Net time-dependent ROC curves: a solution for evaluating the accuracy of a marker to predict disease-related mortality

Marine Lorent, Magali Giral, Yohann Foucher

Department of Biostatistics EA 4275, Clinical Research and Subjective Measures in Health Sciences, University of Nantes, Nantes, France
Transplantation, Urology and Nephrology Institute (ITUN), Nantes Hospital and University, Inserm U1064, Nantes, France

August 22, 2012

Introduction (1)

Prognostic markers of all-cause mortality, essential to :

- identify subjects at high-risk of death
- optimize healthcare management

⇒ The capacity of a score to predict all-cause deaths is evaluated by using **time-dependent ROC curves** [1]

Limits

- an important part of the mortality not related to the chronic disease
- impossibility to identify excess deaths

⇒ Solution : distinguish between the expected mortality and the excess mortality, by using **additive net survival model**.

[1] Heagerty et al., Biometrics, 2000.

Introduction (2)

Objective

- Evaluate the capacity of a marker to predict disease-specific mortality, deaths for which medical specialists can act

Notations

- X : Random variable representing the prognostic marker
- x_j : Observation for the subject j
- n : Sample size, $j = 1, \ldots, n$
- T_{Ej} : Time to death related to the disease
- T_{Pj} : Time to expected death
- $T_j = min(T_{Ej}, T_{Pj})$: Time of the death
- C_j : Time of the last follow-up point (right censoring)
- $H(t)$, $H_E(t)$ and $H_P(t)$: Cumulative hazard functions of T, T_E and T_P at time t

Definition

New estimator : net time-dependent ROC curve

- represents the **net sensitivity** plotted against 1 - the **net specificity** for all the thresholds c of a marker X

  By defining a binary test at the cut-off c,

  - **Net sensitivity** : proportion of positive test ($X > c$) given that the disease-related death occurs before time t :

  $$se_t(c) = Pr(X > c | T_E \leq t)$$

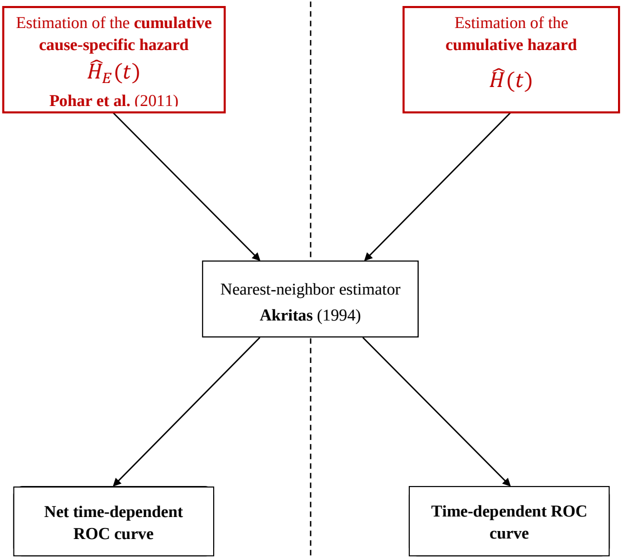
  - **Net specificity** : proportion of negative test ($X \leq c$) given that the disease-related death does not occur before time t :

  $$sp_t(c) = Pr(X \leq c | T_E > t)$$

⇒ **Question** : How estimate the net sensitivity and the net specificity ?

New estimator

Lorent et al. (submited) | **Heagerty et al.** (Biometrics, 2000)



Net sensitivity and net specificity

By adapting the approach of Heagerty (Biometrics, 2000)

the two probabilities can be developed :

- $se_t(c) = \{(1 - G_X(c)) - S_{X,E}(c,t)\}/\{1 - S_{X,E}(-\infty,t)\}$
- $sp_t(c) = 1 - \{S_{X,E}(c,t)/S_{X,E}(-\infty,t)\}$

Estimation of $S_{X,E}(c,t)$: bivariate survival function of X and T_E

⇒ implies to estimate $H_E(t|X = x_j)$, can be obtained from :

- $\hat{H}_E(t)$ [2]
- the calculation of the conditional at-risk and counting process (Use of Akritas estimator [3])
  - $Y_{jl}^{\pi}(t) = I(T_l > t, C_l > t, |\hat{G}_X(x_j) - \hat{G}_X(x_l)| < \pi)/S_{Pl}(t)$
  - $N_{jl}^{\pi}(t) = I(T_l \leq t, C_l \geq T_j, |\hat{G}_X(x_j) - \hat{G}_X(x_l)| < \pi)/S_{Pl}(t)$

[2] Pohar et al., Biometrics, 2011.
[3] Akritas, Annals of Statistics, 1994.

Net time-dependent ROC curve

Estimation of $se_t(c)$ and $sp_t(c)$

- the conditional cumulative excess hazard function is estimated by :

$$\hat{H}_E(t|X = x_j) = \int_0^t \frac{dN_{j.}^{\pi}(u)}{Y_{j.}^{\pi}(u)} - \int_0^t \frac{\sum_{l=1}^{n} Y_{jl}^{\pi}(u) dH_{Pj}(u)}{Y_{j.}^{\pi}(u)}$$

$\Rightarrow$ Allows to obtain :

- an estimation of $S_{X,E}(c, t)$
- an estimation of the net sensitivity and the net specificity for all the thresholds c

$\Rightarrow$ Representation of the net time-dependent ROC curve
Area under the curve = net AUC

Simulations (1)

Objective : validate the proposed estimator

3 different scenarios were considered

- Expected ages of death in general population were simulated to establish life tables
- Excess times-to-death were simulated

$\Rightarrow$ Distinction is possible between expected deaths and excess deaths

$\Rightarrow$ Calculation of the cause-specific AUC by censoring the expected deaths

$\Rightarrow$ Comparison between the traditional AUC, the cause-specific AUC and the net AUC for each sample

Simulations (2)

Censoring rate	Effective	All-cause AUCt	Cause-specific AUCt	Net AUC
≈ 0.30	n=100	0.769 (0.049)	0.955 (0.015)	0.891 (0.089)
	n=250	0.774 (0.035)	0.963 (0.008)	0.906 (0.061)
	n=500	0.773 (0.024)	0.963 (0.006)	0.912 (0.049)
	n=1000	0.772 (0.017)	0.964 (0.004)	0.910 (0.038)
≈ 0.50	n=100	0.756 (0.056)	0.945 (0.017)	0.872 (0.094)
	n=250	0.766 (0.034)	0.954 (0.010)	0.886 (0.067)
	n=500	0.764 (0.024)	0.953 (0.007)	0.888 (0.051)
	n=1000	0.765 (0.018)	0.955 (0.005)	0.889 (0.037)
≈ 0.70	n=100	0.747 (0.063)	0.940 (0.020)	0.839 (0.105)
	n=250	0.754 (0.043)	0.941 (0.014)	0.850 (0.073)
	n=500	0.752 (0.032)	0.944 (0.009)	0.846 (0.057)
	n=1000	0.750 (0.019)	0.943 (0.006)	0.843 (0.034)

Results

⇒ The net AUC provide significant correction of the all-cause AUC

⇒ The net AUC is closer to the cause-specific AUC

Kidney transplantation

Definition

- Disease associated = End stage renal disease
- Choice treatment = Kidney transplantation
- Possible trajectories after kidney graft :
  - return to dialysis
  - patient death related to the disease or not. Distinction is often impossible

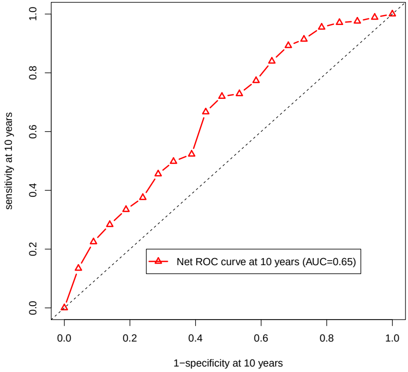
$\Rightarrow$ Endpoint studied in the following applications : patient death related to the disease

Kidney transplant recipients data (1)

Prognostic score of mortality of Hernandez [4]

- tested by using DIVAT cohort from Nantes Hospital (n=1230)
- 10 years prognostic, *net AUC* = 0,65, IC_{95}% = [0,56 - 0,72]

⇒ Difficult to validate the score in the prediction of excess deaths



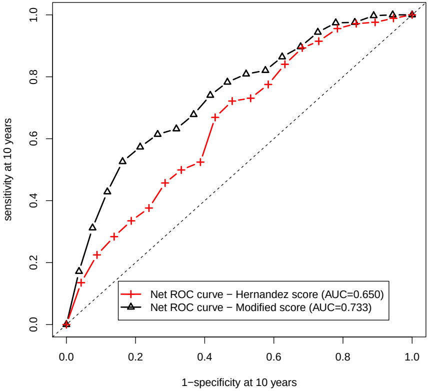
[4] Hernandez, Transplantation, 2009.

Kidney transplant recipients data (2)

Other prognostic score of mortality, created from DIVAT cohort

- 10 years prognostic, *net AUC* = 0,73, $IC_{95}\%$ = [0,64 - 0,80]

⇒ Capacity of this score to predict the disease-related mortality : acceptable



Interest of this method

Net time-dependent ROC curve

- useful when attribution of the deaths is impossible
- *net AUC* at time *t* is interpretable :

  for two patients randomly selected, probability that the patient with the higher value of the marker dies because of the disease, before the patient with the lower value.
- can be applied to others areas of medicine and biology
- implemented in an R package `ROCt` available at : http ://www.divat.fr/en/softwares/roct

Limits

Limits

- When a distinction is feasible between deaths related to the disease and those that are not $\Rightarrow$ competing risk model
- When all the observed mortality is related to the disease $\Rightarrow$ time-dependent ROC curve **OR** net time-dependent ROC curve
- The use of the proposed estimator in the presence of informative censoring $\Rightarrow$ noticeable effect on the results

Thank you for your attention !

Particularly thanks to :

- The EA-4275 Unit
- The DIVAT-Biostatistics Unit :
  - Magali Giral, Scientific council coordinator
  - Pascal Daguin, Computer specialist
  - Sandra Lefloch, Clinical research associate
  - Philippe Tessier, Health Economist
  - Yohann Foucher, Associate professor in Biostatistics
  - Etienne Dantan, Associate professor in Biostatistics
  - Katy Trébern-Launay, PhD Student in Biostatistics
  - Florence Gillaizeau, PhD Student in Biostatistics
  - Florent Leborgne, Trainee in Biostatistics