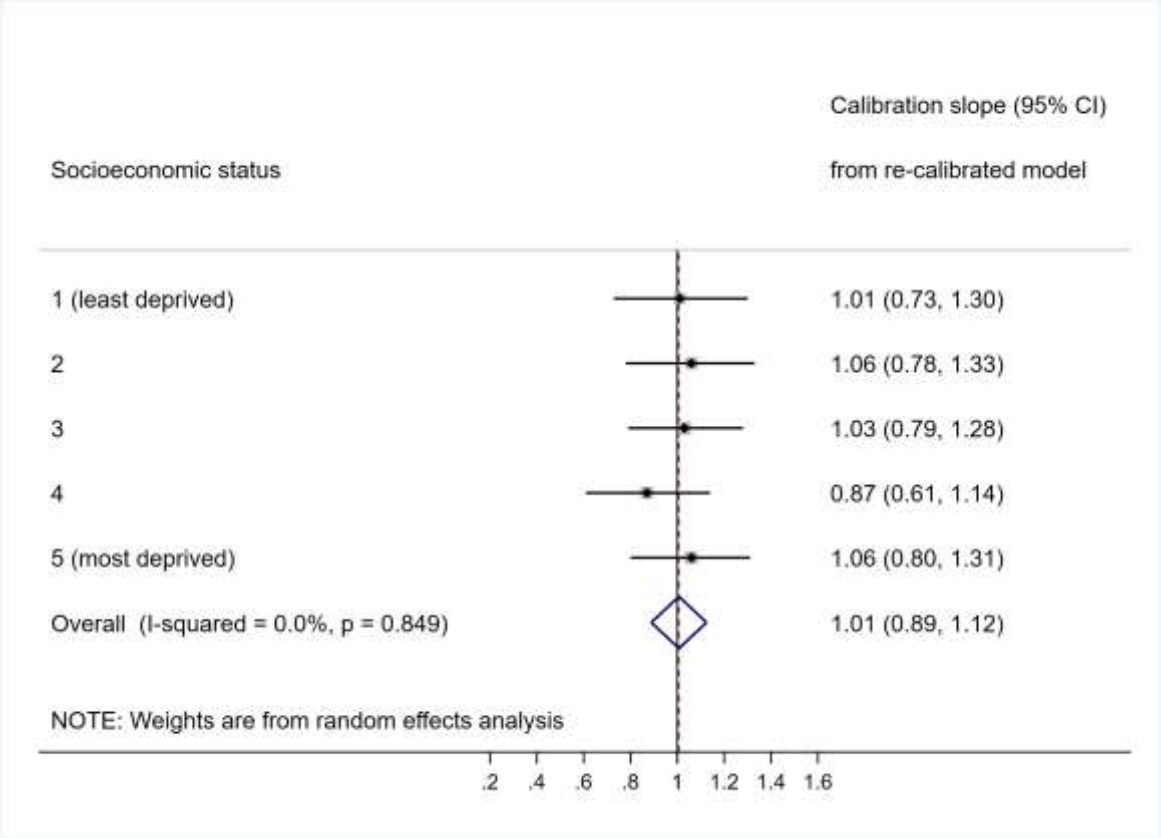


Figure S4 Calibration slope after re- calibration by region

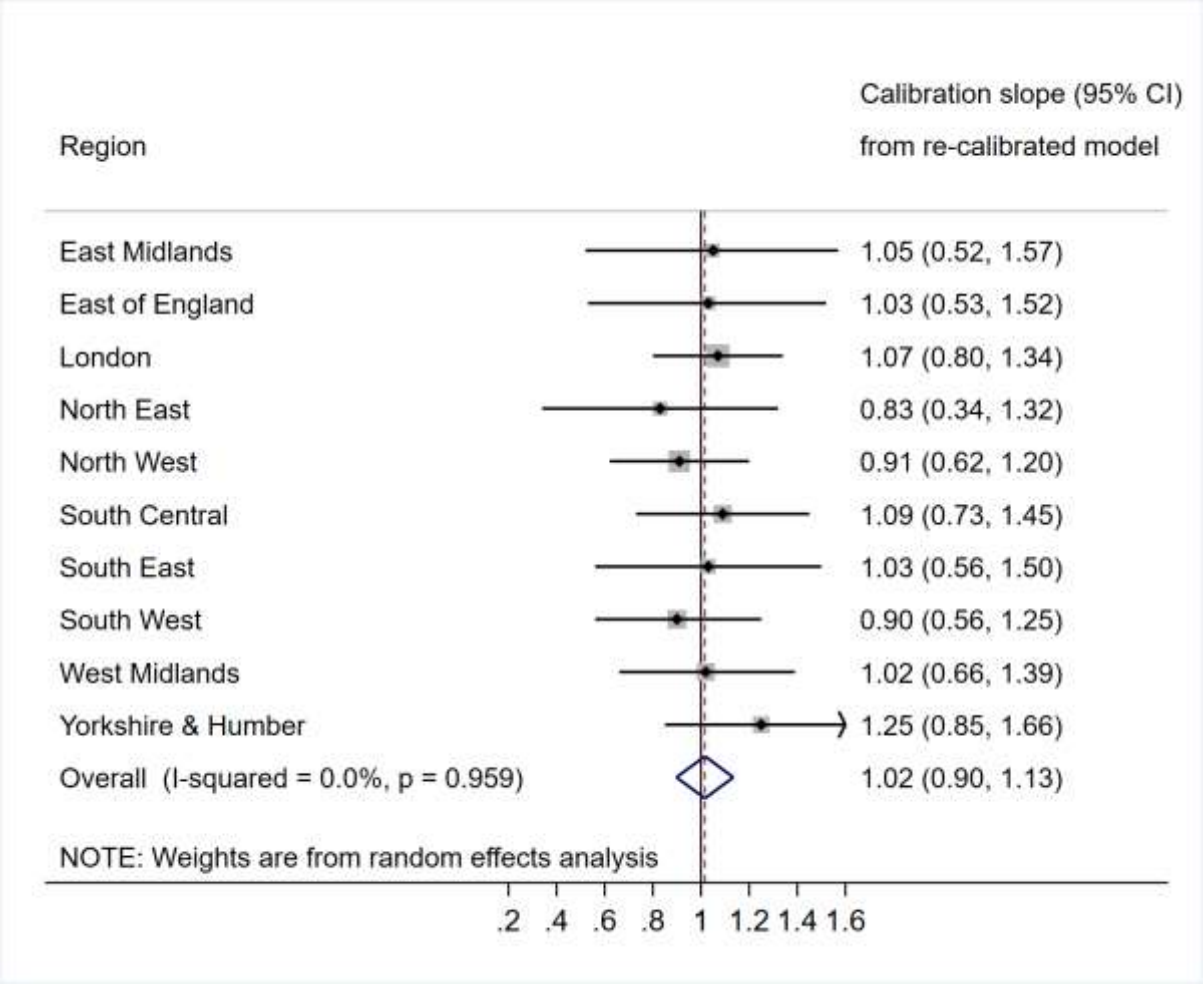


Figure S5 Calibration slope after re-calibration by calendar period

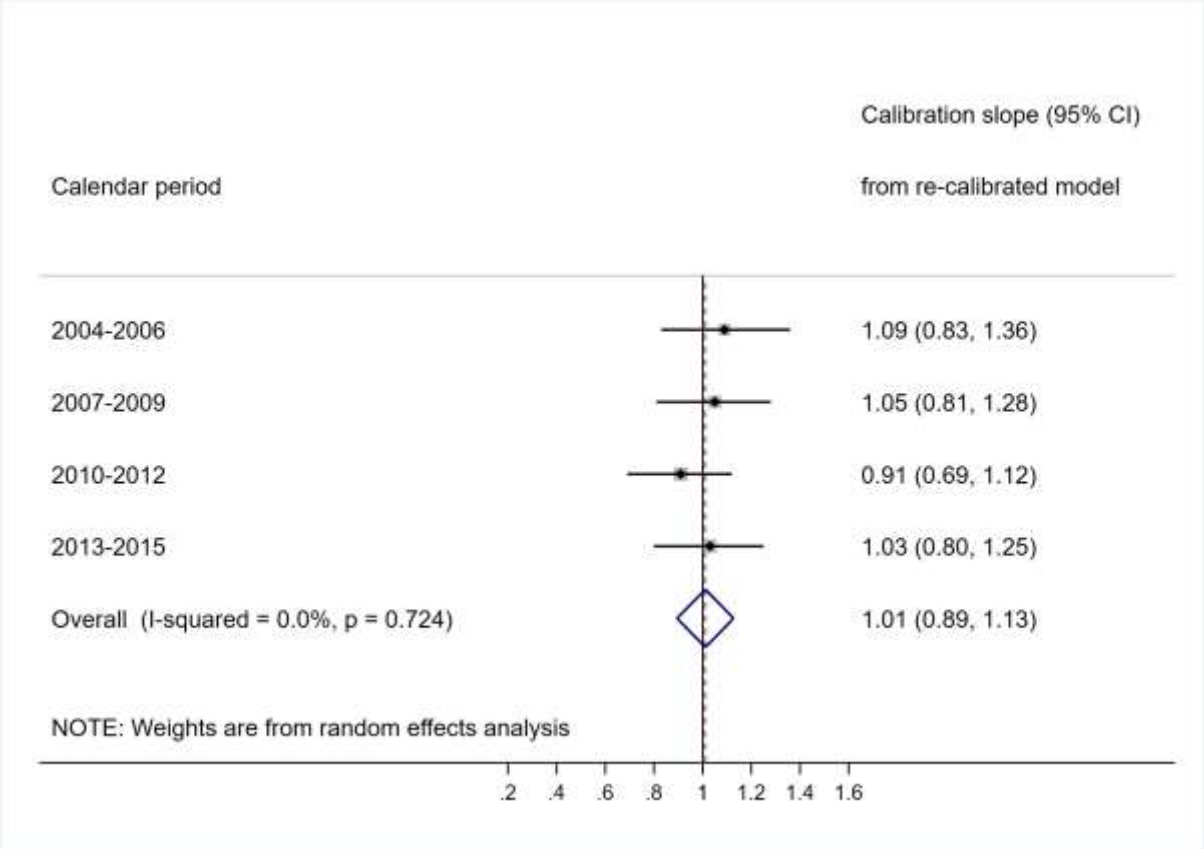


Figure S6 Calibration slope after re-calibration by age group