How to evaluate the prognostic capacity of surrogate makers ?

A methodological review in transplantation litterature

E. Dantan, C. Combescure, J. Ashton-Chess, M. Lorent, M. Giral and Y. Foucher

Yohann.Foucher@univ-nantes.fr
Biostatistics department (EA 4275), Medicine Faculty, University of Nantes, Nantes Hospital, ITUN et INSERM U643

The British Transplant Society - Glasgow 2012

UNIVERSITÉ DE NANTES

CENTRE HOSPITALIER UNIVERSITAIRE DE NANTES

itun institut transplantation urologie néphrologie INSERM - UMR 643

Outcome

1. Background

2. Method

3. Results

4. Conclusion

What is a surrogate marker ? [1]

Definition of surrogate marker

- *"Measurement providing early and accurate prediction of a clinical end point and the effects of treatment on this end point."*

The usual statistical analysis to demonstrate a predictor

- Kaplan-Meier estimator $\rightarrow$ high distance between the survival curves.
- Log-Rank statistic $\rightarrow$ small pvalue.
- Cox model $\rightarrow$ high value of the hazard ratio.

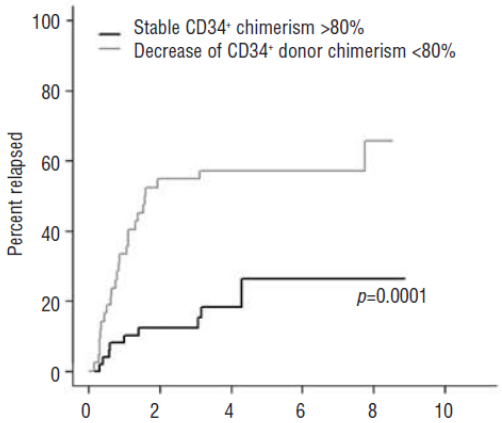
1. Fleming and DeMets. Surrogate End Points in Clinical Trials : Are We Being Misled ? Annals of Internal Medicine (1996)

Example

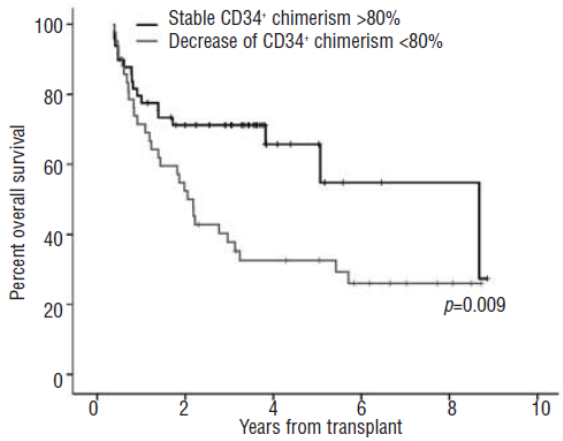
Monitoring of donor chimerism in sorted CD34⁺ peripheral blood cells allows the sensitive detection of imminent relapse after allogeneic stem cell transplantation

Martin Bornhäuser,[1] Uta Oelschlaegel,[1] Uwe Platzbecker,[1] Gesine Bug,[2] Karin Lutterbeck,[1] Michael G. Kiehl,[3] Johannes Schetelig,[1] Alexander Kiani,[1] Thomas Illmer,[1] Markus Schaich,[1] Catrin Theuser,[1] Brigitte Mohr,[1] Cornelia Brendel,[4] Axel A. Fauser,[3] Stefan Klein,[2] Hans Martin,[2] Gerhard Ehninger,[1] and Christian Thiede[1]

A Relapse



B Overall survival



In summary, the results of this study show for the first time that the serial chimerism analysis in CD34⁺ cells sorted out of peripheral blood is a sensitive technique to detect residual or reoccurring disease after allogeneic SCT. Its use allows the prediction of relapsing disease in most cases with CD34⁺ leukemia. The major advantage is the less frequent requirement of bone marrow aspiration and the possibility to start preemptive therapeutic interventions earlier. Since the method is time and labor-intensive, it needs further optimization before entering routine use.

Sensitivity and specificity

Problem

- A marker can be highly correlated with an event but with low predictive capacity.

The sensitivity and specificity for a prognostic up to time t

- SE_t is the proportion of patients that are classified as high risk, among all the patients who experience the event before time t.
- SP_t is the proportion of patients that are classified as low risk, among all the patients who do not experience the event before time t.

ROC curves

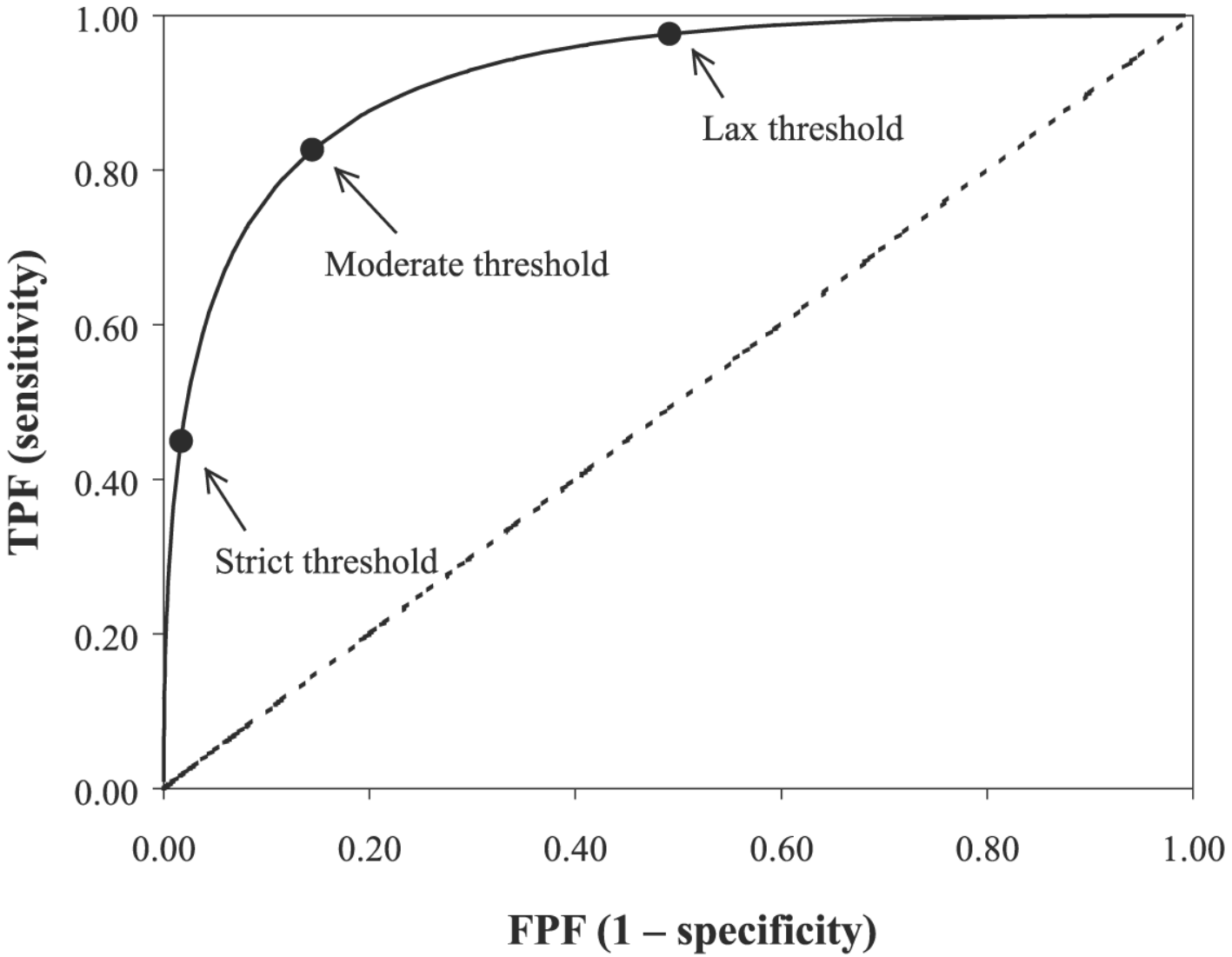
The ROC curve for a prognostic up to time t

- SE_t and SP_t can be computed when the surrogate is binary.
- If the surrogate is continuous : the SE_t and SP_t are computed for the possible thresholds.

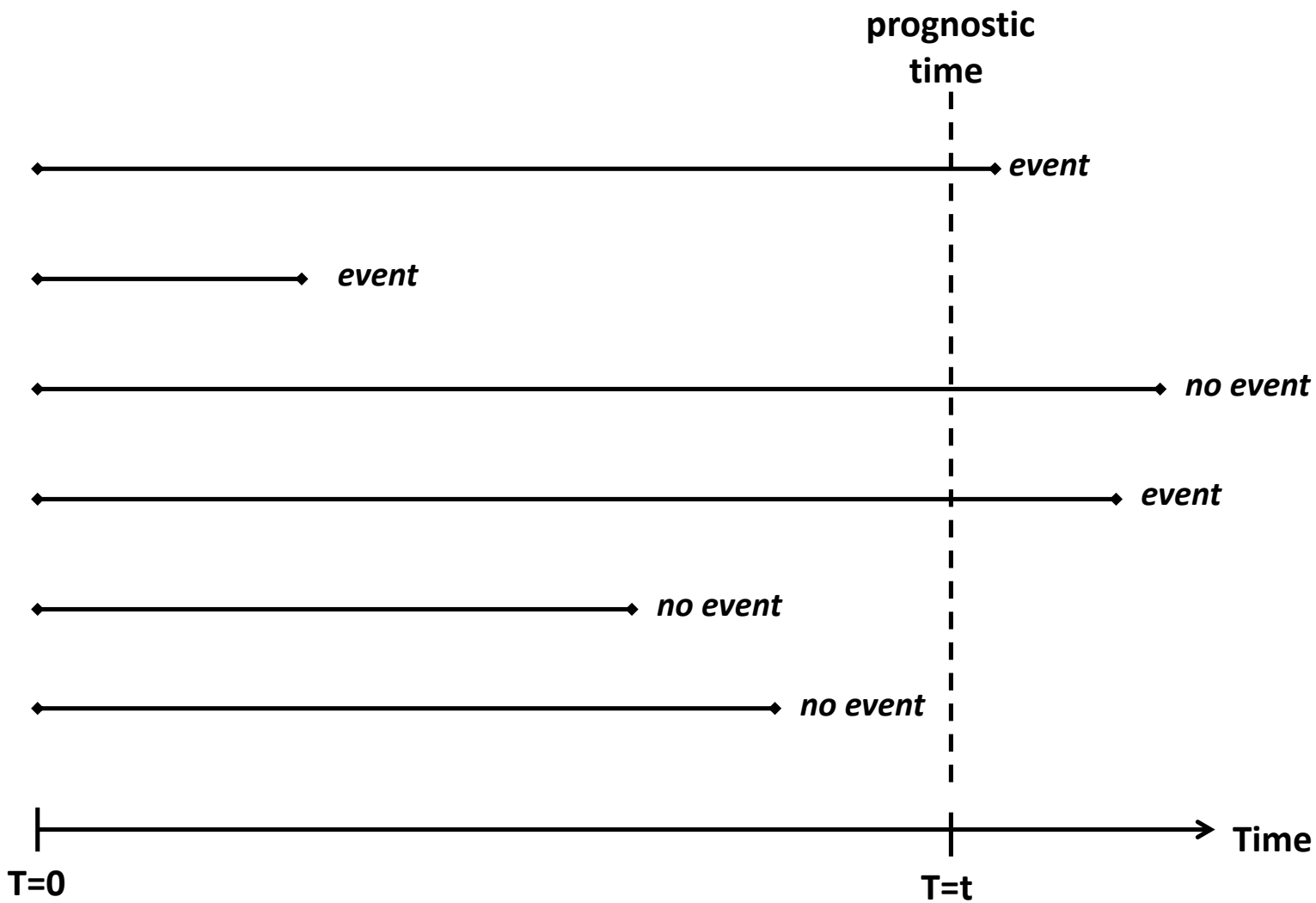
$\Downarrow$

threshold	high risk	low risk	SE_t	SP_t
c_1	$> c_1$	$\leq c_1$	1.00	0.00
c_2	$> c_2$	$\leq c_2$	0.98	0.04
c_3	$> c_3$	$\leq c_3$	0.94	0.06
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
c_k	$> c_k$	$\leq c_k$	0.00	1.00

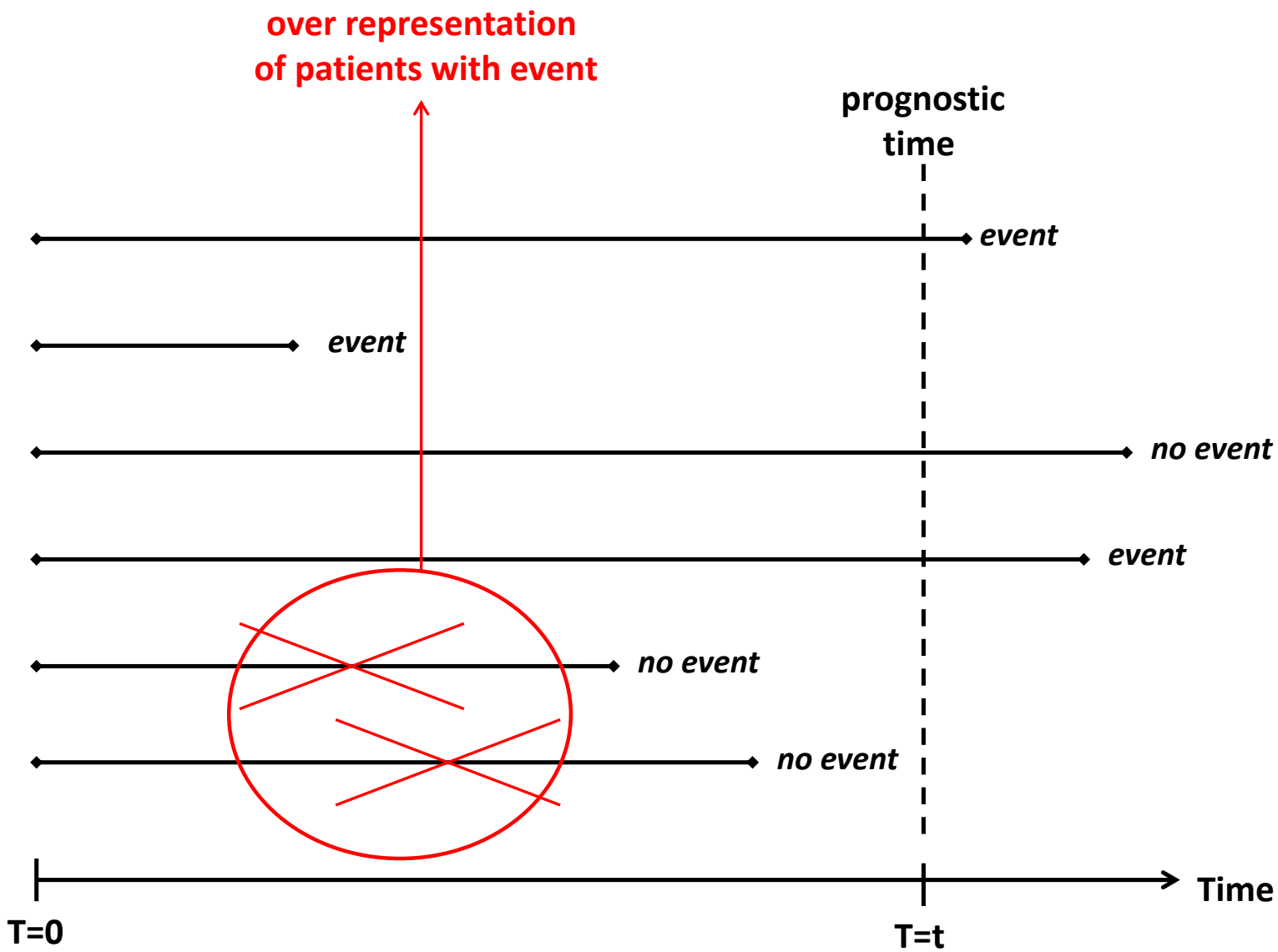
ROC curves


1.00
0.80
0.60
0.40
0.20
0.00
TPF (sensitivity)
Lax threshold
Moderate threshold
Strict threshold
0.00
0.20
0.40
0.60
0.80
1.00
FPF (1 – specificity)

Problem with censored data



Problem with censored data



ROC curve for censored data

- Heagerty et al.[2] proposed a correction for censored data :
  - Time-dependent sensitivity and specificity.
  - Time-dependent ROC curve.
- Method available in a lot of statistical software.
- Nevertheless, the method is not often used in practice....

$\Downarrow$

We proposed a simple methodology for computing time-dependent sensitivity and specificity by using (already) published survival curves.

2. Heagerty PJ, Lumley T, Pepe MS. Time dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 56 :337-344, 2000.

Outcome

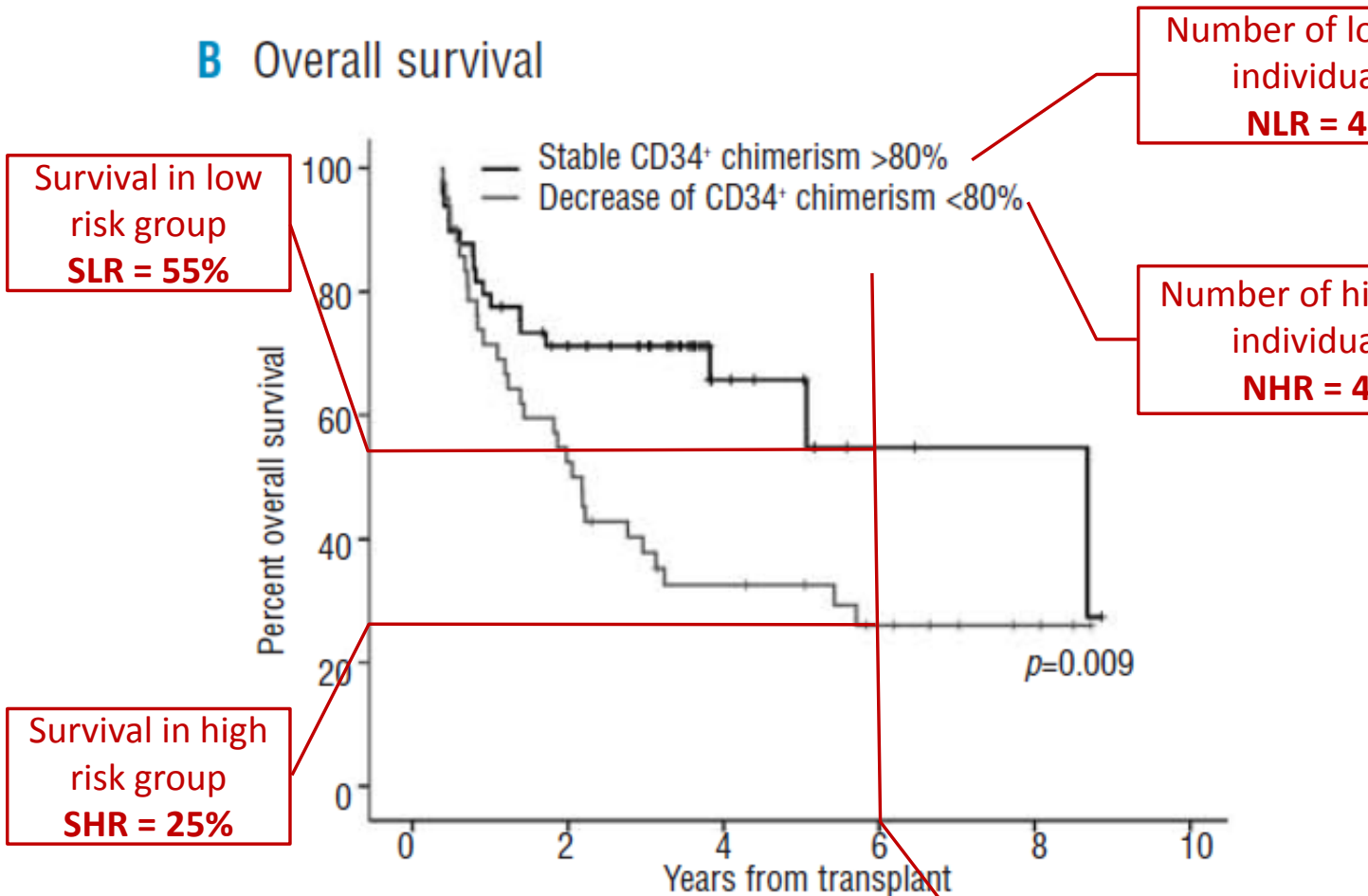
1. Background

2. Method

3. Results

4. Conclusion

Definitions



Number of low risk individuals
NLR = 47

Number of high risk individuals
NHR = 43

Survival in low risk group
SLR = 55%

Survival in high risk group
SHR = 25%

Prognostic time
t = 6

Time-dependent sensitivity [3]

- The proportion of patients that are correctly classified as high risk (HR), among all the patients who experience the event before t :

$$SE_t = \frac{(1 - SHR) \times NHR}{(1 - SHR) \times NHR + (1 - SLR) \times NLR}$$

- Application :

$$SE_{6\ years} = \frac{(1 - 0.25) \times 43}{(1 - 0.25) \times 43 + (1 - 0.55) \times 47} = 60\%$$

3. Foucher Y, Combescure C, Ashton-Chess J, Giral M. Prognostic Markers : Data Misinterpretation Often Leads to Overoptimistic Conclusions. Am J Transplant. 2012 Jan 6.

Time-dependent specificity

- The proportion of patients that are correctly classified as low risk (LR), among all the patients who do not experience the event before before time t :

$$SP_t = \frac{SLR \times NLR}{SHR \times NHR + SLR \times NLR}$$

- Application :

$$SP_{6\,years} = \frac{0.55 \times 47}{0.25 \times 43 + 0.55 \times 47} = 71\%$$

Relationship with predictive values

- Let S_t be the overall survival at time t, which is simply obtained by the weighted mean :

$$S_t = \frac{SHR \times NHR + SLR \times NLR}{NHR + NLR}$$

- The positive predictive value is the proportion of patients who experience the event before t, among all the HR patients :

$$PPV_t = \frac{SE_t \times (1 - S_t)}{SE_t \times (1 - S_t) + (1 - SP_t) \times S_t}$$

- The negative predictive value is the proportion of patients who not experience the event before t, among all the LR patients :

$$NPV_t = \frac{SP_t \times S_t}{SP_t \times S_t + (1 - SE_t) \times (1 - S_t)}$$

Warning

The PPV_t and NPV_t depend on the overall survival.

Outcome

1. Background

2. Method

3. Results

4. Conclusion

Six papers in transplantation

a. Bornhauser et al. Monitoring of donor chimerism in sorted CD34+ peripheral blood cells allows the sensitive detection of imminent relapse after allogeneic stem cell transplantation. Haematologica, 2009. 94(11) : p. 1613-7.

b. Xu et al. An effective model for predicting acute kidney injury after liver transplantation. Hepatobiliary Pancreat Dis Int. 9(3) : p. 259-63.

c. Hauser et al. Prediction of acute renal allograft rejection by urinary monokine induced by IFN-gamma (MIG). J Am Soc Nephrol, 2005. 16(6) : p. 1849-58.

d. Palmer et al. Innate immunity influences long-term outcomes after human lung transplant. Am J Respir Crit Care Med, 2005. 171(7) : p. 780-5.

e. Saracino et al. Early assessment of renal resistance index after kidney transplant can help predict long-term renal function. Nephrol Dial Transplant, 2006. 21(10) : p. 2916-20.

f. Susal et al. Posttransplant sCD30 as a Predictor of Kidney Graft Outcome. Transplantation, 2011. 91(12) : p. 1364-1369.

Six papers in transplantation

Endpoint	Biomarker	The High Risk group	Time	High Risk Group		Low Risk Group		p-value	SE_t	SP_t
				NHR	SHR	NLR	SLR			
Relapse after allogenic stem cell transplantation (a)	CD34+ donor chimerism	CD34+ < 80%	6 years	43	0.42	47	0.75	0.0001	0.68	0.66
			2 years		0.45		0.88		0.81	0.68
Death after allogenic stem cell transplantation (a)	CD34+ donor chimerism	CD34+ < 80%	6 years	43	0.25	47	0.55	0.0090	0.60	0.71
			2 years		0.55		0.72		0.60	0.59
Disease after allogenic stem cell transplantation (a)	CD34+ donor chimerism	CD34+ < 80%	6 years	43	0.25	47	0.57	0.0010	0.61	0.71
			2 years		0.30		0.72		0.70	0.72
Death after liver Transplantation (b)	Acute kidney injury (AKI)	AKI	18 months	33	0.70	69	0.93	0.0010	0.67	0.74
			6 months		0.73		0.99		0.93	0.74
Acute rejection after renal transplantation (c)	Urinary monokine (MIG)	MIG elevation	40 days	20	0.29	49	0.99	0.0018	0.97	0.89
			10 days		0.76		1.00		1.00	0.76
Acute rejection after lung transplantation (d)	TLR4 genotype	no 299/399 heterozygous	1000 days	149	0.30	18	0.53	0.0400	0.92	0.18
			250 days		0.35		0.68		0.94	0.19
Increasing in the creatinine > 50% after renal transplantation (e)	Intrarenal arterial resistance index (RI)	RI > 0.0635	80 months	39	0.68	37	0.90	0.0100	0.77	0.56
			20 months		0.90		1.00		1.00	0.51
Graft failure after renal transplantation (f)	soluble CD30 (sCD30)	sCD30 > 40	36 months	113	0.79	1149	0.92	0.0010	0.21	0.92
			12 months		0.89		0.96		0.21	0.92

Susal et al. Posttransplant sCD30 as a Predictor of Kidney Graft Outcome. Transplantation (2011)

Endpoint	Biomarker	The High Risk group	Time	High Risk Group		Low Risk Group		p-value	SE_t	SP_t
				NHR	SHR	NLR	SLR			
Graft failure after renal transplantation (f)	soluble CD30 (sCD30)	sCD30 > 40	36 months	113	0.79	1149	0.92	0.0010	0.21	0.92
			12 months		0.89		0.96		0.21	0.92

- Conclusion of the paper :
  - → *"... serum sCD30 on day 30 can be used to identify recipients who are at a high risk of subsequent graft loss."*
- But

$$SE_{36\ months} = \frac{(1-0.79)\times 113}{(1-0.79)\times 113 + (1-0.92)\times 1149} = 21\%$$

Warning

- Nearly 80% of kidney rejection at 36 months cases would not be detected.
- High false negative rate.

Endpoint	Biomarker	The High Risk group	Time	High Risk Group		Low Risk Group		p-value	SE_t	SP_t
				NHR	SHR	NLR	SLR			
Acute rejection after renal transplantation (c)	Urinary monokine (MIG)	MIG elevation	40 days	20	0.29	49	0.99	0.0018	0.97	0.89
			10 days		0.76		1.00		1.00	0.76

- Observed values in the paper :
  - $PPV_{40\ days} = 1 - SHR = 1 - 0.29 = 71\%$
  - $NPV_{40\ days} = SLR = 99\%$
- Corresponding sensitivity and specificity :
  - $SE_{40\ days} = 97\%$
  - $SP_{40\ days} = 89\%$
- But the overall survival is :

$$S_{40\ days} = \frac{0.29 \times 20 + 0.99 \times 49}{20 + 49} = 79\%$$

- What would be the predictive values for French recipients ?

- From DIVAT data bank, the overall probability of acute rejection is 9% : $S^{fr}_{40\ days} = 91\%$.

$$PPV^{fr}_{40\ days} = \frac{0.97 \times (1 - 0.91)}{0.97 \times (1 - 0.91) + (1 - 0.89) \times 0.91} = 46.6\%$$

$$NPV^{fr}_{40\ days} = \frac{0.89 \times 0.91}{0.89 \times 0.91 + (1 - 0.97) \times (1 - 0.91)} = 99.7\%$$

Warning

By using this test (MIG elevation) in France, 53% of patients classified in the HR group will have no acute rejection at 40 days (versus 29% in the original sample).

Outcome

1. Background

2. Method

3. Results

4. Conclusion

Conclusion

- When data are censored : the time-dependent ROC theory have to be used.
- If not : you should use the previous formulae.
  1. Compute time-dependent sensitivity and specificity, which do not depend on the event probability.
  2. Compute time-dependent predictive values according to your population.
- If more than two survival curves : you can merge groups together.