Measuring the Prediction Performance for Risk Scores in the Era of Clinical Preventive Care

Yan Yuan, PhD
Assistant Professor
School of Public Health, University of Alberta
2017 Lifetime Data Science Conference

Outline

- Motivation
  - The clinical preventative care focuses on earlier intervention through *personalized* risk prediction
- Measures for evaluating prediction performance of risk scores
- Simulation study
- Data analysis example
- Summary and future work

Examples of Prevention and Early Detection in Clinical Practice

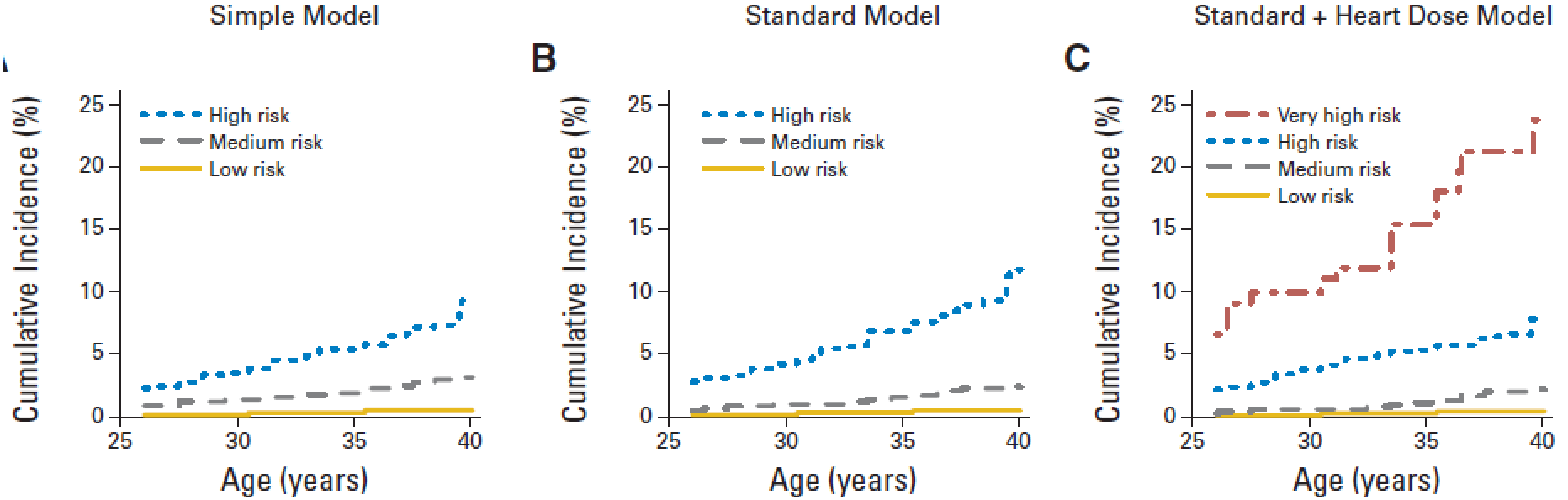
- Framingham risk score for CVD in general population
- CHA_2DS_2-Vasc scores for stroke risk in patients with atrial fibrillation
- Multiple risk score systems (n>40) for diabetes risk in general population
- BIRADS scores for breast cancer early detection

Risk Score as a Screening Tool

- Typical condition that risk scores are used/ developed for have the following characteristics
  - seriousness may result in a high risk of mortality or significantly affect the quality of life;
  - early detection/intervention can make a difference in disease prognosis;
  - the event rate is <u>low</u>

Motivating Data

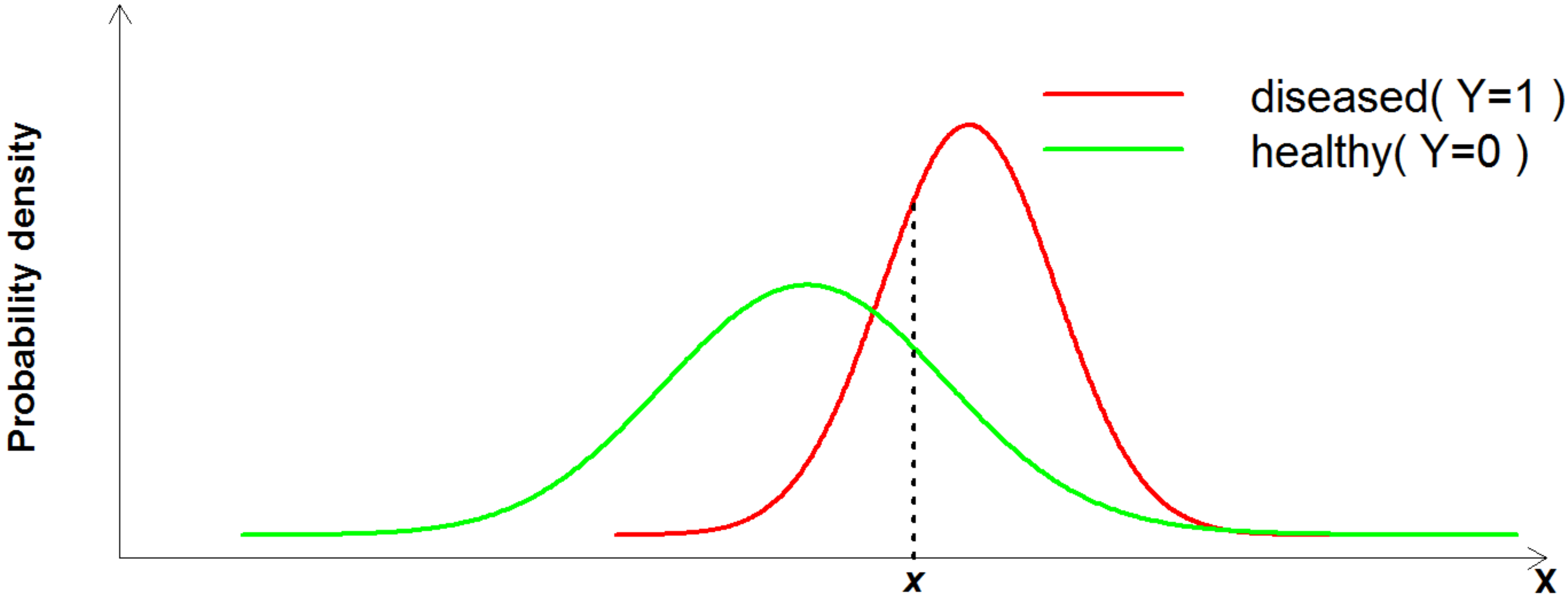
- Late effects of cancer treatments in childhood cancer survivors – e.g. Congestive heart failure (Chow et al. JCO, 2015)
- Cumulative risk of CHF is ~3% by 35 years post diagnosis



Evaluating Model Performance for Predicting Rare Events

- Threshold Dependent Measure (predictor needs to be binary)
  - ~~Misclassification rate~~
  - Sensitivity (TPF): P(test positive | diseased) = $\mathrm{P}(\hat{Y} = 1 \,|Y = 1)$
  - Specificity (FPF): P(test negative | healthy) = $\mathrm{P}(\hat{Y} = 0 \,|Y = 0)$
  - Positive Predictive value (PPV): $\mathrm{P}(Y = 1|\hat{Y} = 1)$
  - Negative Predictive Value (NPV): $\mathrm{P}(Y = 0|\hat{Y} = 0)$

How about when predictor is continuous or ordinal?



Risk score

Threshold Independent Measure

- Area Under the ROC* Curve (AUC, *aROC*)

$$\text{AUC} \equiv \int_0^1 \text{TPF}(s)\text{dFPF}(s)$$

- Extension to event status to accommodate censoring and time to event data -- AUC_{t_0}
- Criticisms of AUC as a measure for risk prediction
  - Retrospective measure
  - Insensitive
  - Over-optimistic

*Receiver Operating Characteristic

Alternatives to AUC_{t_0} for Time-to-event Outcome

- Time-dependent PPV – PPV_{t_0}
  - Needs binary predictor or equivalently a threshold for continuous / ordinal predictor
- Time-dependent Average Positive predictive value (AP_{t_0})

$$\mathrm{AP}_{t_0} = \int_{\mathcal{R}} \mathrm{PPV}_{t_0}(z) d\mathrm{TPF}_{t_0}(z).$$

Note that AP_{t_0} is Threshold Independent

Nonparametric Estimator

Let (X, δ, Z) be the standard survival time notation,
X: the censored event time, δ: the censoring indicator
Z: the risk score

$$\widehat{\mathrm{AP}}_{t_0} = \frac{\sum_{j=1}^{n} I(X_j \leq t_0)\hat{w}_{t_0,j}\widehat{\mathrm{PPV}}_{t_0}(Z_j)}{\sum_{j=1}^{n} I(X_j \leq t_0)\hat{w}_{t_0,j}}.$$

where

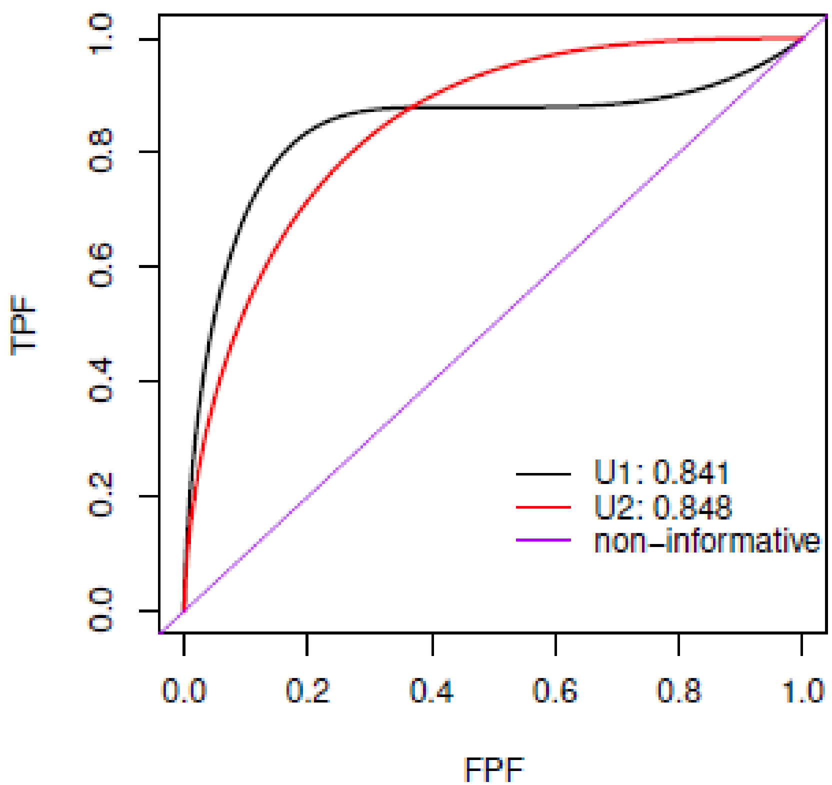
$$\widehat{w}_{t_0,i} = \frac{I(X_i < t_0)\delta_i}{\widehat{\mathcal{G}}(X_i)} + \frac{I(X_i \geq t_0)}{\widehat{\mathcal{G}}(t_0)}$$

$$\widehat{\mathrm{PPV}}_{t_0}(z) = \frac{\sum_{i=1}^{n} \widehat{w}_{t_0,i} I(Z_i \geq z) I(X_i < t_0)}{\sum_{i=1}^{n} I(Z_i \geq z)}$$

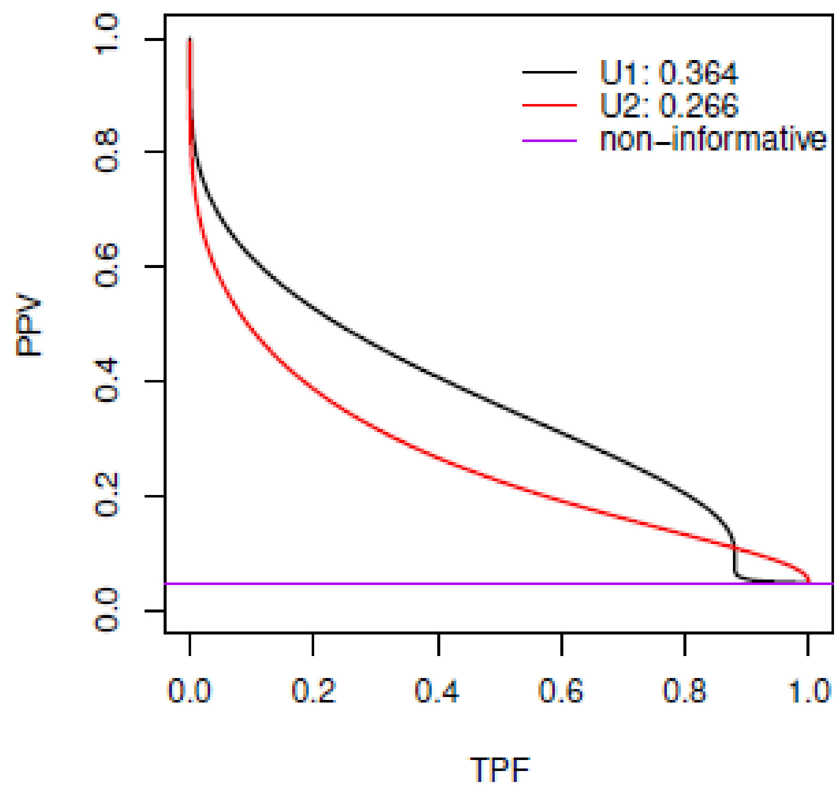
Simulation Study

$$\log(T_i) = 7.2 - 1.1U_{i1} - 2.5U_{i2} - 1.5log(U_{i1}^2) + \epsilon_T,$$

$ROC_{t_0=8}$



$PR_{t_0=8}$



Results (n=2000)

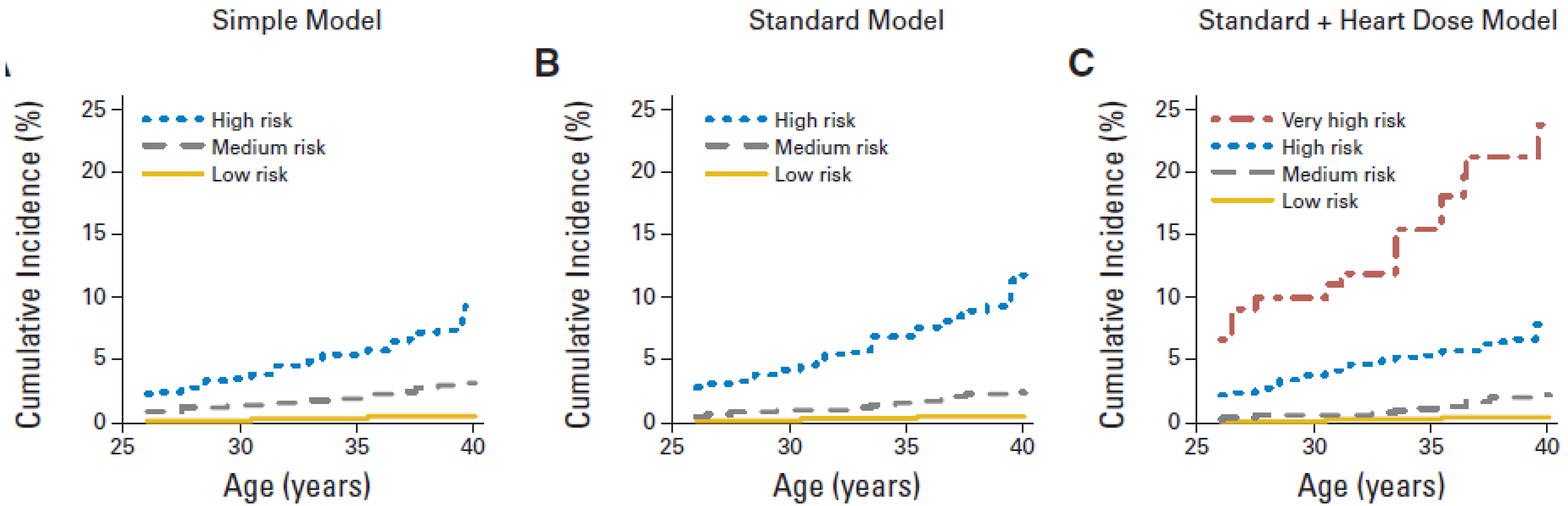
t_0	Event rate		TRUE	BIAS	ESE	ASE^b	$ECOVP^b(\%)$
0.5	0.0101	AP_1	0.182	0.0365	0.0810	0.0795	92.3
		AP_2	0.124	0.0339	0.0689	0.0678	93.0
		rAP	1.47	0.4890	1.5300	1.7600	95.1
8	0.0495	AP_1	0.364	0.0096	0.0527	0.0516	92.5
		AP_2	0.266	0.0129	0.0452	0.0450	93.4
		rAP	1.37	0.0140	0.3290	0.3320	95.7
36	0.0991	AP_1	0.462	0.0098	0.0534	0.0558	95.9
		AP_2	0.375	0.0118	0.0493	0.0501	94.5
		rAP	1.23	0.0135	0.2310	0.2420	94.9

where $rAP_{t_0} = \frac{AP_{u_1,t_0}}{AP_{u_2,t_0}}$

Results (n=5000)

t_0	Event rate		TRUE	BIAS	ESE	ASE^b	$ECOVP^b(\%)$
0.5	0.0101	AP_1	0.182	0.0185	0.0500	0.0504	93.1
		AP_2	0.124	0.0155	0.0416	0.0417	94.8
		rAP	1.47	0.1550	0.7060	0.7600	93.8
8	0.0495	AP_1	0.364	0.0042	0.0337	0.0333	92.9
		AP_2	0.266	0.0049	0.0291	0.0288	93.7
		rAP	1.37	0.0060	0.2160	0.2100	95.4
36	0.0991	AP_1	0.462	0.0034	0.0354	0.0346	95.5
		AP_2	0.375	0.0037	0.0310	0.0313	94.1
		rAP	1.23	0.0051	0.1490	0.1510	95.0

CCSS CHF risk prediction

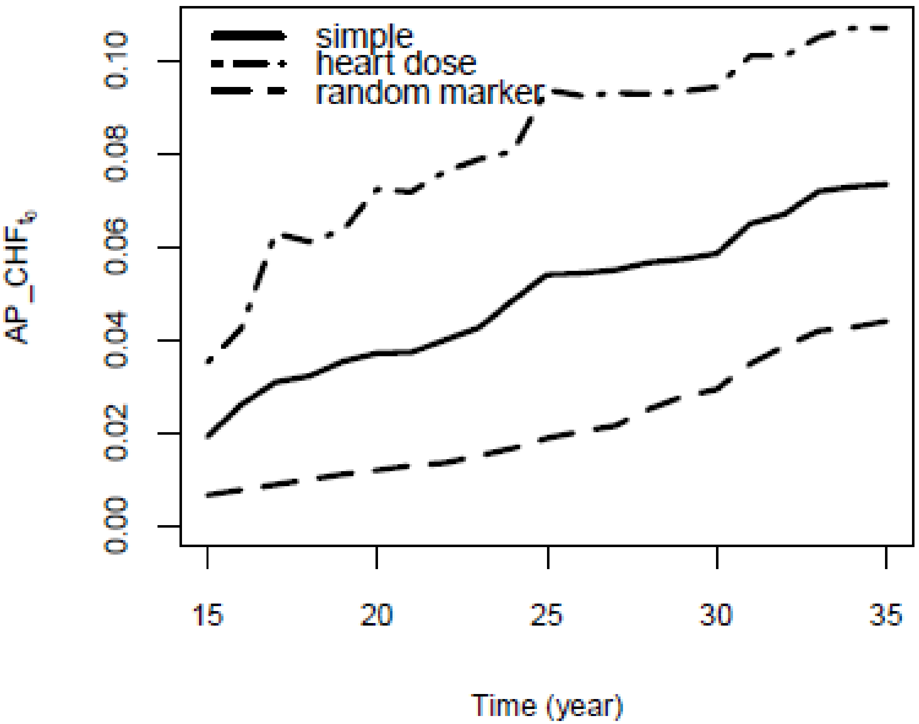


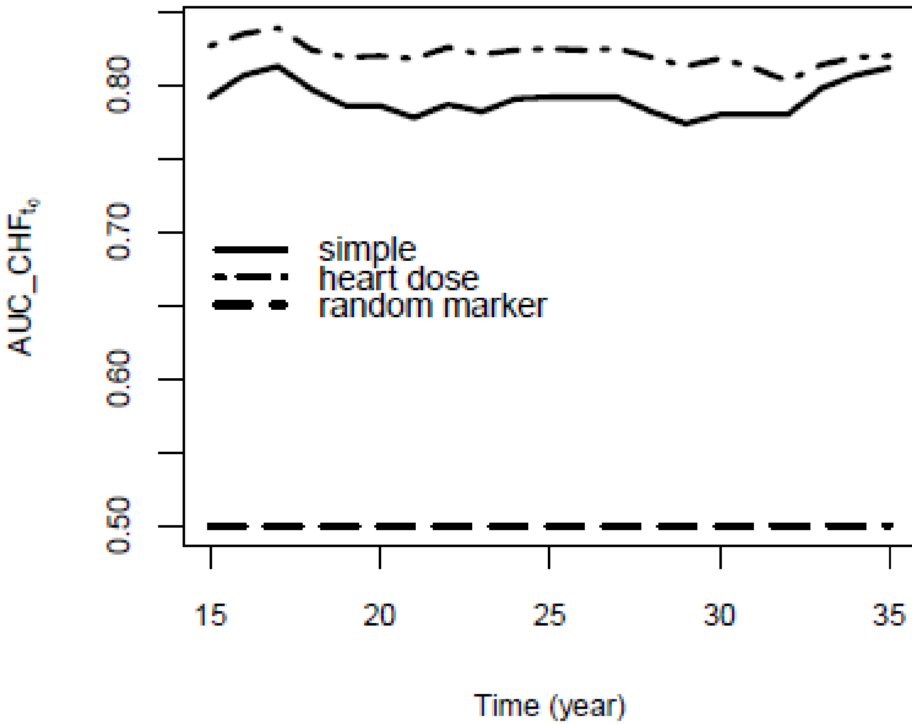
$$\mathrm{PPV}_{t_0}^{\mathrm{CHF}}(z) = Pr\{T < t_0, \Delta = 1 \mid Z \geq z\} \quad \text{and} \quad \mathrm{TPF}_{t_0}^{\mathrm{CHF}}(z) = Pr\{Z \geq z \mid T < t_0, \Delta = 1\}.$$

$$\widehat{\mathrm{PPV}}_{t_0}^{\mathrm{CHF}}(z) = \frac{\sum_{i=1}^{n} \widehat{w}_{t_0,i} I(Z_i \geq z) I(X_i < t_0) I(\Delta_i = 1)}{\sum_{i=1}^{n} I(Z_i \geq z)}$$

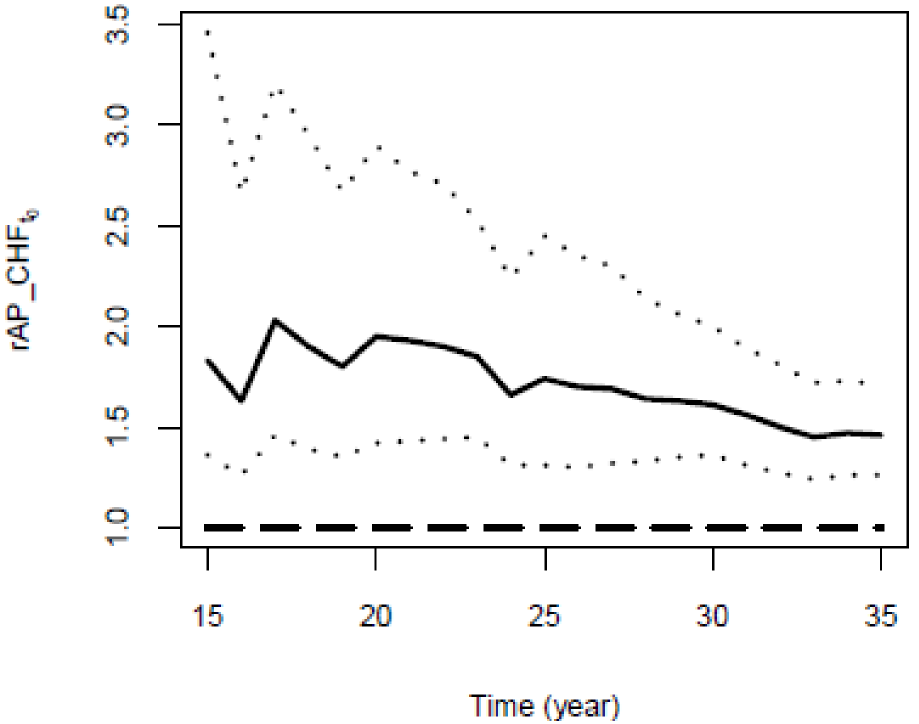
$$\widehat{\mathrm{TPF}}_{t_0}^{\mathrm{CHF}}(z) = \frac{\sum_{i=1}^{n} \widehat{w}_{t_0,i} I(Z_i \geq z) I(X_i < t_0) I(\Delta_i = 1)}{\sum_{i=1}^{n} \widehat{w}_{t_0,i} I(X_i < t_0) I(\Delta_i = 1)}$$

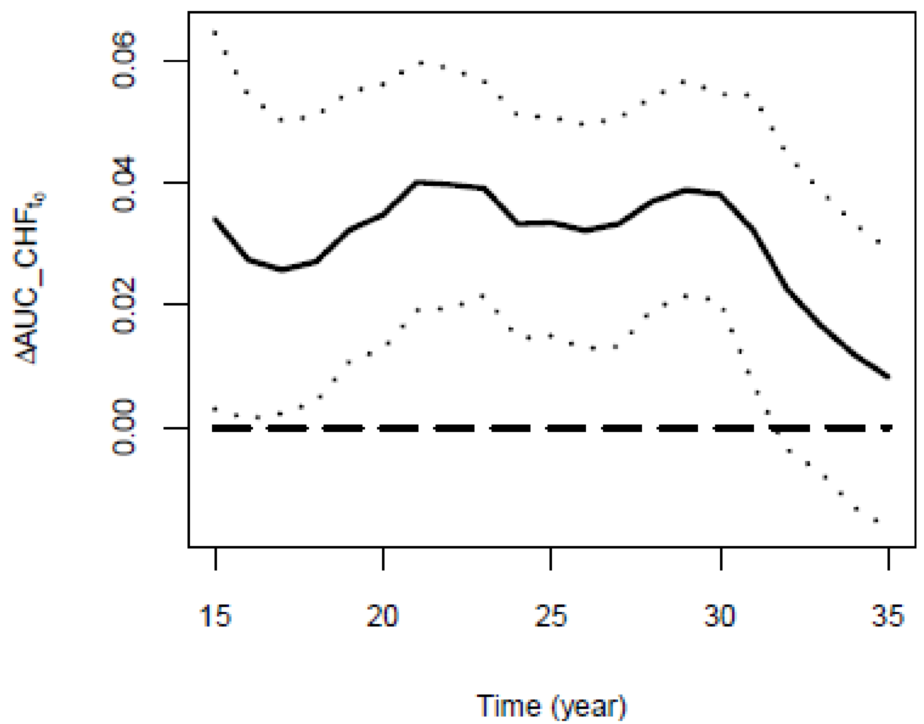
AP_{t_0} vs. t_0 | AUC_{t_0} vs. t_0





Comparison using rAP and ΔAUC





t_0	Event Rate	Risk Score System	AP_{t_0}	AUC_{t_0}
35 years	0.0440	Simple	0.073 (0.062, 0.088)	0.812 (0.778, 0.846)
		Heart dose	0.107 (0.088, 0.135)	0.820 (0.784, 0.856)
		Comparison	1.46 (1.26, 1.71)	0.008 (-0.016, 0.029)

Summary

Contributions

- Nonparametric estimator of AP_{t_0} for censored event status and in the presence of competing risks
- Inference procedure to compare AP_{t_0} for two risk scores
- APtools: an R package for binary and survival time data

Discussion

- AP is a <u>single numerical measure</u>, in this respect it is similar to AUC.
- A summary measure of positive predictive value, better suited in comparing prospective prediction performance of competing risk scores
- More sensitive than AUC as illustrated by the data analysis
- Event rate dependent, AP should be estimated in a prospective cohort or population-based study

Future Work

- Incremental value of biomarkers in risk prediction model as evaluated by AP
- Evaluating the sensitivity of AP with simulated biomarkers that have moderate effect size and are considered clinically significant
- Partial AP

Acknowledgement

Collaborators

- Dr. Qian Michelle Zhou
- Dr. Eric Chow
- Dr. Greg Armstrong

Students

- Doris Li
- Hengrui Cai

CIHR IRSC

Canadian Institutes of Health Research
Instituts de recherche en santé du Canada