On Criteria for Evaluating Risk Prediction Models for Public Health Applications

Ruth Pfeiffer

ROeS Seminar September 2013

Biostatistics Branch
Division of Cancer Epidemiology and Genetics
National Cancer Institute, National Institutes of Health

Motivation: Absolute Risk for Cancer Incidence

$$r(x,a,\tau) = P(T \le a+\tau, cause = C \mid T > a; x)$$

$$= \int_a^{a+\tau} h_C(t,x) \exp\left[-\int_a^t \left\{ h_C(u,x) + h_D(u,x) \right\} du \right] dt$$

T - event time
X - individual risk or protective factors
a - age
τ - length of projection
$h_C(t,x)$ – cancer hazard at age t
$h_D(t,x)$ - mortality hazard from competing risks

Combine Data from Different Sources to Estimate Absolute Risk

model $h_c(t,x) = h_{c0}(t)rr(\beta_c x)$

Cohort, nested case-control, case cohort , case-control data

Cancer Registries

Estimate **relative risk,** $rr(\beta_c x)$ and **attributable risk,** $AR(x)$

$h_{c0}^*(t)$, composite cancer hazard (age spec.)

$$\hat{h}_c(t,x) = h_{c0}^*(t)(1-AR_c)rr(\beta_c x)$$

$$\hat{r}(x,a,\tau) = \int_a^{a+\tau} \hat{h}_C(t,x)\exp\left[-\int_a^t \left\{\hat{h}_C(u,x)+\hat{h}_D(u,x)\right\}du\right]dt$$

Absolute Risk Models for Endometrial and Breast Cancers

We combined data on white non-Hispanic women ages 50+ from two large cohorts to estimate **relative risks** for endometrial and breast cancers

Outcome	Combined cohort size	Total # cases
Endometrial cancer	157,755	**1,656**
Breast cancer	256,522	**8,160**

Relative Risk Estimates

Variable	Endometrial Cancer	Breast Cancer
	RR (95%CIs)	RR (95%CIs)
BMI (<25, 25-<30, 30-<35, 35-<40, 40+)	1.72* (1.65-1.80)	X
Oral contraceptive use (1+ year , <1 year)	1.44(1.28-1.61)	
Menopausal hormone therapy use (never, <10 years, 10+ years)	X	1.40 (1.32-1.49)
Parity (3+, 1-2, 0)	X	X
Family history of breast or ovarian cancer		1.39 (1.32-1.47)
Benign breast disease/biopsy (no/yes)		1.40 (1.33-1.48)
Smoking Never compared to current	1.47 (1.22-1.78)	
Former compared to current	X	
Age first life birth	X	X
Alcohol consumption		X
Age at menopause		X

Absolute Risk Estimates: two 50 year old non-smoking women without history of benign breast disease

	Woman 1	Woman 2
Oral contraceptive use	no	yes
Age at first birth	25	39
# of life births	3	1
BMI, kg/m^2	24	40
Menopausal	no	yes
Alcohol consumption (drinks/day)	0	>1
Family history breast/ovarian cancer	no	yes
HRT use+ duration>1 year	no	yes
10 year absolute breast cancer risk estimate	**1.83%**	**6.83%**
10 year absolute endometrial cancer risk estimate	**0.40%**	**5.91%**

External Validation: Independent Population for Assessment of Model Performance

Assume population of N individuals followed over time period t (cohort data)

Observe disease outcome at end of follow-up

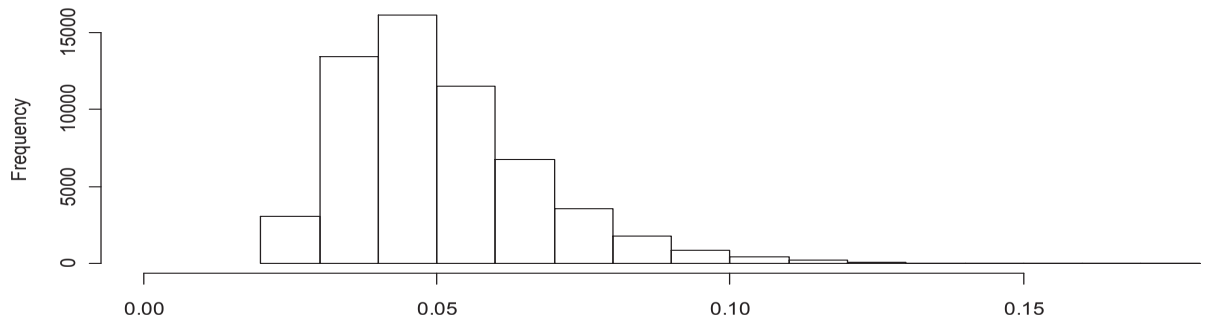
$$Y_i = \begin{cases} 1, & \text{if ith woman develops event during } t \\ 0, & \text{otherwise} \end{cases}$$

$r(x_i) = \hat{P}(Y = 1 \mid x_i)$ absolute risk estimate for ith person with baseline covariates x_i, including age a over time t

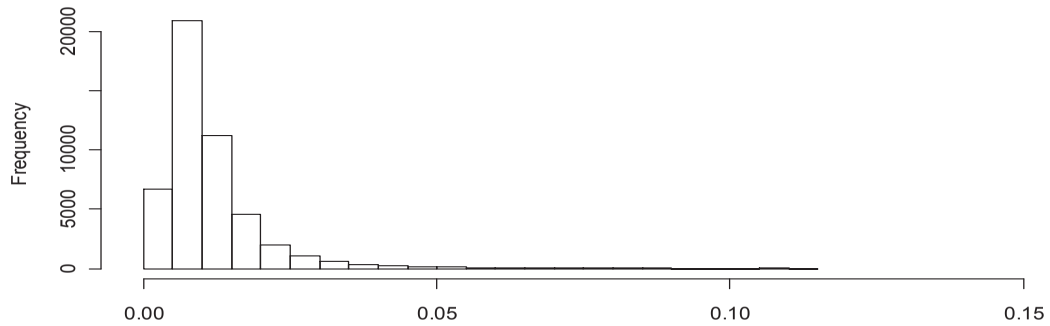
Risk model r known

Risk estimates r have distribution F on $[0, 1]$

Distribution *F* of Risk in Nurses' Health Validation Cohort (N~60,000)



Breast cancer absolute risk estimates



Endometrial cancer absolute risk estimates

Standard Criteria for Evaluating Performance of Risk Prediction Models: Calibration

Assesses model bias

Model r(x) is **well calibrated** if for each x

$$P(Y = 1 \mid r(x) = r) \approx r$$

Then $\mu = E(Y) = P(Y = 1) = \int_0^1 r dF(r) = E(R)$

Model r is unbiased $\left(\text{well calibrated}\right)$ in population if

$$\frac{1}{N}\sum_{i=1}^{N} Y_i \approx \frac{1}{N}\sum_{i=1}^{N} r_i$$

Assess Performance of Models in Nurses' Health Validation Cohort

Ages 51-70 at baseline

Calibration: $O = \sum_{i=1}^{N} Y_i,\ E = \sum_{i=1}^{N} r(X_i)$

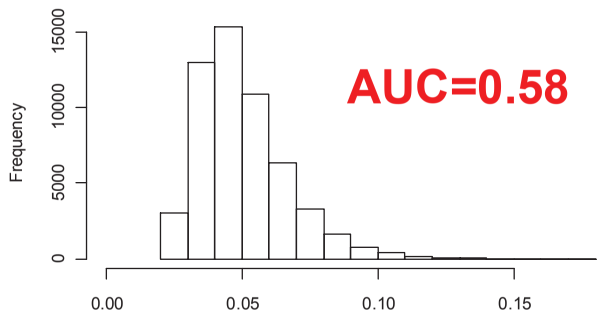
Outcome	# women	O	E	E/O (95%CI)
Breast cancer	57,906	2934	2930	**1.00** (0.96–1.04)
Endometrial cancer	37,241	532	637	**1.20** (1.11-1.29)

Standard Criteria for Evaluating Performance of Risk Prediction Models: Measures of Discrimination

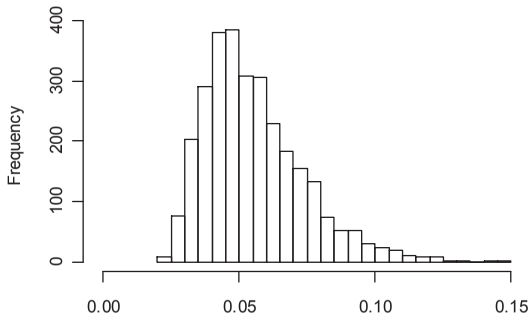
- **Receiver operating characteristic** (ROC) **curve** plots sensitivity (true positive rate) vs. 1 − specificity (false positive rate), for binary classifier as discrimination threshold is varied
- Area under the ROC curve **(AUC) (c-statistic)**

$$AUC=P(r_{Y=1}>r_{Y=0})$$

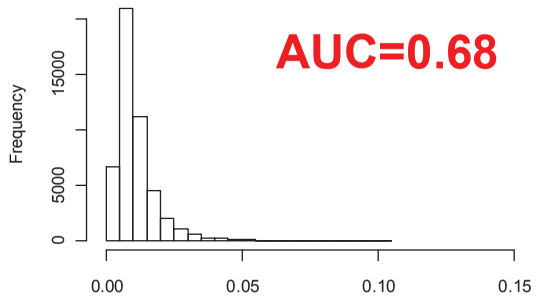
Distribution of risk estimates in independent validation cohort (Nurses' Health Study)



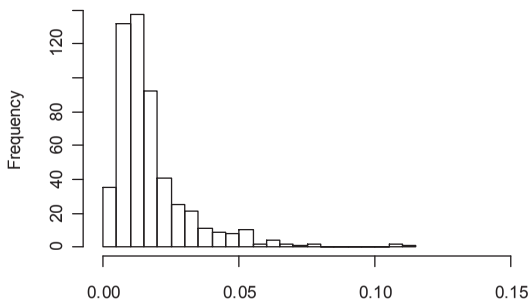
Y=Breast cancer=0



Y=Breast cancer=1



Y=Endometrial Cancer=0



Y=Endometrial Cancer=1

Criteria that Assess Model Performance for Screening Applications

Wish to make **screening recommendations** for population over next five years based on baseline risk assessment from absolute risk model

Compute 5-year cancer risk from model for every woman given baseline covariates X, $r_i = r(X_i)$, i=1,…N

Rank risks from lowest to highest risk: $r_{(1)} \leq r_{(2)} \leq \ldots \leq r_{(N)}$

1. **Proportion of cases followed, PCF(q):** proportion of cases followed-up in screening program that screens proportion *q* of population at highest risk
2. **Proportion needed to follow, PNF(p):** proportion of population at highest risk that needs to get screened so that proportion *p* of future cases will be screened

Distributions of Risk in Cases and Non-cases

$r(x) = \hat{P}(Y = 1 \mid x)$ risk estimate

$F(r^*) = P(r \leq r^*)$ distribution of risk in general population

Distributions of Risk in Cases and Non-cases

$r(x) = \hat{P}(Y = 1 | x)$ risk estimate

$F(r^*) = P(r \le r^*)$ distribution of risk in general population

$G(r^*) = P(r \le r^* | Y = 1)$, distribution of risk in cases

Distributions of Risk in Cases and Non-cases

$r(x) = \hat{P}(Y = 1 | x)$ risk estimate

$F(r^*) = P(r \le r^*)$ distribution of risk in general population

$G(r^*) = P(r \le r^* | Y = 1)$, distribution of risk in cases

$K(r^*) = P(r \le r^* | Y = 0)$, distribution of risk in non-cases

Proportion Cases Followed, PCF(q)

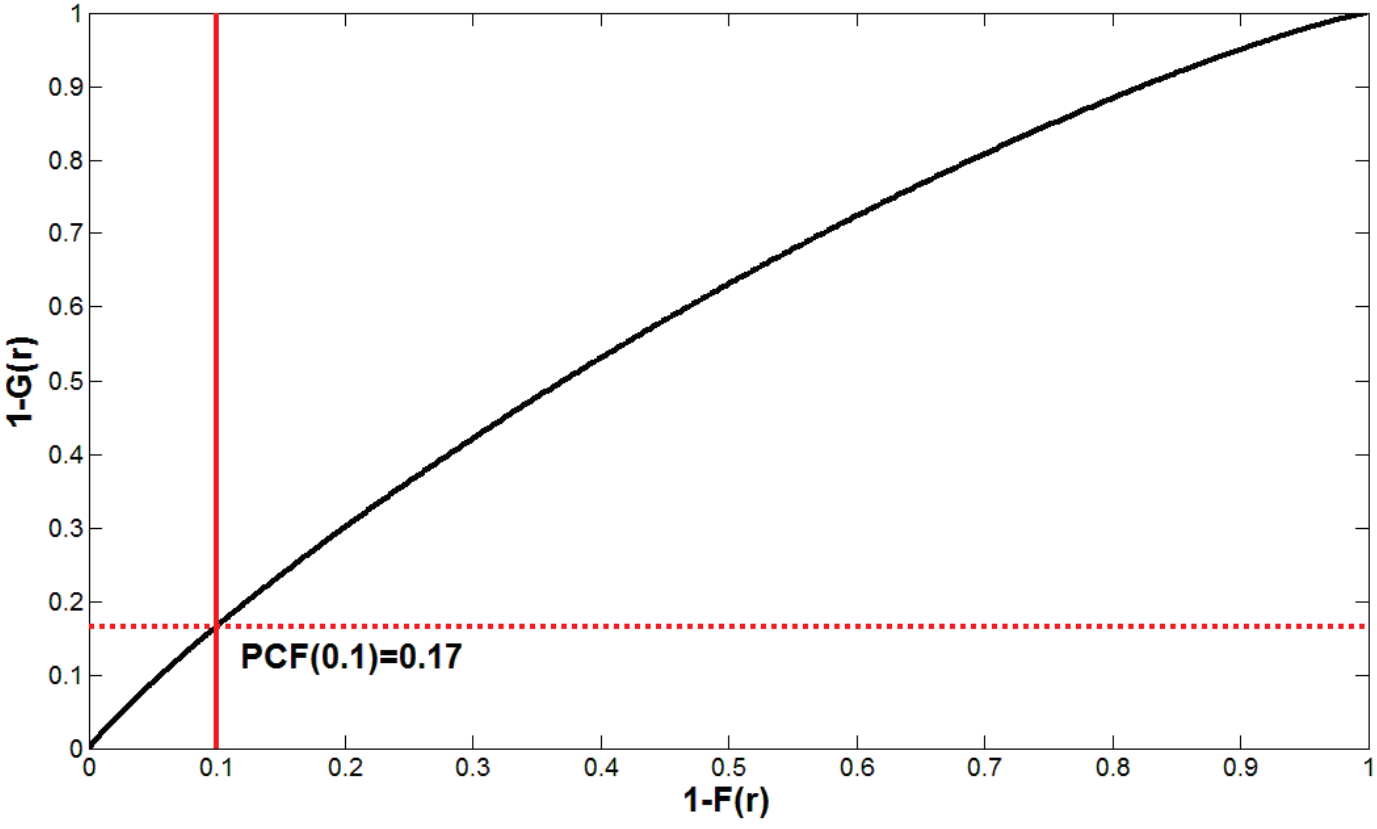
Proportion of cases who are included in proportion *q* of population at highest risk.

$F^{-1}(1-q)$: (1-q)th quantile of population distribution *F*

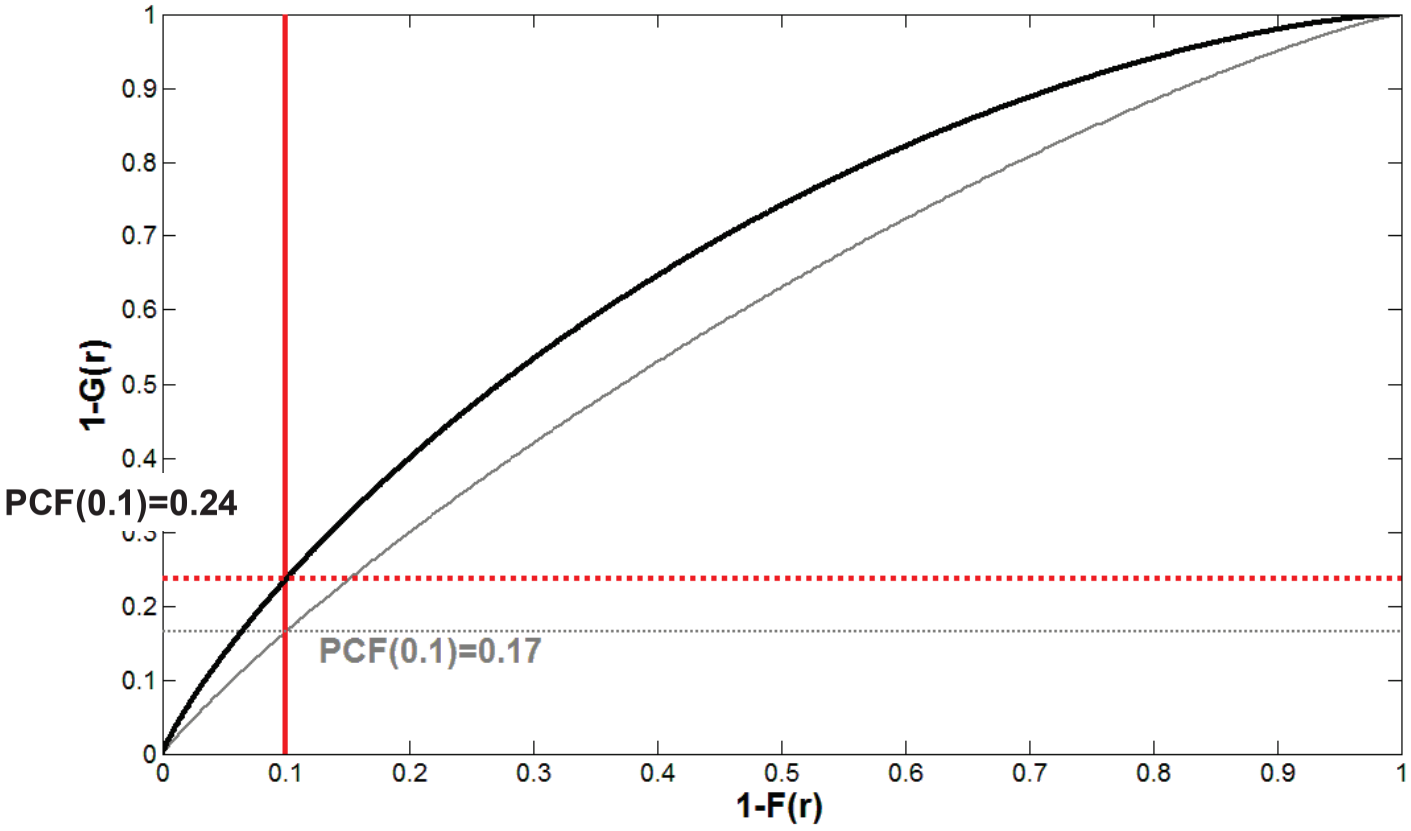
G- distribution of risk in cases

$$PCF(q)=1-G\circ F^{-1}(1-q)$$

PCF, risk distribution *F* in population is Beta(8.5, 161.5) (AUC=0.59)



PCF, risk distribution *F* in population is Beta(2.3, 43.7) (AUC=0.68)



Integrated Proportion of Cases Followed (iPCF)

G- distribution of risk in cases

$PCF(q)=1- G \circ F^{-1}(1-q)$

$$iPCF(q^*)= \int_{q^*}^{1} PCF(q)dq =$$

$$= 1-q^* - \frac{1}{1-q^*} P(R_G \le R_F \mid R_F \in (0, F^{-1}(1-q^*))$$

$$q^* = 0: \; iPCF(0) = P(R_G > R_F)$$

Comparison with ROC curve

F- distribution of risk in population

G- distribution of risk in cases

$$PCF(q)=1-G\circ F^{-1}(1-q)$$

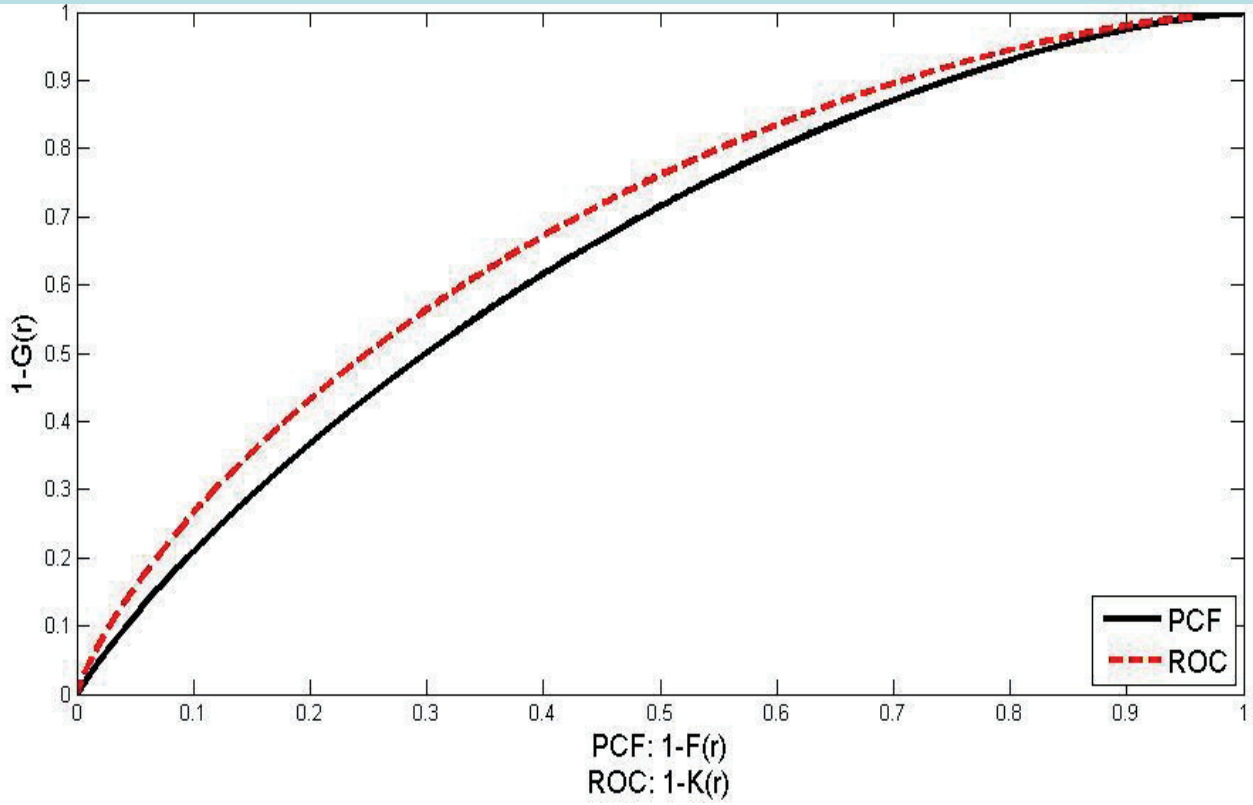
$$iPCF(0)=\int_0^1 PCF(q)dq = P(R_G > R_F)$$

K- distribution of risk in non-cases

$$ROC(q)=1-G\circ K^{-1}(1-q)$$

$$AUC=\int_0^1 ROC(q)dq = P(R_G > R_K)$$

ROC and PCF curves, risk distribution *F* in population is Beta(2.5, 10), AUC=0.69, iPCF=0.66 (mu=0.20)



Proportion Needed to Follow, PNF(p)

Fraction of the general population with highest risks that needs to be screened (followed up) to assure that a given fraction *p* of all cases in population receive screen.

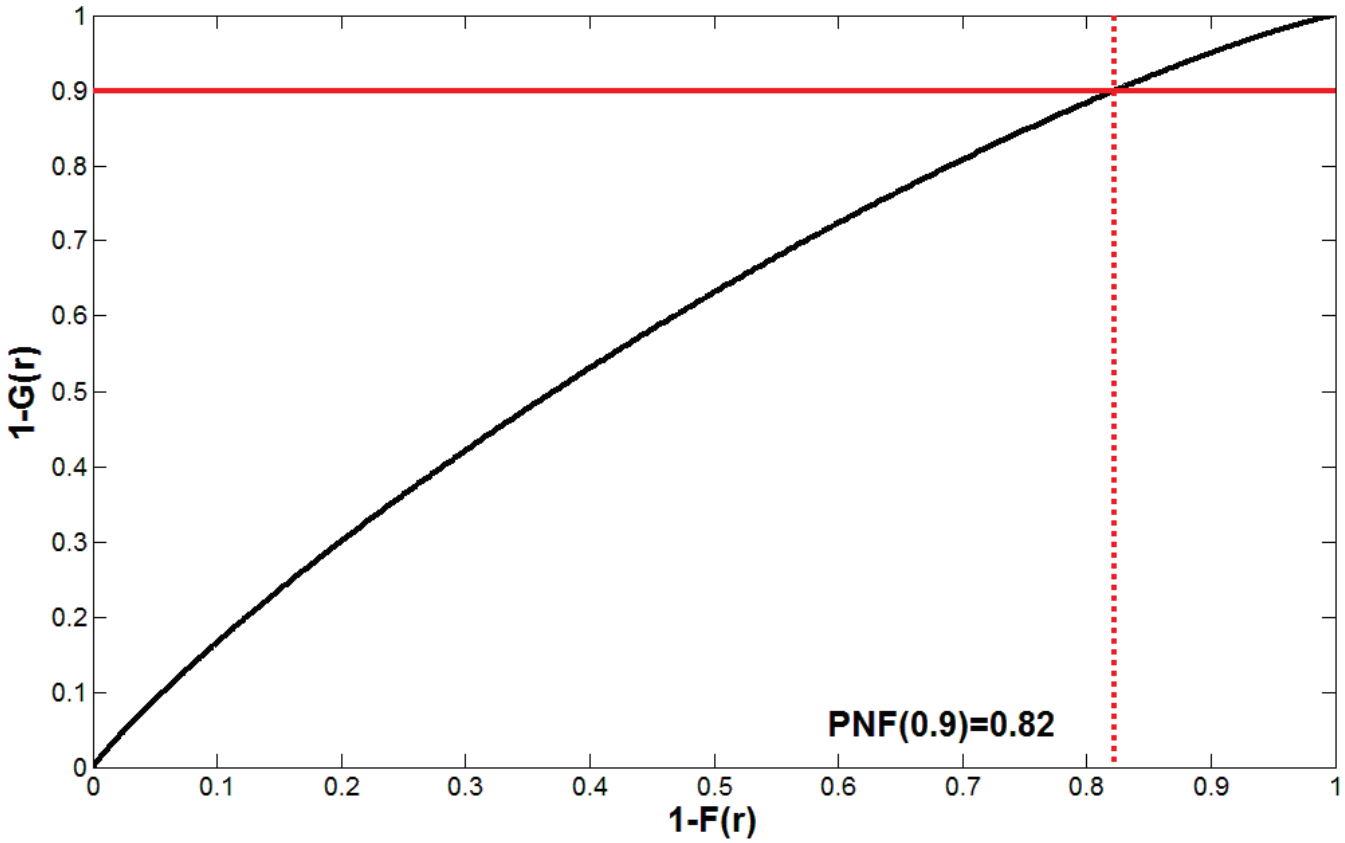
Solve

$$1- G \circ F^{-1}(1-PNF(p)) = p$$

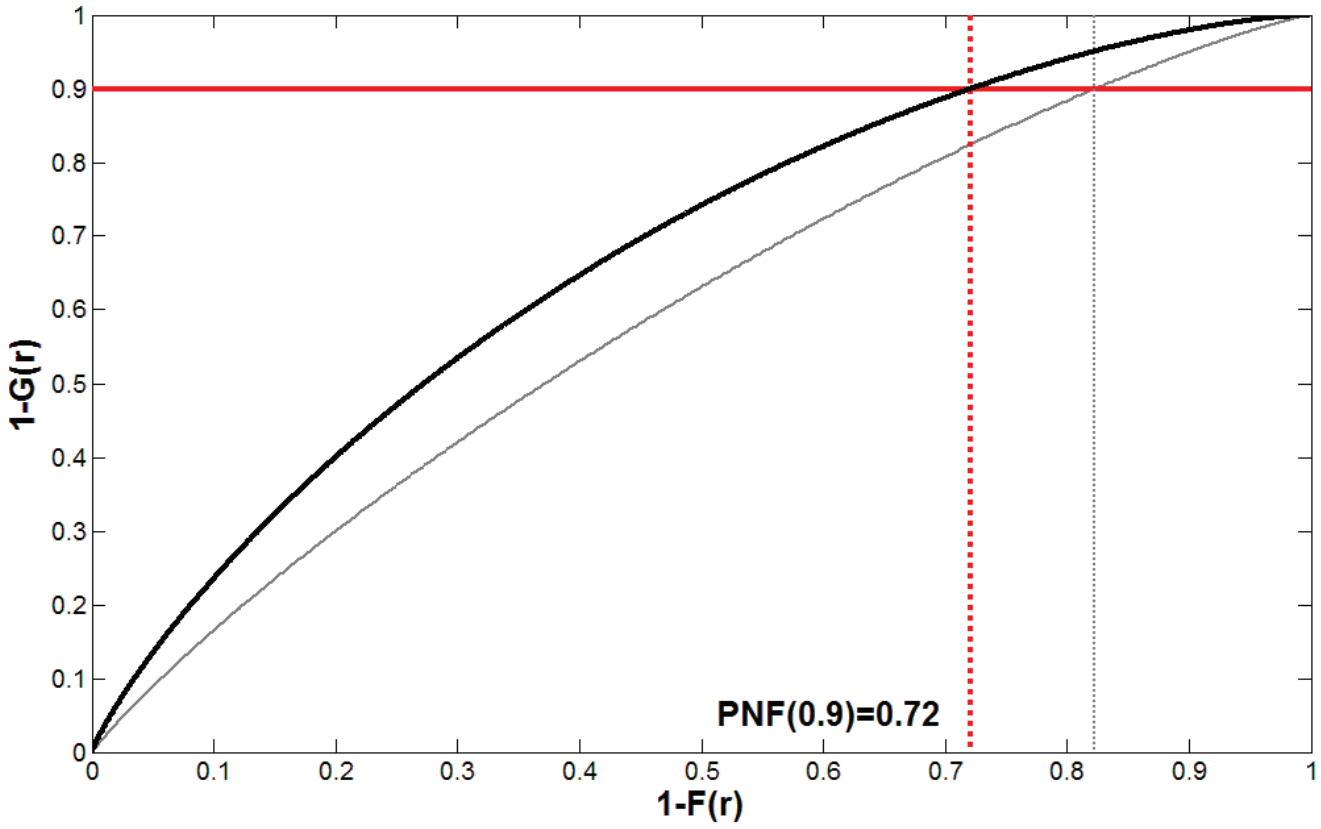
Thus

$$PNF(p) = 1- F \circ G^{-1}(1-p)$$

PNF, risk distribution *F* in population is Beta(8.5, 161.5) (AUC=0.59)



PNF, risk distribution *F* in population is Beta(2.3, 43.7) (AUC=0.68)



Non-parametric estimates of PCF(q), iPCF, PNF(p), iPNF using three different types of data

- Risks $r_1,\ldots,r_N$ in a **cohort under assumption of a well-calibrated model**
- Risks in **case-control study**, $r_1,\ldots,r_{N0}$ in controls ($Y=0$) and $r_1,\ldots,r_{N1}$ in cases ($Y=1$) with **known disease prevalence μ**
- **Risks and outcomes in a cohort**, $(r_1,Y_1),\ldots,(r_N,Y_N)$

Estimate PCF and PNF under assumption of well calibrated model

If model well calibrated : $P(Y \mid r) \sim r$ and $P(Y=1) = E(R) = \mu$

Distribution of risk in cases

$$G(r^*) = P(R \le r^* \mid Y=1) = \frac{1}{\mu}\int_0^{r^*} r dF(r)$$

Estimate PCF and PNF under assumption of well calibrated model

If model well calibrated : $P(Y \mid r) \sim r$ and $P(Y=1) = E(R) = \mu$

Distribution of risk in cases

$$G(r^*) = P(R \le r^* \mid Y=1) = \frac{1}{\mu}\int_0^{r^*} r dF(r)$$

PCF(q)=1- G $\circ$ F^{-1}(1-q) = $1- \frac{1}{\mu}\int_0^{F^{-1}(1-q)} r dF(r) = 1 - L(1-q)$

where L denotes the *Lorenz curve* for F

Estimate PCF and PNF under assumption of well calibrated model

If model well calibrated : $P(Y \mid r) \sim r$ and $P(Y=1) = E(R) = \mu$

Distribution of risk in cases

$$G(r^*) = P(R \le r^* \mid Y = 1) = \frac{1}{\mu}\int_0^{r^*} r dF(r)$$

$$\text{PCF(q)} = 1 - G \circ F^{-1}(1-q) = 1 - \frac{1}{\mu}\int_0^{F^{-1}(1-q)} r dF(r) = 1 - L(1-q)$$

where L denotes the Lorenz curve for F

$$PNF(p) = 1 - F \circ G^{-1}(1-p) = 1 - L^{-1}(1-p)$$

Non-parametric estimates of $\hat{L}$ and $\hat{L}^{-1}$ provided by Goldie (1977)

Inference

- Derived asymptotic distributions for non-parametric estimates of PCF, iPCF, PNF and iPNF for three designs
  - Observed risks alone under assumption of well calibrated model
  - Case-control data with known disease prevalence
  - Observed risks and outcomes in a population
- Variance estimation:
  - Taylor linearization using influence functions
  - Bootstrap procedure
  - **Estimation under assumption of well calibrated model much more efficient**

PCF and PNF Estimates for Endometrial and Breast Models in Validation Cohort (NHS)

	Breast cancer Model (AUC=0.58)	
q	$\hat{PCF}_R$	$\hat{PCF}_{(R,Y)}$
0.10	**0.17 (0.005)**	**0.17 (0.01)**

PCF and PNF Estimates for Endometrial and Breast Models in Validation Cohort (NHS)

	Breast cancer Model (AUC=0.58)		Endometrial cancer Model (AUC=0.68)	
q	$\hat{PCF}_R$	$\hat{PCF}_{(R,Y)}$	$\hat{PCF}_R$	$\hat{PCF}_{(R,Y)}$
0.10	0.17 (0.005)	0.17 (0.01)	**0.27 (0.02)**	**0.30 (0.03)**

PCF and PNF Estimates for Endometrial and Breast Models in Validation Cohort (NHS)

	Breast cancer Model (AUC=0.58)		Endometrial cancer Model (AUC=0.68)	
q	$\hat{PCF}_R$	$\hat{PCF}_{(R,Y)}$	$\hat{PCF}_R$	$\hat{PCF}_{(R,Y)}$
0.10	0.17 (0.005)	0.17 (0.01)	0.27 (0.02)	0.30 (0.03)
p	$\hat{PNF}_R$	$\hat{PNF}_{(R,Y)}$		
0.90	**0.84 (0.02)**	**0.84 (0.03)**		

PCF and PNF Estimates for Endometrial and Breast Models in Validation Cohort (NHS)

	Breast cancer Model (AUC=0.58)		Endometrial cancer Model (AUC=0.68)	
q	$\hat{PCF}_R$	$\hat{PCF}_{(R,Y)}$	$\hat{PCF}_R$	$\hat{PCF}_{(R,Y)}$
0.10	0.17 (0.005)	0.17 (0.01)	0.27 (0.02)	0.30 (0.03)
p	$\hat{PNF}_R$	$\hat{PNF}_{(R,Y)}$	$\hat{PNF}_R$	$\hat{PNF}_{(R,Y)}$
0.90	0.84 (0.02)	0.84 (0.03)	**0.76 (0.02)**	**0.78 (0.04)**

Summary

- Presented PCF, PNF, and their integrated versions (iPCF and iPNF) for assessing performance of risk prediction models for screening applications
- Presented non-parametric estimates for PCF, iPCF, PNF, iPNF and their asymptotic properties using influence function approach for three different designs
- Developed methods for testing differences between two models evaluated on same independent validation data
- High discriminatory accuracy needed for screening applications on population level

References

Pfeiffer RM. Extensions of criteria for evaluating risk prediction models for public health applications, *Biostatistics,* 2013

Pfeiffer RM, Gail MH. Two criteria for evaluating risk prediction models. *Biometrics*, 2011

Pfeiffer RM, Park Y, Kreimer AR et al, Risk predicting for breast, endometrial or ovarian cancer in white women aged 50 years or older: Derivation and validation from population-based cohort studies, *PLOS Medicine, 2013*