

# Predictive accuracy in survival analysis: The ROC curve and related measures

Petra Wolf

1. Dezember 2011

## Einleitung

### Receiver Operating Characteristic Kurven

- ROC Analyse in Fall-Kontroll-Studien

- ROC Analyse bei Überlebenszeiten

  - Heagerty et al. (2000)

  - Neuer Ansatz zur Schätzung

### C-Index

- Harrell's C

- Gönen und Heller's Concordance Probability

- Uno's C

- Vergleich der Methoden

### Integrated Discrimination Improvement

- IDI ohne Zensierungen

- IDI mit Zensierungen

### Ausblick

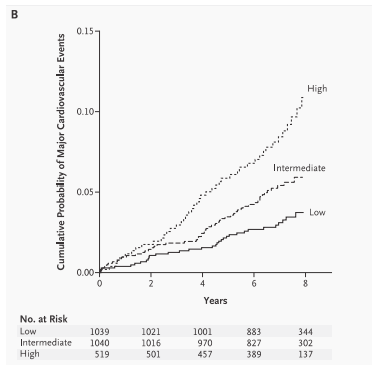
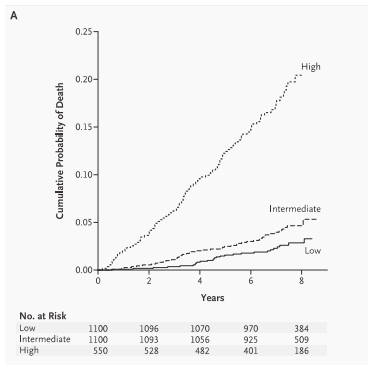
## Multiple Biomarkers for the Prediction of First Major Cardiovascular Events and Death - Wang (2006)

- Framingham Offspring Study
- Zielgrößen: Tod oder schwerwiegendes kardiovaskuläres Ereignis
- bekannte Risikofaktoren: Rauchen, Diabetes, Blutdruck, BMI, Cholesterin, HDL Cholesterin und Kreatininlevel
- zusätzlich 10 neue Biomarker

Statische Methoden:

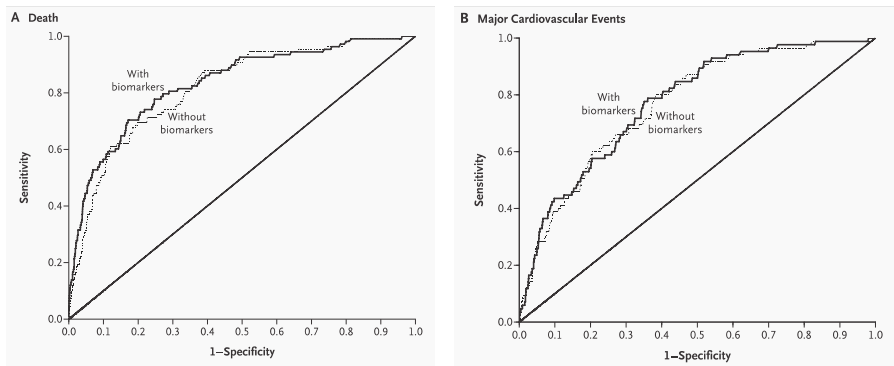
- Kaplan-Meier, Cox Modell
- C-Index
- ROC Kurven

# Multiple Biomarkers for the Prediction of First Major Cardiovascular Events and Death - Wang (2006)



Kaplan-Meier Curves of the Cumulative Probability of Death (Panel A) and Major Cardiovascular Events (Panel B), According to Category of Multimarker Score. Multimarker scores were classified as low, intermediate, or high.

# Multiple Biomarkers for the Prediction of First Major Cardiovascular Events and Death - Wang (2006)



Receiver-Operating-Characteristic Curves for Death (Panel A) and for Major Cardiovascular Events (Panel B) during 5-Year Follow-up. For each end point, curves are based on models of the prediction of risk with the use of conventional risk factors with or without biomarkers (multimarker score). Biomarkers for death were B-type natriuretic peptide, C-reactive protein, the urinary albumin-to-creatinine ratio, homocysteine, and renin. Biomarkers for major cardiovascular events were B-type natriuretic peptide and the urinary albumin-to-creatinine ratio.

# Gütemaße in der Überlebenszeitanalyse

- **Verallgemeinertes  $R^2$ :** Korn & Simon (1990), Nagelkerke (1991), Shemper & Henderson (2000), O'Quigley & Xu (2005), Schmid et al. (2010), ...
- **C-Index:** Harrell et al. (1996), Pencina and D'Agostino (2004), Gönen and Heller (2005), Uno (2011), ...
- **ROC Kurven:** Heagerty et al. (2000), Heagerty & Zheng (2005), ...

## Datensatz Kidtran

```
> library(KMsurv)
> data(kidtran)
> head(kidtran)
```

	obs	time	delta	gender	race	age
1	1	1	0	1	1	46
2	2	5	0	1	1	51
3	3	7	1	1	1	55
4	4	9	0	1	1	57
5	5	13	0	1	1	45
6	6	13	0	1	1	43

- Daten von Überlebenszeiten von 863 Nierentransplantationspatienten
- Transplantation durchgeführt am Ohio State University Transplantationszentrum in der Zeit 1982-1992
- Maximale Follow-up Zeit 9.47 Jahre

# ROC Analyse in Fall-Kontroll-Studien

- Fälle:  $D = 1$  /Kontrollen:  $D = 0$
- Marker  $Z$ , hoher Marker spricht für Krankheit
- Sensitivität/True Positive (TP):

$$TP(c) = P(Z > c | D = 1)$$

- Spezifität/True Negative (1-False Positive(1-FP)):

$$TN(c) = P(Z \leq c | D = 0)$$

- ROC Kurve: monoton wachsende Funktion in  $[0,1]$ :  
[[ $FP(c), TP(c)$ ],  $c \in (-\infty, \infty)$ ]



# ROC Analyse in Fall-Kontroll-Studien

- Vergleich verschiedener Marker über alle möglichen Fehlerkombinationen
- Vergleich der Sensitivität, bei festgesetzter Spezifität (z.B. TP bei FP=10%)
- AUC Interpretation: Wahrscheinlichkeit, dass bei zwei zufällig ausgewählten Personen, diejenige mit Ereignis einen höheren Marker aufweist, als die ohne Ereignis.

# ROC Analyse bei Überlebenszeiten

- statt 2 Kategorien (krank/gesund) Überlebenszeiten (mit Ereignis/zensiert) und vorhergesagte Überlebenszeiten
- welche Bedingung sollte erfüllt sein, dass das Maß sinnvoll ist?
- Was bedeutet „gut diskriminieren“ im Kontext der Überlebenszeiten?
- Wie definiert man *Fälle* und *Kontrollen* bei Ereigniszeiten?
- Der Status kann sich über den Verlauf der Zeit ändern (keine festen Gruppen)
- Problem zensierte Beobachtungen: sind zensierte Beobachtungen Fälle oder Kontrollen?

## Notation

- $Z_i$ : Marker für Person  $i$  bzw.  $Z_i = Z_i^* \beta$
- $T_i$ : Überlebenszeit für Person  $i$
- $C_i$ : (unabhängige) Zensierungszeit
- Beobachtet wird  $X_i = \min(T_i, C_i)$  und Zensierungsindikator  $\delta_i = \mathbb{1}(T_i \leq C_i)$
- $t$ : fester vorgegebener Zeitpunkt

# Definition von „Fällen“ und „Kontrollen“

## 1. kumulativ/dynamisch (k/d):

$$Sens^k = P(Z > z | T \leq t)$$

$$Spez^d = P(Z \leq z | T > t)$$

- kumulative Fälle: Ereignis tritt bis  $t$  ein
- dynamische Kontrollen: kein Ereignis bis  $t$
- zu jedem Zeitpunkt werden Personen in entweder Fälle oder Kontrollen aufgeteilt

Mit *kumulativ/dynamischer* ROC Kurve kann die prognostische Güte eines zu Baseline gemessenen Markers bis zur Zeit  $t$  beurteilt werden

## Definition von „Fällen“ und „Kontrollen“

### 2. **inzident/dynamisch (i/d):**

$$\text{Sens}^i = P(Z > z | T = t)$$

$$\text{Spez}^d = P(Z \leq z | T > t)$$

- inzidente Fälle: Ereignis tritt zum Zeitpunkt  $t$  ein
- dynamische Kontrollen: kein Ereignis bis  $t$
- mit der Zeit abnehmende Menge an Personen, die in die Analyse eingehen.
- zeitvariierende Marker möglich

*inzident/dynamische* ROC Kurve zeigt die Fähigkeit eines Markers zwischen Fällen ( $T = t$ ) und Kontrollen ( $T > t$ ) innerhalb einer Risikomenge  $T \geq t$  zu unterscheiden.

## Ansatz Heagerty et al. (2000)

Satz von Bayes:

$$\text{Sens}(z, t) = P(Z > z | T \leq t) = \frac{P(T \leq t | Z > z) \cdot P(Z > z)}{P(T \leq t)}$$

$$\text{Spez}(z, t) = P(Z \leq z | T > t) = \frac{P(T > t | Z \leq z) \cdot P(Z \leq z)}{P(T > t)}$$

## Ansatz Heagerty et al. (2000): Kaplan-Meier

Einfache Schätzung von Sensitivität und Spezifität für Zeit  $t$  über Kaplan-Meier Schätzer und die empirische Verteilungsfunktion des Markers  $Z$ :

$$\hat{P}_{KM}(Z > z | t \leq t) = \frac{\{1 - \hat{S}_{KM}(t|Z > z)\} \cdot \{1 - \hat{F}_Z(z)\}}{1 - \hat{S}_{KM}(t)}$$

$$\hat{P}_{KM}(Z \leq z | T > t) = \frac{\hat{S}_{KM}(t|Z \leq z) \cdot \hat{F}_Z(z)}{\hat{S}_{KM}(t)}$$

mit  $\hat{F}_Z(z) = \frac{1}{n} \sum I(Z_i \leq z)$

Problem:

Es gilt hier **nicht** der Satz der totalen Wahrscheinlichkeit:

$$P(T \leq t) \neq P(T \leq t | Z \leq z) \cdot P(Z \leq z) + P(T \leq t | Z > z) \cdot P(Z > z)$$

## Satz der totalen Wahrscheinlichkeit bei Ereigniszeiten

$$P(T \leq t) \neq P(T \leq t|Z \leq z) \cdot P(Z \leq z) + P(T \leq t|Z > z) \cdot P(Z > z)$$

T	Total			$Z \leq z$			$Z > z$		
	n	E	S(t)	n	E	S(t)	n	E	S(t)
1	10	1	$\frac{9}{10}$	6	1	$\frac{5}{6}$	4	0	1
2	8	1	$\frac{9}{10} \cdot \frac{7}{8}$	4	0	$\frac{5}{6}$	4	1	$\frac{3}{4}$
...	...	...	...	...	...	...	...	...	...

$$P(T \leq 1) = \frac{9}{10} = \frac{5}{6} \cdot \frac{6}{10} + 1 \cdot \frac{4}{10} = \frac{9}{10}$$

$$P(T \leq 2) = \frac{9}{10} \cdot \frac{7}{8} = \frac{63}{80} \neq \frac{5}{6} \cdot \frac{6}{10} + \frac{3}{4} \cdot \frac{4}{10} = \frac{64}{80}$$



## Ansatz Heagerty et al. (2000): NNE

Bei zensierten Beobachtungen kann eine ROC Kurve über einen Schätzer der bivariaten Verteilungsfunktion von  $[Z, T]$  hergeleitet werden:

$$S(z, t) = P(Z > z, T > t)$$

$$S(z, t) = \int_c^\infty S(t|Z = u) dF_Z(u)$$

mit  $F_Z(u)$ : Verteilungsfunktion von  $Z$

Akritas(1994):

$$\hat{S}_{\lambda_n}(z, t) = \frac{1}{n} \sum_i \hat{S}_{\lambda_n}(t|Z = Z_i) I(Z_i > z)$$

mit  $\hat{S}_{\lambda_n}(t|Z = Z_i)$  einem geeigneten Schätzer der bedingten Survivalfunktion, charakterisiert durch einen Glättungsparameter  $\lambda_n$

## Ansatz Heagerty et al. (2000): NNE

$$\hat{P}_{\lambda_n}(Z > z | T \leq t) = \frac{\{1 - \hat{F}_Z(z)\} - \hat{S}_{\lambda_n}(z, t)}{1 - \hat{S}_{\lambda_n}(t)}$$

$$\hat{P}_{\lambda_n}(Z \leq z | T > t) = 1 - \frac{\hat{S}_{\lambda_n}(z, t)}{\hat{S}_{\lambda_n}(t)}$$

mit  $\hat{S}_{\lambda_n}(t) = \hat{S}_{\lambda_n}(-\infty, t)$ .

- Für NNE  $\lambda_n = o(n^{-1/3})$  ausreichend für schwache Konsistenz
- Bootstrap kann für Inferenz genutzt werden

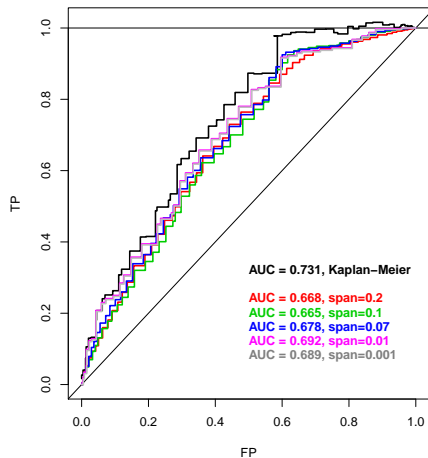
## Ansatz Heagerty et al. (2000)

Probleme der Schätzungen:

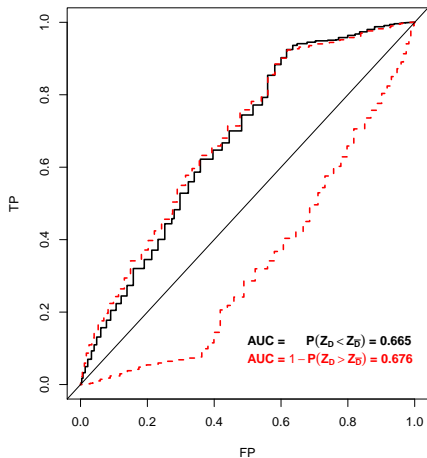
1. einfache Schätzung über **Kaplan-Meier**: führt zu unplausiblen Ergebnissen: ROC Kurve nicht monoton und hält die natürlichen Kriterien  $0 \leq \text{Sens}(z, t), \text{Spez}(z, t) \leq 1$  nicht ein
2. verbesserter Schätzer **Nächster Nachbar Schätzer** für  $S(z, t)$ : hält die Monotonie für eine ROC Kurve ein, aber Wahl des Glättungsparameters ist nicht klar und das Ergebnis ist nicht symmetrisch:  $P(Z_D > Z_{\bar{D}}) \neq 1 - P(Z_D < Z_{\bar{D}})$

# Ansatz Heagerty et al. (2000)

ROC (9 years)



ROC (9 years)



## Neuer Ansatz zur Schätzung I

Schätzung der erwarteten Anzahl an Ereignissen bis zum Zeitpunkt  $t$  ( $e_0(z, t)$  und  $e_1(z, t)$ ):

$$Z \leq z : e_0(z, t) = P(T \leq t | Z \leq z) \cdot n_0(z)$$

$$Z > z : e_1(z, t) = P(T \leq t | Z > z) \cdot n_1(z)$$

mit  $n_0(z)$  und  $n_1(z)$  # Beobachtungen mit  $Z \leq z$  bzw.  $Z > z$   
 $\Rightarrow$

$$Sens(z, t) = \frac{e_1(z, t)}{e_0(z, t) + e_1(z, t)}$$

$$Spez(z, t) = \frac{n_0(z) - e_0(z, t)}{n - (e_0(z, t) + e_1(z, t))}$$

## Neuer Ansatz zur Schätzung II

Schätzung der erwarteten Anzahl an Ereignissen über:

1. Anzahl beobachteter Ereignisse
2. Erwartete Ereignisse unter den zensierten Beobachtungen (Zensierungszeitpunkt  $c$ , Ereigniszeiten  $t_{k-1}$  und  $t_k$ , mit  $t_{k-1} < c < t_k$ ):

$$\begin{aligned}
 e &= P(T \leq t | T > c) \\
 &= P(c < T \leq t_k | T > c) + P(t_k < T \leq t | T > c) \\
 &= P(c < T \leq t_k) + \frac{S(t_k) - S(t)}{S(c)}
 \end{aligned}$$

Schätzung von  $P(c < T \leq t_k)$ 

1. Berechnung über Poisson-Prozess:  $P(c < T \leq t_k) \approx (t_k - c)\lambda_{k-1}$
2. Vereinfachte Berechnung unter der Annahme, das Ereignis hat in der Mitte zwischen  $t_{k-1}$  und  $t_k$  statt gefunden:  

$$P(c < T \leq t_k) \approx \frac{t_k + t_{k-1}}{2} \lambda_{k-1}$$
3. Stärkste Vereinfachung: Intervall  $[t_{k-1}; t_k]$  ist klein  $\Rightarrow$   
 Wahrscheinlichkeit ist nahezu gleich, ob zu Beginn der Intervalls zensiert oder am Ende des Intervalls:

$$\begin{aligned}
 e = P(T \leq t | T > c) &= \frac{S(c) - S(t_k)}{S(c)} + \frac{S(t_k) - S(t)}{S(c)} \\
 &= 1 - \frac{S(t)}{S(c)} \\
 &= 1 - \exp(-(\Lambda(t) - \Lambda(c)))
 \end{aligned}$$

## Neuer Ansatz zur Schätzung

Es besteht auch bei diesem Schätzer ein Problem mit der Monotonie der ROC Kurve. Die Bedingung  $0 \leq \text{Sens}(z, t), \text{Spez}(z, t) \leq 1$  wird aber nun eingehalten.

Das Problem der Monotonie kann nun aber sehr einfach mit Hilfe der isotonen Regression gelöst werden:

PAVA (**P**ooling **A**djacent **V**iolater **A**lgorithm)



## Isotone Regression

Gegeben:  $x_1 \leq \dots \leq x_n$  und  $f(x_i) = y_i$

Formelles Problem:

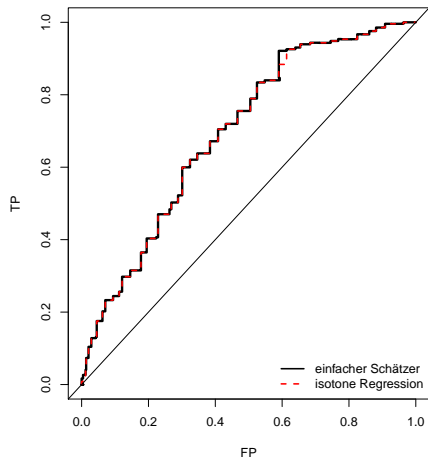
$$\hat{\beta}^{iso} = \underset{\beta \in \mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^n (y_i - \beta_i)^2 \text{ mit } \beta_1 \leq \dots \leq \beta_n$$

Lösung über pool adjacent violators algorithm (PAVA):

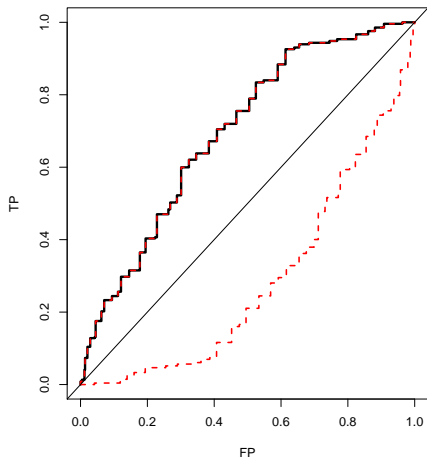
- 1 Beginne links mit  $y_1$ .
- 2 Gehe nach rechts, bis die erste Verletzung  $y_i > y_{i+1}$  auftritt.
- 3 Ersetze dieses Paar durch ihren Durchschnitt.
- 4 Wiederholen den Schritt weiter nach links, falls notwendig.
- 5 Fahre nach dem Schema fort bis  $y_n$ .

# Neuer Ansatz Beispiel

ROC (9 years)



ROC (9 years)



## Berechnung der AUC

- Aus Sensitivität und Spezifität kann über Trapezregel Fläche unter der ROC Kurve berechnen werden:  $AUC = P(Z_1 > Z_2 | T_1 \leq t, T_2 > t)$
- Problem: Schätzung der Varianz
- Ohne Zensierung kann die Varianz z.B. nach Hanley and McNeil (1983) oder DeLong et al. (1988) berechnet werden.
- Methoden lassen sich allerdings nicht einfach auf zensierte Daten übertragen.

## Noch offene Fragen

Problem bei zensierten ROC Analysen ist **Durchführung von Tests**:

Ist Marker 1 besser zur Vorhersage als Marker 2? Ist ein Risikoscore aus einem Modell mit einem zusätzlichen Einflussfaktor besser als ohne diesen? Verändert sich die Informationsgüte im Laufe der Zeit signifikant?

Zur Beantwortung dieser Fragestellungen ist eine Schätzung der Varianz notwendig. Bisher können diese Fragen nur mit Bootstrap-Verfahren beantwortet werden. Diese sind aber bei den hier gegebenen Fragestellungen sehr rechenintensiv, so dass sie für die tägliche Praxis keine adäquaten Verfahren darstellen.

## Harrell's C

Im Gegensatz zur  $AUC(t)$  nicht bestimmte Zeit  $t$  von Interesse, sondern globales Maß: wie gut stimmen vorhergesagte Überlebenszeiten mit den beobachteten Zeiten überein?

$$C_{Harrell} = P(Z_i > Z_j | T_i < T_j)$$

- Ereigniszeiten, nicht Status im Vordergrund
- Zwei Personen ohne Ereignis sind nicht vergleichbar
- Personen, die vor Ereignissen zensiert wurden sind nicht mit Personen mit entsprechenden Ereigniszeiten vergleichbar

# Harrell's C

$$\begin{aligned}
 C_{Harrell} &= P(\text{Paare sind konkordant} | \text{Paare sind vergleichbar}) \\
 &= P(T_i > T_j \text{ und } Z_i < Z_j | \text{Paare sind vergleichbar}) \\
 &\quad + P(T_i < T_j \text{ und } Z_i > Z_j | \text{Paare sind vergleichbar})
 \end{aligned}$$

- Vergleichbare Paare: Zwei Personen sind vergleichbar, wenn  $T_i > T_j$  oder  $T_i < T_j$  oder  $T_i = T_j$
- Konkordante Paare: Zwei Personen sind konkordant, wenn ( $T_i > T_j$  und  $Z_i < Z_j$ ) oder ( $T_i < T_j$  und  $Z_i > Z_j$ )
- mit steigendem Zensierungsanteil wird Schätzung verzerrter
- einfach zu berechnen
- rangbasiert

# Gönen und Heller's Concordance Probability

$$C_{\text{Gönen}} = P(T_i < T_j | Z_i > Z_j)$$

- Umgekehrte Definition von Harrell's C
- Vergleichbar mit Sensitivität vs. positiv prädiktivem Wert

$$\hat{C}_{n,\text{Gönen}}(\hat{\beta}) = \frac{2}{n(n-1)} \sum_{i < j} \sum_{i < j} \left\{ \frac{I(\hat{\beta}'x_{ji} < 0)}{1 + \exp(\hat{\beta}'x_{ji})} + \frac{I(\hat{\beta}'x_{ij} < 0)}{1 + \exp(\hat{\beta}'x_{ij})} \right\}$$

mit  $\hat{\beta}$ : partieller Likelihood Schätzer und  $x_{ij} = x_i - x_j$

- Über PH-Modell einfache Berechnung möglich
- Unabhängig vom Grad der Zensierungen

## Uno's C

$$C_{Uno} = P(Z_i > Z_j | T_i > T_j)$$

$$C_{\tau, Uno} = P(Z_i > Z_j | T_i < T_j, T < \tau)$$

nicht parametrische Schätzung über *inverse probability weighting*:

$$\hat{C}_{\tau} = \frac{\sum_{i=1}^n \sum_{j=1}^n \Delta_i \{\hat{G}_i(X_i)\}^{-2} I(X_i < X_j, X_i < \tau) I(\hat{\beta}' Z_i > \hat{\beta}' Z_j)}{\sum_{i=1}^n \sum_{j=1}^n \Delta_i \{\hat{G}(X_i)\}^{-2} I(X_i < X_j, X_i < \tau)}$$

mit

$I(\cdot)$  Indikatorfunktion

$\hat{G}(\cdot)$  Kaplan-Meier Schätzer für die Verteilung der Zensierungszeiten

$G(t) = P(C > t)$



# Vergleich Harrell's C - Gönen und Heller's C - Uno's C

```
#Harrell's C
> library(KMsurv)
> data(kidtran)
> library(survcomp)
> concordance.index(x = kidtran$age, surv.time = kidtran$time,
  surv.event = kidtran$delta, method = "noether")$c.index
[1] 0.6748258
```

```
#Uno's C
> library(survC1)
> datUno <- kidtran[, c(2, 3, 6)]
> Est.Cval(datUno, tau = max(datUno$time))$Dhat
[1] 0.6826814
```

```
#Gönen's C
> library(clinfun)
> fit <- coxph(Surv(time, delta) ~ age, kidtran)
> coxphCPE(fit)[1]
      CPE
0.6739328
```

## Zusammenhang C-Index - AUC

C-Index entspricht einer gewichteten Fläche unter der inzident/dynamischen ROC Kurve, mit Gewichten, die von der studienspezifischen Zensierungsverteilung abhängen:

$$C = \int_t AUC^{i/d}(t) \cdot w(t) dt$$

mit  $w(t) = 2 \cdot f(t) \cdot S(t)$

Bei Einschränkung der Zeit auf  $(0, \tau)$ :

$$C_\tau = \int_0^\tau AUC^{i/d}(t) \cdot w^\tau(t) dt$$

mit  $w^\tau(t) = w(t) / [1 - S^2(\tau)]$

# Vergleich der Methoden

**ohne Zensierungen** gilt:

$$\text{AUC} = \text{C-Index} = \text{Mann-Whitney U Statistik } P(X > Y)$$

**mit Zensierungen:**

Harrells-C-Index  $\neq$  Mann-Whitney U Statistik (Koziol & Jia, 2009)

$$\text{C-Index} \neq \text{AUC}$$

## Vergleich der Methoden: C-Index - AUC

Was misst der C-Index, was misst die AUC?

$$C - Index_{Harrell, Uno} = P(Z_1 > Z_2 | T_2 > T_1)$$

$$C - Index_{Gönen} = P(T_1 > T_2 | Z_2 > Z_1)$$

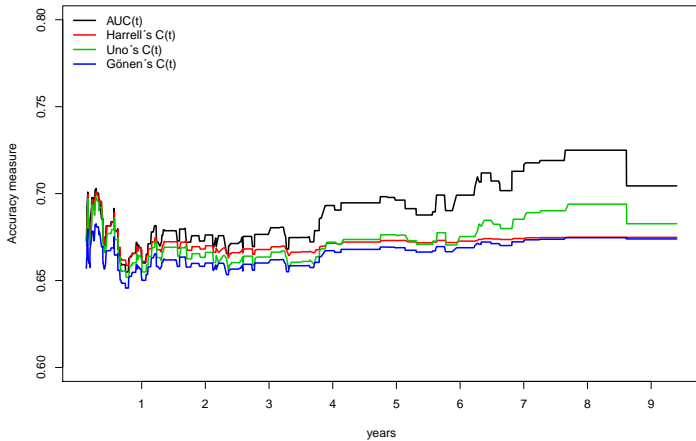
$$AUC = P(Z_1 > Z_2 | T_1 \leq t, T_2 > t)$$

interessierende Aspekte

beim C-Index: Höhe der Markers  $\leftrightarrow$  Überlebenszeit

bei der AUC: Status  $\rightarrow$  Höhe des Marker

# Vergleich der Maßzahlen



# Anwendung in der medizinischen Statistik

Verschiedene Aspekte:

1. Hat der Marker eine prognostische Fähigkeit? → AUC, C-Index
2. Wie kann ich anhand des Markers zwischen Hochrisikopatienten und „normalen“ Patienten unterscheiden? → Sensitivität, Spezifität
3. Ist die prognostische Güte über die Zeit gleichbleibend? → Verlauf der AUC über die Zeit

## IDI - Motivation

**Ausgangspunkt:** Bestehendes prognostisches Modell

**Frage:** kann ein neuer Marker (zusätzlich zu den bekannten Risikofaktoren) die Vorhersage verbessern? Wie nützlich ist der neue Marker?

Marker muss signifikanten Einfluss haben

Statistische Signifikanz aber nicht ausreichend, um klinische Relevanz und Verbesserung der Modellgüte zu gewährleisten

**Frage:** Wie kann am besten die Modellverbesserung beurteilt und der Zugewinn an Vorhersagegenauigkeit bestimmt werden?

# IDI - Motivation

- Klinischer Nutzen eines Überlebenszeitmodells: Kann das Modell Personen mit und ohne Ereignis unterscheiden?
- Beurteilung eines neuen Markers basierend auf der Fähigkeit die AUC zu verbessern:
  - ▶ sehr großer unabhängiger Zuwachs nötig, um die AUC zu verbessern
  - ▶ schwierige Interpretation
- neuer Ansatz: **I**ntegrated **D**iscrimination **I**mprovement
- Differenz zwischen Verbesserung der durchschnittlichen Sensitivität (*IS*) und der potentiellen Erhöhung der durchschnittlichen 1-Spezifität (*IP*)



# NRI

- Idee: Reklassifikation
- Personen mit und ohne Ereignis sollten getrennt betrachtet werden:
  - ▶ eine Klassifizierung in eine höhere Risikoklasse für Personen mit Ereignis bedeutet eine verbesserte Klassifikation
  - ▶ eine Klassifizierung in eine niedriger Risikoklasse beutete eine schlechtere Klassifikation
  - ▶ entsprechend umgekehrt für Personen ohne Ereignis
- Verbesserung in Reklassifikation: Net Reclassification Improvement (NRI)

$$\begin{aligned} NRI &= P(UP|D = 1) - P(DOWN|D = 1) \\ &+ P(DOWN|D = 0) - P(UP|D = 0) \end{aligned}$$

## IDI

Kategorisierung so fein, dass jede Person in eine eigene Risikoklasse fällt: jeder Zuwachs in vorhergesagten Wahrscheinlichkeiten für Personen mit Ereignis bedeutet Aufwärtsbewegung und jeder Abfall bedeutet Abwärtsbewegung.

- $IDI = (IS_{neu} - IS_{alt}) - (IP_{neu} - IP_{alt})$
- $\widehat{IDI} = (\hat{p}_{neu,events} - \hat{p}_{alt,events}) - (\hat{p}_{neu,nonevents} - \hat{p}_{alt,nonevents})$

# IDI - Berechnung mit zensierten Daten Chambless (2011)

Es gilt:

$$IS(t) - IP(t) = \frac{\text{Var}[S(t|Z)]}{S(t)[1 - S(t)]} =: R^2(t)$$

- entspricht dem Anteil der durch das Modell erklärten Varianz
- $0 \leq R^2(t) \leq 1$
- Schätzung einfach über Überlebenszeitmodell möglich:

$$\hat{R}^2(t) = \hat{IS}(t) - \hat{IP}(t) = \frac{\widehat{\text{Var}}(S(t|Z))}{\hat{S}(t) \cdot (1 - \hat{S}(t))}$$

$$IDI(t) = R_{neu}^2(t) - R_{alt}^2(t)$$

## Zusammenhang IDI - AUC

- Beide stellen eine korrigierte durchschnittliche Sensitivität dar: IDI korrigiert durch Abzug des unerwünschten Zuwachses in 1-Spezifität, AUC gewichtet die Sensitivität der beiden interessierenden Modelle durch entsprechende Ableitung der Spezifität:

$$AUC = \int_0^1 TP(u) \frac{d}{du} FP(u)$$

- AUC, IS, 1-IP: gewichtete Durchschnitte über alle möglichen Cutpunkte.
- AUC gewichtet Werte der Sensitivität bei kleinen Cutpunkten stärker → hohe Sensitivitäten werden stärker gewichtet.
- IS gewichtet alle Cutwerte oder Werte der Sensitivität gleich.

## Ausblick - alternative Möglichkeit IDI zu berechnen

- Berechnung des IDI nicht zu fester Zeit  $t$ , sondern zu beobachteter Zeit  $t_i$
- Zensierungen stellen somit kein Problem dar

$$\widehat{IDI}(t_i)_{neu} = (\hat{p}(t_i)_{neu,events} - \hat{p}(t_i)_{alt,events}) - (\hat{p}(t_i)_{neu,nonevents} - \hat{p}(t_i)_{alt,nonevents})$$

# Ausblick

- Inferenz der Maßzahlen:
  - ▶ bisher Bootstrap, sehr rechenintensiv
  - ▶ analytische Lösung oder verbesserte Resampling Verfahren?
- Modellvergleiche:
  - ▶ Einhalten des  $\alpha$ -Niveaus?
  - ▶ Power?
- Kann die  $AUC(t)$  zur Beurteilung von zeitabhängigen Kovariablen eingesetzt werden?

# Literatur

- Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. *Statistics in Medicine* 2011; **30**:22-38.
- Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika Trust* 2005; **92**:965-970.
- Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 1996; **15**:361-387.
- Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000; **56**:337-344.
- Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics* 2005; **61**:92-105.
- Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in Medicine* 2004; **23**:2109-2123.

- Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. (2008): Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in Medicine* 2008; **27**:157-72.
- Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine* 2011; **30**:1105-1117.
- Wang et al. Multiple Biomarkers for the Prediction of First Major Cardiovascular Events and Death. *The New England Journal of Medicine* 2006;**335** 2631-2639.
- Wolf, P, Schmidt G, Ulm K. The use of ROC for defining the validity of the prognostic index in censored data. *Statistics & Probability Letters* 2011; **81**:783-791.