

Biostatistica per la ricerca e la pratica clinica



Validità degli studi predittivi e prognostici:
interpretazione e comunicazione

Ciro Gallo
Seconda Università di Napoli



Biostatistica per la ricerca e la pratica clinica



riflessione sul contributo essenziale della biostatistica nella scienza medica e nella pratica clinica.

taglio più culturale che metodologico

Ho chiesto ai miei amici oncologi: **Cosa vi aspettereste di sentire da me?**

- Nello sviluppo clinico di un farmaco quando vanno studiati i fattori predittivi?
- Qual è il rischio di interpretare erroneamente il significato predittivo?
- Quanto sono affidabili gli studi sul ruolo dei fattori predittivi (ad esempio le analisi di sottogruppi)?

Perché?

Modelli prognostici e pratica clinica



“The reason why physicians collect information about a patient is not simply wanting to know him better but wanting to use this information to do something effective for him”.

“A prognostic model should tell the clinician what to do, i.e. whom to treat how and whom not”

Windeler J. *Statist Med* 2000; 19: 425-30

Modelli prognostici e pratica clinica



Gli studi prognostici si fanno invece più spesso per altre ragioni:

- Migliorare la conoscenza dei processi di malattia
- Pianificare meglio gli studi randomizzati
- Migliorare il confronto di trattamenti in studi non randomizzati
- Informare meglio il paziente (anche se non si può intervenire)

Ogni anno sono pubblicati molti modelli prognostici, ma pochi sono sufficientemente validati e ancora meno trovano un loro spazio nella pratica clinica

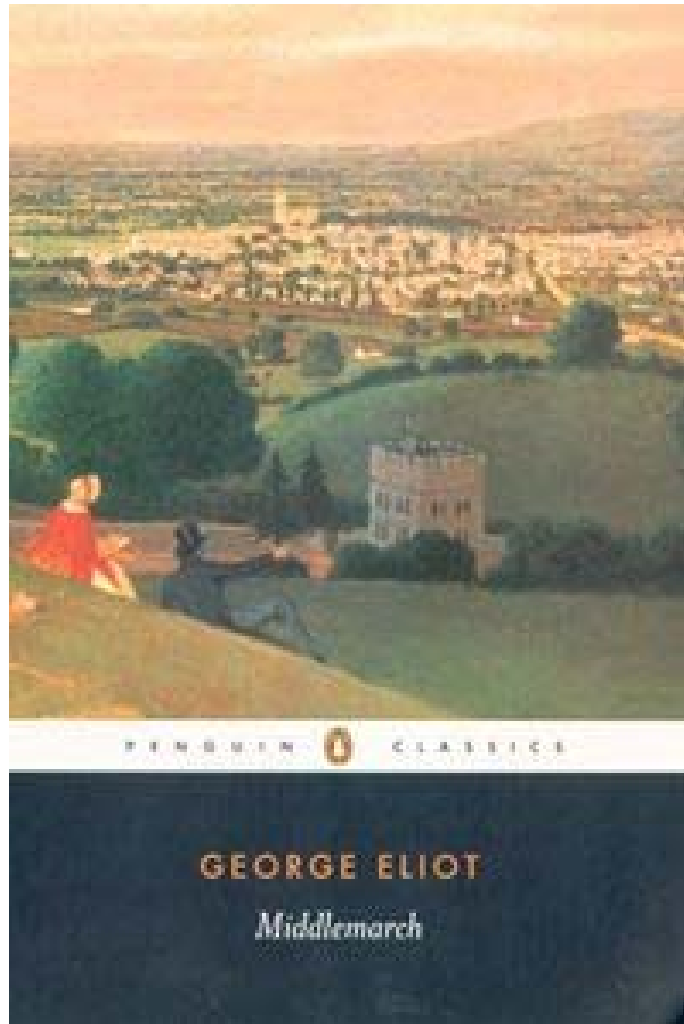
Perché?

Harry Hemingway

Prognosis research:

Why is Dr. Lydgate still waiting?

Journal of Clinical Epidemiology 59 (2006) 1229–1238



Hemingway H. Prognosis research: Why is Dr. Lydgate still waiting?

J Clin Epidemiol 2006;59:1229-1238



- Nella formazione del medico manca la valutazione della prognosi
- Non vi sono test sui principi e i metodi degli studi prognostici
- Nella formazione medica la diagnosi è considerata più importante della prognosi
- La Classificazione Internazionale delle Malattie è largamente basata sui processi patologici senza fare riferimento alla prognosi
- La stima, la comprensione e la comunicazione della prognosi sono raramente incluse nelle indicazioni di buona pratica clinica

Guidelines

REporting recommendations for tumour MARKer prognostic studies (REMARK)

LM McShane^{*,1}, DG Altman², W Sauerbrei³, SE Taube¹, M Gion⁴ and GM Clark⁵ for the Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics⁶

British Journal of Cancer (2005) 93, 387–391

Guidelines for Reporting Health Research: The EQUATOR Network's Survey of Guideline Authors

Iveta Simera^{*}, Douglas G. Altman, David Moher, Kenneth F. Schulz, John Hoey



PLoS Medicine June 2008 | Volume 5 | Issue 6 | e139

doi:10.1371/journal.pmed.0050139



- Insufficiente credibilità clinica
- Insufficiente dimostrazione dell'utilità del modello prognostico
 - ❑ accuratezza
 - ❑ generalizzabilità
 - ❑ efficacia



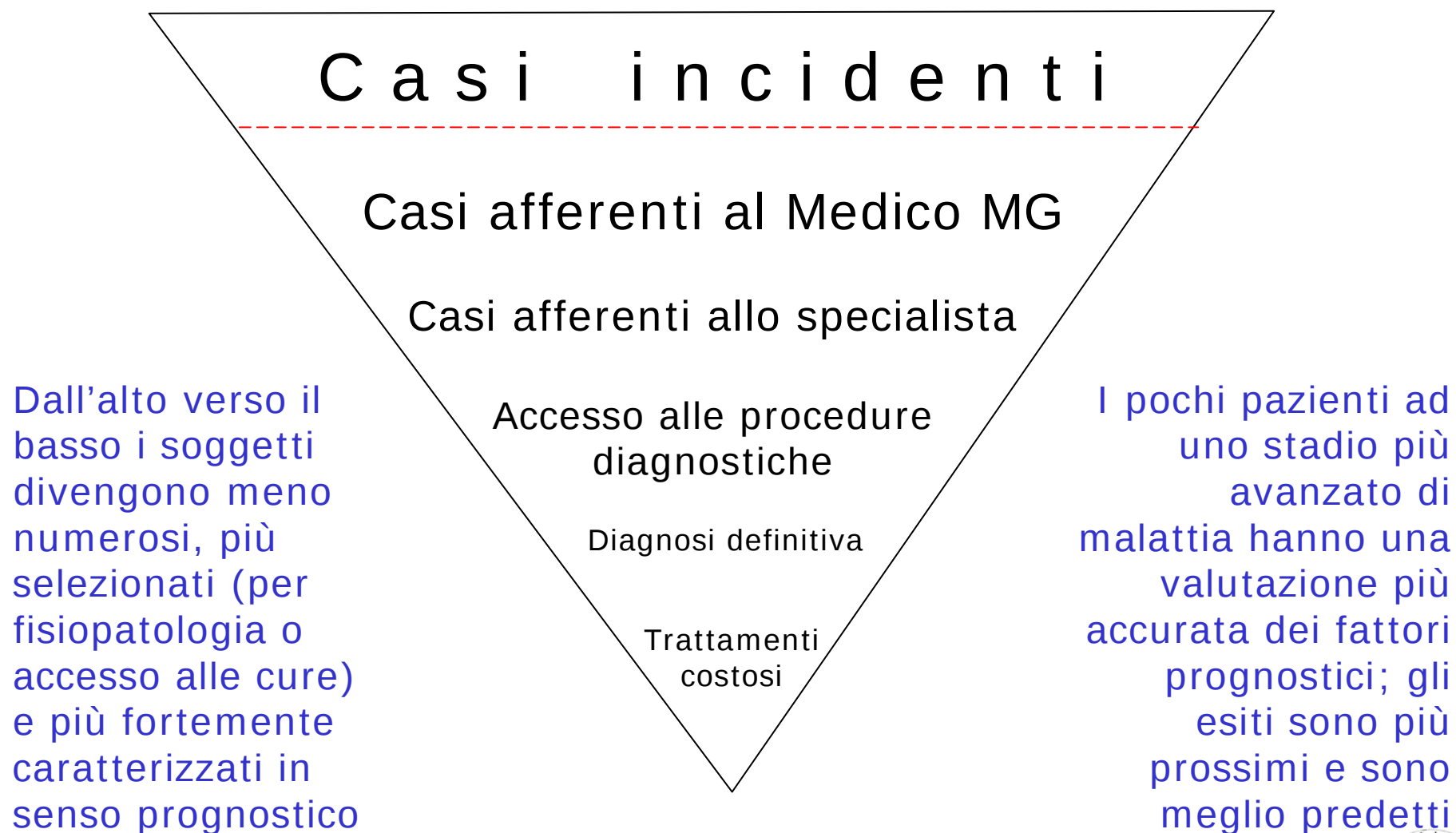
- Insufficiente credibilità clinica
- Insufficiente dimostrazione dell'utilità del modello prognostico
 - ❑ accuratezza
 - ❑ generalizzabilità
 - ❑ efficacia

La credibilità clinica dei modelli prognostici



- Dovrebbero essere inclusi nel modello tutti i pazienti clinicamente rilevanti
- Il modello dovrebbe basarsi su dati semplici da ottenere, di qualità affidabile, che non richiedano costi aggiuntivi e che possano essere ottenuti in tempi sufficientemente brevi da guidare la decisione clinica
- Si dovrebbero evitare i valori-soglia arbitrari per le variabili continue
- Il modello dovrebbe essere biologicamente e clinicamente plausibile (la diffidenza è maggiore per le 'scatole nere')
- Il modello dovrebbe essere facile da usare

La selezione dei casi



Almost all articles on cancer prognostic markers report statistically significant results

Prognostic markers
meta-analyses (n=340)

■ Negative 32 (9%)

Cancer prognostic
markers 2005 (n=1575)

■ Negative 66 (4%)

In the rare articles where no prognostic markers are presented as significant, authors often have other (non-prognostic) statistically significant analyses to show, they expand on the importance of non-significant trends, or defend the importance of the cancer marker with other arguments.

Eventually, totally 'negative' articles on prognostic cancer markers represent less than 1.5%



- Insufficiente credibilità clinica
- Insufficiente dimostrazione dell'utilità del modello prognostico
 - ❑ accuratezza
 - ❑ generalizzabilità
 - ❑ efficacia

L'accuratezza dei modelli prognostici



E' improbabile che un modello prognostico sia utile se non incrementa l'accuratezza prognostica dei medici che dovrebbero usarlo

Tipi di studi

- valutare il significato prognostico di una o più particolari caratteristiche
- sviluppare un modello prognostico in cui combinare insieme variabili rilevanti

L'accuratezza dei modelli prognostici



E' improbabile che un modello prognostico sia utile se non incrementa l'accuratezza prognostica dei medici che dovrebbero usarlo

Tipi di studi

- valutare il significato prognostico di una o più particolari caratteristiche (esplorativi)
 - Il nuovo fattore migliora significativamente la capacità prognostica dei fattori già noti?
 - Quali sono i più importanti fattori prognostici?

A chi credere?

Wild-Type *KRAS* Is Required for Panitumumab Efficacy in Patients With Metastatic Colorectal Cancer

J Clin Oncol 26:1626-1634. © 2008

Multivariate analysis showed that WT *KRAS* status was a predictor for OS in both the panitumumab (HR, 0.64; P .004) and BSC (HR, 0.68; P .007) arms.

K-ras Mutations and Benefit from Cetuximab in Advanced Colorectal Cancer

N Engl J Med 2008;359:1757-65.

In the group of patients receiving best supportive care alone, the mutation status of the *K-ras* gene was not significantly associated with overall survival (hazard ratio for death, 1.01; P = 0.97).



L'accuratezza dei modelli prognostici

Perché i risultati differiscono?

- Bassa numerosità
- Selezione automatica delle variabili
- Correzione per differenti covariate
- Differenti tecniche di misura delle variabili
- Differenti valori-soglia per le variabili continue
- Analisi di differenti sottoinsiemi di pazienti
- Differenti endpoint

Usefulness is determined by how well a model works in practice, not by how many zeros there are in the associated P -values

Altman DG Royston P. Statist Med 2000



Multiple Biomarkers for the Prediction of First Major Cardiovascular Events and Death

Thomas J. Wang, M.D., Philimon Gona, Ph.D., Martin G. Larson, Sc.D.,
Geoffrey H. Tofler, M.D., Daniel Levy, M.D., Christopher Newton-Cheh, M.D., M.P.H.,
Paul F. Jacques, D.Sc., Nader Rifai, Ph.D., Jacob Selhub, Ph.D., Sander J. Robins, M.D.,
Emelia J. Benjamin, M.D., Sc.M., Ralph B. D'Agostino, Ph.D.,
and Ramachandran S. Vasan, M.D.

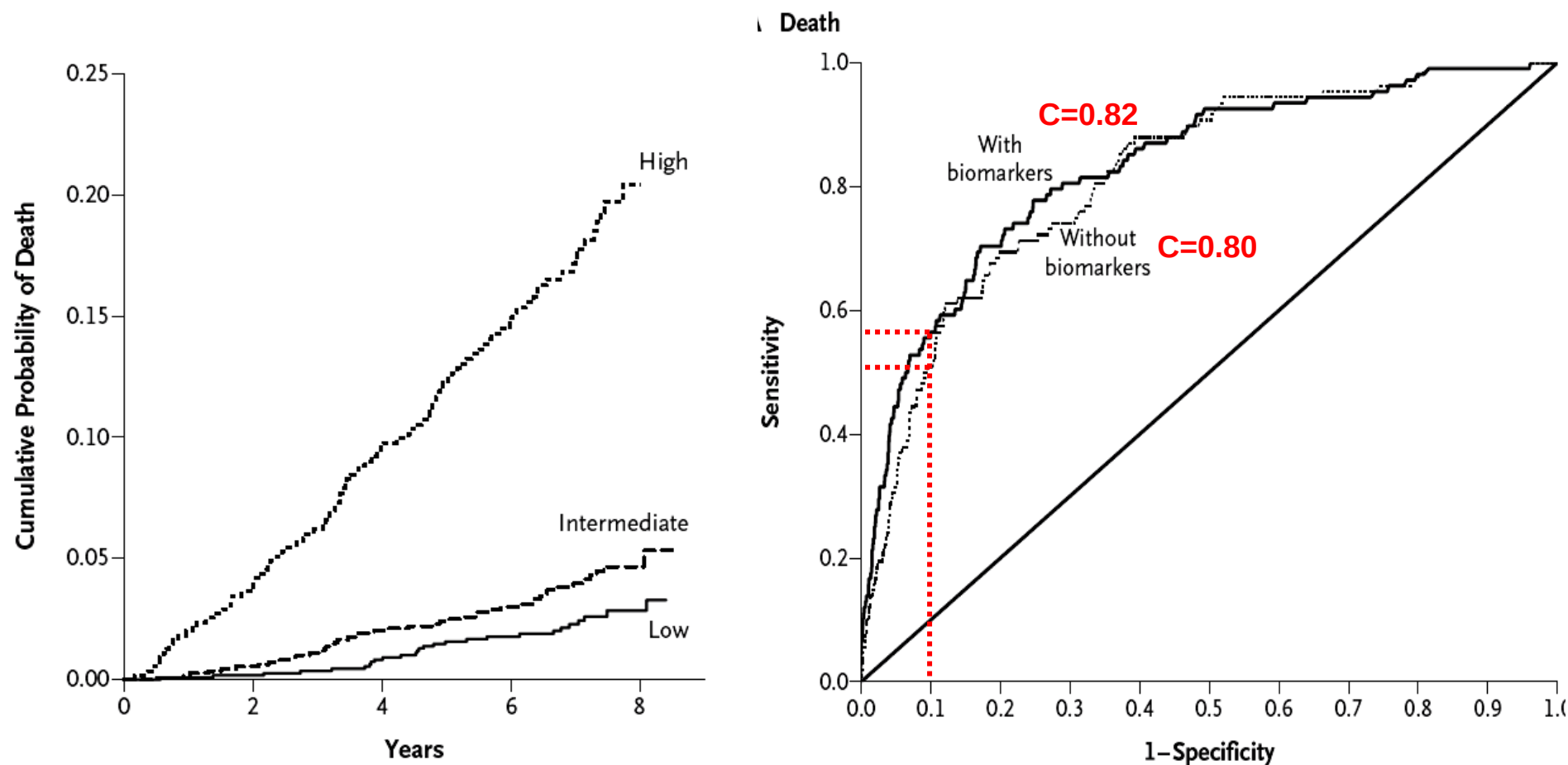
N Engl J Med 2006;355:2631-9

We measured 10 biomarkers in 3209 participants attending a routine examination cycle of the Framingham Heart Study.

Persons with “multimarker” scores (based on regression coefficients of significant biomarkers) in the highest quintile as compared with those with scores in the lowest two quintiles had elevated risks of death (adjusted hazard ratio, 4.08; $P < 0.001$).

However, the addition of multimarker scores to conventional risk factors resulted in only small increases in the ability to classify risk, as measured by the C statistic.

Multiple Biomarkers for the Prediction of First Major Cardiovascular Events and Death



this group of biomarkers makes a substantial contribution to the proportional-hazards model for predicting death from any cause, but it is of limited value for the risk stratification of individual patients.

L'accuratezza dei modelli prognostici



E' improbabile che un modello prognostico sia utile se non incrementa l'accuratezza prognostica dei medici che dovrebbero usarlo

Tipi di studi

- valutare il significato prognostico di una o più particolari caratteristiche
- sviluppare un modello prognostico in cui combinare insieme le variabili rilevanti per predire la probabilità di un dato evento
 - ❑ derivazione del modello
 - ❑ validazione del modello

L'accuratezza dei modelli prognostici



Derivazione del modello

- Sono da preferire i modelli che richiedono informazioni routinarie e affidabili. L'algoritmo dovrebbe essere semplice e facile da utilizzare
- Il principale problema è che la derivazione del modello è fortemente dipendente dai dati disponibili (split-sample, cross validation, leave-one-out, bootstrapping)

L'accuratezza dei modelli prognostici



Derivazione del modello

- Sono da preferire i modelli che richiedono informazioni routinarie e affidabili. L'algoritmo dovrebbe essere semplice e facile da utilizzare
- Il principale problema è che la derivazione del modello è fortemente dipendente dai dati disponibili

Validazione del modello

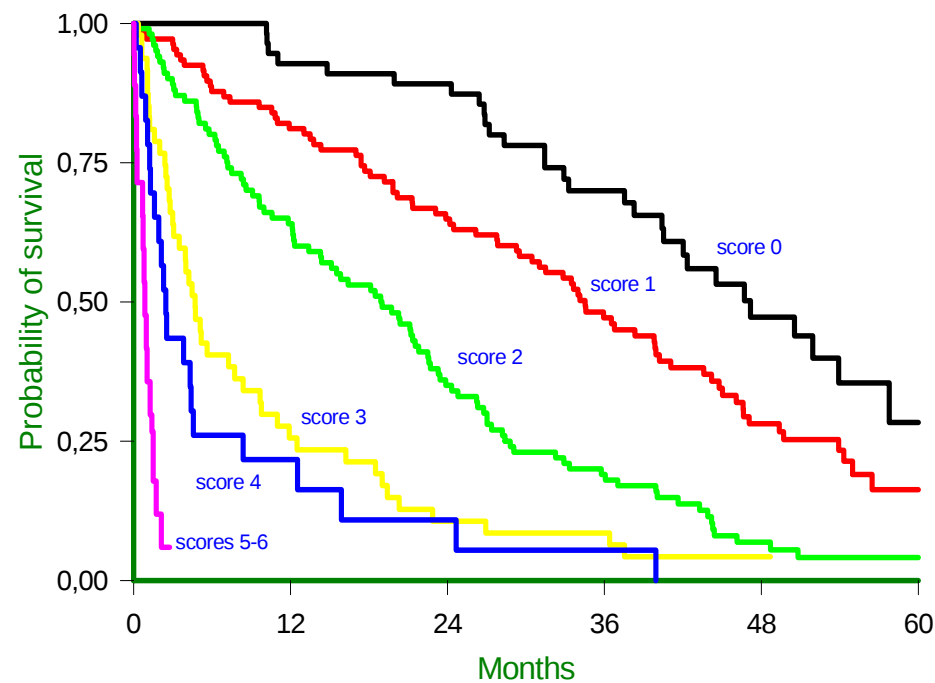
- Quanto è capace il modello di separare i soggetti che presentano l'esito da quelli che non lo hanno (Discriminazione)?
- Quanto concordano le probabilità predette con gli esiti realmente osservati (Calibrazione)?

The CLIP story (1)

A New Prognostic System for Hepatocellular Carcinoma: A Retrospective Study of 435 Patients

THE CANCER OF THE LIVER ITALIAN PROGRAM (CLIP) INVESTIGATORS*

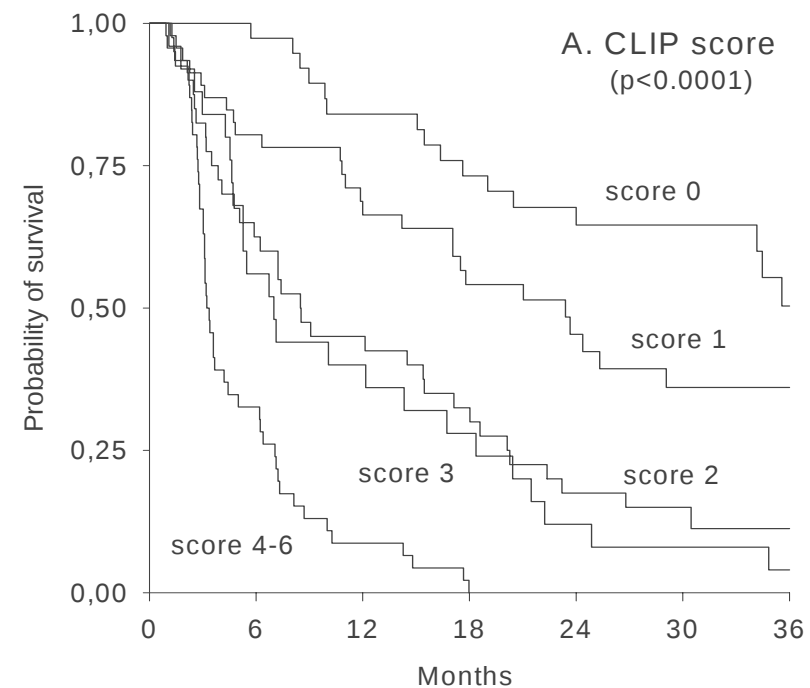
1998



Prospective Validation of the CLIP Score: A New Prognostic System for Patients With Cirrhosis and Hepatocellular Carcinoma

THE CANCER OF THE LIVER ITALIAN PROGRAM (CLIP) INVESTIGATORS*

2000





- Insufficiente credibilità clinica
- Insufficiente dimostrazione dell'utilità del modello prognostico
 - ❑ accuratezza
 - ❑ generalizzabilità
 - ❑ efficacia

La generalizzabilità dei modelli prognostici

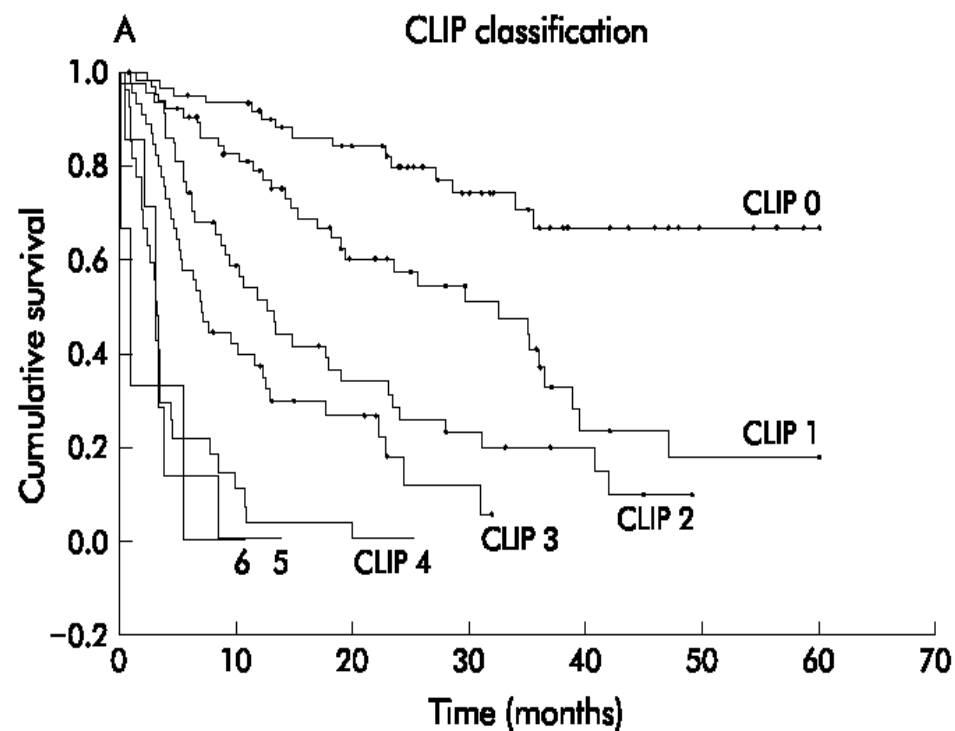


E' improbabile che un modello prognostico sia applicato se non se ne verifica l'utilità nel proprio contesto di pratica clinica

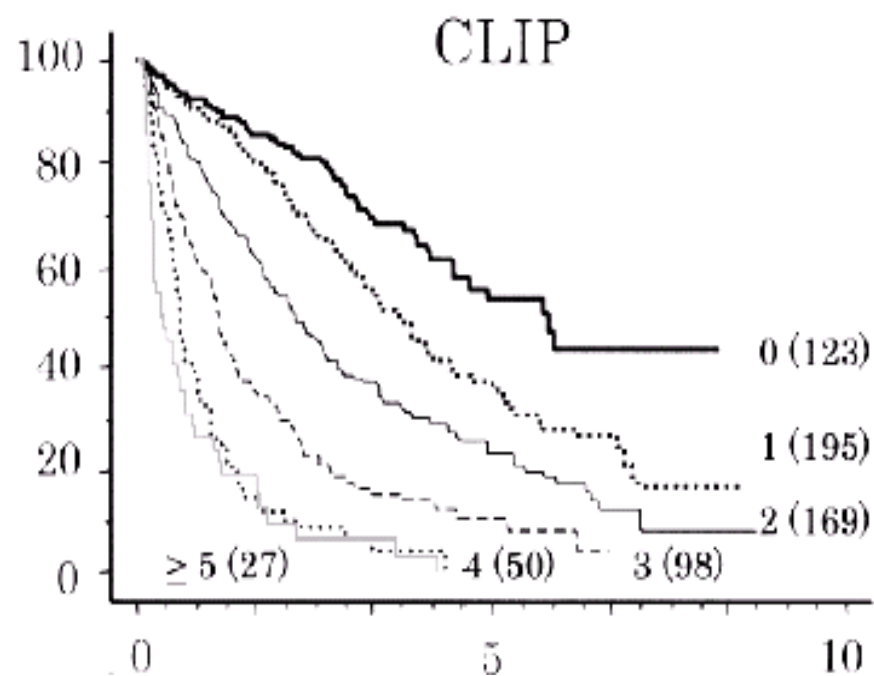
- Quanto è accurato il modello in altri soggetti appartenenti alla stessa popolazione?
- Quanto è accurato il modello in soggetti appartenenti a popolazioni diverse per:
 - ❑ localizzazione geografica
 - ❑ tempo
 - ❑ metodi di misura delle variabili
 - ❑ spettro di severità della malattia
 - ❑ durata del periodo di osservazione

The CLIP story (2)

Canada 2002



Giappone 2001

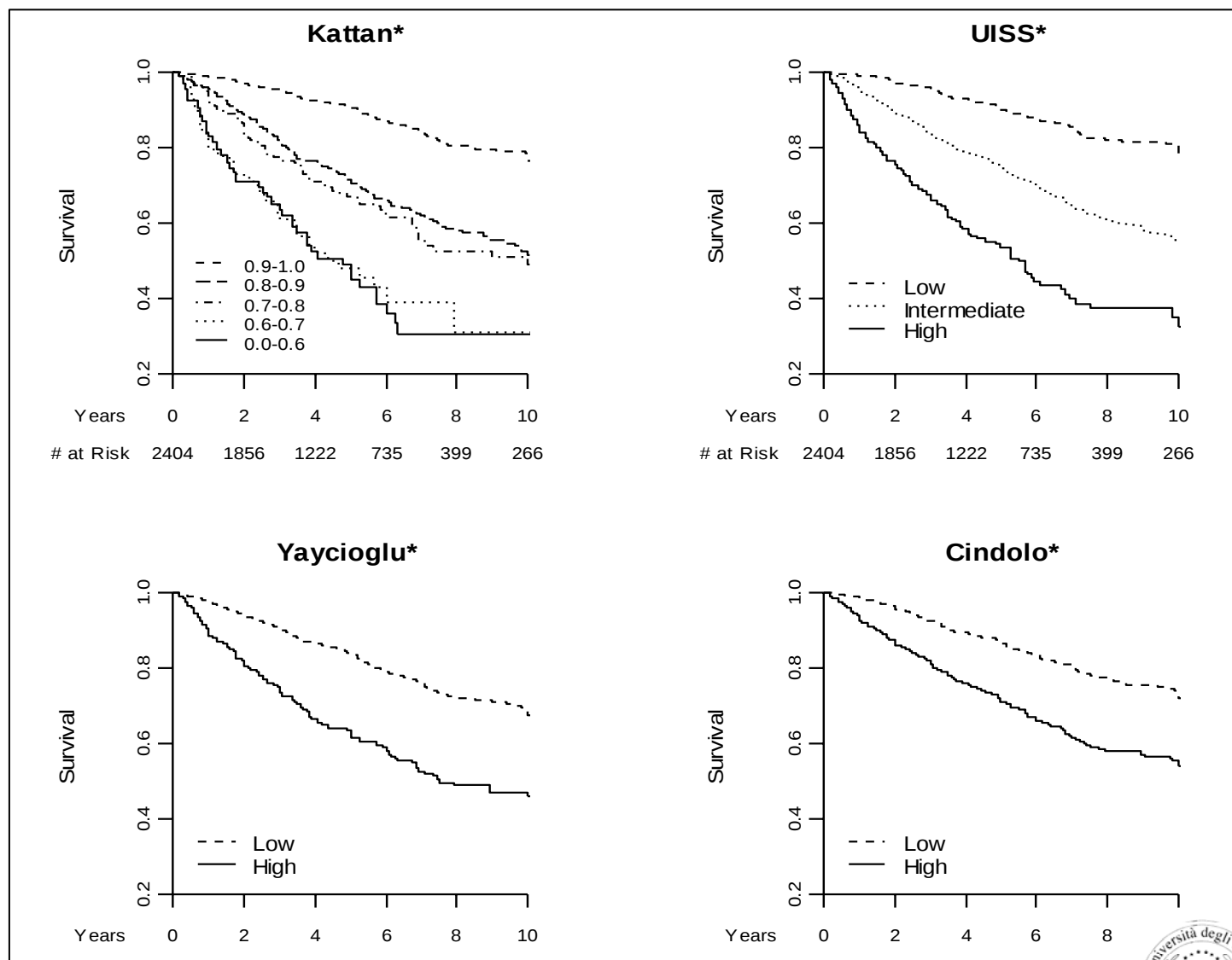


Confronto di 4 modelli prognostici per il Ca renale



	Kattan	UISS	Yaycioglu	Cindolo
Year of publication	2001	2002	2001	2003
Date of Surgery	1989-98	1989-99	1990-99	1987-02
Type of surgery	Part/Rad	Part/Rad	Part/Rad	Part/Rad
Laterality of tumours	unilateral	unilateral	unilateral	unilateral
Histology type	all	all	n.a.	all
1997 T stages	1-3	1-4	1-3	1-3
Nodal involvement or distant metastasis	no	yes	no	no
End points	5y RFS	OS	RFS	RFS
Number of categories	nomogram	3	2	2
Variables	Symptoms Histology Tumour size Pathol. stage	Pathol. Stage Fuhrman's grade ECOG PS	Symptoms Clinical size	Symptoms Clinical size

Confronto di 4 modelli prognostici per il Ca renale



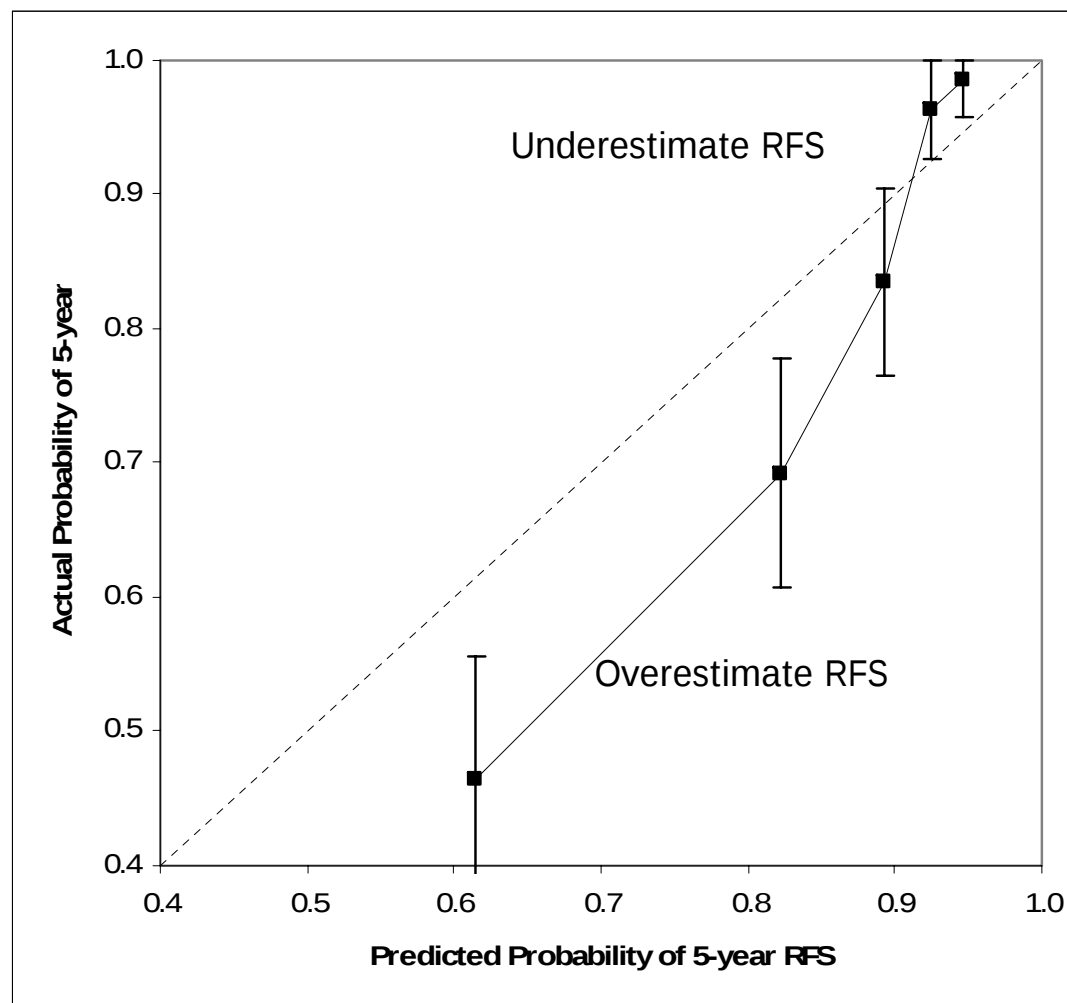
Confronto di 4 modelli prognostici: c-index di Harrell



	Overall survival		Recurrence free survival §	
Kattan	0.706	<i>(0.681 , 0.731)</i>	0.807	<i>(0.777 , 0.835)</i>
UISS	0.683	<i>(0.661 , 0.705)</i>	0.782	<i>(0.752 , 0.812)</i>
Yaycioglu	0.589	<i>(0.566 , 0.611)</i>	0.651	<i>(0.609 , 0.691)</i>
Cindolo	0.615	<i>(0.592 , 0.636)</i>	0.672	<i>(0.640 , 0.704)</i>

bootstrap confidence intervals, § Calculation only on 3 centers

Nomogramma di Kattan: calibrazione

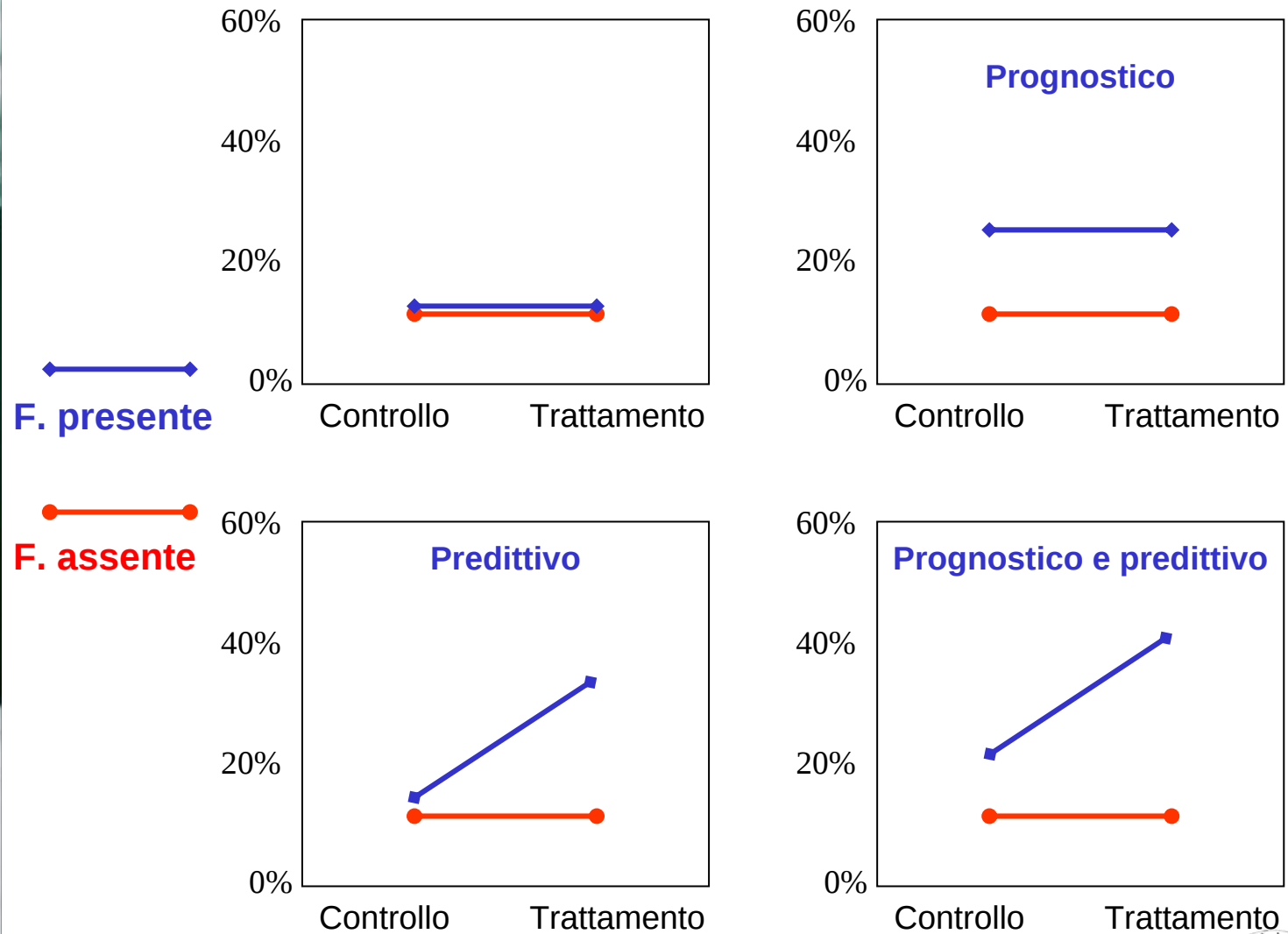


Quintiles of predicted probabilities



- Insufficiente credibilità clinica
- Insufficiente dimostrazione dell'utilità del modello prognostico
 - ❑ accuratezza
 - ❑ generalizzabilità
 - ❑ efficacia

Fattori prognostici e predittivi



Le differenze si ricompongono

Wild-Type *KRAS* Is Required for Panitumumab Efficacy in Patients With Metastatic Colorectal Cancer

J Clin Oncol 26:1626-1634. © 2008

The treatment effect on PFS in the wild-type (WT) *KRAS* group was significantly greater (interaction test $P < .0001$) than in the mutant group.

K-ras Mutations and Benefit from Cetuximab in Advanced Colorectal Cancer

N Engl J Med 2008;359:1757-65.

The effectiveness of cetuximab was significantly associated with *K-ras* mutation status ($P = 0.01$ and $P < 0.001$ for the interaction of *K-ras* mutation status with overall survival and progression-free survival, respectively).

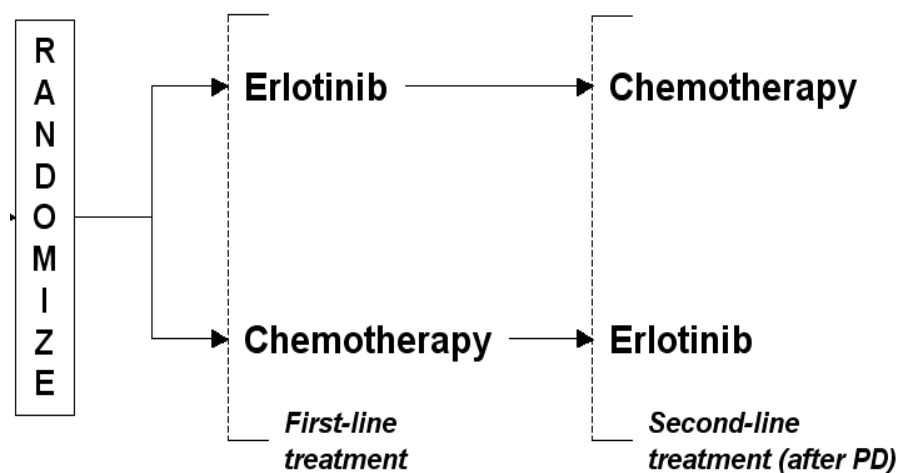




TORCH

TARCEVA OR CHEMOTHERAPY

An international multicenter randomized phase III study of first-line Erlotinib followed by second-line Cisplatin + Gemcitabine versus first-line Cisplatin + Gemcitabine followed by second-line Erlotinib in advanced non small cell lung cancer



Io credo che dovremmo
limitare lo studio ai soli
pazienti EGFR+

L'analisi dei sottogruppi
dimostra che erlotinib non
serve nei pazienti EGFR-

Table 3. Analysis of Responses and Survival.*

Factor	No. of Patients	No. Who Could Be Evaluated	No. of Responses (%)	P Value	Hazard Ratio for Death (95% CI) [†]	P Value
Entire study group	731	427	38 (9)	—	0.70 (0.58–0.85)	<0.001
Sex				0.007		
Male	475	281	17 (6)		0.76 (0.62–0.94)	0.01
Female	256	146	21 (14)		0.80 (0.59–1.07)	0.13
Histopathological subtype				<0.001		
Adenocarcinoma	365	209	29 (14)		0.72 (0.56–0.92)	0.008
Other	366	218	9 (4)		0.81 (0.64–1.02)	0.07
Smoking history				<0.001		
Current or former	545	311	12 (4)		0.87 (0.71–1.05)	0.14
None	146	93	23 (25)		0.42 (0.28–0.64)	<0.001
Unknown	40	23	3 (13)		1.09 (0.54–2.22)	0.80
Race or ethnic group				0.02		
Asian	91	53	10 (19)		0.61 (0.37–1.03)	0.06
Other	640	374	28 (7)		0.79 (0.66–0.95)	0.01
Results of EGFR IHC analysis	325			0.1		
Positive‡	184	106	12 (11)		0.68 (0.49–0.95)	0.02
Negative	141	80	3 (4)		0.93 (0.63–1.36)	0.70
FISH status	125			0.03		
Not amplified	69	41	1 (2)		0.85 (0.48–1.51)	0.59
Amplified§	56	25	5 (20)		0.44 (0.23–0.82)	0.008
Mutational status	170			0.37		
Wild type	137	81	6 (7)		0.73 (0.49–1.10)	0.13
Mutation	40	19	3 (16)		0.77 (0.40–1.50)	0.45

Io credo che dovremmo limitare lo studio ai soli pazienti EGFR+

L'analisi dei sottogruppi dimostra che erlotinib non serve nei pazienti EGFR-

Io credo invece che dovremmo includere tutti i pazienti indipendentemente dalla positività per EGFR.

Stiamo parlando di malati con pochi mesi di sopravvivenza e senza valide alternative terapeutiche.

E poi non è corretto escludere i pazienti EGFR- sulla base dell'analisi dei sottogruppi.

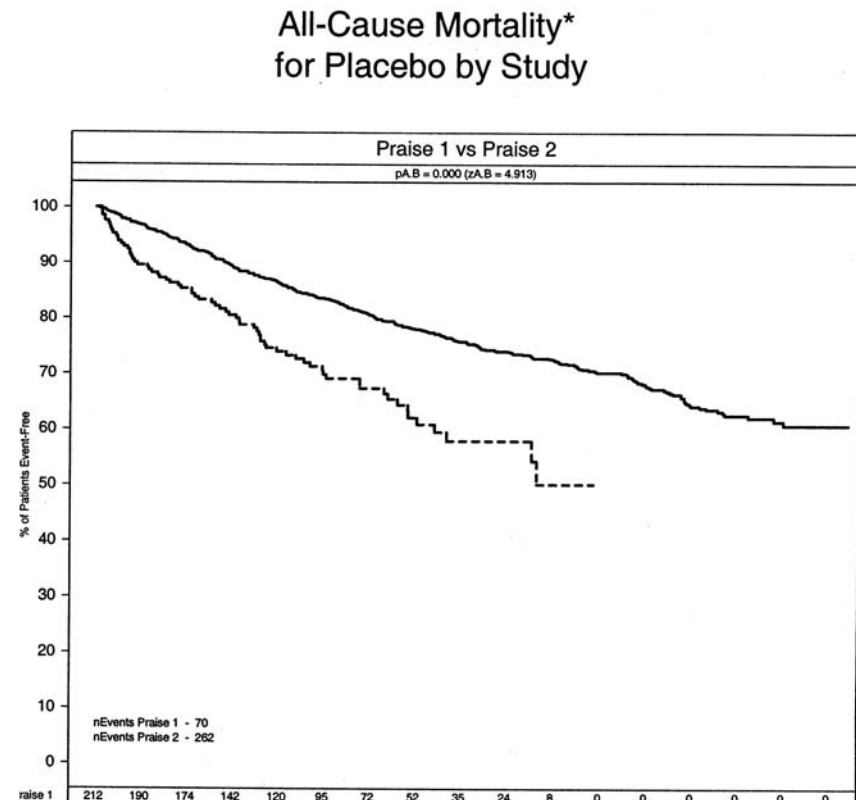
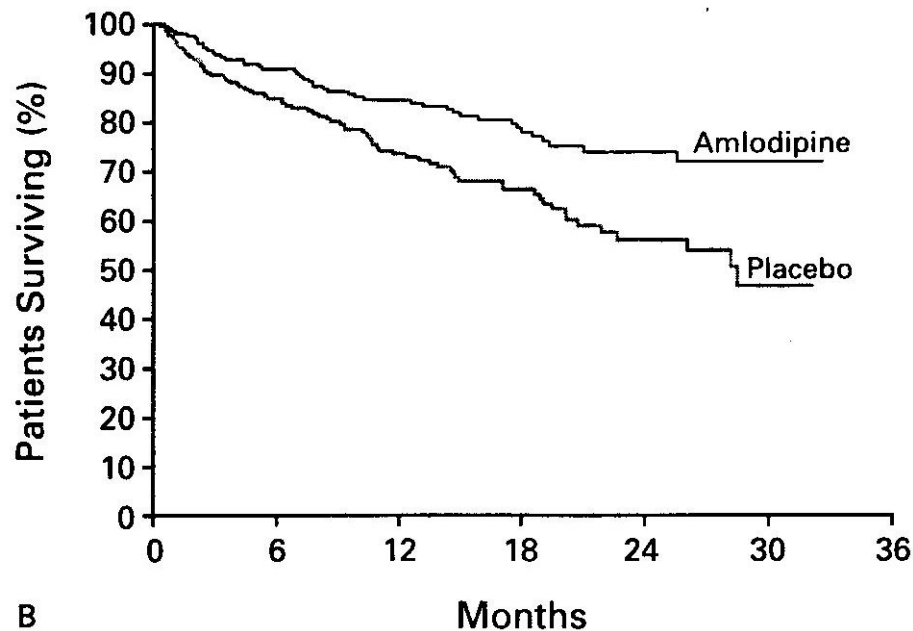
BR.21: summary of treatment effect – **SURVIVAL**

		HR	CI	p*
Gender	Male (475)	0.8	0.6–0.9	0.76
	Female (256)	0.8	0.6–1.1	
Histology	Adenocarcinoma (365)	0.7	0.6–0.9	0.37
	Other (366)	0.8	0.6–1.0	
Ethnicity	Asian (91)	0.6	0.4–1.0	0.44
	Other (640)	0.8	0.7–0.9	
Smoking history	Ever (545)	0.9	0.7–1.0	0.02
	Never (146)	0.4	0.3–0.6	
	Unknown (40)	1.1	0.5–2.6	

*p value for interaction between erlotinib and clinical variables

PRAISE I (1996) e II (2003)

- Amlodipine vs. placebo (n=1153)
- Trattamento x eziologia $P = 0.004$
- Ischemica $P = \text{NS}$
- Non ischemica $P < 0.001$
- Amlodipine vs. placebo (n=1653)
- Solo pz non ischemici
- $RR \cong 1.0$





- Insufficiente credibilità clinica
- Insufficiente dimostrazione dell'utilità del modello prognostico
 - ❑ accuratezza
 - ❑ generalizzabilità
 - ❑ efficacia

Doctors' poor uptake of these tools to aid difficult decisions may result from **healthy scepticism** of undocumented new technology.

