Comparison of existing bleeding risk prediction models to large scale patient-level prediction models

October 12, 2018

Janssen Research & Development

Clair Blacketer MPH, Jenna M. Reps PhD, Patrick B. Ryan PhD, Zhong Yuan MD, PhD

Prediction Problem

Amongst new users of Warfarin or direct oral anticoagulants (DOACs) with prior non-valvular atrial fibrillation, who will develop the outcome of a major bleeding event within a year of the start of treatment?

ATRIA **CHADS$_2$**

TARGET-BLEED **OAC-RISK**

ORBIT **CHA$_2$DS$_2$-VASc**

Methods

Databases Used:

- IBM® MarketScan® Commercial Database (CCAE)
- IBM® MarketScan® Medicare Supplemental Database (MDCR)
- IBM® MarketScan® Multi-state Medicaid Database (MDCD)
- Optum© De-Identified Clinformatics® Data Mart Database – Date of Death (DOD)

Validating the risk scores

The five risk scores were applied to the target populations across the four databases.

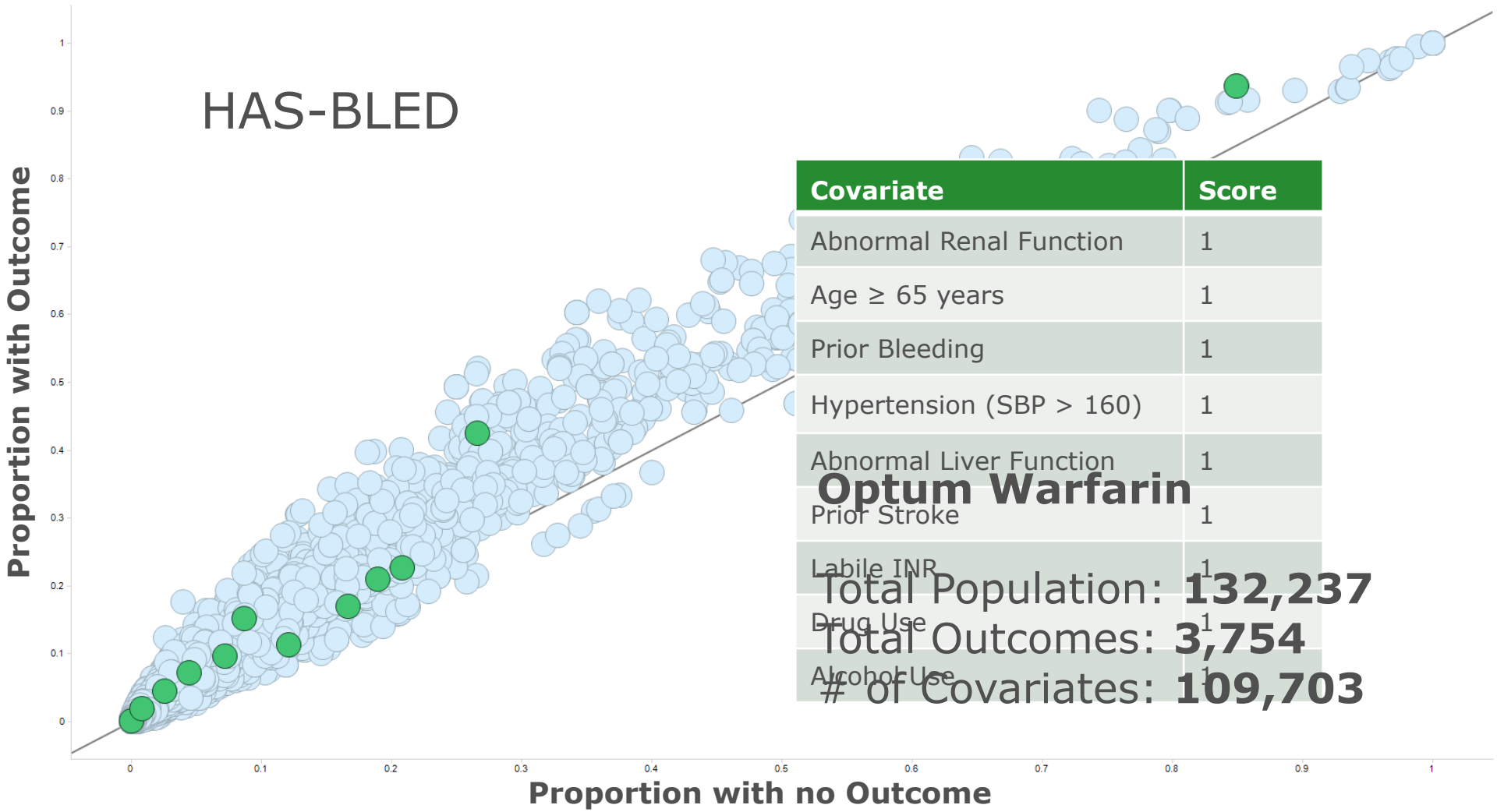
Learning the PLP model

Model: A regularized LASSO logistic regression model was trained separately on the four US claims databases.

Training: 75% of the data from each of the four databases were used to train the models, with 10-fold cross validation implemented to identify the optimal hyper-parameters.

Comparison: Performance of the existing models and learned models were evaluated and compared using the AUC discriminative measure.

Evaluating Candidate Predictors of Bleeding



Covariate	Score
Abnormal Renal Function	1
Age ≥ 65 years	1
Prior Bleeding	1
Hypertension (SBP > 160)	1
Abnormal Liver Function	1
Prior Stroke	1
Labile INR	1
Drug Use	1
Alcohol Use	1

Optum Warfarin

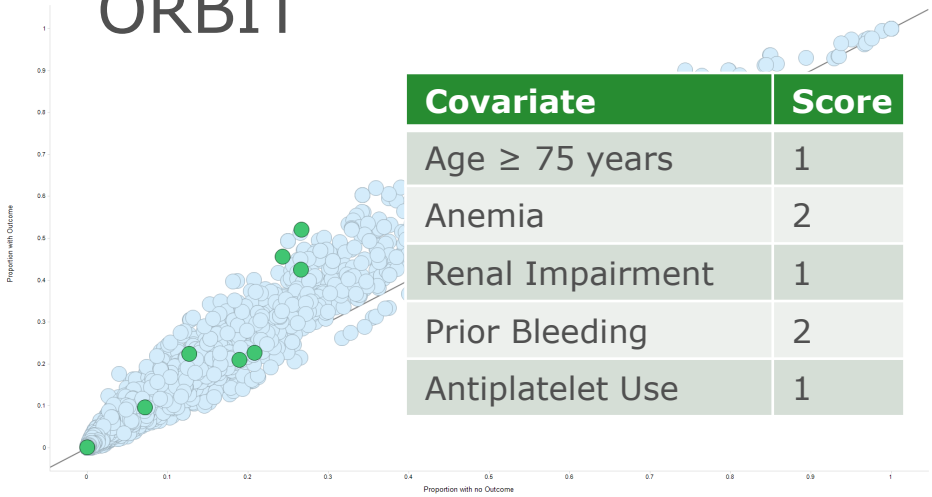
Total Population: **132,237**

Total Outcomes: **3,754**

# of Covariates: **109,703**

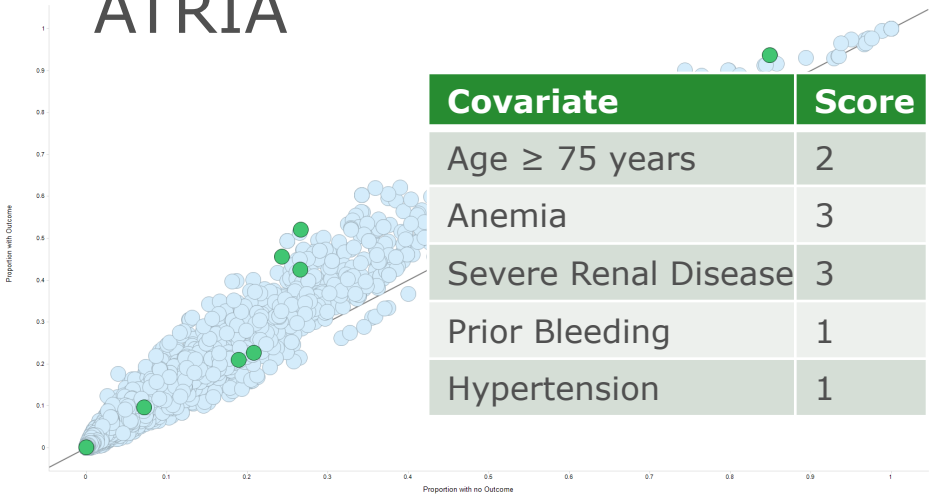
Examining predictors in existing models

ORBIT



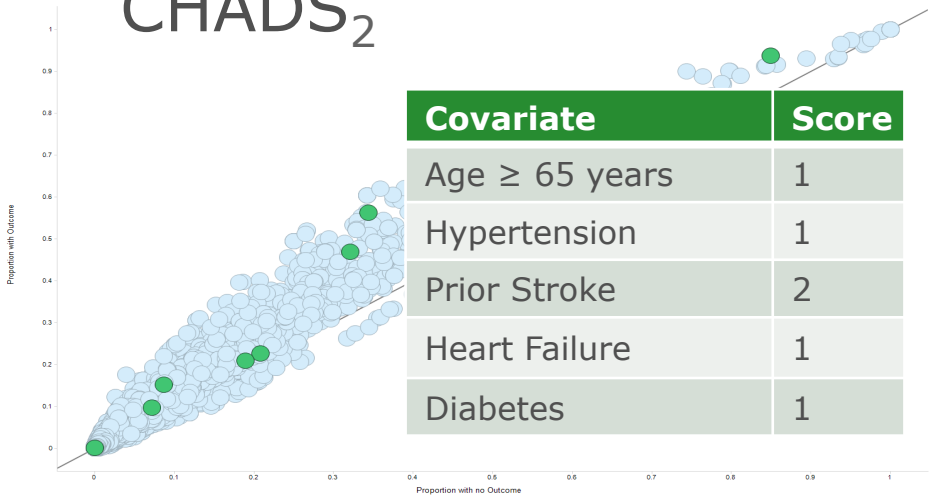
Covariate	Score
Age ≥ 75 years	1
Anemia	2
Renal Impairment	1
Prior Bleeding	2
Antiplatelet Use	1

ATRIA



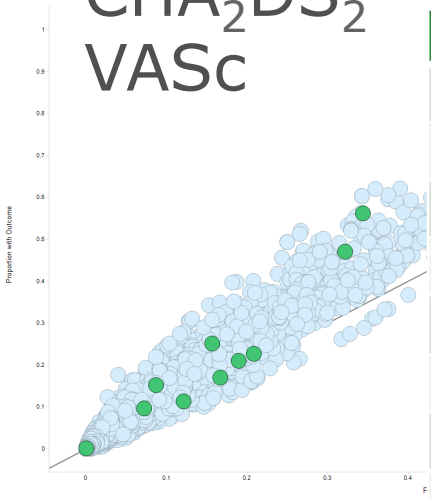
Covariate	Score
Age ≥ 75 years	2
Anemia	3
Severe Renal Disease	3
Prior Bleeding	1
Hypertension	1

$CHADS_2$



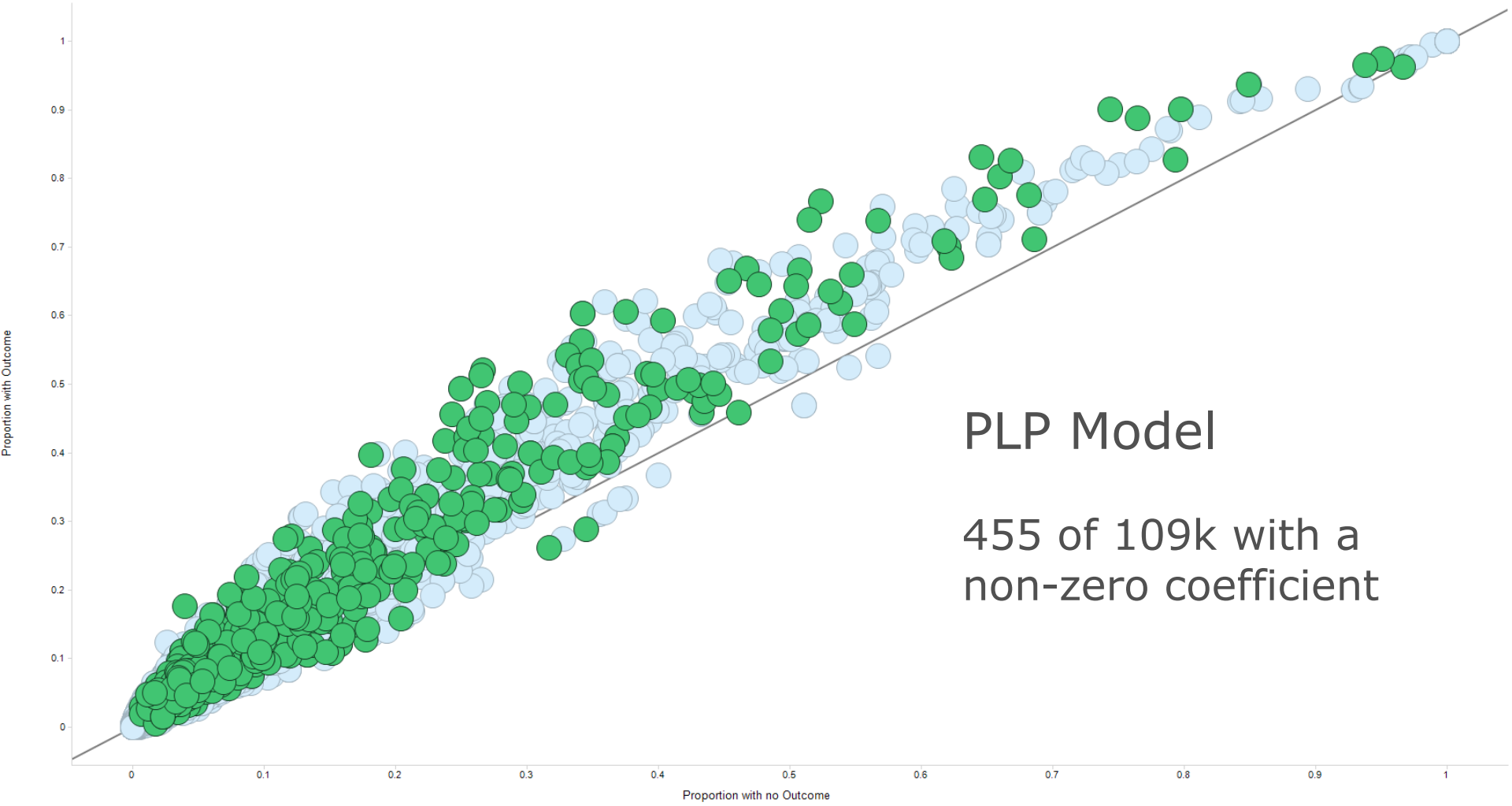
Covariate	Score
Age ≥ 65 years	1
Hypertension	1
Prior Stroke	2
Heart Failure	1
Diabetes	1

CHA_2DS_2-VASc



Covariate	Score
Age ≥ 65 years	1
Age ≥ 75 years	1
Hypertension	1
Prior Stroke	2
Heart Failure	1
Diabetes	1
Vascular Disease	1
Female Gender	1

What Covariates did LASSO Choose?



PLP Model

455 of 109k with a non-zero coefficient

PLP Model vs Risk Scores

PLP Model

Sensitivity: 50%

At-risk threshold: 4%

Specificity: 83%

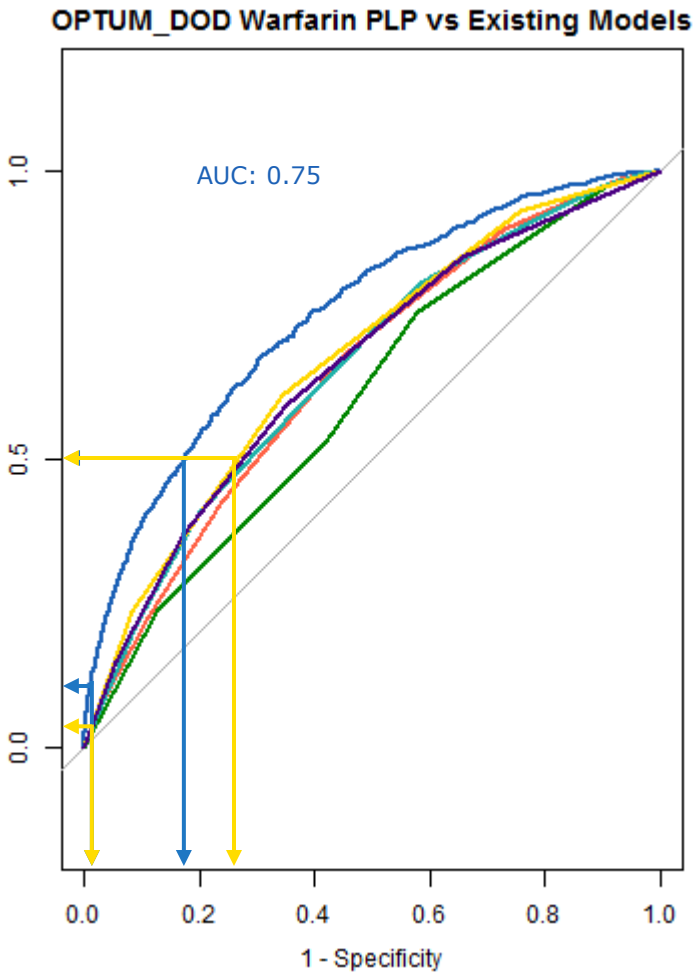
+PV: 8%

Specificity: 99%

At-risk threshold: 18%

Sensitivity: 8%

+PV: 22%



PLP Model: 0.75

HAS-BLED

At-risk threshold: 2.2%

Specificity: 65%

+PV: 4.5%

At-risk threshold: 6.4%

Sensitivity: 0.63%

+PV: 6.6%

.68

Sc: 0.66

5

PLP Model Performance

The Optum Warfarin model for predicting major bleed achieved an internal validation AUC of 0.75 (0.73-0.76)

Figure 1: ROC plot for the Optum Warfarin model

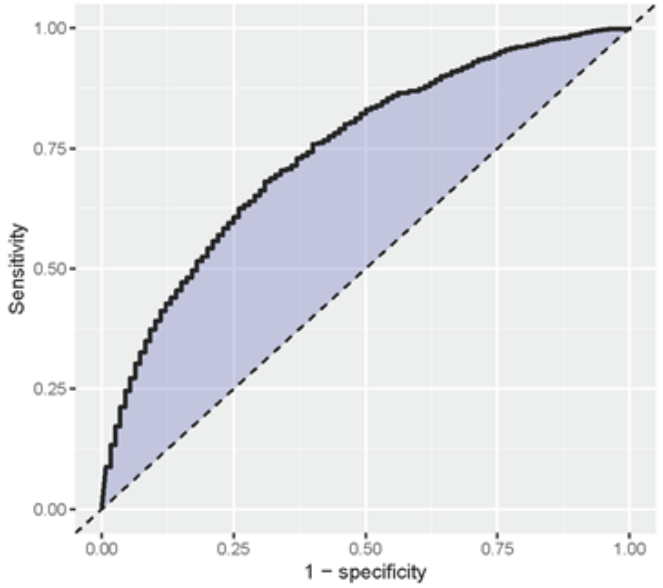


Figure 2: Calibration plot for the Optum Warfarin model

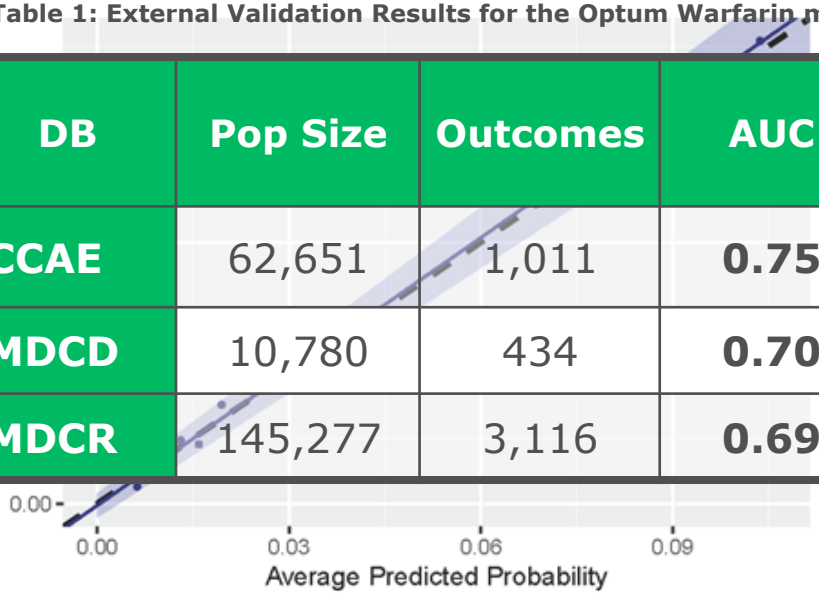


Table 1: External Validation Results for the Optum Warfarin model

DB	Pop Size	Outcomes	AUC
CCAE	62,651	1,011	**0.75**
MDCD	10,780	434	**0.70**
MDCR	145,277	3,116	**0.69**

Conclusion

PLP models developed using the PLP framework to predict major bleeding **consistently outperform** risk scores currently used in clinical practice.

The **patient-level prediction framework offers an efficient way to train, test, and validate large-scale prediction models** across a network of observational databases.