

Evaluating Predictive Accuracy of Survival Models with PROC PHREG

PHUSE Virtual EU Connect 2021

November 15-19, 2021

DEBRUS Roxane
Biostatistics Manager EMEA

JACOBS Lotte
Senior Medical Affairs Specialist

European Medical and Clinical Department, TERUMO EUROPE

TERUM
ARC
HBR

Abstract

Predictive accuracy measures the ability of a model to predict future risks. It can be used to validate a model or to compare new data to a pre-existing model. The concordance statistic (C-statistic) is the most commonly used discrimination measure in the context of logistic regression with binary outcomes. Besides the C-statistic, receiver operator characteristic (ROC) curves and AUC (area under the ROC curve) statistics are also commonly used to assess the discrimination ability of the model with binary outcomes. In the context of survival analysis, various C-statistics have been formulated to deal with right-censored data. In SAS, the PHREG procedure provides a number of state-of-the-art techniques to calculate overall concordance statistics and time-dependent ROC curves and AUC statistics for right-censored data. This presentation describes how to use these criteria to validate and compare fitted survival models and presents an example to illustrate its application.

Confidentiality Disclosure Statement

The following material contains Terumo confidential and proprietary information. No part of this document may be disclosed, printed, copied or disseminated in any form or by any means, without prior permission from Terumo Europe N.V.



Agenda

- **Introduction & Rational**
- **Business Case**
- **Conclusion**

Introduction & Rational

C-Statistic vs ROC and AUC (Logistic Regression)

Predicted Risk	Outcome Yes	Outcome No	Total
Positive	50	20	70
Negative	10	120	130
Total	60	140	200

■ Sensitivity

Proportion of positive predictive results
among people with observed outcome: $Se = 50/60 = 0.83$

■ Specificity

Proportion of negative predictive results in people without observed outcome: $Sp = 120/140 = 0.86$

■ Positive Predictive Value

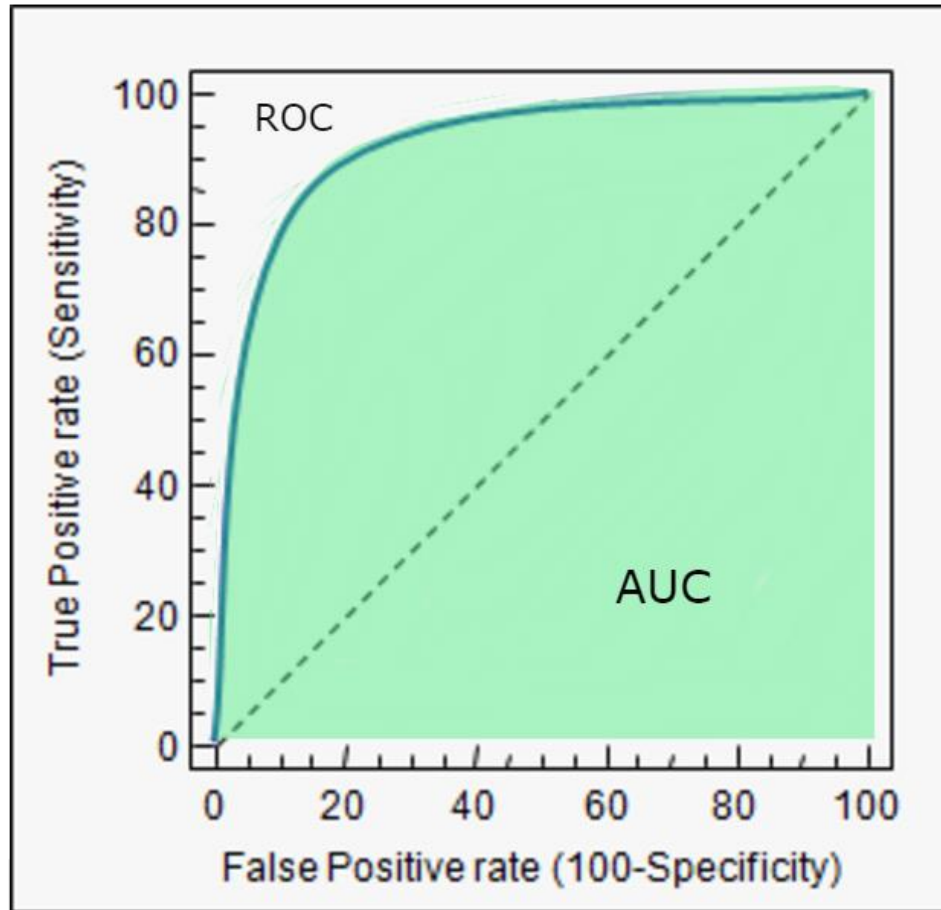
Proportion of people having the outcome among those with positive predictive results: $PPV = 50/70 = 0.71$

■ Negative Predictive Value

Proportion of people without outcome among those with negative predictive results: $NPV = 120/130 = 0.92$

C-Statistic vs ROC and AUC (Logistic Regression)

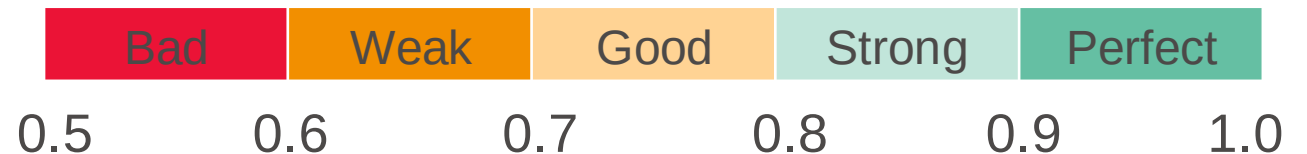
Two important aspects of a prediction model are calibration and discrimination.



Calibration refers to the ability of the model to correctly rank the individuals in the sample by risk.

Discrimination characterizes the model's ability to correctly classify subjects for their actual outcomes.

- **Receiver Operator Characteristic (ROC) curve**
- **Area under the curve (AUC)**
- **C-Statistic** measure of goodness of fit for binary outcomes in a logistic regression model



Harrell's C-index (Survival Analysis)

- Also known as the concordance index
- Measures goodness of fit for models which produce risk scores.
- Commonly used to evaluate risk models in survival analysis (censored data)
- Whereas C-statistics provide overall measures of predictive accuracy, time-dependent ROC curves and AUC functions summarize the predictive accuracy at specific times.

For a given patient, we are interested in predicting the “time-to-event”, i.e. the time duration until event (disease) occurrence, based on (exposure) covariates available at baseline.

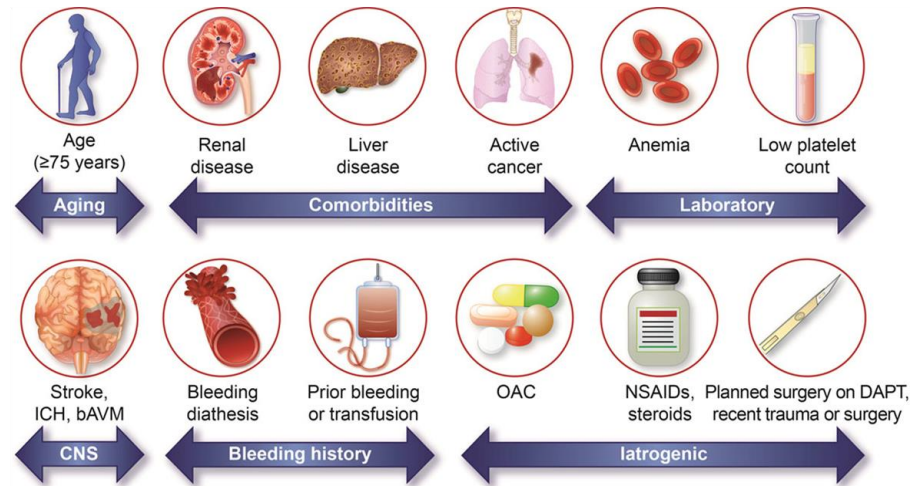
Business Case

To show the practical application of predictive accuracy measures to externally validate an existing risk model in the field of cardiovascular medicine

Description of the predictive risk model

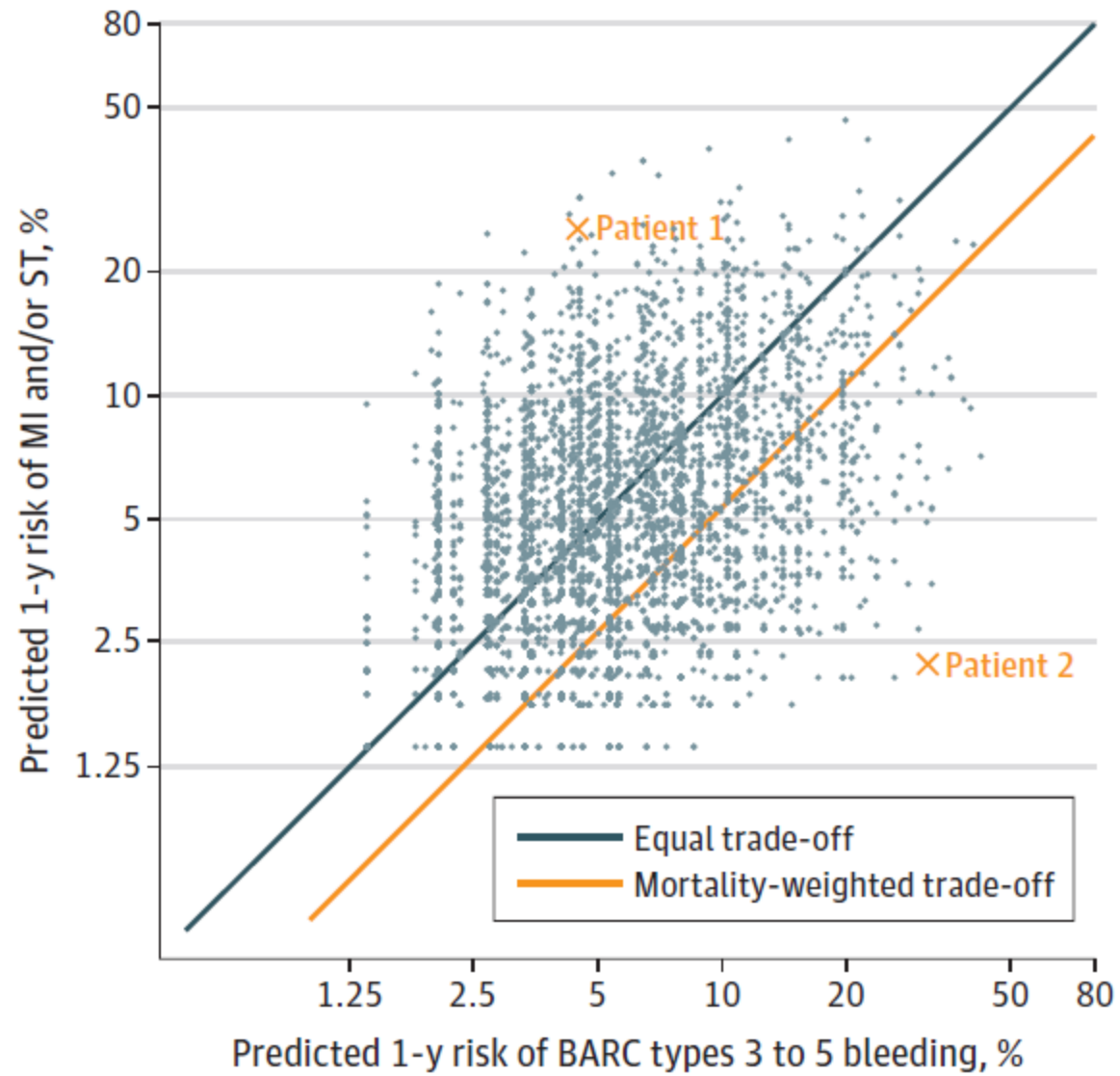
- Developed by Academic Research Consortium High Bleeding Risk (ARC-HBR) expert group
- Population of patients with
 - **Coronary artery disease**, treated with the implantation of coronary stents
 - **High bleeding risk**

Factors associated with increased bleeding risk after PCI



Urban P et al. Eur Heart J 2019;40:2632-2653

A Predicted 1-y risk in patients at high bleeding risk



sk model

sk

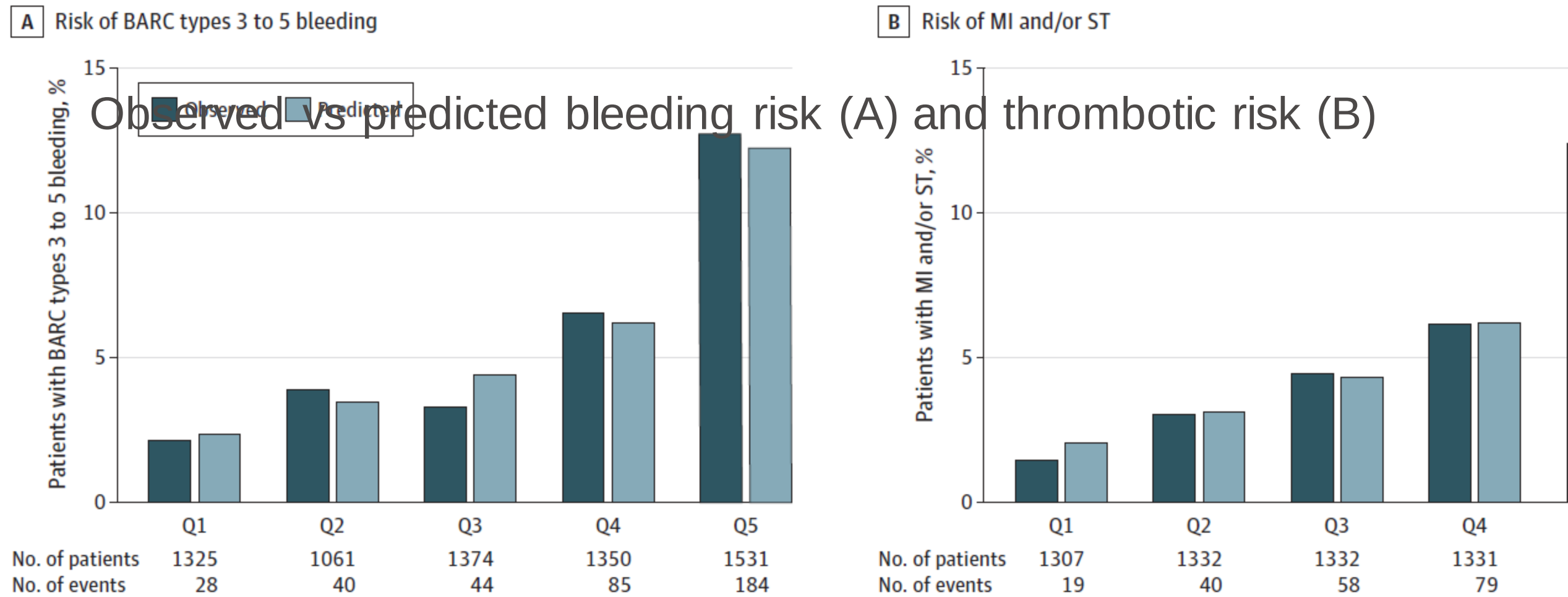
risks of major bleeding and thrombotic events

antithrombotic treatment options

ARC-HBR Trade-off Model - DISCRIMINATION

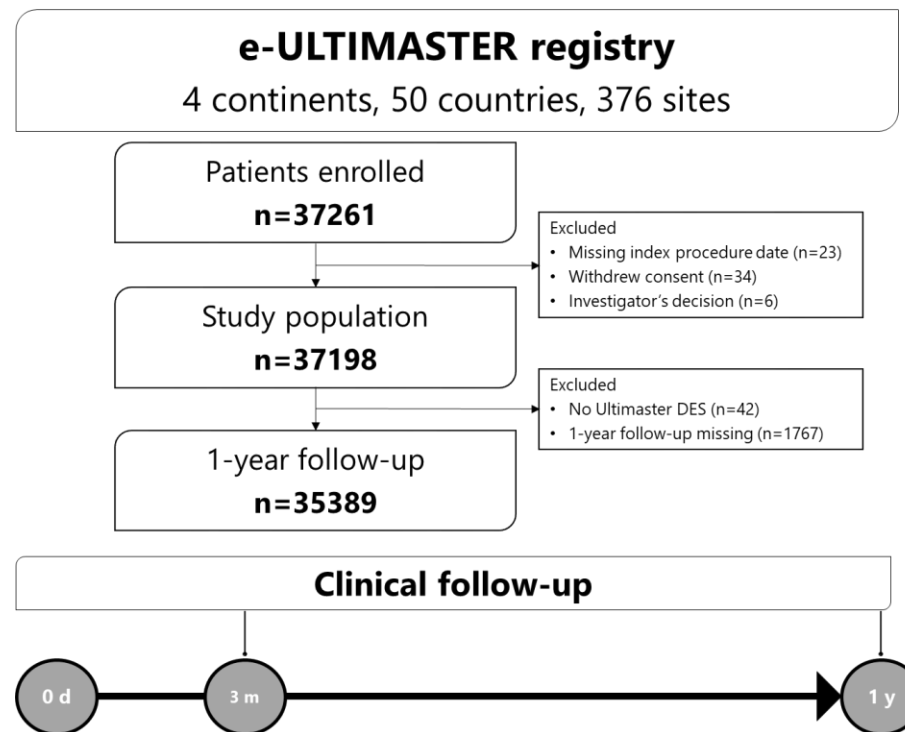
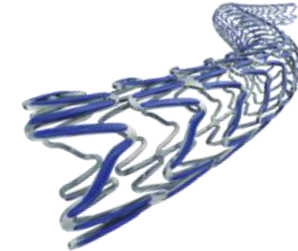
Predictor	Bleeding risk	Thrombotic risk
Age		
Diabetes		
Prior myocardial infarction		
Liver disease, cancer or surgery		
Chronic obstructive pulmonary disease		
Current smoker		
Clinical presentation of disease		
Hemoglobin		
Renal function (eGFR)		
Complex procedure		
Bare metal stent		
Oral anticoagulation medication		
C-statistic	0.68	0.69

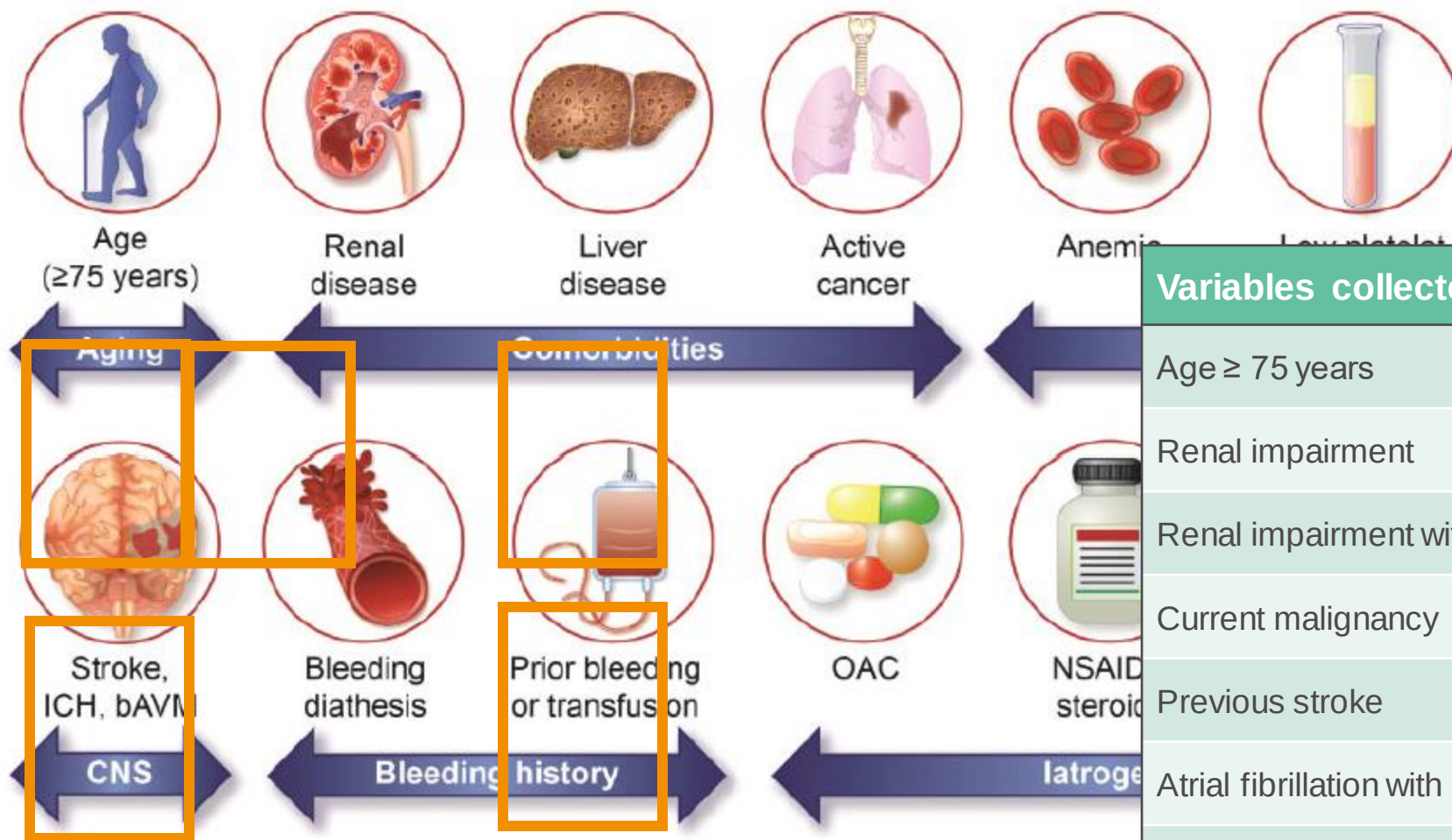
Figure 2. Observed and Predicted Risk of Bleeding Academic Research Consortium (BARC) Types 3 to 5 Bleeding and Myocardial Infarction (MI) and/or Stent Thrombosis (ST) by Risk Strata Quintiles (Q1 to Q5)



e-ULTIMASTER

- Clinical Field: Interventional cardiology
- Investigational Medical Device : Coronary drug-eluting stent
- **37198 patients with coronary artery disease**, 378 sites, 50 countries





ording to

Variables collected in e-ULTIMASTER

Age ≥ 75 years

Renal impairment

Renal impairment with dialysis

Current malignancy

Previous stroke

Atrial fibrillation with oral anticoagulant therapy

Oral anticoagulation intake

Figure Factors associated with an increased bleeding risk after percutaneous coronary intervention. bAVM, brain arteriovenous malformation; CNS, central nervous system; DAPT, dual antiplatelet treatment; ICH, intracranial hemorrhage; NSAID, nonsteroidal anti-inflammatory drug; and OAC, oral anticoagulation.

Urban P et al. Eur Heart J 2019;40:2632-2653



**Selection of 5017
high bleeding risk patients**

e-ULTIMASTER HBR to validate ARC-HBR Trade-off Model

- We used ARC-HBR algorithm based on baseline hazard functions and hazard ratio to calculate a Predicted Risk (PR) for each subject

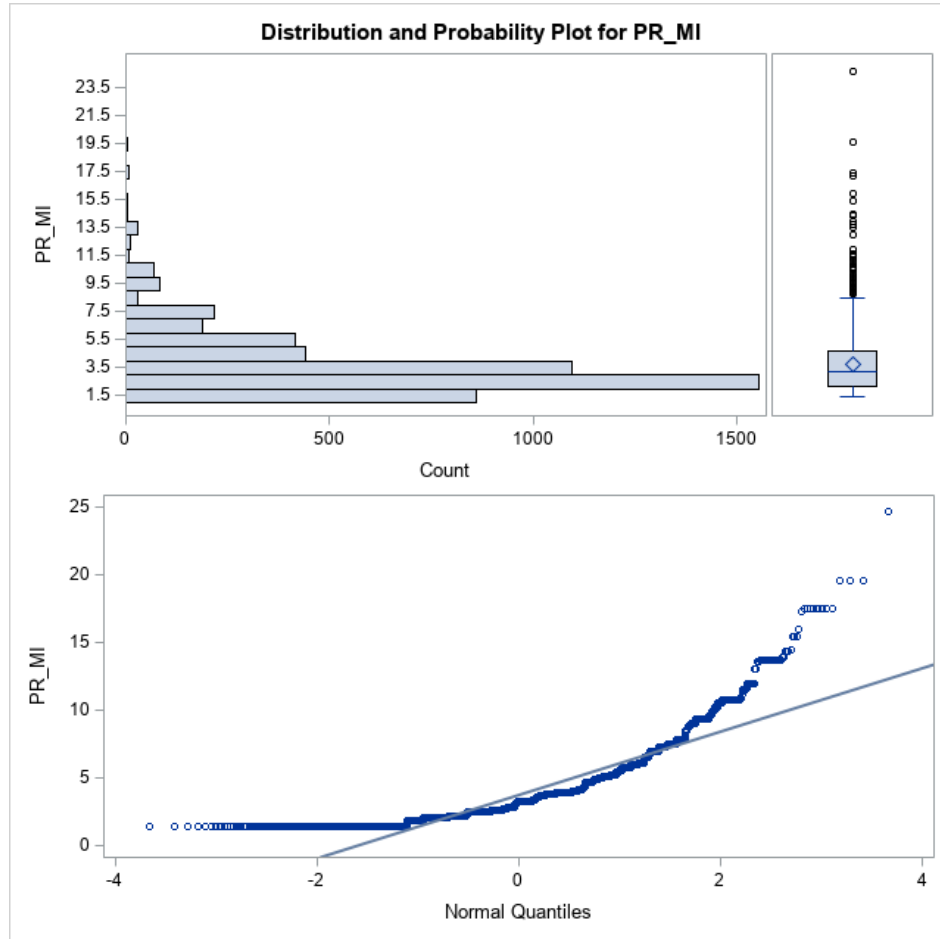
Thrombotic PR	Response	Assigned value
Diabetes	No	1
	Yes	1.56
Smoker	No	1
	Yes	1.48
Prior MI	No	1
	Yes	1.89
NSTEMI – STEMI Presentation	No	1
	Yes	1.82
Hb	≥130	1
	≥110 and <130	1.27
	<110	1.50
EGFR	≥60	1
	≥30 and <59	1.30
	<30	1.69
Complex	No	1
	Yes	1.50
Bare Metal Stent	No	1
	Yes	1.53

Bleeding PR	Response	Assigned Value
Age	<65	1
	≥65	1.50
Smoker	No	1
	Yes	1.47
COPD	No	1
	Yes	1.39
Hb	≥130	1
	≥110 and <130	1.69
	<110	3.99
EGFR	≥60	1
	≥30 and <59	0.99
	<30	1.43
Complex	No	1
	Yes	1.32
OAC	No	1
	Yes	2.0
Liver cancer surgery	(liver disease=Yes) OR (cancer=Yes and years≤3) OR (surgery=Yes)	1.63
	Otherwise	1

Predicted risk

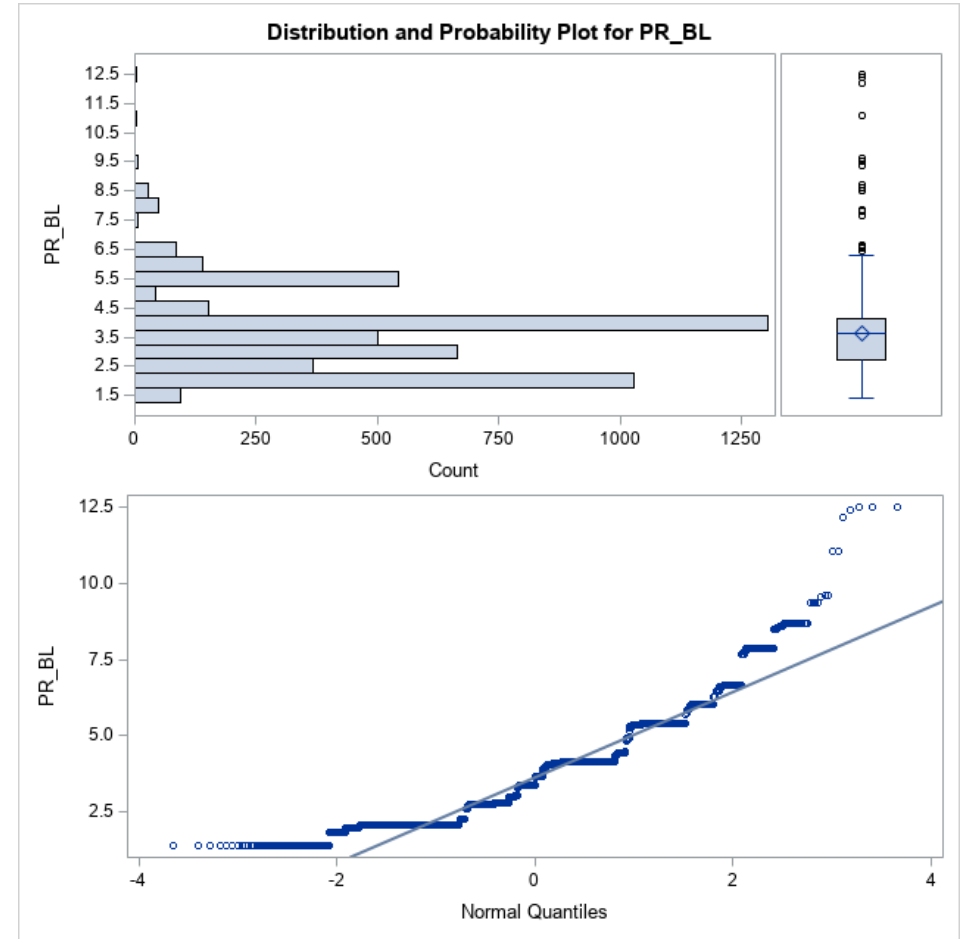
Predicted thrombotic risk (MI/ST)

(Min= 1.40; Max= 24.7; Mean= 3.72, Median= 3.25)

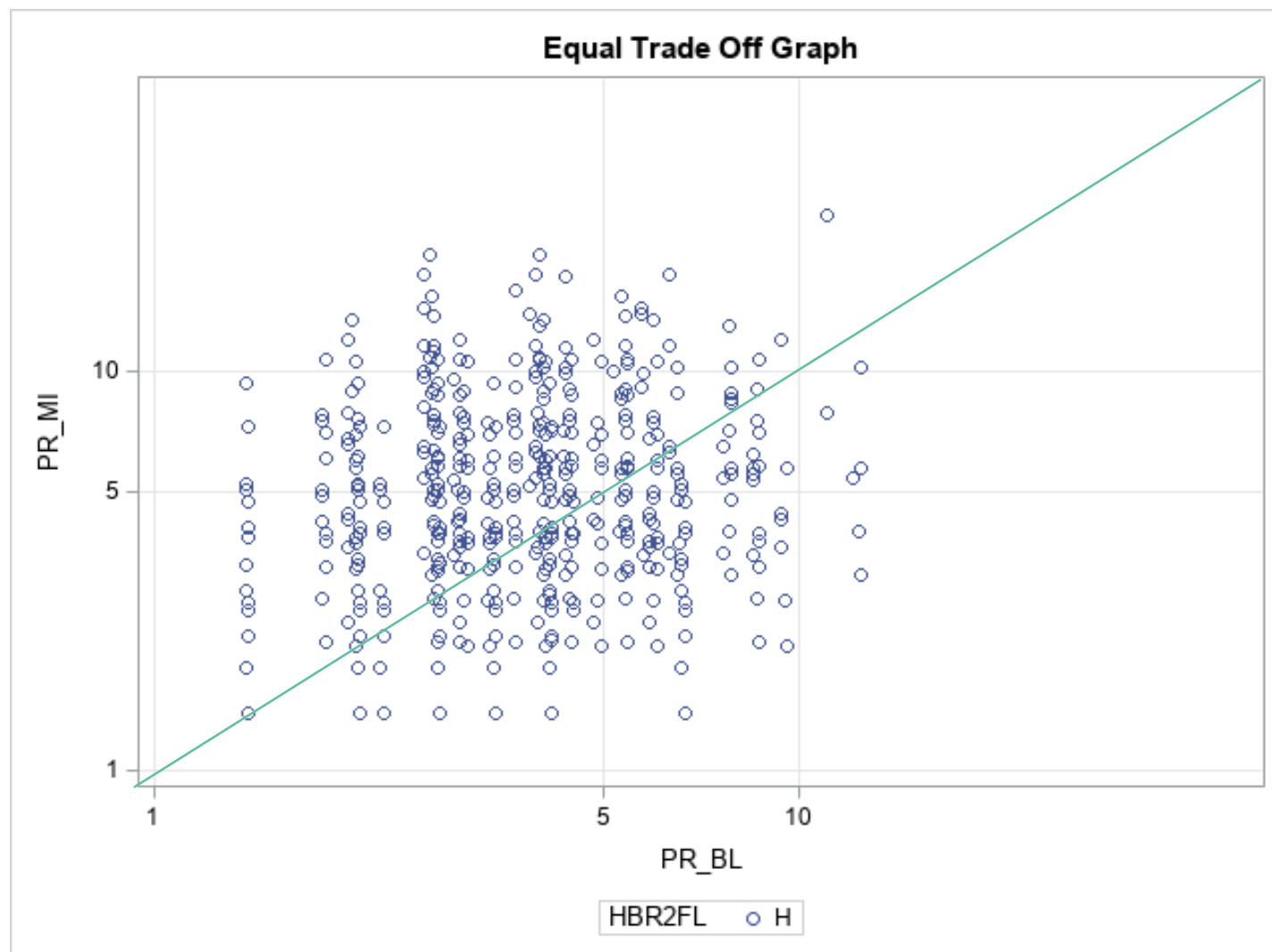


Predicted bleeding risk (BARC3-5 Bleeding)

(Min= 1.39; Max= 12.52; Mean= 3.63, Median= 3.62)



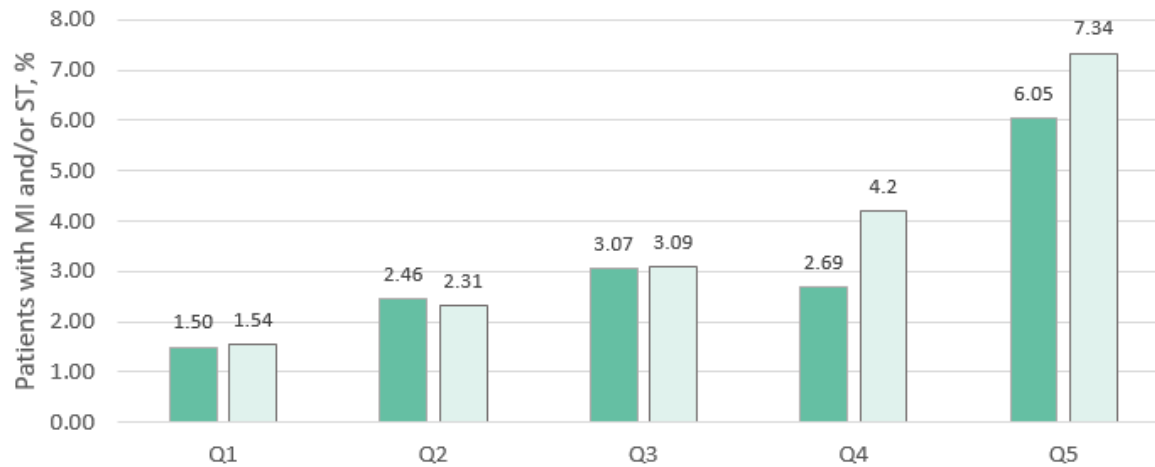
Predicted thrombotic risk versus Predicted bleeding risk



Calibration of the model (Observed vs predicted risks)

Observed and Predicted 1-Year Risk of MI/ST in the Validation Cohort
(Risk Strata Defined by Taking Quintiles of Predicted Risk in eULTIMASTER)

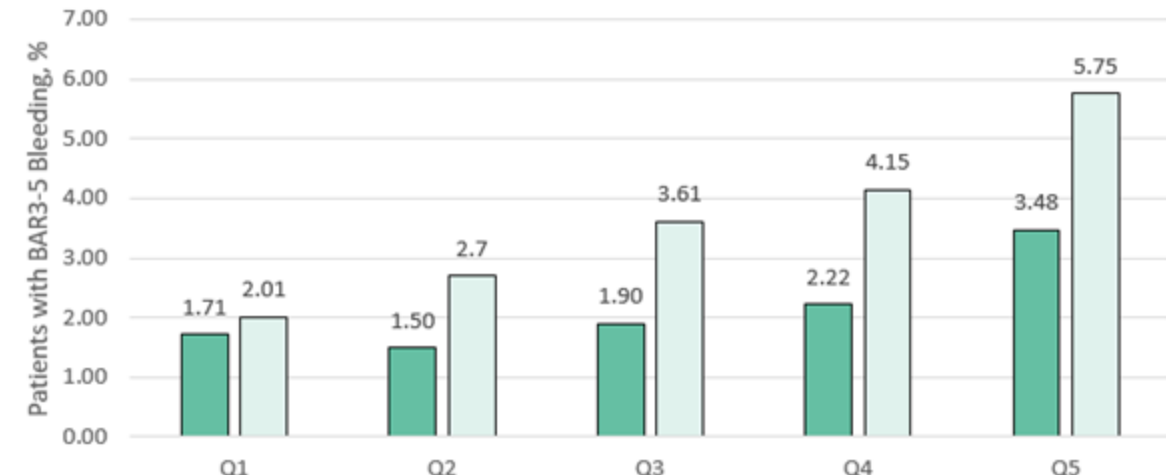
■ Observed □ Predicted



Quintiles Range	1.40 ; 2.07	2.09 ; 2.53	2.62 ; 3.73	3.77 ; 5.06	5.07 ; 24.70
Mean of PR	1.54	2.31	3.09	4.20	7.34
No. of patients	934	1015	1076	968	1024
No. of events	14	25	33	26	62
Obs. Occurrence	1.50	2.46	3.07	2.69	6.05

Observed and Predicted 1-Year Risk of BARC3-5 Bleeding in the Validation Cohort
(Risk Strata Defined by Taking Quintiles of Predicted Risk in eULTIMASTER)

■ Observed □ Predicted



Quintiles Range	1.38 ; 2.09	2.24 ; 2.96	2.97 ; 4.10	4.14 ; 4.40	4.44 ; 12.52
Mean of PR	2.01	2.70	3.61	4.15	5.75
No. of patients	1115	872	1060	951	1019
No. of events	19	13	20	21	35
Obs. Occurrence	1.71	1.50	1.90	2.22	3.48

Discrimination of the model (C-statistic and ROC)

```
proc phreg data=ARC_HBR_MIST concordance
plots=roc rocoptions(at=30 92 366);
model T * status(1,2) = PR_MI ;
run;
```

Number of Observations Read	5017
Number of Observations Used	5017

Summary of the Number of Event and Censored Values

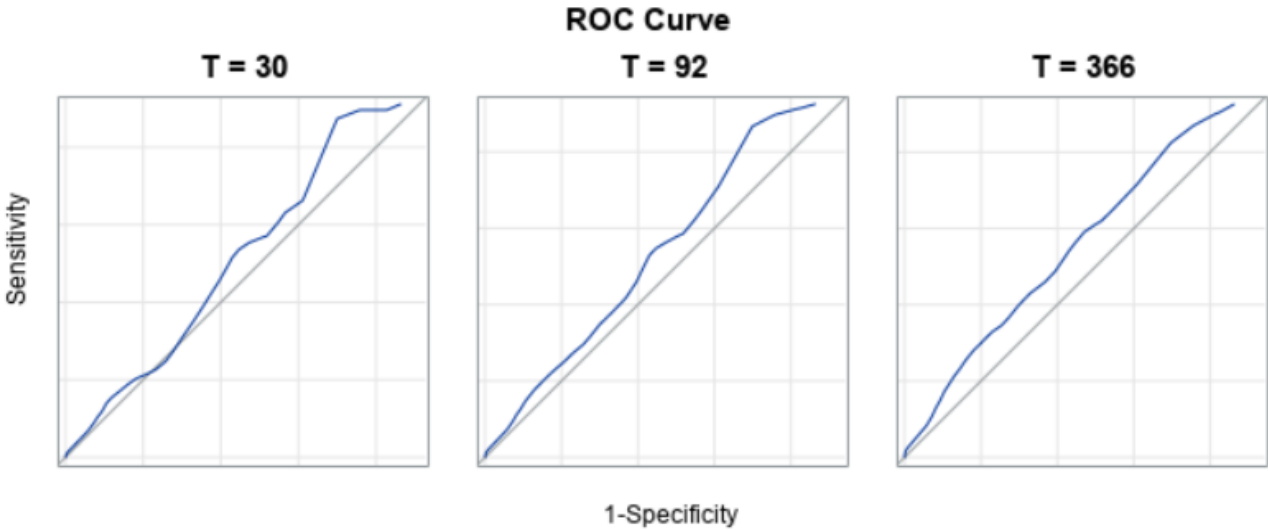
Total	Event	Censored	Percent Censored
5017	160	4857	96.81

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
PR_MI	1	0.17250	0.02132	65.4864	<.0001	1.188

Harrell's Concordance Statistic

Source	Estimate	Comparable Pairs			
		Concordance	Discordance	Tied in Predictor	Tied in Time
Model	0.6348	485215	273958	24596	83

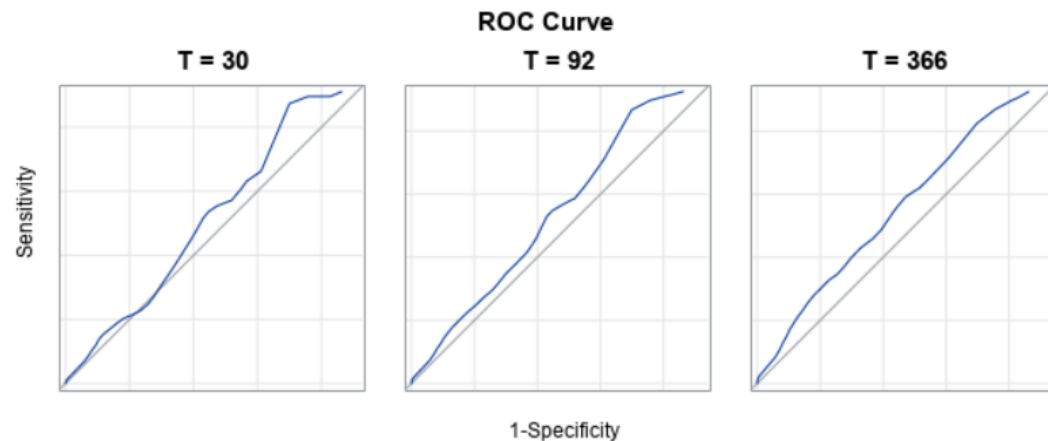


Discrimination of the model (C-statistic and ROC)

- We used these predicted risks to generate 2 Cox models and evaluate the predictive accuracy of these survival models by using Harrell's Concordance Statistic for the HBR subjects (N=5017)

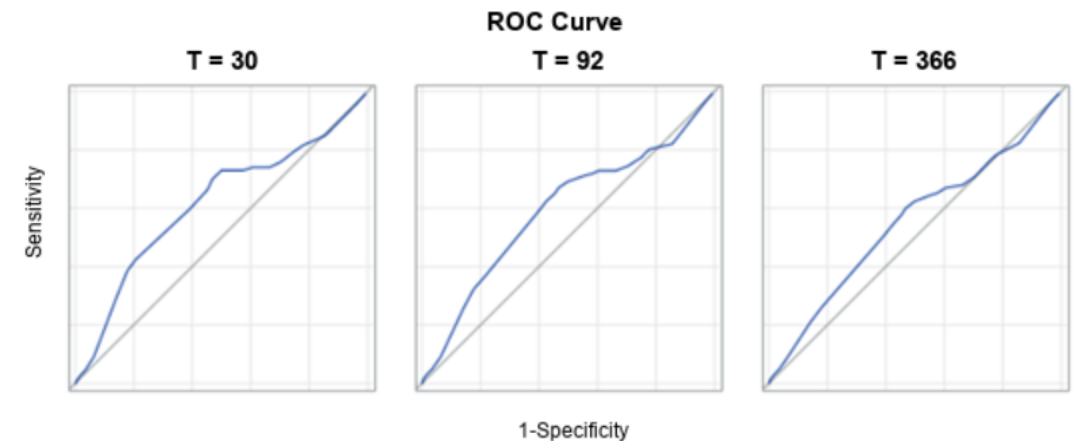
- For Predicted thrombotic risk :

- N= 5017 , event = 160
- c=0.635



- Predicted bleeding risk :

- N= 5017 , event = 108
- c=0.589



Additional PHREG Options

■ Uno's C-Statistic instead of Harrell's C-Statistic

- Harrell's C-statistic (1986) : straightforward approach by discarding the pairs that have become non comparable because of censoring → By default CONCORDANCE = HARRELL
- Uno's C-statistic (2011) : new formulation that models the censoring distribution and uses it to weight the uncensored observations in the estimation, making the estimates censoring-independent
→ Need to specify : CONCORDANCE = UNO

■ Additional Standard Error (SE) calculation

- Based on a perturbation-resampling method proposed by Uno et al. (2011).
- Need to specify CONCORDANCE = HARRELL(SE) or CONCORDANCE = UNO(SE)
- Additional options can control this perturbation process : ITER= , SEED=, DIFF

■ ROC Plot Options

- PHREG procedure use PLOTS=ROC options to produce ROC curves.
Additional ROCOPTIONS can be programmed : AT= , AT=x TO y BY z, AUC, AUCDIFF, IAUC, etc.

Limitations

- **Unconsidered parameters** to define high bleeding risk patients in the external validation cohort of e-ULTIMASTER
- Some **predictors** originally used in the trade-off model were missing in the external validation cohort of e-ULTIMASTER
- Low rate of 1-year bleeding in e-ULTIMASTER
 - No adjudication of bleeding events
 - Possible underreporting of events

Conclusion

Conclusion

- The conventional C-statistic, as implemented in the LOGISTIC procedure, is for binary outcomes. PROC PHREG procedures can be used to calculate equivalent predictive accuracy measures (ROC and C-stat) to validate the survival models
- Additional options can be used in SAS to further analyze it.
- The business case presented shows an example of its use in the field of cardiovascular diseases
- Despite the limitations of the external validation cohort, this analysis shows the practical application and feasibility to externally validate predictive risk models.

References

- Urban P et al. Defining high bleeding risk in patients undergoing percutaneous coronary intervention: a consensus document from the Academic Research Consortium for High Bleeding Risk. *Eur Heart Journal* 2019;**40**:2632–2653
- Urban P et al. Assessing the Risks of Bleeding vs Thrombotic Events in Patients at High Bleeding Risk After Coronary Stent Implantation: The ARC–High Bleeding Risk Trade-off Model. *JAMA Cardiol* 2021;**6**:410–419
- Harrell FE. The PHGLM Procedure. In SUGI Supplemental Library Guide 1986, Version 5 Edition, 1986, SAS Institute Inc., Cary, NC, USA
- Uno H et al. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med* 2011;**30**:1105-17
- Guo C et al. Evaluating Predictive Accuracy of Survival Models with PROC PHREG. Paper SAS462-2017. SAS Institute Inc., Cary, NC, USA
- SAS/STAT® 14.3 User's Guide The PHREG Procedure. 2017, SAS Institute Inc., Cary, NC, USA



Thank you !

Stride Ahead
100th

*Christophe Giot
Christophe Van De Maele
Davide Capodanno
Dragica Paunovic
Kanae Hashimoto
Laurence Seidel
Mamas Mamas
Marie-Claude Morice
Ruth Owen
Shoichi Kuramitsu
Vladimir Borovicanin*

*And all eULTIMASTER study team
and Investigators*