Development and external validation of a multivariable prognostic model to predict the 3 year risk of non-traumatic lower limb amputation in patients starting dialysis

Supplemental materials

Bram Akerboom, Roemer J. Janse, Aurora Caldinelli, Bengt Lindholm, Joris I. Rotmans, Marie Evans, Merel van Diepen

Table of Contents

Supplemental methods

For laboratory values, height, weight and blood pressure measurements the following selection process was used for inclusion:

1. If one or more covariates are measured before starting dialysis and within the predetermined maximum time before baseline (**Supplemental Table 1)** the measurement closest to start dialysis was included. If no measurements were available we proceeded to step 2.
2. If one or more covariates are measured after starting dialysis and within the predetermined maximum time after baseline (**Supplemental Table 1**) the measurement closest to start dialysis was included. If no measurements were available we proceeded to step 3.
3. The covariate was coded and treated as a missing value and handled through multiple imputation.

Supplemental Table 1 Definition of covariates and outcomes and time-windows of inclusion.

Covariates	**Source of data in SNR cohort**	**Maximum time before baseline**	**Maximum time after baseline**
Sex	From SNR registry	-	-
Age	From SNR registry	-	-
Cause of ESKD	From SNR registry	-	-
Dialysis modality at start	From SNR registry	-	-
(Date of) kidney transplantation, recovery, death	From SNR registry	-	-
Date of start kidney replacement therapy	From SNR registry	-	-
Vascular access at start	From SNR registry	-	-
Height	From SNR registry	Any	Any
Weight	From SNR registry	90 days	90 days
Systolic blood pressure	From SNR registry	180 days	180 days
Diastolic blood pressure	From SNR registry	180 days	180 days
BMI	Calculated from Height and Weight	-	-
Laboratory values:			
Albumin	From SNR registry	365 days	365 days
Creatinine at start dialysis	From SNR registry	30 days	0 days
Haemoglobin	From SNR registry	90 days	90 days
Calcium	From SNR registry	90 days	90 days
Phosphate	From SNR registry	90 days	90 days
PTH	From SNR registry	90 days	90 days
C-Reactive Protein	From SNR registry	90 days	90 days
HbA1C	From SNR registry	90 days	90 days
Total cholesterol (TC)	From SNR registry	365 days	365 days
Low density lipoprotein (LDL)	From SNR registry	365 days	365 days
High density lipoprotein (HDL	From SNR registry	365 days	365 days
Total glycerides (TG)	From SNR registry	365 days	365 days
eGFR at start dialysis	Calculated using CKD-EPI 2009	-	-
Medication use at start dialysis:	From national medication registry		
Antiplatelet agents	**ATC code: B01AC**	180 days	0 days
- *Acetylsalicylic acid*	**ATC code: B01AC06**	180 days	0 days
- *Clopidogrel*	**ATC code: B01AC04**	180 days	0 days
- *Dipyridamole*	**ATC code: B01AC07**	180 days	0 days
Antidiabetics			
- *Insulin*	**ATC code: A10A**	180 days	0 days
- *Non-insulin*	**ATC code: A10B**	180 days	0 days
Statins	**ATC code: C10AA**	180 days	0 days
Antihypertensive medication			
- *Ace inhibitor/ARBs*	**ATC code: C09**	180 days	0 days
- *Ca antagonist*	**ATC code: C08**	180 days	0 days
- *B blockers*	**ATC code: C07**	180 days	0 days

- *thiazide diuretics*	**ATC code: C03A, C03B**	180 days	0 days
Comorbidities before baseline:	From national comorbidity registry		
Amputations before/after baseline	**KKÅ code: NFQ, NGQ NHQ.**		
- *Hip*	**KKÅ code: NFQ09**	Any	Any
- *Transfemoral*	**KKÅ code: NFQ19**	Any	Any
- *Knee*	**KKÅ code: NGQ09**	Any	Any
- *Transtibial*	**KKÅ code: NGQ19**	Any	Any
- *Disarticulation of talocrural joint*	**KKÅ code: NHQ09**	Any	Any
- *Forefoot*	**KKÅ code: NHQ11-14**	Any	Any
- *Toe*	**KKÅ code: NHQ16-17**	Any	Any
- *Amputation of limb or limbs*	**ICD-10 code: Y83,5**	Any	Any
Traumatic amputations before/after baseline		Any	Any
- *Hip and thigh*	**ICD-10 code: S78**	Any	Any
- *Lower leg*	**ICD-10 code: S88**	Any	Any
- *Ankle and foot*	**ICD-10 code: S98**	Any	Any
- *Both feet*	**ICD-10 code: T05.3**	Any	Any
- *Both legs*	**ICD-10 code: T05.5**	Any	Any
- *Upper and lower extremities*	**ICD-10 code: T05.6**	Any	Any
- *Multiple, unspecified*	**ICD-10 code: T05.9**	Any	Any
- *Lower limb, level unspecified*	**ICD-10 code: T13.6**	Any	Any
- *Crush injury and then amputation, unspecified*	**ICD-10 code: T14.7**	Any	Any
Obesity	Calculated from BMI (>30 kg/m²). or **ICD-10: E66**	Any	0
Hypertension	**ICD-10 code: I10-I15**	Any	0
Diabetes mellitus	**ICD-10 code: E10-E14**	Any	0
Diabetic retinopathy	**ICD-10 code: E10.3, E11.3, E14.3**	Any	0
Symptomatic peripheral artery disease	**ICD-10 code: I702, I739**	Any	0
Dyslipidemia	**ICD-10 code: E78**	Any	0
Coronary Artery disease	**ICD-10 code: I20-I25**	Any	0
Valve disorders	**ICD-10 code: I134-I137**	Any	0
Atrial fibrillation	**ICD-10 code: I48**	Any	0
Congestive heart failure	**ICD-10 code: I11.0, I13.0, I32, I50**	Any	0
Cerebral vascular disease	**ICD-10 code: I60-I64, I69.0-I69.4 + G45**	Any	0
Malignancy <10 years before start dialysis	**ICD-10 code: C00-C97**	10 years	0

Bone fracture <1 year	**ICD-10 code: S02, S12, S22, S32, S42, S52, S62, S72, S82, S92, T02, T08, T10, T12, M48.4**	1 year	0

BMI = body mass index, CKD = chronic kidney disease, PD = peritoneal dialysis, HD = haemodialysis, AV fistula = arteriovenous fistula, eGFR = estimated glomerular filtration rate, PTH = parathyroid hormone, CRP = C-reactive protein, HDL = high-density lipoprotein, ACE-I = ACE inhibitors, ARBs = angiotensin II receptor blockers.

Supplemental Table 2 Number of missing values and multiple imputation

Covariate	SNR missing values (%)	NECOSAD missing values (%)	Secondary variable minimum days before baseline	Secondary variable maximum days before baseline
Height	683 (7.1)	91 (5.8)	NA	NA
Weight	4489 (46.5)	91 (5.8)	275	455
Systolic blood pressure	2172 (22.5)	13 (0.8)	275	455
Diastolic blood pressure	2174 (22.5)	14 (0.8)	275	455
Albumin	818 (8.4)	13 (0.8)	640	820
Haemoglobin	3979 (41.2)	10 (0.6)	275	455
Calcium albumin corrected	4483 (46.5)	11 (0.7)	275	455
Phosphate	4072 (42.2)	13 (0.8)	275	455
PTH	2409 (25.0)	-	275	455
CRP	4713 (48.9)	923 (55.7)	275	455
Cholesterol total	6268 (65.0)	186 (11.2)	640	820
Cholesterol HDL	6471 (67.1)	-	640	820
Triglycerid	6783 (70.3)	-	640	820
BMI	4531 (47.0)	91 (5.8)	275	455
Obesity	4100 (42.5)	91 (5.8)	275	455
eGFR at start	3775 (39.2)	1139	335	700
Female sex	0 (0)	1 (0.1)	-	-
Diabetes mellitus	0 (0)	146 (8.8)	-	-
Diabetic retinopathy	0 (0)	163 (9.8)	-	-
Peripheral artery disease	0 (0)	147 (8.9)	-	-
Coronary artery disease	0 (0)	147 (8.9)	-	-
Cerebral vascular disease	0 (0)	147 (8.9)	-	-
Cardiovascular disease	0 (0)	147 (8.9)	-	-
Congestive heart failure	0 (0)	147 (8.9)		
Malignancy	0 (0)	147 (8.9)	-	-
Acetylsalicylic acid	0 (0)	164 (9.9)	-	-
Statins	0 (0)	164 (9.9)	-	-
Insulin	0 (0)	171 10.3)	-	-

Some individuals did not have laboratory/demographic/blood pressure values available within our prespecified time. **(Supplemental Table 1)** Therefore, for all individuals secondary variables were created containing the values before our prespecified time, which likely would be informative for what the missing value would have been. These variables and all covariates and outcome variables as described in **Supplemental Table 1** were used during multiple imputation. BMI = body mass index, eGFR = estimated glomerular filtration rate, PTH = parathyroid hormone, CRP = C-reactive protein, HDL = high-density lipoprotein.

Supplemental Table 3 The following candidate coefficients were preselected and are ranked in order of relevance

1. Female sex
2. Age (years)
3. Diabetes Mellitus
4. Symptomatic peripheral artery disease
5. Cardiovascular disease (Cerebral vascular disease or Coronary artery disease)
6. Congestive heart failure
7. Obesity
8. Albumin (g/L)
9. Haemoglobin (mmol/L)
10. Diabetic retinopathy
11. Body mass index (kg/m²)
12. Triglycerides (mmol/L)
13. Hypertension
14. Peritoneal dialysis
15. Phosphate (mmol/L)

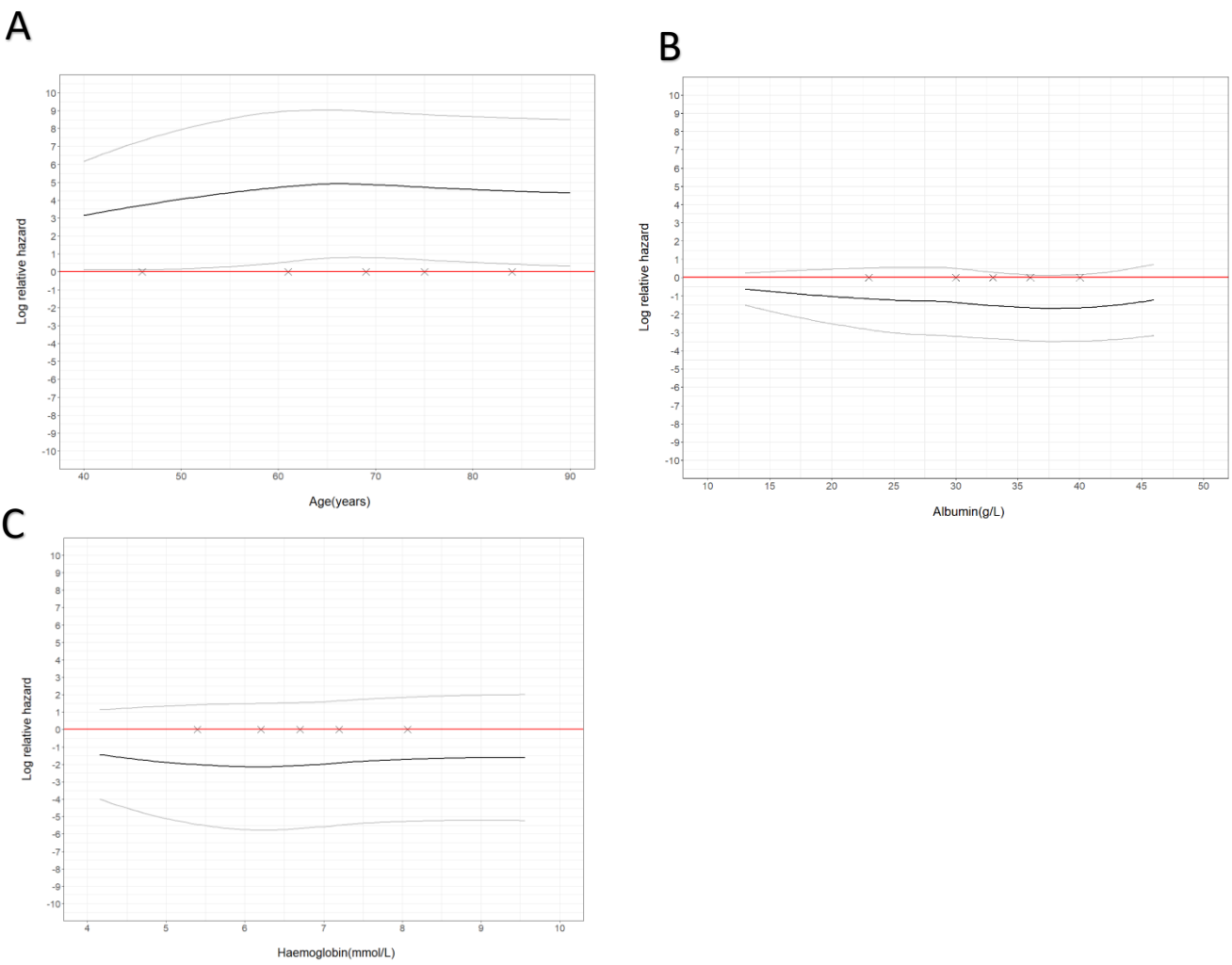
Supplemental Table 4 Correlation matrix

	Female	Age	Obesity	Albumin	Hb	DM	DR	PAD	CVD	CHF
Female		-0.02	0.03	-0.01	0.01	-0.02	0	-0.01	-0.08	-0.02
Age	-0.02		-0.12	-0.04	0.03	0.04	-0.07	0.11	0.24	0.21
Obesity	0.03	-0.12		0.01	0	0.25	0.17	-0.01	0.01	0.07
Albumin	-0.01	-0.04	0.01		0.2	-0.1	-0.1	-0.02	-0.02	-0.07
Hb	0.01	0.03	0	0.2		-0.01	-0.01	0.01	0.03	0.01
DM	-0.02	0.04	0.25	-0.1	-0.01		0.63	0.12	0.2	0.21
DR	0	-0.07	0.17	-0.1	-0.01	0.63		0.1	0.14	0.18
PAD	-0.01	0.11	-0.01	-0.02	0.01	0.12	0.1		0.18	0.16
CVD	-0.08	0.24	0.01	-0.02	0.03	0.2	0.14	0.18		0.32
CHF	-0.02	0.21	0.07	-0.07	0.01	0.21	0.18	0.16	0.32	

We expected to find some collinearity in the shaded coefficients. This is confirmed by some correlation between diabetes mellitus and diabetic retinopathy (0.63). Since we deemed adding both coefficients would provide valuable extra information we included them both in the final model. CHF = Congestive heart failure, CVD = Cardiovascular disease, DM = Diabetes Mellitus, DR = Diabetic retinopathy, Hb = Haemoglobin, PAD = Peripheral artery Disease

Supplemental Table 5 Final model cumulative incidence at 3 years and coefficients

Coefficient	Value	SE	Variable type	Transformation
Cumulative incidence at 3 years	0.0096	-	-	-
Female sex	-0.4652	0.1691	Categorical	-
Age when starting dialysis (years)	-0.0003	0.0068	Continuous	None
Diabetes mellitus	0.7332	0.2405	Categorical	-
Symptomatic peripheral artery disease	0.9705	0.1720	Categorical	-
Cardiovascular disease (coronary artery disease of cerebral vascular disease)	0.3577	0.2328	Categorical	-
Congestive heart failure	0.5456	0.2258	Categorical	-
Obesity	0.2953	0.1943	Categorical	-
Serum albumin (g/L)	-0.0128	0.0170	Continuous	None
Serum haemoglobin (mmol/L)	0.1076	0.0904	Continuous	None
Diabetic retinopathy	0.8981	0.1852	Categorical	-



Supplemental figure 1. Linearity of continuous coefficients

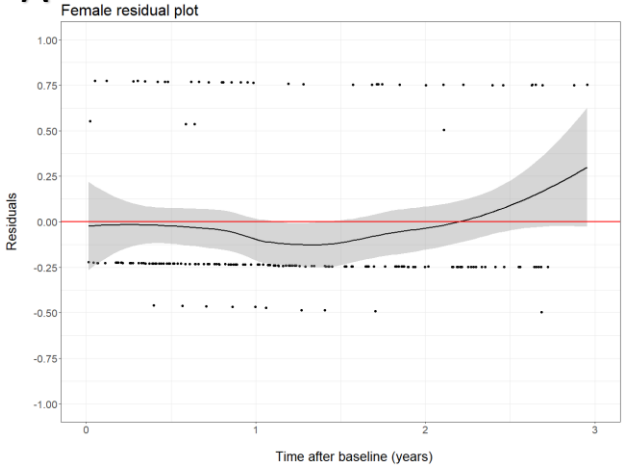
The log relative hazard is plotted as a restricted cubic spline with 5 knots to continuous coefficients. The ability to draw a straight line within the 95CI supports the linearity assumption of all continuous variables.

A: Age (years)

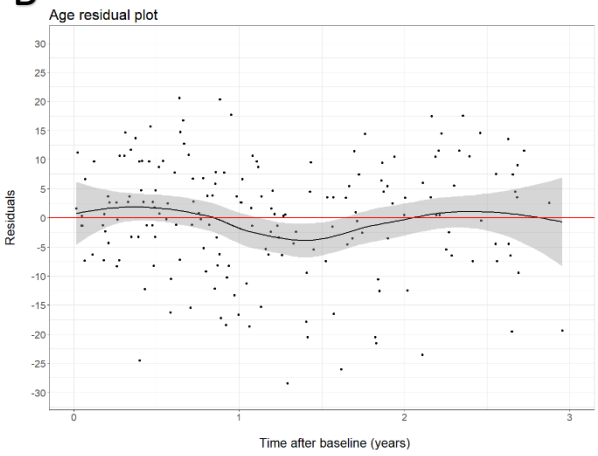
B: Albumin (g/L)

C: Haemoglobin (mmol/L)

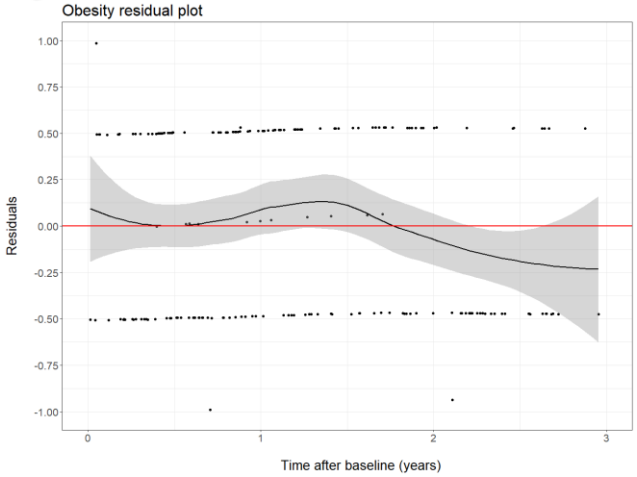
A


Female residual plot
Residuals
Time after baseline (years)

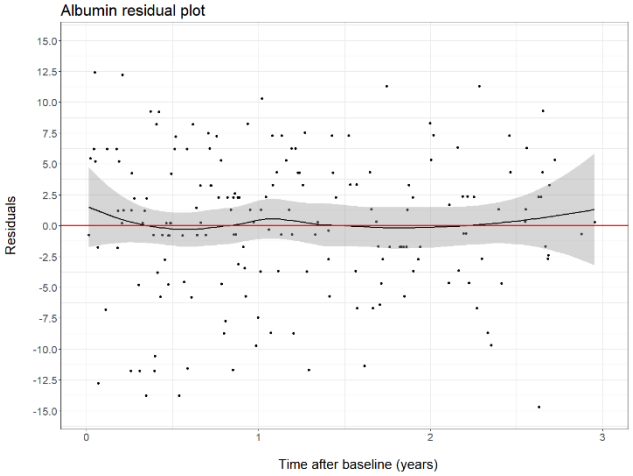
B


Age residual plot
Residuals
Time after baseline (years)

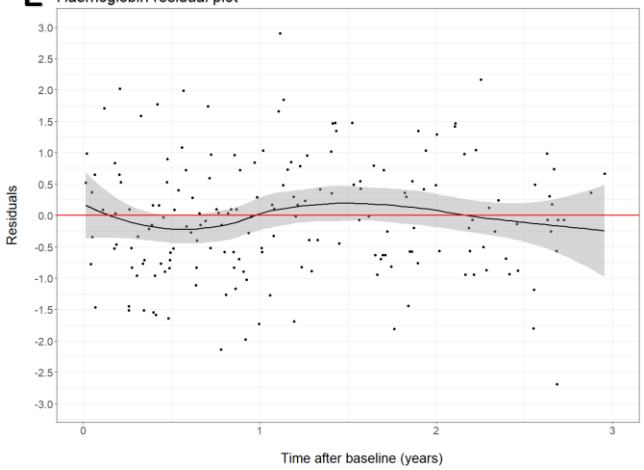
C


Obesity residual plot
Residuals
Time after baseline (years)

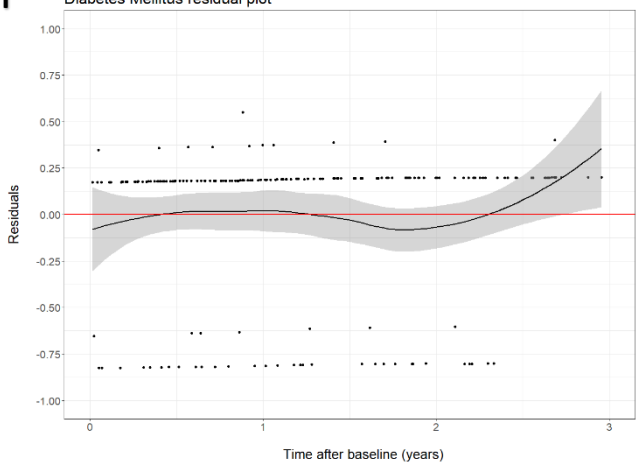
D

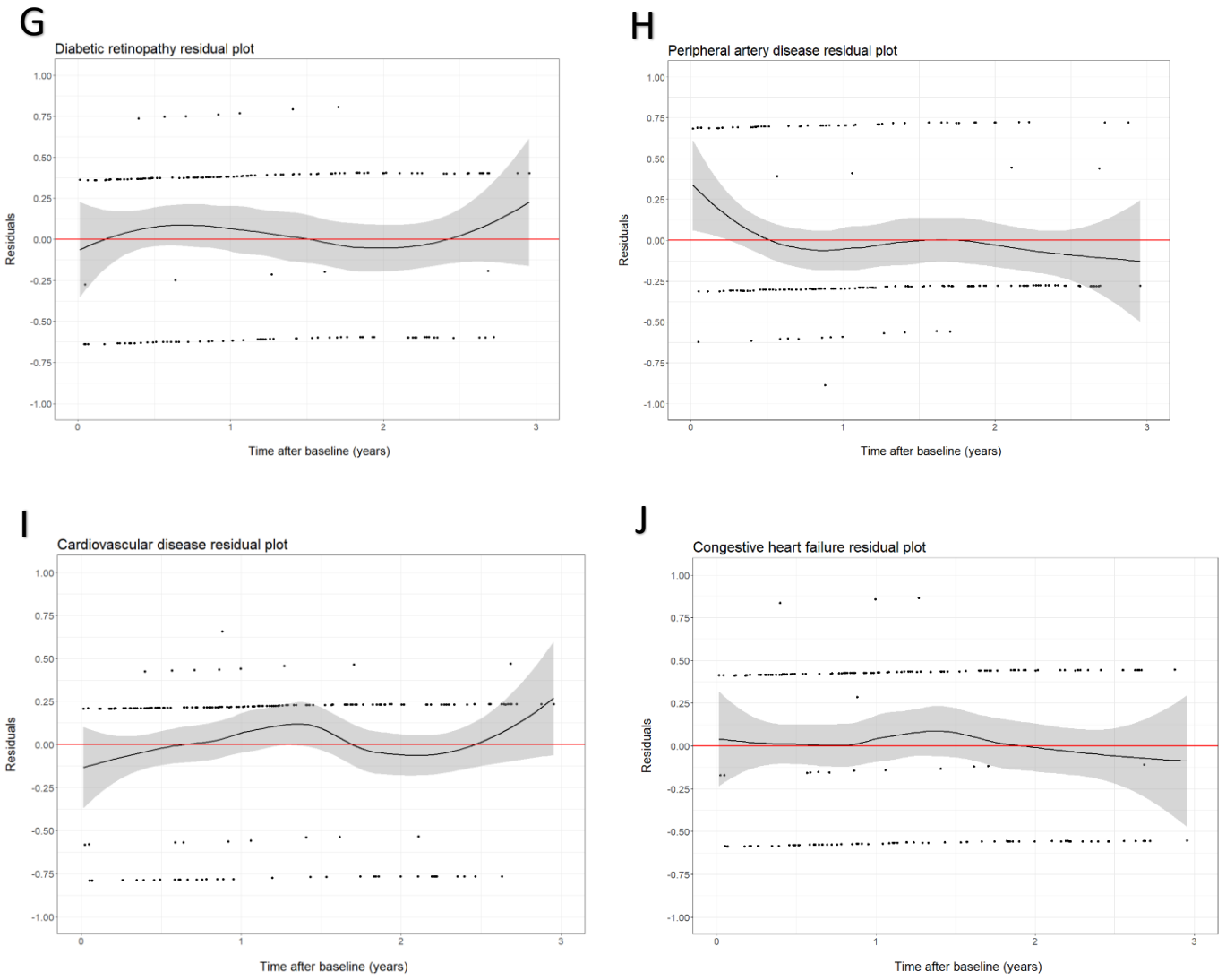

Albumin residual plot
Residuals
Time after baseline (years)

E


Haemoglobin residual plot
Residuals
Time after baseline (years)

F

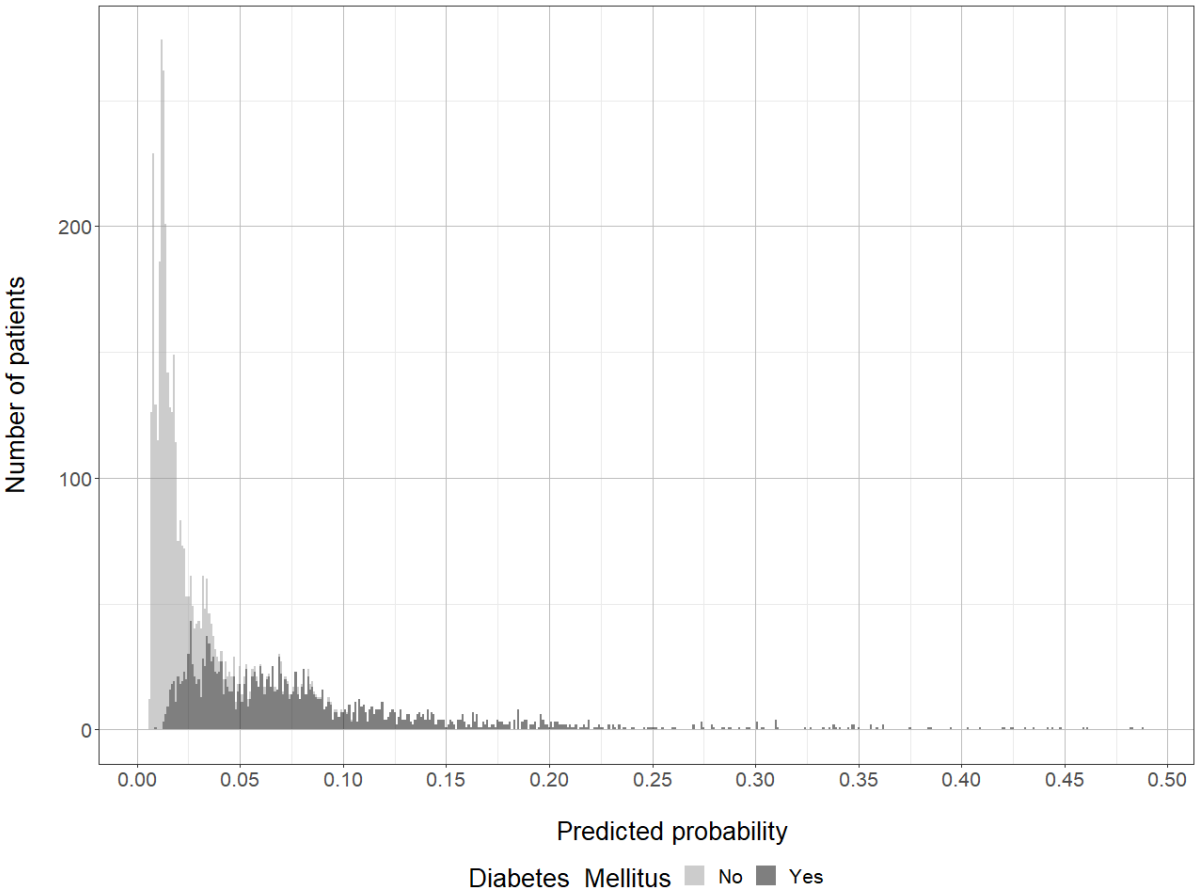

Diabetes Mellitus residual plot
Residuals
Time after baseline (years)



Supplemental figure 2. Residuals plot for proportional hazards

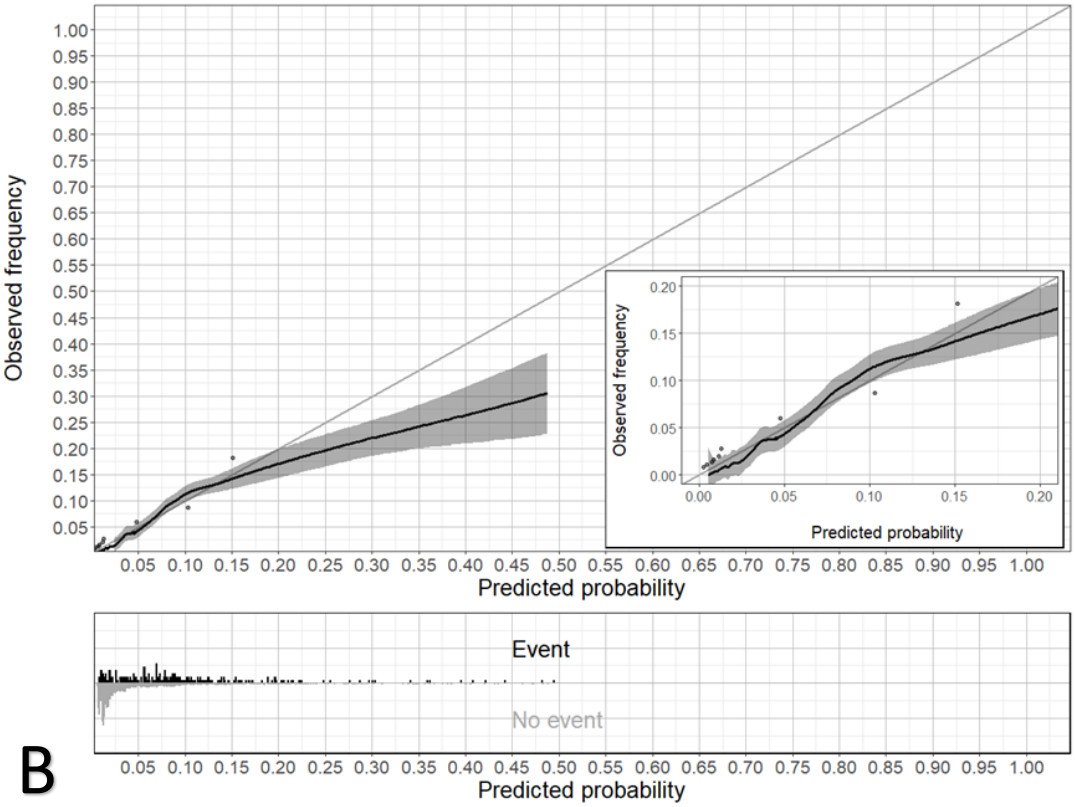
Analogues of the Schoenfeld residuals are plotted to the corresponding time after baseline. If hazards are not proportional the 95CI of the residuals does not contain 0. None of the plots violated the proportional hazards assumption sufficiently to warrant exclusion or transformation.

A: Female sex
B: Age (years)
C: Obesity
D: Albumin (g/L)
E: Haemoglobin (mmol/L)
F: Diabetes Mellitus
G: Diabetic retinopathy
H: Symptomatic peripheral artery disease
I: Cardiovascular disease (Cerebral vascular disease or Coronary artery disease)
J: Congestive heart failure

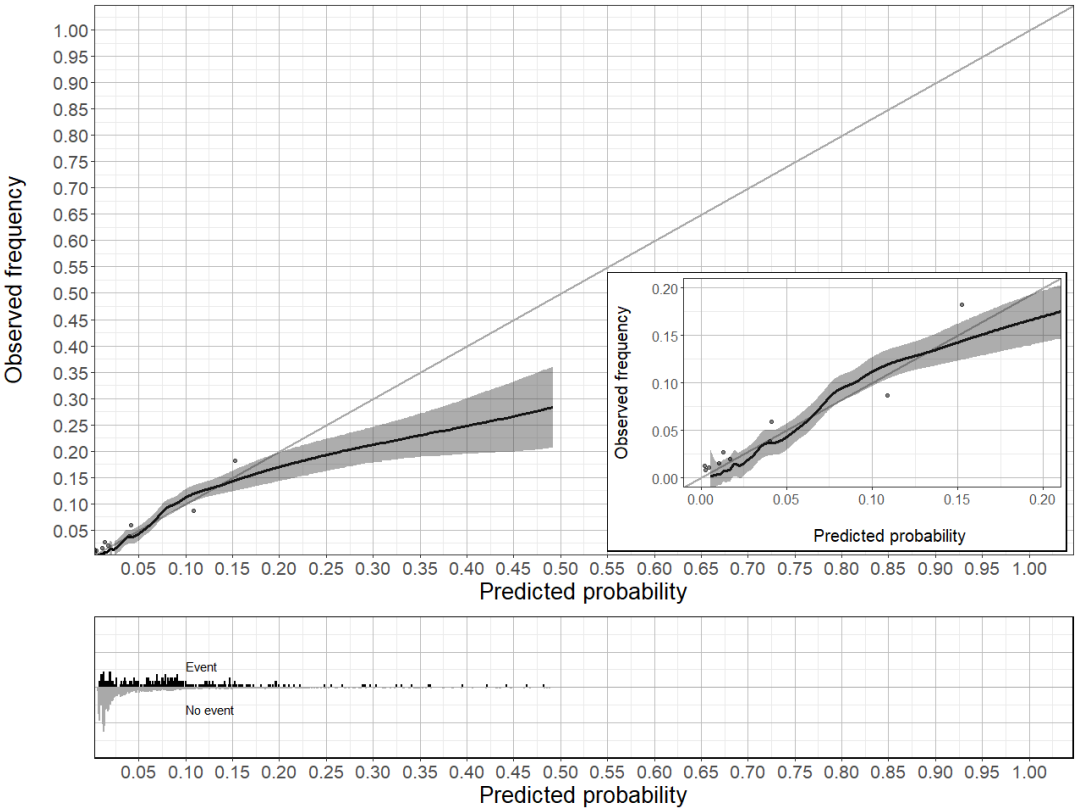


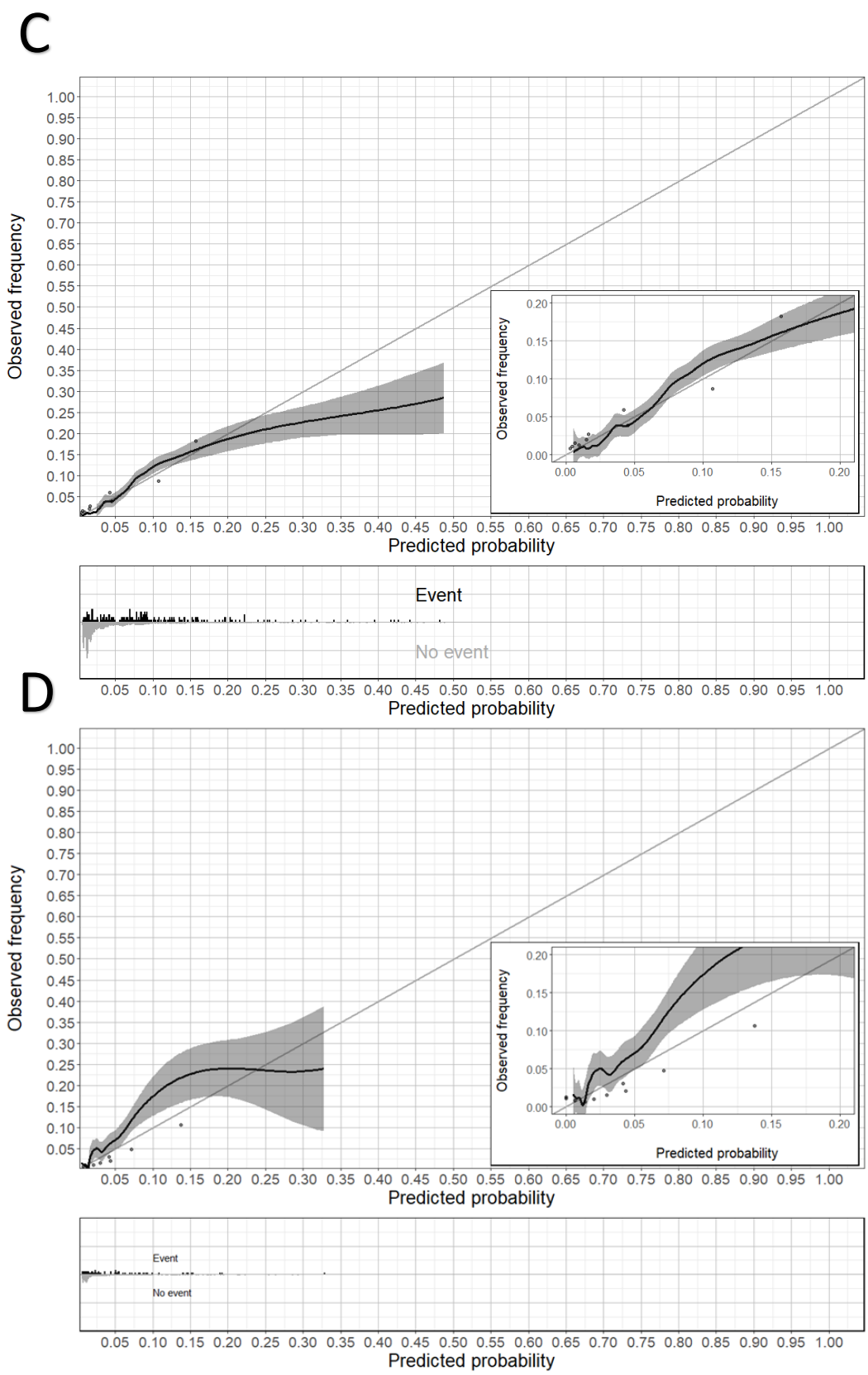
Supplemental figure 3. Distribution of predicted probabilities
Distribution of predicted probabilities for patients with and without diabetes mellitus

A


Observed frequency
Predicted probability
Observed frequency
Predicted probability
Event
No event
Predicted probability

B


Observed frequency
Predicted probability
Observed frequency
Predicted probability
Event
No event
Predicted probability



Supplemental figure 2. Calibration plots during sensitivity analyses

The predicted probability in the cohort is plotted to the observed probability. The shaded area indicates the 95% confidence interval. The grey 45 degree line indicates perfect calibration. The histograms show the relative incidence of either amputation or other type event compared to the predicted probability. The cohort is also divided into 10 percentiles according to their predicted probability. The grey dots show the average predicted probability of each group plotted against the average observed frequency.

A: SNR validation (rural) cohort: Recovery excluded

B: SNR validation (rural) cohort: No censoring after kidney transplant

C: SNR validation (rural) cohort: No exclusion if patient started dialysis after kidney transplant failure

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	3
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	4
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	4
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	5
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	5
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	5
	5b	D;V	Describe eligibility criteria for participants.	5
	5c	D;V	Give details of treatments received, if relevant.	NA
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	5
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	6
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	6
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	6
Sample size	8	D;V	Explain how the study size was arrived at.	6
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	6
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	6
	10c	V	For validation, describe how the predictions were calculated.	6
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	6-7
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	NA
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	NA
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	6
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	8
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	8
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	8
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	8
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	NA
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	8
	15b	D	Explain how to the use the prediction model.	9
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	9
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	NA
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	11
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	NA
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	10-11
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	10
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	13
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	12

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.