Interaction of Treatment with a Continuous Variable:

Issues, Comparison of Approaches and an IPD Meta-analysis to Summarize Results across Several Studies

Willi Sauerbrei[1], **Patrick Royston**[2]

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

wfs@imbi.uni-freiburg.de

Overview

- Interactions in RCTs
- Continuous variables
  - problems of categorization
- Fractional polynominal approach
  - prognostic factor
  - factor * treatment interaction (TEF)
  - simulation
- STEPP
- Meta-analysis of TEFs

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 – Practise

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

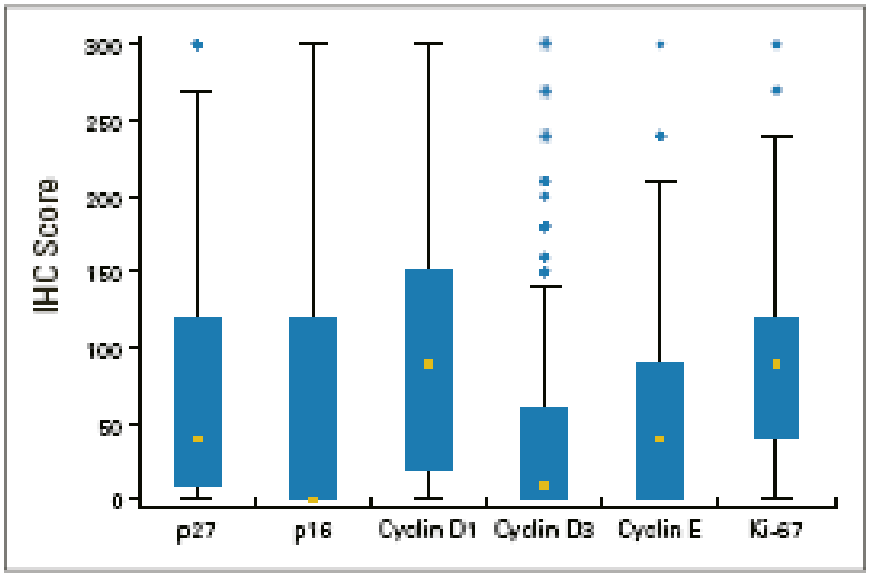
Subgroup analyses commonly lacked statistical power

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	0.97	0.67 to 1.12	
Positive	0.82	0.59 to 1.13	
Cyclin D1			.90
Negative	0.97	0.64 to 1.18	
Positive	0.85	0.65 to 1.11	
Cyclin D3			.79
Negative	0.83	0.63 to 1.09	
Positive	0.89	0.65 to 1.19	
Cyclin E			.33
Negative	0.77	0.57 to 1.02	
Positive	0.94	0.71 to 1.26	
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

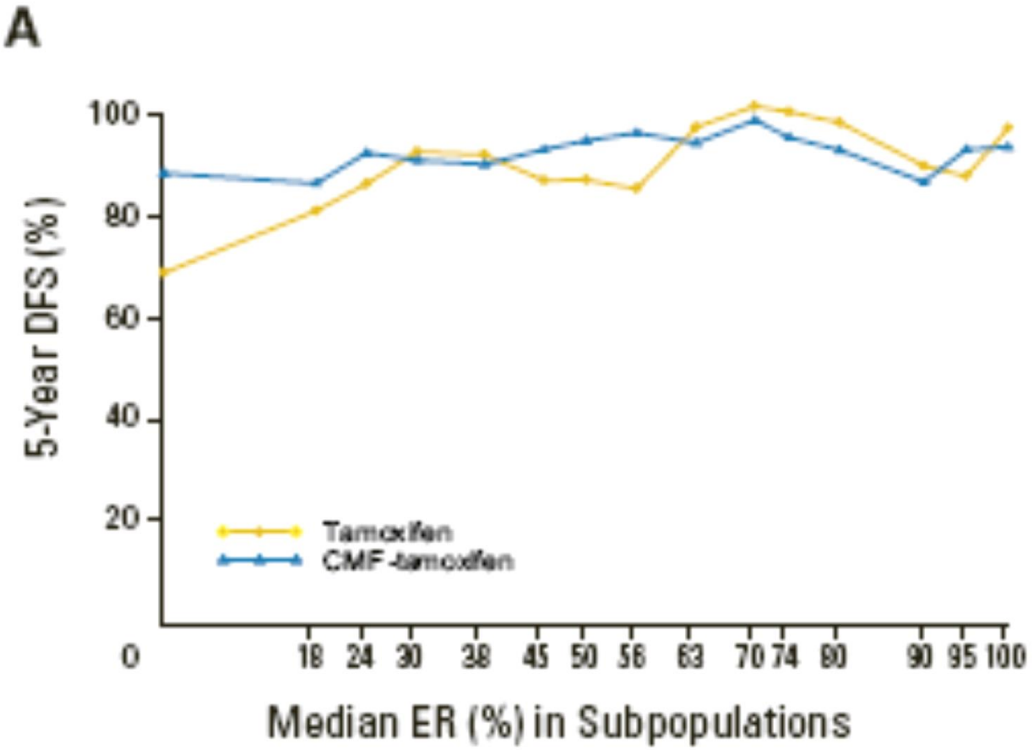
This issue is hardly criticized

Continuous variables – more than two subgroups

STEPP approach (later more)

Subpopulation Treatment Effect Pattern Plots

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

Related issue: Continuous variable as prognostic factor– what functional form?

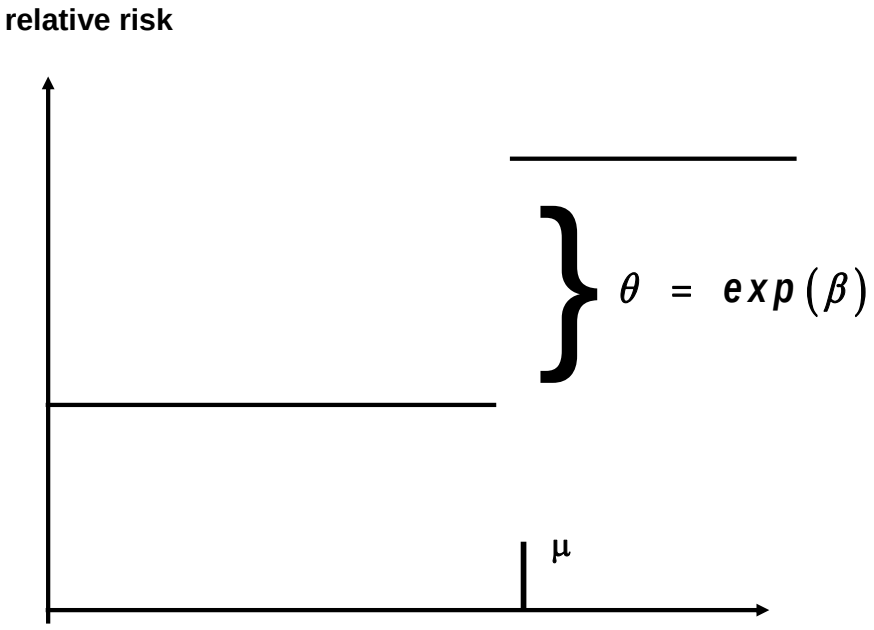
Traditional approaches

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 choic

Step function - the cutpoint problem



In the Cox model

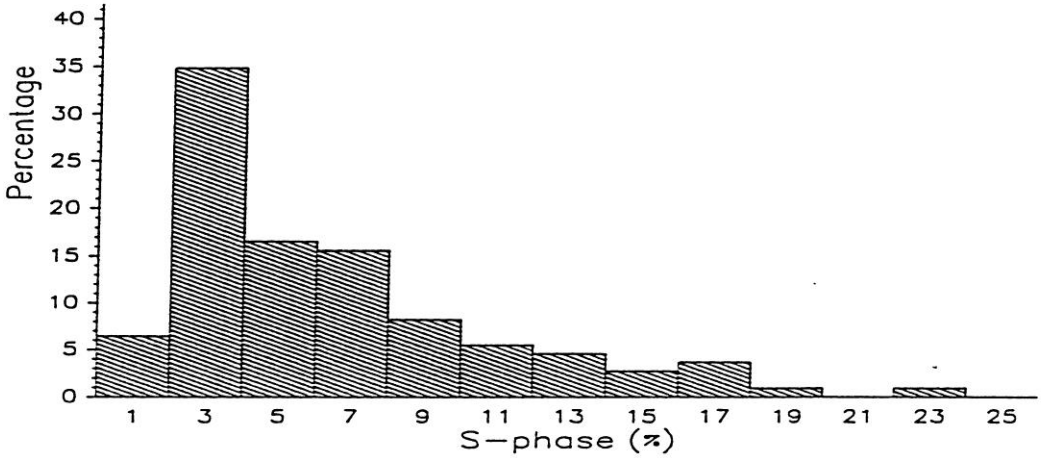
$$\lambda(t \mid X > \mu) = \exp \beta \, \lambda(t \mid X \leq \mu)$$

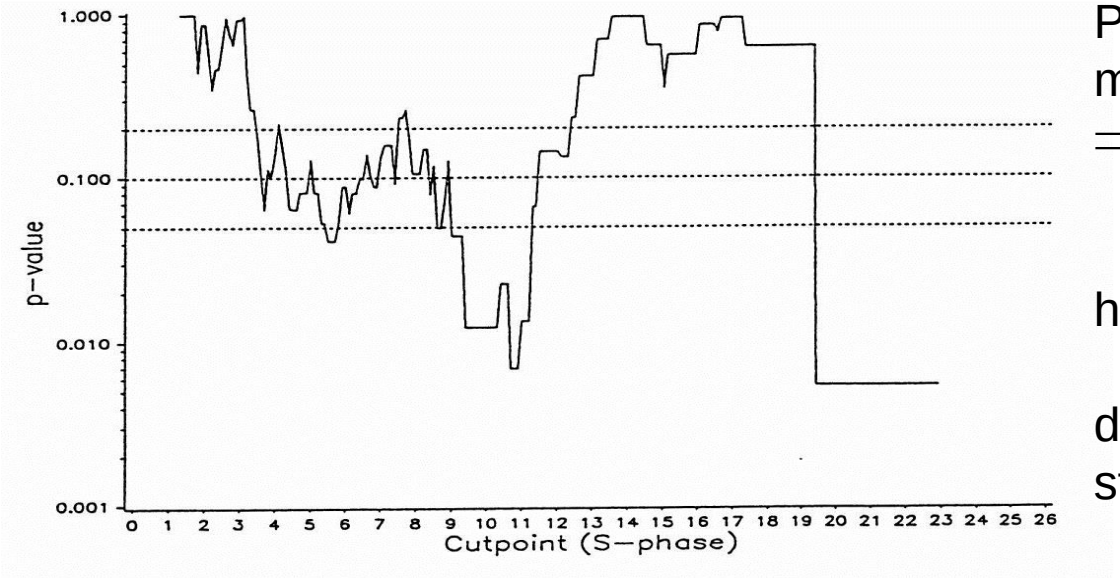
$\hat{\mu}$: estimated cutpoint for the comparison of patients with X above and below μ.

Step function – biologically plausible?

Searching for optimal cutpoint minimal p-value approach

SPF in Freiburg DNA study





Problems

multiple testing

⇒inflated type I error

about 40%-50% instead of 5%

heavily biased estimates

different cutpoints in each study

Example 1: Prognostic factors

GBSG-study in node-positive breast cancer

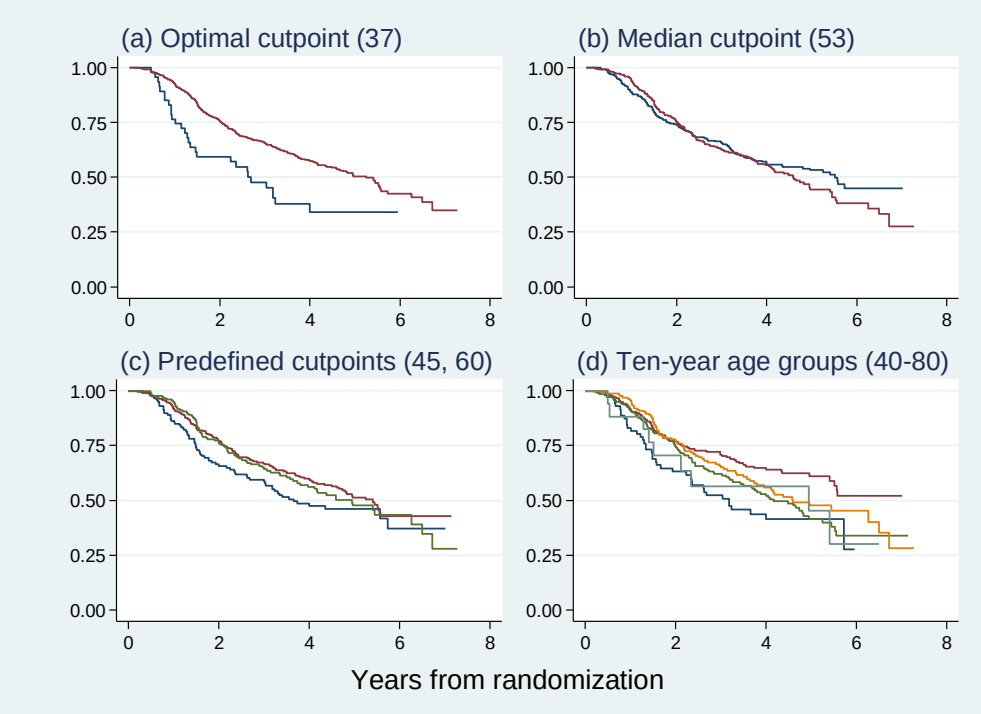
299 events for recurrence-free survival time (RFS) in
686 patients with complete data

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

Tamoxifen yes/no

We will consider
- age as prognostic factor
- estrogen receptor as predictive factor

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); p= 0.2

(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

Modelling continuous covariates has many advantages.

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

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

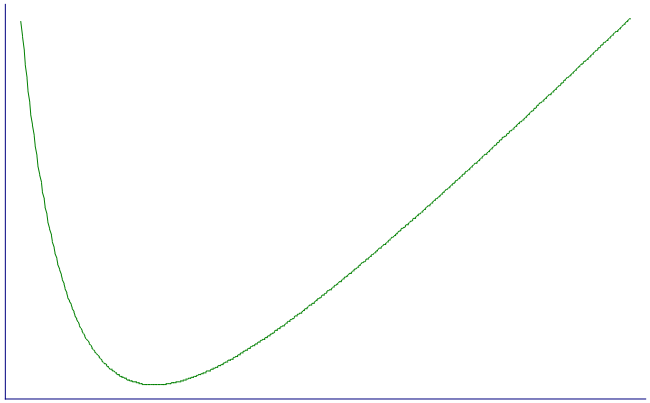
  0- logX

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

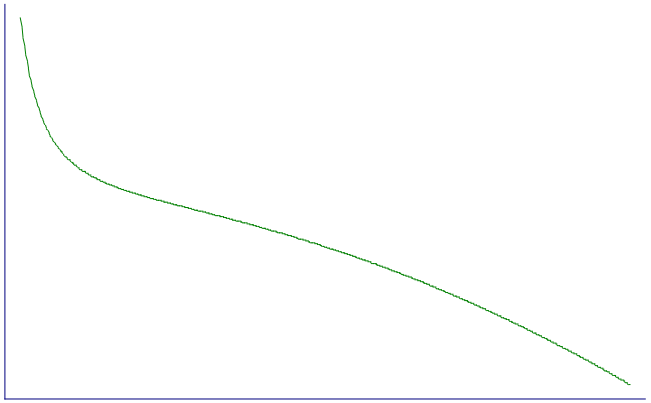
Example $FP2 = \beta_1 X^{0.5} + \beta_2 X^{3}$

Examples of FP2 curves - varying powers

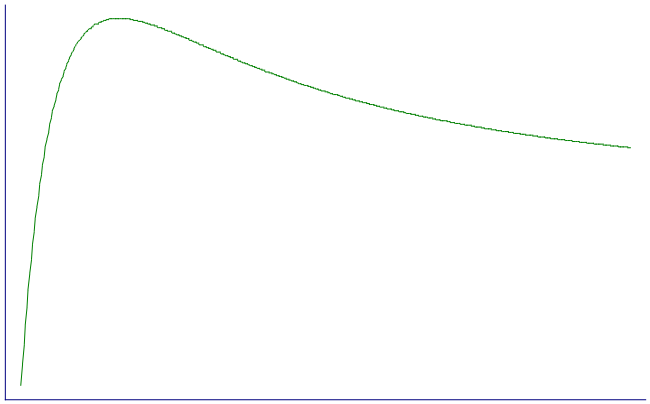
(-2, 1)



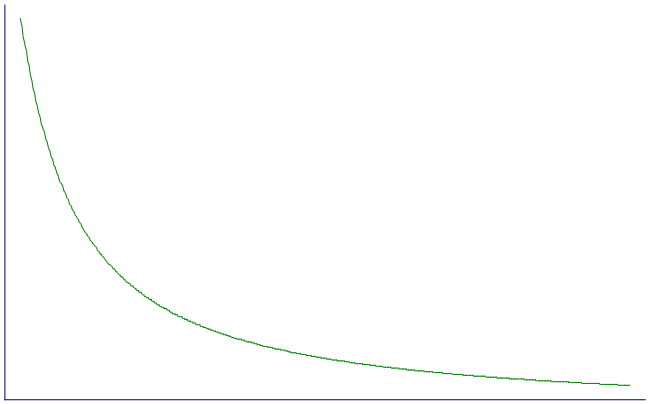
(-2, 2)



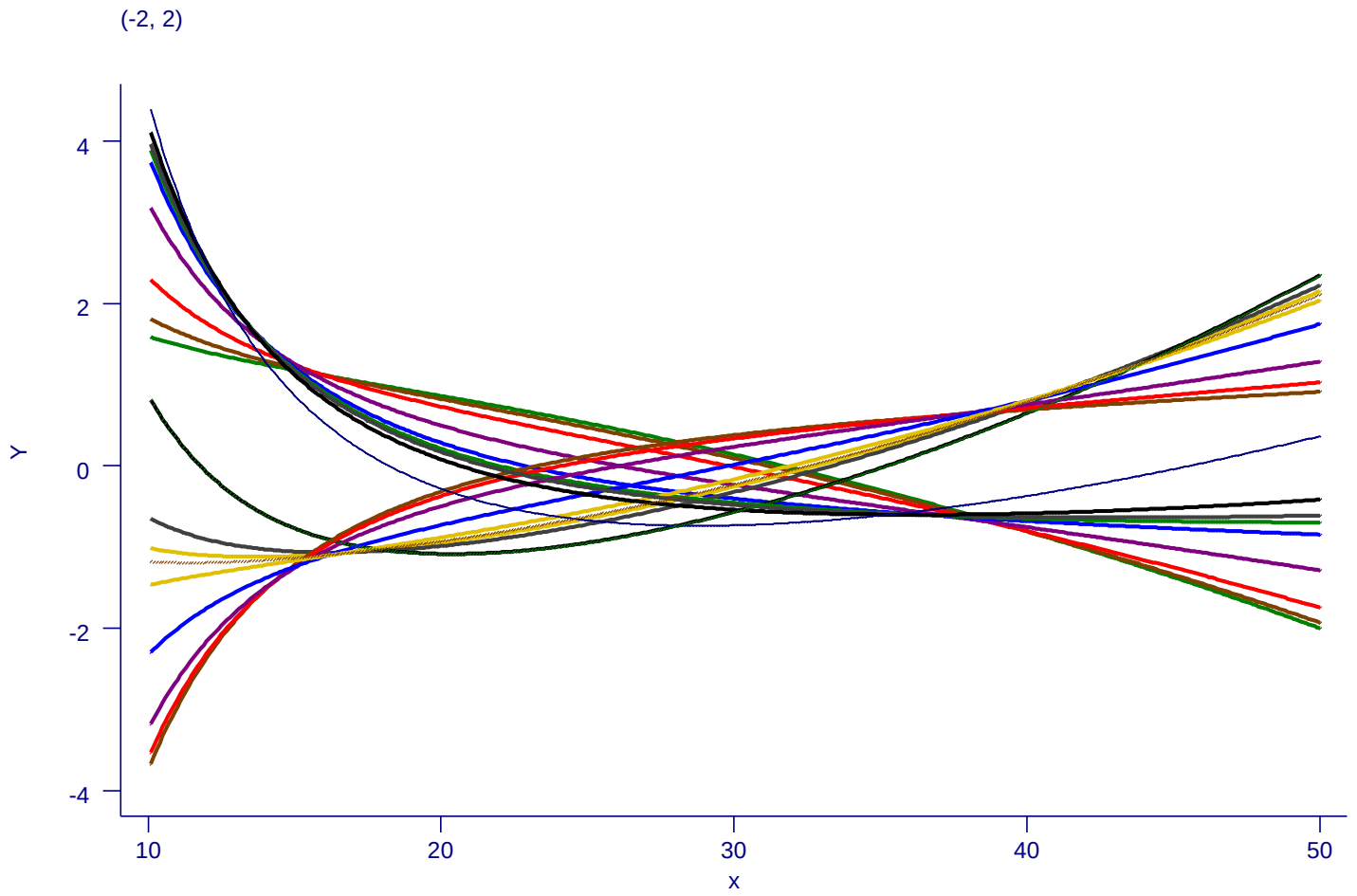
(-2, -2)



(-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 diff chi-square	Powers		Model diff chi-square	Powers		Model diff chi-square	Powers		Model diff 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

7

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

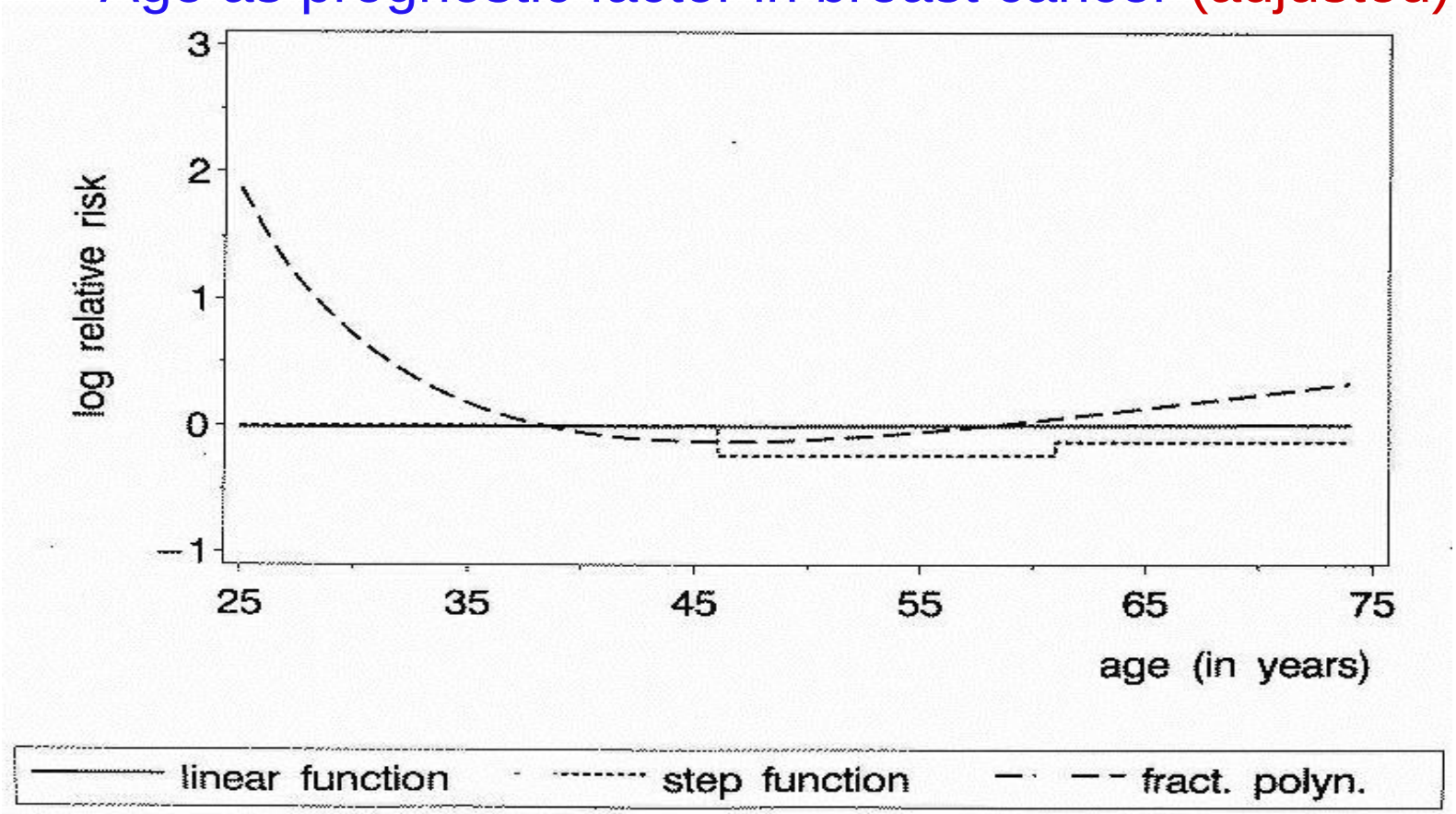
(usually binary and categorical variables are also available)

MFP algorithm combines

- backward elimination with
- FP function selection procedures

Continuous factors
Different results with different analyses

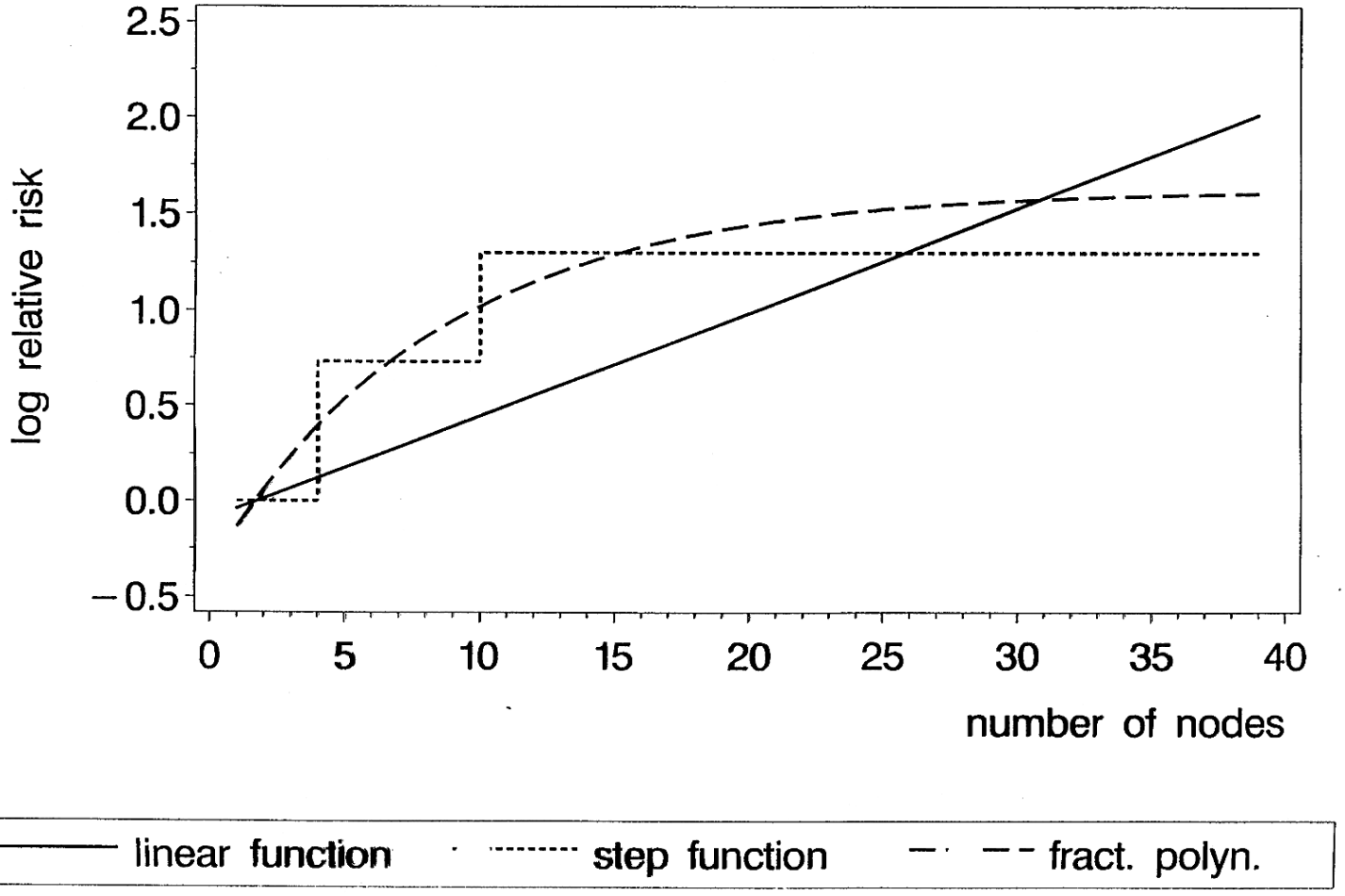
Age as prognostic factor in breast cancer (adjusted)



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 with a Continuous Covariate

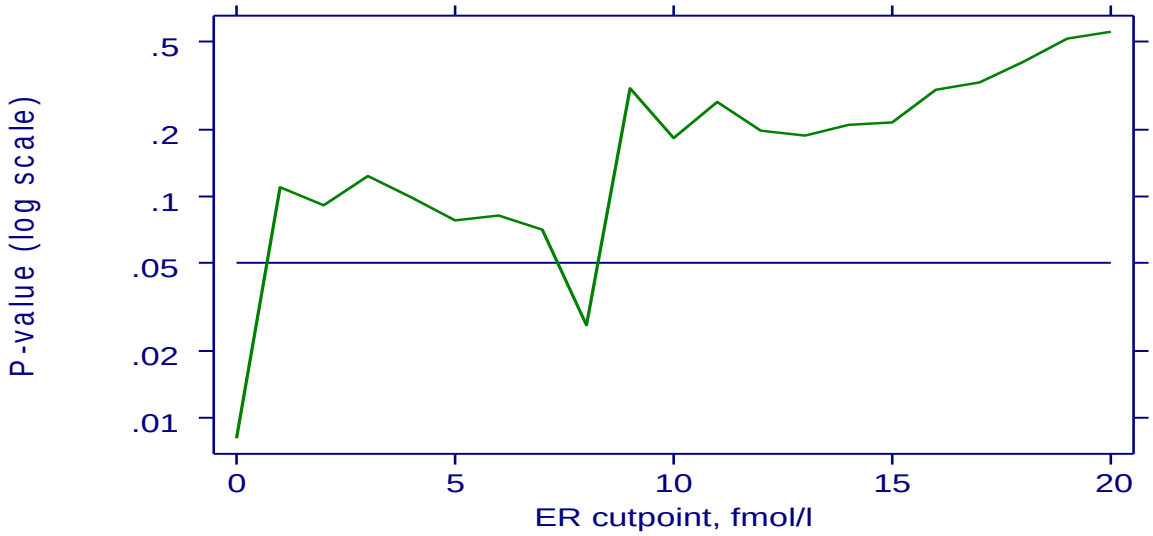
Interaction between treatment and continuous covariate

- 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

Example 2: Interactions in an RCT

Metastatic renal cancer

RCT in UK to compare interferon-α with MPA

N = 347, 322 Death

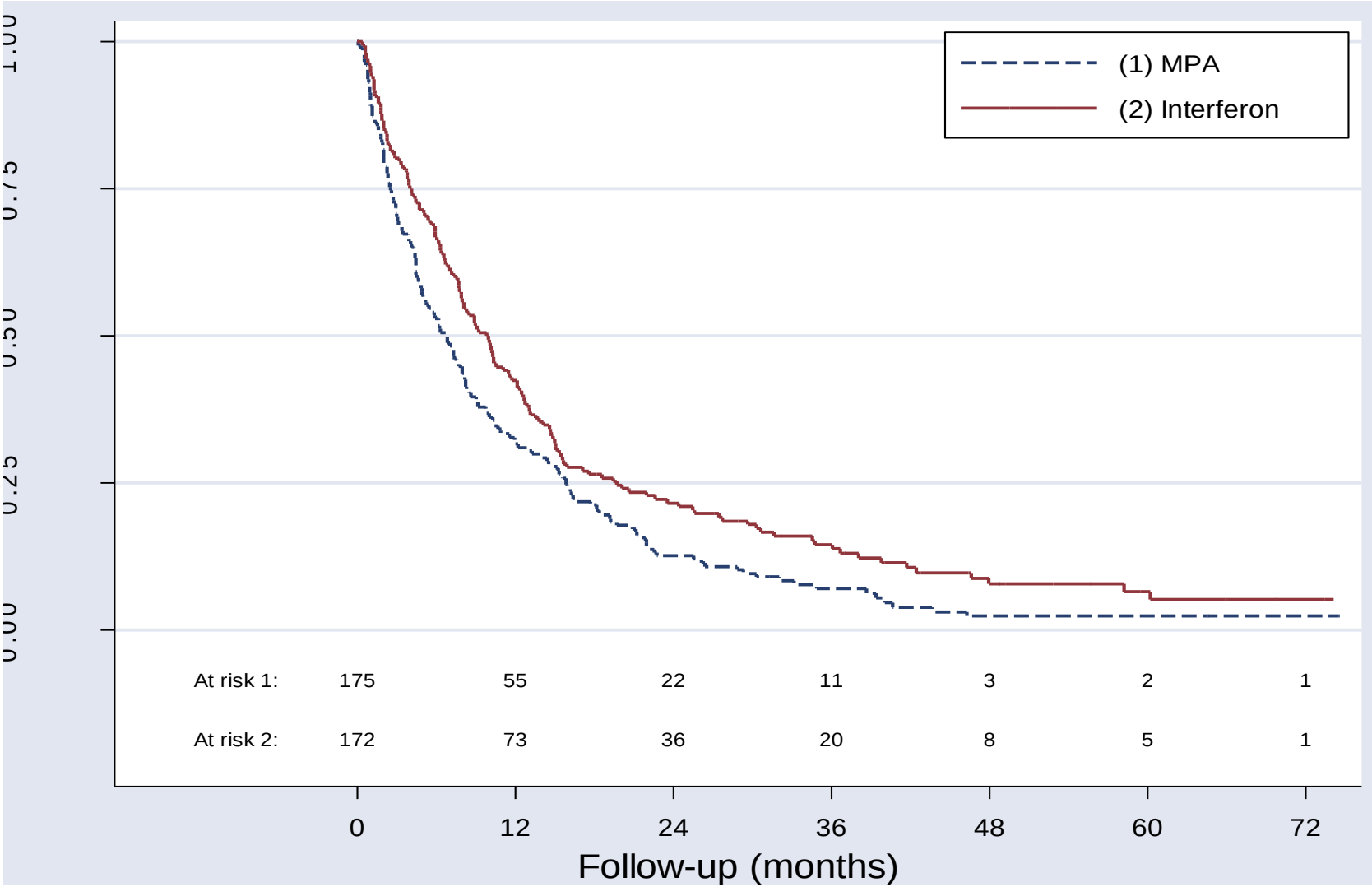
14 potential prognostic factors

Main analysis:

Interferon improves survival

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

Example: MRC trial – MPA and interferon in renal cancer

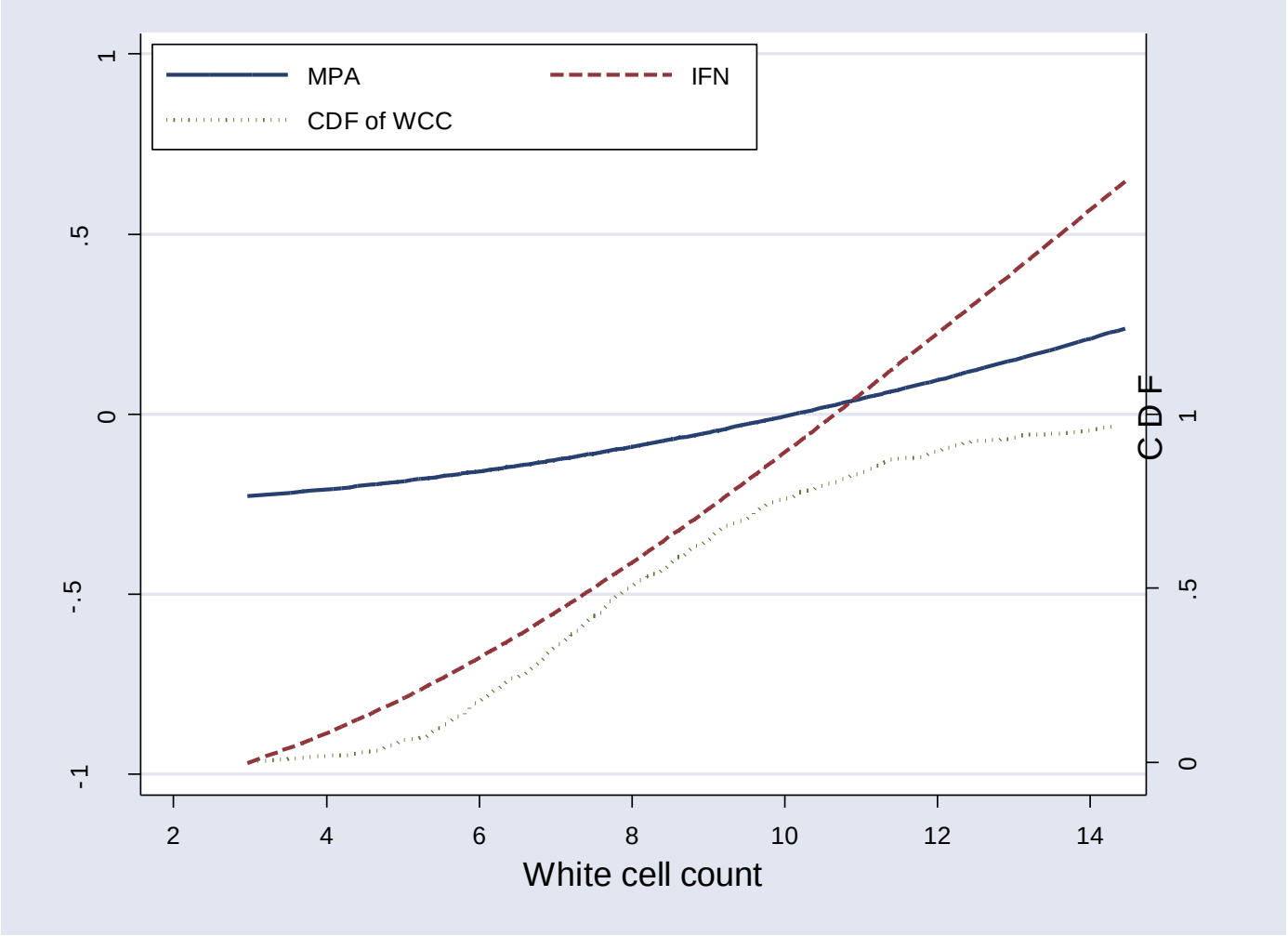


Overall: Interferon is better

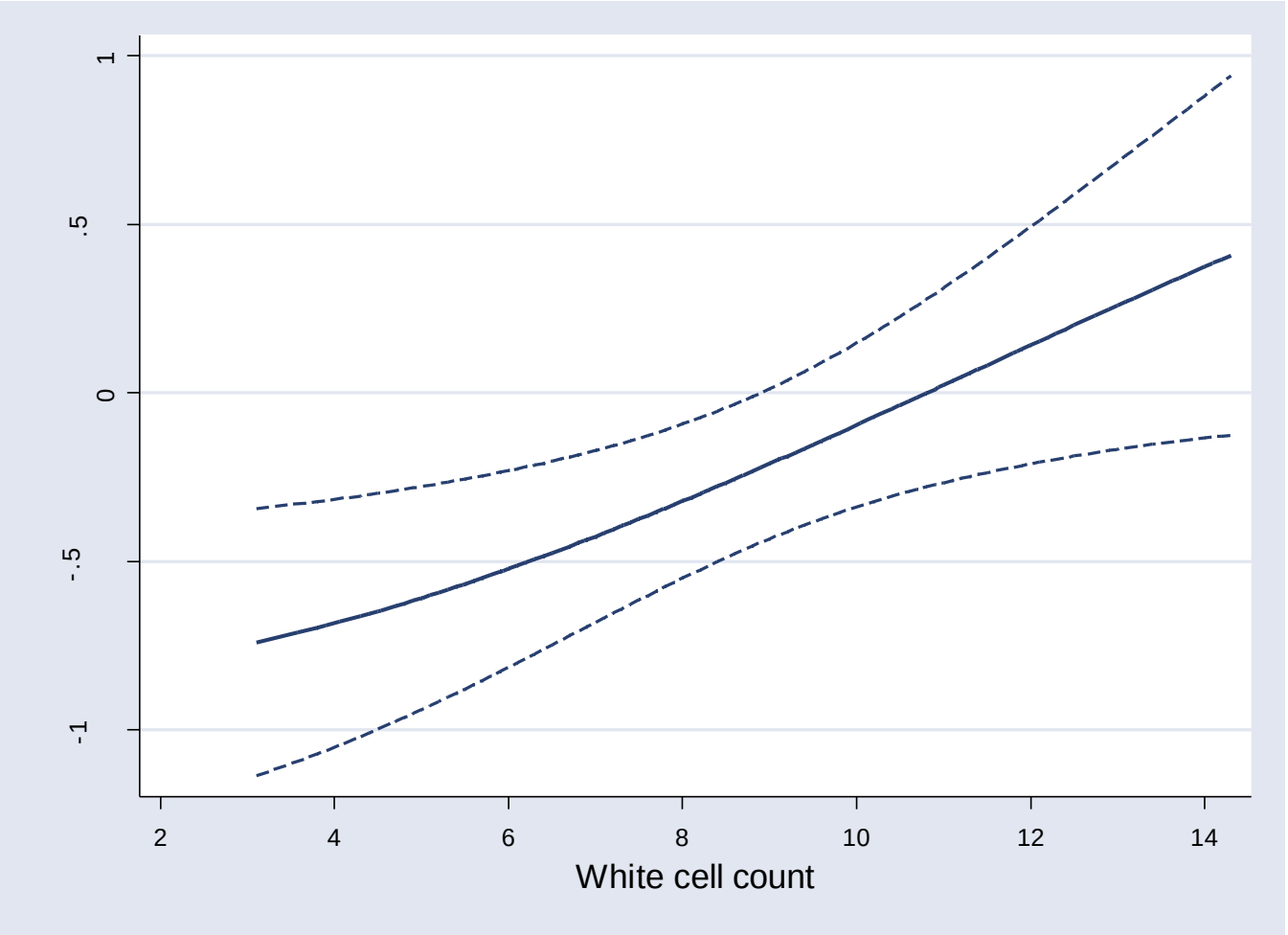
- P < 0.01; HR = 0.75; 95% CI (0.60, 0.93)
- Is the treatment effect similar in all patients? Sensible question?
  - Yes, at least for hypothesis generation
- Ten possible covariates available for the investigation of treatment-covariate interactions – only one is significant (WCC)

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

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



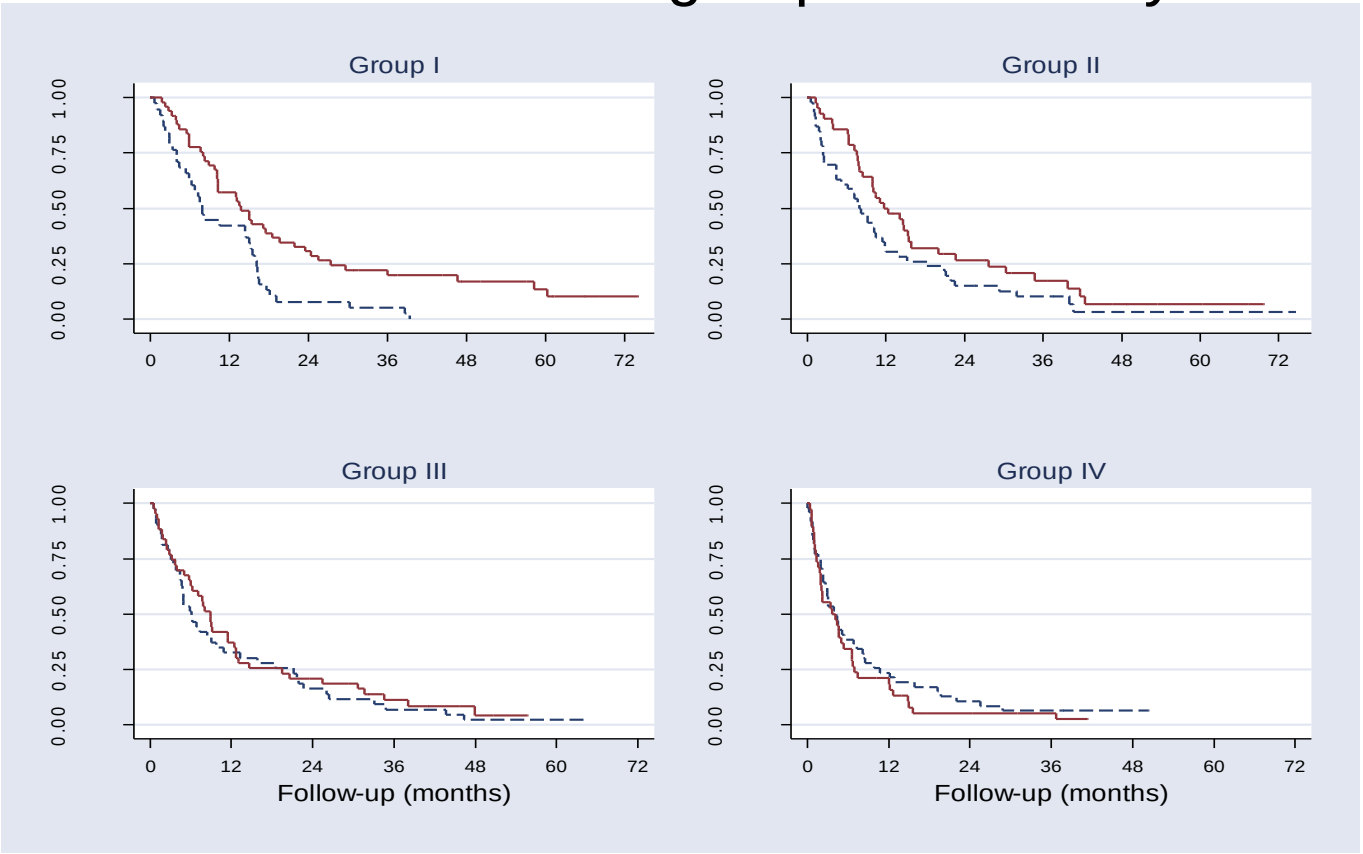
Treatment effect seems to depend on WCC



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)**

How to investigate for interactions between treatment and continuous variables?

(Evidence based) guidance is required.

Possibility to gain information:
large simulation study to compare available methods

Methods of investigating Interaction with Continuous Variables (MICVs)

MICV no.	MICV name	Class	Description
1	`lin`	Linear	Linear function at each level of *t*
2	`cat2`	Categorical	Two equal classes (one dummy variable)
3	`cat3a`	Categorical	Three equal classes (two dummy variables)
4	`cat3b`	Categorical	Three unequal classes ('Cox' cut-points: 27 and 73 centiles)
5	`cat4a`	Categorical	Four equal classes (three dummy variables)
6	`cat4b`	Categorical	Four unequal classes ('Cox' cut-points: 16.3, 50, and 83.7 centiles)
7	`score3a`	Categorical	Linear on `cat3a` scores
8	`score3b`	Categorical	Linear on `cat3b` scores
9	`score4a`	Categorical	Linear on `cat4a` scores
10	`score4b`	Categorical	Linear on `cat4b` scores
11	`fp1`	FP	FP1 function at each level of *t* (with four levels of flexibility)
12	`fp2`	FP	FP2 function at each level of *t* (with four levels of flexibility)
13	`spline`	Splines	Regression splines with 2, 3, or 4 DOF; automatic knot placement

21 approaches, 9 of them with categorized data

MFPI: four variants allowing more or less flexibility

Variant (FLEX)	Description	d.f. FP1	d.f. FP2
1	FP powers determined for main effect of x; use these same powers at each level of t	1	2
2	FP optimized over levels of t; use these same powers at each level of t	1	2
3	As FLEX2, but re-estimate powers for main effect of x	1	2
4	Optimize FP for main effect and separately at each level of t	2	4

MICVs – test for two variables in the renal cancer data

Table I. Significance levels (%) for tests of interaction between treatment and each of two continuous variables (`wcc` and `haem`) in the RE01 kidney cancer dataset.

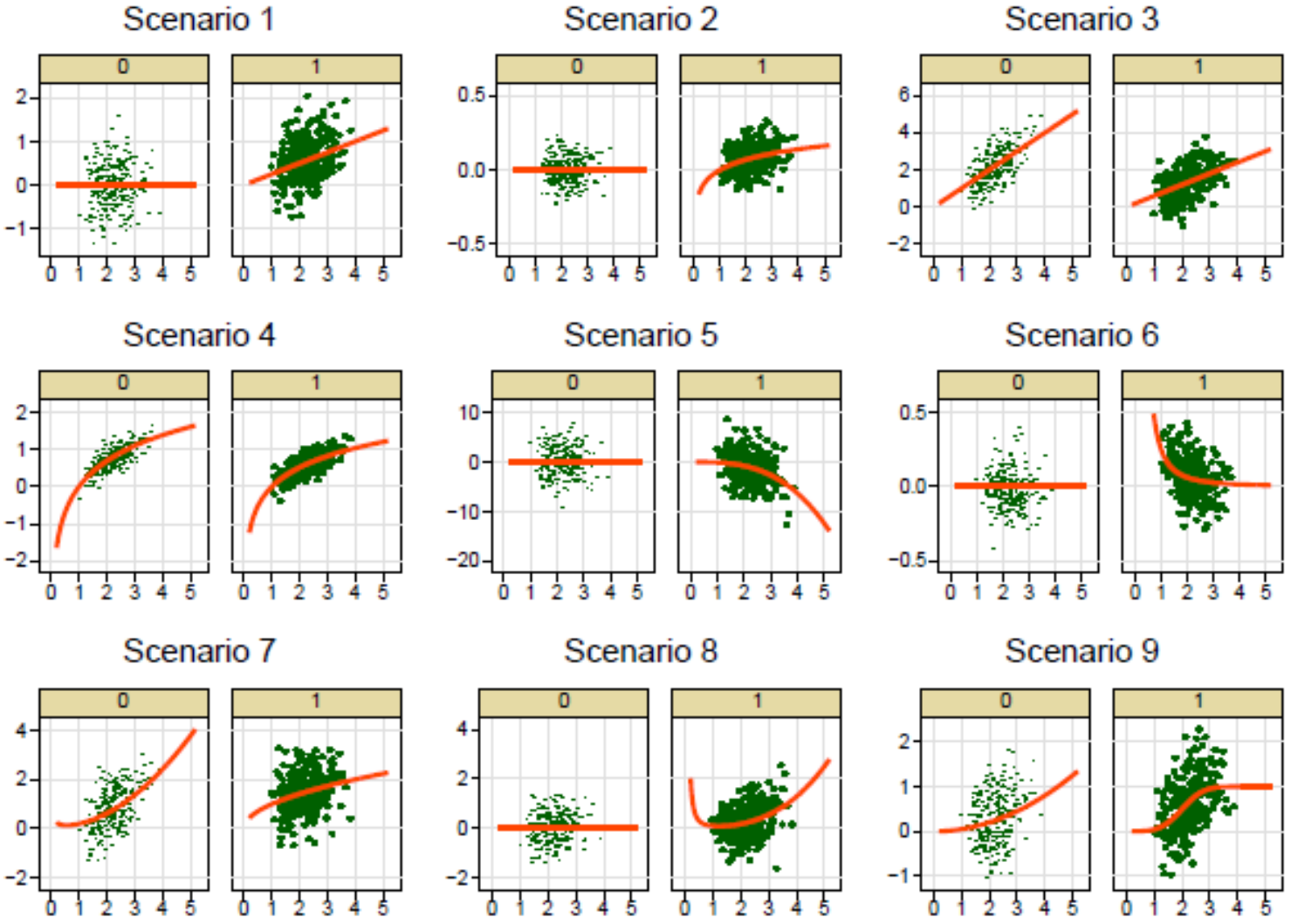
		Variable	
MICV no.	Name	`wcc`	`haem`
1	`lin`	0.5	59.8
2	`cat2`	2.3	24.0
3	`cat3a`	0.8	4.2
4	`cat3b`	2.4	26.9
5	`cat4a`	0.3	37.2
6	`cat4b`	0.3	5.3
7	`score3a`	1.6	9.7
8	`score3b`	3.9	52.6
9	`score4a`	0.8	46.6
10	`score4b`	0.8	85.8
11a	`fp1 (flex1)`	2.1	85.4
11b	`fp1 (flex2)`	0.5	85.4
11c	`fp1 (flex3)`	0.9	85.4
11d	`fp1 (flex4)`	1.7	88.3
12a	`fp2 (flex1)`	1.8	95.0
12b	`fp2 (flex2)`	0.8	95.0
12c	`fp2 (flex3)`	0.9	95.0
12d	`fp2 (flex4)`	5.4	97.7
13a	`spline (2 DOF)`	1.8	95.4
13b	`spline (3 DOF)`	3.2	70.0
13c	`spline (4 DOF)`	4.3	88.7

Royston P, Sauerbrei W (2013), Stat Med, 32: 3788-3803.

Simulation study

- 9 scenarios with different functional forms and size of effects
- Distribution of x: about normal (well behaved), skewed (badly behaved)
- Sample sizes 250 and 500
- 21 MICVs

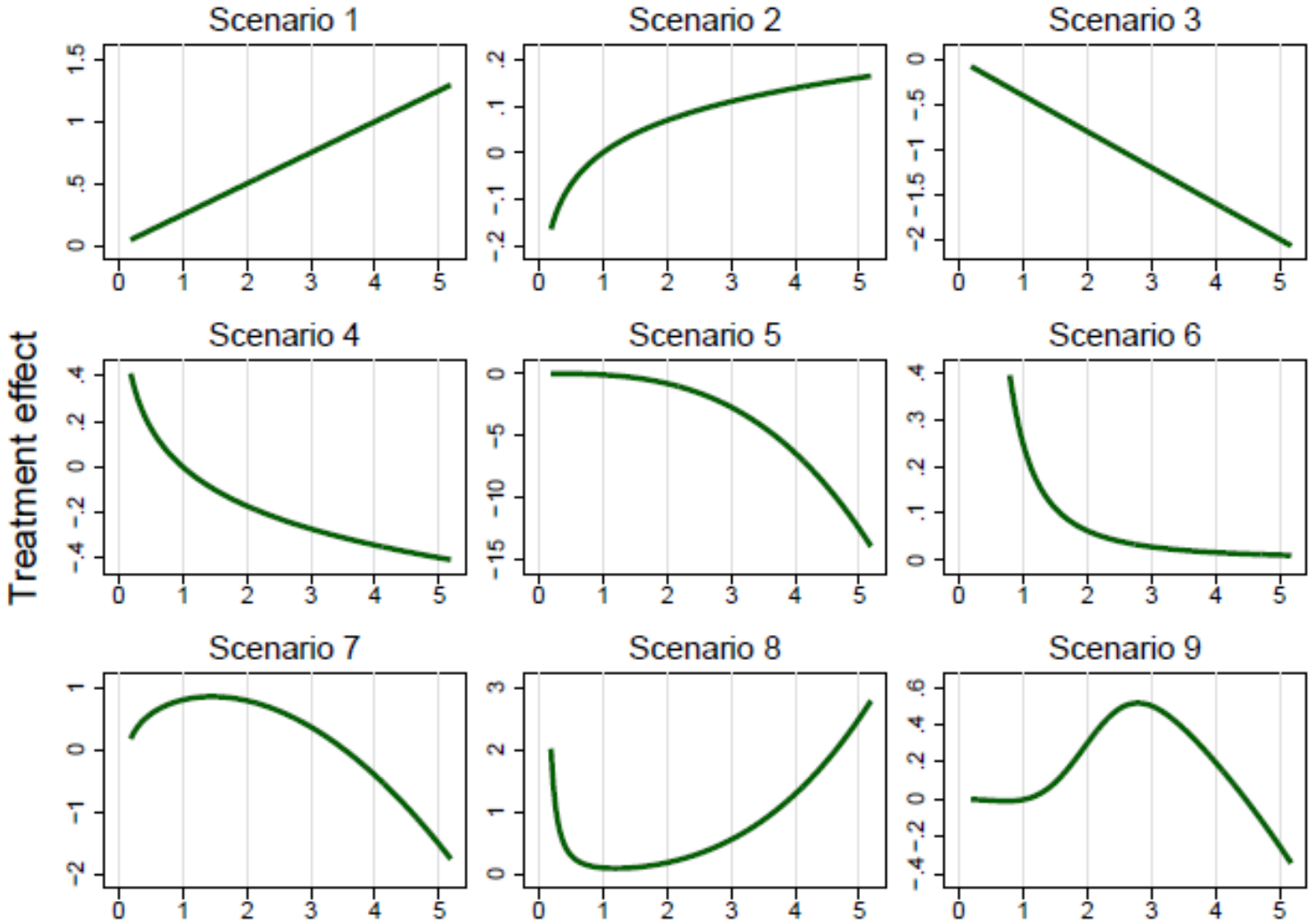
Nine scenarios with interaction



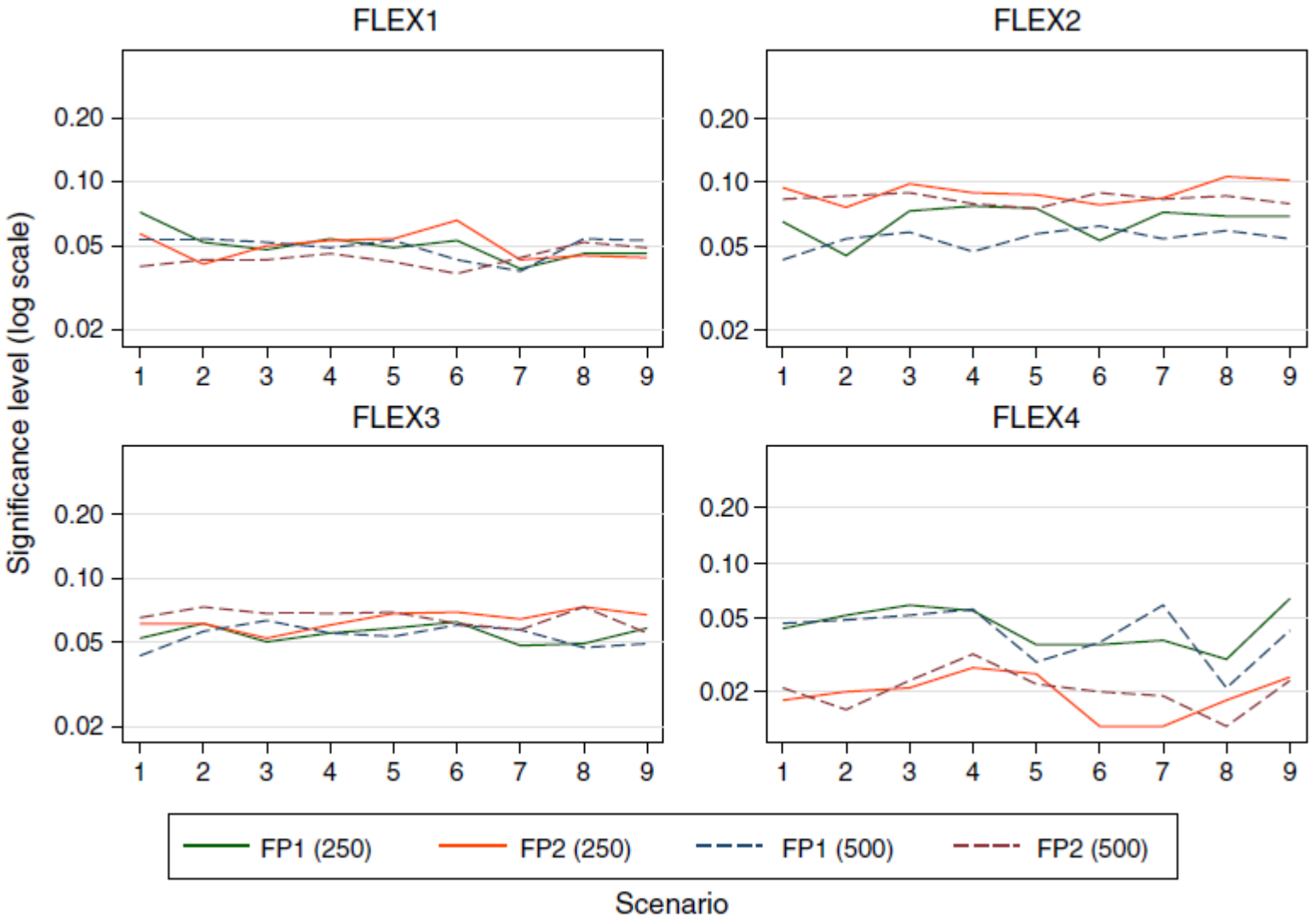
No interaction:

average the prognostic effect across the two groups

Scenarios with interaction: Treatment effect functions

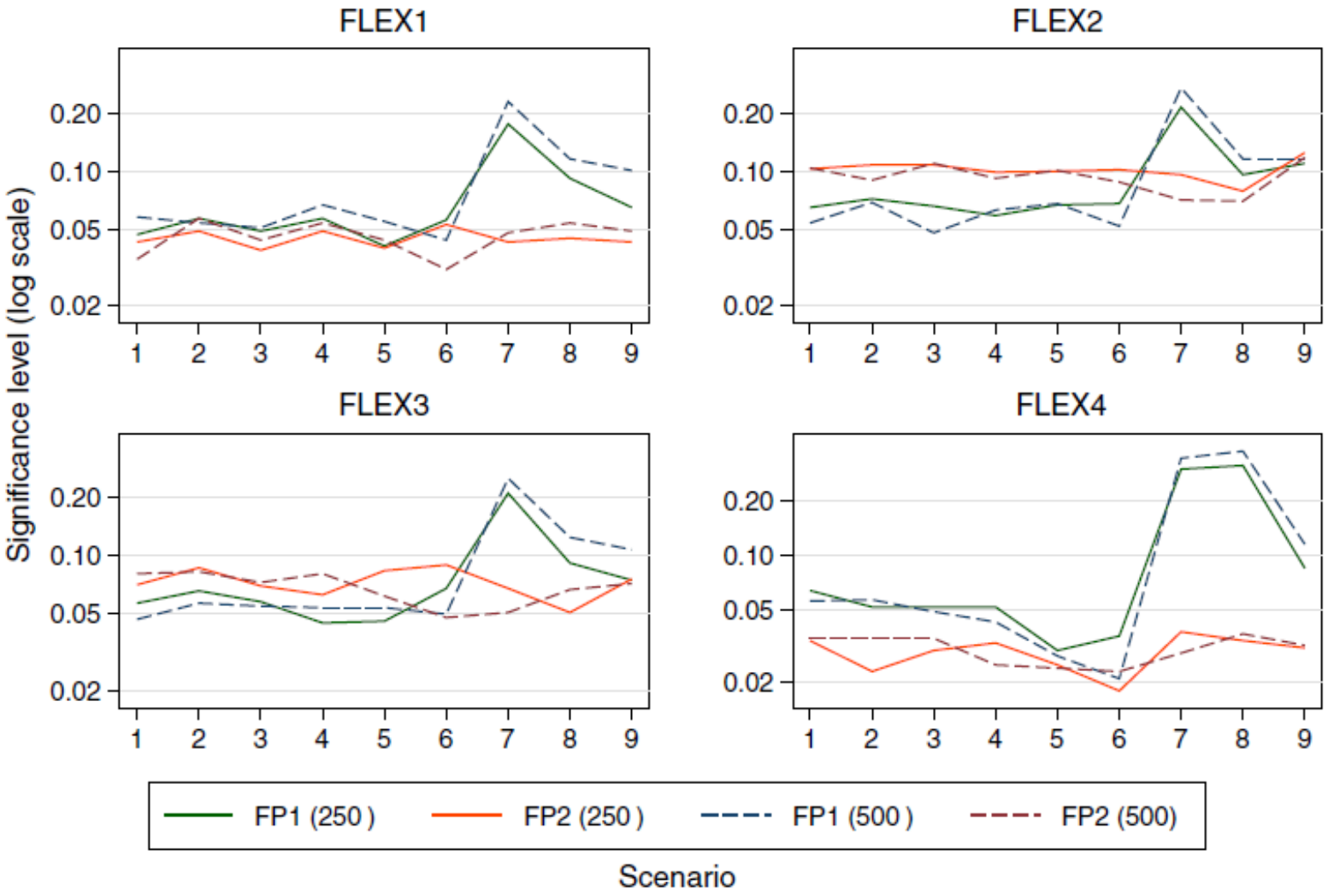


No interaction (nom. sig. level 5%)



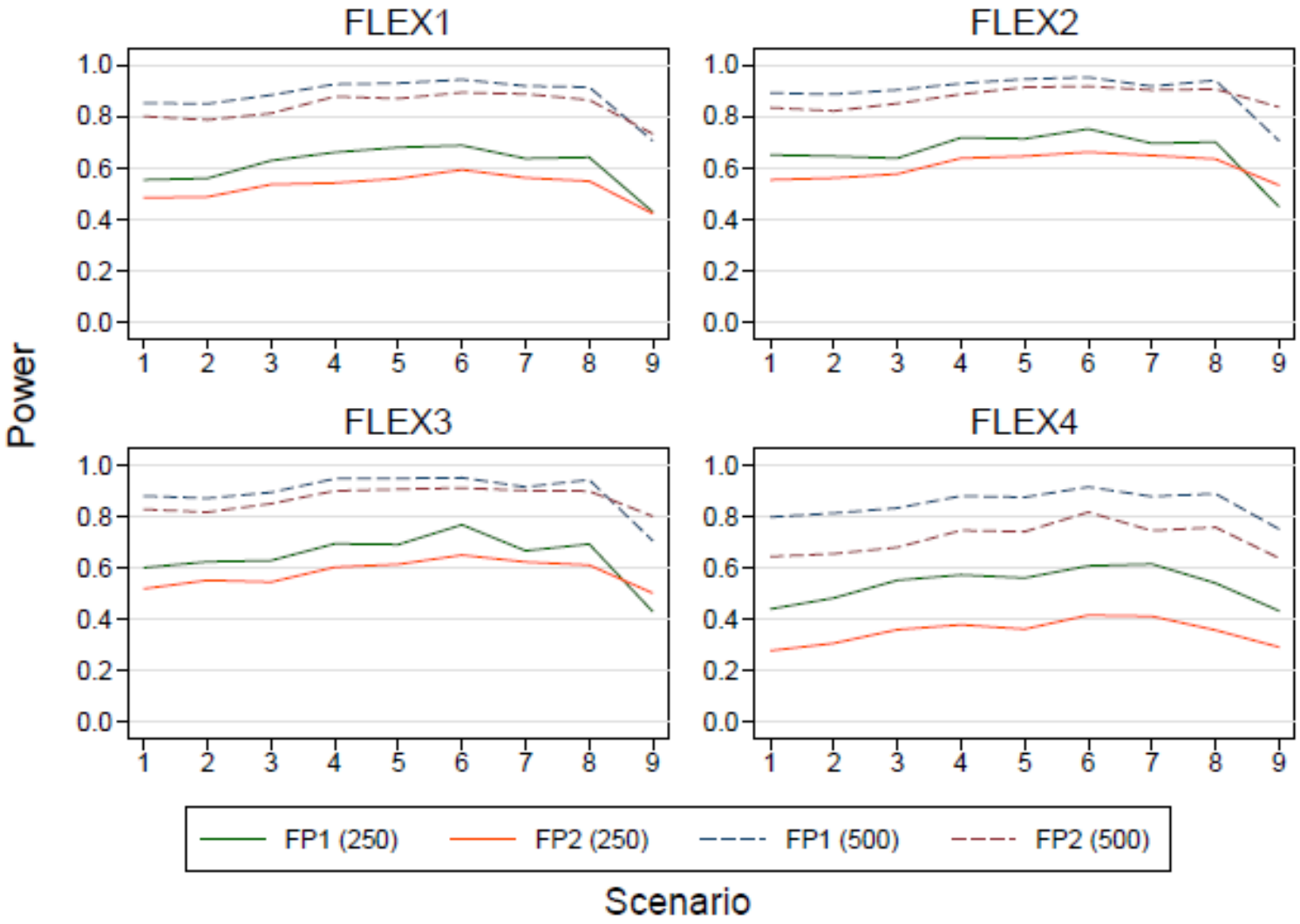
MFPI: Flex1 – Flex4 well behaved case

No interaction (nom. sig. level 5%)



MFPI: Flex1 – Flex4 badly behaved case

Power



MFPI: Flex1 – Flex4 well behaved case

Type I error and Power – MFPI and competitors

Table IV. Mean significance levels (%) for the well-behaved case.

MICV No.	Name		Sample size 250	500
1	lin	★	6.0	5.6
2	cat2		5.7	5.5
3	cat3a		5.1	5.1
4	cat3b		5.7	5.3
5	cat4a		5.5	5.1
6	cat4b		6.1	5.5
7	score3a		5.9	5.0
8	score3b		5.5	5.3
9	score4a		5.9	5.4
10	score4b		6.1	5.5
11a	fp1 (flex1)		5.1	5.0
11b	fp1 (flex2)		6.6	5.4
11c	fp1 (flex3)		5.5	5.4
11d	fp1 (flex4)		5.0	4.4
12a	fp2 (flex1)		4.6	4.4
12b	fp2 (flex2)	★	9.0	8.3
12c	fp2(flex3)	★	6.4	6.5
12d	fp2(flex4)	★	2.0	2.1
13a	spline (2 DOF)	★	5.6	5.3
13b	spline (3 DOF)		5.7	5.5
13c	spline (4 DOF)		5.9	5.4

Values are averages for each sample size over all nine scenarios.

★ Unsuitable significance level, therefore ignored in the summary

Type I (badly behaved)	**Power (well-behaved)**		
	Sample size		
Average (1-6)	250	500	Average
10.1	61	88	74.5
5.3	40	69	54.5
5.3	41	71	56
5.7	43	73	58
5.9	40	70	55
6.4	42	74	58
5.6	49	79	64
6.0	51	80	65.5
5.9	53	82	67.5
6.2	54	83	68.5
5.3	61	88	74.5
6.2	66	90	78
5.5	64	90	77
4.5	53	85	69
4.4	53	84	68.5
10.0	61	88	74.9
7.4	58	87	72.5
2.8	35	71	53
7.2	56	86	71
5.9	49	82	65.5
6.1	45	78	61.5

proposal 12a or 11c

Alternative approach for continuous variables

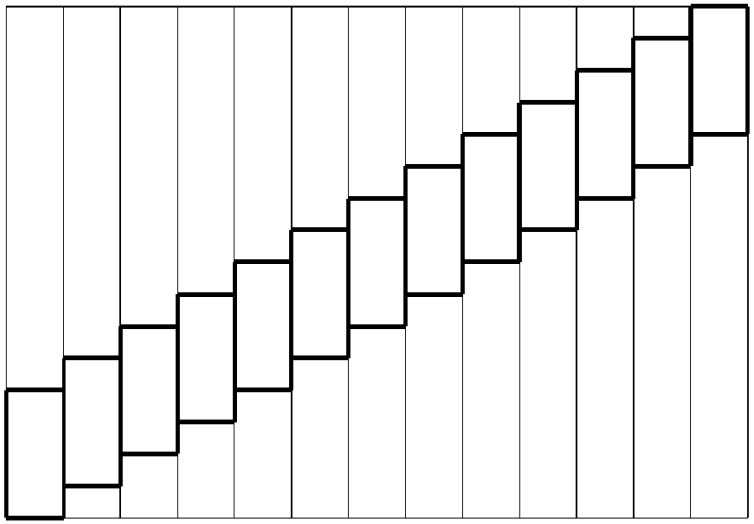
- STEPP

  Subpopulation Treatment Effect Pattern Plots

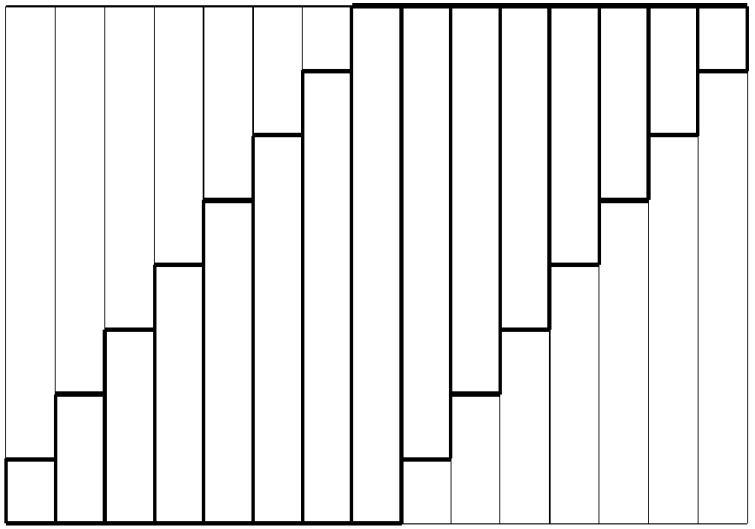
  Bonetti & Gelber 2000, 2004

STEPP

Sequences of overlapping subpopulations



Sliding window



Tail oriented

STEPP

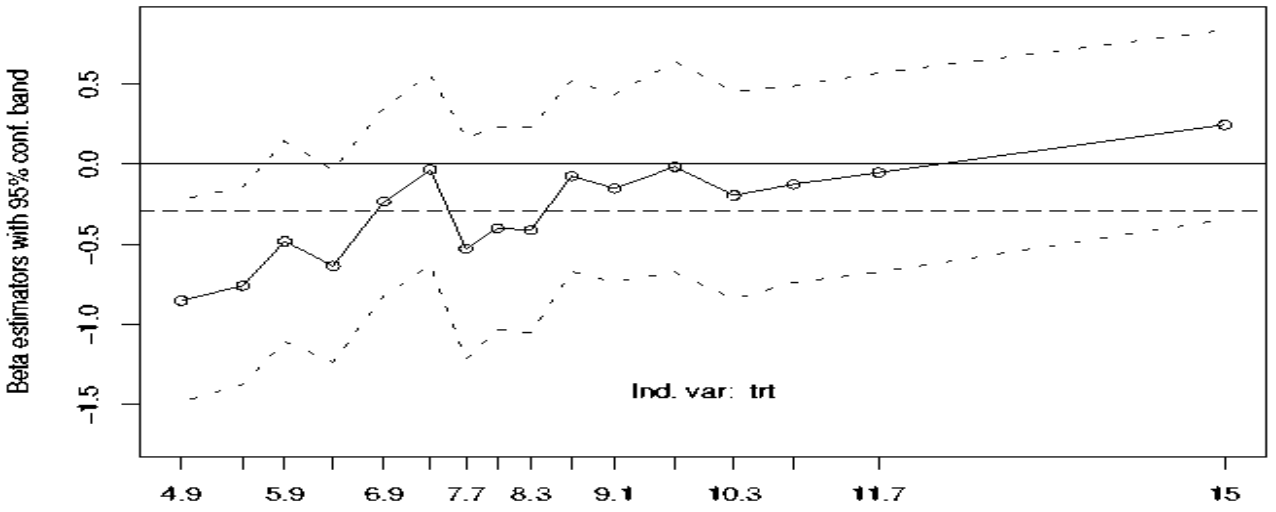
Estimate treatment effect in each subpopulation

Overlapping populations, therefore correlation between the estimates

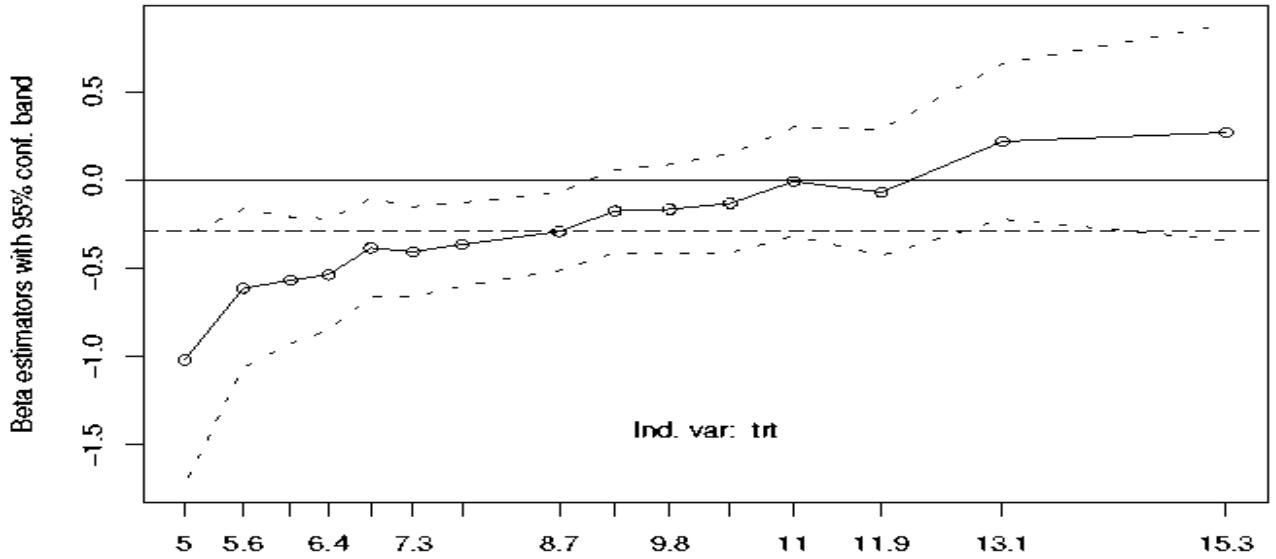
Simultaneous confidence band and tests proposed

STEPP – Interaction with WCC

SLIDING WINDOW (n1 = 25, n2 = 40)

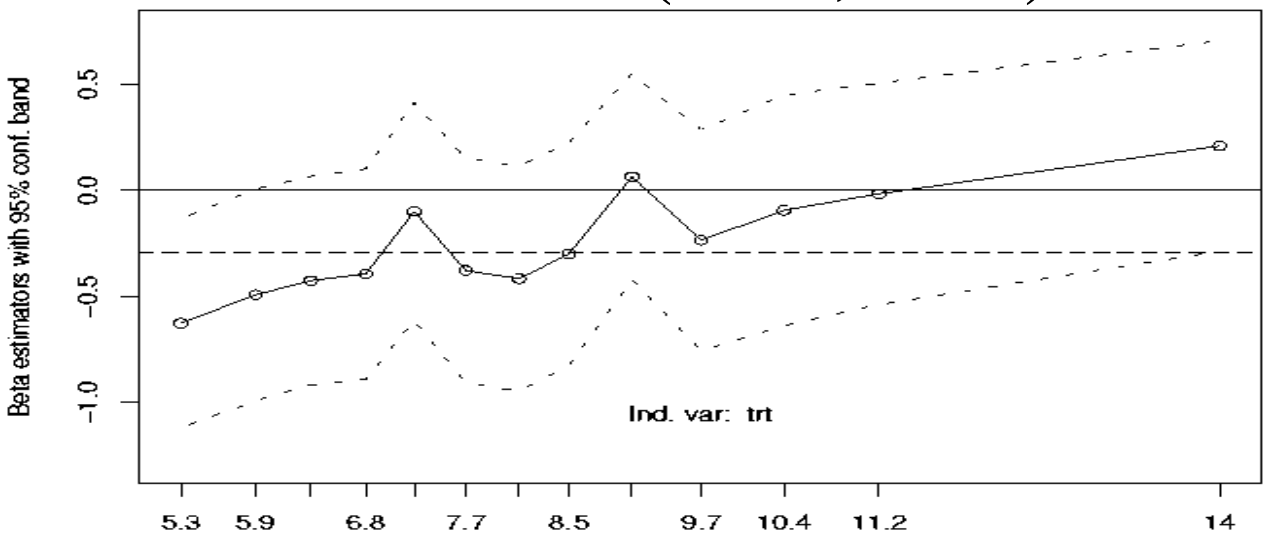


TAIL ORIENTED (g = 8)

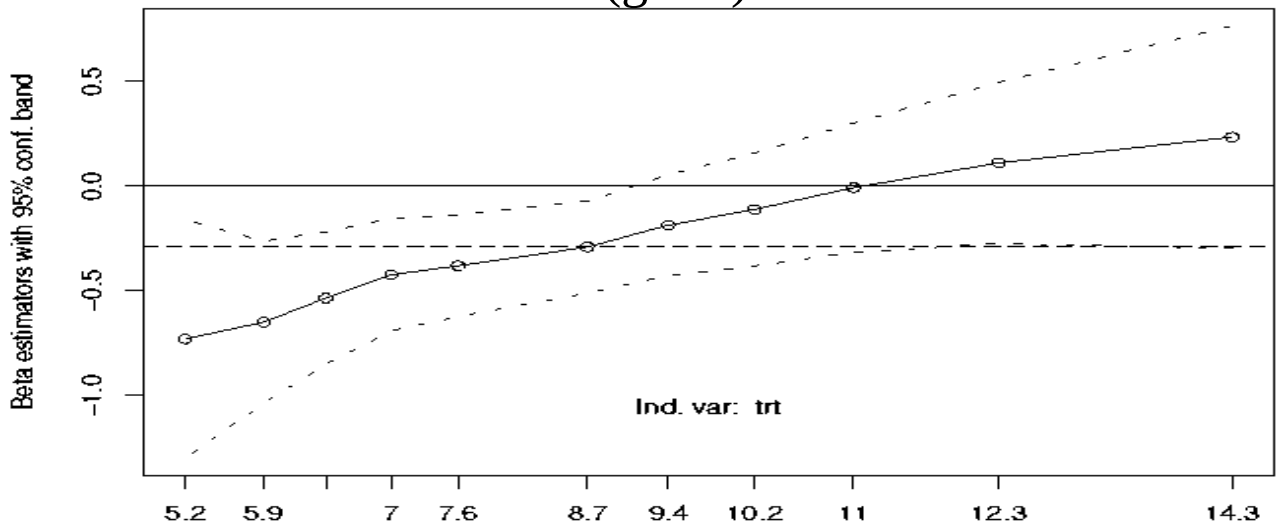


STEPP – Interaction with WCC

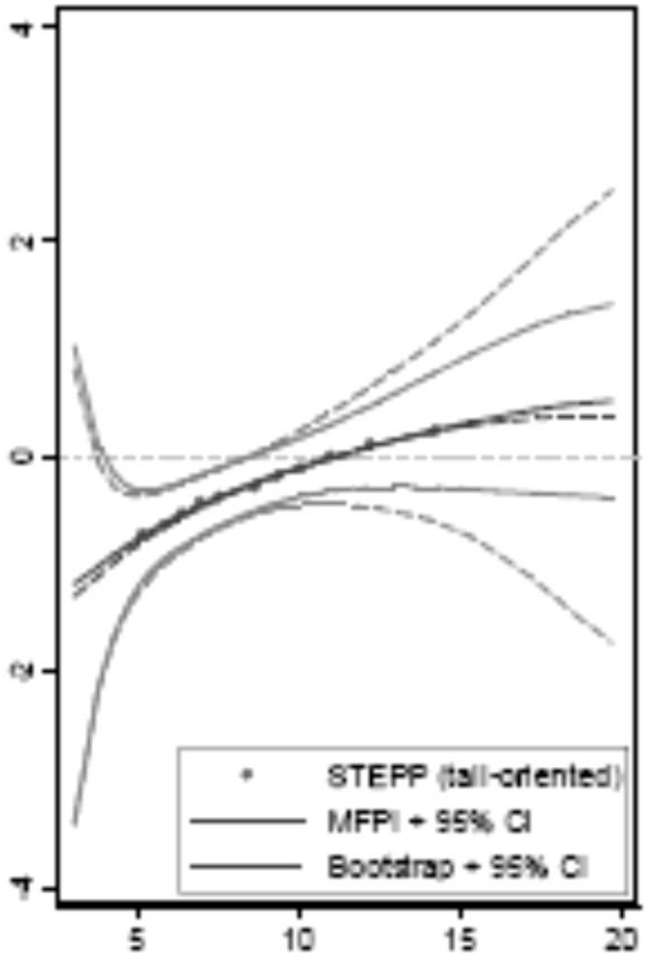
SLIDING WINDOW (n1 = 40, n2 = 60)



TAIL ORIENTED (g = 6)



STEPP as check of MFPI



Meta-analysis of continuous functions – Basic idea

- Meta-analysis of continuous covariates based on **cutpoints** and dummy variables is **not sufficient**
- MA of regression coefficients for **linear functions** is **sometimes acceptable**
- When **non-linearity** is present, MA of FP functions **may offer a way forward**
- Several **critical points** need more consideration, experience and investigation
- Idea developed in the context of observational studies, but useful for MA of **treatment effect functions**

ICEM Study

Investigation of Continuous Effect Modifiers Study (ICEM)

Meta-analysis of treatment effect functions from several RCTs

Patients and Endpoints

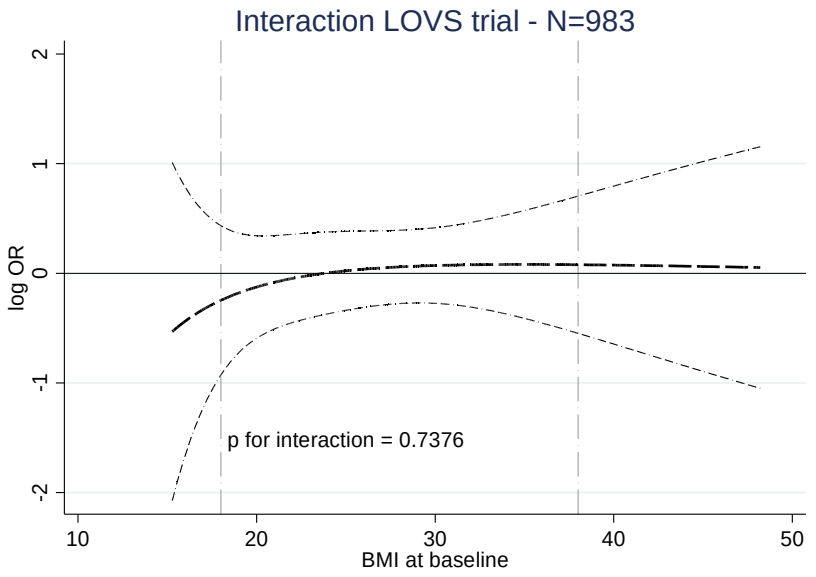
- Patients from 3 different randomized trials
  - Critically ill patients with acute lung injury or ARDS requiring mechanical ventilation
- Intervention
  - Higher PEEP versus Lower PEEP ventilation strategy
- Primary efficacy endpoint
  - In hospital mortality 60 days after randomization (binary, logistic regression)

Outcome 60 days mortality

P-values from tests for interaction in individual studies

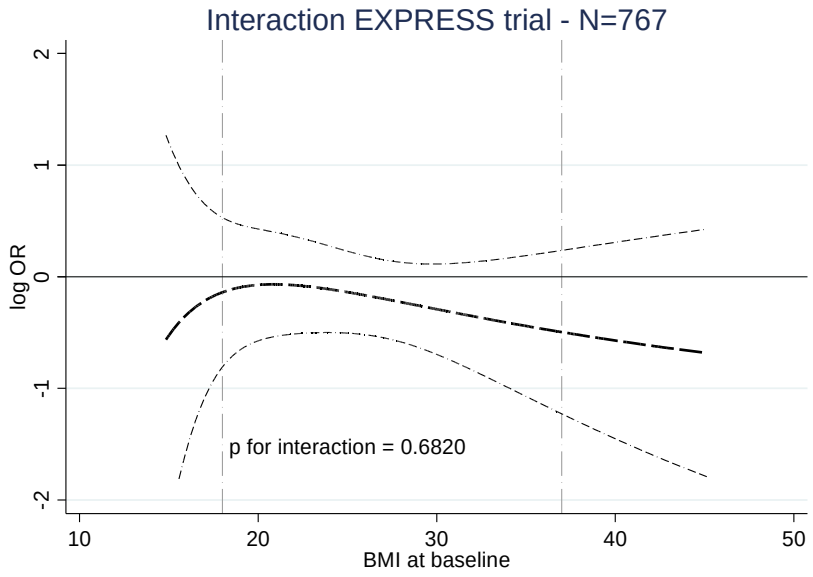
	BMI
LOVS	0.7376
EXPRESS	0.6820
ALVEOLI	0.1842

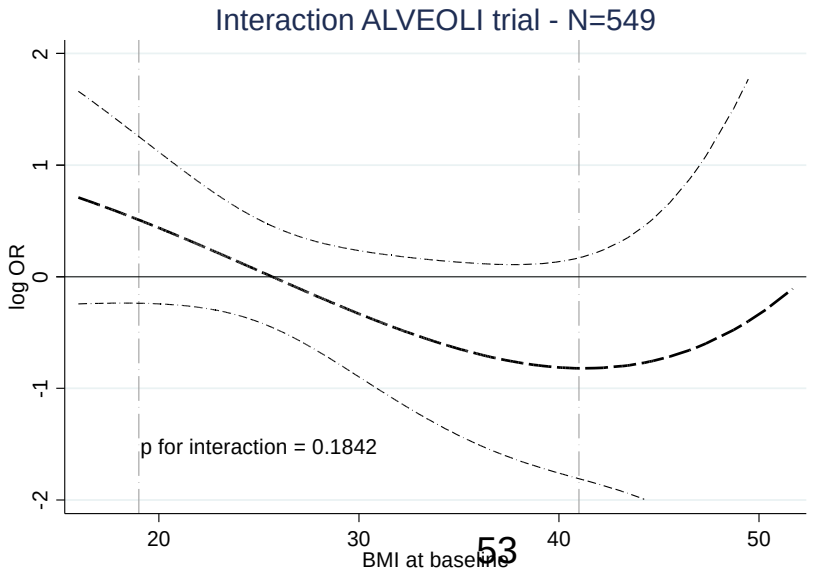
RESULTS – TEF in individual studies
Interaction BMI



log OR<0: Higher PEEP better

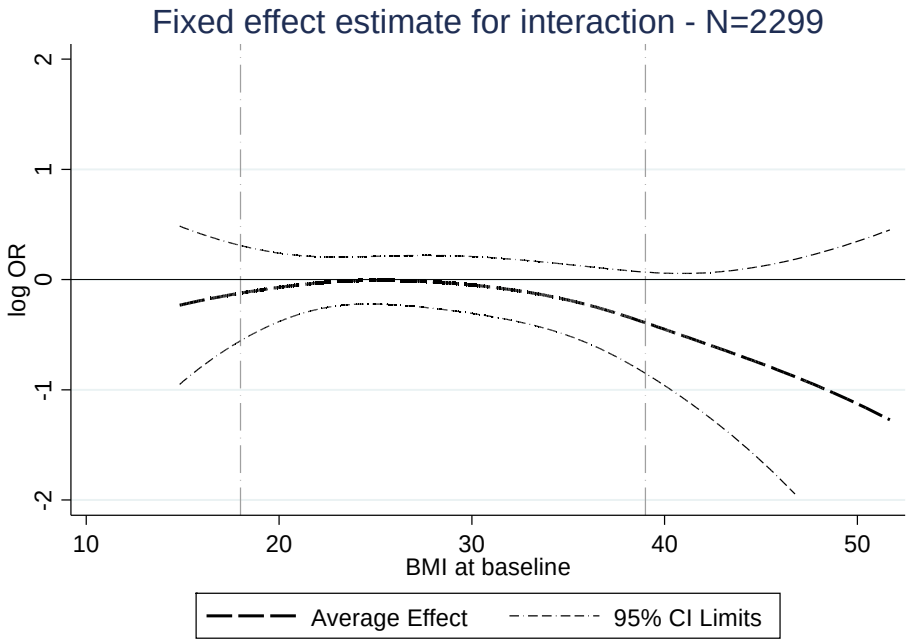
95% CI of TEF model based, therefore too small





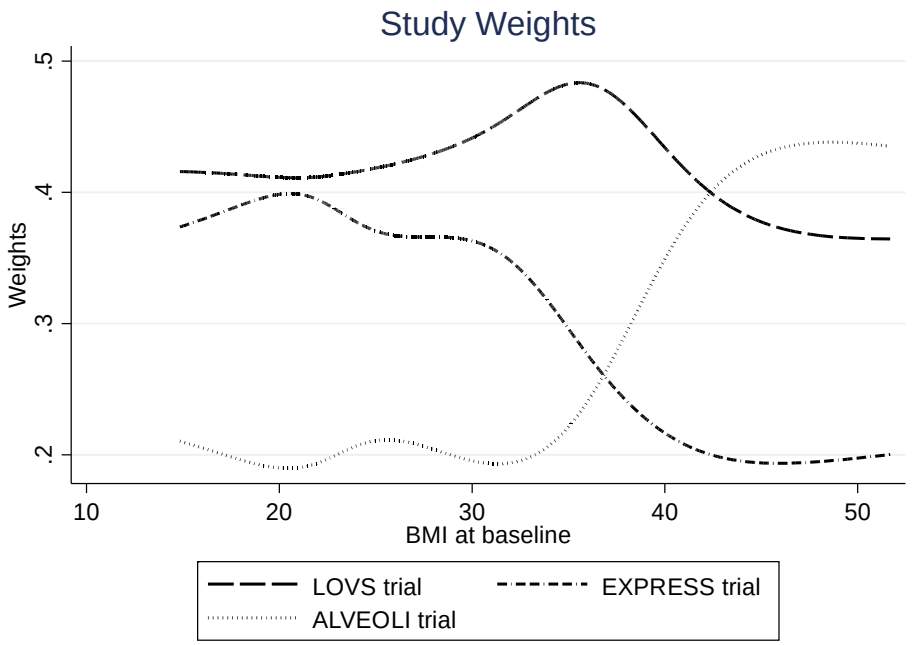
RESULTS – Meta-analysis Interaction BMI

TEF_{mean}



Pointwise average with 95% CI

Weights per study



Weights depend on information (SE) of each study at a specific point

Summary - Interactions

- 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
- Internal check of MFPI result is required,
  external data to validate results
- With IPD a meta-analysis of TEFs provides further insight

References

MFP

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.

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

Subgroups, MFPI and alternatives

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.

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.

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.

Royston P, Sauerbrei W (2013):
Interaction of treatment with a continuous variable: simulation study of significance level for several methods of analysis. Statistics in Medicine, 32: 3788-3803.

Royston P., Sauerbrei W. (2014):
Interaction of treatment with a continuous variable: simulation study of power for several methods of analysis. Statistics in Medicine, 33: 4695-4708.

Meta-analysis

Sauerbrei W, Royston P (2011):
A new strategy for meta-analysis of continuous covariates in observational studies. Statistics in Medicine 30 (28): 3341-60.

Acute Lung Injury

Briel M, Meade M, Mercat A, Brower RG; Talmor D; Walter SD; Slutsky AS; Pullenayegum E; Zhou Q; Cook D; Brochard L; Richard JC; Lamontagne F; Bhatnagar N; Stewart TE; Guyatt G (2010):
Higher vs lower positive end-expiratory pressure in patients with acute lung injury and acute respiratory distress syndrome: systematic review and meta-analysis. JAMA 9 (303): 865–873.

Kasenda B, Sauerbrei W, Royston, Briel M (2014):
Investigation of continuous effect modifiers in a meta-analysis on higher versus lower PEEP in patients requiring mechanical ventilation - protocol of the ICEM study. *Systematic Reviews* 3(46).