Validation of Surrogate Endpoints with Data from Randomized Trials

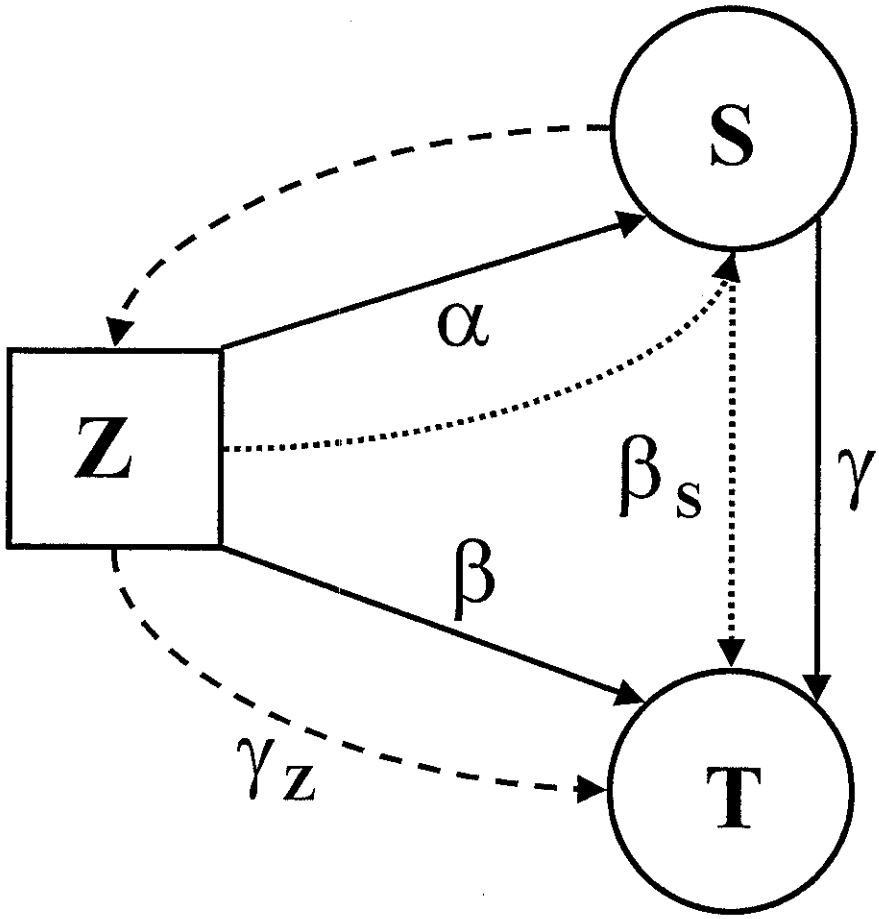
Pascal Piedbois

Marc Buyse

Geert Molenberghs

NIH, December 2, 1998

NOTATION



Ref: ***Biometrics*** **54**:1014 (1998)

DATA

4 meta-analyses in advanced colorectal cancer

Z: ***Treatment indicator***

* 0: standard therapy
* 1: experimental therapy

T: ***True endpoint: survival time***

* continuous
* right-censored
* Time from randomization to death from any cause

S: ***Surrogate endpoint: tumor shrinkage***

* indicator variable based on WHO definition
* 0: CR/PR
* 1: SD/PD/NE

N: ***3895 patients***

* Individual patient data available on every randomized patient, regardless of eligibility, evaluability, or follow-up
* Index *i* refers to patient

Refs: ***JCO*** **10**:896 (1992); **12**:960 (1994); **16**:301 (1998); ***JNCI*** **88**:252 (1996)

PRENTICE'S CRITERIA

Criterion 1 : Treatment Z is prognostic for surrogate S

* $logit\ [\ P(S_i = 0) \mid Z_i\ ] = \mu_S + \alpha\ Z_i$
* $\alpha = -0.755\ (P < 10^{-10})$
* Response odds ratio : $OR = e^{\alpha} = 0.47$
* 95% confidence limits : [0.40; 0.56]

Criterion 2 : Treatment Z is prognostic for true endpoint T

* $log\ [\ h\ (T_i = t \mid Z_i\ )\ /\ h_0\ (T_i = t)] = \mu_T + \beta\ Z_i$
* $\beta = -0.105\ (P = 0.003)$
* Survival hazard ratio : $HR = e^{\beta} = 0.90$
* 95% confidence limits : [0.84; 0.97]

Criterion 3 : Surrogate S is prognostic for true endpoint T

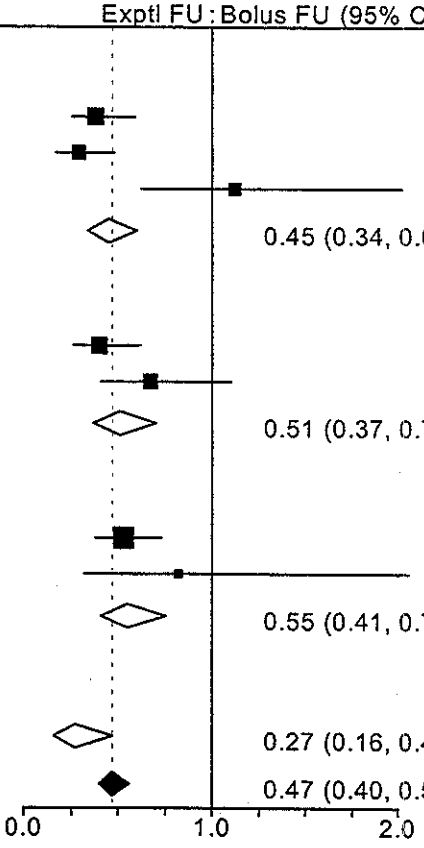
* $log\ [\ h\ (T_i = t \mid S_i\ )\ /\ h_0\ (T_i = t)] = \mu + \gamma\ S_i$
* Caution : length-biased sampling
  $\Rightarrow$ landmark analysis required
* $\gamma = 0.51$ (s.e. 0.027, $P < 10^{-20}$)
  with a landmark time of 3 months

Ref: ***Statist Med*** **8**:431 (1989)

RESPONSE ODDS RATIOS

Response odds ratios

	Exptl FU O/N	Bolus FU O/N	O-E	V	Odds Ratio Exptl FU : Bolus FU (95% C.I.)
(a) 5FU + LV versus bolus FU					
Weekly schedules	95 / 408	22 / 239	-20.6	21.5	
Monthly schedules	55 / 197	18 / 194	-18.3	14.8	
Higher 5FU bolus	31 / 198	24 / 145	1.2	11.0	
Subtotal	181 / 803	64 / 578	-37.7	47.3	0.45 (0.34, 0.60)
(b) 5FU + MTX versus bolus FU					
Same 5FU bolus	64 / 338	28 / 342	-18.1	19.8	
Higher 5FU bolus	51 / 266	30 / 222	-6.4	16.2	
Subtotal	115 / 604	58 / 564	-24.5	36.0	0.51 (0.37, 0.70)
(c) Continuous infusion versus bolus FU					
5FU alone	113 / 483	66 / 475	-23.0	36.0	
5FU + LV	11 / 69	10 / 76	-0.9	4.5	
Subtotal	124 / 552	76 / 551	-23.9	40.5	0.55 (0.41, 0.75)
(d) Hepatic arterial infusion versus bolus FU					
Subtotal	59 / 125	22 / 121	-17.3	13.3	0.27 (0.16, 0.47)
Total	479 / 2084	220 / 1814	-103.4	137.1	0.47 (0.40, 0.56)

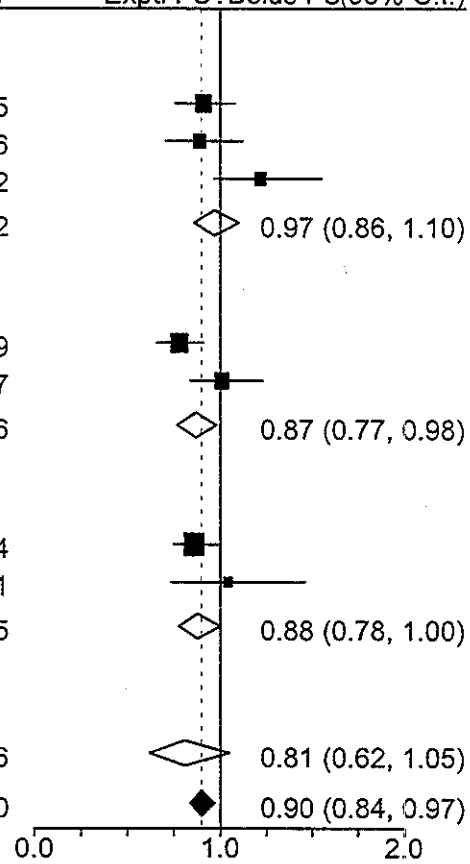


test of treatment effect $X^2_1 = 77.95$, $P < 0.001$ Experimental FU better / Bolus FU better

SURVIVAL HAZARD RATIOS

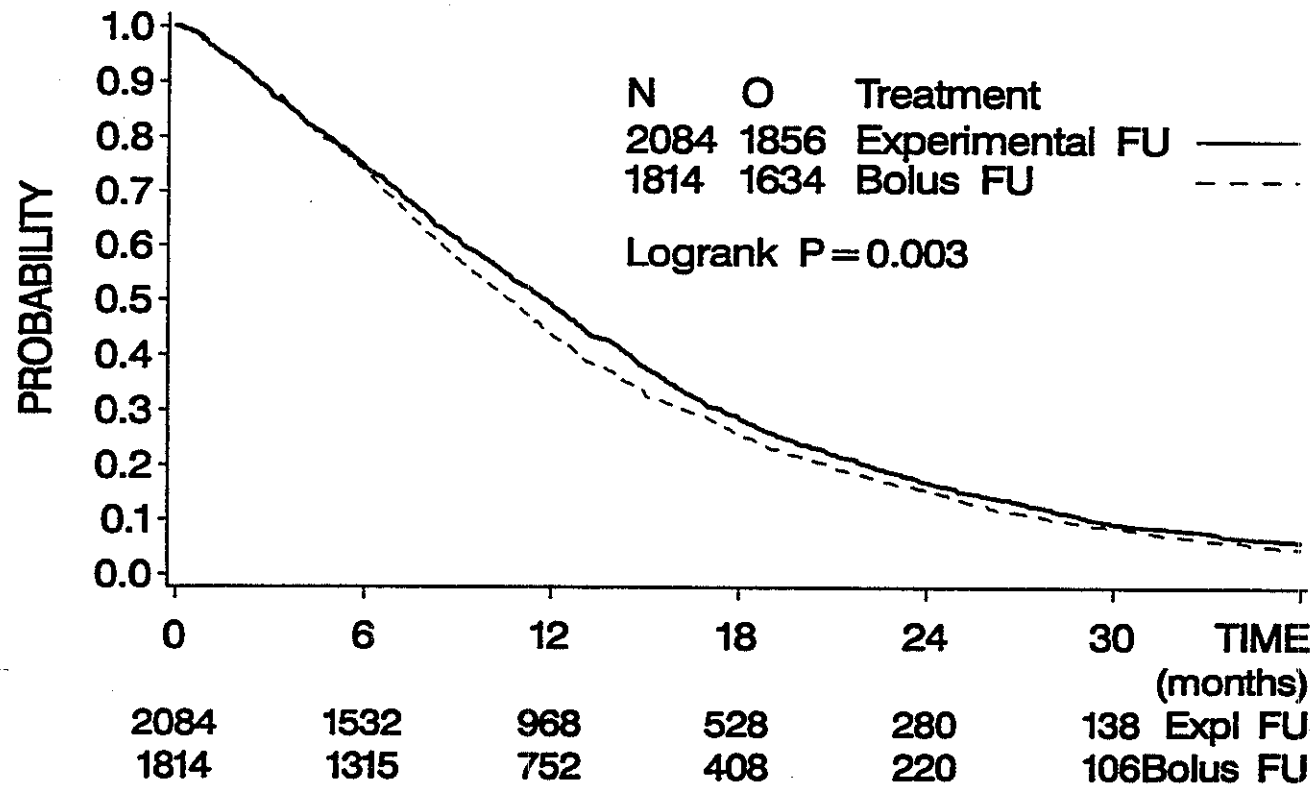
Survival hazard ratios

	Exptl FU O/N	Bolus FU O/N	O-E	V	Hazard Ratio Exptl FU : Bolus FU (95% C.I.)
(a) 5FU + LV versus bolus FU					
Weekly schedules	360 / 408	225 / 239	-12.3	126.5	
Monthly schedules	160 / 197	160 / 194	-8.7	75.6	
Higher 5FU bolus	176 / 198	116 / 145	13.8	69.2	
Subtotal	696 / 803	501 / 578	-7.1	271.2	0.97 (0.86, 1.10)
(b) 5FU + MTX versus bolus FU					
Same 5FU bolus	304 / 338	306 / 342	-37.1	145.9	
Higher 5FU bolus	247 / 266	202 / 222	1.5	104.7	
Subtotal	551 / 604	508 / 564	-35.6	250.6	0.87 (0.77, 0.98)
(c) Continuous infusion versus bolus FU					
5FU alone	428 / 483	437 / 475	-31.2	212.4	
5FU + LV	65 / 69	73 / 76	1.2	33.1	
Subtotal	493 / 552	510 / 551	-30.0	245.5	0.88 (0.78, 1.00)
(d) Hepatic arterial infusion versus bolus FU					
Subtotal	116 / 125	115 / 121	-11.8	55.6	0.81 (0.62, 1.05)
Total	1856 / 2084	1634 / 1814	-84.5	823.0	0.90 (0.84, 0.97)



test of treatment effect $X^2_1 = 8.68$, P= 0.003 Experimental FU better / Bolus FU better

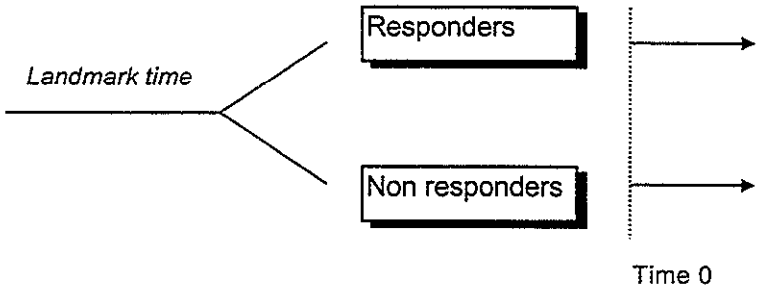
KAPLAN-MEIER PLOT



LANDMARK ANALYSES

The "landmark" time :

* Comparison of survival times after a landmark time sufficient for most responses to have been observed



* Time of individual responses not known ⇒ choice of landmark times from 3 to 12 months
* response strongly prognostic for survival :
  hazard ratios : $HR = e^{\gamma} =$
  1.67 (landmark of 3 months)
  1.45 (landmark of 6 months)
  1.31 (landmark of 9 months)
  1.24 (landmark of 12 months)

Ref: ***Statist Med*** **15**:2797 (1996)

FREEDMAN'S PROPORTION EXPLAINED

Criterion 4 : The full effect of Z on T is captured by S

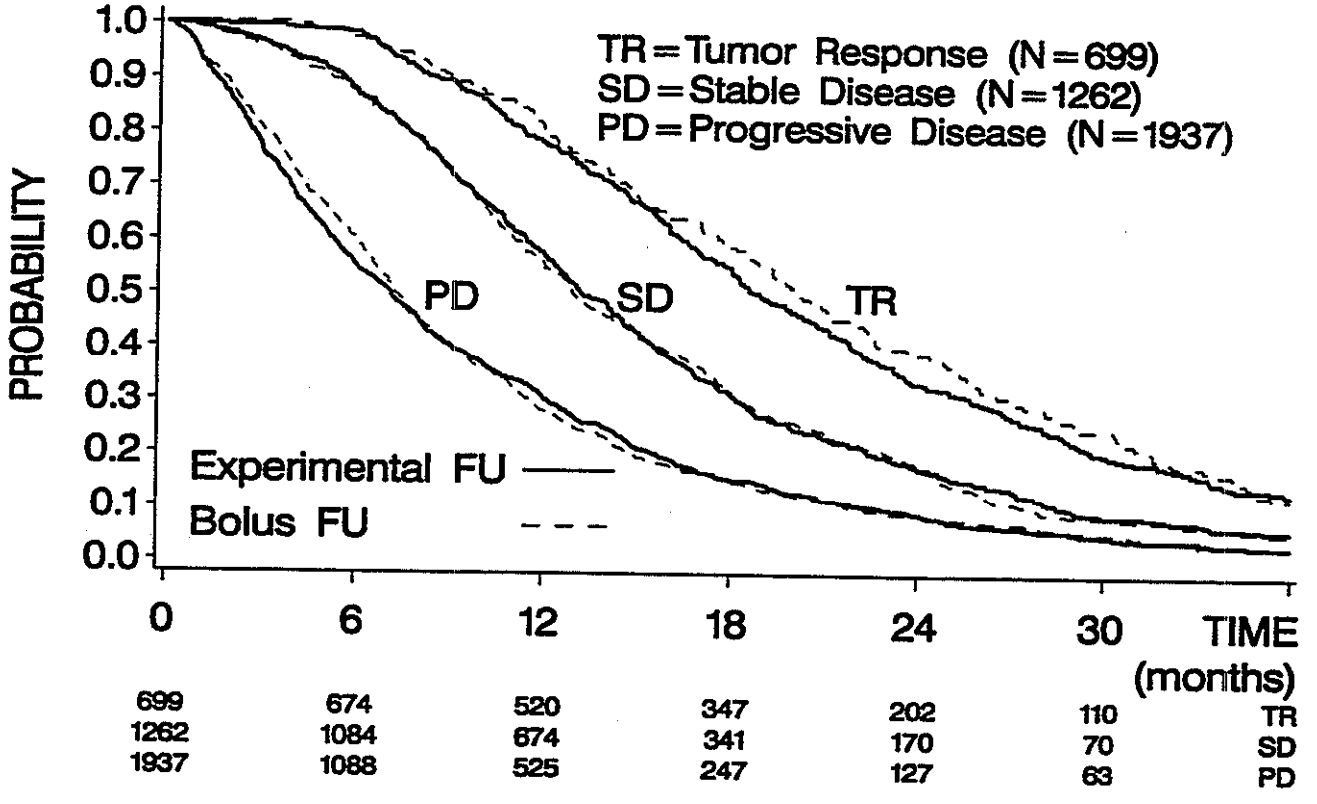
* $\log [\, h\,(T_i = t \mid Z_i\,,\, S_i) \,/\, h_0\,(T_i = t)] = \mu'_T + \beta_S\, Z_i + \gamma_Z\, S_i$
* $H_0 : \beta_S \neq 0$?

Criterion : Proportion explained (PE) must be close to 1

* *Proportion explained:* $PE(T,S,Z) = 1 - \beta_S / \beta$
* $PE = 0.99$ (95% confidence limits [0.54; 4.15]

Refs: ***Statist Med*** **11**:167 (1992); **16**:1515 (1997)

FREEDMAN'S PROPORTION EXPLAINED



THE RELATIVE EFFECT

Problems with PE :

* *PE* not restricted to unit interval
* confidence limits (Fieller or delta method) wide, unless
  - very large sample sizes ($\Rightarrow$ meta-analyses)
  - very strong effect of Z on T (unlikely)

Criterion : The effect of *Z* on *S* predicts clinically useful effects of *Z* on *T*

* *Relative effect:* $RE(T,S,Z) = \beta / \alpha$
* $RE = 0.11$ (95% confidence limits [0.02; 0.22]

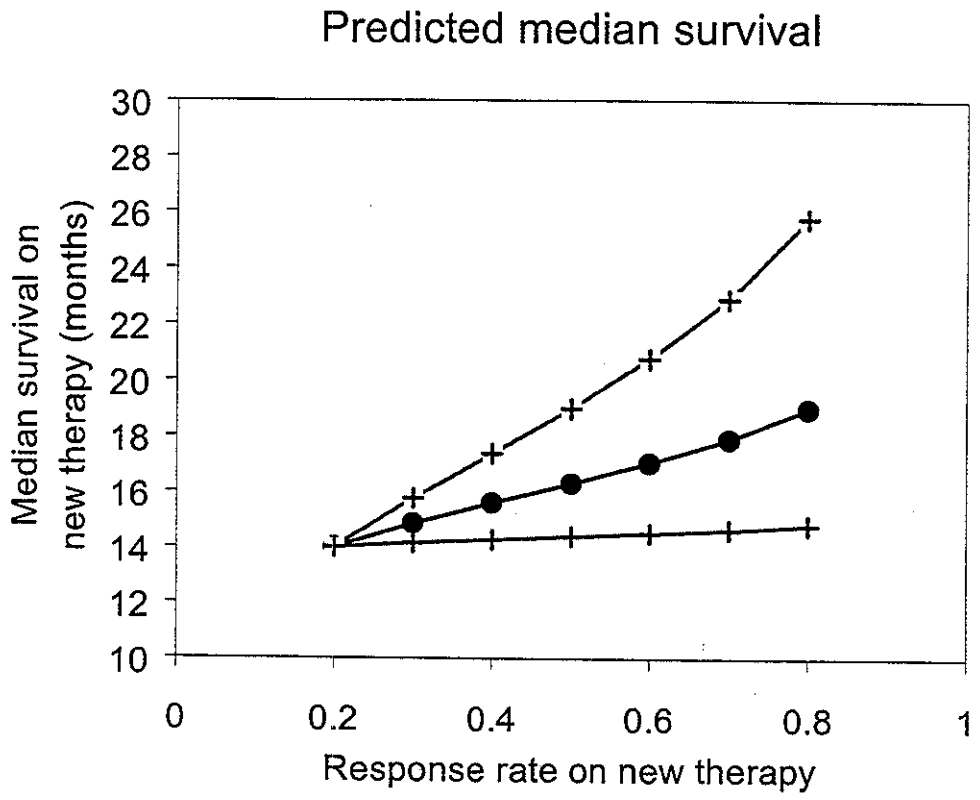
Ref: ***Biometrics* 54**:1014 (1998)

PREDICTION OF SURVIVAL BENEFITS

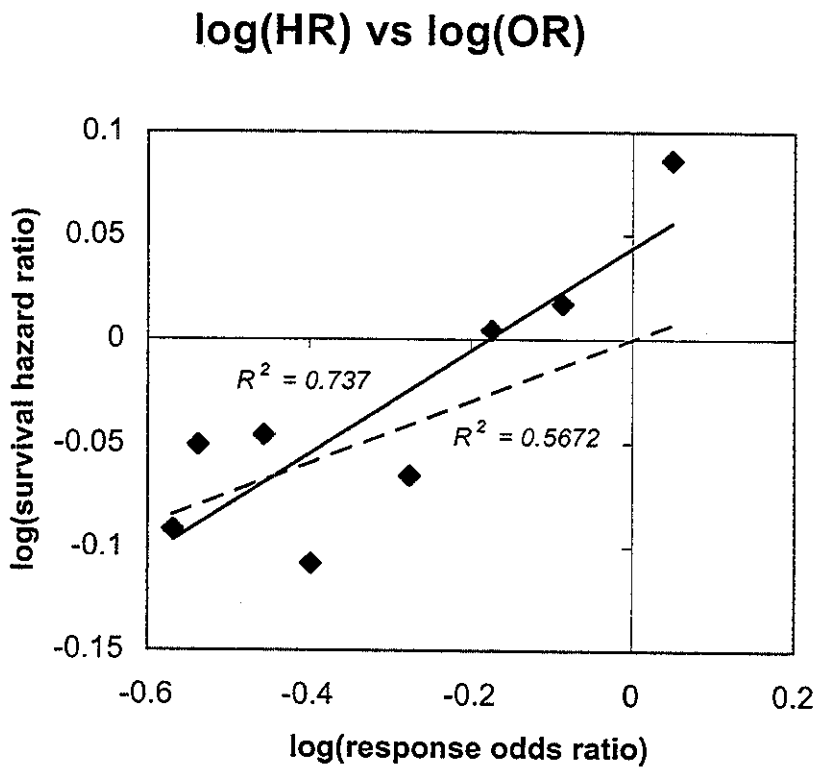
* $RE = \beta / \alpha = \log(HR) / \log(OR)$

Assumptions :

* Response rate of standard therapy = 20%
* Exponential survival
* Median survival on standard therapy = 14 months
* **RE constant !**



IS RELATIVE EFFECT CONSTANT ?



IS RELATIVE EFFECT PREDICTIVE ?

Example 1 :

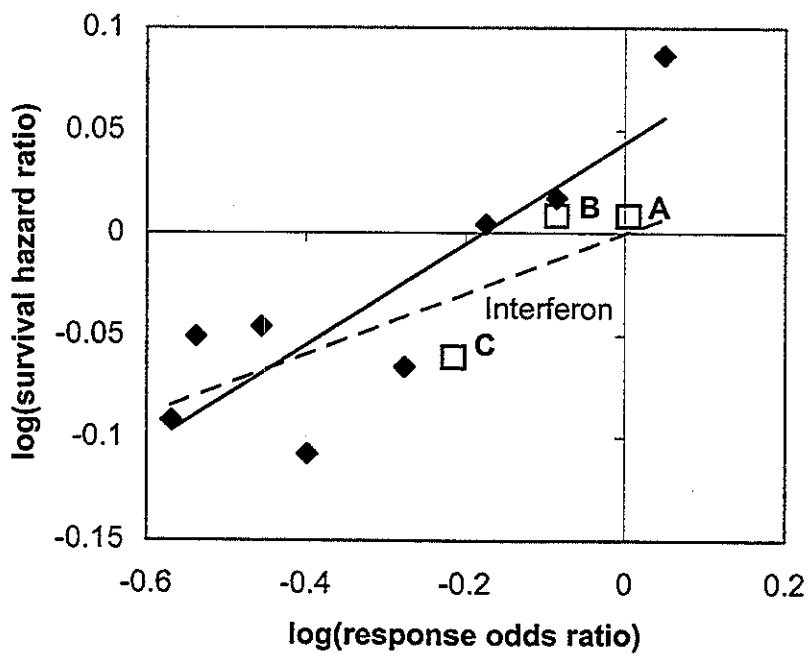
Meta-analysis of 15 trials comparing

A. Interferon-α + 5FU vs 5FU *(N=758)*

B. Interferon-α + 5FU + LV vs 5FU + LV*(N=759)*

C. Interferon-α + 5FU vs 5FU + LV*(N=976)*

log(HR) vs log(OR)

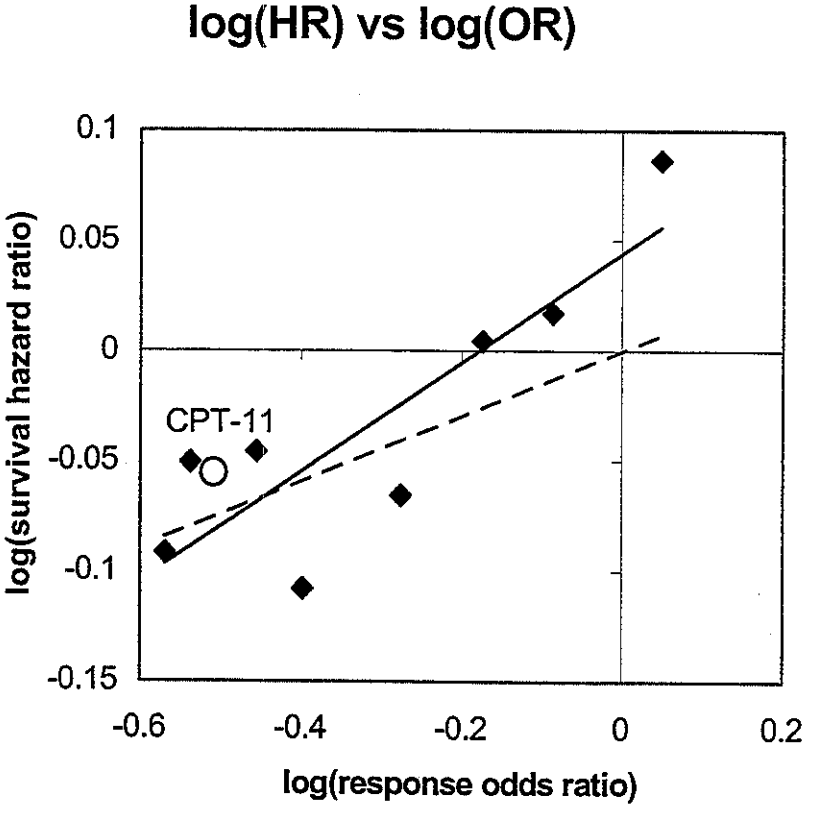


IS RELATIVE EFFECT PREDICTIVE ?

Example 2 (from published data!) :

Second-line therapy comparing

CPT-11 vs 5FU continuous infusion *(N=256)*

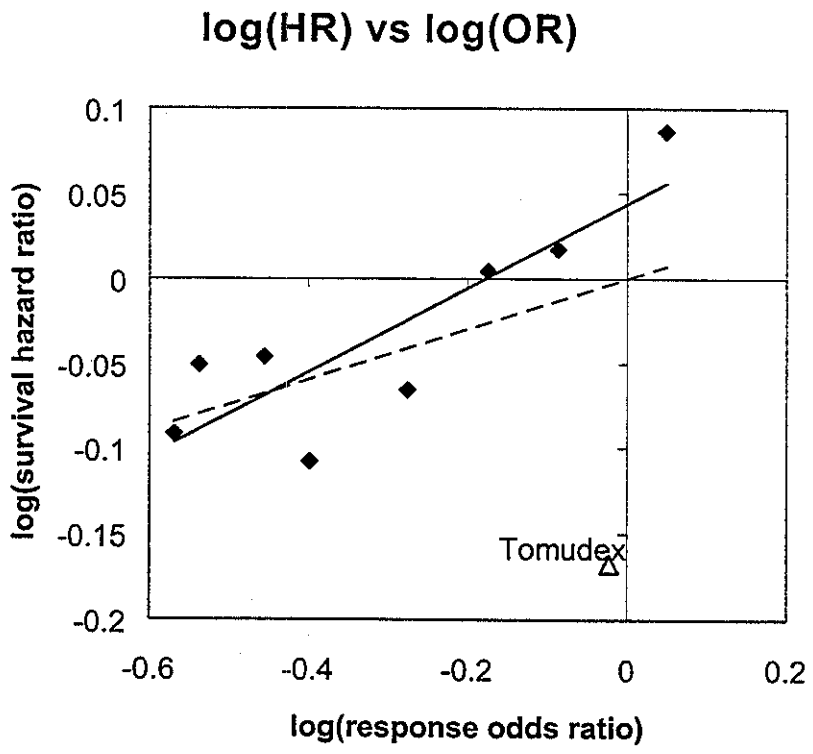


IS RELATIVE EFFECT PREDICTIVE ?

Example 3 (from published data!) :

Meta-analysis of 3 trials comparing

Tomudex vs 5FU + leucovorin *(N=1,361)*



CONCLUSIONS

Primary objectives of validation :

* *NOT* statistical significance
* *Estimation* of treatment effects (α, β) and of relative effect (RE = β / α)
* *Prediction* of β from α

Requirements for informative statistical validation :

* Randomized evidence (*reliable* estimates)
* Large sample sizes (*precise* estimates)
* Individual patient data (*unbiased* estimates)
* Evidence from several sources (*stable* estimates?)