UNIVERSITY OF ALBERTA
SCHOOL OF PUBLIC HEALTH

This is PUBLIC HEALTH

From Cancer Epidemiology to Prediction Accuracy Measure, and Beyond - a biostatistician's eight years in population health

Yan Yuan, PhD

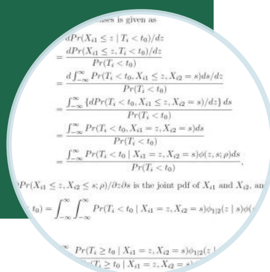
School of Public Health, U Alberta

April 30, 2019 @U Ottawa

Research Program

- Prediction measures and risk prediction
- Gestational weight trajectory modelling
- Relation between association measures

NSERC Theme



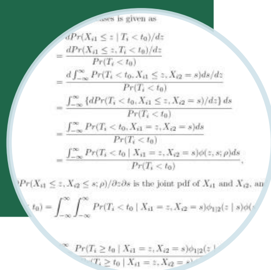
- Health services
- Late effects in survivors
- Rare cancer

CIHR Theme

Research Program

- Prediction measures and risk prediction
- Gestational weight trajectory modelling
- Relation between association measures

NSERC Theme



- Health services
- Late effects in survivors
- Rare cancer

CIHR Theme (Cancer Epi)

Health Services: Breast cancer screening and diagnostic care

- Use administrative data, e.g. physician claims, to identify mode of cancer detection (screen vs. non-screen)? *(Yuan et al. 2015)*
- Time to diagnosis in the two pathways? Any care disparities? *(Yuan et al. 2016)*
- Resolution of cancer screening and rescreening behavior? *(Shen et al. 2018)*
- Quality of breast cancer screening? *(Yuan et al., under review)*
- Breast cancer screening/diagnostic care across Canadian provinces? *(Winget et al., under review)*

Alberta Breast Cancer Screening Program (ABCSP)

Started 2008

12%

88%

Screen Test

- Two clinics: Edmonton, Calgary.
- Mobile units visit rural/remote communities
- Interpreted by sessional radiologists in Edmonton

Fee-for-service Radiologists in Community Practices

Spread through province

Data Sources

- Patient ID
- Demographics
- Tumour details
- Date of cancer diagnosis
- Method of diagnosis

Alberta Cancer Registry

(2007-10)

Database A

- Patient ID
- Date/results of screening and diagnostic mammograms
- Date/results of breast ultrasound, MRI and biopsy

Alberta Society of Radiologists

(2006-10)

Database A & B

- Patient ID
- Date/results of screening mammograms
- Date/results of diagnostic mammogram, breast ultrasound, MRI and biopsy

Screen Test

(2006-10)

Database B

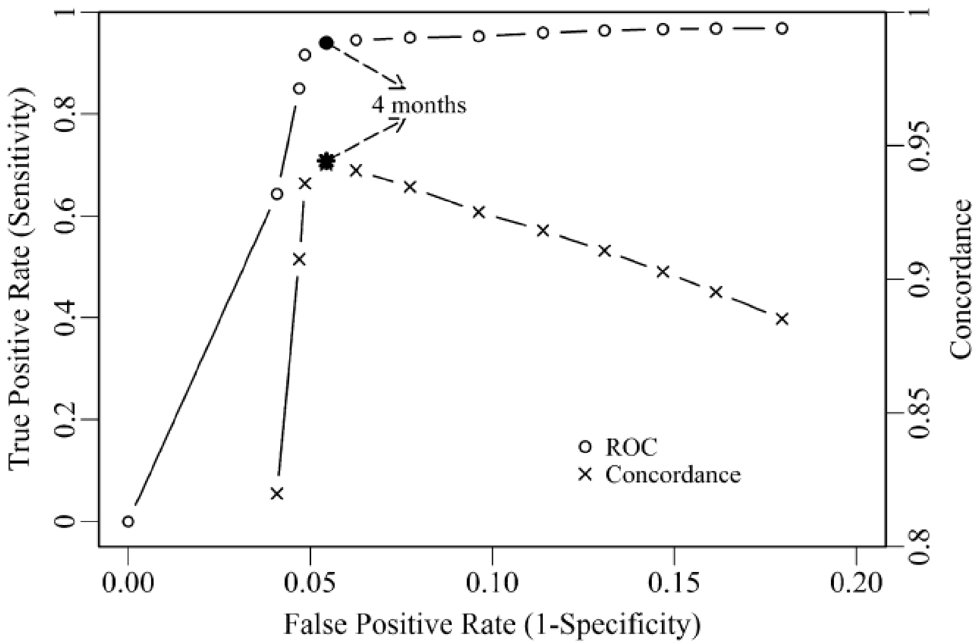
- Patient ID
- Date of screening and diagnostic mammograms
- Date of breast ultrasound, MRI and biopsy

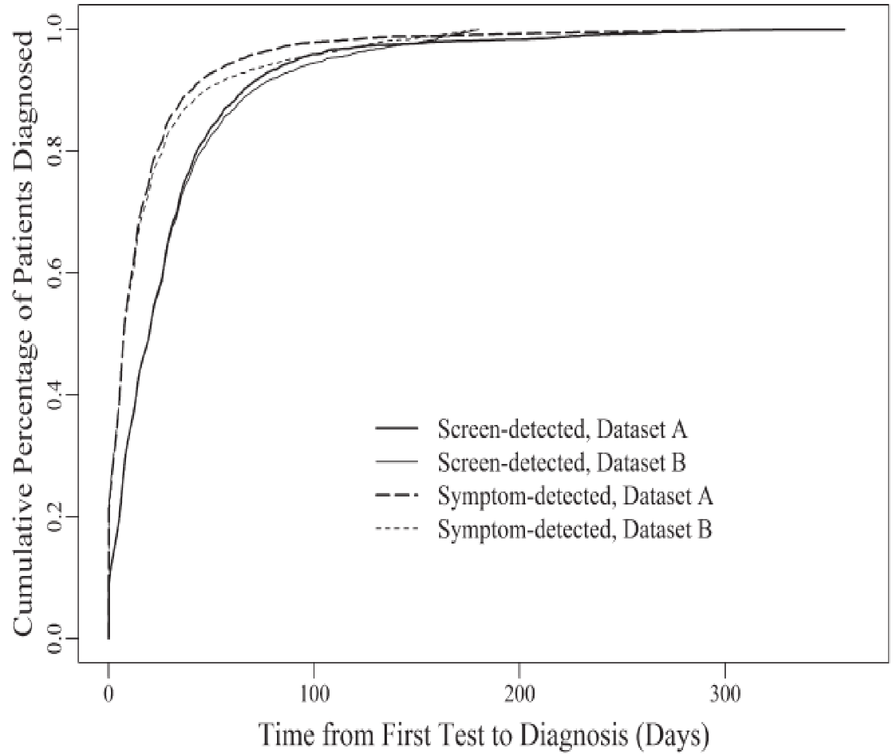
Physician Claims

(2006-10)

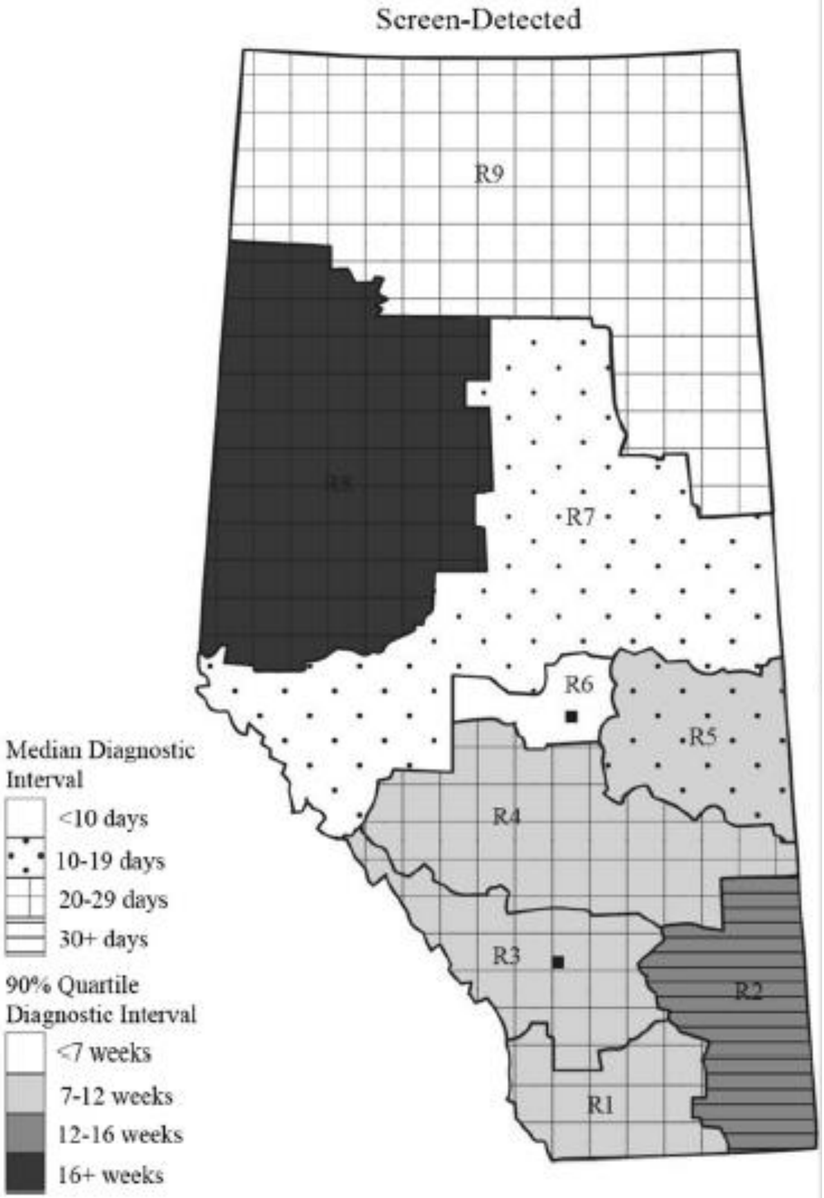
Dataset B: no test results		Dataset A: with test results		
Look-back time		Screen detected n (%)	Symptom detected n (%)	Total n (%)
4 months	Screen detected	2893 (41)	213 (3)	3106 (44)
	Symptom detected	186 (3)	3702 (53)	3888 (56)
6 months	Screen detected	2925 (42)	303 (4)	3228 (46)
	Symptom detected	154 (2)	3612 (52)	3766 (54)
Total		3079 (44)	3915 (56)	6994

Yuan et al. 2015

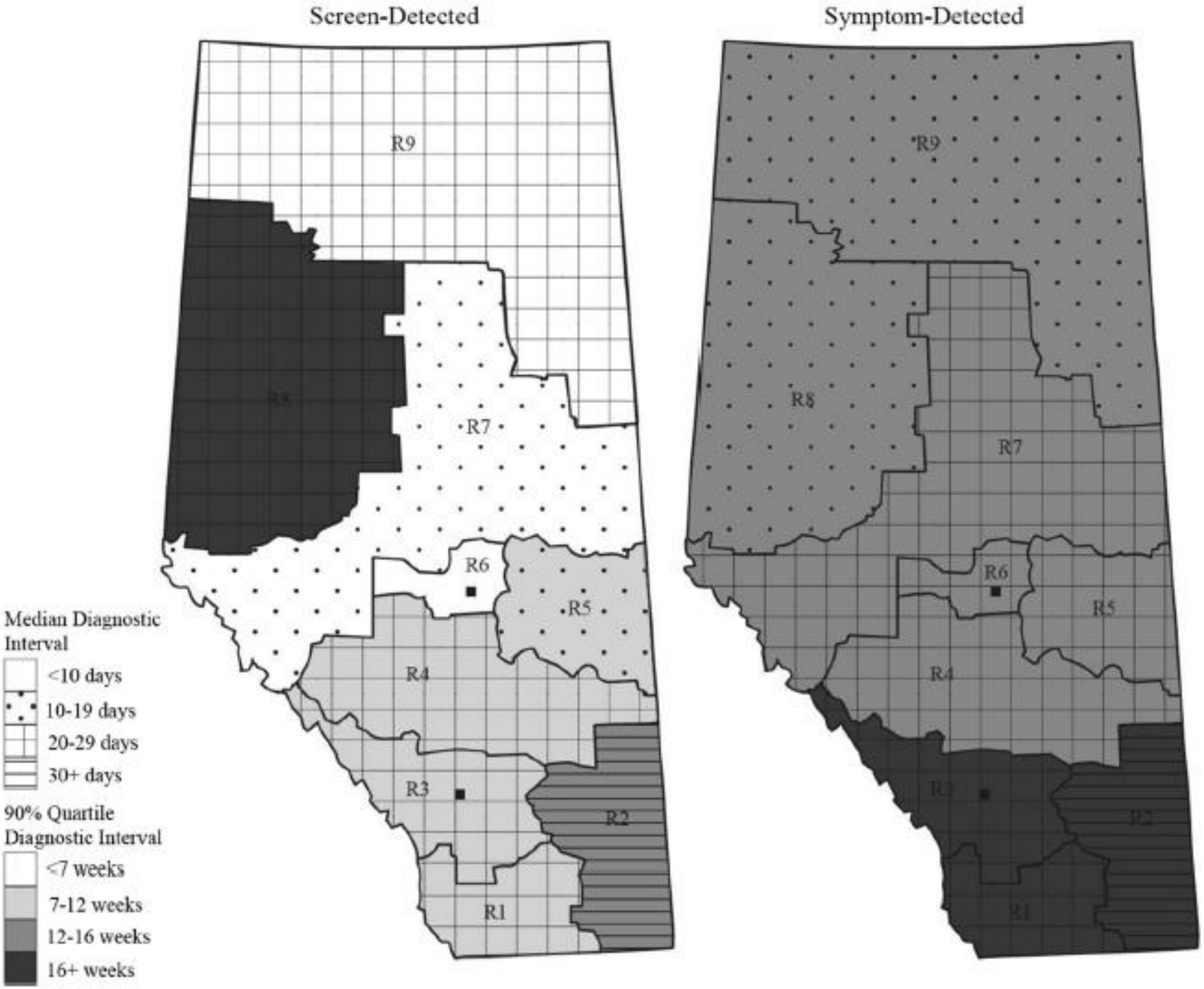




Yuan et al. 2015

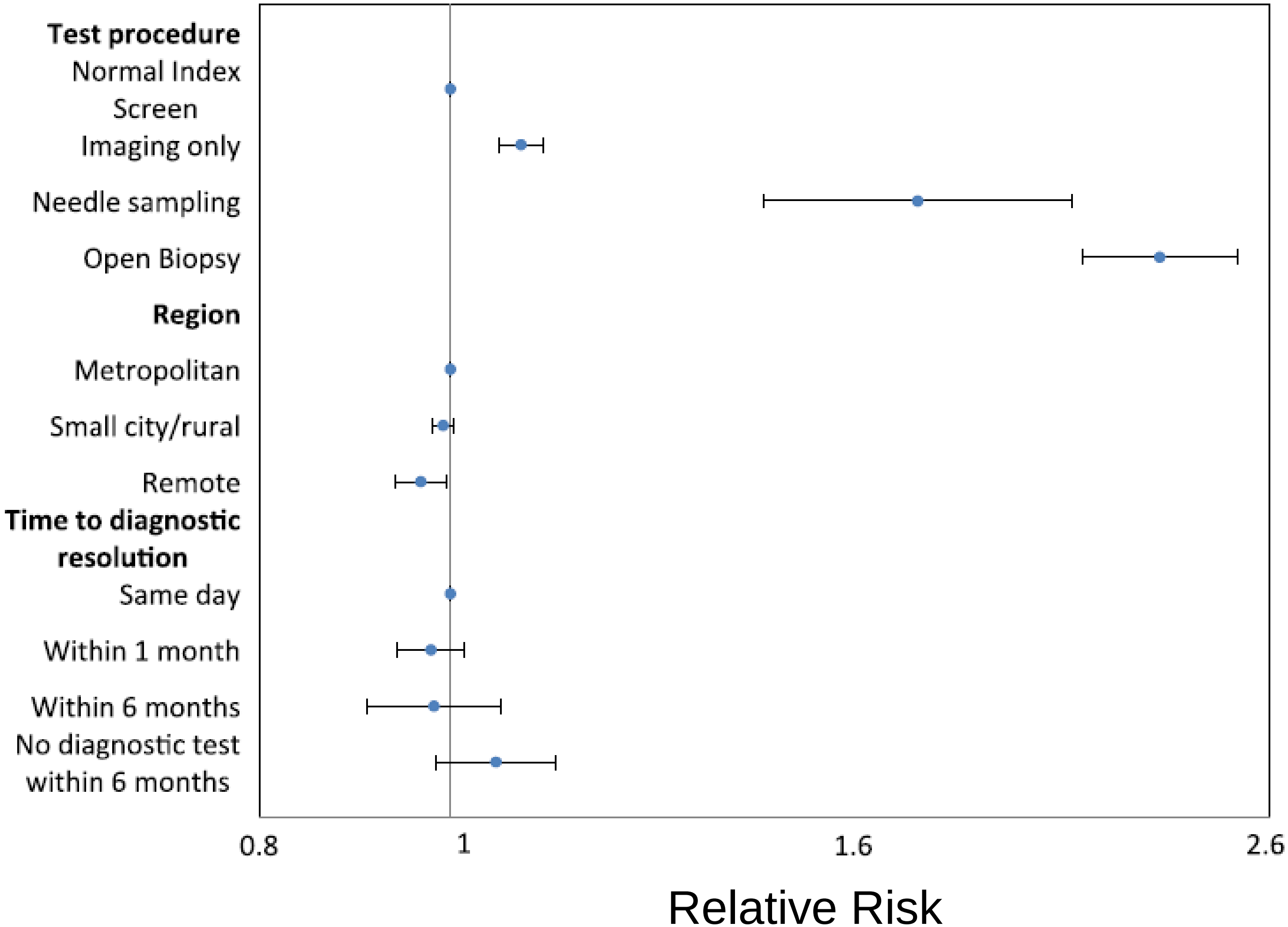


Yuan et al. 2016



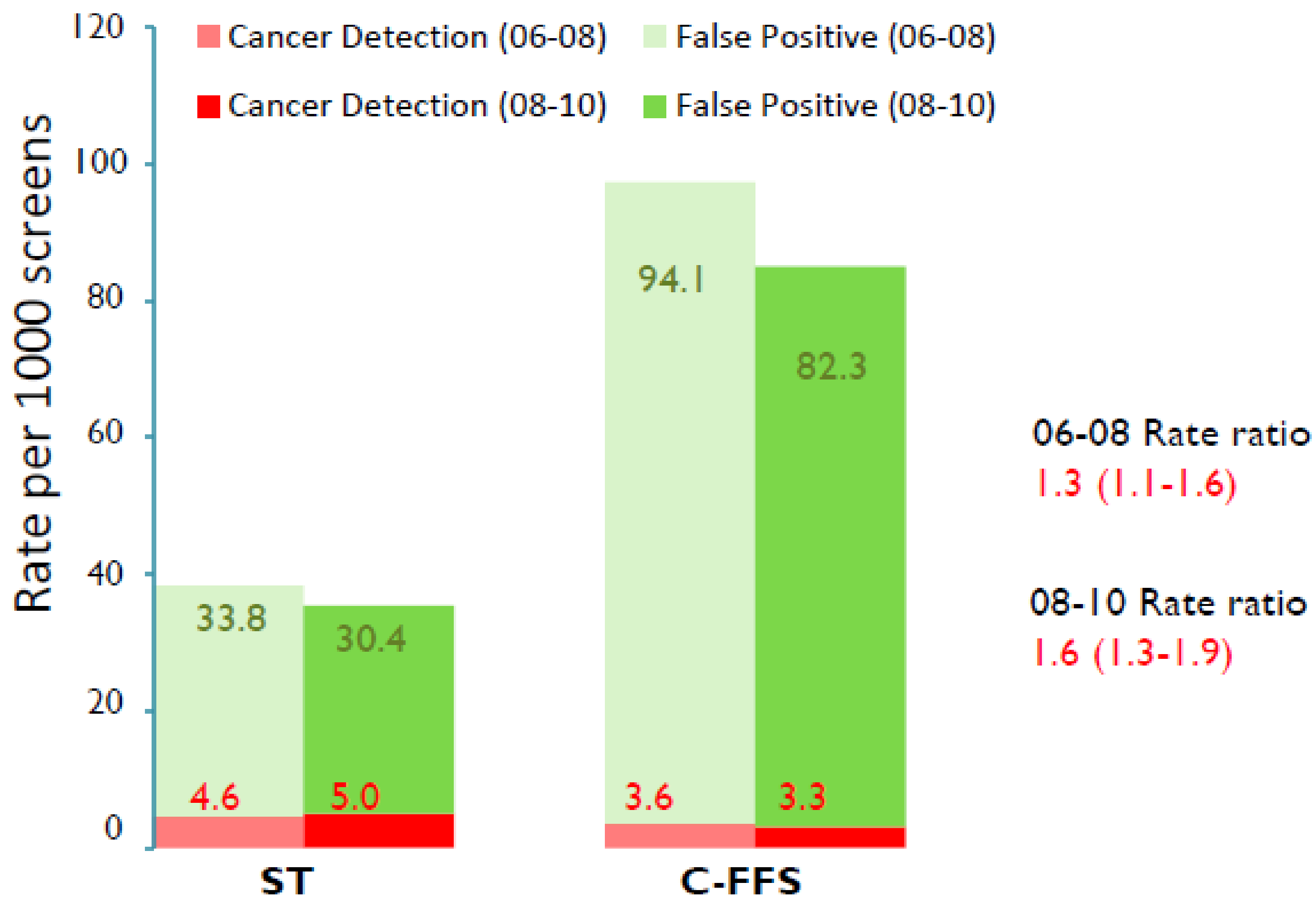
Yuan et al. 2016

Failure to rescreen



Shen et al. 2018

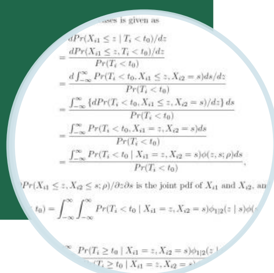
Quality of Breast Cancer Screening



Research Program

- Prediction measures and risk prediction
- Gestational weight / fat trajectory
- Relation between association measures

NSERC Theme



- Health services
- Late effects in survivors
- Rare cancer

CIHR Theme (Cancer Epi)

Motivating Data – Binary outcome

Digital Mammography Imaging Screening Trial (Pisano et al. 2005 *New England Journal of Medicine*)

Malignancy score		7	6	5	4	3	2	1	Total
Digital M	Category Total	11	29	69	1061	2224	6588	32588	42570
	Cancers	10	18	25	85	49	25	122	334
Film M	Category Total	17	29	70	942	2291	6910	32486	42745
	Cancers	13	24	25	74	35	33	131	335

42,760 screening participants underwent two screening technology, 335 were diagnosed with breast cancer by the end of 15 months follow-up.

Performance Evaluation

Predicting Low Prevalence Events

- Threshold Dependent Measure (predictor needs to be binary)
  - ~~Misclassification rate~~
  - Sensitivity (TPF): P(test positive | disease present) = $\mathrm{P}(\hat{Y} = 1 \mid Y = 1)$
  - ~~Specificity (FPF): P(test negative | disease absent) =~~ $\cancel{\mathrm{P}(\hat{Y} = 0 \mid Y = 0)}$
  - Positive Predictive value (PPV): $\mathrm{P}(Y = 1 | \hat{Y} = 1)$
  - ~~Negative Predictive Value (NPV):~~ $\cancel{\mathrm{P}(Y = 0 | \hat{Y} = 0)}$
  - F1 measure: $\dfrac{2}{\frac{1}{\mathrm{PPV}} + \frac{1}{\mathrm{TPF}}}$

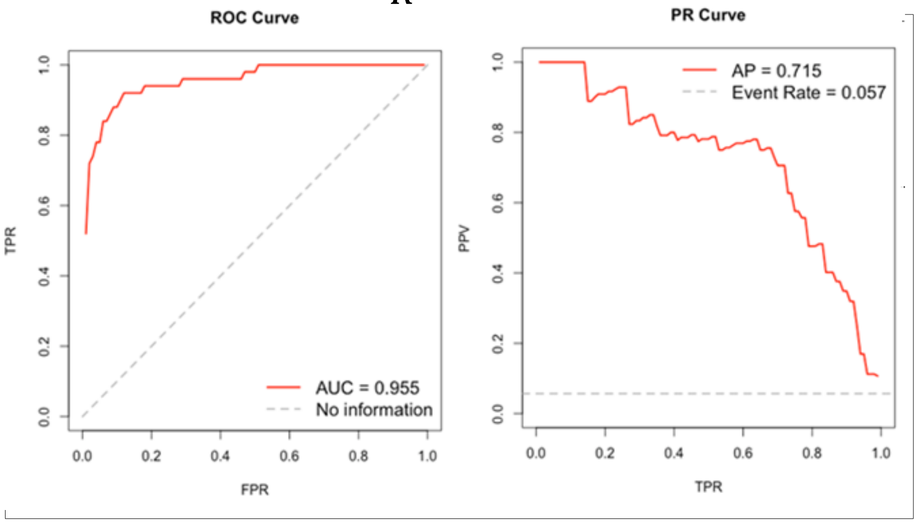
Threshold-free Summary Measure

- Area Under the ROC Curve (AUC)

$$\mathrm{AUC} \equiv \int_R \mathrm{TPF}(z)\, d\mathrm{FPF}(z)$$

- Area under the Precision-Recall curve

$$\mathrm{A}P \equiv \int_R \mathrm{PPV}(z)\, d\mathrm{TPF}(z)$$



Yuan et al. (2015)

AP Estimator (ordinal risk scores)

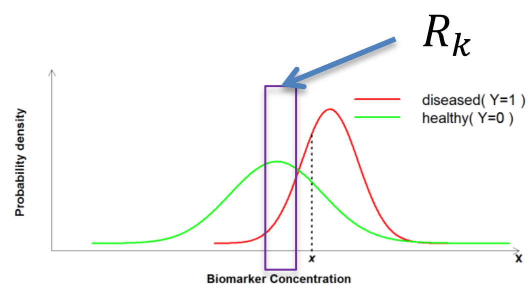
Score	x_1	$>$	x_2	$> \cdots >$	x_k	$>$	x_{k+1}	$> \cdots >$	x_K	
Partition	R_1		R_2	$\cdots$	R_k		R_{k+1}	$\cdots$	R_K	Total
Class-1	Z_1		Z_2	$\cdots$	Z_k		Z_{k+1}	$\cdots$	Z_K	n_1
Class-0	$\bar{Z}_1$		$\bar{Z}_2$	$\cdots$	$\bar{Z}_k$		$\bar{Z}_{k+1}$	$\cdots$	$\bar{Z}_K$	n_0
Total	S_1		S_2	$\cdots$	S_k		S_{k+1}	$\cdots$	S_K	n

Data in the above 2 X K table follow

$$\begin{aligned}
(Z_1, Z_2, ..., Z_K)|n_1 &\sim \text{multinomial}(n_1; p_1, p_2, ..., p_K),\\
(\bar{Z}_1, \bar{Z}_2, ..., \bar{Z}_K)|n_1 &\sim \text{multinomial}(n - n_1; q_1, q_2, ..., q_K),\\
n_1 &\sim \text{binomial}(n, \pi),
\end{aligned}$$

For continuous risk scores

$$p_k = \int_{R_k} f_1(x)dx, \quad q_k = \int_{R_k} f_0(x)dx,$$



Yuan et al. (2015)

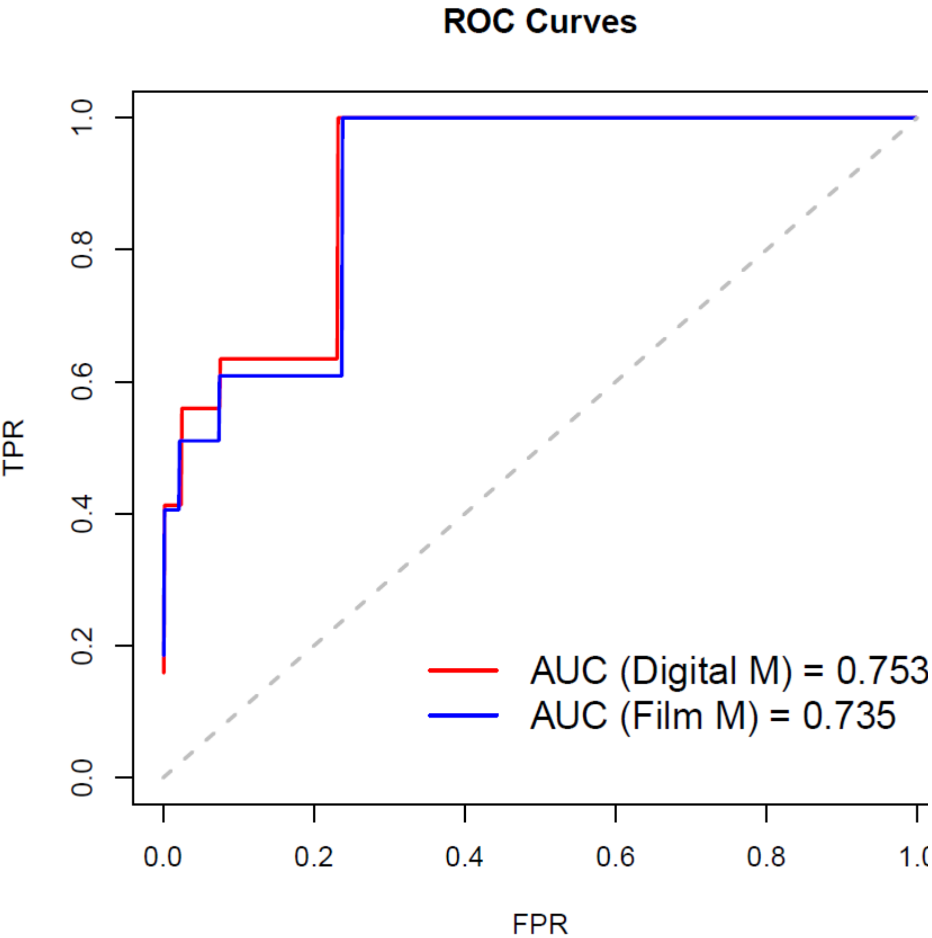
MLE and Variance Estimator

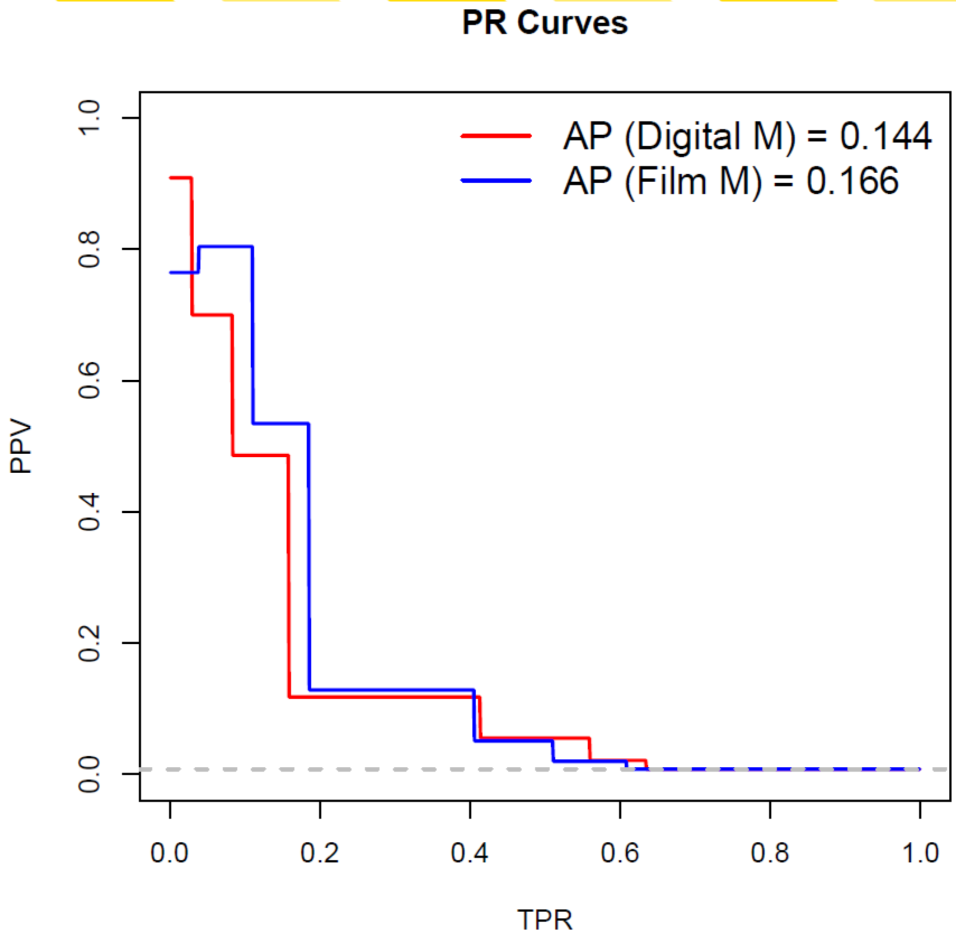
$$\widehat{\mathrm{AP}} = g(\widehat{p}_k, \widehat{q}_k, \widehat{\pi}) = \sum_{k=1}^{K} \left[\widehat{p}_k \left(\frac{\widehat{\pi} \sum_{k' \leq k} \widehat{p}_{k'}}{\widehat{\pi} \sum_{k' \leq k} \widehat{p}_{k'} + (1 - \widehat{\pi}) \sum_{k' \leq k} \widehat{q}_{k'}} \right) \right]$$

Applying the Delta method, the variance estimator is

$$\widehat{var}\left(\widehat{AP}\right) \approx (\nabla g)^T \hat{J}^{-1} (\nabla g)$$

Yuan et al. (2015)





Malignancy score		7	6	5	4	3	2	1	Total
Digital M	Category Total	11	29	69	1061	2224	6588	32588	42570
	Cancers	10	18	25	85	49	25	122	334
Film M	Category Total	17	29	70	942	2291	6910	32486	42745
	Cancers	13	24	25	74	35	33	131	335

Revisiting breast cancer screening example

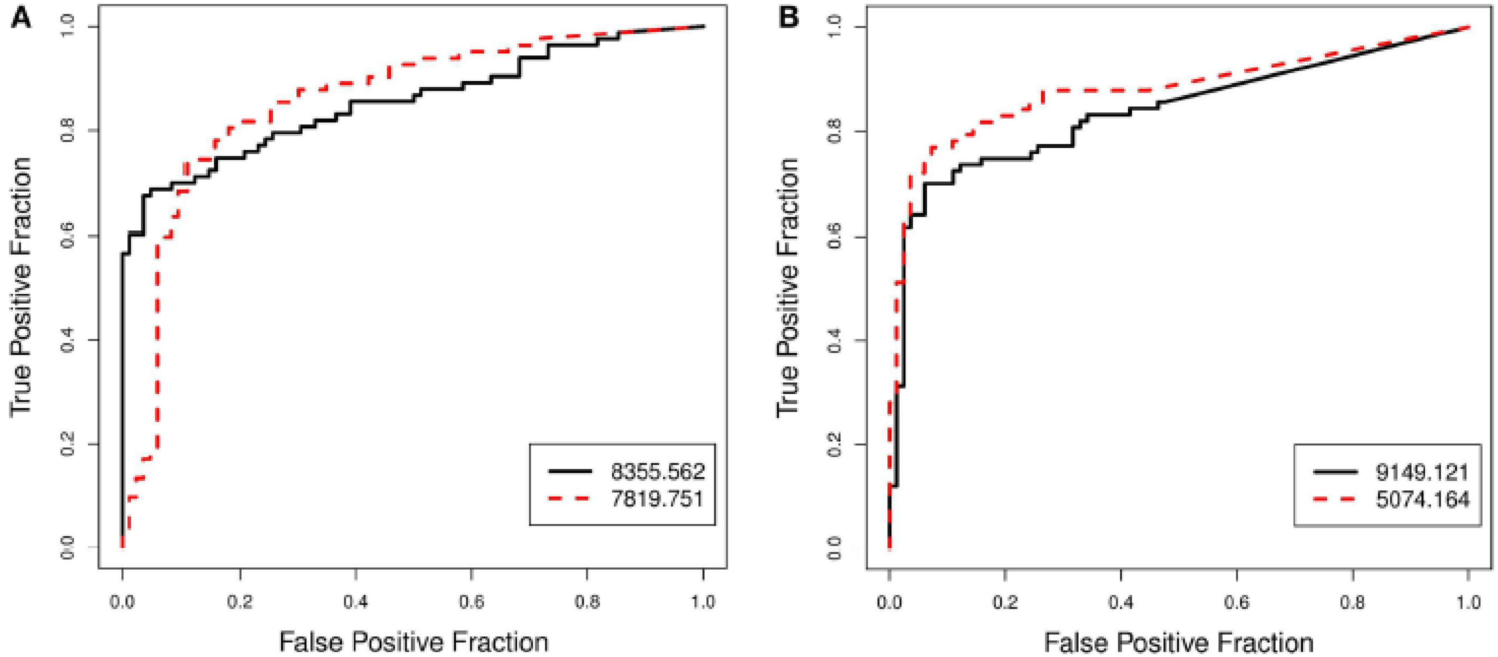


Table 1 | Prostate cancer example.

Biomarkers		AUC			AP		
		$n_0 \times 1$ ($\pi \approx 0.5$)	$n_0 \times 10$ ($\pi \approx 0.09$)	$n_0 \times 100$ ($\pi \approx 0.01$)	$n_0 \times 1$ ($\pi \approx 0.5$)	$n_0 \times 10$ ($\pi \approx 0.09$)	$n_0 \times 100$ ($\pi \approx 0.01$)
A	8355.562	0.849	0.783	0.783	0.856	0.606	0.571
	7819.751	0.850	0.857	0.857	0.802	0.370	0.062
B	5074.164	0.886	0.869	0.869	0.833	0.306	0.043
	9149.121	0.832	0.793	0.793	0.822	0.512	0.225

A simple thought experiment showing changes in the estimated AUC and AP as a result of artificially inflating the number of control subjects (n_0) to mimic real-life screening settings, where the prevalence (π) of disease is low.

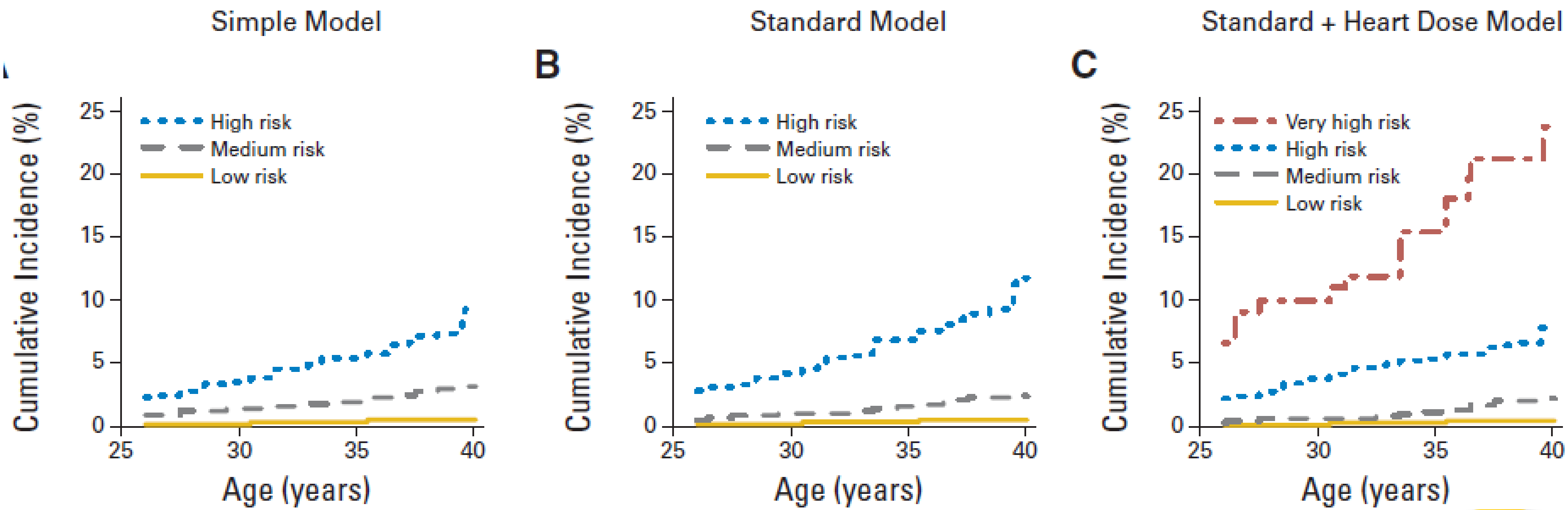
AP – AUC Relationship

- When two risk scores U_1 and U_2 are compared
  - If ROC curve of U_1 dominates that of U_2 everywhere, then PR curve of U_1 dominates that of U_2 everywhere. $AUC_1 > AUC_2$ and $AP_1 > AP_2$
  - If ROC curves of U_1 and U_2 crosses, the ranking of U_1 and U_2 based on of AUC and AP may differ.
- Similar to AUC, AP is a semi-proper scoring rule.

Su et al. 2015
Yuan et al. 2018

Motivating Data – Time to Event outcome

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



AP_{t_0} for Time-to-Event Outcome

- Time-dependent Average Positive predictive value (AP_{t_0}) for event status

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

Yuan et al. 2018

Nonparametric Estimator for Event Status

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

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

where

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

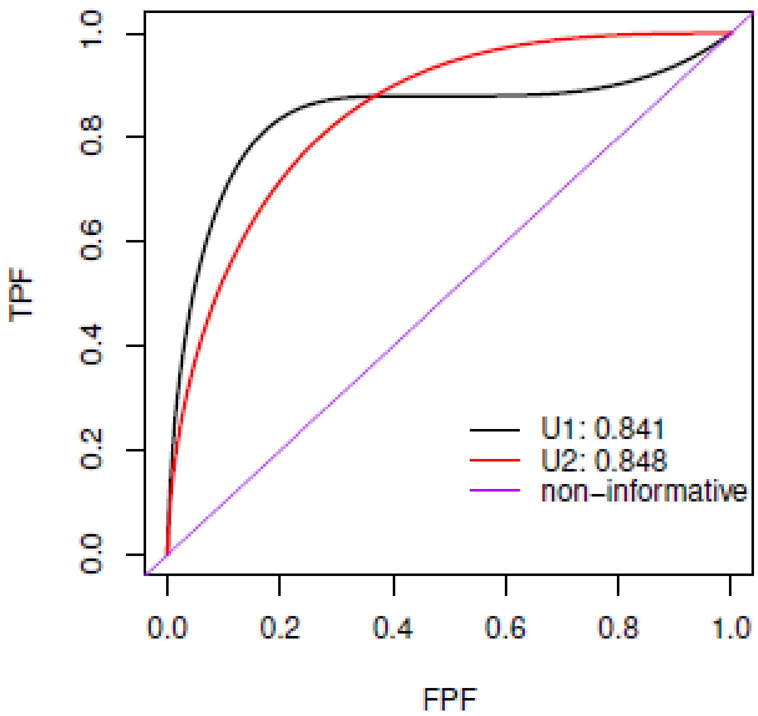
$$\widehat{\mathrm{PPV}}_{t_0}(z) = \frac{\sum_{i=1}^{n} \widehat{w}_{t_0,i} I(Z_i \ge z) I(X_i < t_0)}{\sum_{i=1}^{n} \widehat{w}_{t_0,i} I(Z_i \ge z)}$$

Yuan et al. 2018

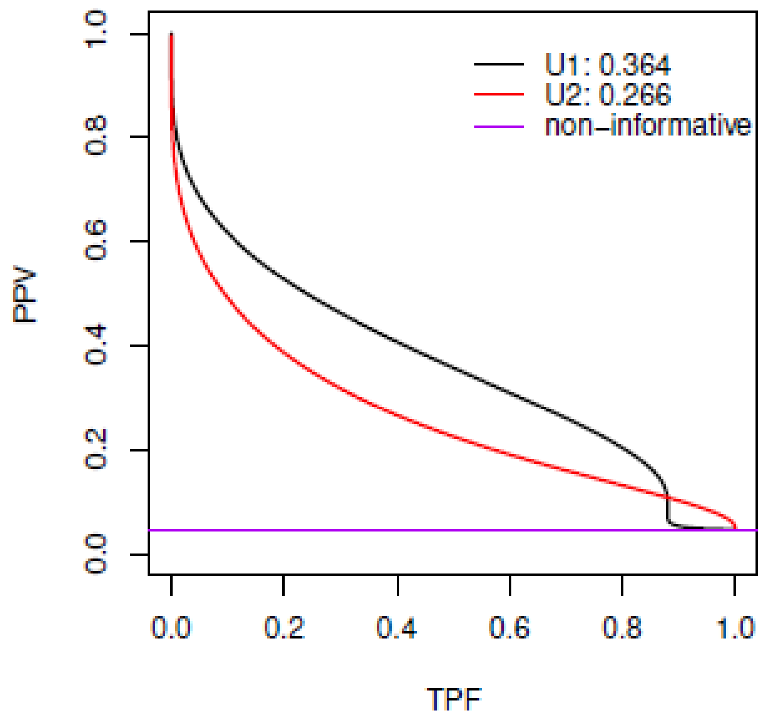
Simulation Study

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

$ROC_{t_0=8}$



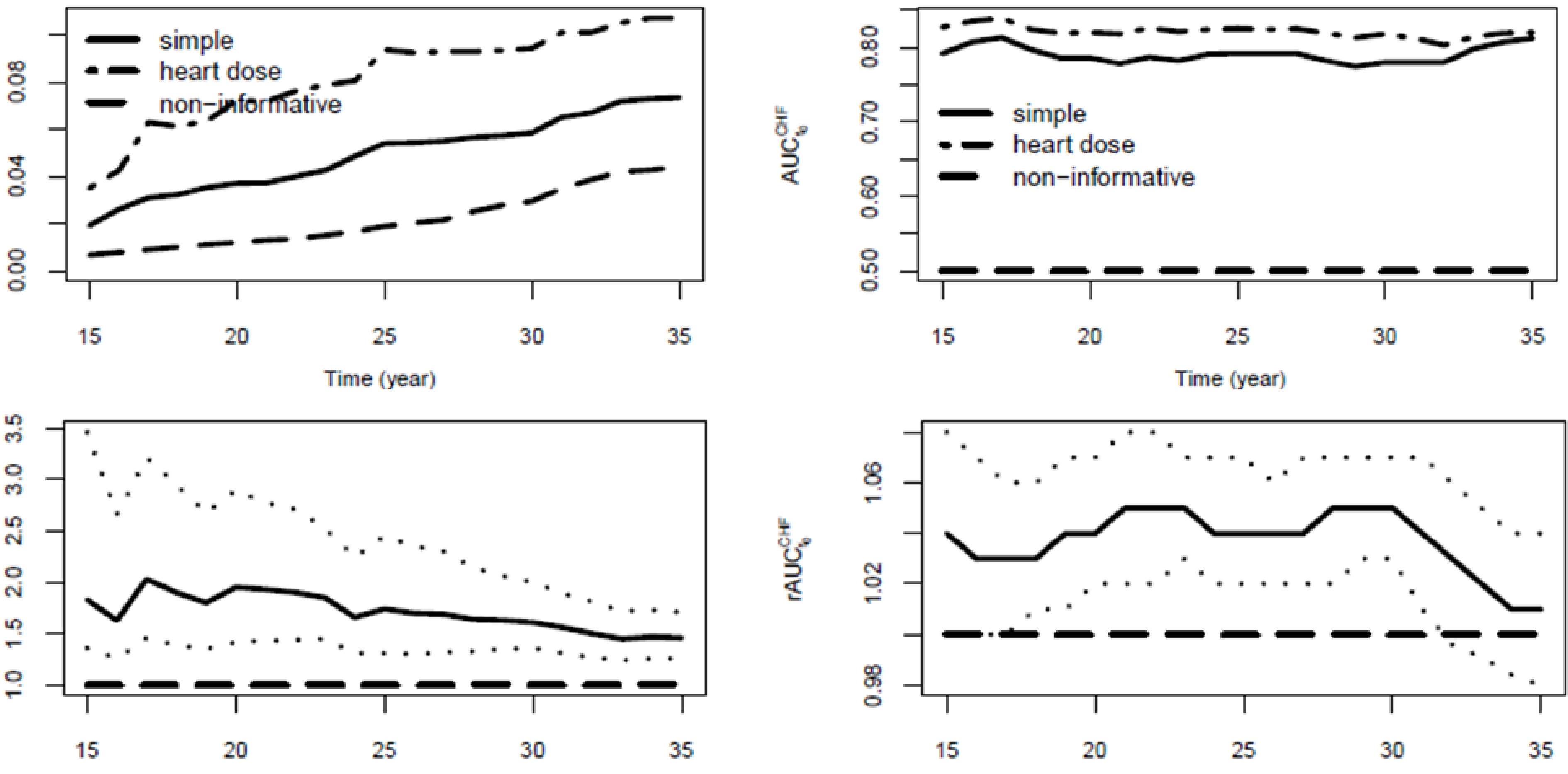
$PR_{t_0=8}$



Results (n=2000)

t_0	Event rate	Risk score	AP TRUE	AP BIAS	AP ESE	AP ASE^b	AP $ECOVP^b(\%)$	AUC TRUE
0.5	0.0101	U_1	0.182	0.0361	0.0806	0.0794	92.2	0.920
		U_2	0.124	0.0339	0.0687	0.0679	94.1	0.904
		Δ	0.058	0.0251	0.102	0.116	96.1	0.016
		Ratio	1.47	0.4820	1.470	1.740	92.4	1.02
8	0.0495	U_1	0.364	0.0085	0.0508	0.0499	94.4	0.841
		U_2	0.266	0.0121	0.0435	0.0439	94.8	0.848
		Δ	0.098	-0.0028	0.0707	0.072	96.3	-0.007
		Ratio	1.37	0.0123	0.310	0.322	95.8	0.99
36	0.0991	U_1	0.462	0.0060	0.0416	0.0431	94.2	0.786
		U_2	0.375	0.0074	0.0387	0.0393	96.3	0.824
		Δ	0.087	-0.0045	0.0655	0.0633	95.7	-0.038
		Ratio	1.23	-0.0010	0.189	0.187	94.5	0.95

$AP_{t_0}\ vs.\ t_0$ $\qquad$ $AUC_{t_0} vs.\ t_0$



R package <APtools> and SAS macro for binary and time to event outcome @ https://sites.ualberta.ca/~yyuan/software.html

Incremental Value

- Risk factor & outcome association vs. information/calibration gain in prediction
- Existing metrics
  - Changes in AUC and Brier scores (BS)
  - NRI (net reclassication improvement)
  - IDI (integrated discrimination improvement)

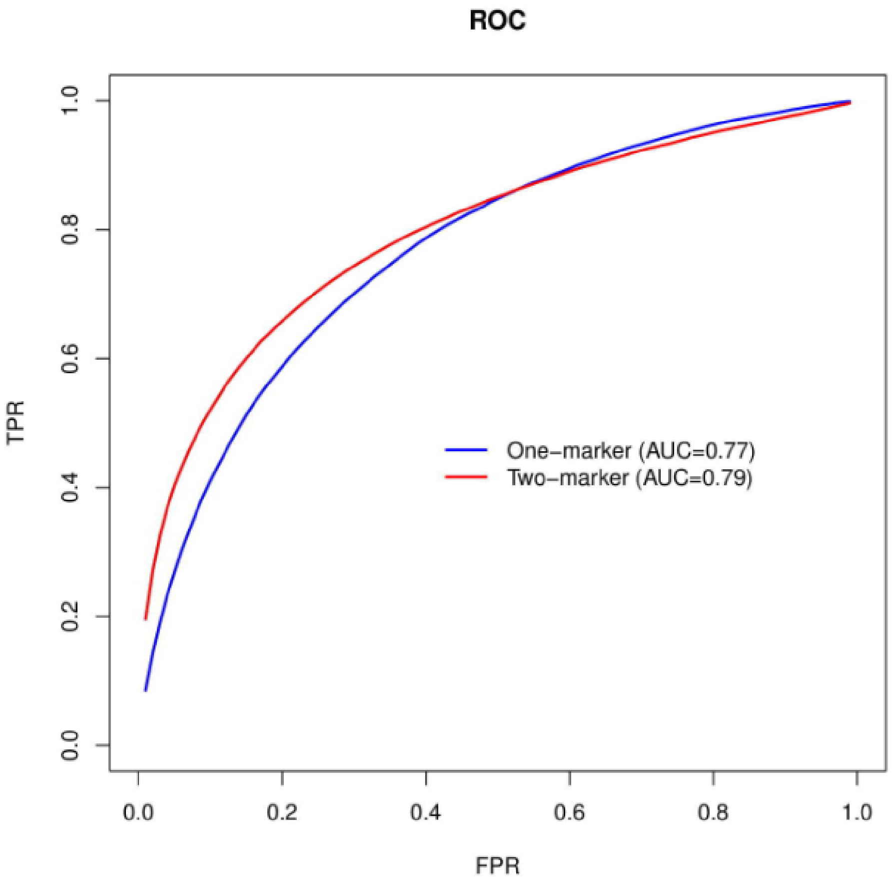
How does AP changes, in comparison to changes in AUC and BS?

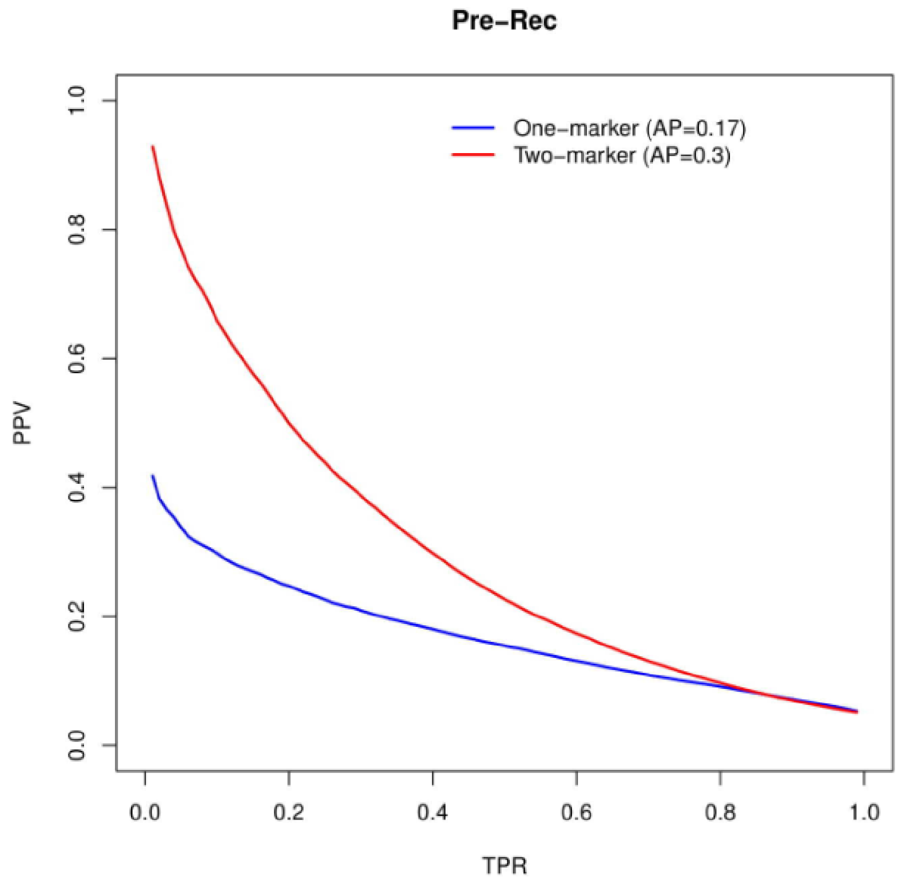
Simulation Study

- True model: $logit(\pi) = \beta_0 + \beta_1 U_1 + \beta_2 U_2 + \beta_3 U_1 U_2,$
  - β_1 and β_2 range: [0.3, 1.2]
  - β_3 range: [-1,1]
  - Independent U1 & U2 ~ iid N(0,1)
  - Event rate: ~5%
- Working model
  - Model 1: $logit(\pi) = \beta_0 + \beta_1 U_1$
  - Model 2: $logit(\pi) = \beta_0 + \beta_1 U_1 + \beta_2 U_2$
- Metrics
  - rAUC, rAP and rBS

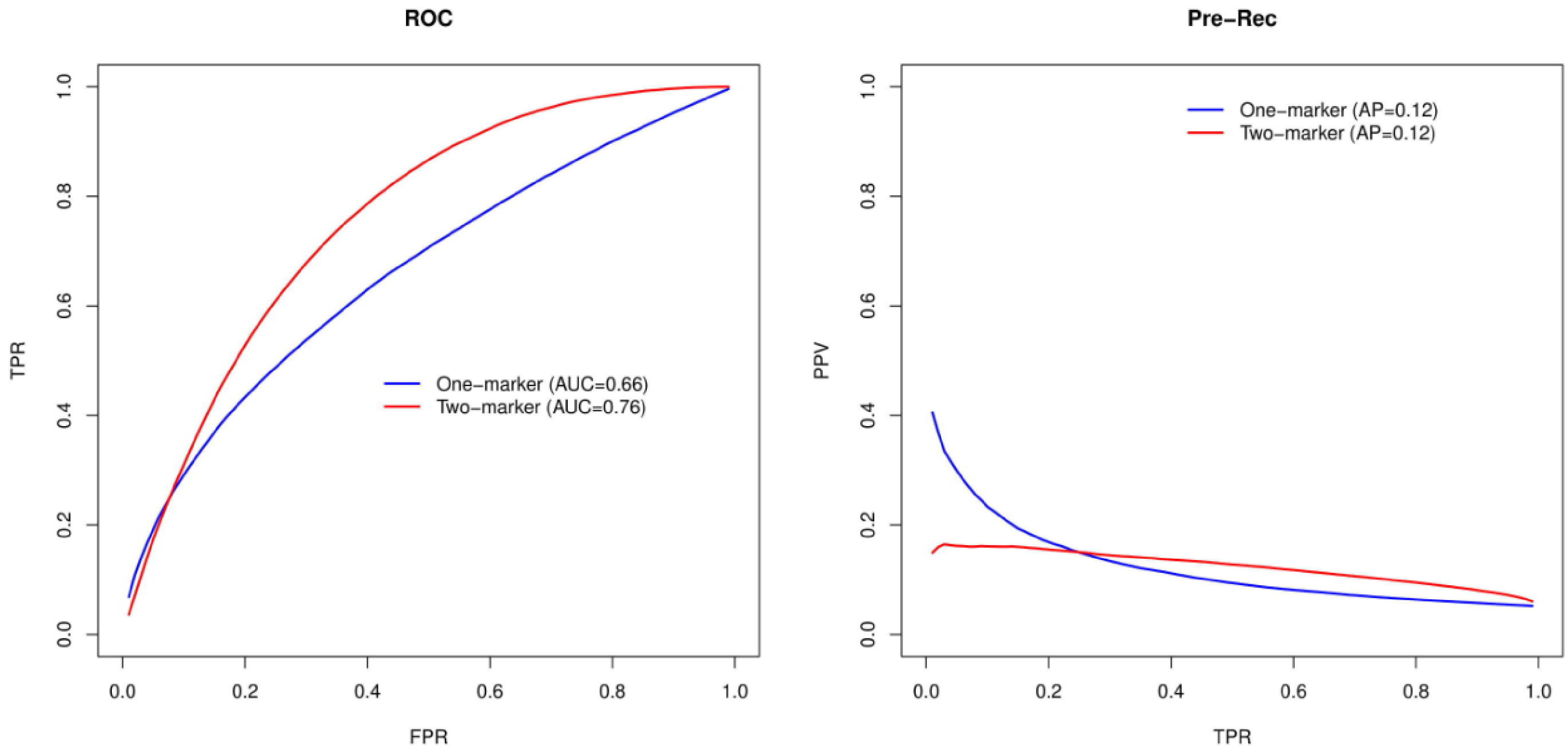
Metrics	Correlation	
Log(ratio of metrics: M2/M1)	Pearson	Spearman
-ln(rBS) and ln(rAUC)	0.083	0.30
-ln(rBS) and ln(rAP)	0.76	0.89
ln(rAUC) and ln(rAP)	0.48	0.51

$$\beta_1 = 0.9, \beta_2 = 0.3, \beta_3 = 0.6$$





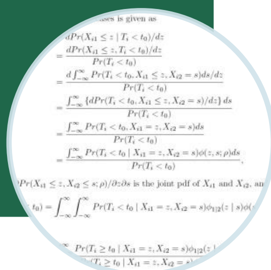
$$\beta_1 = 1, \beta_2 = 1, \beta_3 = -0.6$$



Research Program

- Prediction measures and risk prediction
- Gestational weight / fat trajectory
- Relation between association measures

NSERC Theme

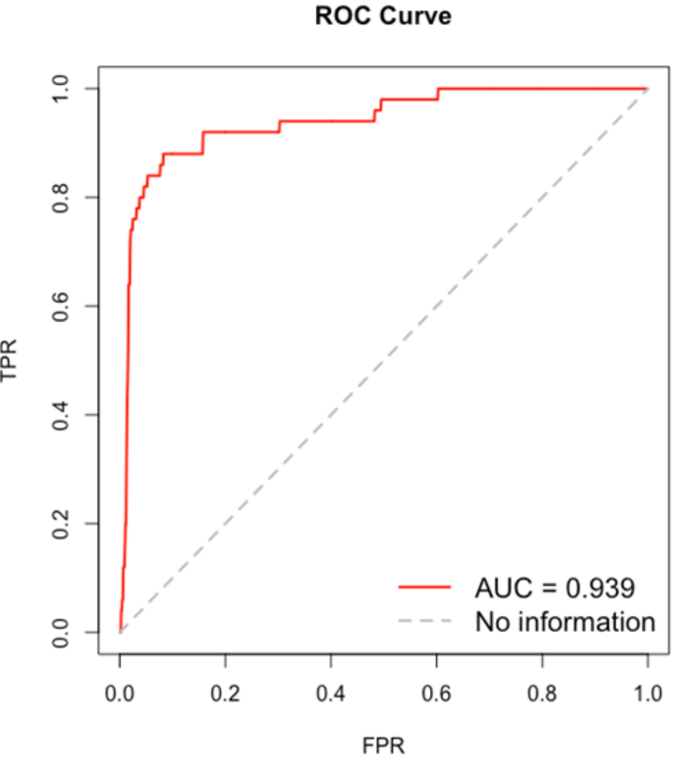


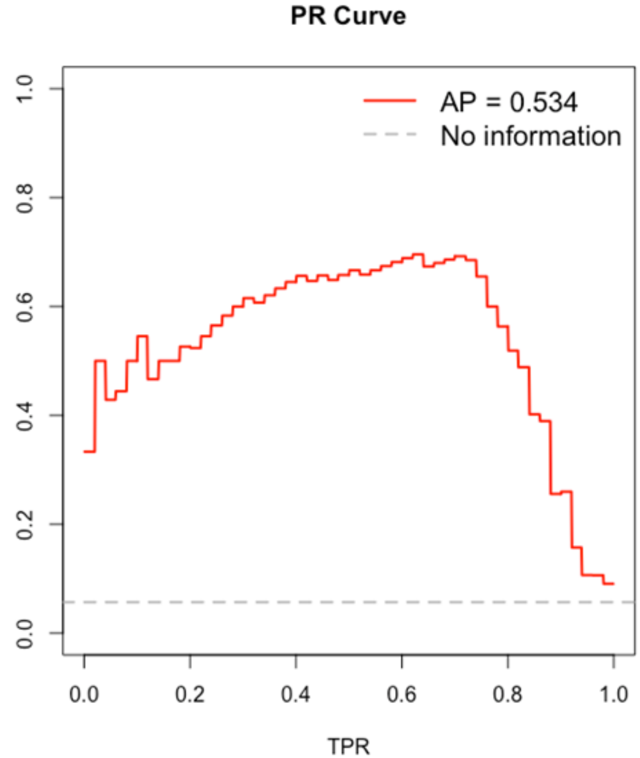
- Health services
- Late effects in survivors
- Rare cancer

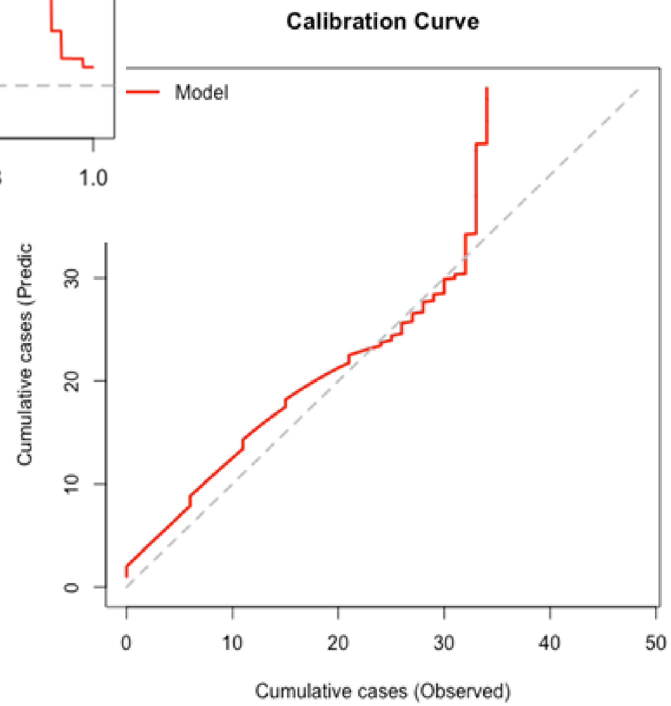
CIHR Theme (Cancer Epi)

Risk Prediction for Ovarian Failure

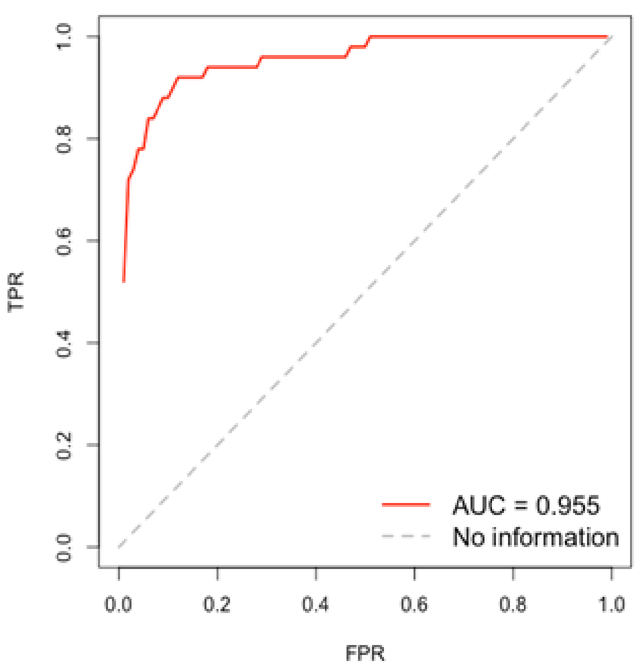
- Goal
  - Developing risk prediction algorithms for ovarian failure (OF) in childhood cancer survivors (CCS)
- Data source
  - ~6000 females (dx 1970-1999) from the CCSS cohort
- Algorithms
  - Logistic regression; Random Forest; and Support Vector Machines
- Performance
  - AUC 0.82, AP 0.50 for Acute OF (Internal validation)


ROC Curve
AUC = 0.939
No information
TPR
FPR
0.0
0.2
0.4
0.6
0.8
1.0

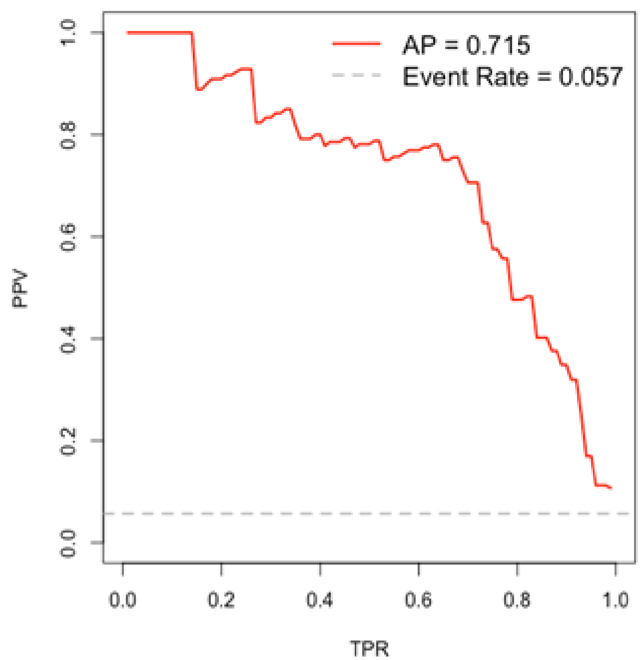

PR Curve
AP = 0.534
No information
TPR
0.0
0.2
0.4
0.6
0.8
1.0


Calibration Curve
Model
Cumulative cases (Predic
Cumulative cases (Observed)
0
10
20
30
40
50

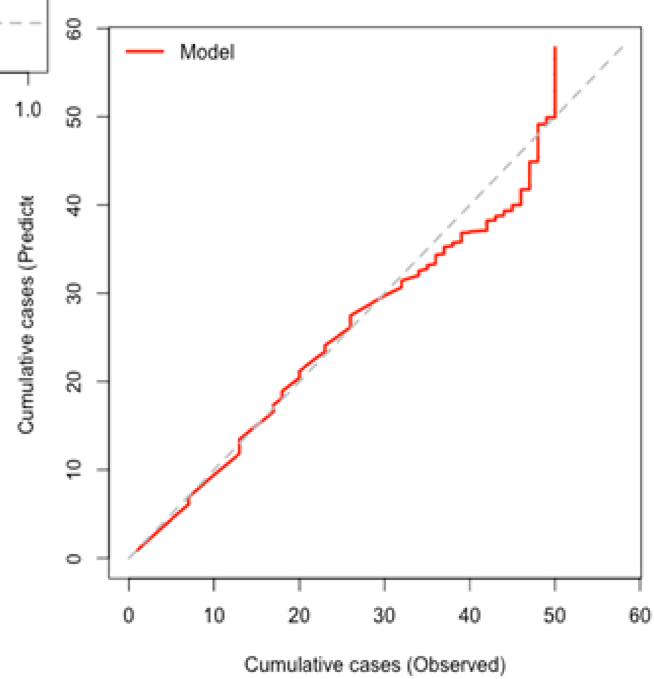
ROC Curve



PR Curve



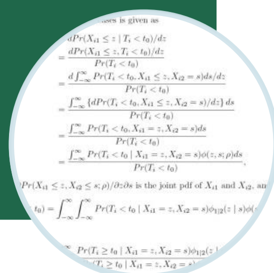
Calibration Curve



Research Program

- Prediction measures and risk prediction
- Gestational weight / fat trajectory
- Relation between association measures

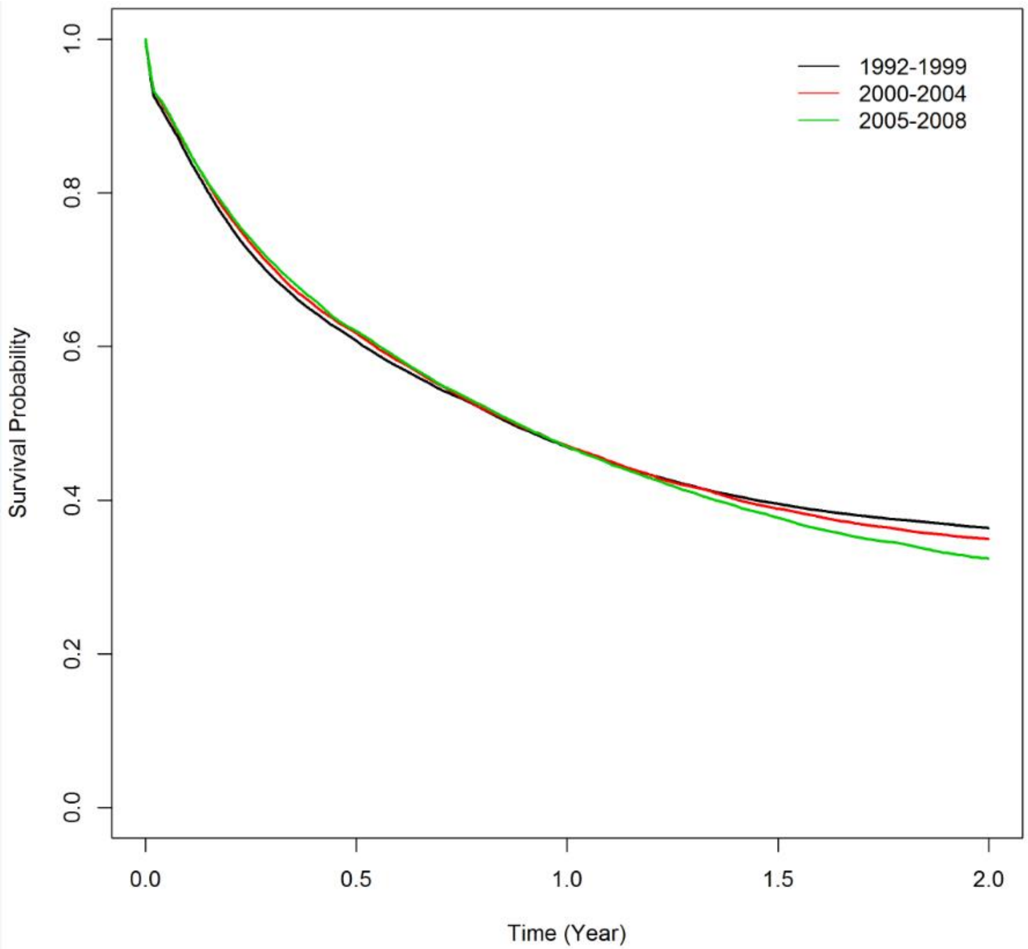
NSERC Theme

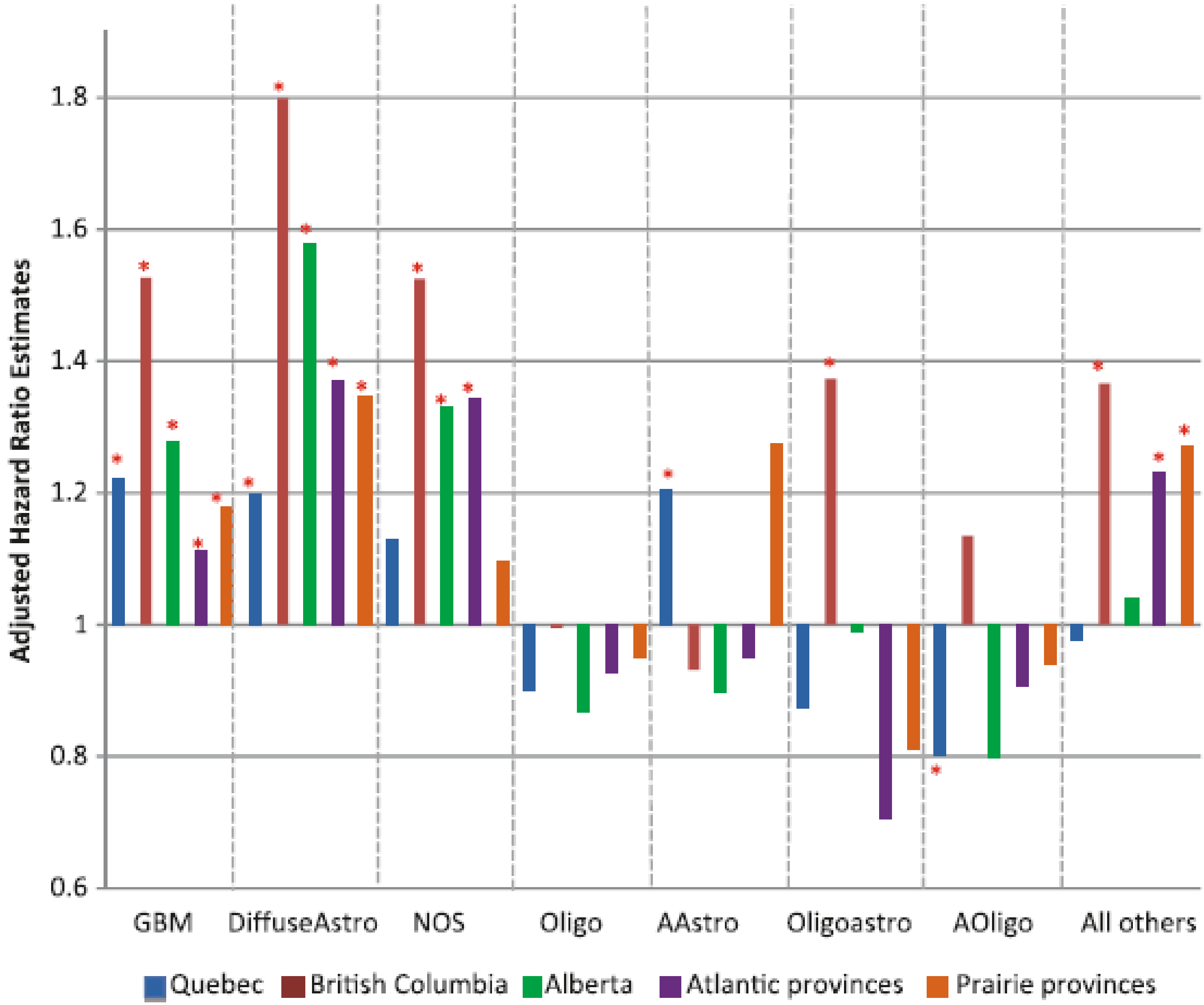


- Health services
- Late effects in survivors
- Rare cancer

CIHR Theme (Cancer Epi)

Brain Tumour Epidemiology: the Canadian Story





Yuan et al. 2016

Capturing Radiological Diagnoses of Brain Tumours in Canada

- Significant underreporting of non-malignant brain tumours (only ~40% expected cases captured)
- Vary by province

Table 2. Method of Diagnosis of Malignant Brain Tumors in 4 Canadian Provinces (2004–2015)

Province	*Microscopic Confirmation n (%)*	*Radiological Confirmation n (%)*	*Other* n (%)*	*Unknown n (%)*	*Total n*
Alberta	2,600 (83.9)	475 (15.3)	25 (0.8)	0 (0)	3,100
British Columbia	3,650 (84.4)	305 (7.1)	370 (8.6)	0 (0)	4,325
Manitoba	885 (79.0)	230 (20.5)	5 (0.5)	0 (0)	1,120
Ontario	11,265 (84.0)	590 (4.4)	215 (1.6)	1,340 (10.0)	13,410
Total	18,400 (83.8)	1,600 (7.3)	615 (2.8)	1,340 (6.1)	21,955

Note: Quebec data is only available up to 2010 and thus is not included in the table. Numbers are randomly rounded in accordance with Statistics Canada requirements.
*Other category includes death certificate, clinically confirmed, surgically confirmed, autopsy, and positive lab marker.

Yuan et al. 2018
Best paper award, NCRA

Table 3. Pathways to Clinical Care and Cancer Registration When a Brain Tumor is Radiologically Diagnosed		
Clinical Care	*Cancer Registries*	
Follow up/Treatment	**Registration**	**Data Quality**
Surgery and/or seen by oncologist	Yes	Accurate
Series of MRI studies over time. Diagnosis may or may not change. Surgery and/or oncology care involved during disease course.	Likely registered when pathology confirms diagnosis or oncologist prescribes treatment.	Level of accuracy varies; initial diagnosis delayed in reporting and information may not be accurately recorded
Series of MRI studies over time only, no surgery.	Not likely	Potential to miss a significant proportion of cases

MRI, magnetic resonance imaging.

Table 4. Legislation and Responsibility for Reporting and Registering Cancer Cases by Province			
Province	*Legislation*	*Healthcare professionals*	*Health Authority*
Alberta	Cancer Act, 2009	"Shall report"	"May request"
British Columbia	Health Act, 2009	Not mandatory but "must comply with request"	"May request"
Manitoba	Public Health Act, 2009	"Must report"	"May request"
Ontario	Cancer Act, 1990	None	Ensure "adequate reporting of cases and the recording and compilation of data"
Quebec	Public Health Act, 2001	"Must report", "in the manner and within the time limits prescribed in the regulation"	"Record"

Yuan et al. 2018
Best paper award, NCRA

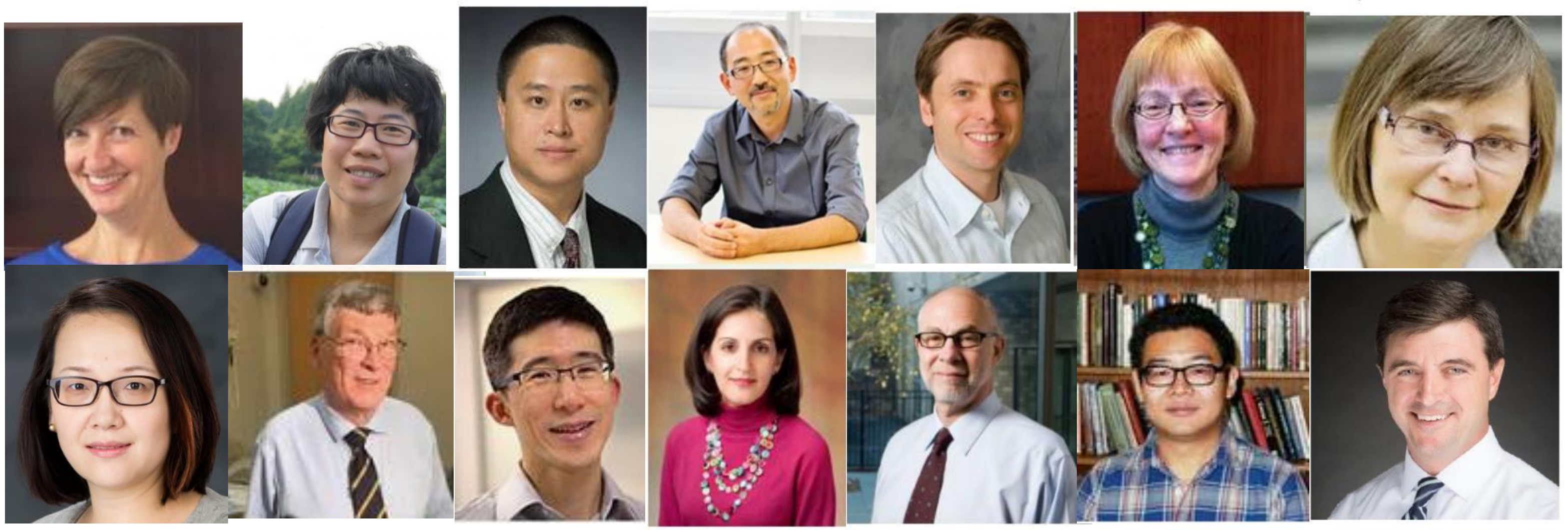
Recommendations

- Algorithmic solution needed (e.g. Natural language processing) for processing the unstructured radiology reports better capture cases
- Synoptic reporting in radiology should be explored

Yuan et al. 2018
Best paper award, NCRA

Funders, Collaborators, and Trainees/Staff

Staff members: Maoji Li, Qian Shi, Dr. Khanh Vu
Ye Shen (Health services**)**
Doris Li, Hengrui Cai, Zorina Han (Risk prediction measures**)**
Rebecca Clark, Michael Lu (Late effects in survivors**)**
Jordan Ross, Sana Amjad, Emily Maplethorpe (Rare cancer**)**



M.S.I. Foundation
LOIS HOLE HOSPITAL FOR WOMEN
nen & children's th research institute
braintumour foundation OF CANADA
UNIVERSITY OF ALBERTA SCHOOL OF PUBLIC HEALTH
CANSSI INCASS
NSERC CRSNG
CIHR IRSC Canadian Institutes of Health Research Instituts de recherche en santé du Canada
This is PUBLIC HEALTH

Thank you!

Questions???