Towards stratified medicine – instead of dichotomization, estimate a treatment effect function for a continuous covariate

Willi Sauerbrei[1], Patrick Royston[2]

[1]IMBI, University Medical Center Freiburg
[2]MRC Clinical Trials Unit, London

wfs@imbi.uni-freiburg.de

Royston, P.; Altman, D. G. (1994): Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametic Modelling. Applied Statistics, 43(3): 429–467.

Altman, D.G., Lausen, B., Sauerbrei, W., Schumacher, M (1994): Dangers of using “Optimal” cutpoints in the evaluation of prognostic factors. Journal of the National Cancer Institute, 86: 829-835.

Sauerbrei, W. (1999): The use of resampling methods to simplify regression models in medical statistics. Applied Statistics, 48: 313-329.

Sauerbrei, W., Royston, P. (1999): Building multivariable prognostic and diagnostic models: Transformation of the predictors by using fractional polynomials’. Journal of the Royal Statistical Society, A. 162: 71-94

Bonetti M.; Gelber R. D. (2000): A graphical method to assess treatment-covariate interactions using the Cox model on subsets of the data. Statistics in Medicine, 19: 2595–2609.

Royston, P., Sauerbrei, W. (2004): A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. Statistics in Medicine, 23: 2509-2525.

Sauerbrei, W., Royston, P., Zapien, K. (2007): Detecting an interaction between treatment and a continuous covariate: a comparison of two approaches. Computational Statistics and Data Analysis, 51: 4054-4063.

Royston P, Sauerbrei W (2008): Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables, Wiley

Royston P., Sauerbrei W. (2009): Two techniques for investigating interactions between treatment and continuous covariates in clinical trials. The Stata Journal, 9: 230-251

Overview

- Multivariable model-building
- Continuous variables
  - Problems of cutpoints
  - Fractional polynomials
- Interactions with continuous variables
  - problems of cutpoints
  - MFPI
  - STEPP
- Reporting

Observational Studies

Several variables, mix of continuous and (ordered) categorical variables

Different situations:

- prediction
- explanation
- confounders only

Explanation is the main interest here:

- Identify variables with (strong) influence on the outcome
- Determine functional form (roughly) for continuous variables

The issues are very similar in different types of regression models (linear regression model, GLM, survival models ...)

Use subject-matter knowledge for modelling ...

... but for some variables, data-driven choice inevitable

Regression models

$X=(X_1, ...,X_p)$ covariate, prognostic factors

g(x) = $ß_1 X_1 + ß_2 X_2 +...+ ß_p X_p$ (assuming effects are linear)

normal errors (linear) regression model

Y normally distributed

E (Y|X) = $ß_0$ + g(X)

Var (Y|X) = $\sigma^2 I$

logistic regression model

Y binary

Logit P (Y|X) = **ln** $\frac{P(Y=1|X)}{P(Y=0|X)} = \beta_0 + g(X)$

survival times

T survival time (partly censored)

Incorporation of covariates

$\lambda(t|X) = \lambda_0(t)\exp(g(X))$

Central issue

To select or not to select (full model)?

Which variables to include?

How to model continuous variables?

Continuous variables – The problem

"Quantifying epidemiologic risk factors using non-parametric regression: model selection remains the greatest challenge"

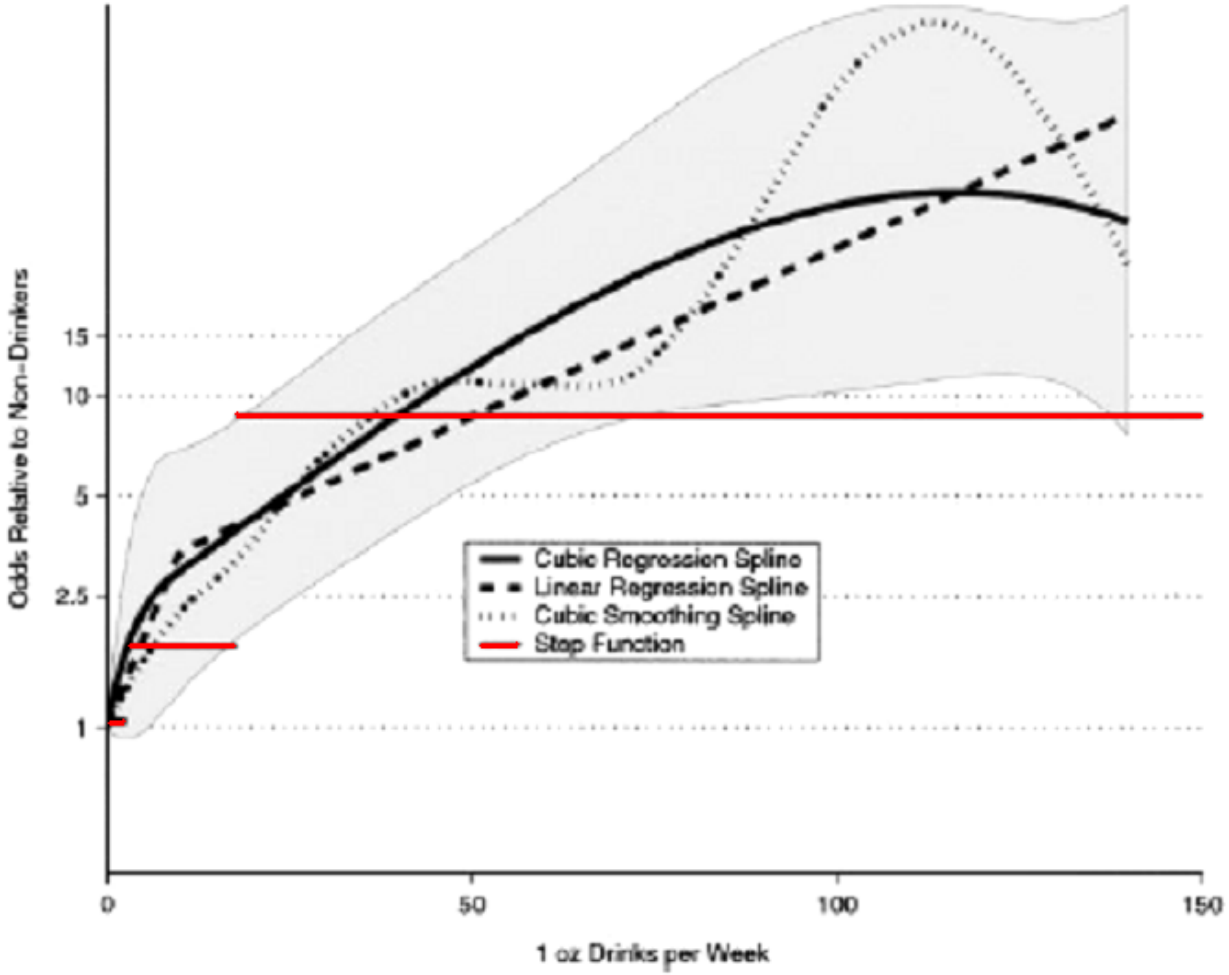
Rosenberg PS et al, Statistics in Medicine 2003; 22:3369-3381

Discussion of issues in (univariate) modelling with splines

Trivial nowadays to *fit* almost any model

To *choose* a good model is much harder

Alcohol consumption as risk factor for oral cancer



Rosenberg et al, StatMed 2003

Multivariable models – methods for variable selection

Full model

- variance inflation in the case of multicollinearity

Stepwise procedures $\Rightarrow$ prespecified (α_{in}, α_{out}) and actual significance level?

- forward selection (FS)
- stepwise selection (StS)
- backward elimination (BE)

All subset selection $\Rightarrow$ which criteria?

- C_p Mallows $= (SSE / \hat{\sigma}^2) - n + p\ 2$
- AIC Akaike Information Criterion $= n \ln (SSE / n) + p\ 2$
- BIC Bayes Information Criterion $= n \ln (SSE / n) + p \ln(n)$

fit penalty

Combining selection with Shrinkage

Bayes variable selection

Recommendations???

Central issue: MORE OR LESS COMPLEX MODELS?

Backward elimination is a sensible approach

- Significance level can be chosen
- Reduces overfitting

Of course required

- Checks
- Sensitivity analysis
- Stability analysis

Continuous variables – what functional form?

Traditional approaches

a) Linear function

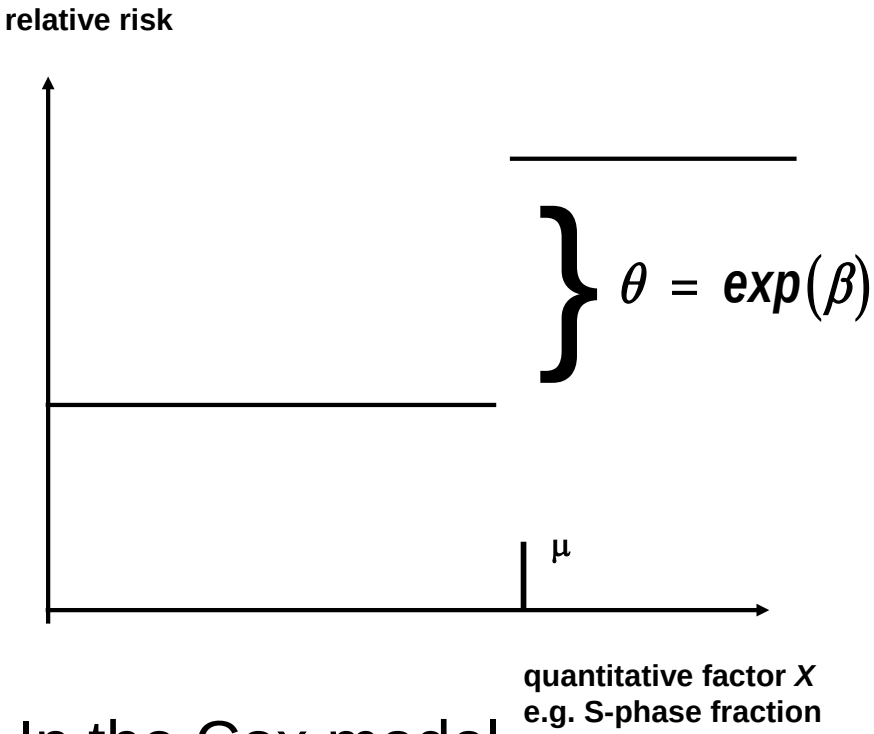
- may be inadequate functional form
- misspecification of functional form may lead to wrong conclusions

b) ‘best‘ ‘standard‘ transformation

c) Step function (categorial data)

- Loss of information
- How many cutpoints?
- Which cutpoints?
- Bias introduced by outcome-dependent choice

Step function - the cutpoint problem



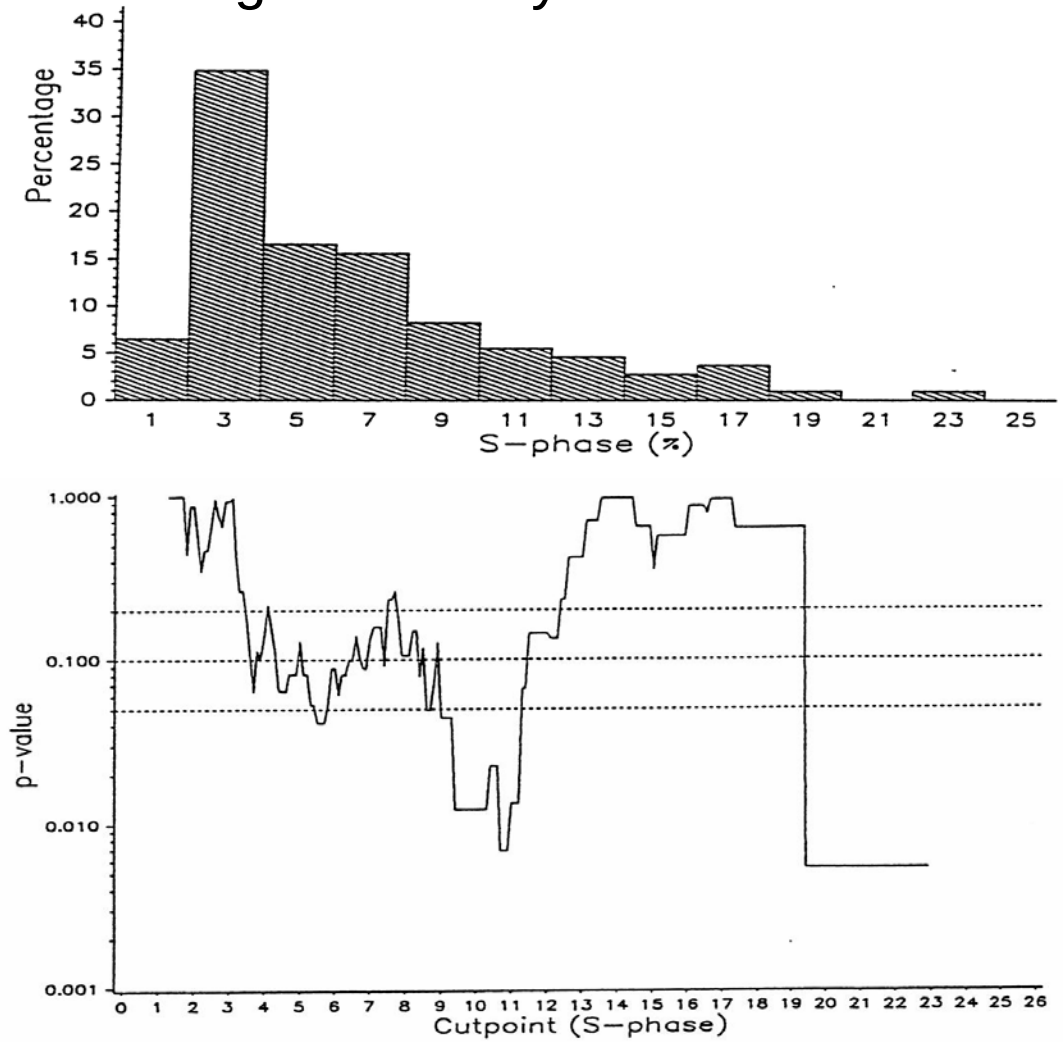
In the Cox model

$$\lambda(t|X > \mu) = \exp \beta \, \lambda(t|X \leq \mu)$$

$\hat{\mu}$: estimated cutpoint for the comparison of patients with X above and below μ.

Searching for optimal cutpoint minimal p-value approach

SPF in Freiburg DNA study

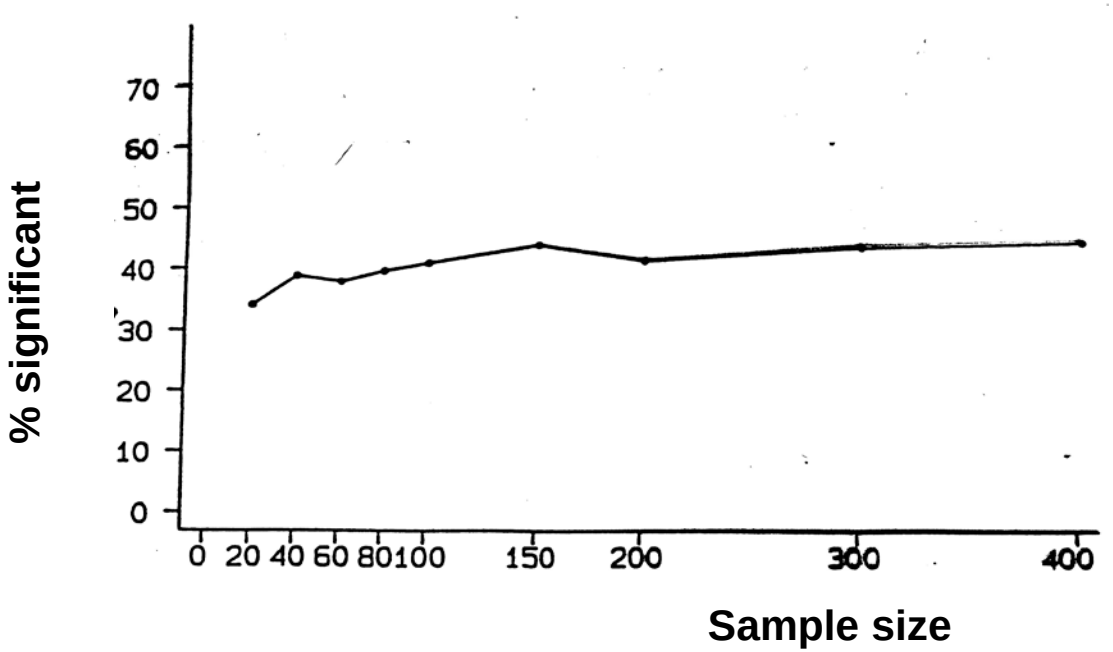


Problem

multiple testing => inflated type I error

Searching for optimal cutpoint

Inflation of type I errors



Simulation study

Type I error about 40% instead of 5%, does not disappear with increased sample size (in contrast to type II error)

Severe bias of estimated effect

Different cutpoints in each study

Step function – biologically plausible??

Example 1: Prognostic factors

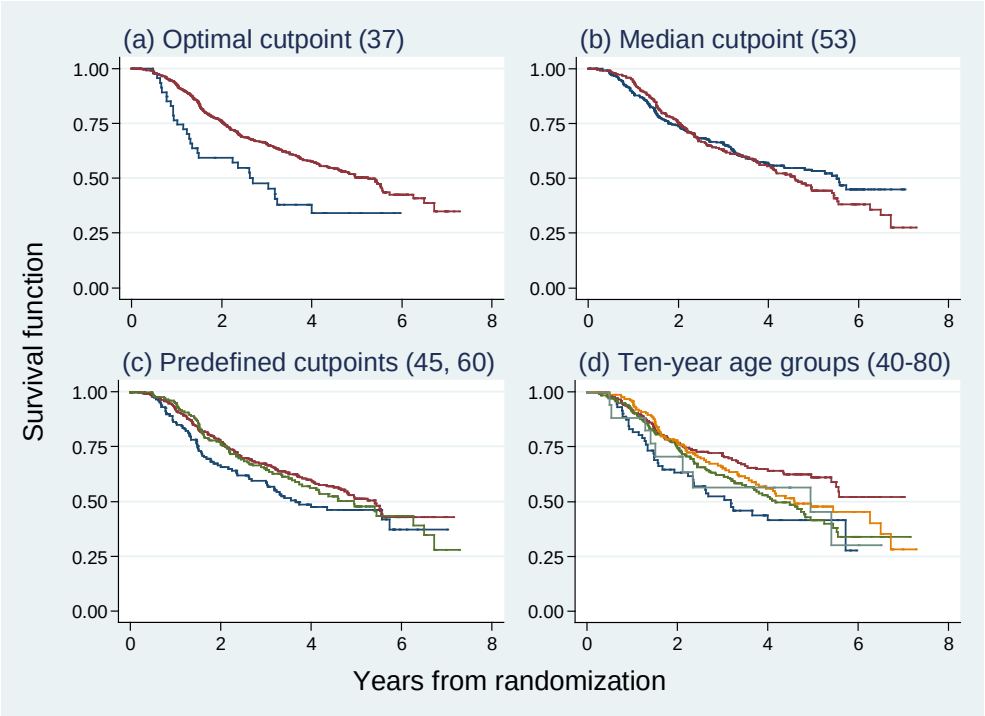
GBSG-study in node-positive breast cancer

299 events for recurrence-free survival time (RFS) in **686** patients with complete data

7 prognostic factors, of which **5** are continuous

Tamoxifen yes/no

Age as prognostic factor – cutpoint analyses



The youngest group is always in blue.

(a)‘Optimal’ (37 years); HR (older vs younger) 0.54, p= 0.004

(b) median (53 years); HR (older vs younger) 1.1, p= 0.4

(c) predefined from earlier analyses (45, 60years);

(d) popular (10-year groups)

Dichotomizing continuous predictors in multiple regression: a bad idea

Patrick Royston[1,*,†], Douglas G. Altman[2] and Willi Sauerbrei[3]

StatMed 2006, 25:127-141

Fractional polynomial models

- *Fractional* polynomial of degree *2* with powers p = (p_1,p_2) is defined as

  $FP2 = \beta_1 X^{p1} + \beta_2 X^{p2}$

- Powers p are taken from a predefined set

  $S = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$

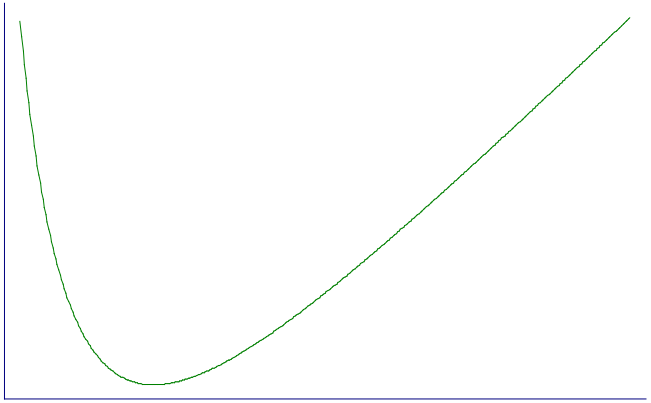
  0- logX

- Repeated powers (p1=p2): $\beta_1 X^{p1} + \beta_2 X^{p1} \log X$
- 8FP1, 36FP2 models

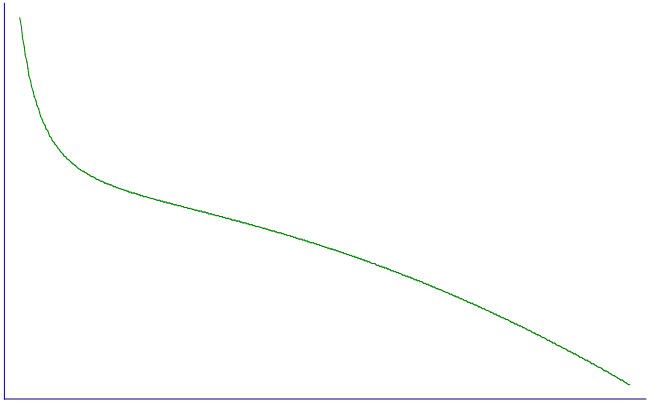
Example $FP2 = \beta_1 X^{0.5} + \beta_2 X^{3}$

Examples of FP2 curves - varying powers

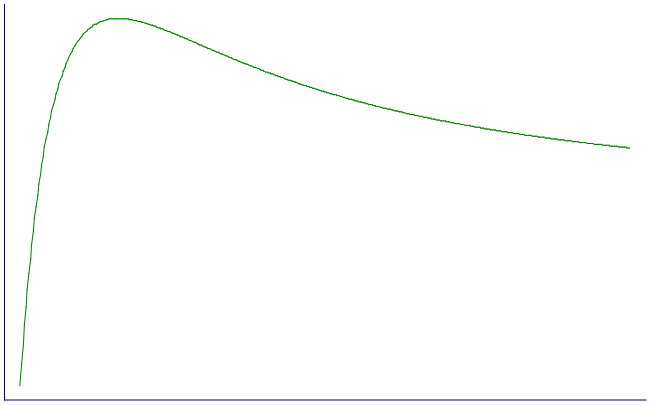
(-2, 1)



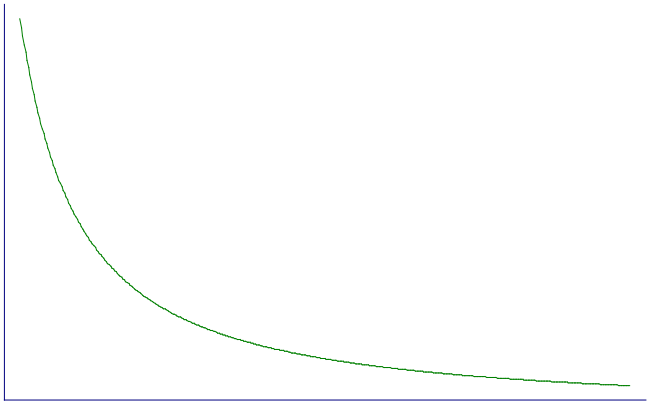
(-2, 2)



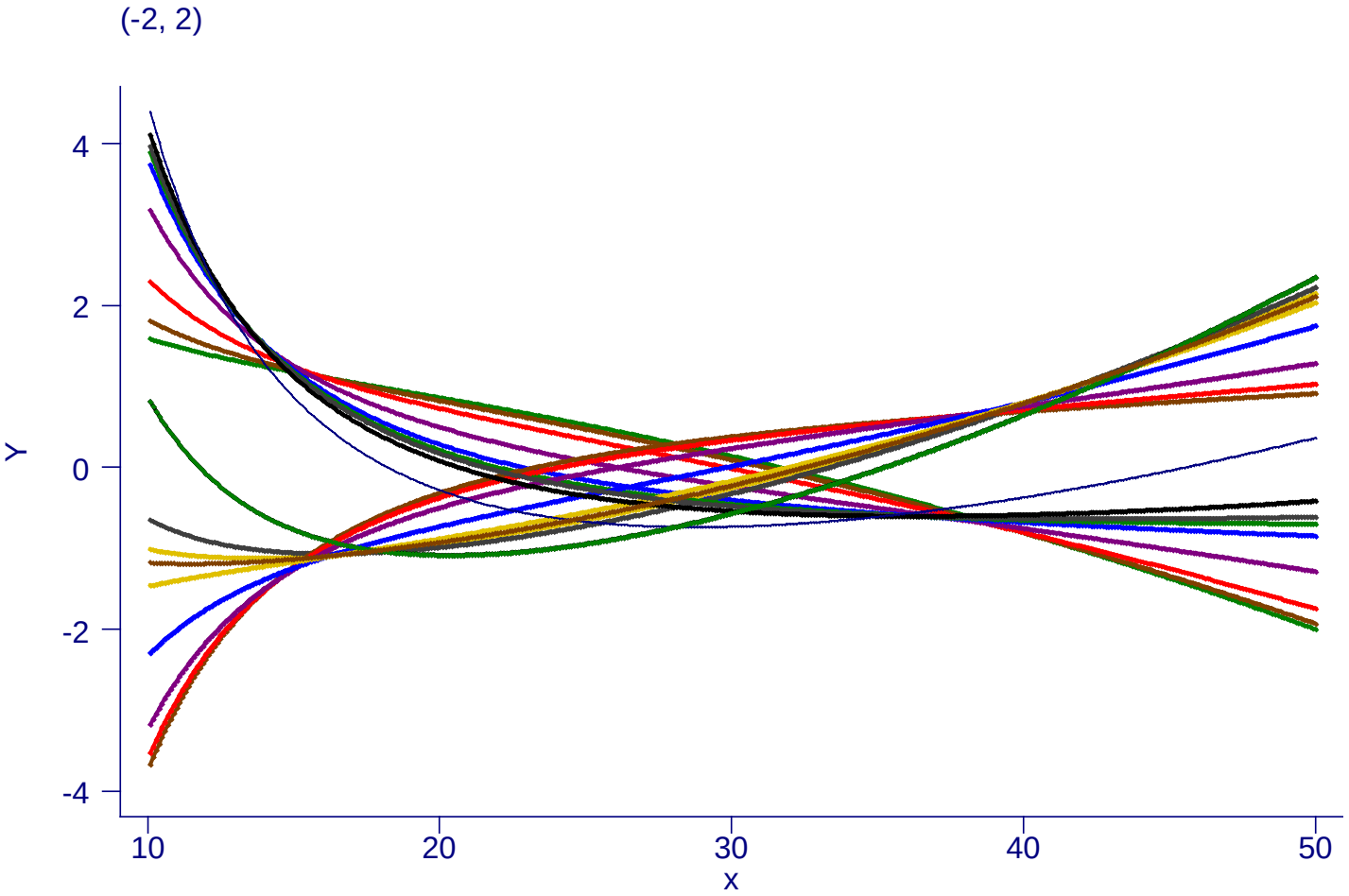
(-2, -2)



(-2, -1)



Examples of FP2 curves - single power, different coefficients



Our philosophy of function selection

- Prefer simple (linear) model
- Use more complex (non-linear) FP1 or FP2 model if indicated by the data
- Contrasts to more local regression modelling
  - Already starts with a complex model

FP analysis for the effect of age

Degree 1		Degree 2							
Power	Model chi-square	Powers		Model chi-square	Powers		Model chi-square	Powers		Model chi-square
-2	6.41	-2	-2	17.09	-1	1	15.56	0	2	11.45
-1	3.39	-2	-1	17.57	-1	2	13.99	0	3	9.61
-0.5	2.32	-2	-0.5	17.61	-1	3	12.37	0.5	0.5	13.37
0	1.53	-2	0	17.52	-0.5	-0.5	16.82	0.5	1	12.29
0.5	0.97	-2	0.5	17.30	-0.5	0	16.18	0.5	2	10.19
1	0.58	-2	1	16.97	-0.5	0.5	15.41	0.5	3	8.32
2	0.17	-2	2	16.04	-0.5	1	14.55	1	1	11.14
3	0.03	-2	3	14.91	-0.5	2	12.74	1	2	8.99
		-1	-1	17.58	-0.5	3	10.98	1	3	7.15
		-1	-0.5	17.30	0	0	15.36	2	2	6.87
		-1	0	16.85	0	0.5	14.43	2	3	5.17
		-1	0.5	16.25	0	1	13.44	3	3	3.67

Function selection procedure (FSP)

Effect of age at 5% level?

	χ^2	df	p-value
Any effect?			
Best FP2 versus null	17.61	4	0.0015
Linear function suitable?			
Best FP2 versus linear	17.03	3	0.0007
FP1 sufficient?			
Best FP2 vs. best FP1	11.20	2	0.0037

Many predictors – MFP

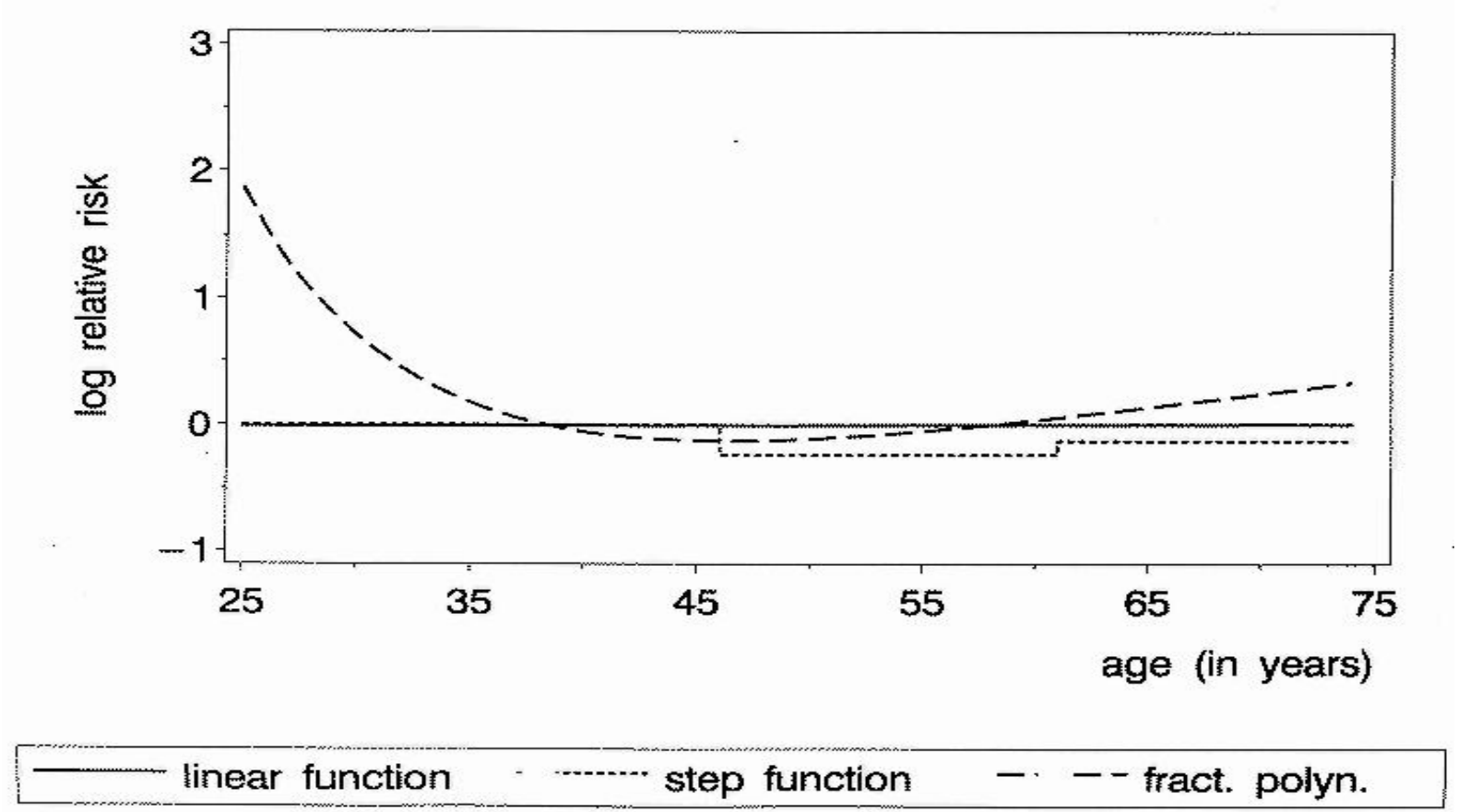
With many continuous predictors selection of best FP for each becomes more difficult $\rightarrow$
MFP algorithm as a standardized way to variable and function selection

(usually binary and categorical variables are also available)

MFP algorithm combines

- backward elimination with
- FP function selection procedures

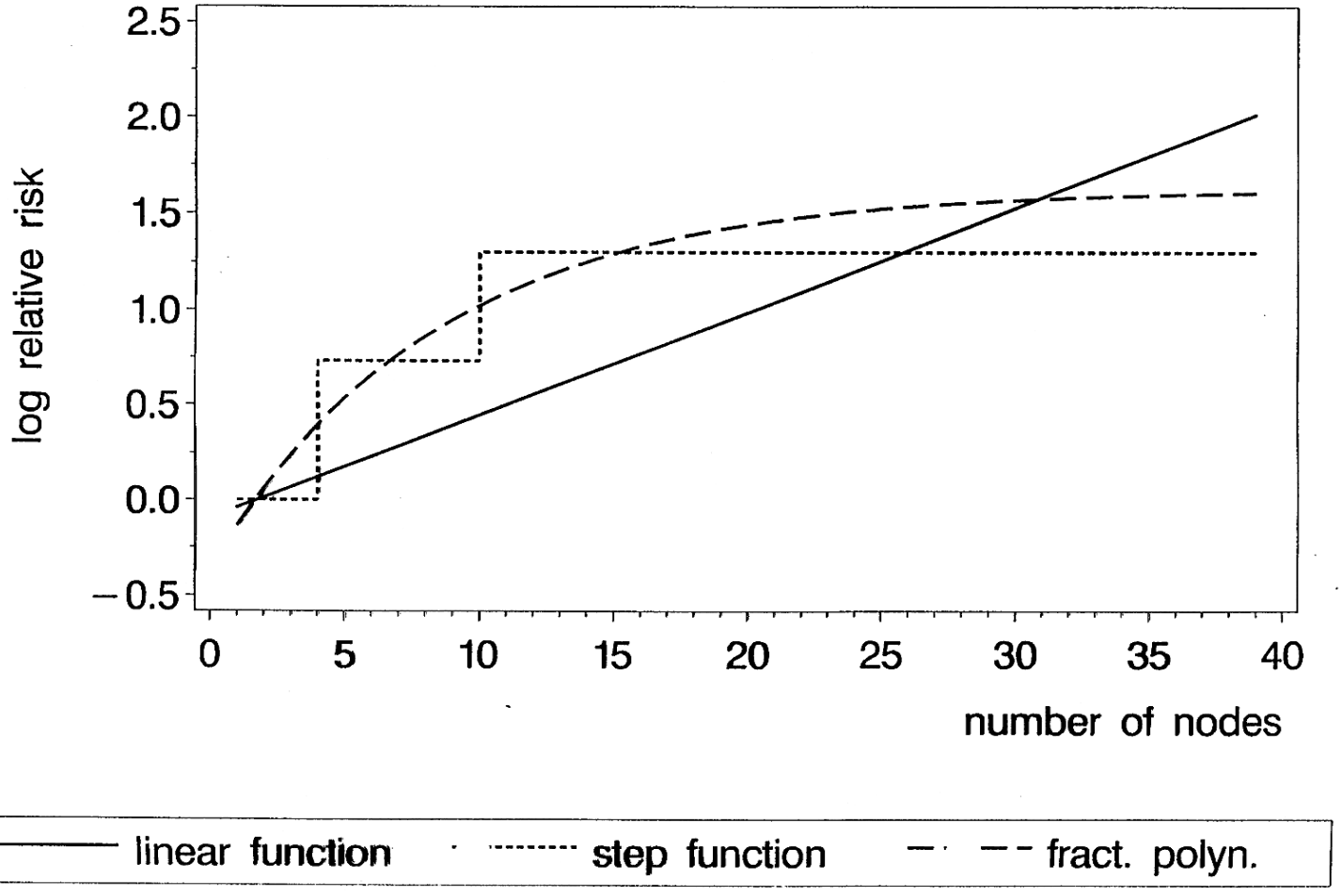
Continuous factors
Different results with different analyses
Age as prognostic factor in breast cancer (adjusted)



P-value 0.9 0.2 0.001

Results similar?

Nodes as prognostic factor in breast cancer (adjusted)



P-value	0.001	0.001	0.001

Interactions between treatment and several markers is the basis of stratified medicine

Interactions in RCTs

- Don't investigate effects in separate subgroups!
- Investigation of treatment/covariate interaction requires statistical tests
- Care is needed to avoid over-interpretation
- Distinguish two cases:
  - Hypothesis generation: searching for interactions with several variables
  - Specific predefined hypothesis

Interactions in RCTs – Practice

Assman et al: Subgroups analysis and other (mis)uses of baseline data in clinical trials. *Lancet 2000.*

50 clinical trial reports from four major medical journals (July to Sept, 1997)

Two-thirds represented subgroup findings, but mostly without appropriate statistical tests for interaction

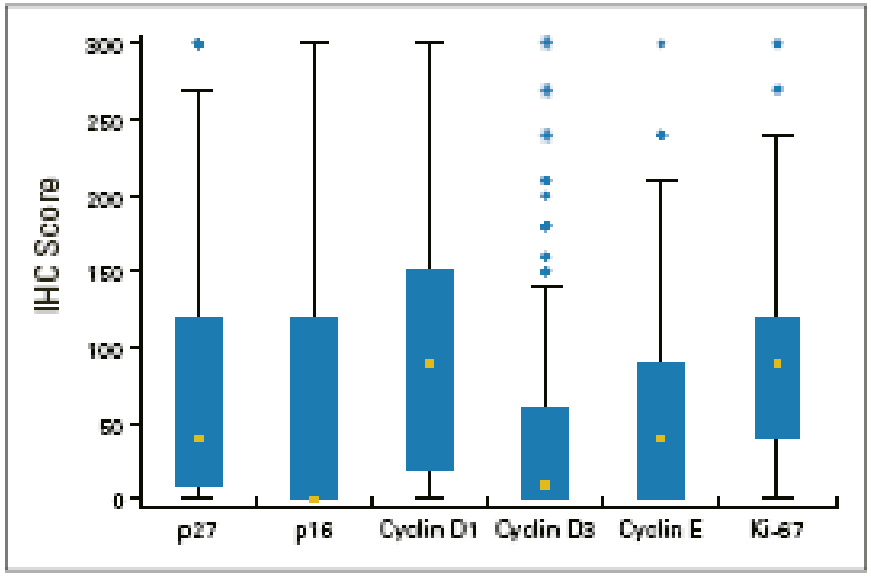
Subgroup analyses commonly lacked statistical power

Of all the various multiplicity problems in clinical trials subgroup analysis remains the most overused and overinterpreted

Similar results in Wang et al NEJM 2007

Interactions in RCTs
Continuous variables – usually dichotomized

Lung cancer



Distribution of 6 ICH scores in 778 patients

Filipits et al, JCO 2007

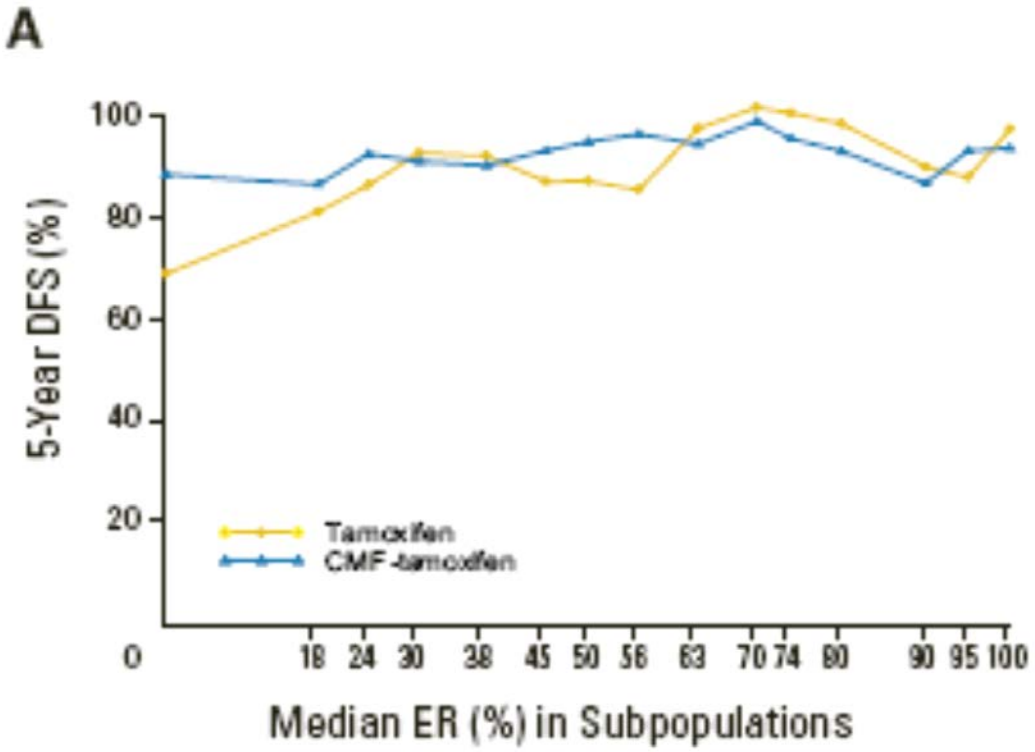
Biomarker	Adjusted HR for Death	95% CI	P*
$p27^{Kip1}$			.02
Negative	0.68	0.50 to 0.93	
Positive	1.09	0.82 to 1.45	
$p16^{INK4A}$			.79
Negative	0.87	0.67 to 1.12	
Positive	0.82	0.59 to 1.13	
Cyclin D1			.90
Negative	0.87	0.64 to 1.18	
Positive	0.95	0.65 to 1.11	
Cyclin D3			.79
Negative	0.83	0.63 to 1.09	
Positive	0.99	0.65 to 1.19	
Cyclin E			.33
Negative	0.77	0.57 to 1.02	
Positive	0.94	0.71 to 1.25	
Ki-67			.45
Negative	0.79	0.59 to 1.04	
Positive	0.92	0.69 to 1.22	

Abbreviation: HR, hazard ratio.
*Adjusted *P* value for interac

This issue is hardly criticized

Continuous variables – more than two subgroups

Breast cancer



Conclusion

'Low levels of ER and PgR are predictive of the benefit of adding chemotherapy to endocrine therapy ….'

Viale et al JCO 2008

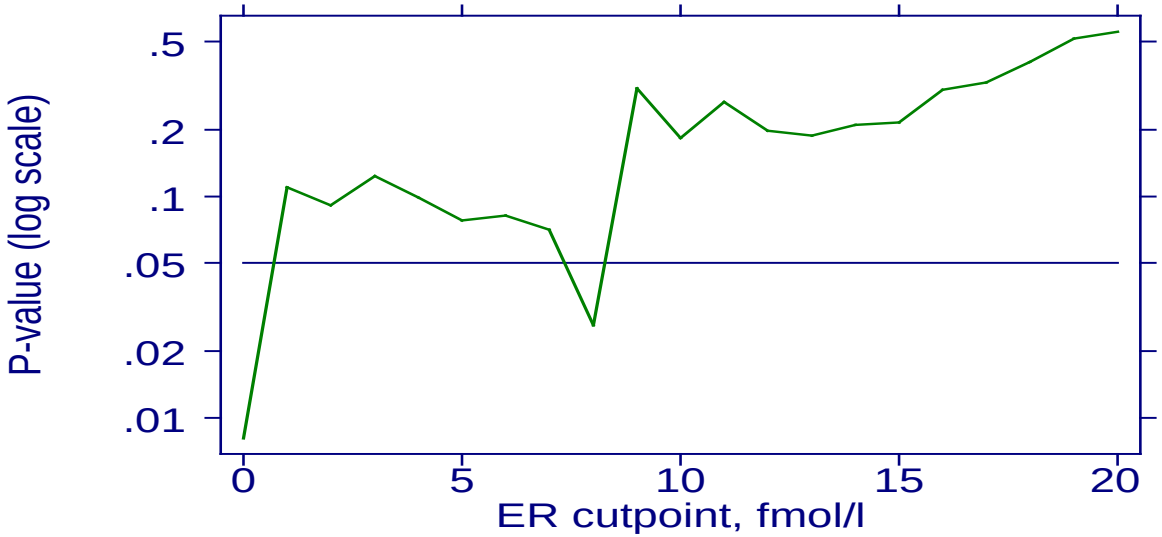
Interaction between treatment and continuous covariate

- GBSG-study in breast cancer
- Hormonal treatment tamoxifen (TAM): yes/no
- Known from overviews that TAM interacts with oestrogen receptor status (ER) of primary tumour
- But the research community needed many years to realize and to accept it

- For illustration: investigate ER $\times$ TAM interaction

Standard approach

- Based on binary predictor
- Need cut-point for continuous predictor
- Illustration - problem with cut-point approach



Interactions – MFPI method

- Have continuous *X* of interest, binary treatment variable *T* and other covariates *Z*
- Select 'adjustment' model *Z** on *Z* using MFP
- Find best FP2 function of *X* (in all patients) adjusting for *Z** and *T*
- Test FP2(*X*) × *T* interaction (2 d.f.)
  - Estimate β's separately in 2 treatment groups
  - Standard test for equality of β's
- May also consider simpler FP1 and linear functions

Interactions – treatment effect function

- Have estimated two FP2 functions – one per treatment group
- Plot difference between functions against *X* to show the interaction
  - i.e. the treatment effect at different *X*
- Pointwise 95% CI shows how strongly the interaction is supported at different values of *X*
  - i.e. variation in the treatment effect

FP analysis

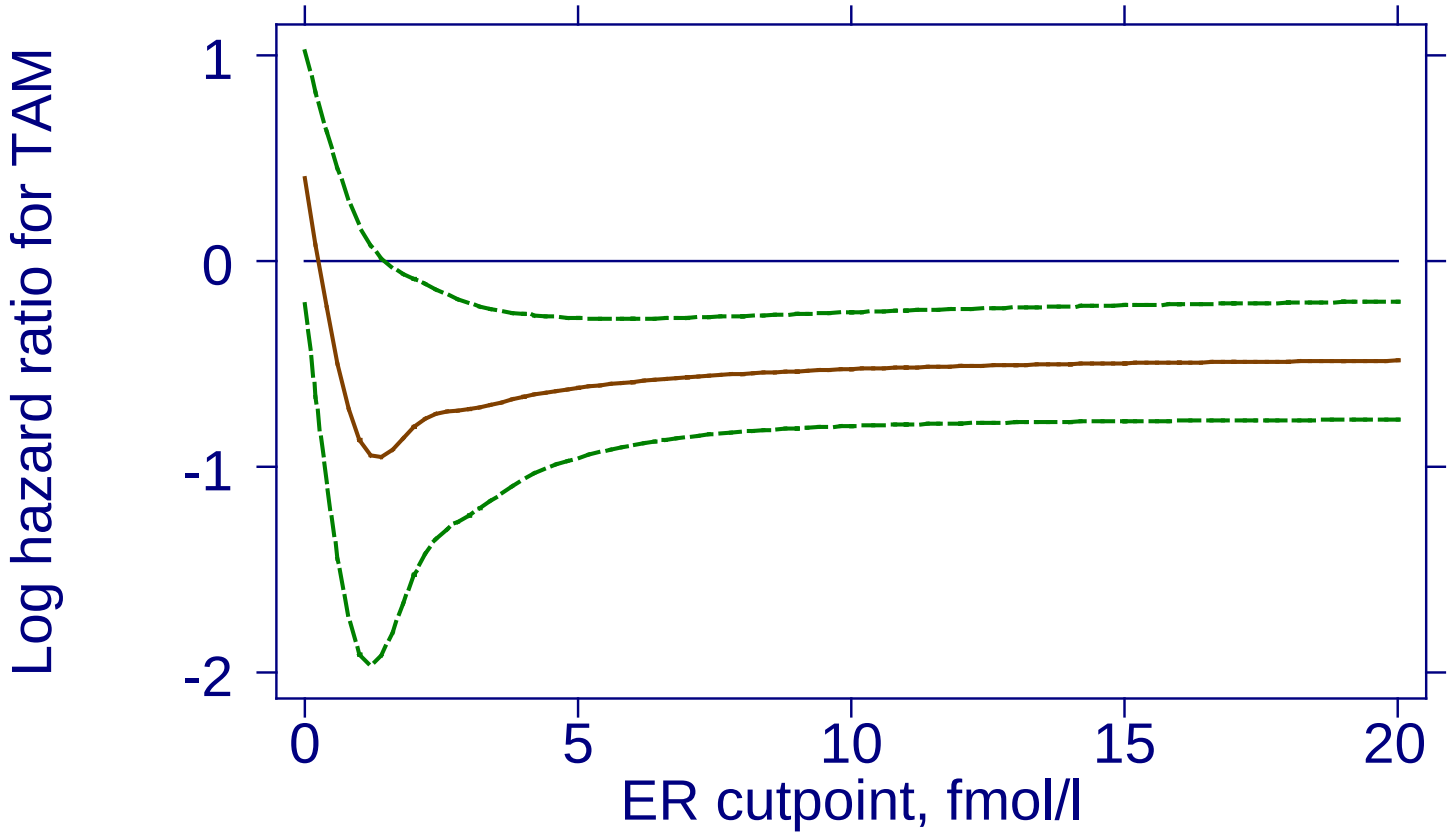
- Adjustment model (prognostic effect of other factors):

Factor	Transformation
Age	**FP2(–2, –1)**
Lymph nodes	**exp(–0.12LN)**
Grade 2/3	**(binary)**

- FP2 model for interaction has powers (–2, –2)
- P value for interaction ER × TAM (2 *df*) is 0.026

Treatment effect function

Effect of TAM by ER



Example: Metastatic renal cancer

RCT in UK to compare interferon-α with MPA

N = 347, 322 Death

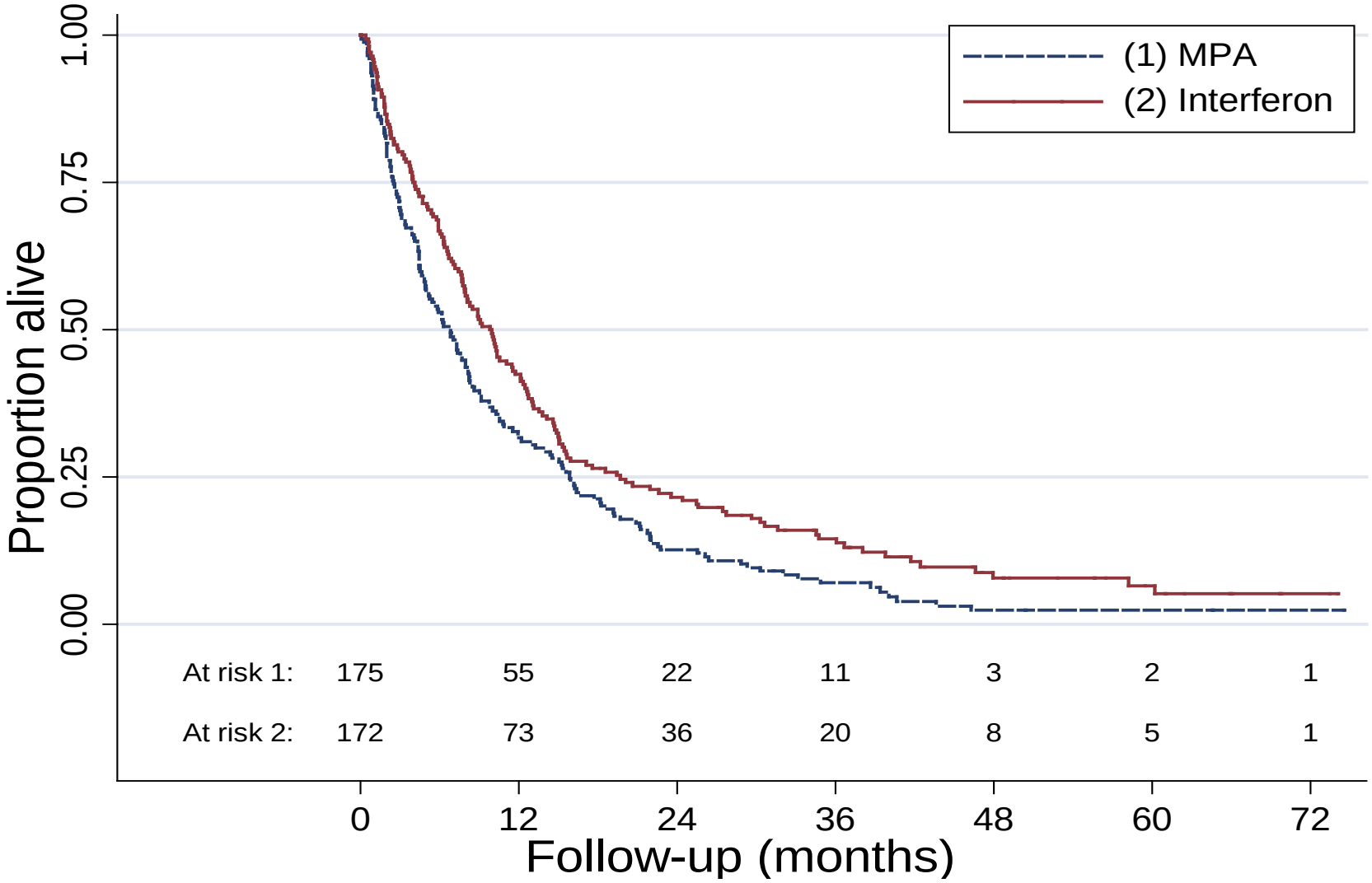
14 potential prognostic factors

Main analysis:

Interferon improves survival

HR: 0.75 (0.60 - 0.93), p = 0.009

Example: MRC trial – MPA and interferon in renal cancer

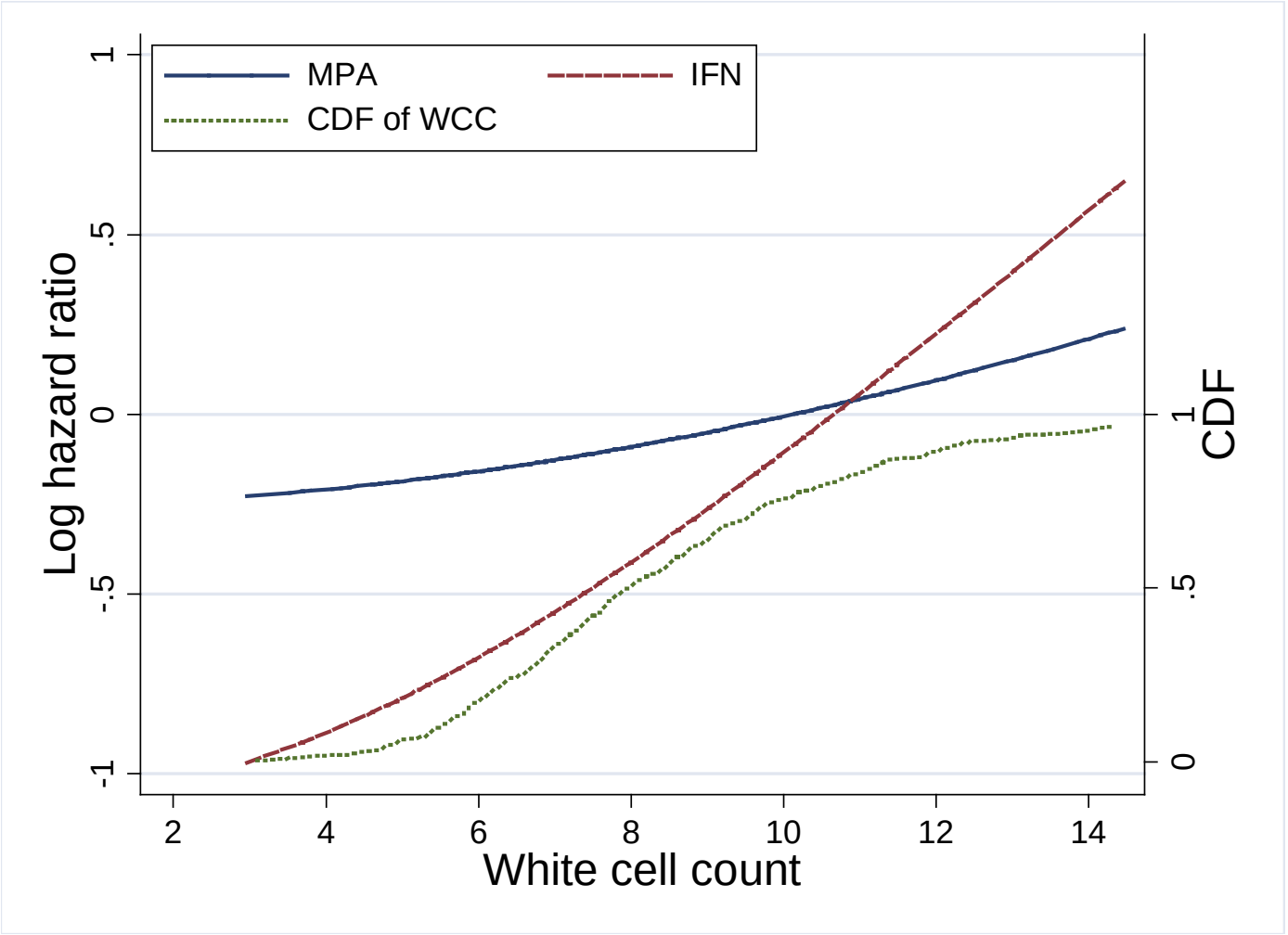


Overall: Interferon is better

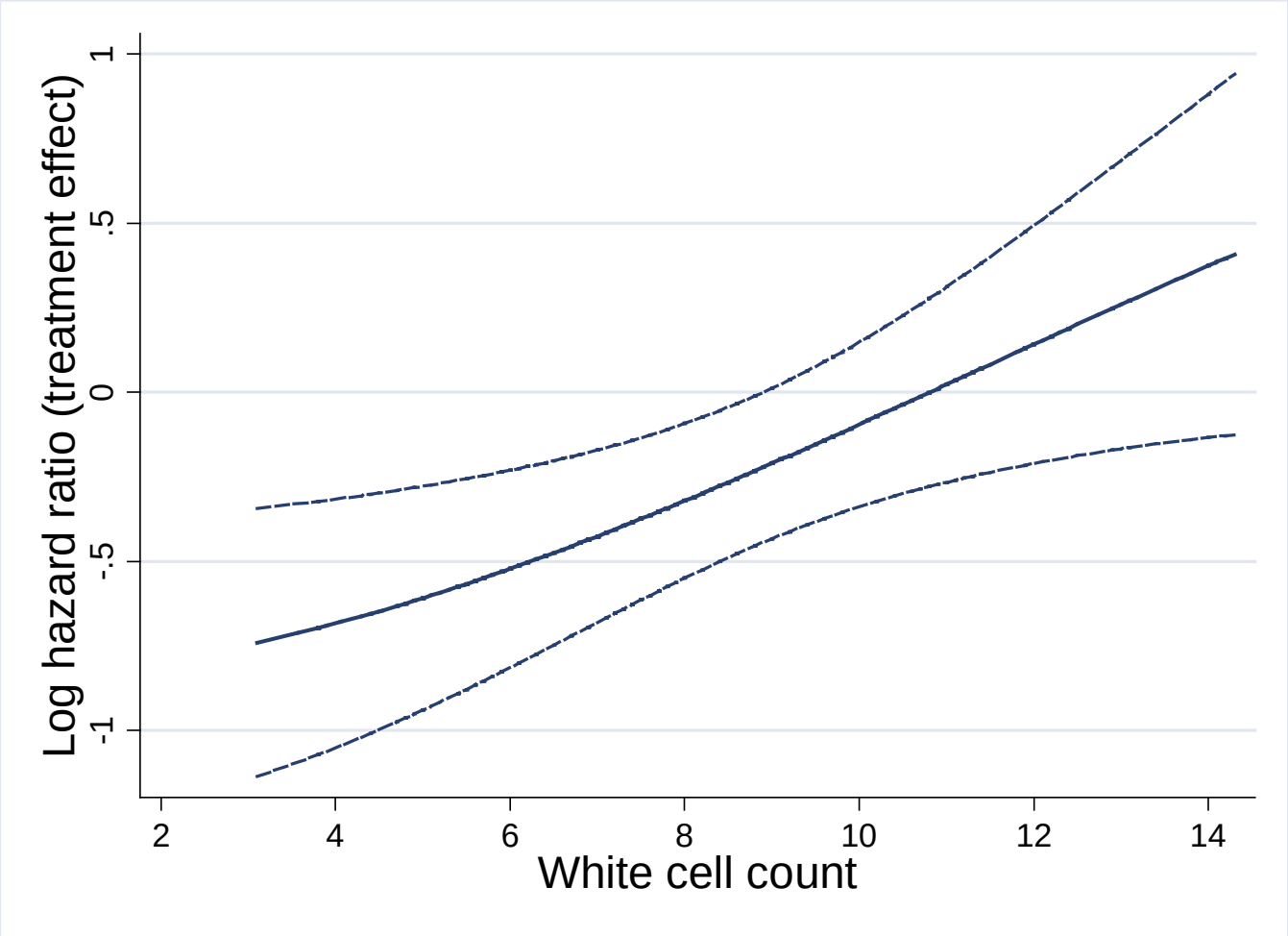
- P < 0.01; HR = 0.75; 95% CI (0.60, 0.93)
- Is the treatment effect similar in all patients? Sensible question?
  - Yes, at least for hypothesis generation
- Ten possible covariates available for the investigation of treatment-covariate interactions – only one is significant (WCC)

Hypothesis generation:
does the treatment effect depend on any factor?

Effect of WCC is best modelled with an FP2 (2, 3).



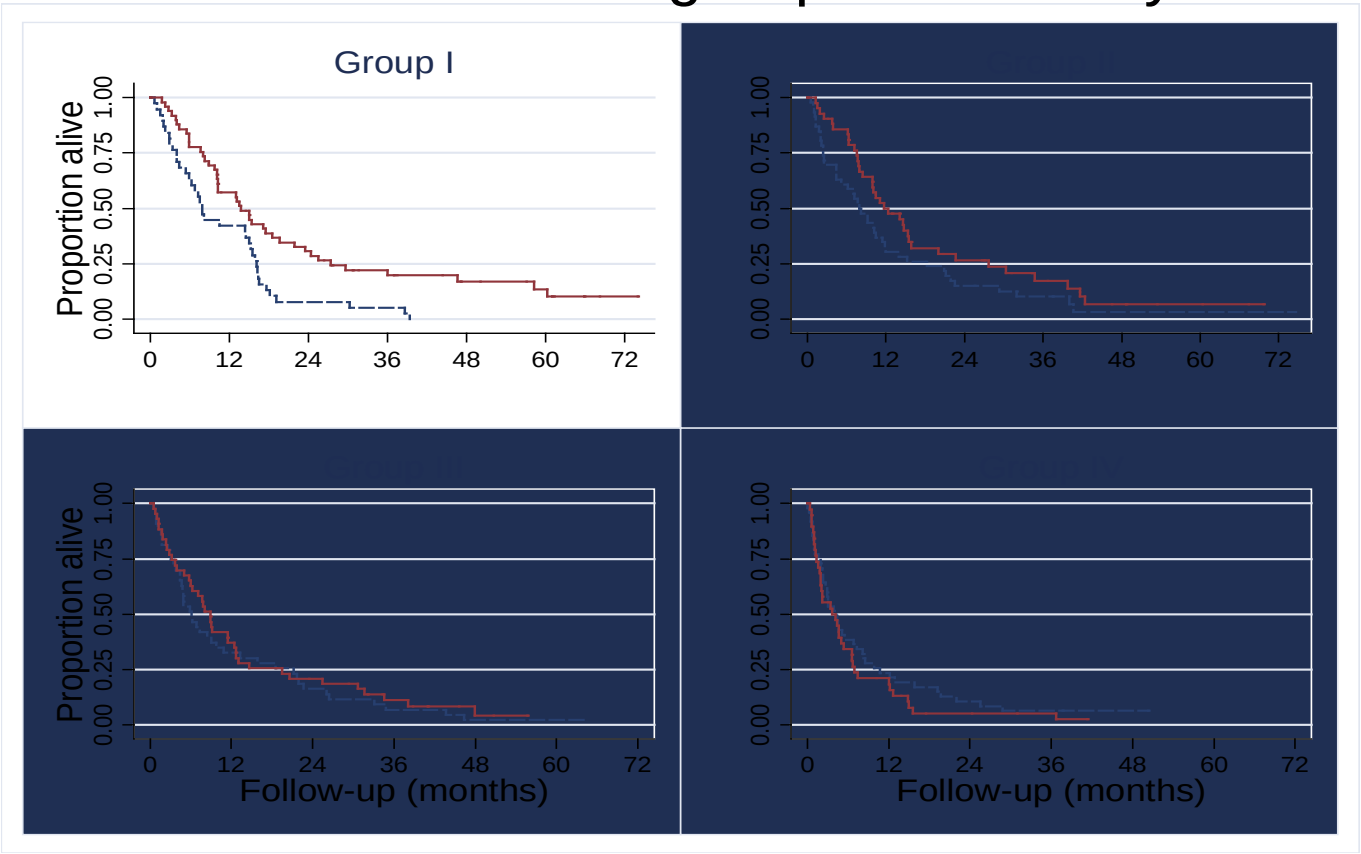
Treatment effect seems to depend on WCC



About 25% of patients with WCC > 10 seem not to benefit from interferon

Does model agree with data?
Check proposed trend

Treatment effect in subgroups defined by WCC



HR (Interferon to MPA; adjusted values similar) **overall: 0.75 (0.60 – 0.93)**

I : 0.53 (0.34 – 0.83) **II : 0.69 (0.44 – 1.07)**

III : 0.89 (0.57 – 1.37) **IV : 1.32 (0.85 –2.05)**

Alternative approach for continuous variables

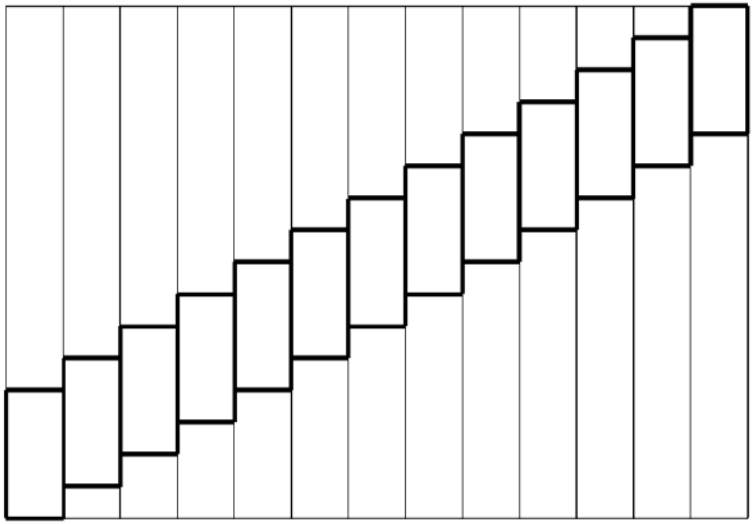
- STEPP

  Subpopulation Treatment Effect Pattern Plots

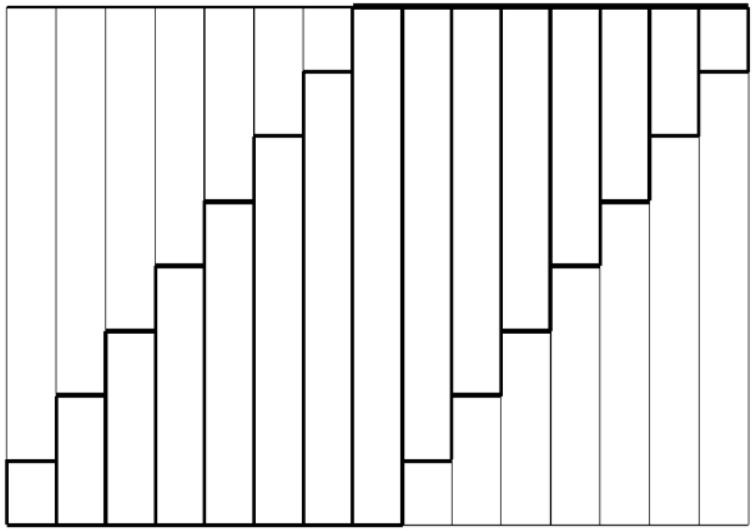
  Bonetti & Gelber 2000, 2004, 2009

STEPP

Sequences of overlapping subpopulations



Sliding window



Tail oriented

STEPP

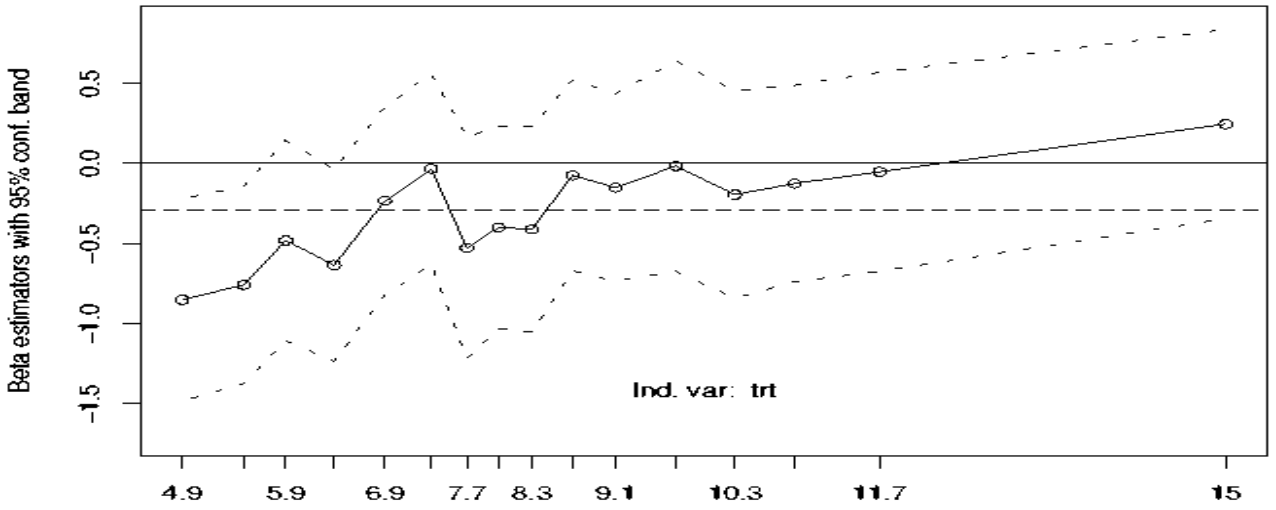
Estimate treatment effect in each subpopulation

Overlapping populations, therefore correlation between the estimates

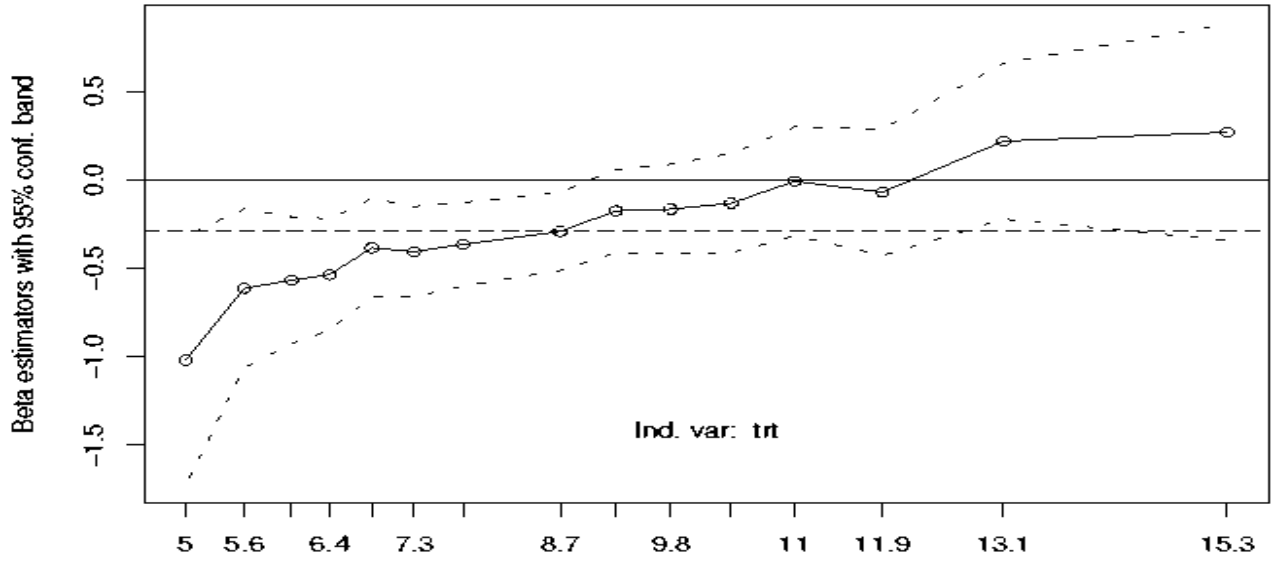
Simultaneous confidence band and tests proposed

STEPP – Interaction with WCC

SLIDING WINDOW (n1 = 25, n2 = 40)

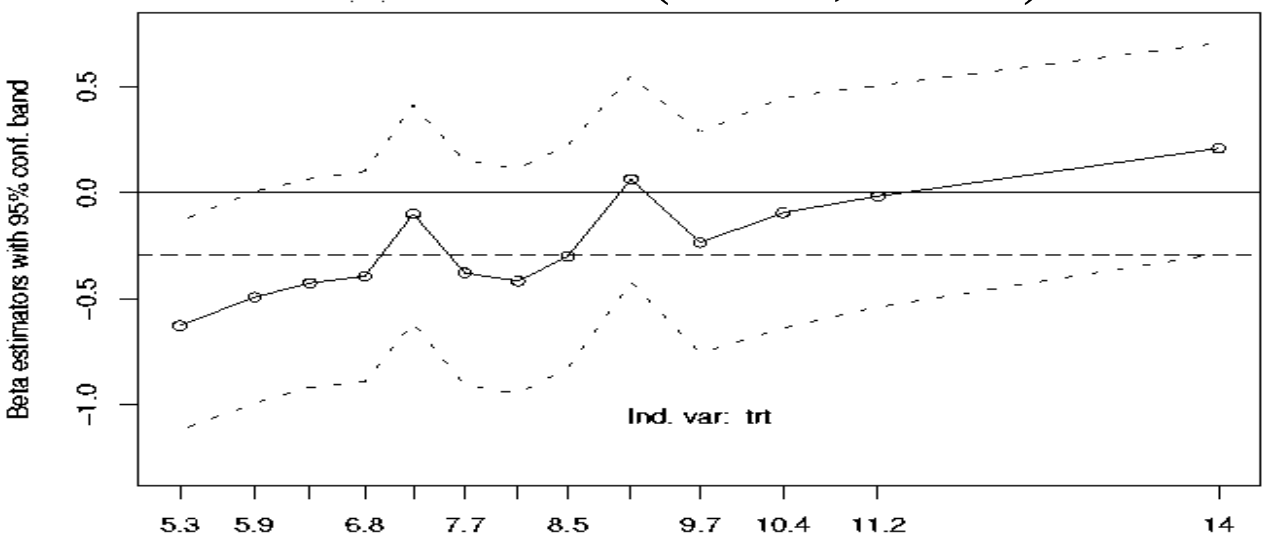


TAIL ORIENTED (g = 8)

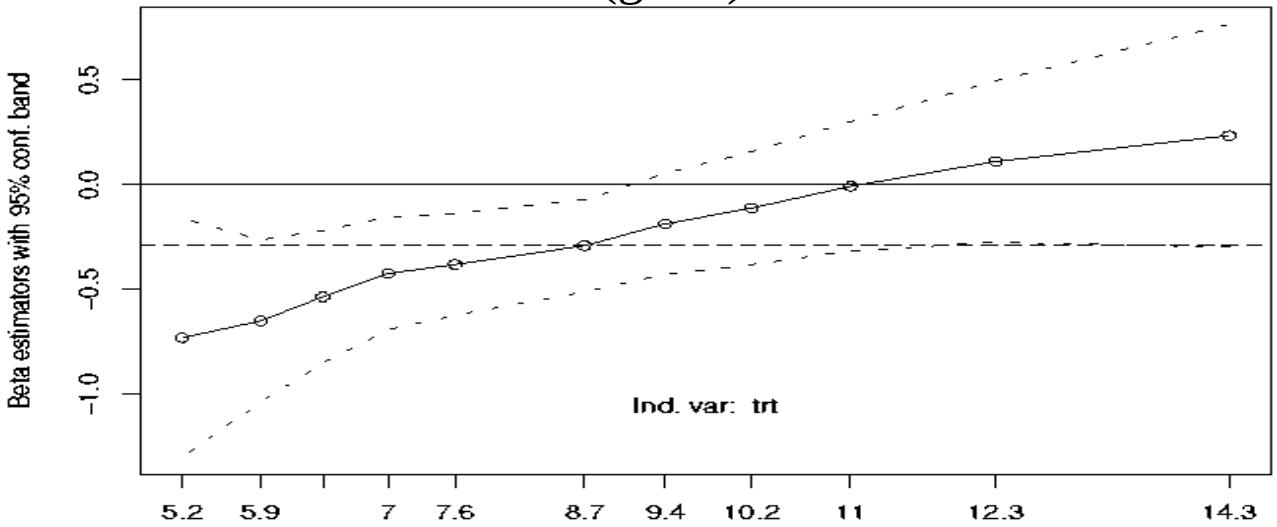


STEPP – Interaction with WCC

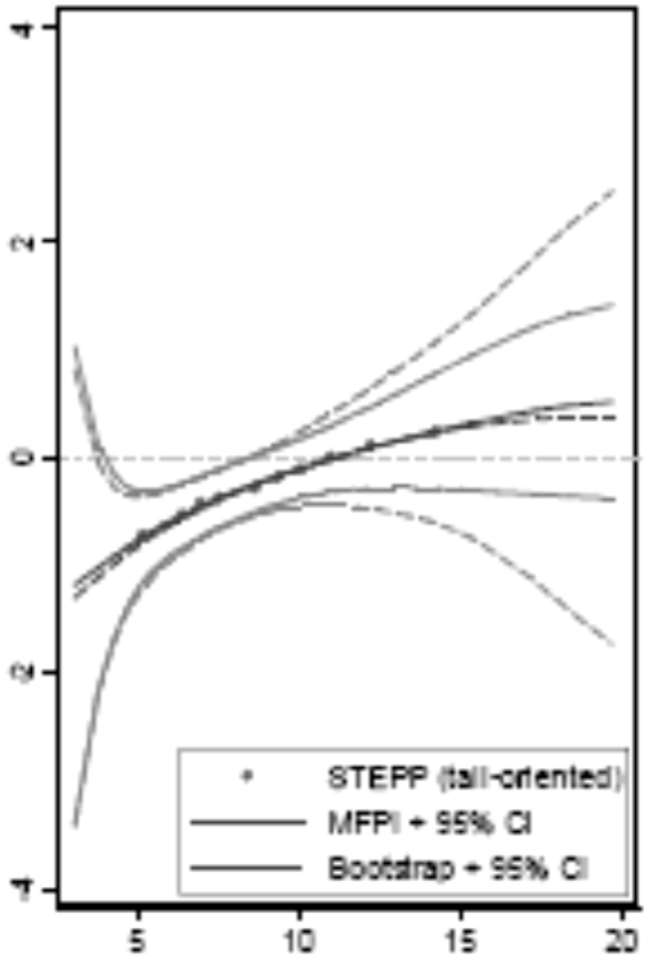
SLIDING WINDOW (n1 = 40, n2 = 60)



TAIL ORIENTED (g = 6)



STEPP as check of MFPI

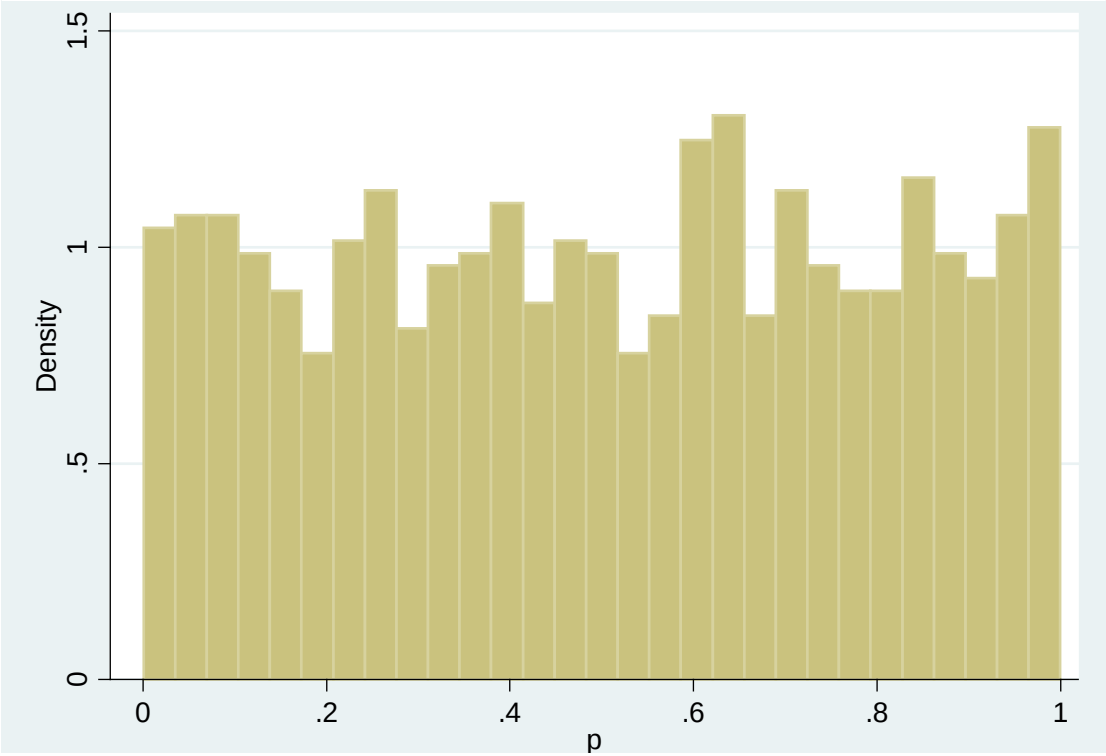


MFPI – Type I error

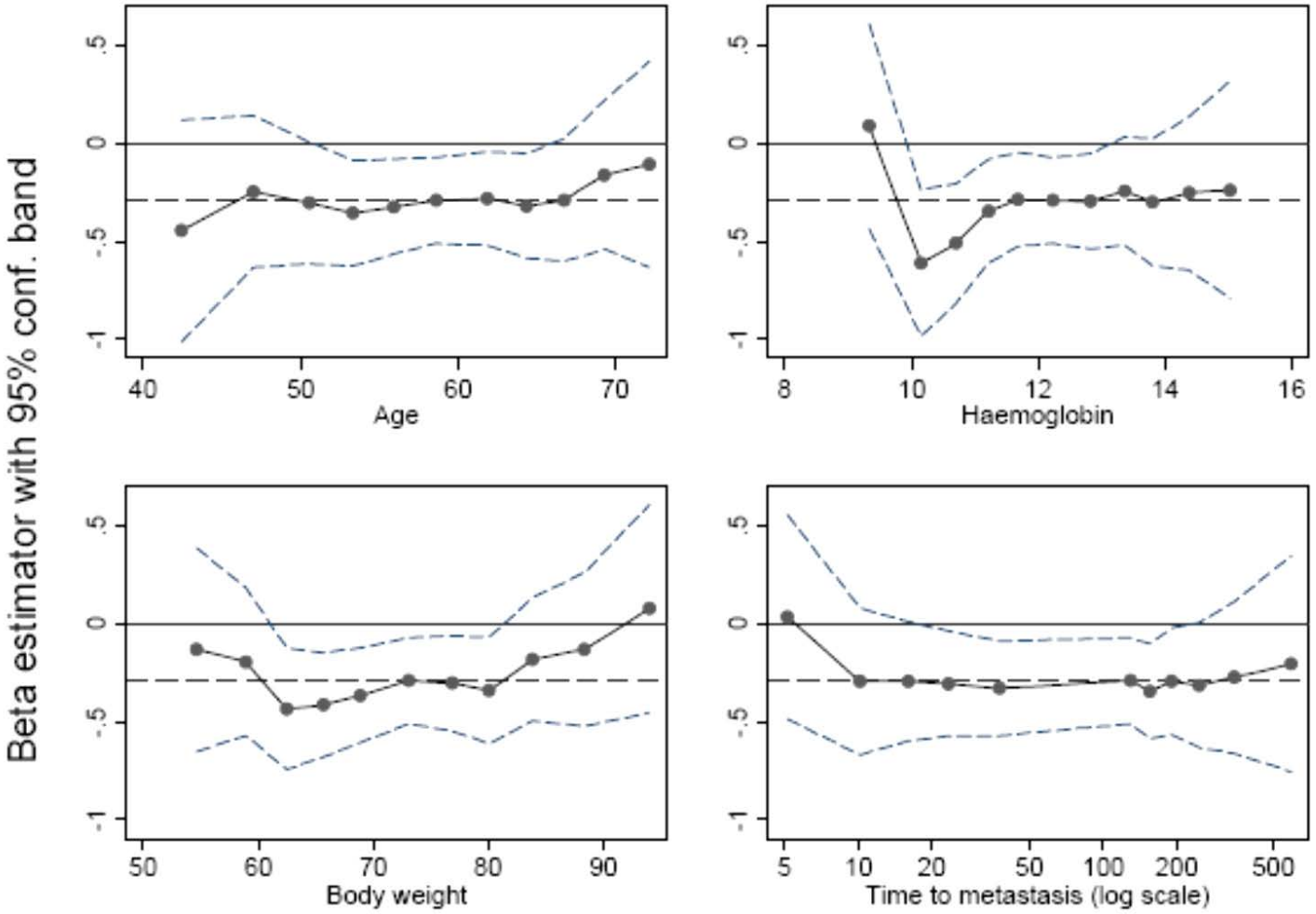
Random permutation of a continuous covariate (haemoglobin)

→ no interaction

Distribution of P-value from test of interaction
1000 runs, Type I error: 0.054



STEPP – No interactions



REPORTING of observational studies

Can we believe in the published literature?

- Selection of published studies
- Insufficient reporting for assessment of quality of
  - planing
  - conducting
  - analysis
- too early publications
- Usefullness for systematic review (meta-analysis)

Begg et al JAMA (1996) Improving the Quality of Reporting of Randomized Controlled Trials – The CONSORT Statement
Moher et al JAMA (2001), Revised Recommendations
Schulz et al Ann Int Med (2010) Updated…

von Elm et al Lancet (2007), STROBE

Reporting of prognostic markers

Riley et al BJC (2003)
Systematic review of tumor markers for neuroblastoma
260 studies identified, 130 different markers

The reporting of these studies was often inadequate, in terms of both statistical analysis and presentation, and there was considerable heterogeneity for many important clinical/statistical factors. These problems restricted both the extraction of data and the meta-analysis of results from the primary studies, limiting feasibility of the evidence-based approach.

Almost all articles on cancer prognostic markers report statistically significant results

Panayiotis A. Kyzas[a], *Despina Denaxa-Kyza*[a], *John P.A. Ioannidis*[a,b,c,*]

[a]*Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece*

[b]*Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina, Greece*

[c]*Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Boston, USA*

EJC 2007, 43:2559-79

Database 1: 340 articles included in meta-analysis

Database 2: 1575 articles published in 2005

We expect some improvements by the REMARK guidelines

REporting recommendations for tumor MARKer prognostic studies (REMARK)

Lisa M McShane*, Douglas G Altman, Willi Sauerbrei, Sheila E Taube, Massimo Gion and Gary M Clark
for the Statistics Subcommittee of the NCI–EORTC Working Group on Cancer Diagnostics

published simultaneously in 5 journals, August 2005

Reporting of MFP and MFPI analysis

The following variables were considered as candidates $x_1, \ldots x_k$

MFP(α_1, α_2) ; FP2 allowed

MFPI (α), adjusted for MFP(α_1, α_2) model

Candidates $x_1, \ldots x_k$

all continuous variables truncated (1%, 99%)

Summary

- Interaction effects are important and require more attention
- To use more information from the data modelling should get a more prominent role
- Type I error is often (over-) controlled at the expense of type II errors!
- Known problems of cutpoint analyses for prognostic factors transfer to the investigation for interactions
- Use full information and derive treatment effect function
  - MFPI well suited
  - STEPP has problems
- Internal check of MFPI result is required, external data to validate results

References

Assman SF, Pocock S, Enos LE, Kasten LE (2000): Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet, 355, 1064-1069.

Bonetti M, Gelber RD (2000): A graphical method to assess treatment-covariate interactions using the Cox model on subsets of the data. Statistics in Medicine, 19, 2595-2609.

Bonetti M, Gelber RD (2004): Patterns of treatment effects in subsets of patients in clinical trials. Biostatistics, 5,465-481.

Royston P, Sauerbrei W (2004): A new approach to modelling interactions between treatment and covariates in clinical trials by using fractional polynomials. Statistics in Medicine, 23, 2509-2525.

Royston P, Sauerbrei W, Ritchie A (2004): Is treatment with interferon-a effective in all patients with metastatic renal carcinoma? A new approach to the investigation of interactions. British Journal of Cancer, 90, 794-799.

Royston P, Sauerbrei W (2008): Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables, Wiley

Sauerbrei W, Royston P (1999): Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. Journal of the Royal Statistical Society A, 162, 71-94.

Sauerbrei W., Royston P., Binder H. (2007): Selection of important variables and determination of functional form for continuous predictors in multivariable model building. Statistics in Medicine,26:5512-5528

Sauerbrei W, Royston P, Zapien K (2007): Detecting an interaction between treatment and a continuous covariate: a comparison between two approaches, Computational Statistics and Data Analysis,51:4054-4063