Absolute Risk: Methods and Applications in Clinical Care and Public Health, Day 2 Materials

Ruth Pfeiffer

Biostatistics Branch
Division of Cancer Epidemiology and Genetics
National Cancer Institute
National Institutes of Health
Bethesda, Maryland, USA
pfeiffer@mail.nih.gov

SISCER Module 12, 2024

Outline Day 2

1. Brief review of Day 1
2. Assessment of risk model performance
   - Assessing impact of population differences
   - Loss function based criteria
   - Comparing two models
3. Applications of absolute risk
4. Miscellaneous topics:
   - Updating risk models

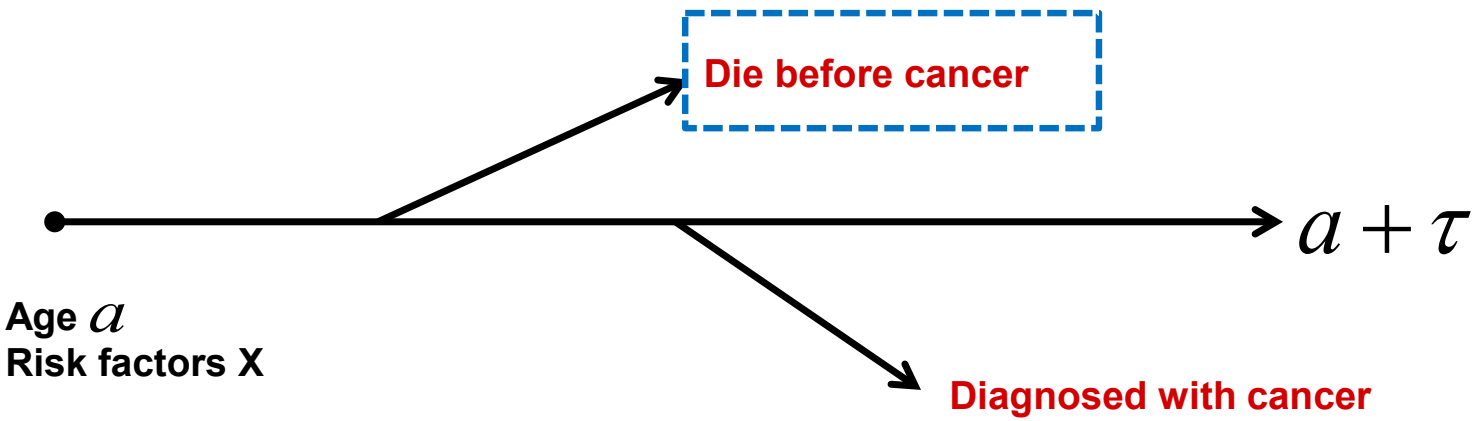
Definitions of "Risk"

- Relative Risk
- Probability of outcome:
  - Prevalence models
  - Probability of future event of cause 1 (e.g. cancer)
    - **"Pure risk":**
      $P(a < T \leq a + \tau, \text{cause=cancer} | T > a$, there are no competing risks)
    - **Absolute risk:** (or "crude risk", or "cumulative incidence")
      $P(a < T \leq a + \tau, \text{cause=cancer} | T > a$, there are competing risks)

Absolute Risk

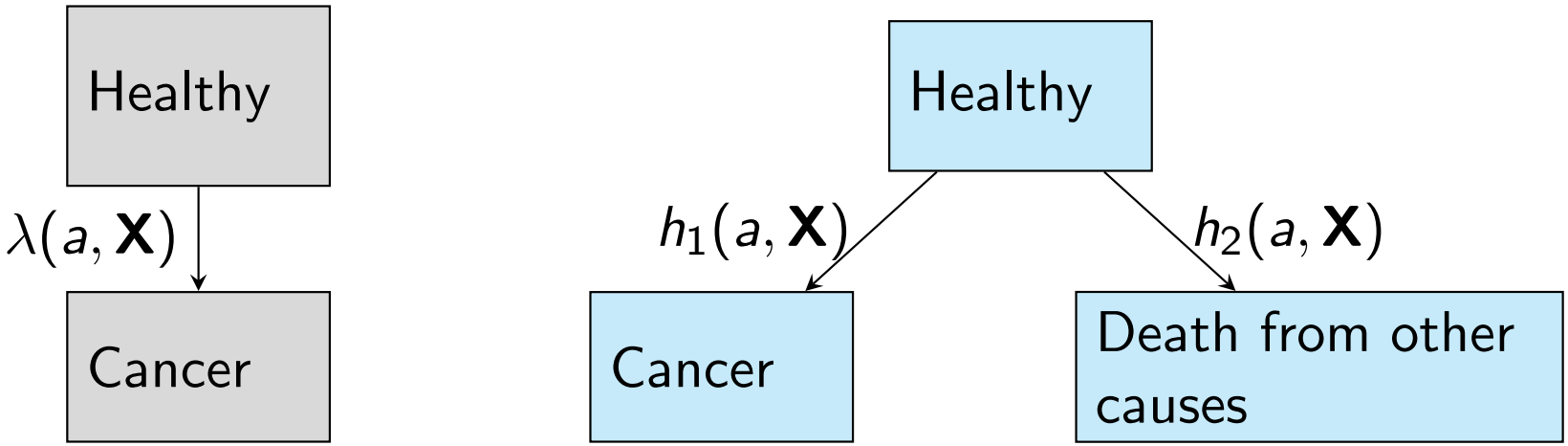
T: time to event

$R(a, a+\tau, \mathbf{X}) = P(a < T \leq a+\tau, \text{cause=1} | T > a,$ there are competing risks)



Modeling and Estimating Absolute Risk: Two Approaches

T: time to event, $\mathbf{X}$: risk factors
$R(a, a+\tau, \mathbf{X}) = P(a < T \leq a+\tau, \text{cause=1} | T > a, \mathbf{X})$



Cumulative incidence regression (Fine & Gray, JASA 1999)

Cause specific approach: model $h_m(a, \mathbf{X}), m = 1, 2$

Absolute Risk Model: Cause-specific Formulation

Focus on 2 types of events: 1=cancer, 2=death from other causes
Observe T= time to first event

$$h_m(t) = \lim_{\epsilon \downarrow 0} \frac{P\left(t \leq T < t + \epsilon,\ \text{cause} = m \middle| T \geq t\right)}{\epsilon}, m = 1, 2$$

$$R(a, a+\tau, \mathbf{X}) = P(T \leq a+\tau, \text{cause}{=}1 \,|\, T > a, \mathbf{X}) = \int_a^{a+\tau} h_1(t, \mathbf{X}) \exp\left[-\int_a^t \{h_1(v, \mathbf{X}) + h_2(v, \mathbf{X})\} \mathrm{d}v\right] \mathrm{d}t$$

- $\mathbf{X}$ - individual risk/protective factors
- τ - projection period

General Strategy to Estimate Cause Specific Absolute Risk

- Model cause specific hazards: $h_i(t, \mathbf{X}) = h_{0i}(t) rr(\beta_i' \mathbf{X}), i = 1, 2$
- Estimate cause specific hazards $\hat{h}_1(t, \mathbf{X})$, $\hat{h}_2(t, \mathbf{X})$
- Obtain "plug in" estimate

$$\hat{R}(a, a+\tau, \mathbf{X}) = \int_a^{a+\tau} \hat{h}_1(t, \mathbf{X}) \exp\left[- \int_a^t \{\hat{h}_1(s, \mathbf{X}) + \hat{h}_2(s, \mathbf{X})\} \mathrm{d}s \right] \mathrm{d}t$$

Comments

- Estimate absolute risk from cohort data
- Estimate absolute risk from sub-samples of cohorts (two-phase studies)
  - **Nested case-control design** (Langholz and Borgan, 1997): at each time a case develops sample individuals from risk set
  - **Case-cohort design** (Prentice and Self, 1988): analyze data from subcohort selected at start of follow-up and all cases observed during follow up
- Estimate absolute risk by combining relative risk estimates with registry data

Combine Data from Different Sources to Estimate $h_1(t, \mathbf{X})$

$h_1(t, \mathbf{X}) = h_{10}(t) \exp(\beta_1' \mathbf{X})$

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

Estimate relative risk, $\exp(\beta' \mathbf{X})$ and attributable risk, $AR(\mathbf{X})$

Disease Registries: data on incidence by age, sex, race, no risk factors

$h_{10}^*(t)$, age, sex, race specific **composite hazard**

Use that $AR(\mathbf{X}) = \frac{h_{10}^*(t) - h_{10}(t)}{h_{10}^*(t)}$ to get $\hat{h}_1(t, \mathbf{X}) = h_{10}^*(t)(1 - \widehat{AR}) \exp(\hat{\beta}_1' \mathbf{X})$

$$\hat{R}(a, a+\tau, \mathbf{X}) = \int_a^{a+\tau} \hat{h}_1(t, \mathbf{X}) \exp\left[- \int_a^t \{\hat{h}_1(s, \mathbf{X}) + \hat{h}_2(s)\} ds \right] dt$$

Example 1: Absolute Risk Model for Breast Cancer in General US Population Developed at NCI (Gail MH, et al., 1989, JNCI)

Risk factors:

- Age at menarche
- Age at first life birth
- Number of first degree relatives with breast cancer
- Number of breast biopsies
- Diagnosis of atypical hyperplasia

http://www.cancer.gov/bcrisktool/

Example 2: Absolute Breast Cancer Risk Model with Modifiable Risk Factors (BCmod) Pfeiffer RM, et al., 2013, PLos Medicine

Risk factors:

- Nulliparous
- Age at first life birth
- Family history of breast/ovarian cancer
- Benign breast disease/biopsy
- Age at menopause
- BMI
- Alcohol consumption
- Hormone replacement therapy use

Risk Model Validation

Before model can be recommended for practical use, need to understand its performance characteristics in independent data for rigorous assessment

- **Calibration (bias)**: Does model correctly predict number of observed events; requires cohort data
- **Classification accuracy:** How well does model categorize/classify individuals
- **Discrimination:** How different are risks in cases compared to non-cases? Quantified by $AUC = P(R_{case} > R_{noncase})$
- **Criteria to assess model performance for screening applications**: Proportion of cases followed (PCF), proportion needed to follow (PNF)

Risk model validation

Factors that can impact performance measures, and possibly cause misleading conclusions:

- Missing data on model predictors
- Incomplete follow-up in validation cohort
- Differences in populations used to build and validate model
  - Operational definition of covariates (e.g. number biopsies)
  - Procedures used to diagnose the disease outcome
  - Screening/outcome verification

Accommodating population differences when validating risk prediction models (Pfeiffer, Chen, Gail, Ankerst, Stat in Med, 2023)

Motivation: model to predict incident prostate cancer

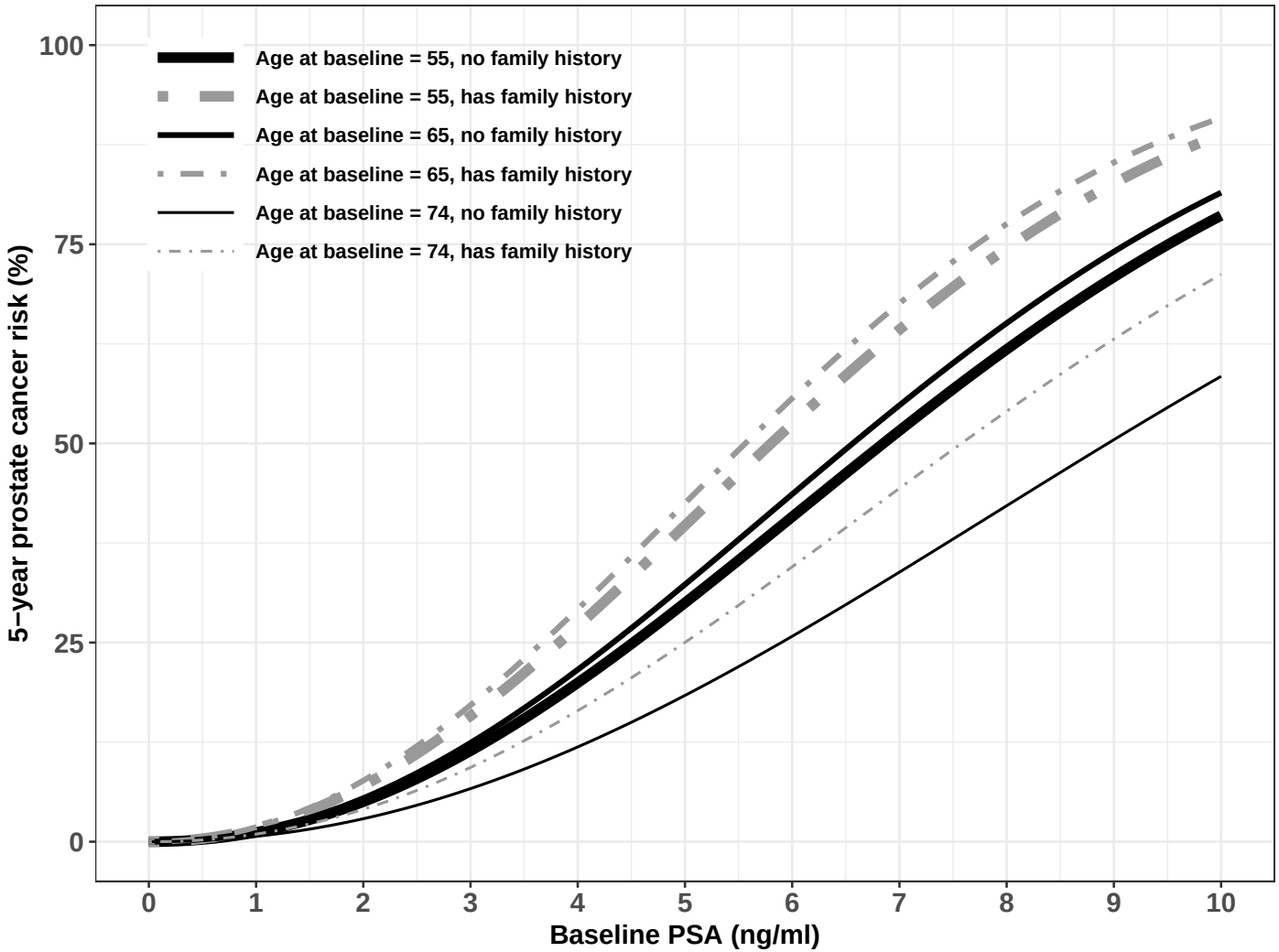
- **Cohort for model building (training data):** 29699 men in screening arm of Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial: annual PSA test for 6 yrs, annual digital rectal exam (DRE) for 4 yrs

Motivation: model to predict incident prostate cancer

- **Cohort for model building (training data):** 29699 men in screening arm of Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial: annual PSA test for 6 yrs, annual digital rectal exam (DRE) for 4 yrs
- **Risk model:** 5-year prostate cancer risk $R(\mathbf{X}, a)$ for man with risk factors $\mathbf{X}$ and age a

Risk factor X	HR (95% CI)
$log_2 PSA$	4.32 (4.09, 4.58)
Family history	1.42 (1.22, 1.65)
Prior negative biopsy	1.83 (1.17, 2.87)
$log_2 PSA * Prior\ negative\ biopsy$	0.61 (0.50, 0.74)

Estimated 5-year prostate cancer risk for select profiles



Motivation: model to predict incident prostate cancer

- **Cohort for model building (training data):** 29699 men in screening arm of PLCO Cancer Screening Trial: annual PSA test for 6 yrs, annual digital rectal exam (DRE) for 4 yrs
- **Risk model:** 5-year prostate cancer risk $R(\mathbf{X}, a)$
- **Validation cohort:** 26422 men in Selenium and Vitamin E Cancer Prevention Trial (SELECT); PSA ≤ 4 and normal DRE at enrollment; 6-month recommended PSA and DRE screens

Motivation: model to predict incident prostate cancer

- **Cohort for model building (training data):** 29699 men in screening arm of PLCO Cancer Screening Trial: annual PSA test for 6 yrs, annual digital rectal exam (DRE) for 4 yrs
- **Risk model:** 5-year prostate cancer risk $R(\mathbf{X}, a)$
- **Validation cohort:** 26422 men in Selenium and Vitamin E Cancer Prevention Trial (SELECT); PSA ≤ 4 and normal DRE at enrollment; 6-month recommended PSA and DRE screens

- Risk factor distributions different because of differing eligibility criteria

Motivation: model to predict incident prostate cancer

- **Cohort for model building (training data):** 29699 men in screening arm of PLCO Cancer Screening Trial: annual PSA test for 6 yrs, annual digital rectal exam (DRE) for 4 yrs
- **Risk model:** 5-year prostate cancer risk $R(\mathbf{X}, a)$
- **Validation cohort:** 26422 men in Selenium and Vitamin E Cancer Prevention Trial (SELECT); PSA ≤ 4 and normal DRE at enrollment; 6-month recommended PSA and DRE screens

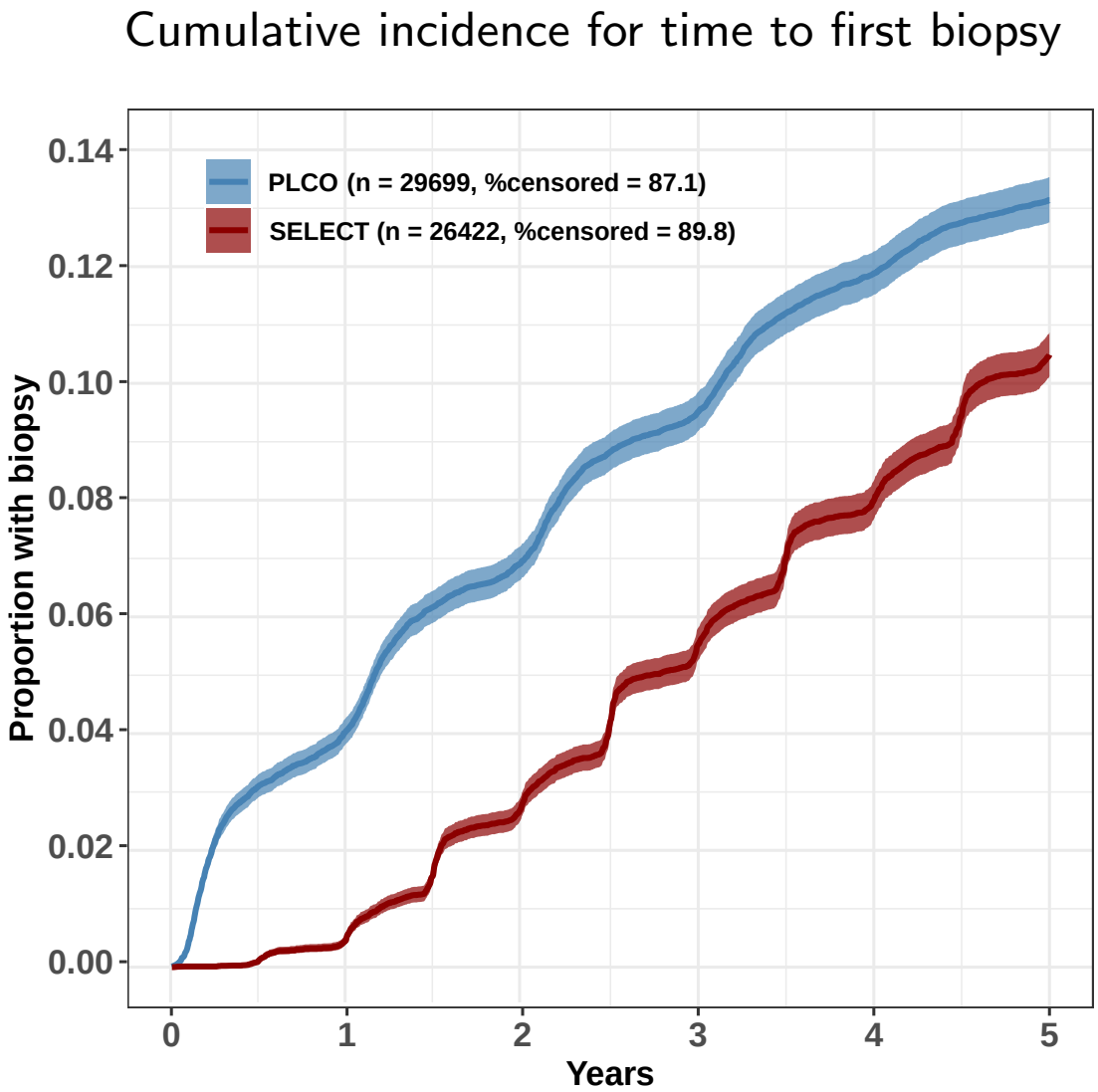
- Risk factor distributions different because of differing eligibility criteria
- Recommended for prostate biopsy if PSA > 4: Diagnosis depends on biopsy; different screening

Differences in training and validation cohorts

Baseline characteristics of PLCO and SELECT participants

	PLCO ($N_1 = 29699$) $n(\%)$	SELECT ($N_0 = 26422$) $n(\%)$
First PSA		
- [0, 1]	13454 (45.3)	12831 (48.6)
⋮		
- (4, 10]	2007 (6.8)	0 (0.0)
Digital rectal exam		
- Suspicious	2119 (7.1)	0 (0.0)
African ancestry		
- Yes	1155 (3.9)	2779 (10.5)
Family history		
- Yes	2288 (7.7)	4623 (17.5)

Differences in training and validation cohorts



Accommodating population differences when validating risk prediction models

Y disease outcome

$R(\mathbf{X}) = \hat{P}(Y = 1|\mathbf{X})$ risk model developed in training data, $T = 1$

$T = 0$ validation population

$\pi_T(\mathbf{X}^*) = P(Y = 1|\mathbf{X}^*, T)$ true disease probabilities, $\mathbf{X}^* = (\mathbf{X}, \mathbf{Z})$

Accommodating population differences when validating risk prediction models

Y disease outcome

$R(\mathbf{X}) = \hat{P}(Y = 1|\mathbf{X})$ risk model developed in training data, $T = 1$

$T = 0$ validation population

$\pi_T(\mathbf{X}^*) = P(Y = 1|\mathbf{X}^*, T)$ true disease probabilities, $\mathbf{X}^* = (\mathbf{X}, \mathbf{Z})$

Model R **well calibrated** if

$$C = \frac{\mathcal{O}}{\mathcal{E}} = \frac{E(Y)}{E(R)} = \frac{\int_{\mathbf{x}^*} \pi(\mathbf{x}^*) dF(\mathbf{x}^*)}{\int_{\mathbf{x}} R(\mathbf{x}) dF(\mathbf{x})} = 1$$

Assumption: R well calibrated in training data, $C_1 = 1$

Motivation: model to predict incident prostate cancer

Calibration in PLCO estimated using 5-fold cross-validation:

$$\hat{C}_1 = \frac{\sum_i Y_i}{\sum_i R(\mathbf{x}_i)} = 0.99 \quad (0.95, 1.04)$$

Motivation: model to predict incident prostate cancer

Calibration in PLCO estimated using 5-fold cross-validation:

$$\hat{C}_1 = \frac{\sum_i Y_i}{\sum_i R(\mathbf{x}_i)} = 0.99 \quad (0.95, 1.04)$$

Calibration in SELECT: $\hat{C}_0 = 1.19 \quad (1.13, 1.26)$

1. Assessing impact of different predictor distributions in training and validation data on performance measures

Assume individual level-data on $\mathbf{X}^*$ for both cohorts available

Define *selection weights*

$$w(\mathbf{X}^*) = \frac{dF_1(\mathbf{X}^*)}{dF_0(\mathbf{X}^*)} = \frac{P(\mathbf{X}^*|T=1)}{P(\mathbf{X}^*|T=0)}$$

1. Assessing impact of different predictor distributions in training and validation data on performance measures

Assume individual level-data on $\mathbf{X}^*$ for both cohorts available

Define *selection weights*

$$w(\mathbf{X}^*) = \frac{dF_1(\mathbf{X}^*)}{dF_0(\mathbf{X}^*)} = \frac{P(\mathbf{X}^*|T=1)}{P(\mathbf{X}^*|T=0)} = \frac{P(T=1|\mathbf{X}^*)P(T=0)}{P(T=0|\mathbf{X}^*)P(T=1)}$$

1. Assessing impact of different predictor distributions in training and validation data on performance measures

Assume individual level-data on $\mathbf{X}^*$ for both cohorts available

Define *selection weights*

$$w(\mathbf{X}^*) = \frac{dF_1(\mathbf{X}^*)}{dF_0(\mathbf{X}^*)} = \frac{P(\mathbf{X}^*|T=1)}{P(\mathbf{X}^*|T=0)} = \frac{P(T=1|\mathbf{X}^*)P(T=0)}{P(T=0|\mathbf{X}^*)P(T=1)}$$

and *selection weighted calibration measure*

$$C_0^W = \frac{E_0[Yw(\mathbf{X}^*)]}{E_0[R(\mathbf{X})w(\mathbf{X}^*)]}$$

Reproducibility and transportability of model $R(\mathbf{X})$ (! Math)

Unweighted (C_0) and selection weighted (C_0^W) calibration measures $\pi_T(\mathbf{X}^*) = P(Y = 1|\mathbf{X}^*, T),\ T = 0, 1,\ \mathbf{X}^* = (\mathbf{X}, \mathbf{Z})$

Assumptions	**Risk factor distributions**	$\mathbf{C_0}$	$\mathbf{C_0^W}$
$\pi_1(\mathbf{X}^*) = \pi_0(\mathbf{X}^*)$ and $F_0(\mathbf{Z}\|\mathbf{X}) = F_1(\mathbf{Z}\|\mathbf{X})$	**Reproducibility** $F_0(\mathbf{X}) = F_1(\mathbf{X})$	$= C_1 = 1$	$= C_1 = 1$
	Transportability $F_0(\mathbf{X}) \neq F_1(\mathbf{X})$ R well calibr.	$\neq C_1$	$= C_1 = 1$

Estimating selection weighted performance measures

Selection weights From combined training and validation data estimate

$$P(T = 1|\mathbf{X}^*, \gamma) = \frac{\exp(\gamma_0 + \gamma_1{}'\mathbf{X} + \gamma_2{}'\mathbf{Z})}{1 + \exp(\gamma_0 + \gamma_1{}'\mathbf{X} + \gamma_2{}'\mathbf{Z})}$$

estimated weights for individuals $i = 1, \ldots, N_0$ in validation data

$$\hat{w}(\mathbf{x}_i^*) = w(\mathbf{x}_i^*, \hat{\gamma}) = \frac{\hat{P}(T = 1|\mathbf{x}_i^*)N_0}{\hat{P}(T = 0|\mathbf{x}_i^*)N_1} = \exp(\hat{\gamma}_0 + \hat{\gamma}_1{}'\mathbf{x}_i + \hat{\gamma}_2{}'\mathbf{z}_i)\frac{N_0}{N_1}$$

Estimating selection weighted performance measures

Selection weights From combined training and validation data estimate

$$P(T = 1|\mathbf{X}^*, \gamma) = \frac{\exp(\gamma_0 + \gamma_1{}'\mathbf{X} + \gamma_2{}'\mathbf{Z})}{1 + \exp(\gamma_0 + \gamma_1{}'\mathbf{X} + \gamma_2{}'\mathbf{Z})}$$

estimated weights for individuals $i = 1, \ldots, N_0$ in validation data

$$\hat{w}(\mathbf{x}_i^*) = w(\mathbf{x}_i^*, \hat{\gamma}) = \frac{\hat{P}(T = 1|\mathbf{x}_i^*)N_0}{\hat{P}(T = 0|\mathbf{x}_i^*)N_1} = \exp(\hat{\gamma}_0 + \hat{\gamma}_1{}'\mathbf{x}_i + \hat{\gamma}_2{}'\mathbf{z}_i)\frac{N_0}{N_1}$$

Weighted calibration measure

$$\widehat{C}_0^W = \frac{\sum_{i=1}^{N_0} Y_i \hat{w}(\mathbf{x}_i^*)}{\sum_{i=1}^{N_0} R(\mathbf{x}_i)\hat{w}(\mathbf{x}_i^*)}$$

Selection weight model, $w(\mathbf{X})$

Odds ratios (ORs) and 95% confidence intervals (CIs) from stepwise logistic selection weight model with cohort selection outcome T (1:in PLCO versus 0: in SELECT) applied to 56121 combined participants (29699 from PLCO, 26422 from SELECT).

Risk factor	OR	95% CI
Intercept	2.14	(1.72, 2.65)
$log_2 PSA$	0.78	(0.64, 0.94)
Age	0.993	(0.990, 0.997)
Family history	0.39	(0.37, 0.41)
African ancestry	0.33	(0.30, 0.35)
Prior negative biopsy	0.43	(0.40, 0.47)
$log_2 PSA * Age$	1.006	(1.003, 1.009)
$log_2 PSA * African\ ancestry$	1.15	(1.08, 1.23)
$log_2 PSA * Prior\ negative\ biopsy$	1.18	(1.11, 1.25)

Assessing impact of different disease verification in training and validation data on performance measures

- $V = 1$ if disease verified, $V = 0$ otherwise
- Observe $P_T(Y = 1, V = 1|\mathbf{X}^*)$ instead of $P_T(Y = 1|\mathbf{X}^*)$
- Assume $\widetilde{C}_1 = \frac{E_1(YV)}{E_1\{R(\mathbf{X})\}} = 1$
- Define verification weight and verification weighted calibration statistic

$$v(\mathbf{X}^*) = \frac{P_1(V = 1|\mathbf{X}^*)}{P_0(V = 1|\mathbf{X}^*)} \text{ and } \widetilde{C}_0^V = \frac{E_0\{YVv(\mathbf{X}^*)\}}{E_0\{R(\mathbf{X})\}}$$

Assessing impact of different disease verification in training and validation data on performance measures

- $V = 1$ if disease verified, $V = 0$ otherwise
- Observe $P_T(Y = 1, V = 1|\mathbf{X}^*)$ instead of $P_T(Y = 1|\mathbf{X}^*)$
- Assume $\widetilde{C}_1 = \frac{E_1(YV)}{E_1\{R(\mathbf{X})\}} = 1$
- Define verification weight and verification weighted calibration statistic

$$v(\mathbf{X}^*) = \frac{P_1(V = 1|\mathbf{X}^*)}{P_0(V = 1|\mathbf{X}^*)} \text{ and } \widetilde{C}_0^V = \frac{E_0\{YVv(\mathbf{X}^*)\}}{E_0\{R(\mathbf{X})\}}$$

If $P_0(Y = 1|\mathbf{X}^*, V = 1) = P_1(Y = 1|\mathbf{X}^*, V = 1)$ or if $P_T(Y = 1|\mathbf{X}^*, V = 1) = P_T(Y = 1|\mathbf{X}^*)$ then $C_0^V = 1$

Estimating verification weighted performance measures

Verification weights Compute $P_T(V = 1|\mathbf{X}^*)$ or $P_T(V(\tau) = 1|\mathbf{X}^*)$ separately in training and verification data $T = 0, 1$, e.g.

$$\hat{P}_T(V = 1|\mathbf{X}^*) = \frac{\exp(\eta_{T0} + +\boldsymbol{\eta}'_{T1}\mathbf{X} + \boldsymbol{\eta}'_{T2}\mathbf{Z})}{1 + \exp(\eta_{T0} + +\boldsymbol{\eta}'_{T1}\mathbf{X} + \boldsymbol{\eta}'_{T2}\mathbf{Z})}$$

Alternatively fit survival model $S(t, \mathbf{X}^*)$ to time to disease verification and compute $\hat{P}_T(V(\tau) = 1|\mathbf{X}^*) = 1 - \hat{S}_T(\tau, \mathbf{X}^*), T = 0, 1$
Verification weights for individuals in validation data

$$\hat{v}(\mathbf{x}_i^*) = \frac{\hat{P}_1(V = 1|\mathbf{x}_i^*)}{\hat{P}_0(V = 1|\mathbf{x}_i^*)}, i = 1, \ldots, N_0$$

Verification probability models, $P(V = 1|\mathbf{X}, T)$

Odds ratios (ORs) and 95% CIs from logistic models for verification of outcome, i.e. biopsy performed within first five years of study (1 yes, 0 no), fit to 29699 PLCO and 26422 SELECT participants

	PLCO ($N_1 = 29699$)	SELECT ($N_0 = 26422$)
Risk factor	OR (95% CI)	OR (95% CI)
Intercept	0.005 (0.002, 0.014)	0.015 (0.013, 0.017)
$log_2 PSA$	4.73 (3.73, 6.03)	4.46 (4.06, 4.92)
$I(PSA > 4)$	350.38 (93.18,1320.18)	28.20 (15.32, 50.92)
DRE	511.01 (364.03, 735.38)	86.35 (69.22, 107.94)
Age	0.98 (0.97, 1.00)	-
Family history	1.22 (1.03, 1.46)	1.37 (1.21, 1.54)
African ancestry	0.010 (0.001, 0.219)	-
Prior negative biopsy	0.81 (0.68, 0.97)	1.77 (1.33, 2.34)
$log_2 PSA * I(PSA > 4)$	0.78 (0.60, 1.01)	0.38 (0.29, 0.49)
$log_2 PSA * DRE$	0.27 (0.21, 0.35)	0.31 (0.26, 0.37)
$log_2 PSA * Prior\ negative\ biopsy$	-	0.76 (0.65, 0.89)
$I(PSA > 4) * DRE$	0.11 (0.07, 0.17)	0.62 (0.35, 1.12)
$I(PSA > 4) * Age$	0.97 (0.95, 0.99)	-
$Age * African\ ancestry$	1.08 (1.02, 1.13)	-

Validation results for PLCO prostate cancer risk model in 26422 SELECT participants

Estimated unweighted and weighted calibration ratios (C_0, C_0^W, C_0^V, C_0^{WV}) with bootstrap 95% confidence intervals (CIs)

	Estimate	95% CI
C_0	1.19	(1.13, 1.26)
C_0^W	1.16	(1.09, 1.22)
C_0^V	0.89	(0.84, 0.95)
C_0^{WV}	0.88	(0.82, 0.94)

Validation results for PLCO prostate cancer risk model in 26422 SELECT participants

Estimated unweighted and weighted calibration ratios (C_0, C_0^W, C_0^V, C_0^{WV}) with bootstrap 95% confidence intervals (CIs)

	Estimate	95% CI
C_0	1.19	(1.13, 1.26)
C_0^W	1.16	(1.09, 1.22)
C_0^V	0.89	(0.84, 0.95)
C_0^{WV}	0.88	(0.82, 0.94)
AUC_0	0.828	(0.817, 0.840)
AUC_0^W	0.824	(0.812, 0.835)
AUC_0^V	0.853	(0.842, 0.865)
AUC_0^{WV}	0.851	(0.839, 0.862)

Comment: key assumption is that weights are modeled correctly

Summary

Proposed selection and verification weights measures of risk model performance so that weighted validation data more closely resemble training data w.r.t. risk factor distributions and disease ascertainment

- Defined selection and verification weighted measures
- Formalized notion of reproducibility and transportability of a risk model
- Assumption: correctly modeled weights

Back to Model Performance Assessment

Model Assessment Based on Expected Costs for Particular Application

Recall: Using risk threshold t for decision making

Measure how well true outcome predicted

- Sensitivity $sens(t) = P(\hat{Y} = 1 | Y = 1) = P(R \geq t | Y = 1)$

Recall: Using risk threshold t for decision making

Measure how well true outcome predicted

- Sensitivity $sens(t) = P(\hat{Y} = 1|Y = 1) = P(R \geq t|Y = 1)$
- Specificity $spec(t) = P(\hat{Y} = 0|Y = 0) = P(R < t|Y = 0)$

Model Assessment Based on Expected Costs for Particular Application, the 2×2 setting

Two health states, $Y = 0, 1$, and two intervention options; $\pi = P(Y = 1)$, R= risk estimate, C= cost

Intervene	**Disease**	**Costs**	**Risk criterion at threshold t**	**Outcome probability**
Yes	Yes	C_{TP}	$R > t$	$\pi \times sens(t)$
No	Yes	C_{FN}	$R \leq t$	$\pi \times \{1 - sens(t)\}$
Yes	No	C_{FP}	$R > t$	$(1 - \pi)\{1 - spec(t)\}$
No	No	C_{TN}	$R \leq t$	$(1 - \pi)spec(t)$

TP, true positive; FN, false negative; FP, false positive; TN, true negative

Model Assessment Based on Expected Costs for Particular Application

Intervene	**Disease**	**Costs**	**Risk criterion at threshold t**	**Outcome probability**
Yes	Yes	C_{TP}	$R > t$	$\pi \times sens(t)$
No	Yes	C_{FN}	$R \leq t$	$\pi \times \{1 - sens(t)\}$
Yes	No	C_{FP}	$R > t$	$(1 - \pi)\{1 - spec(t)\}$
No	No	C_{TN}	$R \leq t$	$(1 - \pi)spec(t)$

TP, true positive; FN, false negative; FP, false positive; TN, true negative

Expected loss in population:

$$\bar{C}(t) = \pi \times sens(t)C_{TP} + \pi \times \{1 - sens(t)\}C_{FN} + (1 - \pi)\{1 - spec(t)\}C_{FP} + (1 - \pi)\{spec(t)\}C_{TN} + C_{test}$$

Threshold t^* to Minimize Expected Cost

$B_{case} = C_{FN} - C_{TP} \geq 0$ is net benefit of intervening on a case
$B_{non-case} = C_{FP} - C_{TN} \geq 0$ is net cost from intervening on a non-case
Risk threshold t^* that minimizes $\bar{C}(t)$ is

$$t^* = \frac{B_{non-case}}{B_{non-case} + B_{case}}$$

This threshold does not depend on risk model R, only on costs

Threshold t^* to Minimize Expected Cost

$B_{case} = C_{FN} - C_{TP} \geq 0$ is net benefit of intervening on a case
$B_{non-case} = C_{FP} - C_{TN} \geq 0$ is net cost from intervening on a non-case
Risk threshold t^* that minimizes $\bar{C}(t)$ is

$$t^* = \frac{B_{non-case}}{B_{non-case} + B_{case}}$$

This threshold does not depend on risk model R, only on costs

Key assumption: Model R is well calibrated

Pauker and Kassirer, NEJM 1975

Minimal Expected Cost (at t^*)

$$\bar{C}_{min} = \pi C_{FN} + (1-\pi) C_{FP} + C_{test} - \pi \times sens(t^*) B_{case} - (1-\pi) \times spec(t^*) B_{non-case}$$

$$\bar{C}_{perfect} = \pi C_{TP} + (1-\pi) C_{TN} + C_{test}$$

$$\bar{C}_{all} = \pi C_{TP} + (1-\pi) C_{FP}$$

$$\bar{C}_{none} = \pi C_{FN} + (1-\pi) C_{TN}$$

Example: Decision to Take Tamoxifen in 100,000 Women with Uteri, Aged 50-59

Relative risks (RRs) from tamoxifen treatment for various health outcomes and absolute numbers of health outcomes expected in 5 years with and without tamoxifen

Health events	**RR** [a]	**# cases no tamoxifen**	**# cases tamoxifen**
Invasive breast cancer	0.51	246	125.8
Hip fracture	0.55	101.6	55.9
Endometrial cancer	44.01	81.4	326.4
Stroke	1.59	110	174.9
Pulmonary emboli	3.01	50	150.5
Total		589.6	833.5
Total non-breast cancer		343	707.7

Optimal Threshold for Tamoxifen for Women Aged 50-54 Years

Aggregated non-breast cancer rates:
343.0 without tamoxifen
707.7 with tamoxifen
Cost = total number of life-threatening events in 100 000 women

Decision for threshold t	Breast cancer	No breast cancer
$r \geq t$ (give tamoxifen)	51 000+707.7=51 707.7	707.7
$r < t$ (dont give tamoxifen)	100 000+343=100 343	343

$B_{case} = 100,343 - 51,707.7 = 48,635.8$
$B_{non-case} = 707.7 - 343 = 364.7$
$t^* = 364.7/(364.7 + 48635.3) = 744x10^{-5}$ (=3.72% in 5 years)

Minimized expected losses (NHS validation data)

$\bar{C} = \frac{1}{n}\sum_i \{I(r_i \geq t^*, Y_i = 1) \times 51707.7 + I(r_i \geq t^*, Y_i = 0) \times 707.7 + I(r_i < t^*, Y_i = 1) \times 100343 + I(r_i < t^*, Y_i = 0) \times 343\}$

For BCmod $\bar{C} = 1818.17$

With $\pi = 0.01475$, a perfect risk model would have expected loss
$0.01475 \times 51,707.7 + (1 - 0.01475) \times 343 = 1100.6$

Using no tamoxifen yields expected loss
$0.01475 \times 100,343 + (1 - 0.01475) \times 343 = 1818.00$

Example: Multiple outcomes in prevention trials

Relative risks (RRs) from tamoxifen treatment for various health outcomes and absolute numbers of health outcomes expected in 5 years with and without tamoxifen in population of 10,000 white 40-year-old women with uteri and a projected breast cancer risk of 2%

Health events	**RR** [a]	**None get tamoxifen**	**All get tamoxifen**	**Prevented by tamoxifen**
Invasive breast cancer	0.51	200	103	97
Hip fracture	0.55	2	1	1
Endometrial cancer	2.53	10	26	-16[b]
Stroke	1.59	22	35	-13
Pulmonary emboli	3.01	7	22	-15
Net life-threatening events		241	187	**54**
In situ breast cancer	0.50	106	53	53
Deep vein thrombosis	1.60	24	39	-15
Net serious events		130	92	**38**

[a] From Fisher et al., 1998.

[b] Minus sign means tamoxifen increases number of events

Example: Multiple outcomes in prevention trials, cont.

Under costs c_k for outcomes $k = 1, 2, \ldots, 7$, can determine net benefit from taking tamoxifen by testing

$$\text{Net Benefit} = \sum_{k=1}^{K} c_k P_{0k} - \sum_{k=1}^{7} c_k P_{1k} = \sum_{k=1}^{7} c_k (P_{0k} - P_{1k}) > 0.$$

P_{0k}/P_{1k}: probability of outcome k without/with tamoxifen
Net benefit: expected cost without tamoxifen minus expected cost under tamoxifen
Gail et al, 1999 used $P_{1k} = RR_k P_{0k}$
$c_k = 1$ for $k = 1, 2, \ldots, 5$ (life-threatening events) and $c_k = 0.5$ for $k = 6, 7$ (serious events)
Young women with high BC risk had positive net benefits, older women tended to have negative net benefit due to higher baseline risks P_{0k} for stroke, pulmonary emboli, deep vein thromboses and endometrial cancer (Gail et al, 1999); for raloxifene see Freedman et al. (2011)

Benefit/risk indices for tamoxifen and raloxifene for white non-Hispanic women with a uterus

	Tamoxifen			Raloxifene		
5-year risk	50-59	60-69	70-79	50-59	60-69	70-79
1.5	-133	-310	-325	21	-11	-15
2.0	-105	-283	-298	43	11	7
2.5	-78	-255	-271	65	33	29
3.0	-51	-228	-244	86	55	51
3.5	-25	-202	-217	108	76	71
4.0	3	-175	-190	128	97	93
4.5	29	-148	-164	150	119	115
5.0	56	-121	-137	172	140	136
5.5	83	-95	-111	193	161	157
6.0	109	-69	-84	214	183	179
6.5	135	-42	-58	236	204	199
7.0	162	-15	-32	256	225	221

- Strong evidence of benefits outweighing risks
- Moderate evidence of benefits outweighing risks
- Benefits do not outweigh risks

5-year projected risk of IBC is ≥ 1.67%.

Using BCPT data and WHI baseline rates

Combining RR from BCPT and STAR using WHI baseline rates

Freedman AN et al. JCO 2011;29:2327-2333

JOURNAL OF CLINICAL ONCOLOGY

What if Costs (and Threshold) Hard to Define? Use "Decision Curve"

Solve

$$t^* = \frac{B_{non-case}}{B_{non-case} + B_{case}}$$

for $B_{non-case}/B_{case}$, get

$$B_{non-case}/B_{case} = t^*/(1 - t^*)$$

What if Costs (and Threshold) Hard to Define? Use "Decision Curve"

Solve

$$t^* = \frac{B_{non-case}}{B_{non-case} + B_{case}}$$

for $B_{non-case}/B_{case}$, get

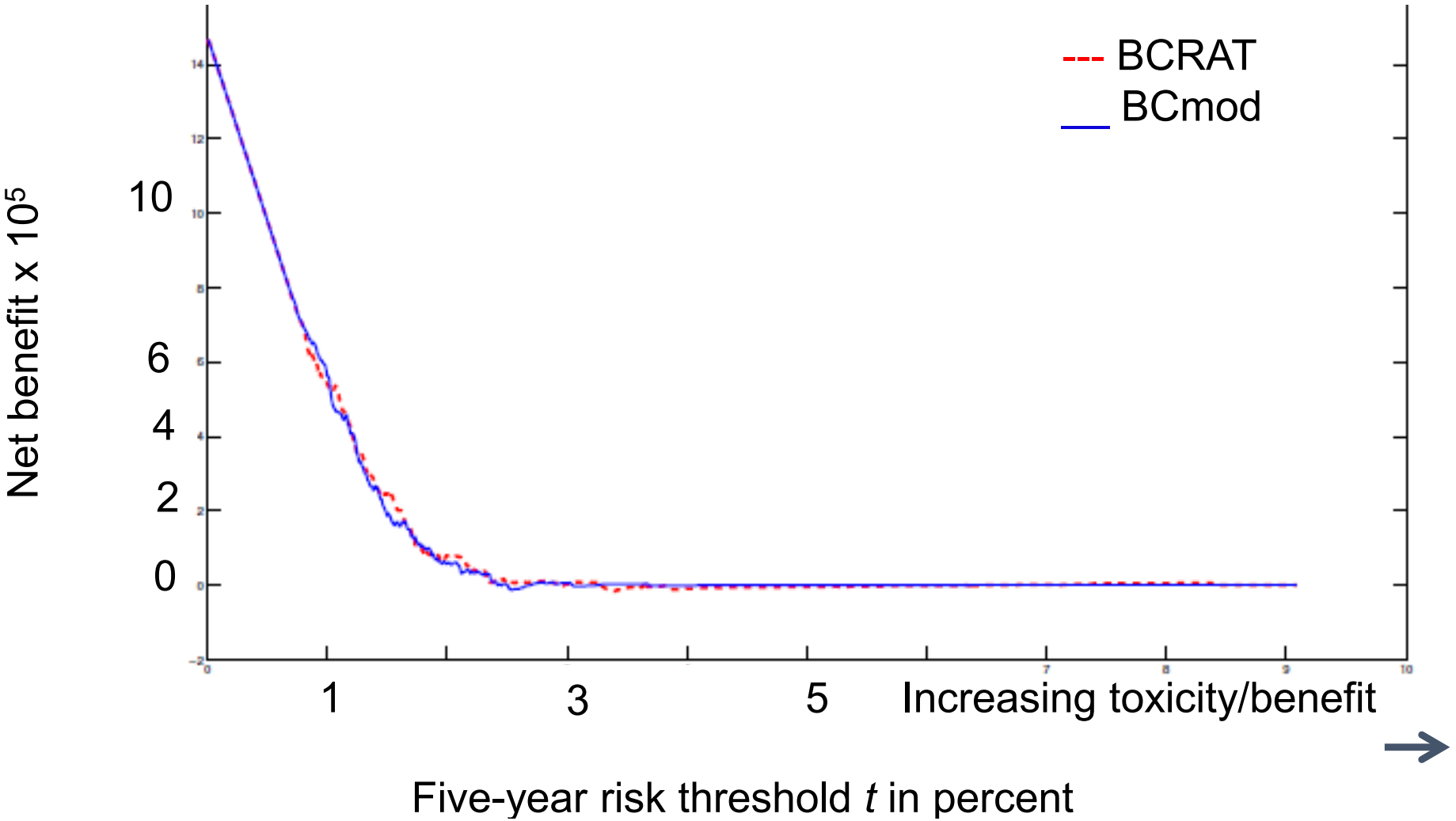
$$B_{non-case}/B_{case} = t^*/(1 - t^*)$$

Net benefit $NB(t)$ studies a range of $B_{non-case}/B_{case}$ ratios

$$NB(t) = \pi \times sens(t) - (1 - \pi)\{1 - spec(t)\}\{t/(1 - t)\}$$

is implicitly a function of cost ratios and sensitivity and specificity as threshold t

Net Benefit or Relative Utility Curve



Some Comments on Net Benefit Curve

- Miscalibration of risk model in target population lowers the decision curve (reduces net benefit)
- If the model is mis-calibrated, thresholds will not correspond to correct cost ratios $B_{non-case}/B_{case} = t^*/(1-t^*)$ leading to incorrect estimates of net benefit
- If there are subgroups in a population with different cost ratios, separate decision curves should be used for each subgroup.

Kerr et al. JCO 2017; Pepe et al. Stat Bioscience 2015; Van Calster and Vickers Med Decision Making, 2015

Comparing Two Absolute Risk Prediction Models

NCI Absolute Breast Cancer Risk Models

Empirical models; estimated by combining relative risks with cancer registry data on incidence

- BCRAT ("Gail model"; Gail et al, 1989)
  - Risk factors: age, age at menarche, age at first life birth, number affected mother or sister, number of breast biopsies+diagnosis of atypical hyperplasia
- BCmod (Pfeiffer et al, 2013)
  - Adds modifiable risk factors to BCRAT factors: body mass index (BMI), alcohol consumptions, hormone replacement therapy use (HRT)

Absolute Risk Estimates: Two 60 Year-old Women

Model Predictor	**Woman 1**	**Woman 2**
Age at menarche	12	12
Age at first birth	25	25
Number of life births	3	3
Past diagnosis of benign breast disease	no	no
Family history breast/ovarian cancer	no	no
Menopausal	yes (age 50)	yes (age 55)
Hormone replacement therapy use (duration)	no	yes (5yrs)
BMI, kg/m^2	24	35
Alcohol consumption (drinks/day)	0	> 1
5 year BCRAT absolute risk estimate	**1.6%**	**1.6%**
5 year BCmod absolute risk estimate	**1.1%**	**3.8%**

Comparing Two Absolute Risk Prediction Models

- Compare calibration
- Compare accuracy measures
- Compare AUC values
- Compare PCF and PNF or iPCF and iPNF
- Criteria based on reclassification tables
- Net reclassification improvement
  - Integrated discrimination improvement
  - Compare expected losses

Compare calibration

Comparison of observed (O) and expected (E) incident breast cancers with 95% CIs based on 5-year predictions from BCmod and BCRAT in women ages 50–55 in NHS validation cohort.

			BCmod		BCRAT
	O	**E**	**E/O (95% CI)**	**E**	**E/O (95% CI)**
All women	252	231	0.92 (0.72, 1.04)	238	0.94 (0.84, 1.07)
Age at menarche					
< 12	59	53	0.90 (0.70, 1.16)	59	1.00 (0.77, 1.29)
$12 - 13$	146	136	0.93 (0.79, 1.09)	140	0.96 (0.81, 1.12)
≥ 14	47	42	0.89 (0.67, 1.19)	40	0.85 (0.64, 1.13)
BMI					
$< 25 kg/m^2$	144	122	0.85 (0.72, 1.00)	130	0.90 (0.77, 1.06)
25 to $< 30 kg/m^2$	77	69	0.89 (0.71, 1.12)	70	0.91 (0.73, 1.13)
30 to $< 35 kg/m^2$	18	27	1.48 (0.93, 2.35)	26	1.43 (0.90, 2.27)
$\geq 35 kg/m^2$	13	3	0.26 (0.15, 0.44)	12	0.95 (0.55, 1.64)
Benign breast disease					
no	93	105	1.13 (0.92, 1.38)	106	1.14 (0.93, 1.40)
yes	159	126	0.79 (0.68, 0.93)	132	0.83 (0.71, 0.97)
# 1^{st} degr relat w BC					
0	208	200	0.96 (0.84, 1.10)	193	0.93 (0.81, 1.06)
1	43	30	0.69 (0.51, 0.93)	41	0.97 (0.72, 1.30)
2	1	1	1.18 (0.17, 8.39)	3	3.49 (0.49, 24.78)

Comparing Two Absolute Risk Prediction Models

- Compare calibration
- Compare accuracy measures
- Compare AUC values
- Compare PCF and PNF or iPCF and iPNF
- Criteria based on reclassification tables
- Net reclassification improvement
  - Integrated discrimination improvement
  - Compare expected losses

Classification of women in NHS cohort ages 50-55 at baseline, who got breast cancer (events) and those who did not (non-events) for BCmod and BCRAT with 5-year risk threshold $r^* = 1.66\%$

	$\leq$ **1.66%**	$>$ **1.66%**	**Total**
	5-yr risk from BC2013		
n	13,624	3,461	17,085
Events	168	84	252
Non-events	13,456	3,377	16,833
Percentage with events(%)	1.23	2.43	1.47
	5-yr risk from BCRAT		
n	13,449	3,636	17,085
Events	164	88	252
Non-events	13,285	3,548	16,833
Percentage with events(%)	1.22	2.42	1.47

Accuracy measures for estimates from BCmod and BCRAT models using 5-year absolute risk threshold $r^* = 0.0166$ for $50 - 55$ year old women

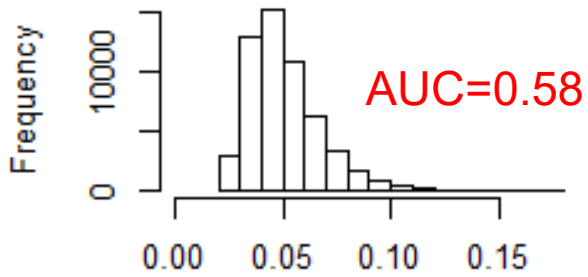
Measure	BCmod	BCRAT	
	Estimate (95% CI)	**Estimate (95% CI)**	**Difference (95% CI)**
Sens	0.333 (0.28, 0.40)	0.349 (0.29, 0.41)	0.0159 (-0.041, 0.073)
Spec	0.799 (0.79, 0.81)	0.789 (0.78, 0.80)	-0.0102 (-0.017, -0.004))
PPV	0.024 (0.019, 0.030)	0.024 (0.020, 0.030)	-0.0001 (-0.0008, 0.0007)
NPV	0.988 (0.986, 0.990)	0.988 (0.986, 0.990)	0.0001 (-0.0009, 0.0012)
PCC	0.793 (0.786, 0.799)	0.783 (0.777, 0.789)	-0.0098 (-0.0159, -0.0037)

Abbreviations: PPV= Positive predictive value; NPV= Negative predictive value; PCC= Probability of correct classification; Sens= sensitivity; Spec=specificity.

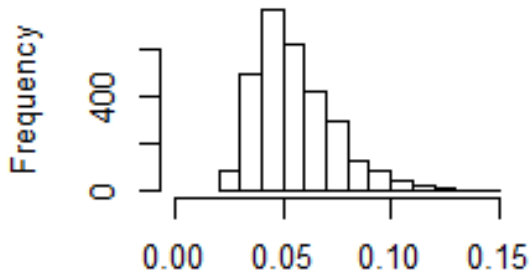
Comparing Two Absolute Risk Prediction Models

- Compare calibration
- Compare accuracy measures
- Compare AUC values
- Compare PCF and PNF or iPCF and iPNF
- Criteria based on reclassification tables
- Net reclassification improvement
  - Integrated discrimination improvement
  - Compare expected losses

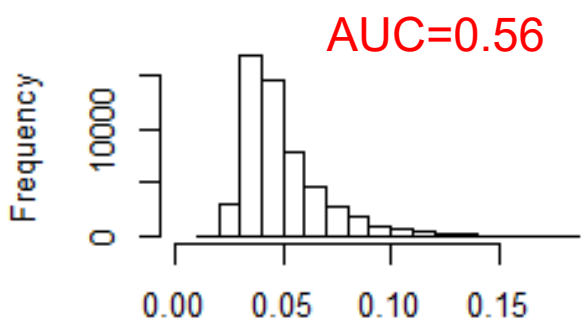
Distribution of Risk Estimates in NHS Validation Cohort by Breast Cancer Status



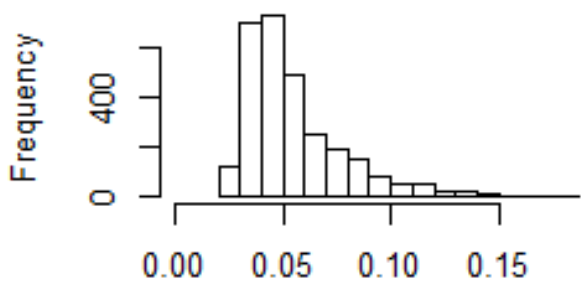
BCmod, no breast cancer



BCmod, breast cancer



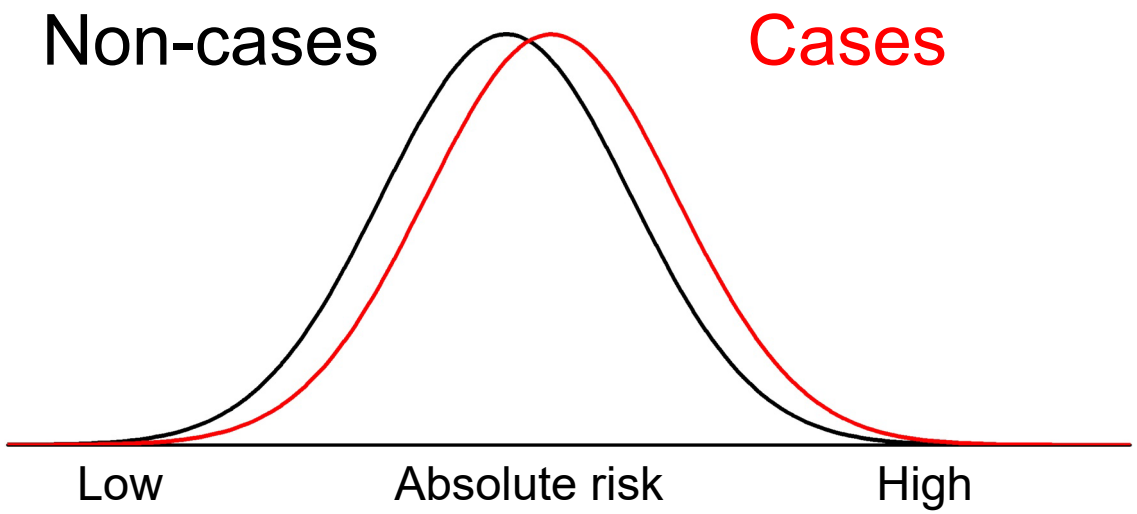
BCRAT, no breast cancer



BCRAT, breast cancer

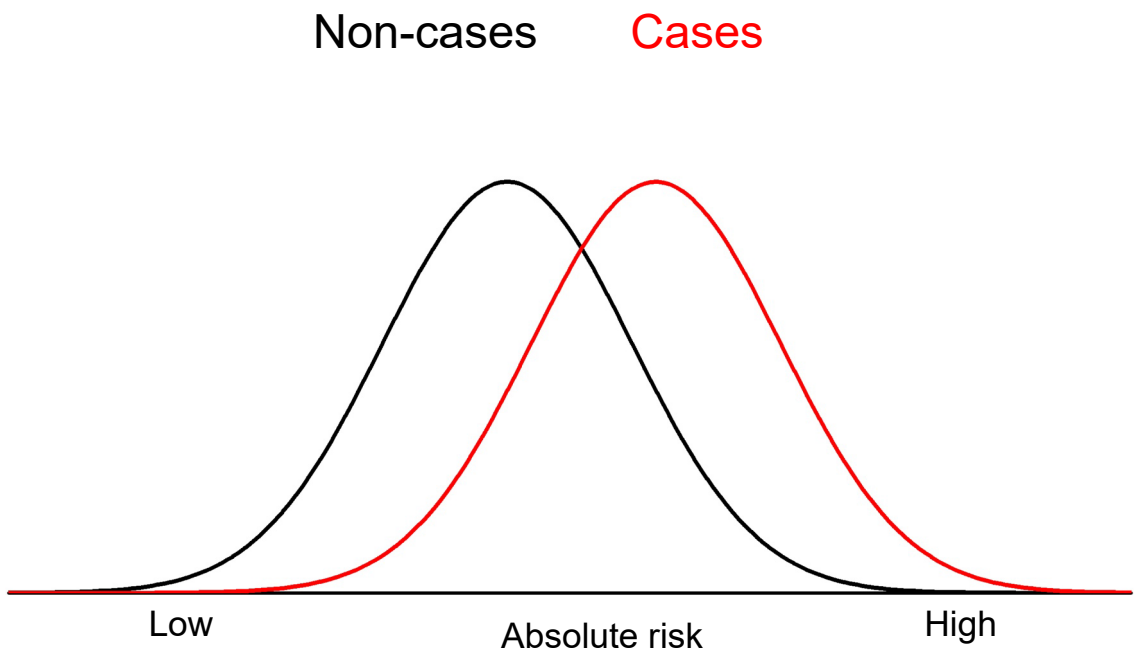
[1]

Sidetrack: Example of Distributions of Log-normal Risk Estimates with modest AUC



AUC=0.6, like BCmod

Sidetrack: Example of Distributions of Log-normal Risk Estimates with good AUC



Good discriminatory accuracy, AUC=0.8

Compare AUC values

Compute AUC using disease placement value
$P(R \geq r|Y = k) = S_k(r), k = 0, 1$ (DeLong, 1988; Pepe, 2003)
n_k - number of individuals with $Y = k, k = 0, 1$
r_{ij}^k risk from model i in person with $Y_j = k$

$$\widehat{AUC}_i = \frac{1}{n_0} \sum_{j=1}^{n_0} S_1(r_{ij}^0), i = 1, 2$$

$$\Delta = \widehat{AUC}_1 - \widehat{AUC}_2$$

If $\widehat{AUC}_i$ computed using paired samples (i.e. based on same individuals in the cohort) then

$$var(\Delta) \approx \frac{var(S_1^1(r_1^0) - S_1^2(r_2^0))}{n_0} + \frac{var(S_0^1(r_1^1) - S_0^2(r_2^1))}{n_1}$$

Don't despair, bootstrap also works!

Compare AUCs in NHS Women Ages 50-55

BCRAT: $AUC_1 = 0.623(0.0173)$

BCmod: $AUC_2 = 0.617(0.0167)$

$\Delta = \widehat{AUC}_1 - \widehat{AUC}_2 = 0.006(95\%CI; -0.0355, 0.0282)$

Reclassification Tables Cook et al, 2006

Risk estimates from 2 models are divided into risk categories and cross tabulated

Typical application: model 2 is a refinement of model 1, e.g. additional risk factor is added

Reclassification Tables Cook et al, 2006

Risk stratification based on 5 year absolute risks from BCRAT and BCmod models for women in NHS validation cohort ages 50-55 years at baseline

5-yr risk from BCRAT		**5-year risk from BCmod**				
		0 to $<1\%$	1 to $\leq 1.66\%$	>1.66 to $<2.5\%$	$\geq 2.5\%$	Total
0 to $<1\%$	n	2,291	1,675	114	2	4,082
	Events[a]	13	16	1	0	30
	Non-events	2,278	1,659	113	2	4,052
	Prop. Events (%)	0.57	0.96	0.88	0.00	0.73
1% to $\leq 1.66\%$	n	1387	6632	1299	49	9367
	Events	10	100	24	0	134
	Non-events	1,377	6,532	1,275	49	9,233
	Prop. Events (%)	0.72	1.51	1.85	0.00	1.43
$>1.66\%$ to $<2.5\%$	n	9	1,522	1,120	78	2,729
	Events	0	27	38	0	65
	Non-events	9	1,495	1,082	78	2,664
	Prop. Events (%)	0.00	1.77	3.39	0.00	2.38
$\geq 2.5\%$	n	1	107	598	201	907
	Events	0	2	13	8	23
	Non-events	1	105	585	193	884
	Prop. Events (%)	0.00	1.87	2.17	3.98	2.54
Total	n	3,688	9,936	3,131	330	17,085
	Events	23	145	76	8	252
	Non-events	3,665	9,791	3,055	322	16,833
	Prop. Events (%)	0.62	1.46	2.43	2.42	1.47

Net Reclassification Improvement, NRI Pencina et al, 2008

- "up": model 2 moves person to higher risk category than model 1
- "down": model 2 moves person to lower risk category than model 1

$$NRI = \{P(up|Y=1) - P(down|Y=1)\} + \{P(down|Y=0) - P(up|Y=0)\}$$

= improvement for cases + improvement for controls

Comment: NRI is function of ranks of predicted probabilities

Net Reclassification Improvement, NRI Pencina et al, 2008

$$P(up|Y=i) = \frac{\text{number in group } Y=i \text{ moving up}}{\text{number in group } Y=i}$$

Breast cancer model example:

$$P(up|Y=1) - P(down|Y=1) = \frac{16+1+0+24+0}{252} - \frac{10+0+27+0+0+2+13}{252} = -0.0437$$

$$P(down|Y=0) - P(up|Y=0) = \frac{1377+9+1495+1+105+585}{16833} - \frac{1659+113+2+1275+49+78}{16833} = 0.0235$$

$$NRI = -0.0437 + 0.0235 = -0.0201$$

Issue with NRI: Sensitivity to Mis-calibration

Assume we compare two models:

- **Well calibrated model:** Risks uniform on [0,0.1] in $Y = 0$ and uniform on [0,0.2] in $Y = 1$ (cases)
- **New poorly calibrated model** adds 0.05 to all risks: $r\ U[0.05, 0.15]$ in $Y = 0$ and U[0.05,0.25] in $Y = 1$
- For risk threshold 0.2, old model has sens=0 and spec=1.0; new model has sens=.25 and spec =1.0.
- AUC=0.75 for both models
- NRI=(.25-0)+(1.0-1.0)=0.25 **only because new model is biased**

Continuous Net Reclassification Improvement (cNRI) Pencina et al, 2011

$$NRI = \{P(up|Y=1) - P(down|Y=1)\} + \{P(down|Y=0) - P(up|Y=0)$$
$$= 2P(R_2 > R_1|Y=1) - 1 - 2P(R_2 > R_1|Y=0) - 1$$

Continuous NRI

$$cNRI = P(R_2 > R_1|Y=1) - P(R_2 > R_1|Y=0)$$

$$AUC_1 - AUC_2 = P(R_2^{Y=1} > R_2^{Y=0}) - P(R_1^{Y=1} > R_1^{Y=0})$$

Applications of Absolute Risk Models

Applications of Absolute Risk Models

- Counseling patients
  - General perspective on disease risk
  - Weigh risks and benefits of interventions such as tamoxifen
  - Modify known risk factors
  - Prognosis for dying of a disease after disease onset

Should a Woman in her Forties Have Screening Mammography?

- US Prev. Services Task Force (2016): "The decision to start screening mammography in women prior to age 50 years should be an individual one. Women who place a higher value on the potential benefit than the potential harms may choose to begin biennial screening between the ages of 40 and 49 years."
- No mention of risk

Should a Woman in her Forties Have Screening Mammography? Wu, Graubard, Gail, AIM 2012

- A 40-year old woman is uncertain whether to have screening mammograms. Her mother and sister had breast cancer. Her 5-year absolute risk (1.8%) exceeds that of a 50-year old woman without risk factors (0.6%).
- 11.4 million white women in their forties in US (74%) have risks above that of 50-year old women without risk factors.

Applications of Absolute Risk Models

- Counseling patients
  - General perspective on disease risk
  - Weigh risks and benefits of interventions such as tamoxifen
  - Modify known risk factors
  - Prognosis for dying of a disease after disease onset

Example of Use of Risk Model: Weighing the Risks and Benefits of Tamoxifen for Breast Cancer Chemoprevention

Gail, Costantino, Bryant, Croyle, Freedman, Helzlsouer, Vogel, JNCI 1999; 91:1829-46.
Based on Breast Cancer Prevention Trial (P1 Trial) by Fisher et al., JNCI, 1998

10,000 40-Year-old White Women with Uteri. 5-Year Absolute Risk of Invasive Breast Cancer 2%*

LIFE-THREATENING	NO TAMOXIFEN	PREVENTED BY TAMOXIFEN
Invasive Breast Cancer	200	97
Hip Fracture	2	1
Endometrial Cancer	10	-16
Stroke	22	-13
Pulmonary Embolus	7	-15
		net prevented 54
SEVERE EVENTS		
In situ Breast Cancer	106	53
Deep Vein Thrombosis	24	-15
		net prevented 38

*Average 5-year BC risk of 40 year old White woman in US population is 0.6%

Applications of Absolute Risk Models

- Counseling patients
  - General perspective on disease risk
  - Weigh risks and benefits of interventions such as tamoxifen
  - Modify known risk factors
  - Prognosis for dying of a disease after disease onset
- Public health/prevention
  - Designing prevention trials
  - Assessing reduction in population absolute risk from decreased exposure to modifiable risk factors
  - "High risk" prevention strategy
  - Assessing risk based screening strategies

Designing prevention trials

- Statistical power
  - Depends on the number of events
  - Number of events is proportional to average absolute risk of trial participants
- Eligibility criteria
  - Select subjects who stand to benefit from intervention
  - Increase efficiency of trial by including high risk subjects
- Examples: BCPT (P-1) Trial, STAR Trial

Applications of Absolute Risk Models

- Counseling patients
  - General perspective on disease risk
  - Weigh risks and benefits of interventions such as tamoxifen
  - Modify known risk factors
  - Prognosis for dying of a disease after disease onset
- Public health/prevention
  - Designing prevention trials
  - Assessing reduction in population absolute risk from decreased exposure to modifiable risk factors
  - "High risk" prevention strategy
  - Assessing risk based screening strategies

Models with Modifiable Risk Factors

- Breast cancer
  - Petracci et al. JNCI 2011(exercise, BMI, alcohol)
  - Pfeiffer et al Plos Med 2013(BMI, HRT, alcohol)
  - Maas et al JAMA Oncol 2016 (BMI, HRT,alcohol, smoking)
- Heart disease (Framingham models) (blood pressure, smoking, HDL, LDL, diabetes)

Assess Disease Preventable Due to Modifiable Risk Factors

Non-modifiable factors ($\mathbf{X}$): e.g. parity, age 1st birth, benign breast disease/biopsy,
Modifiable factors ($\mathbf{Z}$): BMI, alcohol, HRT use
Absolute risk reduction for individual woman:

$$R(a, a+\tau, \mathbf{X}, \mathbf{Z}) - R(a, a+\tau, \mathbf{X}, \mathbf{Z}_0)$$

Mean absolute risk reduction in population

$$E\{R(a, a+\tau, \mathbf{X}, \mathbf{Z}) - R(a, a+\tau, \mathbf{X}, \mathbf{Z}_0)\}$$

Fractional reduction in mean absolute risk

$$\frac{E\{R(a, a+\tau, \mathbf{X}, \mathbf{Z}) - R(a, a+\tau, \mathbf{X}, \mathbf{Z}_0)\}}{E\{R(a, a+\tau, \mathbf{X}, \mathbf{Z})\}}$$

Modifying Alcohol Consumption, HRT use, BMI: Changes in 20-year BCmod Risk for Women Aged 50 Years in NHS Validation Cohort

Subset of population	Initial abs risk (%)	Absolute mean risk reduct (%)	Fractional reduct in mean risk (%)
Entire population (N=2447)	5.9	1.0	17.0
Women with positive family history (N=295)	8.0	1.5	18.8
Women in top 10% of population risk (N=246)	9.2	2.5	27.2

Comments on Modifiable Risks

- Key assumptions for predicting preventive effects
  - There are interventions to effect these changes
  - Interventions have their predicted effects
  - People will comply with interventions
- More disease prevented by treating the entire population, rather than high risk subset
- Changes in absolute risk offer different perspective than attributable risk

Applications of Absolute Risk Models

- Counseling patients
  - General perspective on disease risk
  - Weigh risks and benefits of interventions such as tamoxifen
  - Modify known risk factors
  - Prognosis for dying of a disease after disease onset
- Public health/prevention
  - Designing prevention trials
  - Assessing reduction in population absolute risk from decreased exposure to modifiable risk factors
  - High risk prevention strategy
  - Assessing risk based screening strategies

Example of using well calibrated risk model: Estimating the breast cancer (BC) burden in Germany and implications for risk-based screening

Collaboration with German Cancer Research Center

- Used data from 22 098 women aged 40+ years enrolled in EPIC Germany cohort (European Prospective Investigation into Cancer & Nutrition)
- 745 breast cancers occurred during median follow-up 12 years
- Predicted breast cancer risk using BCmod

Example of using well calibrated risk model: Estimating the breast cancer (BC) burden in Germany and implications for risk-based screening

Collaboration with German Cancer Research Center

- Used data from 22 098 women aged 40+ years enrolled in EPIC Germany cohort (European Prospective Investigation into Cancer & Nutrition)
- 745 breast cancers occurred during median follow-up 12 years
- Predicted breast cancer risk using BCmod
- **Calibration:** $\mathcal{O}/\mathcal{E}_{BCmod} = 0.93$ (95% CI: $0.83 - 1.05$)

Example of using well calibrated risk model: Estimating the breast cancer (BC) burden in Germany and implications for risk-based screening

Collaboration with German Cancer Research Center

- Used data from 22 098 women aged 40+ years enrolled in EPIC Germany cohort (European Prospective Investigation into Cancer & Nutrition)
- 745 breast cancers occurred during median follow-up 12 years
- Predicted breast cancer risk using BCmod
- **Calibration:** $\mathcal{O}/\mathcal{E}_{BCmod} = 0.93 \quad (95\%\ \text{CI}: 0.83 - 1.05)$
- **Discrimination:**
  $AUC_{BCmod} = P(R_{case} > R_{noncase}) = 0.61 \quad (95\%\text{CI}: 0.58 - 0.63)$

(Hüsing, . . . , Pfeiffer, Cancer Causes & Control, 2020)

Example of using well calibrated risk model: Estimating the breast cancer (BC) burden in Germany and implications for risk-based screening (Quante, . . . , Pfeiffer, 2021)

Collaboration with Women's Hospital of Munich, Germany
Currently: Mammographic screening starts at age 50 in Germany
Estimated $\mathcal{E}_{\tau=1}$ using 1-year BCmod risks for women in DEGS survey and number needed to screen (NNS) to detect one BC case

	Age 40-44	Age 45-49	Age 50-69
	total population		
Total count	3,261,000	3,461,000	10,498,000
$\mathcal{E}_{\tau=1}$	3,400	5,500	35,900
NNS	953	631	292

Estimating the breast cancer burden in Germany and implications for risk-based screening (Quante, . . . , Pfeiffer, 2021)

Int. guidelines: start screening when a woman's 5-year BC risk $> 1.7\%$

	Age 40-44	Age 45-49	Age 50-69
	total population		
Total count	3,261,000	3,461,000	10,498,000
$\mathcal{E}_{\tau=1}$	3,400	5,500	35,900
NNS	953	631	292
	5-year BC risk $> 1.7\%$		
Total count	0	39,000	4,761,000
$\mathcal{E}_{\tau=1}$	0	140	22,000
NNS		282	217

Miscellaneous Topics

Updating Prediction Models with New Risk Factors

Examples of new molecular risk factors for breast cancer:

- Polygenic risk score based on SNPs
- Mammographic density
- Involution measures
- Inflammation markers

Improve Models: Updating Risk Models with New Predictors

- Many risk prediction models available that use well-established risk factors
- Combine new information with information from existing models to improve predictions
- Particularly relevant when new molecular markers measured in

Approaches for incorporating new information into existing relative risk models

Y - disease outcome, binary
$\mathbf{X} = (X_1, \ldots, X_p)$ - vector of original old predictors
Z - new marker (could also be vector)
AIM: use information from "old model" model with predictors $\mathbf{X}$

$$R_{\mathbf{X}} = P(Y = 1|\mathbf{X}) = \frac{\exp(\gamma_0 + \boldsymbol{\gamma}'\mathbf{X})}{1 \exp(\gamma_0 + \boldsymbol{\gamma}'\mathbf{X})}$$

New model

$$R_{\mathbf{X},Z} = P(Y = 1|\mathbf{X}, Z) = \frac{\exp(\beta_0 + \boldsymbol{\beta}'\mathbf{X} + \beta_z Z)}{1 + \exp(\beta_0 + \boldsymbol{\beta}'\mathbf{X} + \beta_z Z)}$$

Approaches for incorporating new information into existing relative risk models

Y - disease outcome, binary
$\mathbf{X} = (X_1, \ldots, X_p)$ - vector of original old predictors
Z - new marker (could also be vector)

$$rr(\mathbf{X}, Z) = rr(\boldsymbol{\beta}'\mathbf{X}, \beta_Z Z) = \exp(\boldsymbol{\beta}'\mathbf{X} + \beta_Z Z)$$

Updating Methods, General Idea

Bayes Theorem:

$$P(Y|\mathbf{X}, Z) = \frac{P(Y|\mathbf{X})P(Z|\mathbf{X}, Y)}{\sum_Y P(Y|\mathbf{X})P(Z|\mathbf{X}, Y)}$$

Thus

$$\log \frac{P(Y=1|\mathbf{X}, Z)}{P(Y=0|\mathbf{X}, Z)} = \log \frac{P(Y=1|\mathbf{X})}{P(Y=0|\mathbf{X})} + \log \frac{P(Z|Y=1, \mathbf{X})}{P(Z|Y=0, \mathbf{X})}$$

posterior odds, new model $R(\mathbf{X}, Z)$	=	prior odds, old model $R(\mathbf{X})$	+	likelihood ratio (LR)

Updating Methods: Estimate LR from Case-Control Data with Information on $Z, \mathbf{X}$

$$LR(Z|Y, \mathbf{X}) = \frac{P(Z|Y = 1, \mathbf{X})}{P(Z|Y = 0, \mathbf{X})}$$

- Independence Bayes: Assume $Z, \mathbf{X}$ independent, model $P(Z|Y)$
- Independence Bayes with shrinkage: $\theta \log LR(Z|Y)$
- Separate estimation of LR: model $P(Z|\mathbf{X})$ separately in cases and controls
- Joint estimation of LR: model distribution of marker Z as function of case-control status Y and old predictors $\mathbf{X}$ model $P(Z|Y, \mathbf{X})$

Updating Methods: Fit $R(\mathbf{X}, Z)$ to new data only

$$R(\mathbf{X}, Z) = \hat{P}(Y = 1|\mathbf{X}, Z) = \frac{\exp(\beta_0 + \boldsymbol{\beta}'\mathbf{X} + \beta_z Z)}{1 + \exp(\beta_0 + \boldsymbol{\beta}'\mathbf{X} + \beta_z Z)}$$

Completely ignore information from old model $R(\mathbf{X})$ and fit logistic regression model to new data only

Updating Methods: Offset

$$R(\mathbf{X}, Z) = \hat{P}(Y = 1|\mathbf{X}, Z) = \frac{\exp(\gamma'\mathbf{X} + \delta_0 + \delta_1 Z)}{1 + \exp(\gamma'\mathbf{X} + \delta_0 + \delta_1 Z)}$$

Include prior odds as offset term in a logistic regression model that is fit to new data set

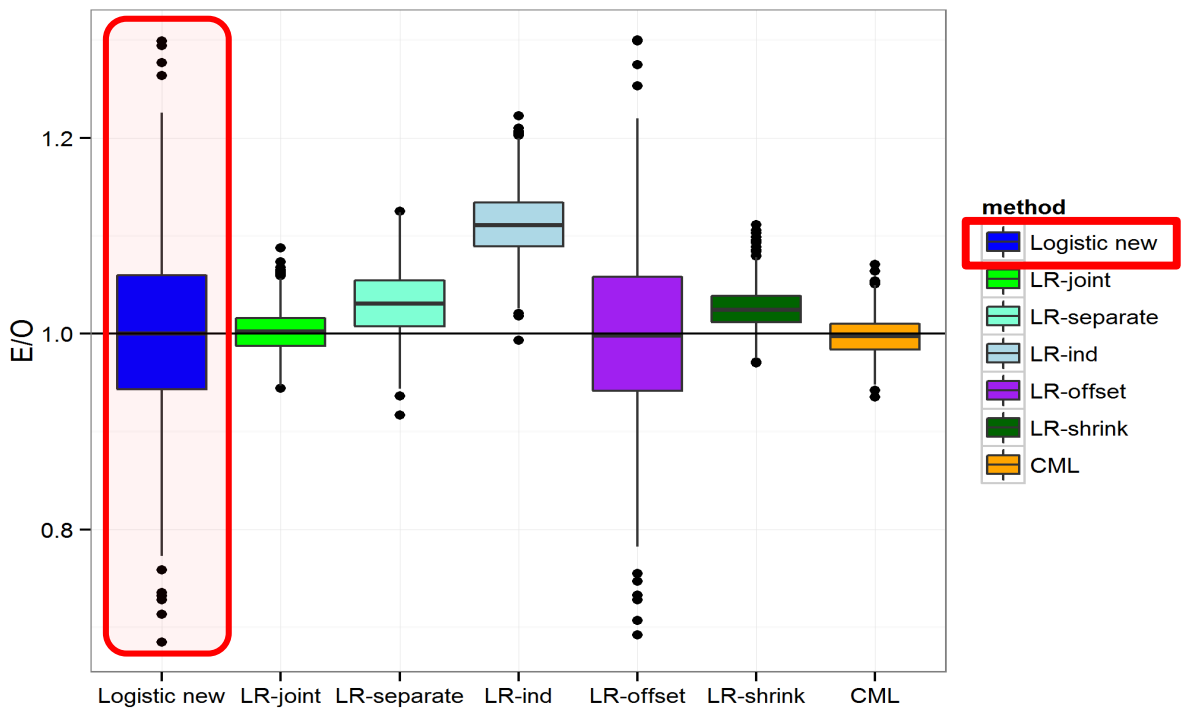
Updating Methods: Constraint Maximum Likelihood

Chatterjee et al. (JASA 2016) identified set of general constraints that link full model $P(Y|\mathbf{X}, Z)$ and reduced model $P(Y|\mathbf{X})$ and used them to propose semiparametric maximum likelihood estimate for updated model.

Most efficient estimate for updated model parameters

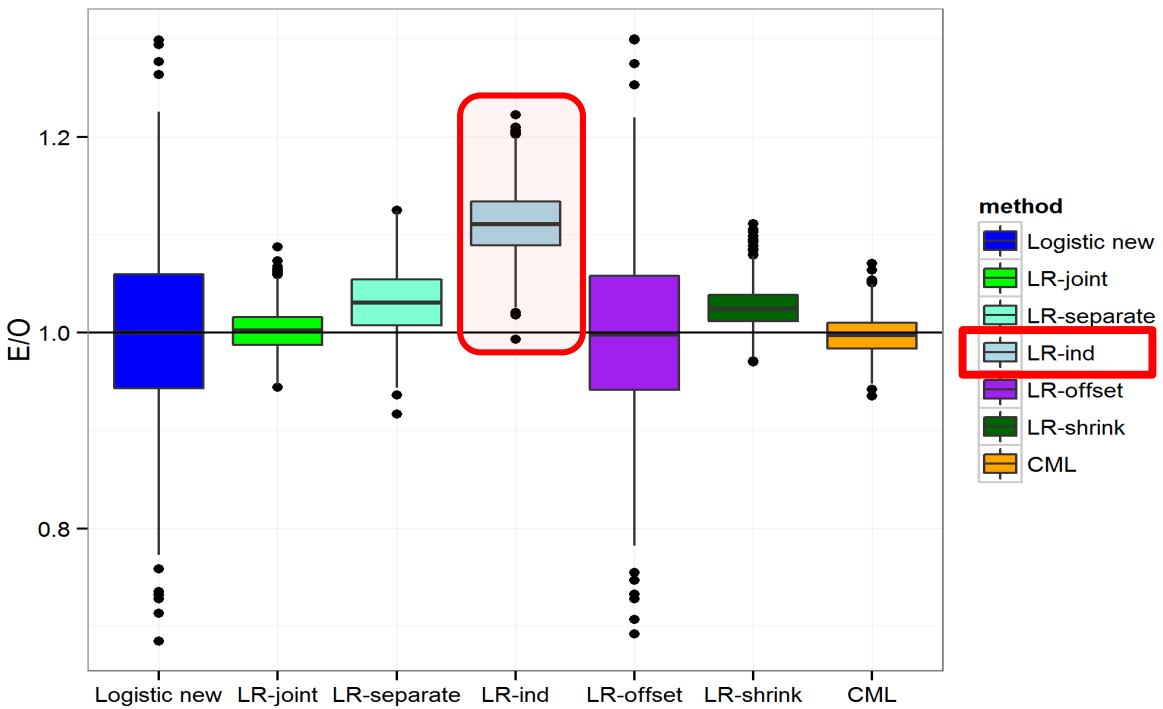
Example: Compared calibration (E/O) of methods for updating old model to predict HCV treatment response with 2 new markers (IFNL4 genotype, HCV-RNA)

Cohort setting; Two new markers: Z_1 (binary) and Z_2 continuous



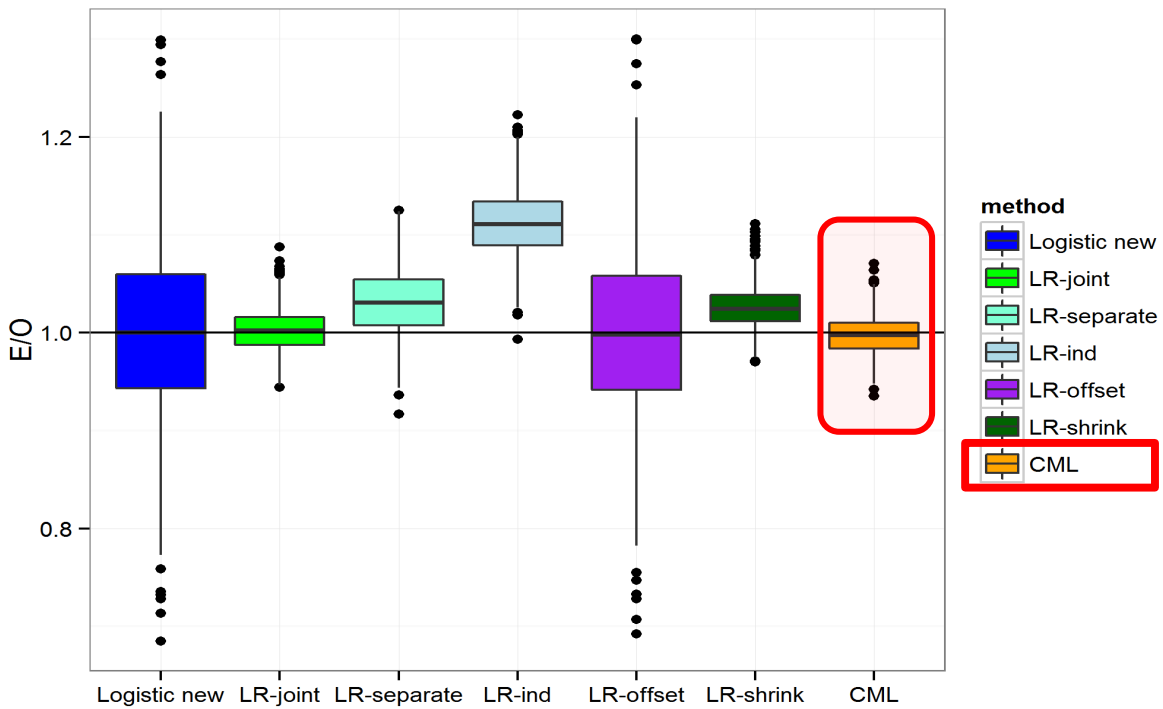
Example: Compared calibration (E/O) of methods for updating old model to predict HCV treatment response with 2 new markers (IFNL4 genotype, HCV-RNA)

Cohort setting; Two new markers: Z_1 (binary) and Z_2 continuous



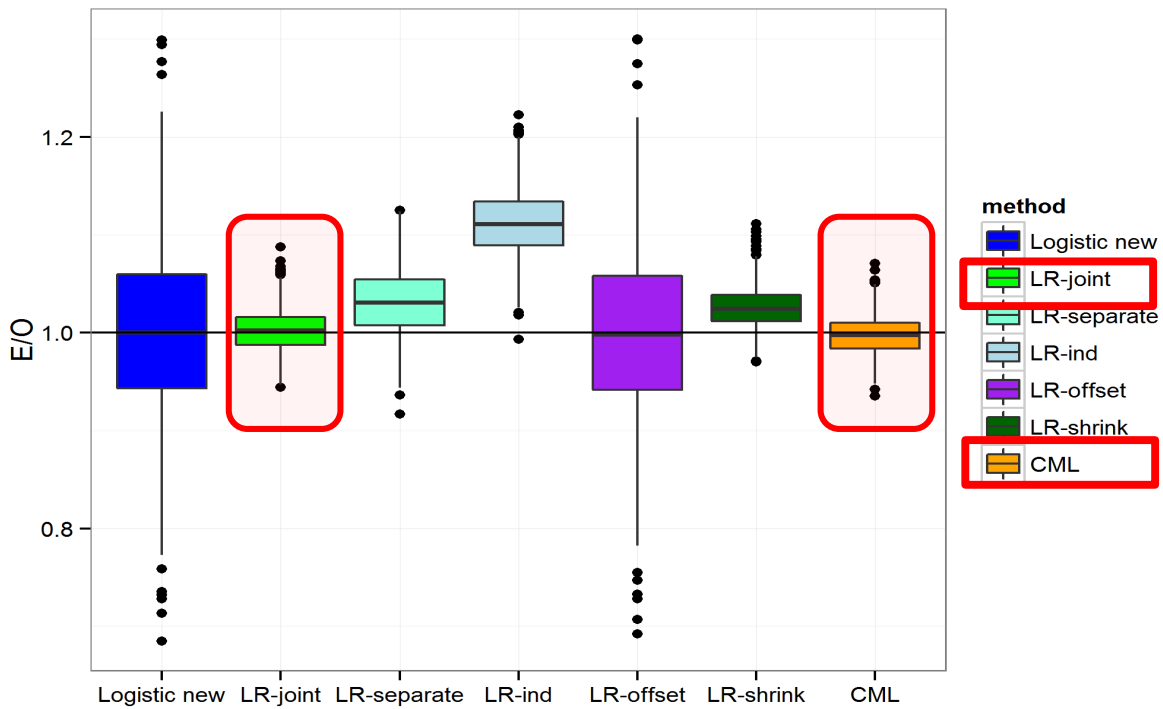
Example: Compared calibration (E/O) of methods for updating old model to predict HCV treatment response with 2 new markers (IFNL4 genotype, HCV-RNA)

Cohort setting; Two new markers: Z_1 (binary) and Z_2 continuous



Example: Compared calibration (E/O) of methods for updating old model to predict HCV treatment response with 2 new markers (IFNL4 genotype, HCV-RNA)

Cohort setting; Two new markers: Z_1 (binary) and Z_2 continuous



Summary of Model Updating Results

Combine new predictor Z with information on "established predictors $\mathbf{X}$

- Fitting model to new data only (ignoring available information) yields unbiased predictions with large variability
- Likelihood ratio (LR) updating assuming independence of Z and $\mathbf{X}$ typically biases updated model predictions
- Logistic new and LR offset show large variability in the predictions
- CML and LR methods allowing for dependence between Z and $\mathbf{X}$ yield unbiased predictions with similar variance
- Recommendation: LR updating by modeling $P(Z|\mathbf{Y}, \mathbf{X})$ (easy to implement in standard software)

Workshop Goals/Summary

- Learn what "absolute risk" ("crude risk' or "cumulative incidence') is
- Learn how to estimate it from data from various designs
- Learn how to assess the validity and usefulness of a model of absolute risk
- Learn what it can be used for

Selected references not in book by Pfeiffer & Gail, 2018

- Shin YE, Gail MH, Pfeiffer RM. Assessing risk model calibration with missing covariates.Biostatistics. 2022 18;23(3)
- Quante AS, . . ., Pfeiffer RM. Estimating the breast cancer burden in Germany and implications for risk-based screening. Cancer Prev Res. 2021;14(6):627-634.
- Hüsing A, . . ., Pfeiffer RM. Validation of two US breast cancer risk prediction models in German women. Cancer Causes Control. 2020;31(6):525-536.
- Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *JASA* **87,** 376–382.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *JASA* **47,** 663–685.
- Samuelsen, S. O. (1997). A pseudolikelihood approach to analysis of nested case-control studies. *Biometrika* **84,** 379–394.
- Pfeiffer RM, Chen Y, Gail MH, Ankerst DP. Accommodating population differences when validating risk prediction models. Stat Med. 2022 30;41(24):4756-4780.

Selected references

- Andersen, P. K. and Keiding (2012), Interpretability and importance of functionals in competing risks and multistate models, *Statistics in Medicine,* 31, 1074–1088
- Benichou, J. and Gail, M. H. (1995), Methods of inference for estimates of absolute risk derived from population-based case-control studies, *Biometrics*, 51, 182–194
- Gail, M. (1975), Review and critique of some models used in competing risk analysis, *Biometrics*, 31, 209–222
- Langholz, B. and Thomas, D. (1990). Nested case-control and case-cohort methods of sampling from a cohort: A critical comparison. *AJE* **131,** 169–176.
- Prentice, R. L. and Kalbfleisch, J. D. and Peterson, A. V. and Flournoy, N. and Farewell, V. T. and Breslow, N. E. (1978) Analysis of failure times in presence of competing risks, *Biometrics*, 34, 541–554
- Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, 73, 1–11.
- Tsiatis, A. (1975) Nonidentifiability aspect of problem of competing risks, *Proceedings of the National Academy of Sciences of the United States of America*, 72, 20–22