Optimal Combination of Multiple Markers and Applications to Survival Data with Disease-Free Sampling

Mei-Cheng Wang
Dept. of Biostatistics, Johns Hopkins University

2017 Workshop on Perspectives and Analysis Methods for Personalized Medicine, Singapore

Prospective studies of Alzheimer's Disease

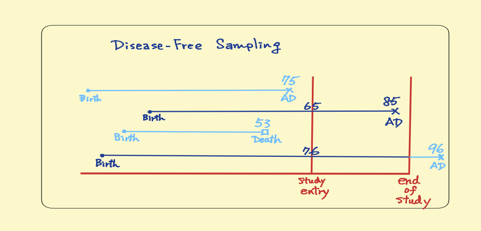
The BIOCARD project at Johns Hopkins studies biomarkers in relation to cognitive decline among normal individuals for progression to Alzheimer's disease (AD). The prospective study started in 1995 and recruited cognitively normal individuals who were free of AD at baseline.

- **Sampling**: Recruit disease-free individuals into the study - this is a common sampling scheme for many AD studies!
- **Biomarkers**: cognition, imaging MRI, genetics and CSF (Cerebrospinal fluid) measures
- **Research aims**:

  - Identify prognostic biomarkers and risk factors to predict AD.
  - Identify protective factors that delay the onset of AD.
  - Help study design for secondary (or even primary) prevention trials.
- **Outcome?**

  T = time from study entry to disease incidence (AD) - for treatment or prevention trials
  T = individual's age at disease incidence (AD) - for natural history study of AD
- **Importance**
  - Alzheimer's disease is the most common type of dementia.
  - Thus far, no treatments stop or reverse its progression, though some may temporarily improve symptoms.
  - In developed countries, AD is one of the most financially costly diseases.
  - In U.S. very large funding allocated to AD research; Aiming to effectively treat and prevent AD by 2025.

The BIOCARD Study of Alzheimer's Disease



- **Outcome?**

  T = time from study entry to disease incidence - RC survival data

  T = individual's age at disease incidence - LTRC survival data

- **An important analytical question:**

  How to optimally use multiple markers to predict AD? Rigorous optimality inference?

- **Proportional hazards models**

  Z: baseline markers, $Z(t)$: time-dependent markers, W: covariates

  $\lambda(t|Z_{hist}(t), W) = \lambda_0(t)exp\{Z(t)^T\beta + W^T\gamma\}$ - not of interest!

  $\lambda(t|Z, W) = \lambda_0(t)exp\{Z^T\beta + W^T\gamma\}$ - more relevant and interesting!

  Keep in mind that biomarkers are typically or usually collected after an individual entered the study.

Talk Content

- **Basics** - notation; background and existing results
- **How to combine multiple biomarkers for optimal prediction of disease incidence?**

  $Z_1, ..., Z_r$: biomarkers, $W_1, ..., W_q$: baseline covariates

  - How to optimally combine $Z_1, ..., Z_r, W_1, ..., W_q$ (markers & covariates)?

  - How to optimally combine $Z_1, ..., Z_r$ (markers only) conditioning on $W_1, ..., W_q$?
- **proportional/additive hazards model**

  - extension to landmark hazards model

  - time-dependent optimality; global optimality
- **Data applications to the BIOCARD study**

  T = time from study entry to disease incidence - RC data

  T = individual's age at disease incidence - LTRC data

Basics for binary outcome

Z: $1 \times r$ baseline markers
W: $1 \times q$ baseline covariates
M: univariate marker

- **When disease outcome D is binary:**

  e.g. logistic regression model: $P(D = 1 \mid Z, W) = \frac{\exp\{\alpha+z^T\beta+w^T\gamma\}}{1+\exp\{\alpha+z^T\beta+w^T\gamma\}}$

  e.g. density ratio model for case-ctrl data $f_1(z, w)$=$f_0(z, w) \cdot \exp\{\alpha^* + z^T\beta + w^T\gamma\}$

  Neyman-Pearson lemma $\Longrightarrow$ The composite marker M=$Z^T\beta$+$W^T\gamma$
  maximizes ROC and AUC.

  McIntosh and Pepe (2002); Qin and Zhang (2009)

- **When disease outcome is time-to-event?**

Model-based optimality for time-to-event outcome

When disease outcome is time-to-event, T :

- $TP_t(m) = P(M > m|T = t)$: true positive rate (incident sensitivity)
  $FP_t(m) = P(M > m|T > t)$: false positive rate (1 $-$ dynamic specificity)
  $ROC_t(p) = TP_t\{FP_t^{-1}(p)\}, 0 \leq p \leq 1$: time-dependent ROC function
  $AUC_t = \int I(0 \leq p \leq 1) ROC_t(p) dp$: time-dependent AUC

- Time-dependent likelihood ratio: $LR_t(z, w) = \frac{p(z,w|T=t)}{p(z,w|T>t)}$

  By the Bayes' Rule, $LR_t(z, w) = \lambda(t|z, w) \cdot \lambda(t)^{-1}$

  Thus, for fixed t, $LR_t(z, w)$ is a monotone function of the hazard $\lambda(t|z, w)$.

- Classification rule $LR_t(z, w) > c$ produces maximized ROC and AUC

  nonparametric $\Longrightarrow$ too messy! (or, survival tree methods?)

  model-based $\Longrightarrow$ which models? how to combine markers?

Model-based optimality

- Model 1 $\boldsymbol{\lambda(t|z,w) = \lambda_0(t) exp\{\eta(z,w)\}}$ *Cox(1972)*

  e.g. $\boldsymbol{\eta(z,w) = z^T\beta + w^T\gamma}$

- Model 2 $\boldsymbol{\lambda(t|z,w) = \lambda_0(t) exp\{\eta_t(z,w)\}}$ *Zucker and Karr (1990)*

  e.g. Time-varying coefficient model, $\boldsymbol{\eta_t(z,w) = z^T\beta(t) + w^T\gamma(t)}$

- If Model 1 holds,

  - Composite marker $M = \eta(Z,W)$ yields maximized ROC and AUC at each t

    thus, if $\eta(z,w) = z^T\beta + w^T\gamma \Longrightarrow Z^T\beta + W^T\gamma$ is optimal at each t

  - Conditional on W=w, $M_w = \eta(Z,w)$ yields maximized ROC and AUC at each t;

    thus, if $\eta(z,w) = z^T\beta + w^T\gamma \Longrightarrow$ at each t and given w, $Z^T\beta + w^T\gamma$ is an optimal composite marker

    $\Longleftrightarrow$ at each t, $Z^T\beta$ is optimal (constant shift) which is free of w.

- If Model 2 holds,

  - Composite marker $M_t = \eta_t(Z,W)$ yields maximized ROC and AUC at t

    thus, if $\eta_t(z,w) = z^T\beta(t) + w^T\gamma(t) \Longrightarrow Z^T\beta(t) + W^T\gamma(t)$ is optimal at t

  - Conditional on W=w, $M_{w,t} = \eta_t(Z,w)$ yields maximized ROC and AUC at t;

    thus, if $\eta_t(z,w) = z^T\beta(t) + w^T\gamma(t) \Longrightarrow$ at t and given w, $Z^T\beta(t) + w^T\gamma(t)$ is an optimal composite

    $\Longleftrightarrow$ at t, $Z^T\beta(t)$ is optimal (constant shift) which is free of w.

Optimality based on a global index

- Suppose (T, Z, W, M) , (T_1, Z_1, W_1, M_1) and (T_2, Z_2, W_2, M_2) are iid.
  M: a univariate marker
- A global index due to Heagerty and Zheng (2005) is
  $\pi(M) =$ concord. prob. = $P\{M_1 > M_2 \mid T_1 < T_2\}$
- $\mathcal{M}_L = \{g(Z, W)\}$ is the large collection of all real-valued functions of (Z, W);
  $\mathcal{M}_S = \{g(Z)\}$ is the smaller collection of all real-valued functions of Z.

Property. Under Model 1: $\lambda(t|z, w) = \lambda_0(t) exp\{\eta(z, w)\}$:
(i) $\eta(Z, W) = aug\ max_M \{\pi(M) : \ M \in \mathcal{M}_L\}$.
(ii) $\eta(Z, w) = aug\ max_M \{\pi_w(M) : \ M \in \mathcal{M}_S\}$.

e.g. with additional assumption $\eta(z, w) = z^T\beta + w^T\gamma$,
$Z^T\beta + w^T\gamma = aug\ max_M \{\pi_w(M) : \ M \in \mathcal{M}_S\}$ is optimal
$\Longleftrightarrow Z^T\beta$ is optimal (constant shift) which is free of w.

Extension to additive hazards models

All the properties hold if proportional hazards models are replaced by additive hazards model:

- Model $\mathbf{1}^*$ $\boldsymbol{\lambda(t|z,w) = \lambda_0(t) + \eta(z,w)}$
- Model $\mathbf{2}^*$ $\boldsymbol{\lambda(t|z,w) = \lambda_0(t) + \eta_t(z,w)}$

Landmark models for dynamic prediction

- **Use biomarkers at pre-specified time points:**

  $0 = s_0 < s_1 < ... < s_K < s_{K+1} = \infty$

- **LM-Model 1** $\lambda(t|z(s_k), w) = \lambda_0(t) \exp\{\eta(z(s_k), w)\},\ s_k \leq t < s_{k+1}$

  LM-Model 2 $\lambda(t|z, w) = \lambda_0(t) exp\{\eta_t(z, w)\},\ s_k \leq t < s_{k+1}$

- **Compared with time-dependent marker PHM:**

  - only need marker information at pre-specified time points
  - ideal for dynamic prediction; e.g., using marker info at age 60 to predict AD at age 60∼70; using marker info at age 70 to predict AD at age 70∼80

- **Optimality properties extend to LM-specific $\eta(Z(s_k), W)$.**

Landmark models: a new global index

Define the global index as the sum of the interval-specific indices:

$\vec{M} = (M(s_0), ..., M(s_K))$: vector of markers defined at $(s_0, ..., s_K)$

Interval-specific global index:

$$\pi(M(s_k)) = P\{M_1(s_k) > M_2(s_k), s_k \le T_1 \le s_{k+1} \mid T_1 < T_2\}, \quad k = 0, ..., K.$$

A new global index:

$$\pi(\vec{M}) = \sum_{k=1}^{K} \pi(M(s_k)).$$

Property. Under LM-Model 1

(i) $\eta(\vec{Z}, W) = aug\ max_M\{\pi(\vec{M}) : \ \vec{M} \in \vec{\mathcal{M}}_L\}$

(ii) $\eta(\vec{Z}, w) = aug\ max_M\{\pi_w(\vec{M}) : \ \vec{M} \in \vec{\mathcal{M}}_S\}$

e.g. with additional assumption $\eta(\vec{Z}, w) = z(s_k)^T\beta + w^T\gamma, \ \ s_k \le t < s_{k+1}$

$(Z(s_0)^T\beta, ..., Z(s_K)^T\beta) = aug\ max_M\{\pi_w(M) : \ M \in \mathcal{M}_S\}$,

which is free of w.

Estimation of unknown parameters/functions

- Estimation of unknown parameters/functions - abundant in literature
- Risk-set-based estimation of TP_t - Xu and O'Quigley (2000, JRSS-B), Heagerty & Zheng (Biometrics, 2005)
- Nonparametric estimation of FP_t - Heagerty et al. (Biometrics, 2000)
- Model-based estimation of TP_t and FP_t

  Song and Zhou (2008)

Application to BIOCARD data

The BIOCARD project at Johns Hopkins studies biomarkers in relation to cognitive decline among normal individuals for progression to Alzheimer's disease (AD).

N≈ 300
No. of disease incidences observed ≈ 60

- **Sampling**: Recruit disease-free individuals into the study
- **Longitudinal biomarkers**: Longitudinally collect cognition, imaging, CSF (Cerebrospinal fluid) since study entry.
- **Outcome**

  T = time from study entry to disease incidence (AD) ; or
  T = individual's age at disease incidence (AD)

Application to BIOCARD data: Approach 1

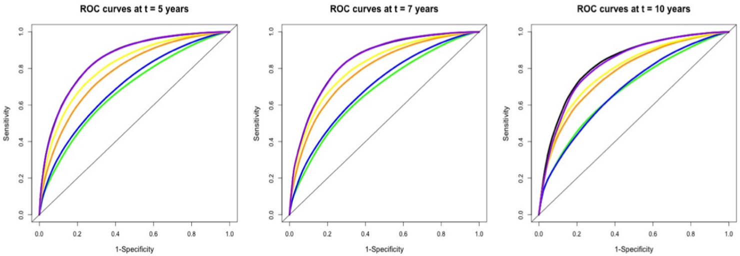
T = time from study entry to AD diagnosis

Use Model 1: $\boldsymbol{\lambda(t|z,w) = \lambda_0(t)exp\{z^T\beta + w^T\gamma\}}$

- **MRI-imaging** Right Hippocampus Volume, Right Entorhinal Cortex Thickness
- **CSF (Cerebrospinal fluid)** Abeta, Ptau, Ptau/Abeta
- **Cognition** Paired Associates Immediate, Digital Symbol Substitution
- **Genetics** APOE4
- **Demographics** Baseline age, Education

Application to BIOCARD data: Approach 1

T = time from study entry to AD diagnosis



Black: (Full model: demographics, ApoE-4, Cognition, MRI, CSF)

Purple: (Reduced model by removing Z with insignificant p-values: demographics, ApoE-4, Cognition, MRI, CSF)

*Note: Black and purple curves overlap at 5, 7 years

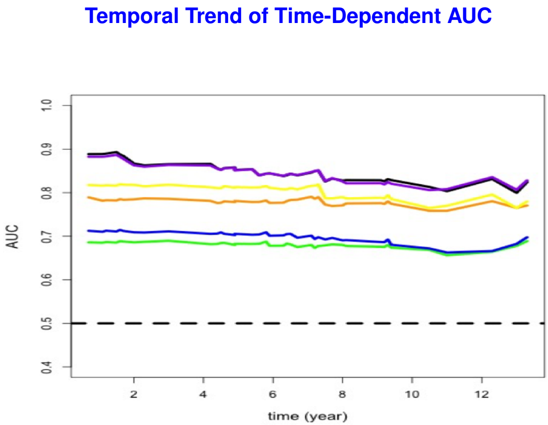
Yellow: (demographics, ApoE-4, Cognition, MRI)

Orange: (demographics, ApoE-4, Cognition)

Blue: (demographics, ApoE-4)

Green: demographics

Application to BIOCARD data: Approach 1



T = time from study entry to AD diagnosis

Application to BIOCARD data: Approach 2

T = individual's age at AD diagnosis

- **Keep in mind that longitudinal markers are typically collected after study entry**

  Model 1: $\lambda(t|z,w) = \lambda_0(t) exp\{z^T\beta + w^T\gamma\}$ does NOT directly work.

  Model 2: $\lambda(t|z,w) = \lambda_0(t) exp\{z^T\beta(t) + w^T\gamma(t)\}$ does NOT directly work.

- **Use landmark models instead:**

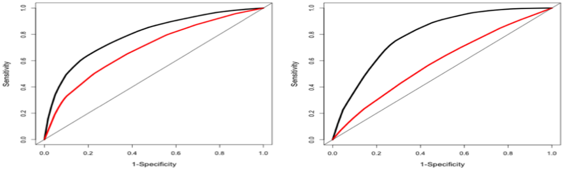
  LM-Model 1: $\lambda(t|z(s_k),w) = \lambda_0(t)\exp\{z(s_k)^T\beta + w^T\gamma\},\ s_k \le t < s_{k+1}$

  LM-Model 2: $\lambda(t|z,w) = \lambda_0(t) exp\{z(s_k)^T\beta(t) + w^T\gamma(t)\},\ s_k \le t < s_{k+1}$

  - A: age at study entry, C: time from study entry to censoring
    $X = \min(T, A+C)$, $\Delta = I(T < A+C)$.
  - LTRC risk set $\mathcal{R}(t) = \{i:\ A_i \le t \le X_i, i = 1, \ldots, n\}$
    Modify it as $\mathcal{R}_k^*(t) = \{i:\ A_i \le s_k \le t \le X_i, i = 1, \ldots, n\}$, $s_k \le t < s_{k+1}$
    so any subject in $\mathcal{R}_k^*(t)$ has marker information at s_k
  - $\mathcal{R}_k^*(t)$ has the RC risk set structure under indep censoring/truncation conditions
    at the price of 'loosing cases'

Application to BIOCARD data: Approach 2

Markers at age 60 predicts t=65 **Markers at age 70 predicts t=75**



T = age at AD diagnosis

black: education, ApoE-4, MRI

red: education, ApoE-4

Use LM-Model 1: $\lambda(t|z(s_k), w) = \lambda_0(t)\exp\{z(s_k)^T\beta + w^T\gamma\},\ s_k \leq t < s_{k+1}$

Markers at age 60 predicts t=65: red AUC=0.691, black AUC=0.792

Markers at age 70 predicts t=75: red AUC=0.588, black AUC=0.797

Conclusion

- **This talk:** Explore theory and methods for optimal combination of multiple markers for predicting the disease incidence *(Wang and Zhu, 2017, manuscript)*

  Ongoing research: AUC-guided survival tree approach for optimal prediction of time-to-disease *(Sun and Wang, 2017, manuscript)*

- **Other topics related to Biomarkers and AD research:**

  - Backward biomarker process

  Chan and Wang (2010, AAS; 2017, JASA), Cai, Wang & Chan (2017, Biometrics)

  - Change-point problems in backward biomarker process

- **Important applications to AD or other diseases**

  - Biomarker precision medicine with time-to-event outcome

  - Biomarker precision medicine with disease-free sampling

Thank you!