Criteria for Assessment of Performance of Cancer Risk Prediction Models: Overview

Ruth Pfeiffer

Cancer Risk Prediction Workshop, May 21, 2004

Division of Cancer Epidemiology and Genetics
National Cancer Institute

Cancer Risk Prediction Models

- Model input:
  - Individual's age and risk factors
  - Age interval at risk
- Model output:
  - Estimate of individual's absolute risk of developing cancer over a given time period (e.g. the next 5 years).

Definition of Absolute Risk for Cancer in [a, a+τ]

$$r(x;a,\tau) = P(T \le a+\tau, cause = c_1 \mid T > a; x)$$

$$= \int_a^{a+\tau} h_1(t,x)\exp\left[-\int_a^t \{h_1(u,x)+h_2(u)\}du\right]dt$$

$h_1(t,\ x)$ - hazard of cancer incidence at age t

$h_2(t)$ - competing mortality hazard

x - individual risk or protective factors

Applications of absolute risk prediction models

- Population level:
  - Estimate population disease burden
  - Estimate impact of changing the risk factor distribution in the general population
  - Plan intervention studies
- Individual level:
  - Clinical decision-making:
    - Modification of known risk factors (diet, exercise)
    - Weighing risks and benefits of intervention (eg chemoprevention)
  - Screening recommendations

Evaluating the performance of risk models

- How well does model predict for groups of individuals: **Calibration**
- How well does model categorize individuals: **Accuracy scores**
- How well does model distinguish between individuals who will and will not experience event: **Discriminatory Accuracy**

Independent population for validation

Assume population of N individuals followed over time period ☺

Define

$$Y_i = \begin{cases} 1, & \text{if ith subject develops cancer in } \tau \\ 0, & \text{otherwise} \end{cases}$$

$r_i = r(x_i; a) =$ absolute risk for ith subject with baseline covariates x_i, age a

Assessing Model Calibration

Goodness-of-fit criteria based on comparing observed (O) with expected (E) number of events overall and in subgroups of risk factors of the population

$$O = \sum_{i=1}^{N} Y_i, \quad E = \sum_{i=1}^{N} r_i$$

Use Poisson approximation to sum of independent binomial random variables with $r_i << 1$

Assessing Model Calibration, cont.

Unbiased (well calibrated)

$$\frac{1}{N}\sum_{i=1}^{N} Y_i \approx \int_0^1 r dF(r) = \mu$$

Remark: $\int_0^1 r dF(r) = \int_X r(x) dG(x)$

Brier Score

$$BS = \frac{1}{N}\sum_{i}(O_i - r_i)^2$$

Brier Score = Mean Squared Error (measure of accuracy)

Brier, 1950

Comparison of observed (O) and expected (E) cases of invasive breast cancer (Gail et al Model 2) in placebo arm of Breast Cancer Prevention Trial (Table 4, Costantino et al, JNCI, 1999)

Age Group	# women	O	E	E/O
<=49	2332	60	55.9	0.9
50-59	1807	43	48.4	1.1
>=60	1830	52	54.7	1.1
All ages	5969	155	159.0	1.0

Assess model performance for clinical decision making

For clinical decision making a decision rule is needed

$$\delta_i = \begin{cases} 1, & \text{if } r_i > r^* \\ 0, & \text{otherwise} \end{cases}$$

for some threshold r*

For given threshold r* define sensitivity and specificity of decision rule as

$$\text{sens}(r^*) = \frac{1}{\mu} \int_{r^*}^{1} r\,dF(r),$$

proportion of cases with $r > r^*$

$$\text{spec}(r^*) = \frac{1}{1-\mu} \int_{0}^{r^*} (1-r)\,dF(r)$$

$$\text{with } \mu = \int_{0}^{1} r\,dF(r)$$

Problem: sensitivity and specificity not always appropriate measures

Example: rare disease π=P(Y=1)=0.01

Sensitivity =0.95, specificity=0.95

$$ppv = P(Y = 1 \mid \delta = 1) = \frac{P(\delta = 1 \mid Y = 1)\pi}{P(\delta = 1 \mid Y = 1)\pi + P(\delta = 1 \mid Y = 0)(1-\pi)} = \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + (1-0.95) \cdot 0.99} = 0.16$$

Accuracy Scores

Measure how well true disease outcome predicted

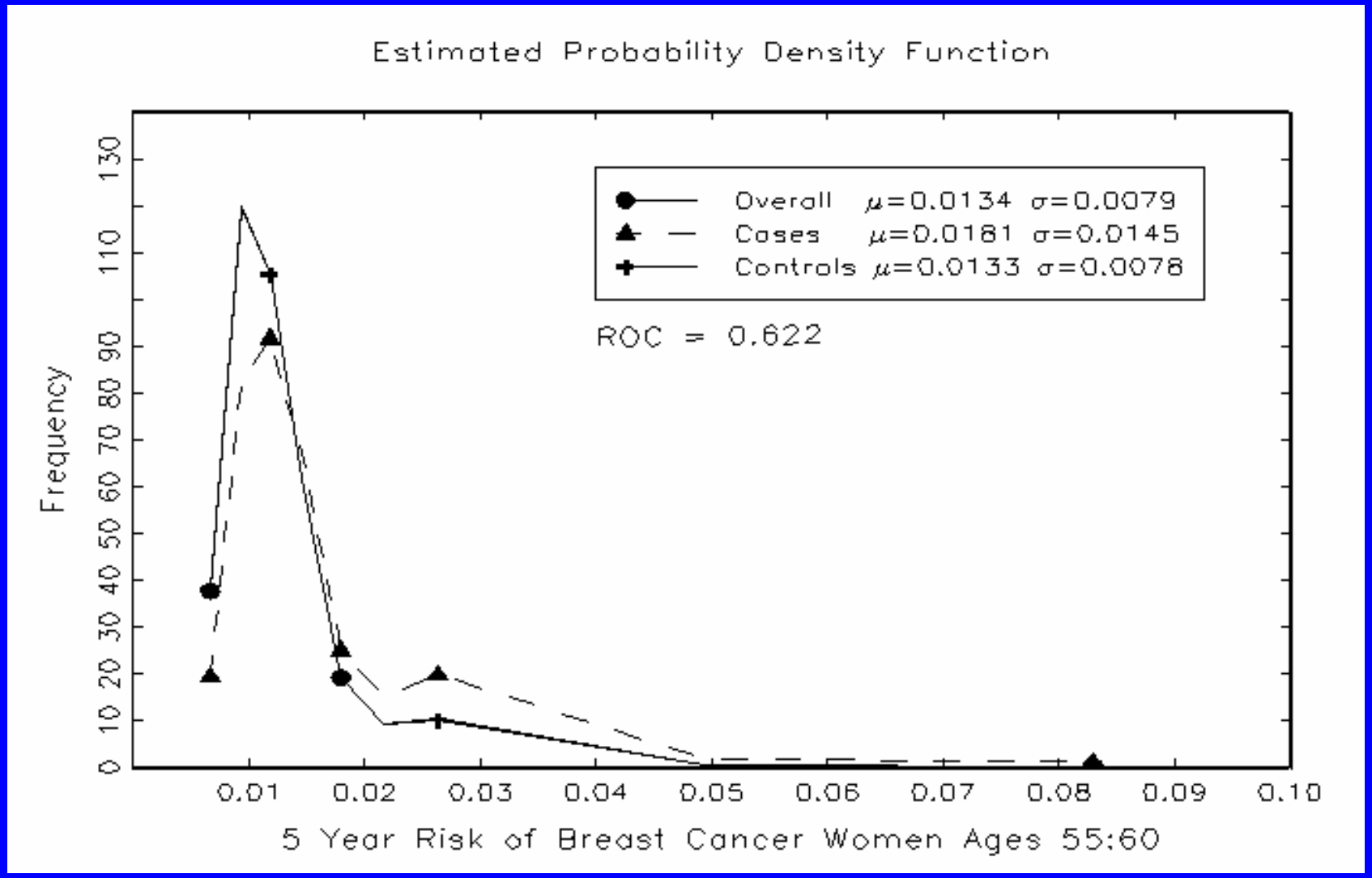
Quantify clinical value of decision rule (Zweig & Campbell, 1993)

- Positive predictive value $ppv = P(Y = 1 | \delta = 1)$
- Negative predictive value $npv = P(Y = 0 | \delta = 0)$
- Weighted combinations of both

Depend on sensitivity, specificity, disease prevalence

Measures of Discrimination for Range of Thresholds

- ROC curve (plots sensitivity against 1-specificity)
- Area under the ROC curve (AUC) ~Mann-Whitney-Wilcoxon Rank Sum Test ~ Gini index for rare events
- Concordance statistic (Rockhill et al, 2001; Bach et al, 2003)
- Partial area under the curve (Pepe, 2003; Dodd&Pepe, 2003)


Estimated Probability Density Function
Overall μ=0.0134 σ=0.0079
Cases μ=0.0181 σ=0.0145
Controls μ=0.0133 σ=0.0078
ROC = 0.622
Frequency
0 10 20 30 40 50 60 70 80 90 110 130
0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10
5 Year Risk of Breast Cancer Women Ages 55:60

Decision Theoretic Framework

Specify loss function for each combination of true disease status and decision:

	$Y=0$	$Y=1$
$\delta=0$	C_{00}	C_{01}
$\delta=1$	C_{10}	C_{11}

Known Loss Function

$$EL = C_{11}\int_{r^*}^{1} r\,dF(r) + C_{01}\int_{0}^{r^*} r\,dF(r) + C_{10}\int_{r^*}^{1} (1-r)\,dF(r)$$

$$+ C_{00}\int_{0}^{r^*} (1-r)\,dF(r) \rightarrow \min$$

$$r^* = \frac{C_{10} - C_{00}}{C_{10} + C_{01} - C_{00} - C_{11}}$$

$$EL = C_{11}\mu\ \text{sens}(\text{r}^*) + C_{01}\mu(1-\text{sens}(\text{r}^*)) +$$
$$C_{10}(1-\mu)(1-\text{spec}(\text{r}^*)) + C_{00}(1-\mu)\text{spec}(\text{r}^*)$$

If sens(r*)=1 and spec(r*)=1

$$EL = C_{11}\mu\ + C_{00}(1-\mu)$$

Special Cases

1. $C_{00}=C_{11}=0$; $C_{10}=C_{01}$

overall loss=misclassification rate:

$$EL=\int_0^{r^*} r\,dF(r)+\int_{r^*}^{1}(1-r)\,dF(r)$$

EL minimized for r*=0.5

Special Cases, cont

2. $C_{00} = C_{11} = 0; C_{10} << C_{01},$

$$r^* = \frac{C_{10}}{C_{01} + C_{10}} << 1$$

Recall:

$$EL = C_{11}\mu\ \text{sens}(r^*) + C_{01}\mu(1-\text{sens}(r^*)) +$$
$$C_{10}(1-\mu)(1-\text{spec}(r^*)) + C_{00}(1-\mu)\text{spec}(r^*)$$

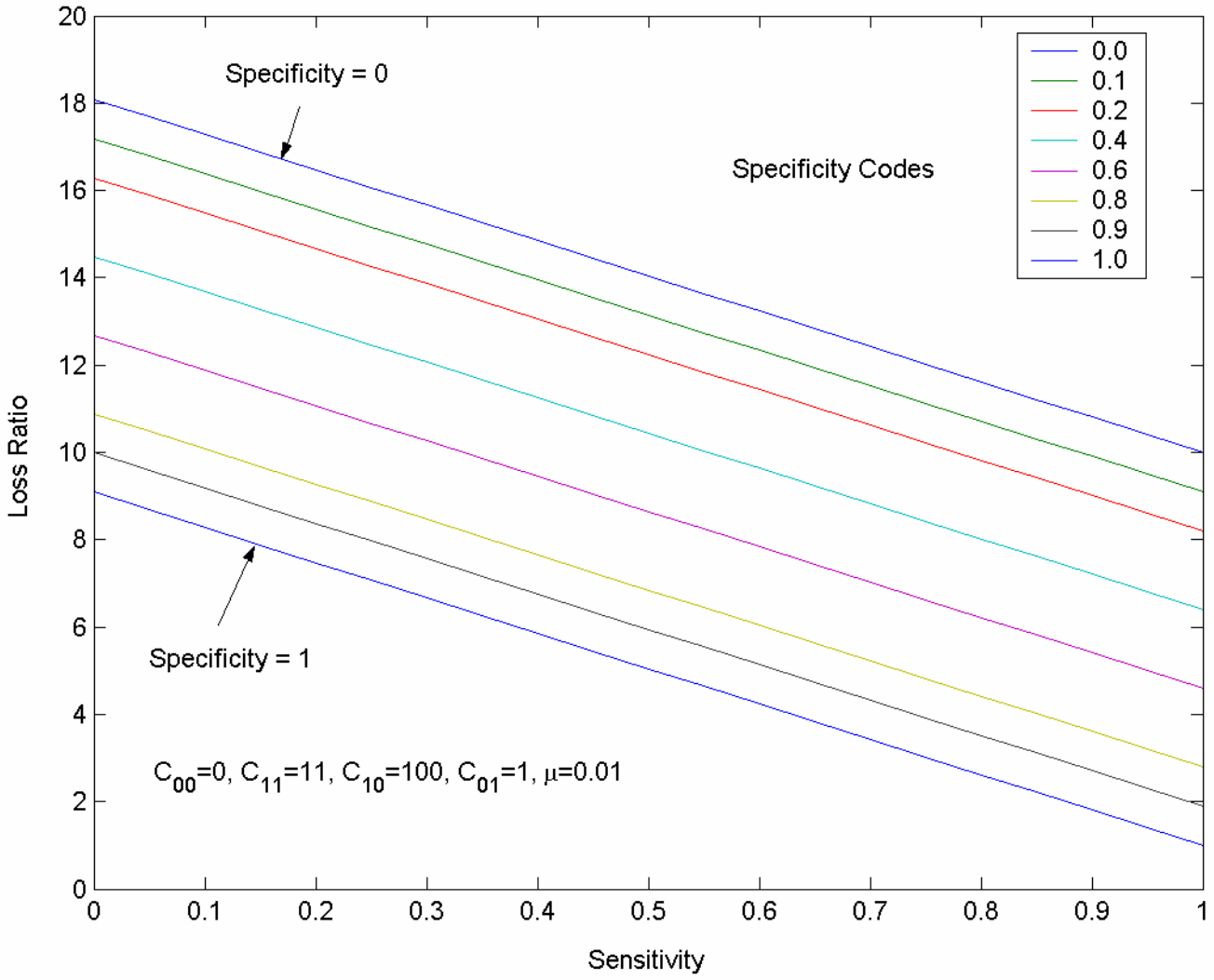
If sens(r*)=1 and spec(r*)=1

$$EL = C_{11}\mu\ + C_{00}(1-\mu)$$

Should Mammographic Screen be Recommended Based on a Risk Model?

Outcome over next 5 Years	No Screen	Screen
Y=0 (no cancer)	0	1
Y=1 (cancer)	100	11

Ratio of Expected Loss to Minimum Expected Loss vs Sensitivity



Intervention Setting

Two outcomes: eg Y_1=breast cancer

Y_2=stroke

Loss

	$Y_1=0$	$Y_1=1$
$Y_2=0$	C_{00}	C_{01}
$Y_2=1$	C_{10}	C_{11}

Intervention Setting

Intervention does not change cost, it changes probability function of joint outcomes

No intervention: $P_{\delta=0}(Y_1, Y_2)$

Intervention: $P_{\delta=1}(Y_1, Y_2)$

$$EL = \sum_{i,j,k} C_{ji} P\delta = k(Y_1 = i, Y_2 = j)$$

Ideally we would have joint risk model for both outcomes, Y_1, Y_2

Simplification: $P_i(Y_1=1, Y_2=1|x) = p_{2i}\, r_i(x)$

$p_{21} = p_{20}\ \rho_2$

$r_1(x) = r_0(x)\rho_1$

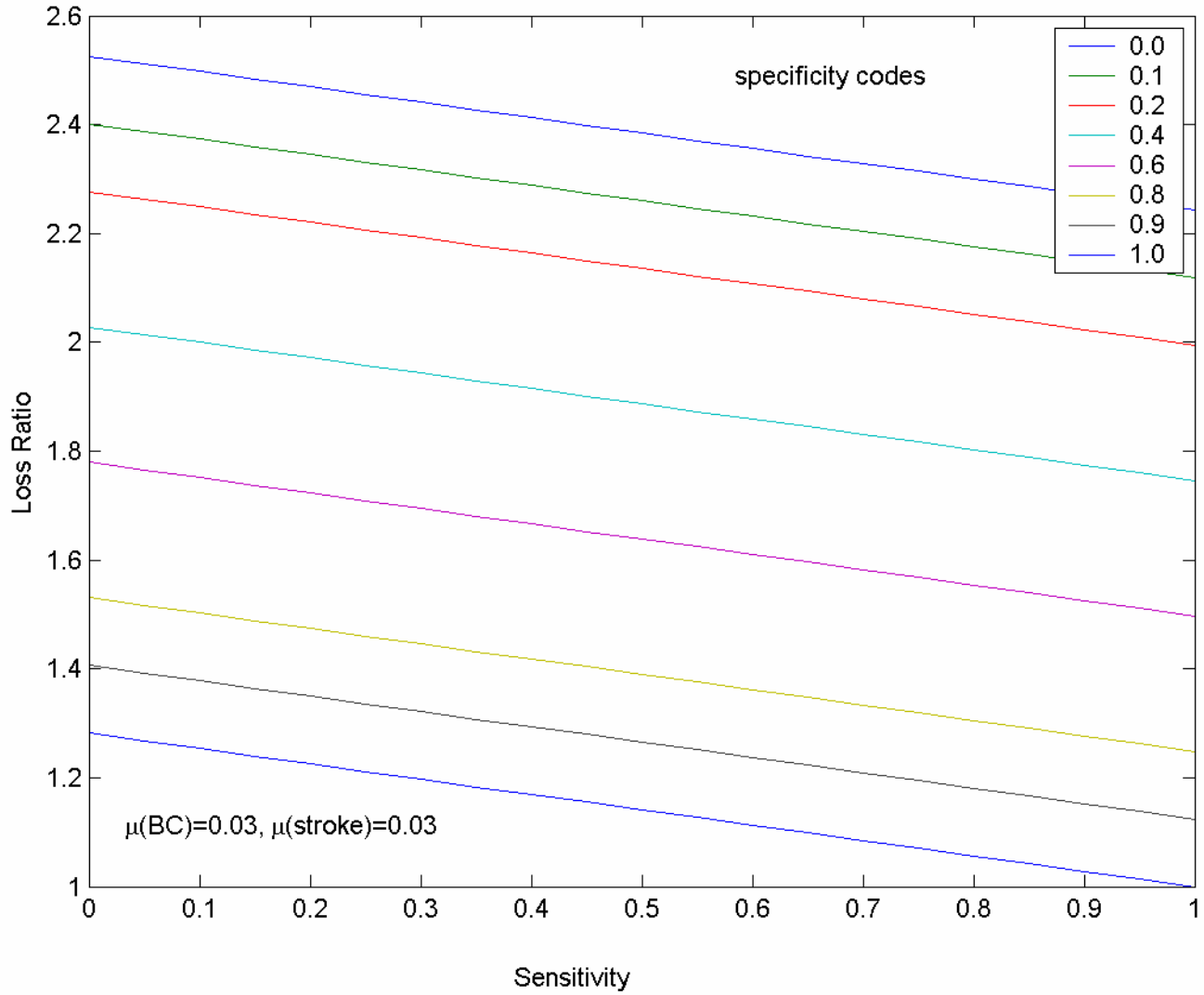
$$EL = \sum_{i,j,k} C_{ki} P\delta = k(Y_1 = i, Y_2 = j) \rightarrow \min r^*$$

Loss function for clinical decision: should woman take Tamoxifen for breast cancer prevention?

$\rho_1=0.5$, $\rho_2=3$

Over next 5 years	No Breastcancer	Breastcancer
No Stroke	0	1
Stroke	1	2

Ratio of Expected Loss to Expected Loss with sens=spec=1 vs Sensitivity



Summary

- For certain applications (screening) high sensitivity and specificity more important than others (clinical decision making)
- Always want a well calibrated model
- Discriminatory aspects of models may be less important than accuracy and calibration

Collaborators

Mitchell Gail, NCI

Andrew Freedman, NCI
Patricia Hartge, NCI

References

Brier GW, 1950, Monthly Weather Review, 75, 1-3

Dodd LE, Pepe M, 2003, JASA 98 (462): 409-417

Efron B, 1986, JASA 81 (394): 461-470

Efron B, 1983, JASA 78 (382): 316-329

Gail MH et al, 1999, JNCI, 91 (21): 1829-1846

Hand DJ, 2001, Statistica Neerlandica, 55 (1): 3-16

Hand DJ, 1997, *Construction and assessment of classification rules*, Wiley.

Pepe MS 2000, JASA, 95 (449): 308-311

Schumacher M, et al, 2003, Methods of information in medicine 42: 564-571

Steyerberg EW, et al, 2003, Journal of Clinical Epidemiology 56: 441-447

AUC value for the Gail et al Model 2

0.58

Relative Risk Estimates for "Gail Model"

Risk Factor	
Age at menarche (yrs.) (>14, 12-13, <12)	1.00-1.21
Number of Biopsies (0, 1, 2+)	1.00-2.88
Age at first live birth (yrs.) (<20, 20-24, 25-29, ≥ 30)	1.00-1.93
# of first degree relatives with breast cancer (0, 1, 2+)	1.00-6.80

Intervention Setting

Two outcomes: eg Y_1=breast cancer

Y_2=stroke

Loss

	$Y_1=0$	$Y_1=1$	$Y_2=0$	$Y_2=1$
$\delta=0$	C^1_{00}	C^1_{01}	C^2_{00}	C^2_{01}
$\delta=1$	C^1_{10}	C^1_{11}	C^2_{10}	C^2_{11}