# Design and Analysis of Biomarker Studies for Risk Prediction

Yingye Zheng

Fred Hutchinson Cancer Research Center, Seattle WA

November 18, 2015

**Introduction: Overview**

**Prediction with a Single Marker**

- *defining proper measures of accuracy*
- *estimating accuracy summaries*

**Prediction with Multiple Markers**

- *constructing composite scores through regression*
- *evaluating the accuracy of composite scores*
- *evaluating the incremental value of new markers*
- *additional considerations*

**Design and Analytical Strategies**

- *motivations*
- *nested case control & case cohort designs*
- *estimation and inference*
- *additional considerations*
    - *efficiency: design and estimation perspectives*
    - *other sampling strategies*

**Motivating Examples**

- Framingham Offspring Study
- Breast Cancer Gene-expression Profile Studies

- Examples of biomarkers
  - *genetic markers such as SNPs*
  - *protein markers such as PSA, CA125, CRP*
  - *risk scores:*
    - *Framingham Risk score (Anderson et al, 1991; Wilson et al, 1998): age, total cholesterol, HDL, blood pressure, present smoking status, and diabetes mellitus.*
    - *Breast Cancer Risk Assessment Tool (BCRAT) (Gail et al, 1989: age at menarche, age at first live birth, number of previous breast biopsies, and number of first- degree relatives with BC.*
- Biomarkers are used in clinical settings
  - *as a surrogate endpoints/exposures*
  - *for risk prediction and stratification*
  - *as a treatment-selection tool*

- Risk prediction and stratification play a central role in medical decision making
  - *Predicted risks $\rightsquigarrow$ appropriate intervention.*
  - *Example: prevention strategies according to predicted CHD risks by the AHA*
  - *BCRAT: identify high risk women for MRI screening and chemoprevention*

- We are still far behind in molecular diagnosis and prognosis: Accurate risk assessment is a difficult task

- To develop prediction rules for optimal risk assessment, we need to
  - *Identify important predictors*
  - *Develop risk prediction models*
  - *Evaluate and compare risk prediction rules with rigorous assessment (beyond p value)*

# General Framework

- Study Design:
  - *Risk factors measured at baseline*
  - *Subjects followed for the occurrence of an event*
  - *Outcome of interest: whether an event occurs within t-years*
- Goal:
  - *predict the risk of developing an event within t-years*
  - *evaluate the performance of such a risk prediction rule*
  - *compare to the existing prediction rules*
- Challenges:
  - *incorporating the time domain*
  - *censoring*
  - *competing risks*
  - *biomarkers too expensive to measure and samples are valuable: optimal study designs*

# Survival Prediction with a Single Marker

In many clinical studies, the outcome of interest $T$ is time to the occurrence of a clinical condition.

- Examples: time to disease diagnosis; onset of a CVD event; death.

Marker of interest $Y$ is measured at baseline

- Examples: Framingham Risk Score; CRP; gene expression signature score.

To assess the accuracy of a marker $Y$ in predicting the event time $T$, various accuracy measures have been suggested:

- Calibration: the ability to correctly predict the proportion of subjects within any given group who will experience disease events
  - **Prediction Error (Brier score)** (Graf et al 1999; Begg et al, 2000)
- Discrimination:the ability to distinguish between patients who are at higher compared with lower risk.
  - **Time-dependent Classification/Predictive measures: TPR, FPR, PPV, NPV** (Heagerty & Pepe, 2000; Heagerty & Zheng, 2005; Cai et al, 2005; Zheng et al, 2008, 2010)
  - **Mean Risk Difference (MRD), Net Benefit (NB), Proportion of Cases Followed (PCF), Proportion Need to Follow-up (PNF)**

    Vickers & Elkin, 2006; Gu & Pepe, 2009; Pfeiffer & Gail, 2011; Zheng et al, 2012
  - **Overall concordance measures** (Harrell et al, 1982; Begg et al, 2000; Uno et al, 2010)
- Reclassification of new models versus existing one

- One approach to quantifying the predictiveness of a marker $Y$ for a survival outcome $T$ is to consider the prediction of $t$-year survival, i.e. the prediction of a binary outcome

$$D_t = I(T \leq t)$$

by constructing binary prediction rules $I(Y \geq c)$ with some threshold value $c$.

- Many of the existing prediction accuracy measures are developed by examining the ability of $I(Y \geq c)$ in predicting $D_t$.

The classification accuracy of $I(Y \geq c)$ in predicting $D_t$ may be summarized by

$$\mathrm{TPR}_t(c) = P(Y \geq c \mid D_t = 1), \qquad \mathrm{FPR}_t(c) = P(Y \geq c \mid D_t = 0),$$

This corresponds to a time dependent ROC curve

$$\mathrm{ROC}_t(c) = \mathrm{TPR}_t\left\{\mathrm{FPR}_t^{-1}(u)\right\}$$

The prediction accuracy measures can be defined as

$$\mathrm{PPV}_t(c) = P(D_t = 1 \mid Y > c) \qquad \mathrm{NPV}_t(c) = P(D_t = 0 \mid Y \leq c)$$

(Zheng & Heagerty 2004; Heagerty & Zheng 2005; Cai et al 2006; Zheng et al 2006)

In general, several types of time dependent ROC curves have been proposed by defining $D_t$ and the populations of interest differently.

$$\begin{aligned}
\text{Entire Population}: &\quad D_t = 1 \quad \text{if } T \le t, \quad D_t = 0 \quad \text{if } T > t \\
\{T \le t\} \cup \{T > \tau\}: &\quad D_t = 1 \quad \text{if } T \le t, \quad D_t = 0 \quad \text{if } T > \tau \\
\\
\{T \ge t\}: &\quad D_t = 1 \quad \text{if } T = t, \quad D_t = 0 \quad \text{if } T > t \\
\{T = t\} \cup \{T > \tau\}: &\quad D_t = 1 \quad \text{if } T = t, \quad D_t = 0 \quad \text{if } T > \tau
\end{aligned}$$

- $\tau$ is a pre-defined time point such that $T > \tau$ is considered *controls*.

Classification accuracy measures can be defined accordingly.

For example, Heagerty and Zheng (2005) and Cai et al (2006) defined various types of ROC curves.

- **Cumulative / Dynamic** $\mathbf{ROC}_t(u) = \mathbf{TPR}_t\{\mathbf{FPR}_t^{-1}(u)\}$

$$\mathbf{TPR}_t(c) = P(Y \geq c \mid T \leq t), \qquad \mathbf{FPR}_t(c) = P(Y \geq c \mid T > t)$$

- **Incident / Dynamic** $\mathbf{ROC}_t^{\mathbb{I}\mathbb{D}}(u) = \mathbf{TPR}_t^{\mathbb{I}}\{\mathbf{FPR}_t^{\mathbb{D}^{-1}}(u)\}$

$$\mathbf{TPR}_t^{\mathbb{I}}(c) = P(Y \geq c \mid T = t), \qquad \mathbf{FPR}_t^{\mathbb{D}}(c) = P(Y \geq c \mid T > t)$$

Time dependent *overall accuracy measures*, such as the AUC, could also be derived from the corresponding definitions of time dependent ROC curves.

- The Cumulative / Dynamic ROC curve leads to

$$\text{AUC}_t = \int \text{ROC}_t(u)du = P(Y_1 \geq Y_2 \mid T_1 \leq t, T_2 > t)$$

- The Incident / Dynamic ROC curve leads to

$$\text{AUC}_t^{\mathbb{ID}} = \int \text{ROC}_t^{\mathbb{ID}}(u)du = P(Y_1 \geq Y_2 \mid T_1 = t, T_2 > t)$$

  - $\text{AUC}_t^{\mathbb{ID}}$ is closely related to the standard concordance measure $\mathcal{C}$, and Kendal's $\tau$, $\mathcal{K}$ $P(Y_1 > Y_2 \mid T_1 < T_2)$. (Heagerty and Zheng, 2005).

In practice, it is often of interest to consider these accuracy measures at the risk scale based on

$$\mathcal{R}_t(Y) = P(T \leq t \mid Y)$$

The classification and predictive accuracy functions can be defined accordingly:

$$\text{TPR}_t(p) = P\{\mathcal{R}_t(Y) > p \mid T \leq t\}, \qquad \text{FPR}_t(p) = P\{\mathcal{R}_t(Y) \geq p \mid T > t\}$$
$$\text{PPV}_t(p) = P\{T \leq t \mid \mathcal{R}_t(Y) > p\}, \qquad \text{NPV}_t(p) = P\{T > t \mid \mathcal{R}_t(Y) \leq p\}$$

(Pepe et al 2008; Zheng et al 2010; Gu & Pepe 2011)

- The mean risk difference $\text{MRD}_t = E\{\mathcal{R}_t(Y) \mid D_t = 1\} - E\{\mathcal{R}_t(Y) \mid D_t = 0\} \equiv \text{ITP}_t - \text{IFP}_t$ summarizes the difference in the mean risk between the cases and the controls.
- The net benefit at time $t$ is a point on the decision curve (Vickers & Elkin, 2006; Baker, 2009), with $\rho_t = P(T \leq t)$,

$$\text{NB}_t(p) = \rho_t \text{TPR}_t(p) - \frac{p}{1-p}(1-\rho_t)\text{FPR}_t(p)$$

,

- Define $\bar{\mathcal{V}}(p) \equiv \text{P}[\mathcal{R}_t(\mathbf{Y}) > p]$: Pfeiffer &Gail (2011) proposed proportion of case followed (PCF)

$$\text{PCF}_t(v) = \text{TPR}_t\{\bar{\mathcal{V}}^{-1}(v)\}$$

and proportion needed to follow-up (PNF)

$$\text{PNF}_t(p) = \text{PCF}_t^{-1}(p).$$

In most studies with event time outcomes, the event time is subject to *censoring* due to loss to follow up or end of study. Consequently, for event time $T$, we observe

$$(X, \Delta), \quad \text{where} \;\; X = \min(T, C), \;\Delta = I(T \leq C)$$

where $C$ is the follow-up (censoring) time.

Estimation of the accuracy measures requires assumptions about $C$:

- *A stronger assumption requires $C$ to be independent of both $T$ and $Y$ with a common survival function $S_C(t) = P(C \geq t)$.*
- *A weaker assumption requires $C$ to be independent of the event time $T$ conditional on the marker value $Y$, but may depend on $Y$.*

Suppose we are interested in estimating

$$\mathrm{TPR}_t(c) = P(Y \geq c \mid T \leq t) = \frac{P(T \leq t \mid Y \geq c)P(Y \geq c)}{P(T \leq t)}$$

Due to censoring, $D_t = I(T \leq t)$ is not always observable.

Various approaches may be taken to account for censoring.

- *Inverse probability weighted (IPW) estimator*
- *Robust estimator based on conditional Nelson Aalen (CNA)*

If $C \perp (T, Y)$, $\text{TPR}_t(c)$ may be consistently estimated based on

- Kaplan-Meier estimates of $P(T \leq t)$ and $P(T \leq t \mid Y \geq c)$. For any $c$, $P(T \leq t \mid Y \geq c)$ may be estimated using observations from the subset of patients with $\{Y \geq c\}$.
- An IPW approach with weights

$$W_{Ci}(t) = \frac{I(X_i \leq t)\delta_i}{S_C(X_i)} + \frac{I(X_i > t)}{S_C(t)}$$

Note that $I(T_i \leq t)$ is observable if $I(X_i \leq t)\delta_i = 1$ or $I(X_i > t) = 1$.

For the IPW approach, one may show that

$$E\{W_{Ci}(t)I(T_i \le t, Y_i \ge c) \mid T_i, Y_i\} = I(T_i \le t, Y_i \ge c)$$

and hence

$$\frac{\sum_{i=1}^n W_{Ci}(t)I(Y_i \ge c, T_i \le t)}{\sum_{i=1}^n W_{Ci}(t)I(T_i \le t)} \to \frac{E\{W_{Ci}(t)I(Y_i \ge c, T_i \le t)\}}{E\{W_{Ci}(t)I(T_i \le t)\}} = \mathrm{TPR}_t(c).$$

Thus, $\mathrm{TPR}_t(c)$ may be estimated by

$$\widehat{\mathrm{TPR}}_t(c) = \frac{\sum_{i=1}^n \widehat{W}_{Ci}(t)I(Y_i \ge c, T_i \le t)}{\sum_{i=1}^n \widehat{W}_{Ci}(t)I(T_i \le t)}.$$

where $\widehat{W}_{Ci}(t)$ **is obtained by replacing** $S_C(\cdot)$ **or** $S_{C,Y_i}(\cdot)$ **in** $W_{Ci}(t)$ **by their respective estimates,** $\widehat{S}_C(\cdot)$ **(e.g. Kaplan Meier).**

If $C$ depends on $Y$ but is independent of $T$ conditional on $Y$, one may estimate $\text{TPR}_t(c)$ by first estimating

$$S_y(t) = P(T \leq t \mid Y = y)$$

and subsequently constructing a plug in estimate of $\text{TPR}_t(c)$ based on

$$P(T \leq t \mid Y \geq c) = \frac{\int_c^\infty S_y(t) dF(y)}{1 - F(c)}, \quad \text{where } F(y) = P(Y \leq y)$$

$S_y(t)$ may be estimated

- *semi-parametrically by assuming a regression model for $T \mid Y$ such as the Cox and the AFT model* *(Kalbfleish & Prentice, 2002)*
- *non-parametrically via conditional Kaplan-Meier (Nelson Aalen) with kernel weights $K_h(Y_i - y)$* *(Dabrowska 1989; Du & Akritas, 2002)*

# Framingham Offspring Study for CVD Prediction

**Framingham Heart Study**:

- Goal: identifying risk factors for CVD
- Framingham Risk Score for CHD/Stroke prediction
- 3 generations
  - *original cohort (1948)*
  - *Offspring cohort (1971): ¿5000 followed prospectively*
  - *3rd generation cohort (2002)*

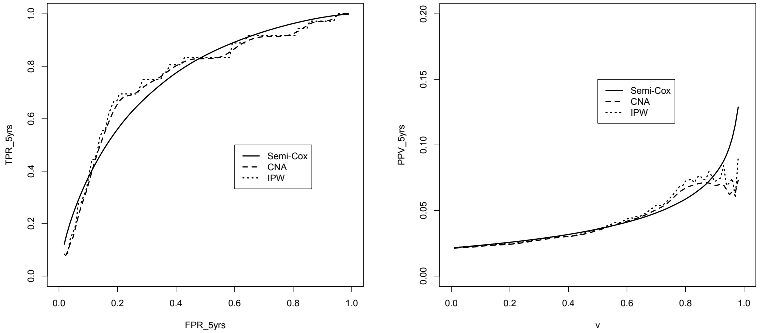
**Framingham Offspring Study Female Participants**

- 1687 female out of a total 5124 participants
- 261 events (death/CVD) with 10-year event rate 6%
- Framingham risk score (Wilson et al. 1998)
- C-reactive protein (CRP) (Cook et al, 2006; Ridker et al, 2007)

Table: Non-parametric estimates (Est) and standard errors (SE) of accuracy measures ($\times$ 100) for 5-year survival based on the conditional Nelson Aalen (CNA), IPW method and the semi parametric Cox model. Here $c_p$ is the $p$th percentile of the observed risk score in the full cohort.

| | CNA | | IPW | | Semi-Cox | |
|---|---|---|---|---|---|---|
| | Est | SE | Est | SE | Est | SE |
| $\text{FPR}_5(c_{.2})$ | 79.7 | 1.0 | 79.7 | 1.0 | 79.6 | 1.0 |
| $\text{FPR}_5(c_{.8})$ | 19.1 | 1.0 | 18.8 | 0.9 | 19.3 | 1.0 |
| $\text{TPR}_5(c_{.2})$ | 92.8 | 4.5 | 91.9 | 4.3 | 96.2 | 0.6 |
| $\text{TPR}_5(c_{.8})$ | 61.2 | 7.9 | 62.2 | 7.7 | 54.9 | 3.0 |
| $\text{NPV}_5(c_{.2})$ | 99.2 | 0.5 | 99.1 | 0.5 | 99.2 | 0.1 |
| $\text{NPV}_5(c_{.8})$ | 99.0 | 0.3 | 99.0 | 0.3 | 98.8 | 0.2 |
| $\text{PPV}_5(c_{.2})$ | 2.5 | 0.4 | 2.5 | 0.4 | 2.6 | 0.4 |
| $\text{PPV}_5(c_{.8})$ | 6.5 | 1.3 | 6.8 | 1.4 | 5.9 | 1.0 |
| AUC | 75.2 | 4.1 | 75.8 | 3.9 | 75.7 | 1.5 |
| $\text{FPR}_{\text{TPR}=.9}$ | 65.0 | 13.9 | 58.7 | 8.4 | 61.8 | 3.0 |
| $\text{NPV}_{\text{TPR}=.9}$ | 99.4 | 0.3 | 99.4 | 0.7 | 99.4 | 0.5 |
| $\text{PPV}_{\text{TPR}=.9}$ | 2.9 | 0.8 | 3.2 | 0.2 | 3.1 | 0.5 |

Figure: Time-dependent ROC curve (a) and PPV curve (b) of the risk score for predicting 5-year CVD events.




(a) (b)

# Survival Prediction with Multiple Markers

When there are multiple markers available to assist in prediction, one may construct a composite score as for binary outcomes.

A wide range of survival regression models have been proposed in the literature.

- Cox proportional hazards model;
- Proportional odds model;
- Semi-parametric transformation model;
- Accelerated Failure Time (AFT) model;
- non-parametric transformation model;
- time-specific generalized linear model.

## Cox Proportional Hazards (PH) Model (Cox, 1972)

$$\lambda_{\mathbf{Y}}(t) = \frac{f_{\mathbf{Y}}(t)}{S_{\mathbf{Y}}(t)} = \lambda_0(t) \exp(\boldsymbol{\beta}_0^\top \mathbf{Y})$$

- $\lambda_{\mathbf{Y}}(t)$ is the hazard function for a subject with marker value $\mathbf{Y}$, $f_{\mathbf{Y}}(t)$ is the density of $T$ given $\mathbf{Y}$ and $S_{\mathbf{Y}}(t) = P(T > t \mid \mathbf{Y})$, and $\lambda_0(t)$ is the baseline hazard function.
- An equivalent form of the model is

$$P(T \leq t \mid \mathbf{Y}) = g(h_0(t) + \boldsymbol{\beta}_0^\top \mathbf{Y})$$

  where $g(x) = 1 - e^{-e^x}$ and $h_0(\cdot)$ is an unknown increasing function.
- $\boldsymbol{\beta}_0$ may be estimated by maximizing the partial likelihood.

## Proportional Odds (PO) Model

$$\text{logit } P(T \leq t \mid \mathbf{Y}) = h_0(t) + \boldsymbol{\beta}_0^\mathsf{T}\mathbf{Y}$$

- For any fixed $t \Rightarrow$ logistic regression with response $I(T \leq t)$.
- Rank based estimator (Pettitt, 1984) and non-parametric maximum likelihood estimator (Murphy et al, 1997) have been proposed for $\boldsymbol{\beta}_0$.

## Semi-parametric Transformation Model

$$P(T \leq t \mid \mathbf{Y}) = g\left\{h_0(t) + \boldsymbol{\beta}_0^\mathsf{T}\mathbf{Y}\right\}, \qquad g(\cdot) \text{ known and } \uparrow$$

- An equivalent form of the model is

$$h_0(T) = -\boldsymbol{\beta}_0^\mathsf{T}\mathbf{Y} + \epsilon \quad \text{with} \quad P(\epsilon \leq x) = g(x)$$

- Estimation equation based estimators for $\boldsymbol{\beta}_0$ have been proposed by Cheng et al (1995) and Chen et al (2002). Zeng & Lin (2006) developed a non-parametric maximum likelihood estimator.

## Accelerated Failure Time Model

$$\log(T) = \boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{Y} + \epsilon, \qquad \epsilon \sim F(\cdot) \text{ unknown}$$

- Since the model may be written as $T = T_0 e^{\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{Y}}$, $\boldsymbol{\beta}_0$ can be interpreted as the acceleration rate.
- To estimate $\boldsymbol{\beta}_0$, Buckley and James (1979) proposed iterative weighted least square estimator; Tsiatis (1990) and Jin et al (2003) studied rank based estimators.

## Non-parametric Transformation Model

$$P(T \leq t \mid \mathbf{Y}) = g\left\{h_0(t) + \boldsymbol{\beta}_0^\top \mathbf{Y}\right\} \quad \text{or} \quad h_0(T) = -\boldsymbol{\beta}_0^\top \mathbf{Y} + \epsilon$$

- Both the link function $g(\cdot)$ and the baseline function $h_0(\cdot)$ are completely unspecified.
- Maximum rank correlation based estimator for $\boldsymbol{\beta}_0$ Khan & Tammer, 2007; Cai & Cheng, 2008).
- Under the general transformation framework, across all time $t$, $\boldsymbol{\beta}_0^\top \mathbf{Y}$
  - is the optimal score in distinguishing $\{T \leq t\}$ from $\{T > t\}$.
  - achieves the highest $\text{ROC}_t(\cdot)$ among all scores of $\mathbf{Y}$.
- Under the PH and PO models, across all time $t$, $\boldsymbol{\beta}_0^\top \mathbf{Y}$ is also
  - the optimal score in distinguishing $\{T = t\}$ from $\{T > t\}$
  - achieves the highest $\text{ROC}_t^{\mathbb{ID}}(\cdot)$ among all scores of $\mathbf{Y}$.

## Time-specific Generalized Linear Model

Markers useful for identifying short term survivors may be not be useful for identifying long term survivors.

To construct time-dependent optimal score, one may consider *time-specific* generalized linear models:

$$P(T \leq t \mid \mathbf{Y}) = g\left\{h_0(t) + \boldsymbol{\beta}_0(t)^{\mathsf{T}}\mathbf{Y}\right\}$$

- Without censoring, for any given time $t$, one may fit a usual GLM to the data $\{D_t, \mathbf{Y}\}$ to obtain an estimate of $\boldsymbol{\beta}_0(t)$.
- To incorporate censoring, Zheng et al (2006) and Uno et al (2007) considered IPW estimators for time-specific logistic regression model.
- $\boldsymbol{\beta}_0(t)^{\mathsf{T}}\mathbf{Y}$ is the optimal score in distinguishing $\{T \leq t\}$ from $\{T > t\}$ and achieves the highest $\mathrm{ROC}_t(\cdot)$ .

By fitting the survival models, one may obtain an estimate of the regression coefficient. For example,

- For the PH model, one may estimate $\boldsymbol{\beta}_0$ as the maximizer of the log partial likelihood function,

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[ \boldsymbol{\beta}^{\mathsf{T}} \mathbf{Y}_i - \log \left\{ n^{-1} \sum_{j=1}^{n} I(X_j \geq X_i) e^{\boldsymbol{\beta}^{\mathsf{T}} \mathbf{Y}_j} \right\} \right]$$

- For the time-specific GLM, one may estimate $\boldsymbol{\beta}_0(t)$ as the solution to the weighted estimating equation

$$\sum_{i=1}^{n} \widehat{W}_{Ci}(t) \begin{pmatrix} 1 \\ \mathbf{Y}_i \end{pmatrix} \left\{ I(T_i \leq t) - g(\alpha + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{Y}_i) \right\} = 0$$

Suppose $\widehat{\beta}(t)$ is the estimator for the effect of $\mathbf{Y}$ and let $\boldsymbol{\beta}_0(t)$ denote its limit.

- *For many of the existing estimators, $\widehat{\beta}(t)$ is unique and converges to a deterministic vector $\boldsymbol{\beta}_0(t)$ regardless of model adequacy.*
- *When the fitted models hold, these estimators are consistent for the true model parameter; when the fitted models fail to hold, these estimators are consistent for a limiting vector $\boldsymbol{\beta}_0(t)$.*

The accuracy of the composite score $\boldsymbol{\beta}_0(t)^\top \mathbf{Y}$ may be estimated non-parametrically by replacing $\boldsymbol{\beta}_0(t)^\top \mathbf{Y}$ as $\widehat{\beta}(t)^\top \mathbf{Y}$.

- For example, assuming that the censoring is independent of $T$ and $\mathbf{Y}$,

$$\mathrm{TPR}_t\{c; \boldsymbol{\beta}_0(t)\} = P\{\boldsymbol{\beta}_0(t)^\top \mathbf{Y} \geq c \mid T \leq t\}$$

  may be estimated by

$$\widehat{\mathrm{TPR}}_t\{c; \widehat{\beta}(t)\} = \frac{\sum_{i=1}^{n} \widehat{W}_{Ci}(t) I(\widehat{\beta}(t)^\top \mathbf{Y}_i \geq c, T_i \leq t)}{\sum_{i=1}^{n} \widehat{W}_{Ci}(t) I(T_i \leq t)}$$

  where $\widehat{W}_{Ci}(t) = \frac{I(X_i \leq t)\delta_i}{\widehat{S}_C(X_i)} + \frac{I(X_i > t)}{\widehat{S}_C(t)}$ and $\widehat{S}_C(t)$ is the Kaplan-Meier estimator of $S_C(t) = P(C > t)$.

# Estimating the Accuracy of the Composite Score

- Alternatively, a more robust estimator for $\mathrm{TPR}_t\{c; \boldsymbol{\beta}_0(t)\}$ may be constructed as

$$\widehat{\mathrm{TPR}}_t\{c; \widehat{\boldsymbol{\beta}}(t)\} = \frac{\int_c^\infty \widehat{S}_y\{t; \widehat{\boldsymbol{\beta}}(t)\} d\widehat{F}(y; \widehat{\boldsymbol{\beta}}(t)\}}{1 - \widehat{F}\{c; \widehat{\boldsymbol{\beta}}(t)\}}.$$

  where $\widehat{S}_y(t; \boldsymbol{\beta})$ is the conditional Kaplan Meier estimator of $P(D_t = 1 \mid \boldsymbol{\beta}^\mathsf{T}\mathbf{Y} = y)$ based on synthetic data $\{(X_i, \delta_i, \boldsymbol{\beta}^\mathsf{T}\mathbf{Y}_i)\}$ with kernel weights $K_h(\boldsymbol{\beta}^\mathsf{T}\mathbf{Y}_i - y)$.

- With either type of estimators,

$$\mathrm{ROC}_t\{u; \boldsymbol{\beta}_0(t)\} = \mathrm{TPR}_t\left[\mathrm{FPR}_t^{-1}\{u; \boldsymbol{\beta}_0(t)\}; \boldsymbol{\beta}_0(t)\right]$$

  may be estimated by plugging in $\widehat{\mathrm{TPR}}_t\{c; \widehat{\boldsymbol{\beta}}(t)\}$ and $\widehat{\mathrm{FPR}}_t\{c; \widehat{\boldsymbol{\beta}}(t)\}$.

# Inference for the Accuracy Parameters

- The asymptotic distribution of these accuracy estimators can be shown to normal. However, explicit variance estimation may be difficult especially under model mis-specification.
- Resampling procedures can be used to approximate the distribution.
- Example: suppose $\widehat{\boldsymbol{\beta}}(t) = \widehat{\boldsymbol{\beta}}$ is obtained through fitting the Cox PH model, then the distribution of $n^{\frac{1}{2}}\{\widehat{\mathrm{TPR}}_t(c;\widehat{\boldsymbol{\beta}}) - \mathrm{TPR}_t(c;\boldsymbol{\beta}_0)\}$ can be approximated by the distribution of $n^{\frac{1}{2}}\{\widehat{\mathrm{TPR}}_t^*(c;\widehat{\boldsymbol{\beta}}^*) - \widehat{\mathrm{TPR}}_t(c;\widehat{\boldsymbol{\beta}})\} \mid$ the observed data, where

$$\widehat{\mathrm{TPR}}_t^*\{c;\widehat{\boldsymbol{\beta}}^*(t)\} = \frac{\sum_{i=1}^n \widehat{W}_{Ci}^*(t) I(\widehat{\boldsymbol{\beta}}^*(t)^{\mathsf{T}}\mathbf{Y}_i \geq c, T_i \leq t)\mathcal{V}_i}{\sum_{i=1}^n \widehat{W}_{Ci}^*(t) I(T_i \leq t)\mathcal{V}_i}$$

  - $\{\mathcal{V}_i, i = 1, ..., n\}$ i.i.d with mean 1 and variance 1;
  - $\widehat{\boldsymbol{\beta}}^*$ obtained by fitting the Cox PH with weights $\{\mathcal{V}_i, i = 1, ..., n\}$;
  - $\widehat{W}_{Ci}^*(t) = \frac{I(X_i \leq t)\delta_i}{\widehat{S}_C^*(X_i)} + \frac{I(X_i > t)}{\widehat{S}_C^*(t)}$ and $\widehat{S}_C^*(t)$ is the Kaplan-Meier estimator with weights $\{\mathcal{V}_i, i = 1, ..., n\}$.

# Example: Gene Expression Markers for Predicting Breast Cancer Survival

The New England
Journal of Medicine

Copyright © 2002 by the Massachusetts Medical Society

VOLUME 347 DECEMBER 19, 2002 NUMBER 25

**A GENE-EXPRESSION SIGNATURE AS A PREDICTOR OF SURVIVAL IN BREAST CANCER**

MARC J. VAN DE VIJVER, M.D., PH.D., YUDONG D. HE, PH.D., LAURA J. VAN 'T VEER, PH.D., HONGYUE DAI, PH.D., AUGUSTINUS A.M. HART, M.SC., DORIEN W. VOSKUIL, PH.D., GEORGE J. SCHREIBER, M.SC., JOHANNES L. PETERSE, M.D., CHRIS ROBERTS, PH.D., MATTHEW J. MARTON, PH.D., MARK PARRISH, DOUWE ATSMA, ANKE WITTEVEEN, ANNUSKA GLAS, PH.D., LEONIE DELAHAYE, TONY VAN DER VELDE, HARRY BARTELINK, M.D., PH.D., SJOERD RODENHUIS, M.D., PH.D., EMIEL T. RUTGERS, M.D., PH.D., STEPHEN H. FRIEND, M.D., PH.D., AND RENÉ BERNARDS, PH.D.

- 295 breast cancer patients who were diagnosed with breast cancer between 1984 and 1995. The median survival time is 3.8 years for these patients.
- Outcome: time to death
- Markers: gene expression markers
    - The gene expression measurement is the logarithm of the intensity ratios between the red and the green fluorescent dyes, where green dye is used for the reference pool and red is used for the experimental tissue.
    - The prognosis rule developed by van't veer et al (2002) and Vijver et al (2002) was derived based on a 70 gene expression markers.
    - For illustration, we selected 6 out of 70 gene expression markers for prediction.

# Example: Gene Expression Markers for Predicting Breast Cancer Survival

- Obtain a linear score $\widehat{\beta}(t)^{\mathsf{T}}\mathbf{Y}$ for classifying $I(T \leq t)$ by fitting various regression models:
  - proportional hazards model $\lambda_{\mathbf{Y}}(t) = \lambda_0(t)e^{\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{Y}}$
  - proportional odds model $\text{logit}P(T \leq t \mid \mathbf{Y}) = h_0(t) + \boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{Y}$.
  - time-specific logistic regression model $\text{logit}P(T \leq t \mid \mathbf{Y}) = h_0(t) + \boldsymbol{\beta}_0(t)^{\mathsf{T}}\mathbf{Y}$
  - AFT model: $\log T = \boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{Y} + \epsilon$
- Estimate the ROC curve,

$$\text{ROC}_t(\cdot),$$

for distinguishing $\{T \leq t\}$ from $\{T > t\}$ by estimating

$$\text{TPR}_t(c), \quad \text{and} \quad \text{FPR}_t(c)$$

non-parametrically using inverse-probability weighting approach.

- Summarize the overall accuracy of $\widehat{\beta}(t)^{\mathsf{T}}\mathbf{Y}$ by estimating

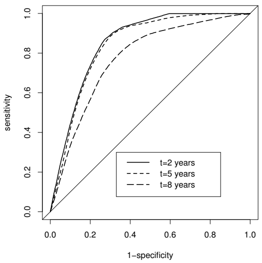
$$\text{AUC}_t = \int_0^1 \text{ROC}_t(u)du.$$

# Example: Gene Expression Markers for Predicting Breast Cancer Survival

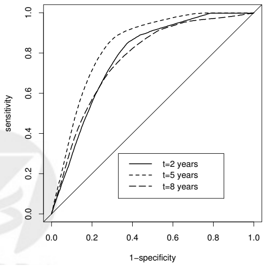
Table: Estimated $\text{AUC}_t$ (95% CI) at $t = 2$, 5 and 8 years after diagnosis using a 6-gene classifier with linear composite scores derived from different regression models.

| | $t = 2$ years | $t = 5$ years | $t = 8$ years |
|---|---|---|---|
| Cox | .78(.62, .87) | .84(.78, .88) | .77(.71, .84) |
| Proportional Odds | .78(.59, .87) | .83(.68, .88) | .77(.65, .84) |
| Time-specific Logistic | .85(.80, .91) | .84(.80, .89) | .77(.71, .84) |
| AFT | .81(.70, .88) | .84(.81, .89) | .78(.72, .84) |

# Example: Gene Expression Markers for Predicting Breast Cancer Survival




(a) Logistic




(b) Cox

When the sample size $n$ is not large with respect to the number of markers, one may use cross-validation methods to obtain less biased accuracy estimators.

- Randomly split the data into $K$ disjoint sets of about equal size and label them as $\mathcal{I}_k, k = 1, \cdots, K$.
- For each $k$,
  - an estimate for the model parameters may be obtained based on, $\mathcal{I}_{(-k)}$, all observations which are not in $\mathcal{I}_k$;
  - the accuracy of the resulting risk score trained in $\mathcal{I}_{(-k)}$ may be estimated based on data in $\mathcal{I}_k$.
- A bias corrected estimator of the accuracy measure may be obtained by averaging over the $K$ accuracy estimates.

- In addition to obtaining a point estimator for the accuracy, it is crucial to assess the **variability** in the estimated accuracy measure.
  - The variability may be assessed via procedures such as the bootstrap although theoretical justification may be difficult.
  - Other types of resampling methods such as the aforementioned perturbation have also been considered in the literature. (Parzen et al, 1994; Jin et al, 2003; Cai et al, 2005; Tian et al, 2007; Uno et al, 2007).

- Results given in Tian et al (2007) & Uno et al (2007) imply that the confidence intervals (CI) can be constructed as follows:
  - center of the CI: the cross-validated estimators
  - width of the CI: using the resampling procedure to assess the variability of the apparent accuracy

## Step (I) Risk Modeling

- Fitting survival models such as the Cox PH and time-specific logistic regression model

$$P(D_t = 1 \mid \mathbf{Y}) = g(\mathbf{Y}; \boldsymbol{\beta}_t)$$

- Risk Score for a future patient with $\mathbf{Y}^0$: $\widehat{\mathcal{R}}_t(\mathbf{Y}^0) = g(\mathbf{Y}^0; \widehat{\boldsymbol{\beta}}_t)$

$$\widehat{\mathcal{R}}_t(\mathbf{Y}^0) > c \Rightarrow T^0 \le t_0; \quad \widehat{\mathcal{R}}_t(\mathbf{Y}^0) \le c \Rightarrow T^0 > t_0$$

## Step (II) Evaluation of Prediction Accuracy

Estimating accuracy measures such as

$$\mathrm{TPR}_t(c) = P\{\mathcal{R}_t(\mathbf{Y}^0) > c \mid T^0 \le t_0\} \quad \mathrm{FPR}_t(c) = P\{\mathcal{R}_t(\mathbf{Y}^0) > c \mid T^0 > t_0\}$$

as well as other measures such as $\mathrm{ROC}_t(\cdot)$, $\mathrm{AUC}_t$, $\mathrm{MRD}_t$.

# Summary

- The choice of the accuracy measure may depend on the clinical questions of interest.
- To obtain estimators for the classification accuracy measures with survival outcomes, one needs to incorporate censoring appropriately.
- When there are multiple markers available, various survival regression models may be used to construct composite scores for prediction. Such scores may be optimal with respect to certain accuracy measures when the imposed model holds.
- Bias correction and variance estimation should be considered when assessing the accuracy.
- When assessing subgroup specific incremental values, it is crucial to account for multiple comparisons.

**Marker too expensive to be measured on all study participants?**

$\downarrow$

**Two-Phase Study Designs**

# Motivating Example: the Breast Cancer Study

Research article **Open Access**

**A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients**

Laurel A Habel[1], Steven Shak[2], Marlena K Jacobs[1], Angela Capra[1], Claire Alexander[2], Mylan Pho[2], Joffre Baker[2], Michael Walker[2], Drew Watson[2], James Hackett[2], Noelle T Blick[1], Deborah Greenberg[3], Louis Fehrenbacher[4], Bryan Langholz[5] and Charles P Quesenberry[1]

- Aim: to evaluate the prognostic capacity of a tumor gene expression based Recurrence Score for breast cancer mortality.
- Recurrence Score (Oncotype DX): 21-gene assay developed based on 250 genes assembled from various sources.

- For most women with early-stage invasive breast cancer adjuvant hormonal and/or chemotherapy are recommended. Both have adverse effects.
- Most patients with node-negative disease who receive chemotherapy will not derive benefit, because they would not go on to have a recurrence even without such treatment.
- Treatment decisions are based on age, node status, tumor size, and some histologic information.
- Multigene assays may provide information on patient prognosis and response to therapy that is superior /complementary to standard clinical information.
- Multiple studies in independent populations are needed to establish the clinical usefulness of these assays.

# Motivating Example (Habel et al., 2006)

- Study cohort: about 5000 Kaiser Permanente patients diagnosed with node-negative invasive breast cancer from 1985 to 1994 (Habel et al., 2006)
- Standard full cohort analysis not feasible: new markers too expensive
- NCC design: controls are individually matched to cases with respect to age, race, adjuvant tamoxifen, diagnosis year, and were alive at the date of death of their matched cases.
- Habel et al., 2006 went as far as reporting OR and absolute risk using a conditional logistic regression.

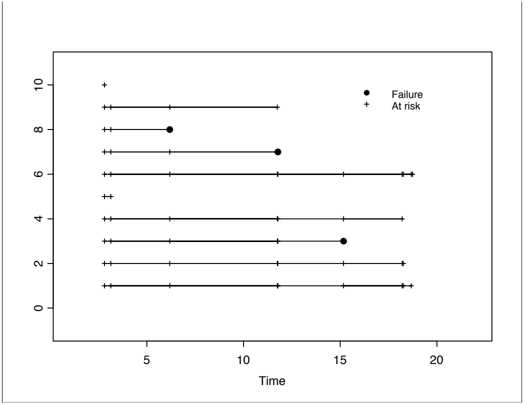
# Motivating Research Questions

- How to analyze data collected with complex cohort study designs?
- How to design cohort study that allows more efficient marker evaluation?

# Study Designs

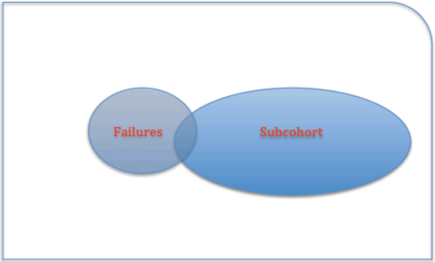
- Prospective cohort studies are desirable:
  - easy calculation of absolute risks at various time points
  - avoid selection bias
- Cohort study may not always be feasible:
  - rare disease outcome lead to a big cohort with few cases and many controls
  - biomarker can be expensive to measure
  - biospecimens are of limited quantity

- Cost-effective two-phase designs
- Widely adopted in large cohort studies:
  - Atherosclerosis Risk in Communities (ARIC) study(Folsom et al. 2002)
  - Nurse's Health Study ($\sim$ 2000 publications) (Colditz et al. 1997)
  - Women's Health Initiative ($\sim$ 1000 publications)(Anderson et al. 2003)
- Are of great value for biomarker studies:
  - avoid having to measure expensive markers on all subjects
  - achieve similar efficiency compared with a full cohort analysis




- Nested case-control study (NCC) (Thomas, 1977; Prentice & Breslow; 1978)
- Covariate matched NCC

# Two-phase Designs: Illustration




- Case-cohort study (CCH) (Prentice, 1986)
- Stratified CCH (Borgan et al., 2000; Gray 2008)

- CCH:
  Pseudo-likelihood for estimating relative risk parameters under CCH design (Prentice, 1986) with asymptotic properties developed (Self & Prentice, 1988)
- NCC:
  Conditional logistic regression for estimating relative risk parameters (Thomas, 1977). Asymptotic properties have been formally derived for estimators of hazard ratios (Goldstein and Langholz, 1992) and absolute risk (Langholz and Borgan,1997).
- Marker evaluation adds another level of complexity (population distribution of marker associated risks). Different approach is needed for estimating prediction performance summaries.

- Cast the problem within the general framework of failure time analysis with missing covariates under MAR
  - **Inverse probability weighted approach**
  - **Likelihood based approach**

- N individuals, each followed to $X_i$, $X_i = \min(T_i, C_i)$, $\delta_i = I(X_i = T_i)$.
- $\mathbf{Z}_i$ covariate measures for all; $Y_i$ sampled only for a subset at the second phase. $V_i = 1$ if $Y_i$ is measured.
- **stratified CCH:**
  $L$ strata are defined based on $(\delta_i, \mathbf{Z}_i)$; sample $n_l$ out of set $\mathcal{R}_l$ with size $N_l$ in each $l$.
- **covariate-matched NCC:**
  For $j$th selected failure, covariate-specific risk set,
  $\mathcal{R}_{\mathbf{Z}}(X_j) = \{i : I(X_i \geq X_j)I(\mathbf{Z}_j = \mathbf{Z}_i)\}$. with size $n_{\mathbf{Z}}(X_j)$.
  $m$ 'controls' are sampled without replacement from $\mathcal{R}_{\mathbf{Z}}(X_j) \setminus j$.

## Inverse probability weighted (IPW) Estimators

- IPW: based on only selected observations ($V_i = 1$)
- Weighing the contributions from selected observations with weight $\widehat{w}_i = V_i/\widehat{p}_i$
- $\widehat{p}_i$: the probability of the $i$th subject ever being selected based on the sampling scheme of the study design.
- For CCH, $\widehat{p}_i = \sum_{l=1}^{L} I(i \in \mathcal{R}_l) n_l/N_l$;
- For NCC, $\widehat{p}_i = \delta_i + (1-\delta_i)\{1 - \widehat{G}(X_i)\}$

$$\widehat{G}(X_i) = \Pi_{X_j < X_i} \left\{ 1 - \frac{m \Delta_j V_j I(\mathbf{Z}_j = \mathbf{Z}_i)}{n_{\mathbf{Z}}(X_j) - 1} \right\}$$

- $E\{V_i/\widehat{p}_i \mid (X_i, \delta_i)\} = 1$.

Plug-in estimators for risk distribution indices under a NCC study:

$$
\begin{aligned}
\widehat{\mathrm{TPF}}_t^{\mathrm{NCC}}(p) &= \frac{\int_{\widehat{\mathcal{R}}_t^{\mathrm{NCC}-1}(p)}^{\infty} \widehat{\mathcal{R}}_t^{\mathrm{NCC}}(y) d\widehat{F}^{\mathrm{NCC}}(y)}{\int_{-\infty}^{\infty} \widehat{\mathcal{R}}_t^{\mathrm{NCC}}(y) d\widehat{F}^{\mathrm{NCC}}(y)} \\
\widehat{\mathrm{FPF}}_t^{\mathrm{NCC}}(p) &= \frac{\int_{\widehat{\mathcal{R}}_t^{\mathrm{NCC}-1}(p)}^{\infty} \{1-\widehat{\mathcal{R}}_t^{\mathrm{NCC}}(y)\} d\widehat{F}^{\mathrm{NCC}}(y)}{\int_{-\infty}^{\infty} \{1-\widehat{\mathcal{R}}_t^{\mathrm{NCC}}(y)\} d\widehat{F}(y)} \\
\widehat{\mathrm{MDR}}_t^{\mathrm{NCC}} &= \int \widehat{\mathrm{TPF}}_t^{\mathrm{NCC}}(p) dp - \int \widehat{\mathrm{FPF}}_t^{\mathrm{NCC}}(p) dp
\end{aligned}
$$

where

- $\widehat{F}^{\mathrm{NCC}}(y) = \sum_{i=1}^{N} \frac{V_i/\widehat{p}_i}{\sum_j V_j/\widehat{p}_j} I(Y_i > y)$.
- $\widehat{\mathcal{R}}_t^{\mathrm{NCC}}(y)$ can be obtained
  - *semi-parametrically; or*
  - *non-parametrically*

- Under a Cox model, $\widehat{\beta}_{\mathrm{NCC}}$ is obtained from a weighted partial likelihood (Samuelsen, 1997):

$$\mathcal{L}(\beta) = \sum_{i=1}^{N} \widehat{w}_i \delta_i \left\{ \beta Y_i - \log \sum_{j=1}^{N} \widehat{w}_j I(X_j \geq X_i) \exp(\beta Y_j) \right\}.$$

- $\widehat{\mathcal{R}}_t^{\mathrm{NCC}}(y) = 1 - \exp\{-\widehat{\Lambda}_0^{\mathrm{NCC}}(t) \exp(\widehat{\beta}^{\mathrm{NCC}} y)\}$, where

$$\widehat{\Lambda}_0^{\mathrm{NCC}}(t) = \sum_{i=1}^{N} \frac{\widehat{w}_i I(X_i \leq t) \delta_i}{\sum_{j \in \mathcal{R}_i} \widehat{w}_j \exp(\widehat{\beta}_{\mathrm{NCC}} Y_j)}$$

(Cai and Zheng, 2011a)

- $\widehat{\mathcal{R}}^{LB}(t|y)$ uses individuals in $\widetilde{\mathcal{R}_i}$ (Langholz and Borgan,1997).

# Non-parametric Estimation of $\mathcal{R}_t^{\mathsf{NCC}}(y)$

With a single marker $Y$, the conditional risk $\widehat{\mathcal{R}}_t^{\mathsf{NCC}}(y)$ can be obtained via IPW kernel smoothing based on the data

$$\{(X_i, \delta_i, Y_i), i = 1, ..., N\}$$

using *weighted conditional Kaplan-Meier* or Nelson Aalen estimator with weights

$$K_h(Y_i - y)\widehat{w}_i$$

(Cai and Zheng, 2011b)

# Double IPW Estimation

Under the independent censoring assumption $C \perp (Y, T)$, the accuracy parameters could be estimated using double IPW with weights

- $\widehat{W}_{ci}(t)$ to account for censoring; and
- $\widehat{w}_i$ to account for the missingness in $Y$

For example, $\text{TPR}_t(p)$ can be estimated as

$$\widehat{\text{TPR}}_t(p) = \frac{\sum_{i=1}^{n} \widehat{W}_{Ci}(t) \widehat{w}_i I\left\{\widehat{F}^{\text{NCC}}(Y_i) \geq p, T_i \leq t\right\}}{\sum_{i=1}^{n} \widehat{W}_{Ci}(t) \widehat{w}_i I(T_i \leq t)}.$$

(Cai and Zheng, 2011b)

- Under the finite population sampling scheme, $V_i$ are weakly dependent conditional on data.
- General theory for IPW estimator under stratified CCH design was developed for $\beta$ (Breslow & Wellner, 2006)
- For any summary measures of interest, denoted by a generic term $\mathcal{A}_t$, Under sCCH, $\widehat{W}_{\mathcal{A}_t} = N^{\frac{1}{2}}\{\widehat{\mathcal{A}}_t - \mathcal{A}_t\} = N^{-\frac{1}{2}}\sum_{i=1}^{N} w_i \widetilde{U}_{\mathcal{A}_t}(\mathcal{H}_i) + o_p(1)$,
- $N^{\frac{1}{2}}\{\widehat{\mathcal{A}}_t - \mathcal{A}_t\}$ has variance function $\Sigma_{\widehat{\mathcal{A}}_t} = Q\{\widetilde{U}_{\mathcal{A}_t}\}$, where

$$Q(U) = \mathrm{Var}(U) + \sum_{l=1}^{L} v_l(\pi_l^{-1} - 1)\mathrm{Var}_l(U)$$

with $\mathrm{Var}_l$ denoting the variance within the $l$th stratum.

(Liu, Cai and Zheng, 2012)

- General theory for IPW estimator of NCC design is not well developed.
- Establish asymptotic properties using results on the strong and weak convergence of weighted sums of negatively associated dependent variables (Liang and Baek, 2006).

- For any summary measures of interest, denoted by a generic term $\mathcal{A}^w$,

$$N^{1/2}(\widehat{\mathcal{A}}^w - \mathcal{A}^w) = \mathrm{N}^{-1/2} \sum_{i}^{N} \widehat{w}_i U_{\mathcal{A}i} + o_p(1),$$

which is asymptotically normal with mean 0 and variance

$$\sigma_{\mathcal{A}}^2 = E\left(\frac{U_{\mathcal{A}i}^2}{p_i}\right) - m\mathcal{R}_{U_{\mathcal{A}}}^2 = E(U_{\mathcal{A}i}^2) + E\left\{\sigma_{\widehat{w}_i|\mathcal{D}}^2 U_{\mathcal{A}i}^2\right\} - m\mathcal{R}_{U_{\mathcal{A}}}^2,$$

where $\mathcal{R}_{U_{\mathcal{A}}}$ is some complicated function.

(Cai and Zheng, 2011b)

# Inverse Probability Weighted (IPW) Estimators

- We have developed IPW estimators for estimating many summary indices under different designs.
- Flexible and simple to implement; Robust to censoring assumptions.
- Theoretical justification is difficult with finite sampling, but is needed as standard Bootstraps does not work (recent development with resampling method exist: (Cai & Zheng, 2013; Huang, 2015)
- Not fully efficient. There are ways to improve. (more discussions later)

- Nonparametric maximum likelihood estimators (NPMLE) for for hazard ratio parameters under the Cox model under case-cohort (Scheike, Martinussen, 2004) and nested case control study (Scheike, Juul, 2004) have been developed.
- The work can be extended to the estimation of various summary indices of prediction performance.

Table: Simulation Results from NPMLE Estimators

| | | | | | |
|---|---|---|---|---|---|
| $\text{TPR}_t(c) = 0.953$ | | | | | |
| | % bias | ESD | ASE | CP(%) | RE(%) |
| IPW | -0.01 | 0.0108 | 0.0106 | 93.2 | 100 |
| MLE | -0.01 | 0.0086 | 0.0086 | 94.3 | 63.4 |
| $\text{FPR}_t(c) = 0.715$ | | | | | |
| | % bias | ESD | ASE | CP(%) | RE(%) |
| IPW | 0.10 | 0.0235 | 0.0229 | 94.5 | 100 |
| MLE | 0.08 | 0.0213 | 0.0210 | 94.6 | 82.2 |
| $\text{MDR}_t = 0.292$ | | | | | |
| | % bias | ESD | ASE | CP(%) | RE(%) |
| IPW | -0.004 | 0.023 | 0.022 | 93.1 | 100 |
| MLE | -0.005 | 0.017 | 0.017 | 93.4 | 54.6 |

- **Inverse probability weighted approach**
  - allows for both nonparametric and semiparametric procedures
  - flexible in model specification and censoring assumption
  - may be less efficient
- **Likelihood based approach**
  - fully efficient
  - computationally intensive; infeasible for missing in multiple markers
  - biased when censoring is dependent on marker Y

# Additional Considerations

- How to design the study to achieve optimal efficiency?
  - *NCC or CCH?*
  - *Match or no match?*
- Leverage auxiliary information to improve estimation efficiency?
- Alternative estimation procedures?
- Possible practical complications in design?

# Which to Use, NCC or CCH?

- Practical considerations (Wacholder, 1991; Barlow et al. 1999)
  - ease of planning;
  - ease of analysis;
  - ease of reuse samples for future study;
  - batch effects, storage effects, and freezethaw cycles (Rundle et al. 2005).
- Statistical relative efficiency (Langholz &Thomas 1990).

Table: Estimate (SD) of prediction performance indices based on IPW estimators

| | NCCz | | CCHz | | Full Cohort | |
|---|---|---|---|---|---|---|
| $\hat{\beta}$ | 1.101 | (0.046) | 1.101 | (0.046) | 1.101 | (0.039) |
| AUC | 0.787 | (0.009) | 0.788 | (0.009) | 0.788 | (0.008) |
| MDR | 0.176 | (0.014) | 0.177 | (0.014) | 0.177 | (0.012) |
| NB($\rho_t$) | 0.062 | (0.005) | 0.063 | (0.005) | 0.063 | (0.005) |
| TPR(0.05) | 0.946 | (0.006) | 0.946 | (0.005) | 0.946 | (0.005) |
| FPR(0.05) | 0.692 | (0.028) | 0.692 | (0.027) | 0.691 | (0.024) |
| PPV(0.05) | 0.190 | (0.007) | 0.190 | (0.006) | 0.190 | (0.006) |
| NPV(0.05) | 0.971 | (0.001) | 0.971 | (0.001) | 0.971 | (0.001) |
| TPR(0.25) | 0.481 | (0.025) | 0.48 | (0.024) | 0.481 | (0.022) |
| FPR(0.25) | 0.116 | (0.009) | 0.115 | (0.009) | 0.116 | (0.008) |
| PPV(0.25) | 0.415 | (0.010) | 0.416 | (0.010) | 0.416 | (0.009) |
| NPV(0.25) | 0.909 | (0.004) | 0.909 | (0.003) | 0.909 | (0.003) |
| PCF(0.20) | 0.531 | (0.014) | 0.532 | (0.013) | 0.531 | (0.012) |
| PNF(0.85) | 0.475 | (0.015) | 0.475 | (0.015) | 0.476 | (0.013) |

- Why match: eliminate confounding; gain in efficiency
- Should match on confounding factors to improve efficiency in evaluating risk model?

Table: ARE (Full cohort versus specific design) for estimates under CCH design

| | Model 1 | | Model 2 | | Model 2 - Model 1 | |
|---|---|---|---|---|---|---|
| Matching | NO | YES | NO | YES | NO | YES |
| $\beta_1$ | 0.502 | 0.762 | 0.479 | 0.528 | | |
| $\beta_2$ | | | 0.481 | 0.531 | | |
| AUC | 0.603 | 0.341 | 0.645 | 0.34 | 0.523 | 0.119 |
| DMR | 0.598 | 0.888 | 0.633 | 0.877 | 0.541 | 0.692 |
| NB($\rho_t = 0.18$) | 0.847 | 0.929 | 0.854 | 0.911 | 0.450 | 0.597 |
| TPR($p = 0.05$) | 0.625 | 0.309 | 0.599 | 0.272 | 0.238 | 0.066 |
| FPR($p = 0.05$) | 0.610 | 0.647 | 0.623 | 0.611 | 0.383 | 0.209 |
| TPR($p = 0.25$) | 0.622 | 0.800 | 0.628 | 0.707 | 0.357 | 0.388 |
| FPR($p = 0.25$) | 0.644 | 0.785 | 0.616 | 0.710 | 0.320 | 0.372 |
| PCF(0.20) | 0.654 | 0.850 | 0.694 | 0.842 | 0.501 | 0.632 |
| PNF(0.85) | 0.619 | 0.835 | 0.657 | 0.804 | 0.472 | 0.535 |

Model 1: $Y^{old}$; Model 2: $Y^{old} + Y^{new}$; Z: $Y^{old} > 0$

Table: ARE (Full cohort versus specific design) for estimates under NCC design

| | Model 1 | | Model 2 | | Model 1 - Model 2 | |
|---|---|---|---|---|---|---|
| Matching | NO | YES | NO | YES | NO | YES |
| $\beta_1$ | 0.51 | 0.775 | 0.46 | 0.581 | | |
| $\beta_2$ | | | 0.455 | 0.586 | | |
| AUC | 0.578 | 0.081 | 0.611 | 0.067 | 0.506 | 0.036 |
| DMR | 0.591 | 0.766 | 0.616 | 0.743 | 0.495 | 0.623 |
| NB($\rho_t = 0.18$) | 0.667 | 0.484 | 0.661 | 0.467 | 0.358 | 0.478 |
| TPR($p = 0.05$) | 0.499 | 0.135 | 0.47 | 0.122 | 0.199 | 0.041 |
| FPR($p = 0.05$) | 0.541 | 0.267 | 0.555 | 0.257 | 0.354 | 0.142 |
| TPR($p = 0.25$) | 0.526 | 0.495 | 0.488 | 0.471 | 0.289 | 0.279 |
| FPR($p = 0.25$) | 0.513 | 0.341 | 0.478 | 0.319 | 0.230 | 0.245 |
| PCF(0.20) | 0.569 | 0.411 | 0.589 | 0.379 | 0.420 | 0.466 |
| PNF(0.85) | 0.572 | 0.400 | 0.583 | 0.375 | 0.485 | 0.364 |

Model 1: $Y^{old}$; Model 2: $Y^{old} + Y^{new}$; Z: $Y^{old} > 0$

# How to Design a Study Using Auxiliary Covariate Information to Improve Study Efficiency

- Under sCCH, $\widehat{W}_{\mathcal{A}_t} = N^{\frac{1}{2}}\{\widehat{\mathcal{A}}_t - \mathcal{A}_t\} = N^{-\frac{1}{2}} \sum_{i=1}^{N} w_i \widetilde{U}_{\mathcal{A}_t}(\mathcal{H}_i) + o_p(1)$,
- $N^{\frac{1}{2}}\{\widehat{\mathcal{A}}_t - \mathcal{A}_t\}$ has variance function $\Sigma_{\mathcal{A}_t} = Q\{\widehat{U}_{\mathcal{A}_t}\}$, where

$$Q(U) = \mathrm{Var}(U) + \sum_{l=1}^{L} v_l(\pi_l^{-1} - 1)\mathrm{Var}_l(U)$$

  with $\mathrm{Var}_l$ denoting the variance within the $l$th stratum.
- The overall sampling fraction $\pi$, is predetermined and a stratified cohort sampling design will be adopted.
- One can gain efficiency by minimizing the second terms of the asymptotic variances, subject to the constraint that $\pi = \sum_{l=1}^{L} v_l \pi_l$.

- The optimal sampling fraction for strutum $l$ for an accuracy measure $\mathcal{A}$

$$\widetilde{\pi}_l = \pi \frac{\mathrm{Var}_l(\widehat{U}_{\mathcal{A}})^{1/2}}{\sum_{j=1}^{L} v_j \mathrm{Var}_j(\widehat{U}_{\mathcal{A}})^{1/2}}.$$

- The formula is similar to the 'Neyman allocation' in survey studies.
- Practical implication.

# Summary of Design Options

- Both NCC and CCH designs offer logistic efficiency compared with full cohort. The statistical efficiency achieved often are comparable in many situations.
- Efficiency can be improved by considering
  - matching in some situations for some measures;
  - more efficient estimation procedures: e.g., augmented weights (Breslow &Wellner (2007)) or MLE. This may achieve similar efficiency while preserving simplicity in design implement.

## Summary

Challenges and important considerations in biomarker evaluation for risk prediction

- *incorporating the time domain & censoring when building and evaluating the risk prediction models*
- *choice of the accuracy parameters*
- *robust/efficient estimation of the accuracy parameters*
- *two-phase design issues*
  - *for both CCH and NCC designs, we considered methods that varies in terms of flexibility, robustness and efficiency.*
  - *investigators can now take advantage of various two-phase designs and conduct analysis for more efficient and rigorous biomarker validation.*
  - *the methods also easily extend to more complicated yet more flexible study designs.*
- *Software available at: http://www.fredhutch.org/en/labs/profiles/zheng-yingye.html*

**Andersen, P. and Gill, R.** (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics* **10**, 1100–1120. ISSN 0090-5364.

**Anderson, G., Manson, J., Wallace, R., Lund, B., Hall, D., Davis, S., Shumaker, S., Wang, C., Stein, E. and Prentice, R.** (2003). Implementation of the Women's Health Initiative study design. *Annals of Epidemiology* **13**, S5–17. ISSN 1047-2797.

**Anderson, K., Wilson, P., Odell, P. and Kannel, W.** (1991). An updated coronary risk profile. A statement for health professionals. *Circulation* **83**, 356–362.

**Baker, S., Cook, N., Vickers, A. and Kramer, B.** (2009). Using relative utility curves to evaluate risk prediction. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **172**, 729–748.

**Barlow, W., Ichikawa, L., Rosner, D. and Izumi, S.** (1999). Analysis of case-cohort designs. *Journal of clinical epidemiology* **52**, 1165–1172.

**Begg, C. B., Cramer, L. D., Venkatraman, E. S. and Rosai, J.** (2000). Comparing tumour staging and grading systems: A case study and a review of the issues, using thymoma as a model. *Statistics in Medicine* **19**, 1997–2014.

**Borgan, O., Goldstein, L. and Langholz, B.** (1995). Methods for the analysis of sampled cohort data in the cox proportional hazards model. *The Annals of Statistics* , 1749–1778.

**Breslow, N., Lumley, T., Ballantyne, C., Chambless, L. and Kulich, M.** (2009). Improved Horvitz–Thompson Estimation of Model Parameters from Two-phase Stratified Samples: Applications in Epidemiology. *Statistics in Biosciences* **1**, 32–49. ISSN 1867-1764.

**Breslow, N. and Wellner, J.** (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to cox regression. *Scand. J. Statist.* **34**, 86–102.

**Buckley, J. and James, I.** (1979). Linear regression with censored data. *Biometrika* **66**, 429–436.

**Bura, E. and Gastwirth, J.** (2001). The binary regression quantile plot: assessing the importance of predictors in binary regression visually. *Biometrical Journal* **43**, 5–21.

**Cai, T. and Cheng, S.** (2008). Robust combination of multiple diagnostic tests for classifying censored event times. *Biostatistics* **9**, 216.

**Cai, T., Pepe, M., Zheng, Y., Lumley, T. and Jenny, N.** (2006). The sensitivity and specificity of markers for event times. *Biostatistics* **7**, 182–97.

**Cai, T. and Zheng, Y. a.** (2011a). Evaluating Prognostic Accuracy of biomarkers under Nested Case-control Studies. *Biostatistics* **000**, 000.

**Cai, T. and Zheng, Y. b.** (2011b). Nonparametric evaluation of biomarker accuracy under nested case-control studies. *Journal of the American Statistical Association* , 1–12.

**Chen, K., Jin, Z. and Ying, Z.** (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89**, 659–668.

**Cheng, S., Wei, L. and Ying, Z.** (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835.

**Cheng, S., Wei, L. and Ying, Z.** (1997). Predicting survival probabilities with semiparametric transformation models. *Journal of the American Statistical Association* , 227–235.

**Colditz, G., Manson, J. and Hankinson, S.** (1997). The nurses' health study: 20-year contribution to the understanding of health among women. *Journal of Women's Health* **6**, 49–62. ISSN 1059-7115.

**Cook, N.** (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* **115**, 928.

# References III


**Cook, N.** (2008). Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clinical chemistry* **54**, 17.

**Cook, N., Buring, J. and Ridker, P.** (2006). The effect of including c-reactive protein in cardiovascular risk prediction models for women. *Annals of Internal Medicine* **145**, 21.

**Cox, D.** (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)* , 187–220.

**Cui, J.** (2009). Overview of risk prediction models in cardiovascular disease research. *Annals of Epidemiology* **19**, 711–717.

**Dabrowska, D. M.** (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *The Annals of Statistics* **17**, 1157–1167.

**Du, Y. and Akritas, M. G.** (2002). Uniform strong representation of the conditional kaplan-meier process. *Mathematical Methods of Statistics* **11**, 152–182.

**Folsom, A., Aleksic, N., Catellier, D., Juneja, H. and Wu, K.** (2002). C-reactive protein and incident coronary heart disease in the atherosclerosis risk in communities (aric) study. *American Heart Journal* **144**, 233–238. ISSN 0002-8703.

**Freedman, A., Slattery, M., Ballard-Barbash, R., Willis, G., Cann, B., Pee, D., Gail, M. and Pfeiffer, R.** (2009). Colorectal cancer risk prediction tool for white men and women without known susceptibility. *Journal of Clinical Oncology* **27**, 686.

**Gail, M., Brinton, L., Byar, D., Corle, D., Green, S., Schairer, C. and Mulvihill, J.** (1989). Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *JNCI Journal of the National Cancer Institute* **81**, 1879.

**Goldstein, L. and Langholz, B.** (1992). Asymptotic theory for nested case-control sampling in the Cox regression model. *The Annals of Statistics* **20**, 1903–1928. ISSN 0090-5364.

**Graf, E., Schmoor, C., Sauerbrei, W. and Schumacher, M.** (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statist. Med.* **18**, 2529–45.

**Gu, W. and Pepe, M.** (2009). Measures to summarize and compare the predictive capacity of markers. *International Journal of Biostatistics* **5**, 27.

**Habel, L., Shak, S., Jacobs, M., Capra, A., Alexander, C., Pho, M., Baker, J., Walker, M., Watson, D., Hackett, J. et al.** (2006). A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res* **8**, R25.

**Harrell Jr, F., Lee, K., Califf, R., Pryor, D. and Rosati, R.** (1984). Regression modelling strategies for improved prognostic prediction. *Statistics in medicine* **3**, 143–152.

**Heagerty, P. J., Lumley, T. and Pepe, M. S.** (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56**, 337–344.

**Heagerty, P. J. and Pepe, M. S.** (1999). Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. *Appl. Statist.* **48**, 533–51.

**Heagerty, P. J. and Zheng, Y.** (2005). Survival model predictive accuracy and ROC curves. *Biometrics* **61**, 92–105.

**Henderson, R.** (1995). Problems and prediction in survival-data analysis. *Statistics in Medicine* **14**, 161–184.

**Huang, Y., Sullivan Pepe, M. and Feng, Z.** (2007). Evaluating the predictiveness of a continuous marker. *Biometrics* **63**, 1181–1188.

**Jin, Z., Lin, D., Wei, L. and Ying, Z.** (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–353.

**Kalbfleisch, J. D. and Prentice, R. L.** (2002). *The statistical analysis of failure time data*. John Wiley & Sons.

**Kannel, W., D'Agostino, R., Sullivan, L. and Wilson, P.** (2004). Concept and usefulness of cardiovascular risk profiles* 1. *American Heart Journal* **148**, 16–26.

**Kannel, W., Feinleib, M., McNamara, P., Garrison, R. and Castelli, W.** (1979). An investigation of coronary heart disease in families. *American Journal of Epidemiology* **110**, 281.

**Khan, S. and Tamer, E.** (2007). Partial rank estimation of duration models with general forms of censoring. *Journal of Econometrics* **136**, 251–280.

**Korn, E. L. and Simon, R.** (1990). Measures of explained variation for survival data. *Statistics in Medicine* **9**, 487–503.

**Langholz, B. and Borgan, Y.** (1995). Counter-matching: A stratified nested case-control sampling method. *Biometrika* **82**, 69–79.

**Langholz, B. and Borgan, Y.** (1997). Estimation of absolute risk from nested case-control data. *Biometrics* **53**, 767–774.

**Liang, H. and Baek, J.** (2006). Weighted sums of negatively associated random variables. *Aust. N. Z. J. Stat.* **48**, 21–31.

**Lloyd-Jones, D.** (2010). Cardiovascular Risk Prediction: Basic Concepts, Current Status, and Future Directions. *Circulation* **121**, 1768.

**Murphy, S., Rossini, A. and Van der Vaart, A.** (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association* , 968–976.

# References VI


**Murphy, S. and Van der Vaart, A.** (2000). On profile likelihood. *Journal of the American Statistical Association* , 449–465.

**Parzen, M., Wei, L. and Ying, Z.** (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–350.

**Pearson, T., Blair, S., Daniels, S., Eckel, R., Fair, J., Fortmann, S., Franklin, B., Goldstein, L., Greenland, P., Grundy, S. et al.** (2002). AHA guidelines for primary prevention of cardiovascular disease and stroke: 2002 update: consensus panel guide to comprehensive risk reduction for adult patients without coronary or other atherosclerotic vascular diseases. *Circulation* **106**, 388.

**Pencina, M., D'Agostino Sr, R., D'Agostino Jr, R. and Vasan, R.** (2008). Evaluating the added predictive ability of a new marker: from area under the roc curve to reclassification and beyond. *Statistics in medicine* **27**, 157–172.

**Pencina, M., D'Agostino Sr, R. and Steyerberg, E.** (2011). Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in Medicine* **30**, 11–21.

**Pepe, M., Feng, Z., Huang, Y., Longton, G., Prentice, R., Thompson, I. and Zheng, Y.** (2008a). Integrating the predictiveness of a marker with its performance as a classifier. *American Journal of Epidemiology* **167**, 362.

**Pepe, M., Feng, Z., Janes, H., Bossuyt, P. and Potter, J.** (2008b). Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *Journal of the National Cancer Institute* **100**, 1432–1438.

**Pepe, M., Zheng, Y., Jin, Y., Huang, Y., Parikh, C. and Levy, W.** (2008c). Evaluating the roc performance of markers for future events. *Lifetime data analysis* **14**, 86–113.

**Pettitt, A.** (1984). Proportional odds models for survival data and estimates using ranks. *Applied Statistics* **1**, 169–175.

**Pfeiffer, R. and Gail, M.** (2011a). Two criteria for evaluating risk prediction models. *Biometrics* **67**, 1057–1065.

**Pfeiffer, R. and Gail, M.** (2011b). Two criteria for evaluating risk prediction models. *Biometrics* **67**, 1057–1065.

**Pollard, D.** (1990). *Empirical processes: theory and applications*. Institute of Mathematical Statistics.

**Prentice, R.** (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1. ISSN 0006-3444.

**Ridker, P., Buring, J., Rifai, N. and Cook, N.** (2007). Development and validation of improved algorithms for the assessment of global cardiovascular risk in women. *JAMA: the journal of the American Medical Association* **297**, 611–619.

**Rundle, A., Vineis, P. and Ahsan, H.** (2005). Design options for molecular epidemiology research within cohort studies. *Cancer Epidemiology Biomarkers & Prevention* **14**, 1899–1907.

**Saha, P. and Heagerty, P.** (2010). Time-dependent predictive accuracy in the presence of competing risks. *Biometrics* **66**, 999–1011.

**Samuelsen, S. O.** (1997). A pseudolikelihood approach to analysis of nested case-control studies. *Biometrika* **84**, 379–394.

**Scheike, T. and Juul, A.** (2004). Maximum likelihood estimation for cox's regression model under nested case-control sampling. *Biostatistics* **5**, 193–206.

**Scheike, T. and Martinussen, T.** (2004). Maximum likelihood estimation for cox's regression model under case–cohort sampling. *Scandinavian Journal of Statistics* **31**, 283–293.

**Schemper, M. and Stare, J.** (1996). Explained variation in survival analysis. *Statistics in Medicine* **15**, 1999–2012.

**Schuster, E. and Sype, W.** (1987). On the negative hypergeometric distribution. *International Journal of Mathematical Education in Science and Technology* **18**, 453–459.

**Self, S. and Prentice, R.** (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *The Annals of Statistics* , 64–81.

**Smith, E.** (2009). Risk prediction in cardiovascular disease–current status and future challenges. *The Canadian journal of cardiology* **25**, 7A.

**Steyerberg, E. and Pencina, M.** (2010). Reclassification calculations for persons with incomplete follow-up. *Annals of internal medicine* **152**, 195–196.

**Thomas, D. C.** (1977). Addendum to "Methods of cohort analysis: Appraisal by application to asbestos mining". *Journal of the Royal Statistical Society, Series A, General* **140**, 483–485.

**Thompson, I., Ankerst, D., Chi, C., Goodman, P., Tangen, C., Lucia, M., Feng, Z., Parnes, H. and Coltman Jr, C.** (2006). Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. *JNCI Journal of the National Cancer Institute* **98**, 529.

**Tian, L., Cai, T., Goetghebeur, E. and Wei, L.** (2007). Model evaluation based on the sampling distribution of estimated absolute prediction error. *Biometrika* **94**, 297.

**Tsiatis, A.** (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics* **18**, 354–372.

**Uno, H., Cai, T., Pencina, M., D'Agostino, R. and Wei, L.** (2011). On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in medicine* **30**, 1105–1117.

**Uno, H., Cai, T., Tian, L. and Wei, L.** (2010). Graphical procedures for evaluating overall and subject-specific incremental values from new predictors with censored event time data. *Biometrics* .

**Uno, H., Cai, T., Tian, L. and Wei, L. J.** (2007). Evaluating prediction rules for t-year survivors with censored regression models. *J. Am. Statist. Assoc.* **102**, 527–37.

**Van De Vijver, M., He, Y., Van't Veer, L., Dai, H., Hart, A., Voskuil, D., Schreiber, G., Peterse, J., Roberts, C., Marton, M. et al.** (2002). A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine* **347**, 1999–2009.

**Van't Veer, L., Dai, H., Van De Vijver, M., He, Y., Hart, A., Mao, M., Peterse, H., Van Der Kooy, K., Marton, M., Witteveen, A. et al.** (2002). Gene expression profiling predicts clinical outcome of breast cancer. *nature* **415**, 530–536.

**Vickers, A. and Elkin, E.** (2006). Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making* **26**, 565–574.

**Wacholder, S.** (1991). Practical considerations in choosing between the case-cohort and nested case-control designs. *Epidemiology* , 155–158.

**Wilson, P., D'Agostino, R., Levy, D., Belanger, A., Silbershatz, H. and Kannel, W.** (1998a). Prediction of coronary heart disease using risk factor categories. *Circulation* **97**, 1837.

**Wilson, P., D'Agostino, R., Levy, D., Belanger, A., Silbershatz, H. and Kannel, W.** (1998b). Prediction of coronary heart disease using risk factor categories. *Circulation* **97**, 1837–47.

**Zeng, D. and Lin, D.** (2006). Maximum likelihood estimation in semiparametric transformation models for counting processes. *Biometrika* **93**, 627–40.

**Zeng, D. and Lin, D.** (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**, 507–564.

**Zheng, Y., Cai, T. and Feng, Z.** (2006). Application of the Time-Dependent ROC Curves for Prognostic Accuracy with Multiple Biomarkers. *Biometrics* **62**, 279–287.

# References X


**Zheng, Y., Cai, T., Jin, Y. and Feng, Z.** (2012). Evaluating Prognostic Accuracy of Biomarkers under Competing Risk. *Biometrics* **68**, 388–96.

**Zheng, Y., Cai, T., Pepe, M. and Levy, W.** (2008). Time-dependent predictive values of prognostic biomarkers with failure time outcome. *Journal of the American Statistical Association* **103**, 362.

**Zheng, Y., Cai, T., Stanford, J. and Feng, Z.** (2010). Semiparametric models of time-dependent predictive values of prognostic biomarkers. *Biometrics* **66**, 50–60.

**Zheng, Y. and Heagerty, P.** (2004). Semiparametric estimation of time-dependent roc curves for longitudinal marker data. *Biostatistics* **5**, 615–632.