GOOD SCIENCE
BETTER MEDICINE
BEST PRACTICE

Statistical validation of biomarkers and surogate endpoints

Marc Buyse, ScD
IDDI, Louvain-la-Neuve & Hasselt University, Belgium

marc.buyse@iddi.com

Adding Value to Clinical Data

universiteit hasselt

OUTLINE

1. Setting the scene: definitions and types of biomarkers
2. Pharmacodynamic biomarkers
3. Prognostic biomarkers
4. Predictive biomarkers
5. Surrogate biomarkers

BIOMARKER

A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

Examples:

PSA, CTCs (prostate cancer)

KRAS mutation (colorectal cancer)

HER2-neu amplification (breast cancer)

Gene signatures

Tumour measurements (advanced tumors)?

Ref: Biomarkers Definition Working Group, Clin Pharmacol Ther 2001, ***69****: 89*

CLINICAL ENDPOINT

A clinical endpoint is a characteristic or variable that reflects how a patient feels, functions, or survives.

Examples:

disease-free or progression-free survival

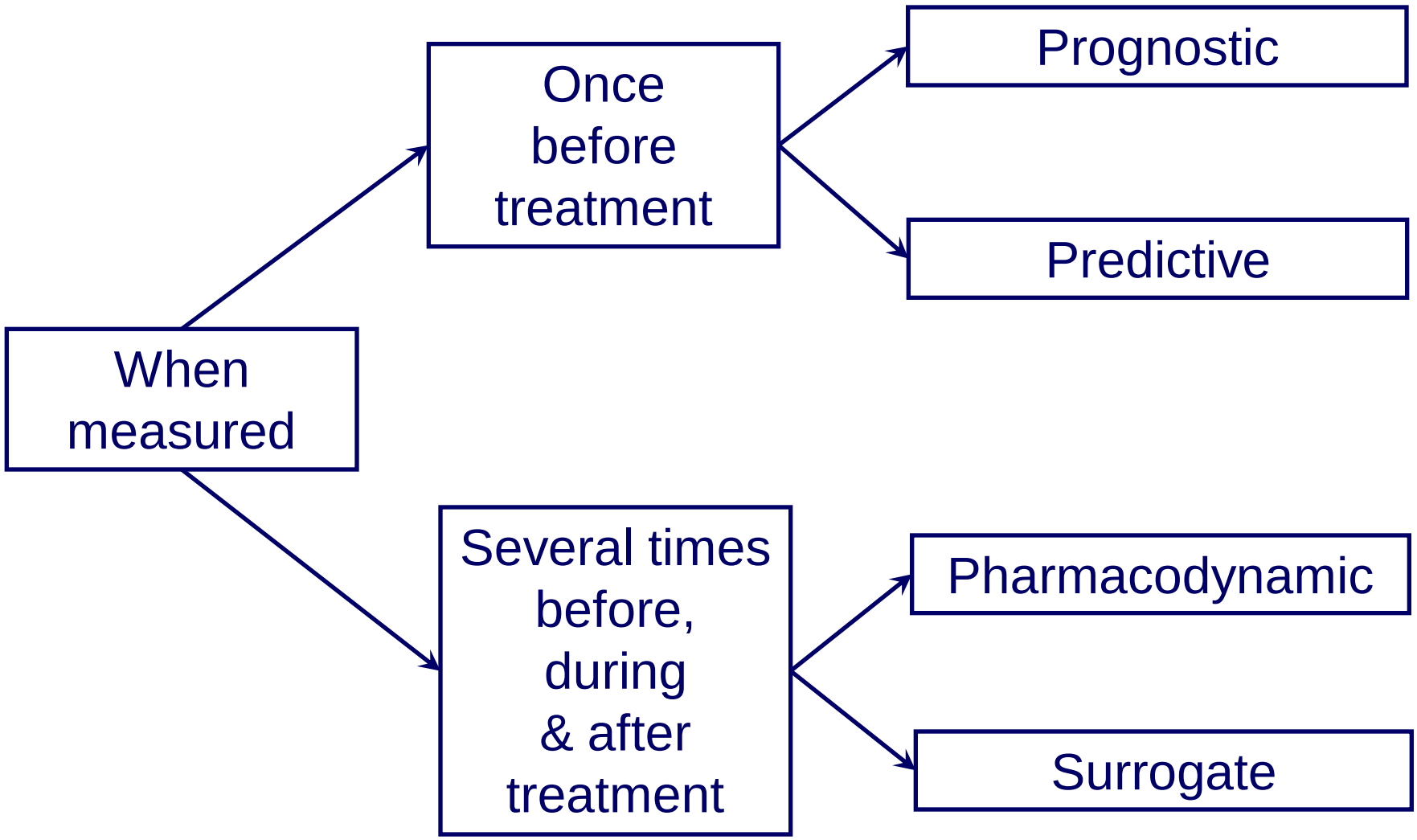
survival

quality of life

tumor response (non-solid tumors)?

Ref: Biomarkers Definition Working Group, Clin Pharmacol Ther 2001, ***69****: 89*

TYPES OF BIOMARKERS



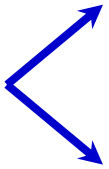
1. PHARMACODYNAMIC BIOMARKERS

Potential uses	Example: a protein kinase inhibitor
Proof of local exposure	Tumor penetration
Proof of mechanism	Inhibition of phosphorylated protein
Proof of principle (pathway activity)	Change in cell turnover
Proof of concept (clinical activity)	Tumor shrinkage

A PHASE II BIOMARKER-BASED TRIAL

Phase II trial of Interleukin-2 + a viral suspension of a recombinant vaccinia vector containing the sequence coding for the human MUC1 antigen

21 patients with elevated PSA after prostatectomy and histological documentation of MUC1 antigen expression



Weekly schedule

Three-weekly schedule

BIOMARKER AND CLINICAL OUTCOMES

Biomarker

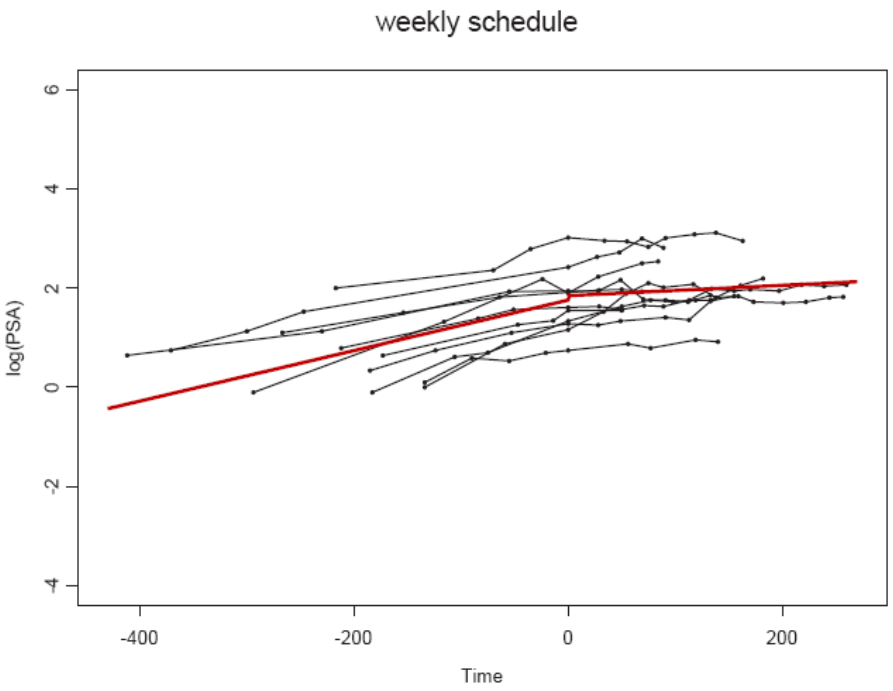
- PSA measurements over time

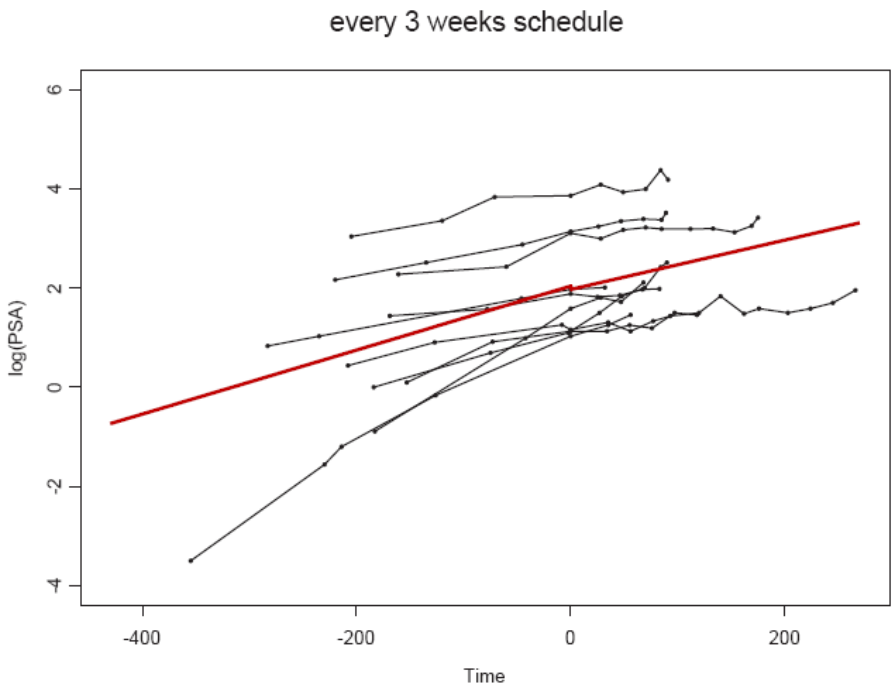
Protocol-defined outcomes

- PSA response rate*
- Duration of PSA response
- Time to PSA progression

* *PSA decreased to < 4 ng/ml or to < 50% of baseline level for at least 4 weeks*

PSA MEASUREMENTS OVER TIME





MODELLING OF PSA MEASUREMENTS

$$\log(PSA_{ij}) = \beta_1 T_i + \beta_1 W_i + \beta_3 t_{ij} + \beta_4 P_i + \beta_5 P_i t_{ij} + \beta_6 T_i t_{ij} + \beta_7 T_i P_i + \beta_8 T_i P_i t_{ij} + b_{0i} + b_{1i} t_{ij} + \varepsilon_{ij}$$

Model contains the following terms:

- Randomized treatment (Weekly or Three-weekly)
- Time
- Period (pre- vs. post-treatment)
- Interactions

A PHASE II BIOMARKER-BASED TRIAL

Table 3: *Effects of interest with corresponding p-values.*

Effect	p-value
Pre-baseline *vs* post-baseline (both schedules)	< 0.0001
Pre-baseline *vs* post-baseline (every 3 weeks schedule)	0.038
Pre-baseline *vs* post-baseline (weekly schedule)	< 0.0001
Weekly *vs* every 3 weeks (pre-baseline)	0.26
Weekly *vs* every 3 weeks (post-baseline)	0.0056

A PHASE II BIOMARKER-BASED TRIAL

Table 3: *Effects of interest with corresponding p-values.*

Effect	*p*-value
Pre-baseline *vs* post-baseline (both schedules)	< 0.0001
Pre-baseline *vs* post-baseline (every 3 weeks schedule)	0.038
Pre-baseline *vs* post-baseline (weekly schedule)	< 0.0001
Weekly *vs* every 3 weeks (pre-baseline)	0.26
Weekly *vs* every 3 weeks (post-baseline)	0.0056

Treatment had an overall effect

A PHASE II BIOMARKER-BASED TRIAL

Table 3: *Effects of interest with corresponding p-values.*

Effect	p-value
Pre-baseline *vs* post-baseline (both schedules)	< 0.0001
Pre-baseline *vs* post-baseline (every 3 weeks schedule)	0.038
Pre-baseline *vs* post-baseline (weekly schedule)	< 0.0001
Weekly *vs* every 3 weeks (pre-baseline)	0.26
Weekly *vs* every 3 weeks (post-baseline)	0.0056

Weekly schedule had a more pronounced effect on PSA levels

A PHASE II BIOMARKER-BASED TRIAL

Table 3: *Effects of interest with corresponding p-values.*

Effect	p-value
Pre-baseline *vs* post-baseline (both schedules)	< 0.0001
Pre-baseline *vs* post-baseline (every 3 weeks schedule)	0.038
Pre-baseline *vs* post-baseline (weekly schedule)	< 0.0001
Weekly *vs* every 3 weeks (pre-baseline)	0.26
Weekly *vs* every 3 weeks (post-baseline)	0.0056

There were no pre-treatment differences in PSA levels between the two schedules (as expected)

A PHASE II BIOMARKER-BASED TRIAL

Table 3: *Effects of interest with corresponding p-values.*

Effect	p-value
Pre-baseline *vs* post-baseline (both schedules)	< 0.0001
Pre-baseline *vs* post-baseline (every 3 weeks schedule)	0.038
Pre-baseline *vs* post-baseline (weekly schedule)	< 0.0001
Weekly *vs* every 3 weeks (pre-baseline)	0.26
Weekly *vs* every 3 weeks (post-baseline)	0.0056

The weekly schedule had a significantly larger effect on PSA levels as compared with the three-weekly schedule

PD BIOMARKERS IN EARLY TRIALS

- Early trials may use PD biomarkers to confirm a treatment's activity and select a dose for further testing
- Randomized phase II trials using PD biomarkers are more informative than uncontrolled trials looking just at « response rate » (a poor endpoint, statisticallly)

BUT

- Is biomarker predictive of clinical efficacy?

2. PROGNOSTIC BIOMARKERS

Potential uses:

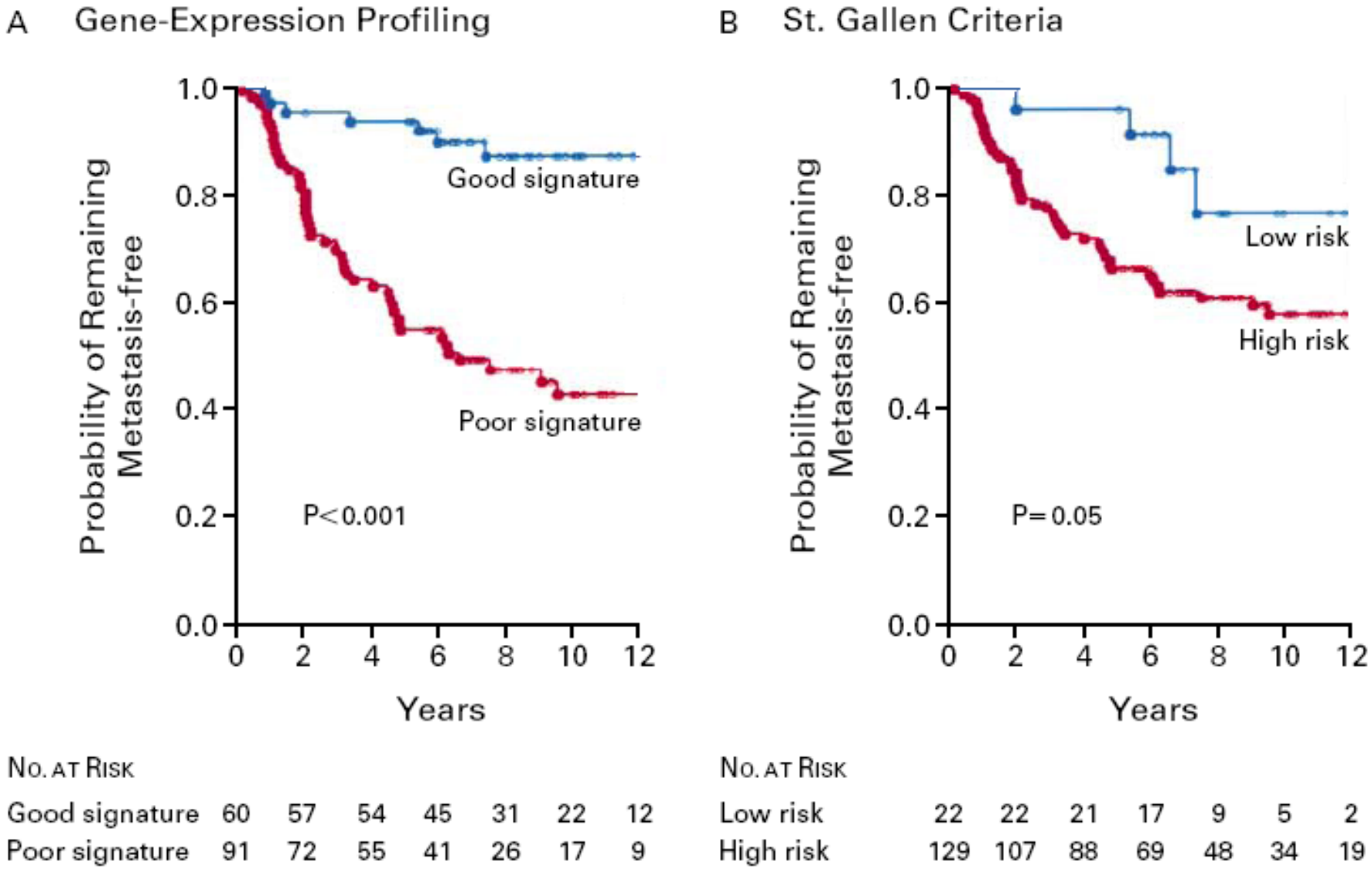
- Patient stratification in trials (no big deal)
- Treatment decisions

Difficulties:

- Is biomarker prognostic impact sufficient?

GENE SIGNATURES IN BREAST CANCER

- 70-gene « Amsterdam » signature (*Mammaprint™, Agendia*)
- 76-gene « Rotterdam » signature (*Veridex*)
- 21-gene assay (*Oncotype-DX™, Genomic Health*)
- 97-gene « genomic grade » (*Mapquant Dx™, Ipsogen*)
- Many others…



Ref: van de Vijver et al, NEJM 2002;347,1999

ALL RISK CLASSIFICATIONS HAVE POOR PREDICTIVE ACCURACY

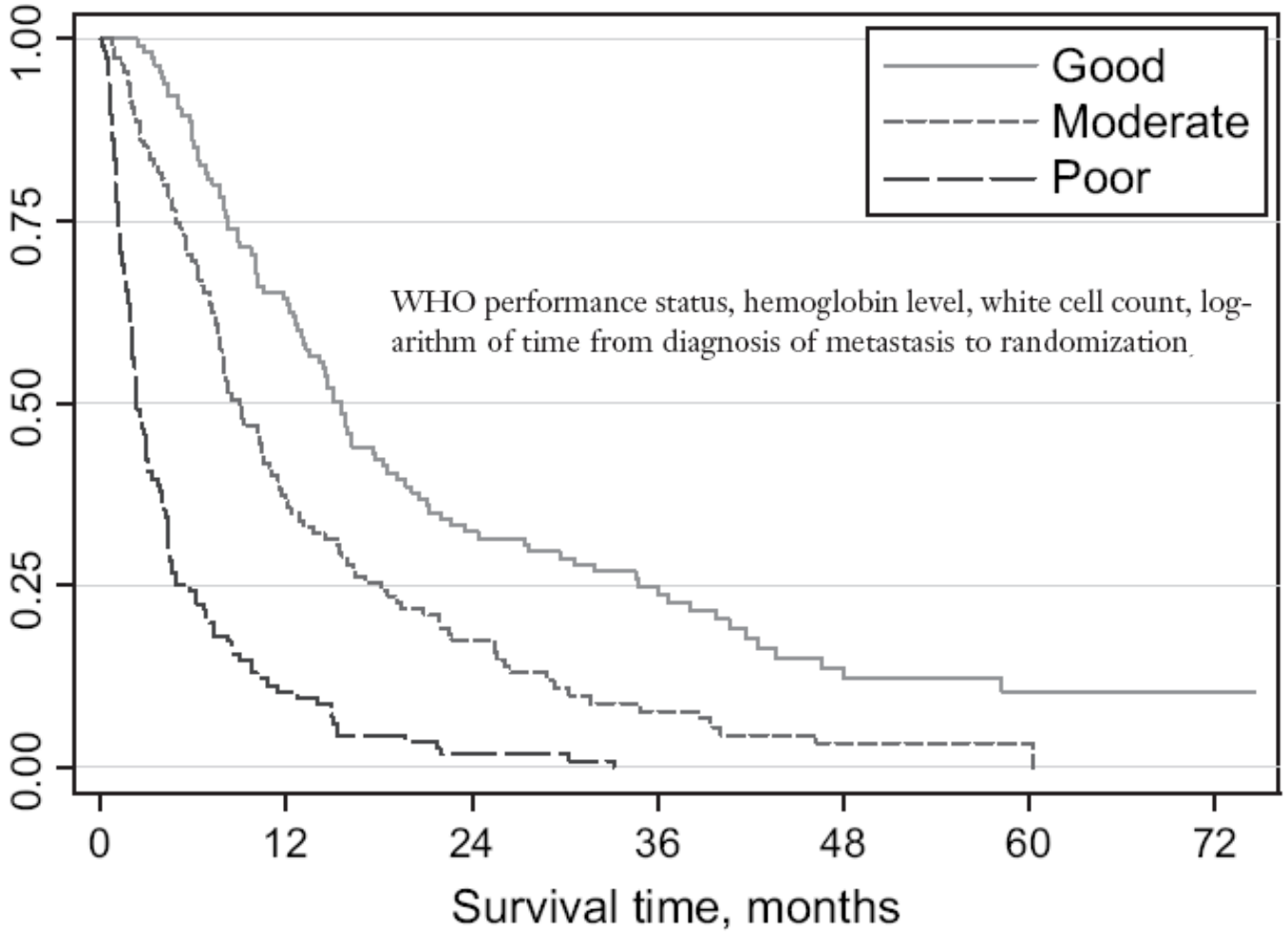
Metastases within 5 years	Sensitivity*	Specificity**
Gene signature	0.90	0.42
Adjuvant! software	0.87	0.29
NPI	0.91	0.32
St Gallen criteria	0.96	0.10

* Sensitivity = Proportion of patients with distant mets within 5 years who are classified high risk

** Specitivity = Proportion of patients without distant mets within 5 years who are classified low risk

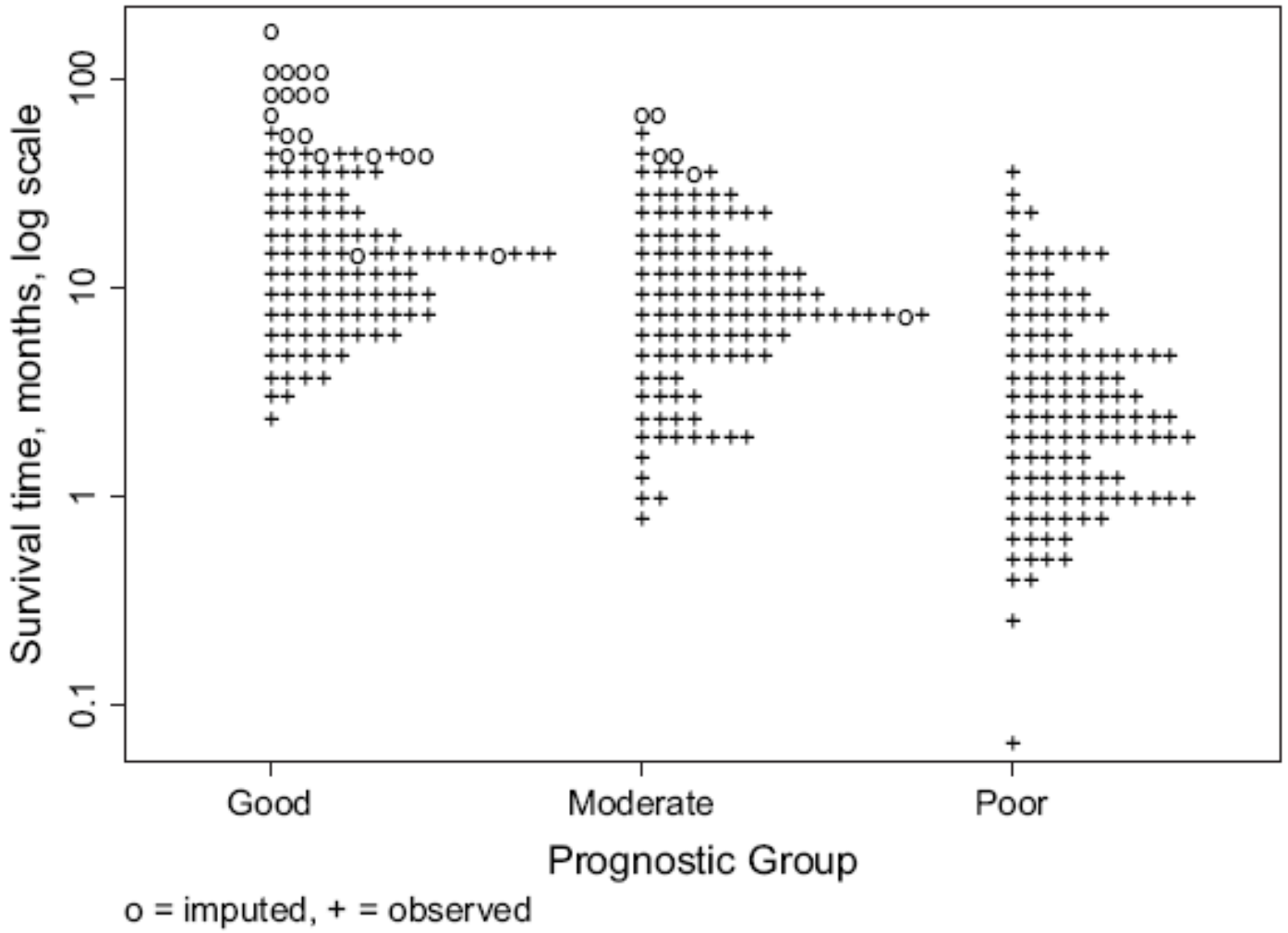
Ref: Buyse et al, JNCI 2006; 98:1183.

EVEN THE BEST PROGNOSTIC MODELS HAVE POOR DISCRIMINATION POWER



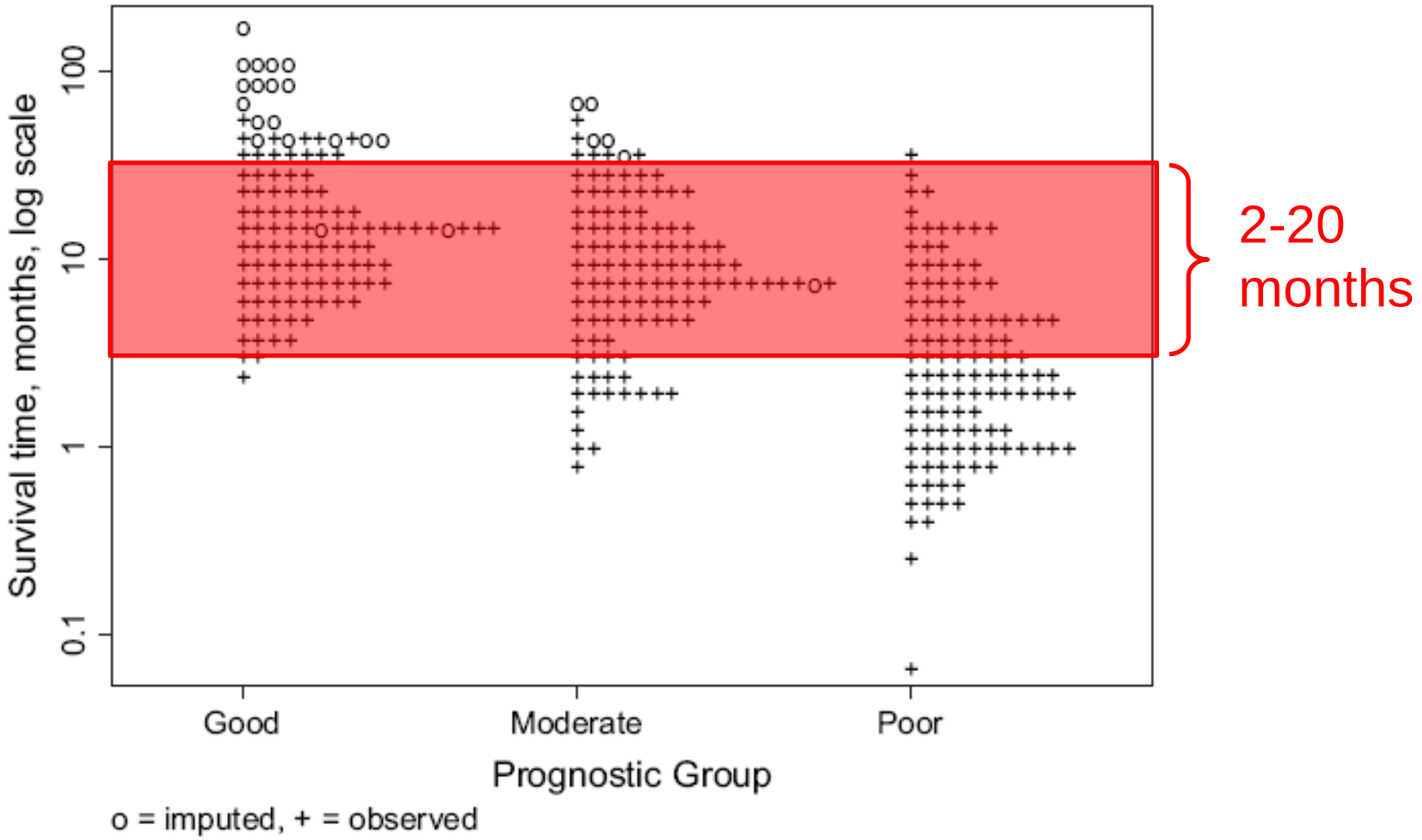
Ref: Royston et al, JNCI 2008; 100:92.

EVEN THE BEST PROGNOSTIC MODELS HAVE POOR DISCRIMINATION POWER



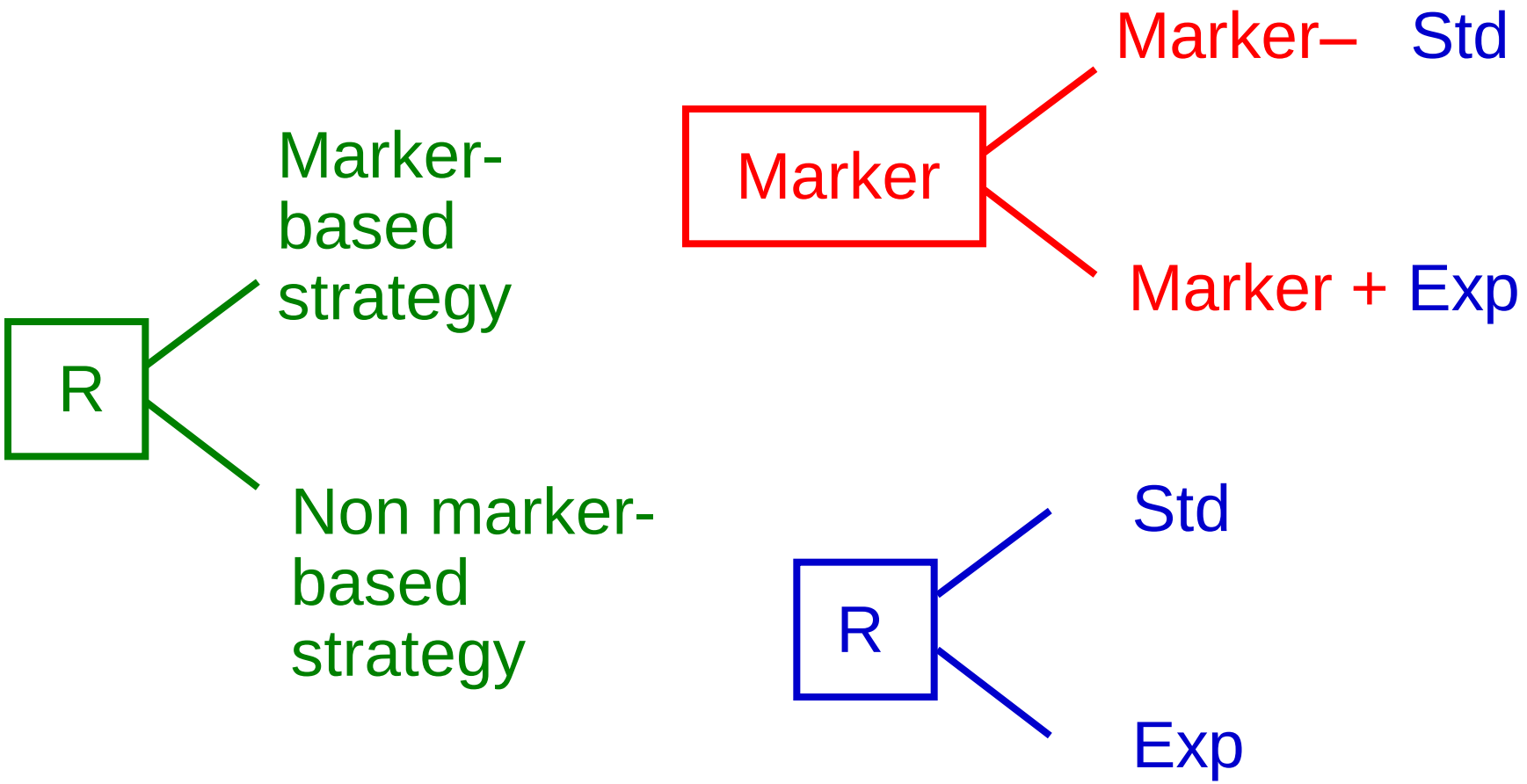
Ref: Royston et al, JNCI 2008; 100:92.

EVEN THE BEST PROGNOSTIC MODELS HAVE POOR DISCRIMINATION POWER



Ref: Royston et al, JNCI 2008; 100:92.

TRIAL DESIGN TO VALIDATE CLINICAL UTILITY OF PROGNOSTIC MARKER

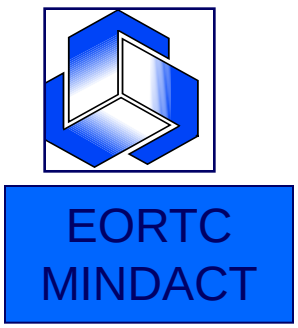


PROBLEM WITH PROGNOSTIC MARKER VALIDATION TRIALS

- Few patients benefit from a marker-based treatment optimization (as compared to a random choice)
- The power of the trial is reduced by patients for whom both strategies lead to the same treatment
- Treatment benefits are small

$\rightarrow$ Such a trial would require exceedingly large numbers to show any difference

Ref: Bogaerts et al, Nature Clinical Practice 2006;3:540.



Evaluate Clinical-Pathological risk and 70-gene signature risk in 6000 patients

60%

Clinicopathological and 70-gene both HIGH risk

30%

Clinicopathological and 70-gene risks discordant

10%

Clinicopathological and 70-gene both LOW risk

R1

Use clinicopathological risk to decide Chemo or not

Use 70-gene signature risk to decide Chemo or not

Clin-Path High
70-gene Low: CTx

Clin-Path Low
70-gene High: no CTx

Clin-Path High
70-gene Low: no Ctx

Clin-Path Low
70-gene High: Ctx

Chemotherapy 4350 patients

R2

Anthracycline -based

Taxane Capecitabine-based

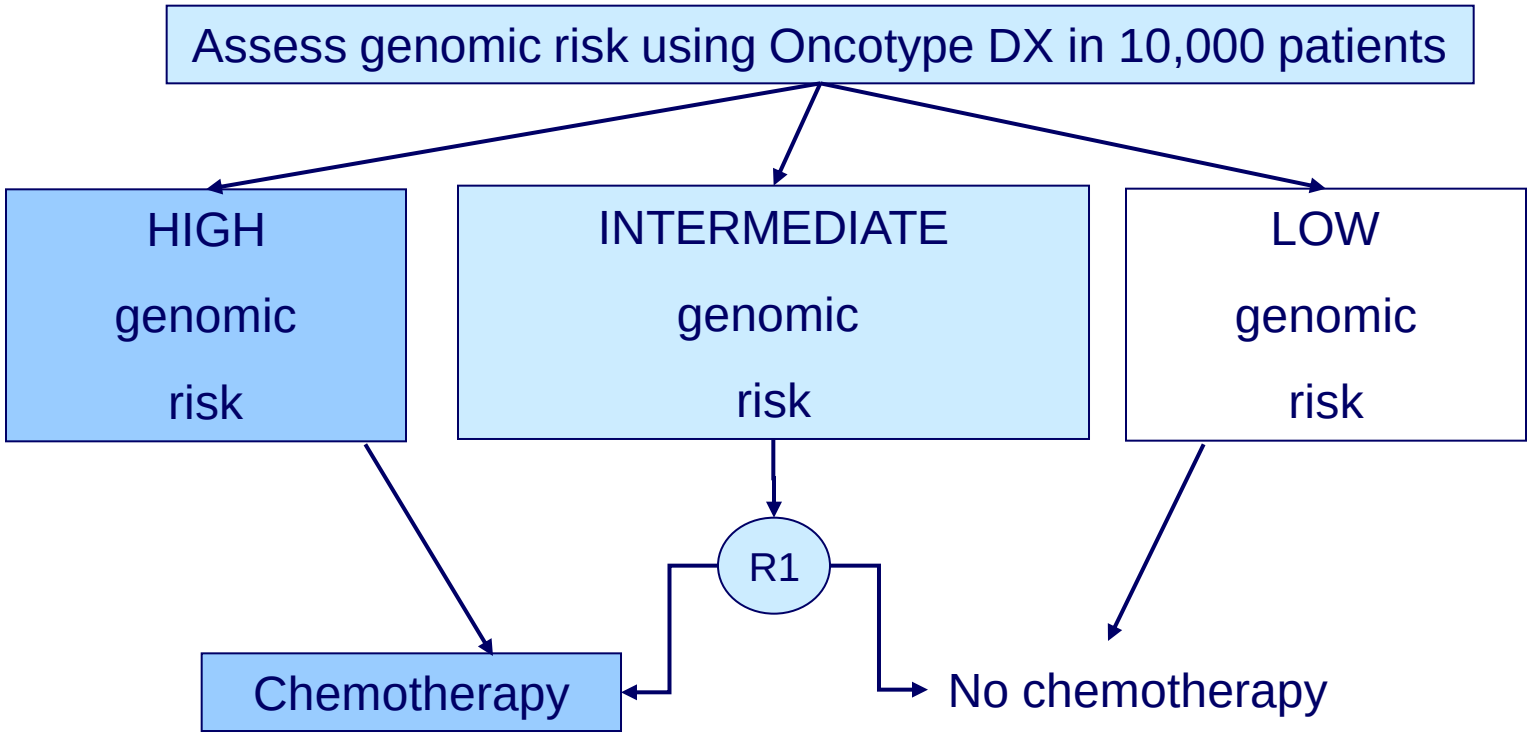
Endocrine therapy (≤ 6000 patients)

R3

2yrs Tam → 5yrs Letrozole

7yrs Letrozole

THE TAILOR-X TRIAL



3. PREDICTIVE BIOMARKERS

Potential uses:

- Patient stratification for trials
- Patient selection for trials
- Treatment decisions

Difficulties:

- Is biomarker truly predictive?

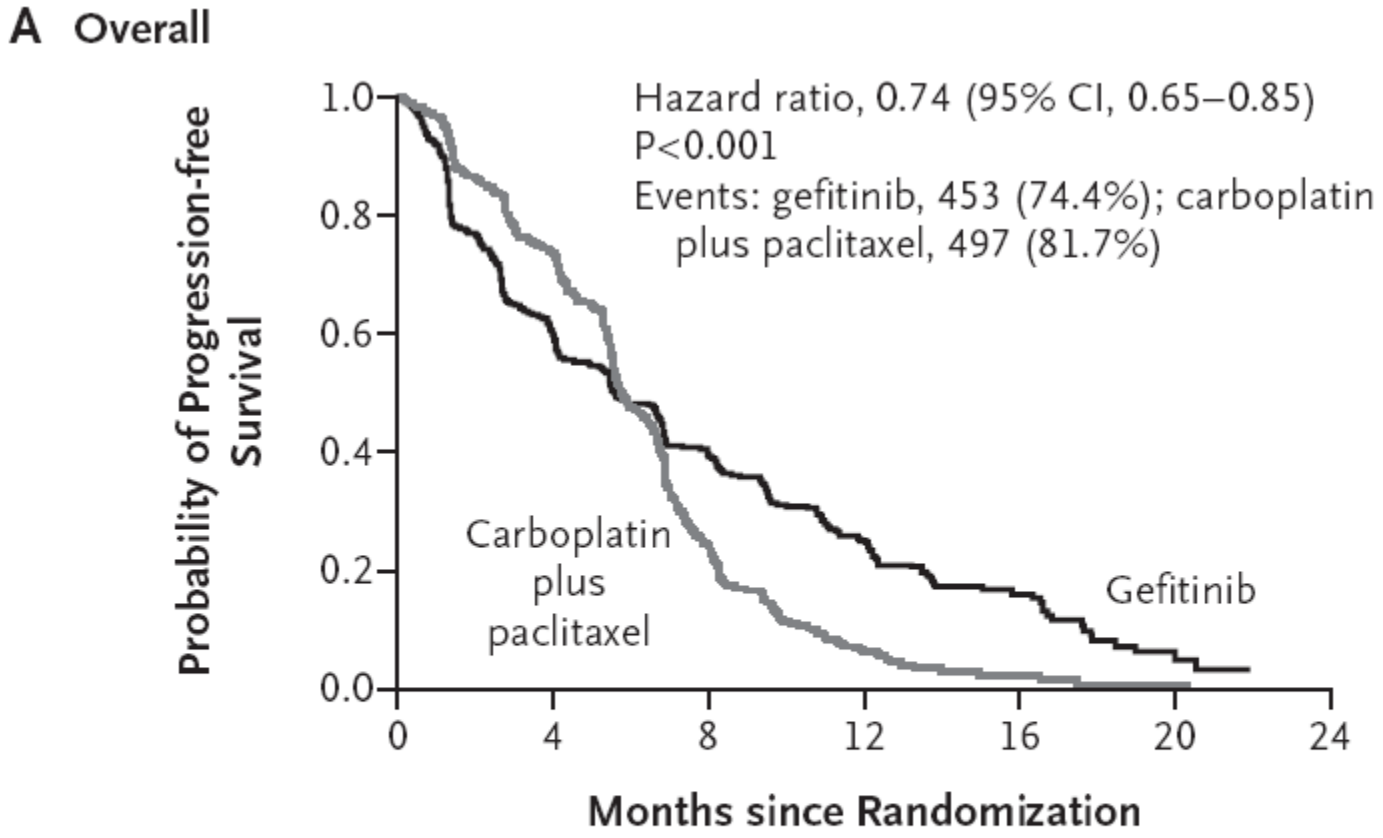
A PREDICTIVE MARKER IN NON-SMALL CELL LUNG CANCER

The NEW ENGLAND JOURNAL of MEDICINE

Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma

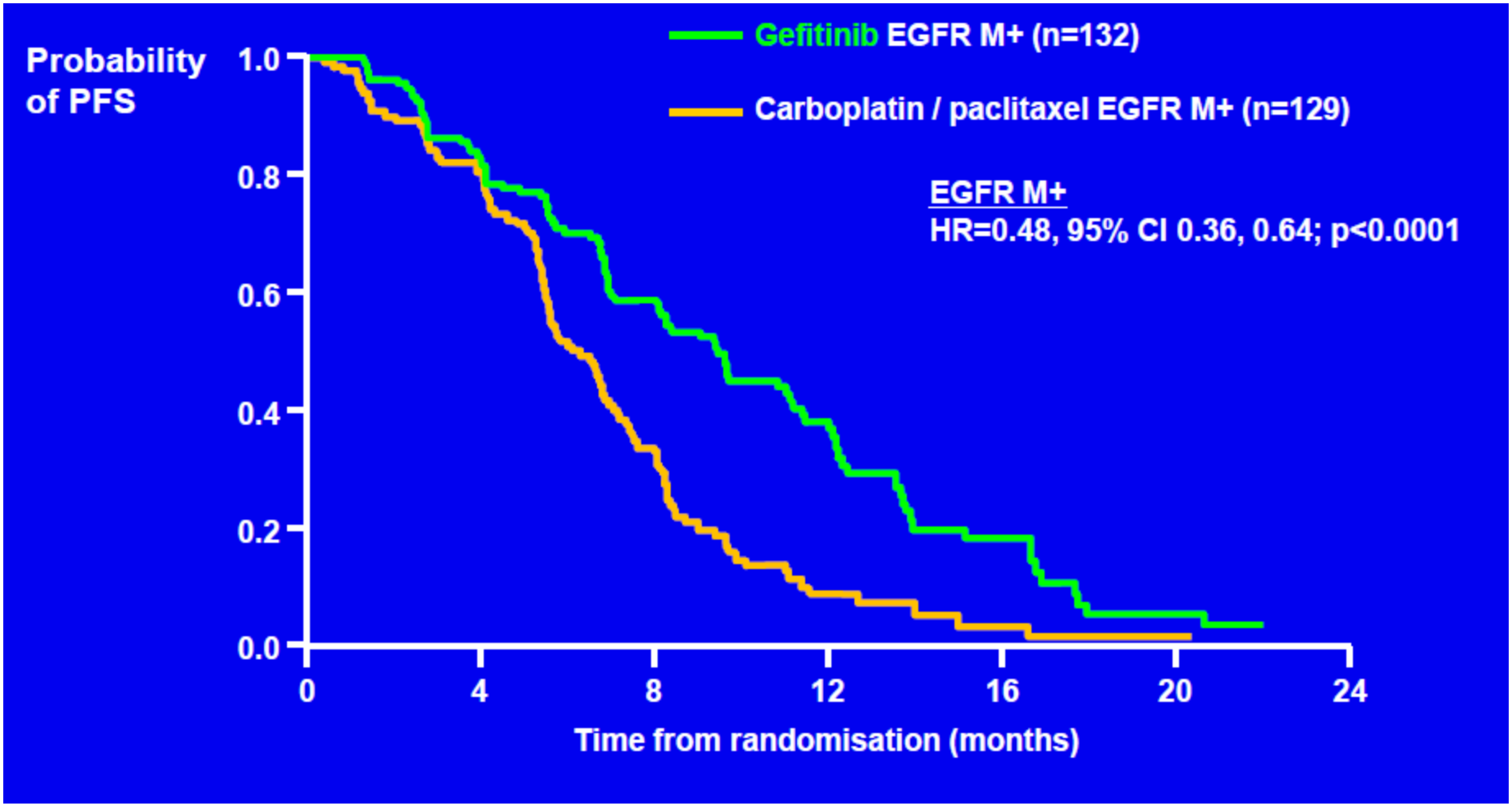
Tony S. Mok, M.D., Yi-Long Wu, M.D., F.A.C.S., Sumitra Thongprasert, M.D., Chih-Hsin Yang, M.D., Ph.D., Da-Tong Chu, M.D., Nagahiro Saijo, M.D., Ph.D., Patrapim Sunpaweravong, M.D., Baohui Han, M.D., Benjamin Margono, M.D., Ph.D., F.C.C.P., Yukito Ichinose, M.D., Yutaka Nishiwaki, M.D., Ph.D., Yuichiro Ohe, M.D., Ph.D., Jin-Ji Yang, M.D., Busyamas Chewaskulyong, M.D., Haiyi Jiang, M.D., Emma L. Duffield, M.Sc., Claire L. Watkins, M.Sc., Alison A. Armour, F.R.C.R., and Masahiro Fukuoka, M.D., Ph.D.

A PREDICTIVE MARKER IN NON-SMALL CELL LUNG CANCER



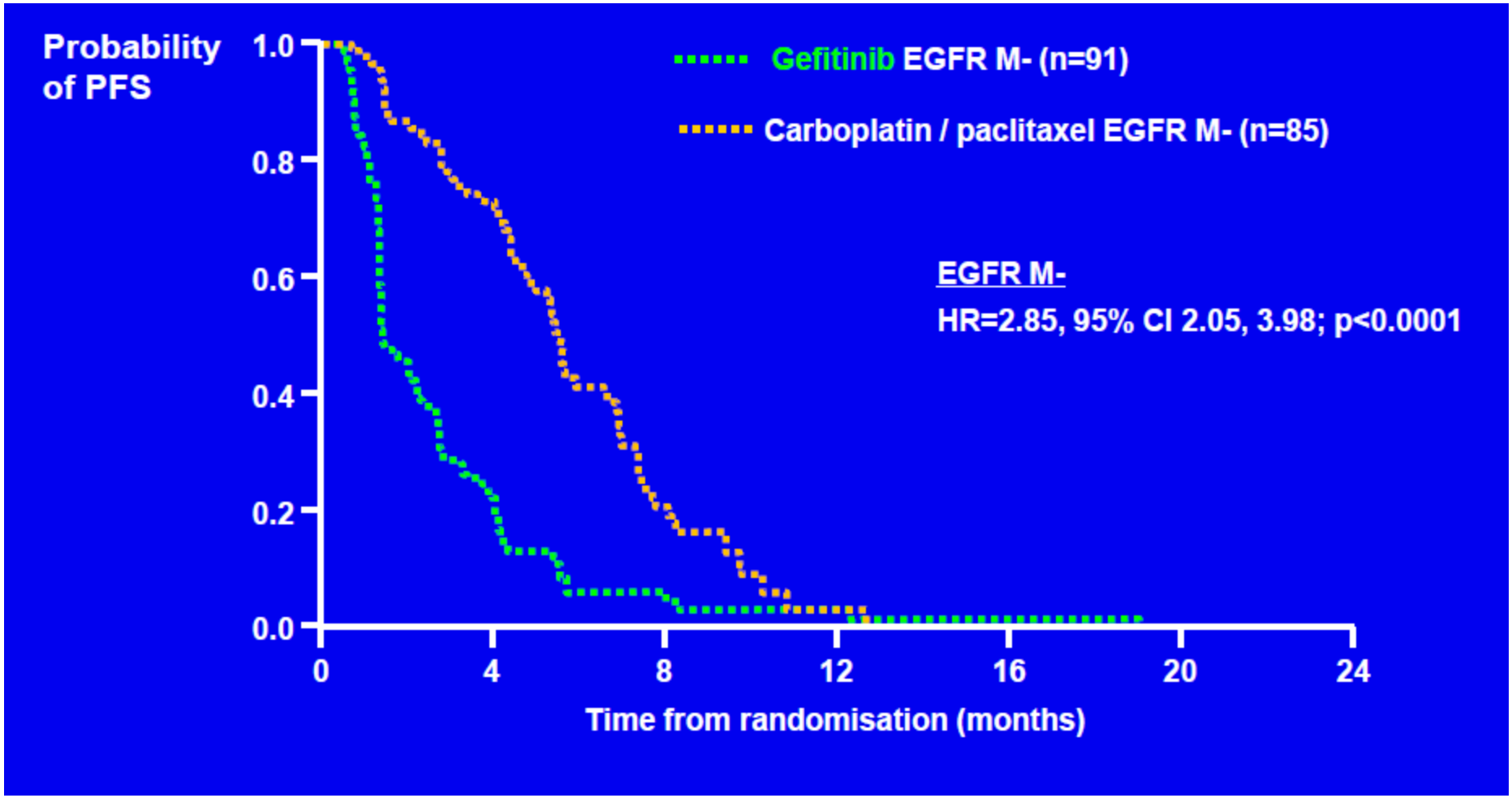
Ref: Mok et al, NEJM 2009;361:947

A PREDICTIVE MARKER IN NON-SMALL CELL LUNG CANCER



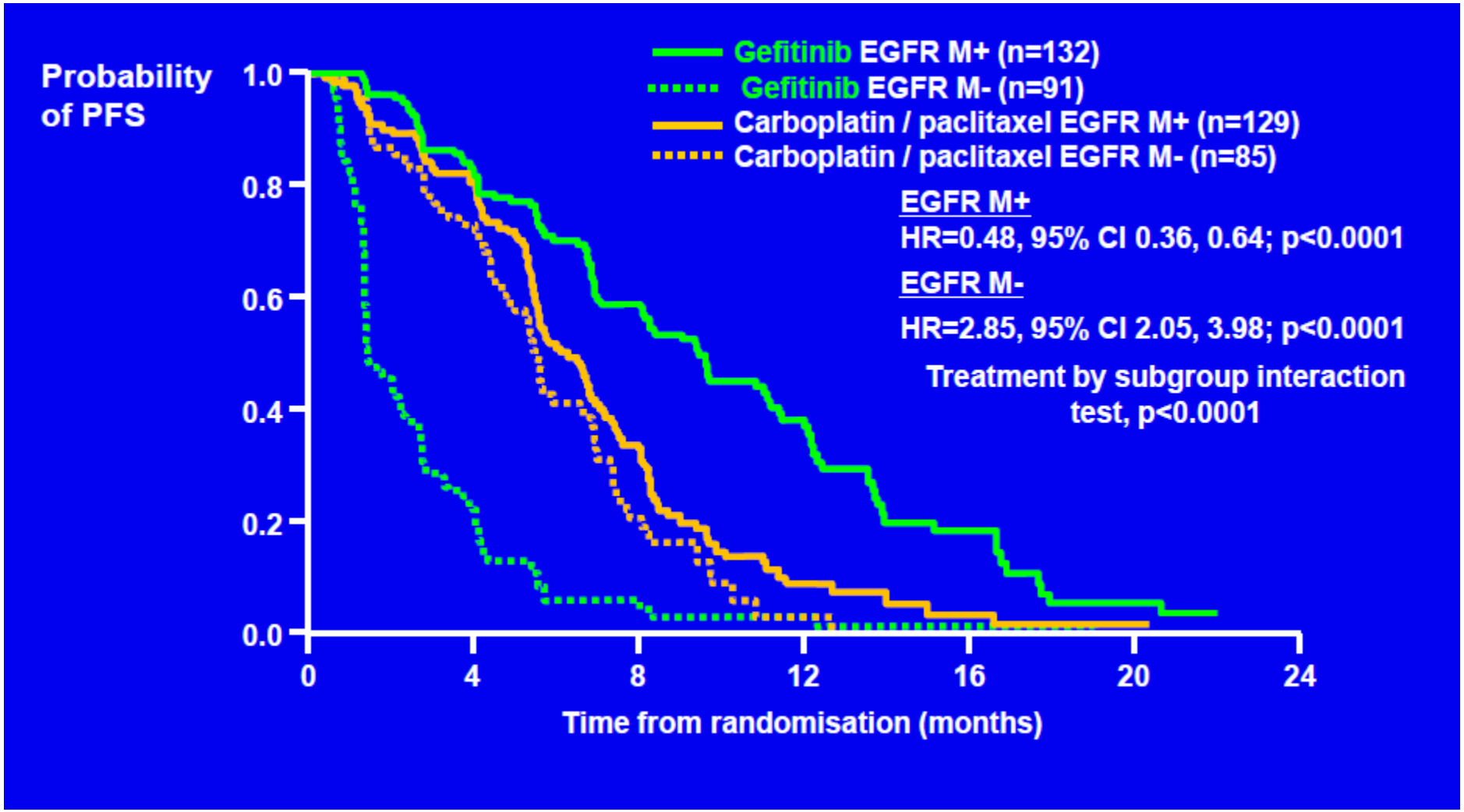
Ref: Slides by courtesy of Astra-Zeneca

A PREDICTIVE MARKER IN NON-SMALL CELL LUNG CANCER



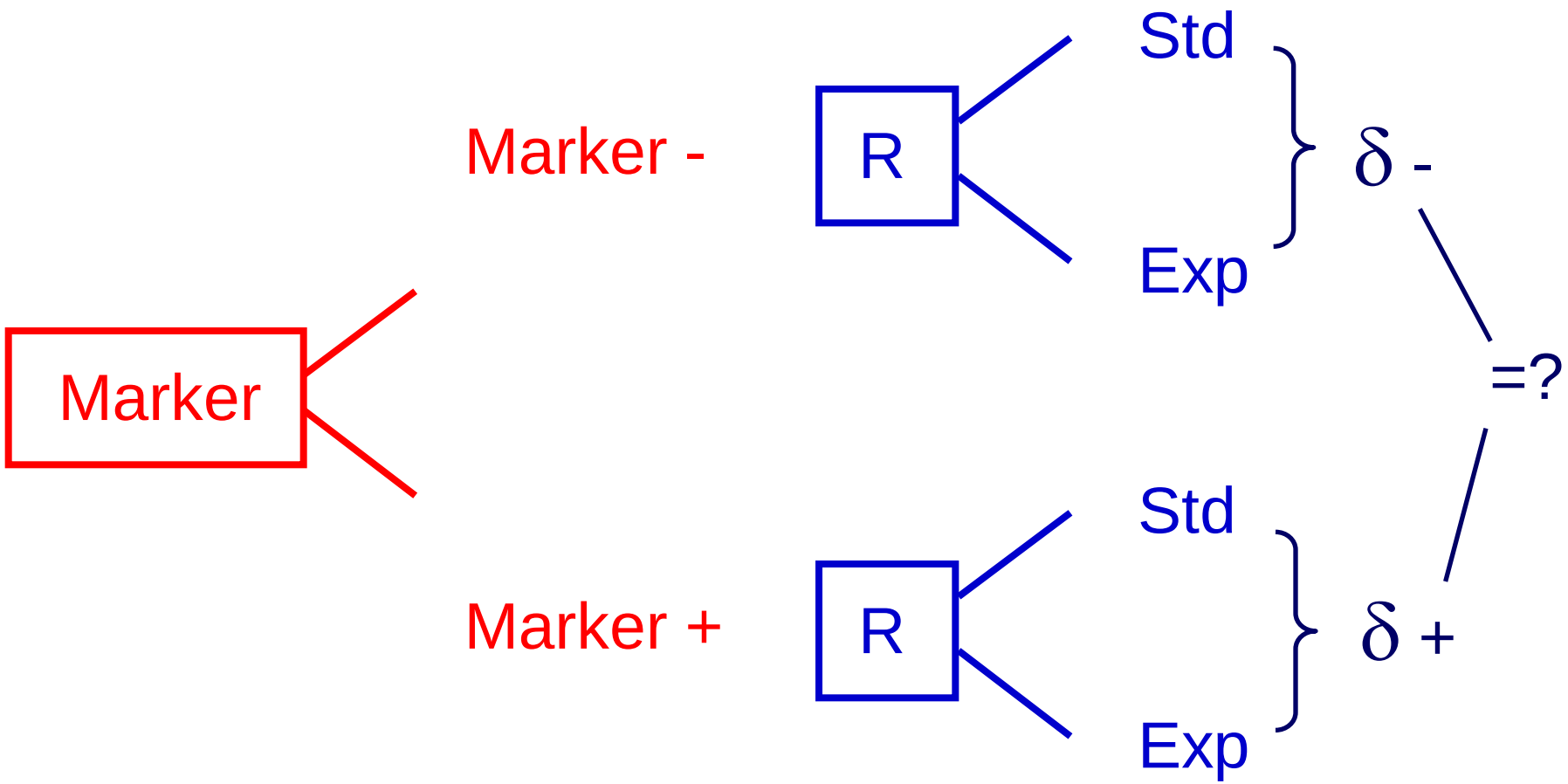
Ref: Slides by courtesy of Astra-Zeneca

A PREDICTIVE MARKER IN NON-SMALL CELL LUNG CANCER



Ref: Slides by courtesy of Astra-Zeneca

TRIAL DESIGN TO VALIDATE PREDICTIVE MARKER



PROBLEM WITH PREDICTIVE MARKER VALIDATION TRIALS

- Marker often unknown or poorly defined (e.g. EGFR mutations in NSCLC, *KRAS* mutations in colorectal cancer) for prospective stratification
- The power of the “interaction test” is low

→ Such a trial would require very large numbers to conclude to a statistically significant interaction *unless* a sensitive biomarker was used as the outcome of interest

→ Perhaps different hypotheses should be tested?

Ref: Buyse et al, Nature Reviews Clinical Oncology 2010 (in press).

4. SURROGATE BIOMARKERS

Potential uses:

- Assessment of treatment effects on earlier / more sensitive endpoint (based on biomarker) than the ultimate clinical endpoint of interest

Difficulties:

- Does treatment effect on biomarker reliably predict treatment effect on clinical endpoint?

Ref: *Burzykowski, Molenberghs, Buyse.*
The Evaluation of Surrogate Endpoints. Springer, Heidelberg, 2005

VALIDATION OF SURROGATE ENDPOINTS

Randomized treatment

Trt

Potential surrogate (« intermediate »)

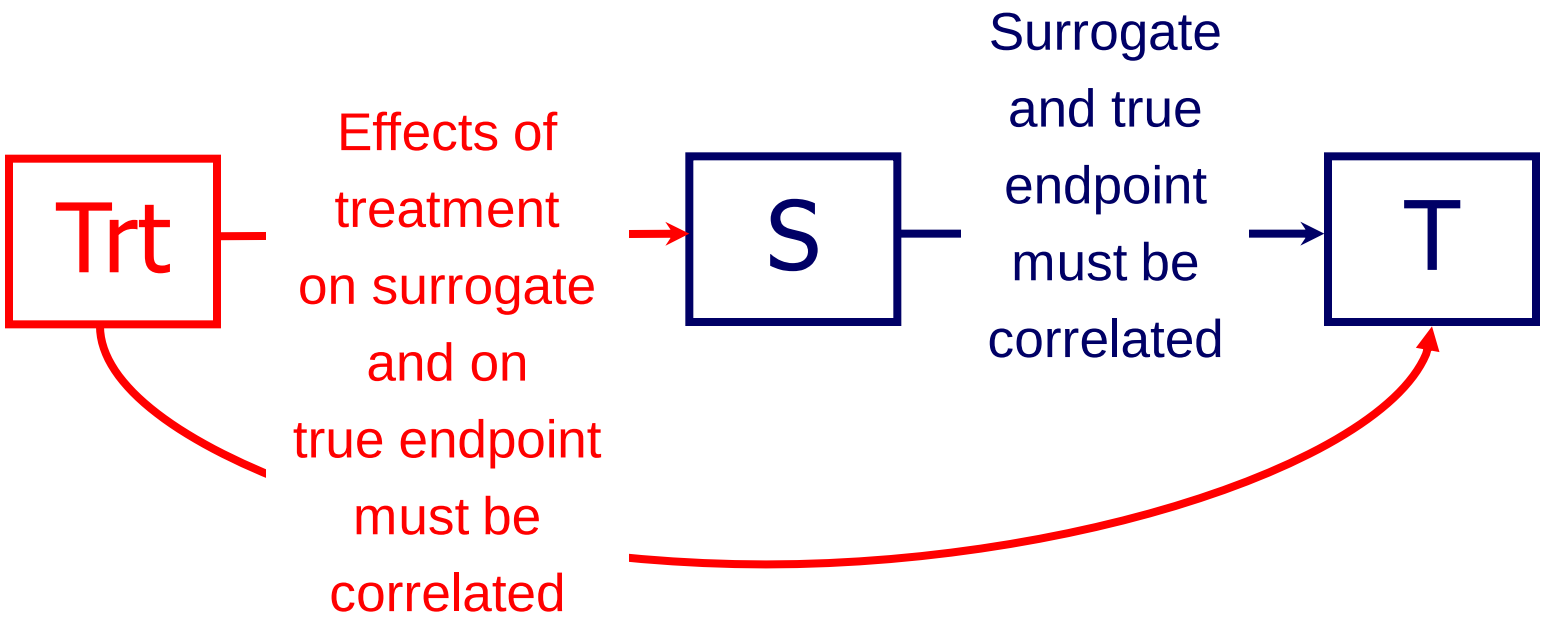
S

True endpoint

T

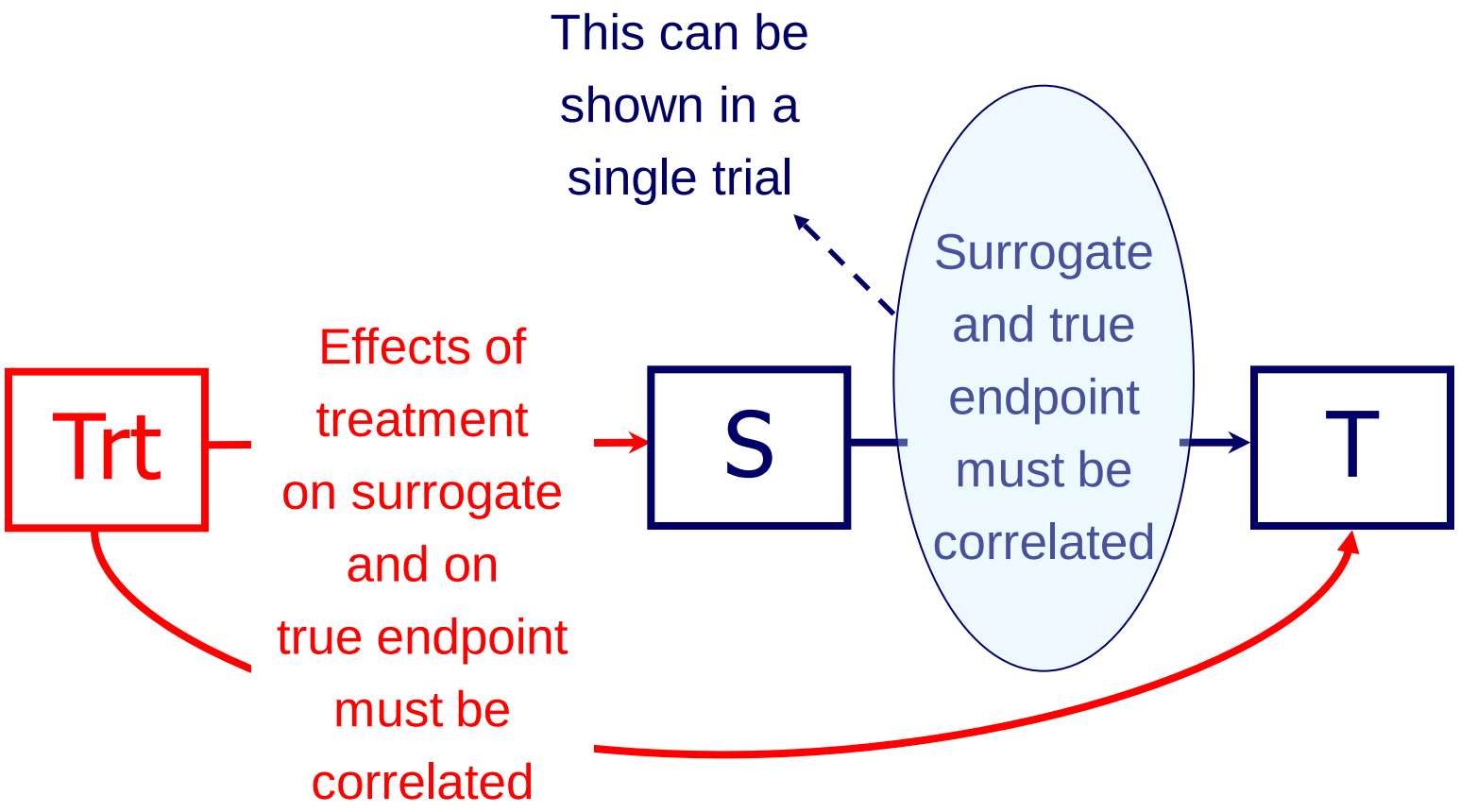
Ref: *Buyse and Molenberghs, Biometrics 1998, **54**: 1014*

VALIDATION OF SURROGATE ENDPOINTS

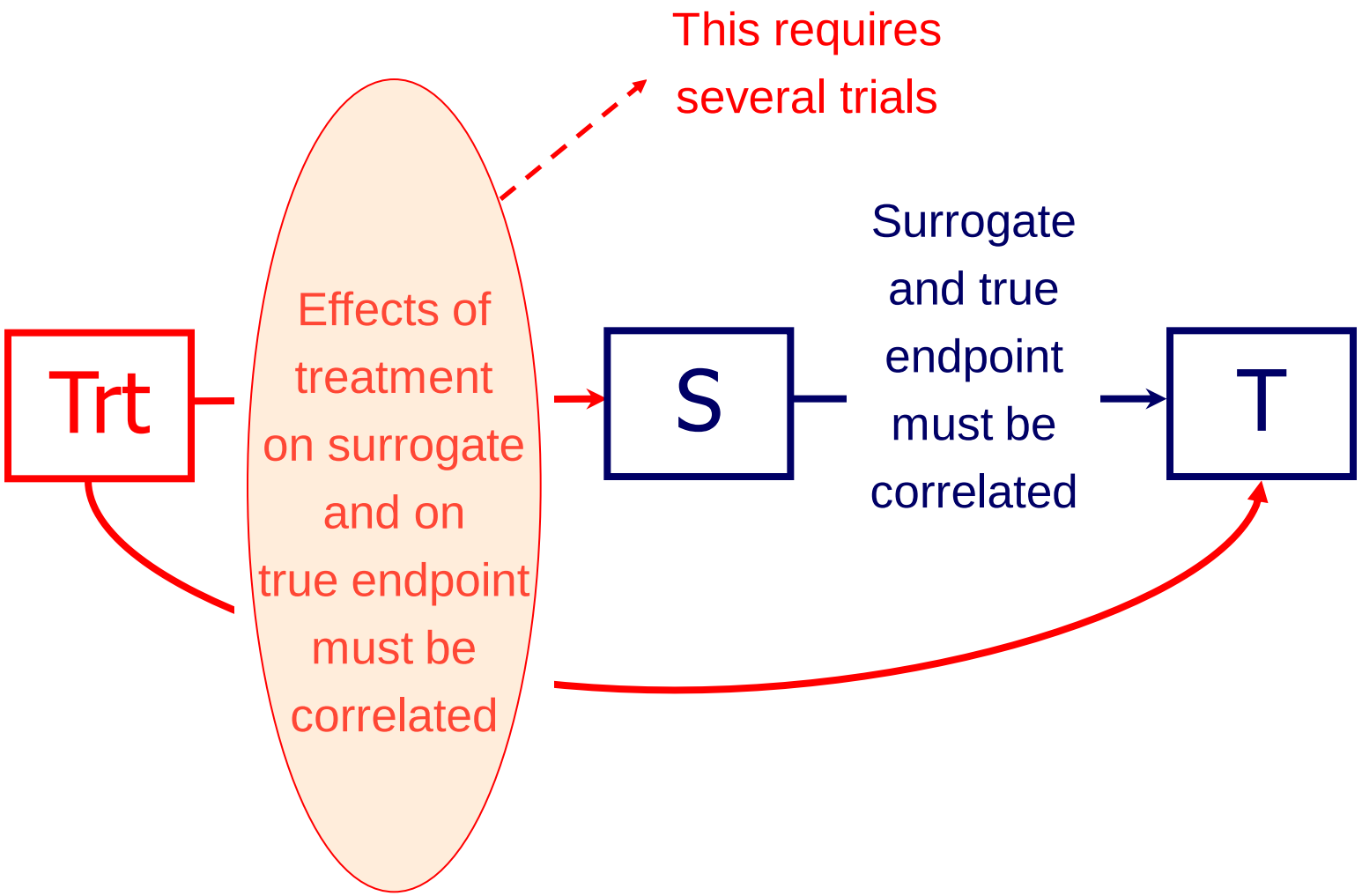


*Ref: Buyse et al, Biostatistics 2000;**1**:49.*

VALIDATION OF SURROGATE ENDPOINTS



VALIDATION OF SURROGATE ENDPOINTS

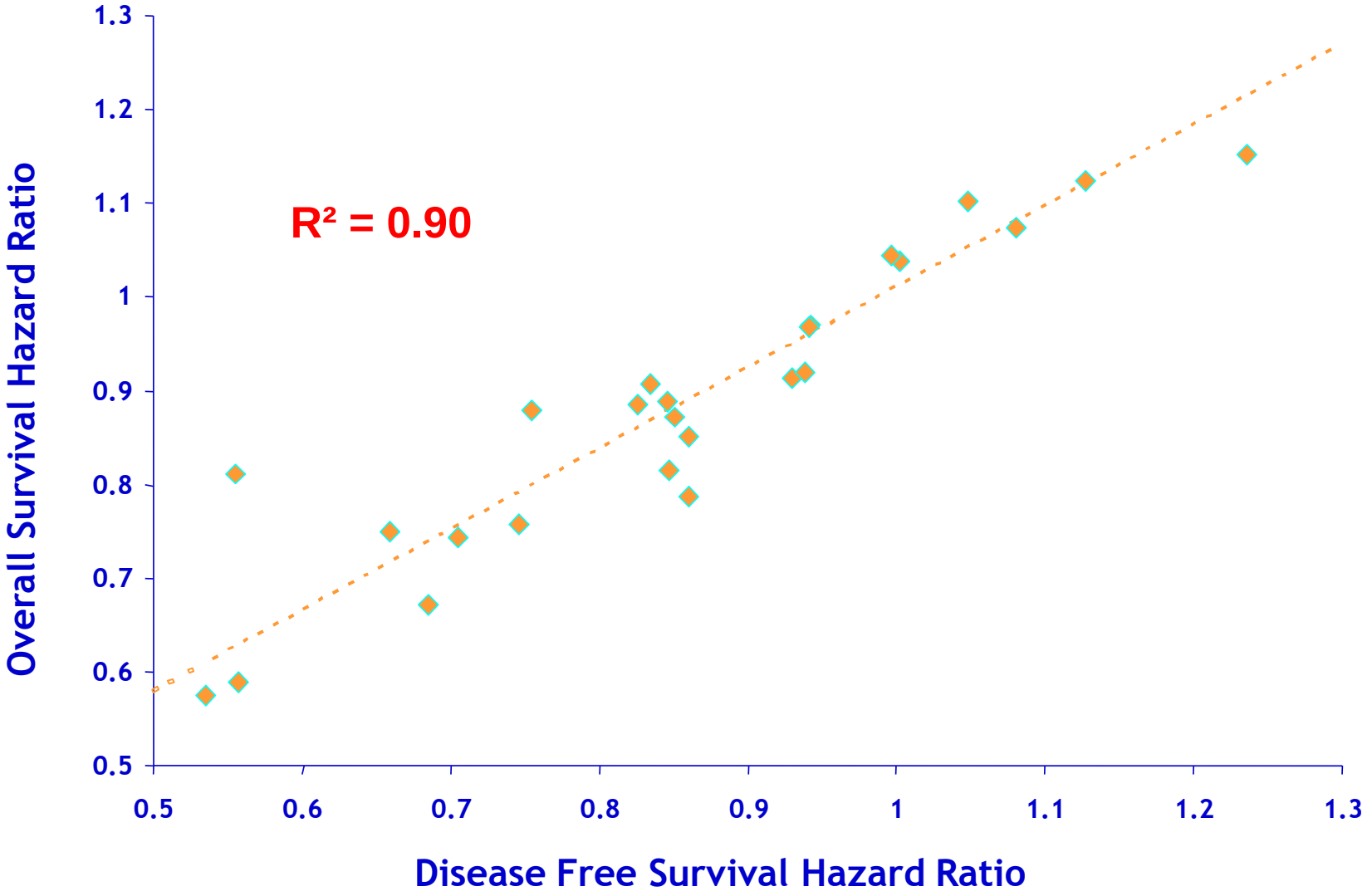


EXAMPLE IN RESECTED COLORECTAL CANCER

- 43 treatment arms in 18 randomized trials (20,898 patients)
  - 9 surgery alone control groups
  - 34 5FU-based experimental treatment groups

*Ref: Sargent et al, JCO 2005;**23**:8664.*

DFS AS SURROGATE FOR OS

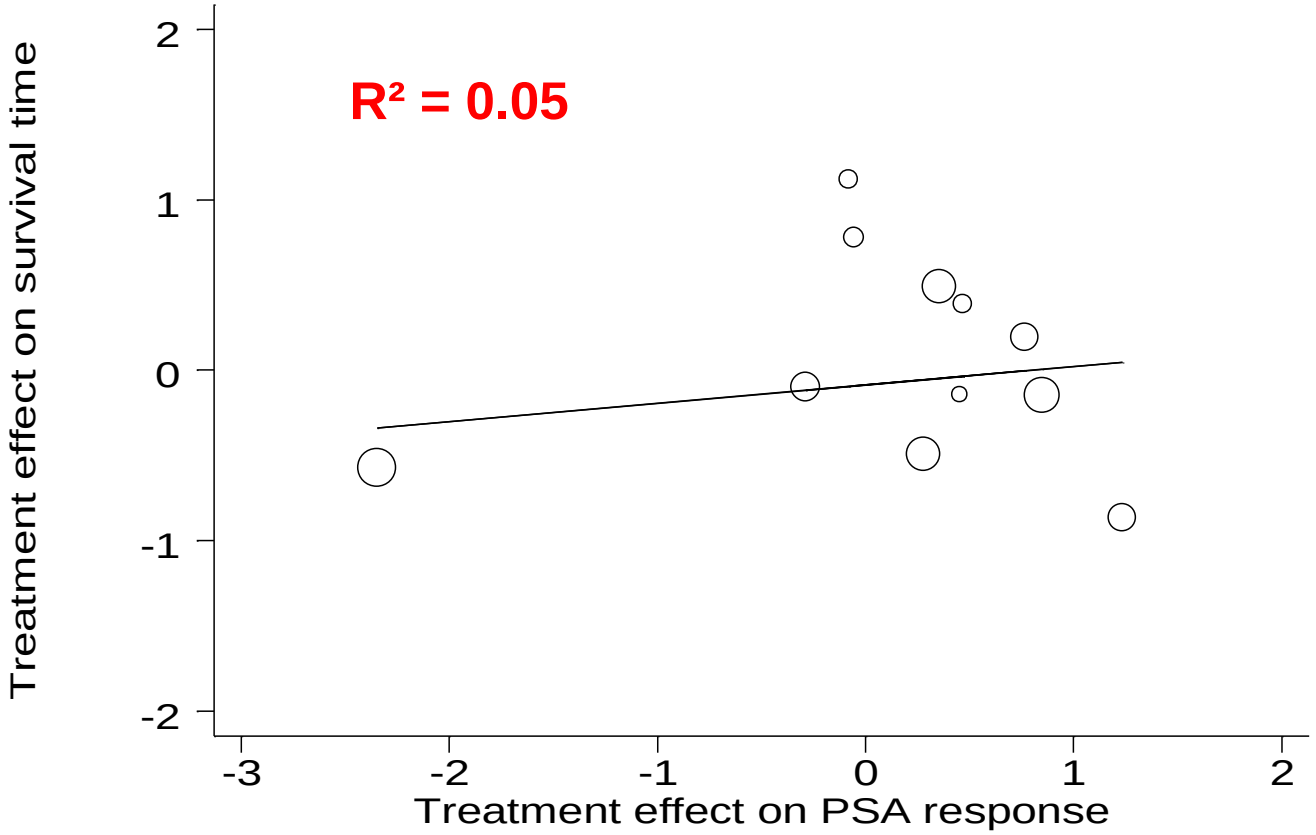


EXAMPLE IN ADVANCED PROSTATE CANCER

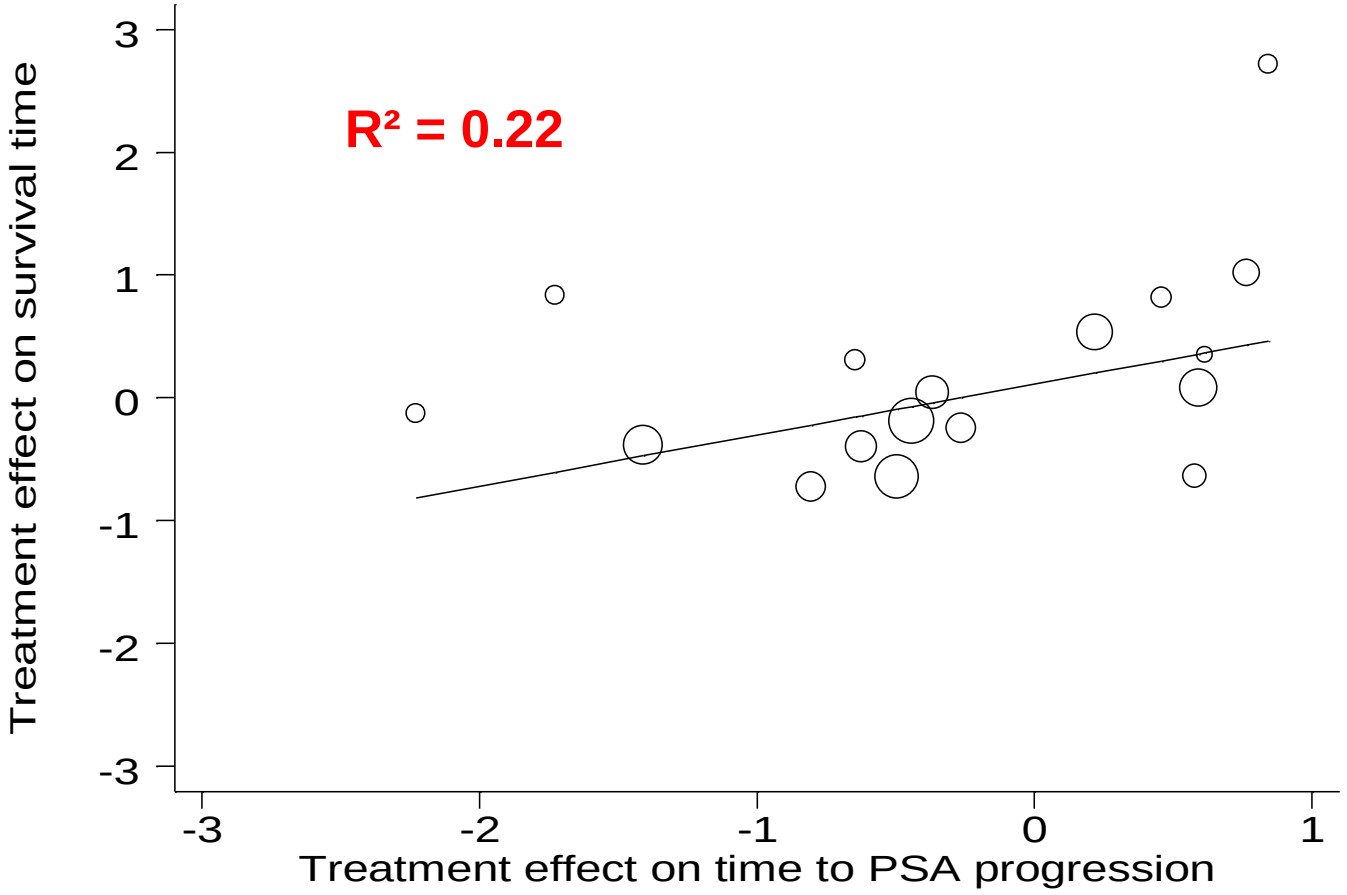
- Two multicentric trials carried out in 19 countries for patients in relapse after first-line endocrine therapy (596 patients)
- Treatments:
  - Experimental (retinoic acid metabolism-blocking agent)
  - Control (anti-androgen)

Ref: Buyse et al, in: Biomarkers in Clinical Drug Development (Bloom JC, ed.): Springer-Verlag, New York, 2003.

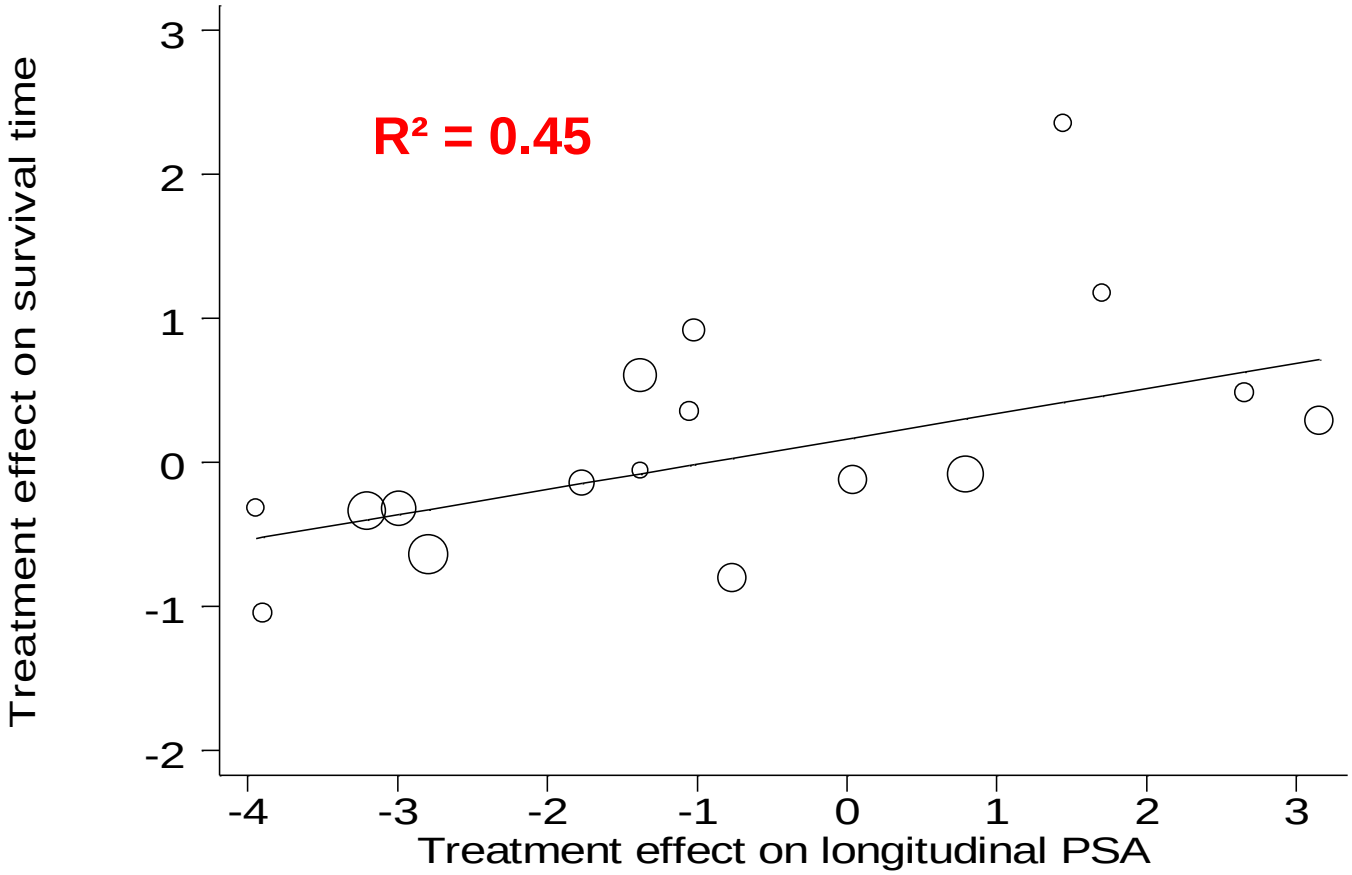
PSA RESPONSE AS SURROGATE FOR SURVIVAL



TIME TO PSA PROGRESSION AS SURROGATE FOR SURVIVAL



LONGITUDINAL PSA AS SURROGATE FOR SURVIVAL



CONCLUSIONS

1. Pharmacodynamic biomarkers
2. Prognostic biomarkers
3. Predictive biomarkers
4. Surrogate biomarkers

} Need large randomized trials