Lecture 4
Multivariable Fractional Polynomials (MFP)
An approach to select variables and derive functional forms for continuous variables

Willi Sauerbrei, Edwin Kipruto

Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center -University of Freiburg, Germany

WS 2025/26

Learning objectives

- Understand that correlation between variables influences the selection of variables and of the functional form
- Introduce MFP as a sensible strategy for the selection of variables and functional form for continuous variables
- Understand how subject-matter knowledge can be incorporated into model building
- Understand that model complexity is a key issue of models selected and how to control it for MFP models
- Understand that every selected model is based on a variety of assumptions and that a critical assessment is required
- Understand that it is important to get the BIG picture right

Plan

1. Multivariable analysis required
2. The MFP algorithm
3. Presentation of FP models
4. Tuning parameters for variable and function selection
5. Model criticism
6. Contribution of individual components
7. Big Data
8. Further comments on selection of variables and functional forms in multivariable analysis

Much of the material is taken from Royston and Sauerbrei (2008) book

1. Multivariable analysis required

In the first lecture we have shown how to derive an FP function in a univariate analysis.
However, usually we have further variables and need to build a multivariable model.
In the prostate data we estimated the functional form for the influence of
cancer volume (cavol) on the outcome log PSA concentration.

We have seen that log(cavol) fits the data well, certainly much better than the linear function.

We have 6 additional variables

Age, svi, pgg45, weight, bph and cp.

Which factors have an influence on the outcome?
What is the functional form for continuous variables?

MFP models

Typically, a mix of continuous and binary covariates is available

- Dummy variables for categorical predictors

– Dummy coding should reflect ordered nature

Wish to find ‘best’ multivariable FP model

Impractical to try all combinations of powers for all continuous covariates

Requires iterative fitting procedure

Implicit assumptions:

- Subject matter knowledge is restricted (can be included otherwise)
- Main interest in effect of individual variables, overall predictor alone is not sufficient

Some aims of MFP models

- One continuous variable of main interest, but adjustment for other variables is required
  - in principle ‚simple‘ extension of univariate dose response modelling. In epidemiology popular to use several cutpoints for a continuous variable.
  - next analysis: main variable of interest is one of the adjustment variables from above.
  - next analysis (and next paper): main variable is another ‚old‘ adjustment variable

Many papers but scientifically to criticize

- One model including all relevant predictors of interest.
  - Keep the full information for each continuous variable and determine the dose-response functions simultaneously. Irrelevant predictors (according to subject-matter and significance value) should be excluded.

2. The MFP algorithm

Definition

Re-analyses of several studies

- multivariable instead of univariate
- non-linear instead of linear

different assumptions => different results?

Presentation of (M)FP models

The MFP algorithm

- COMBINE backward elimination with a search for the best FP functions
- START: Determine fitting order from linear model
- UPDATE: Apply univariate FP model selection procedure to each continuous X in turn, adjusting for (last FP function of) each other X
- UPDATE: Categorical covariates similarly – but just in/out of model
- CYCLE: until convergence – usually 2-3 cycles

Tuning parameters of MFP

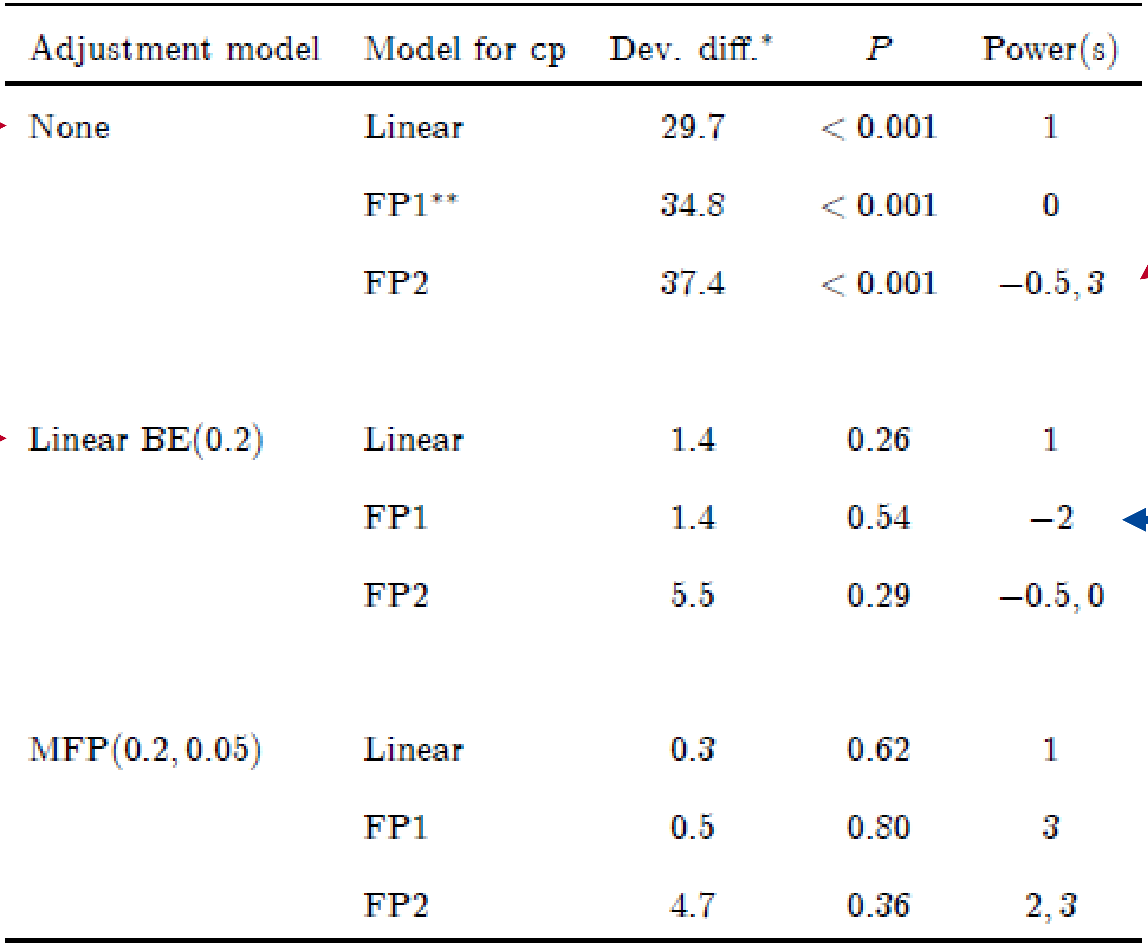
- Significance level for
  variable selection
  function selection
  may be different

- Notation MFP(0.2,0.05)

- Subject-matter knowledge can (**and should**) be included
  eg. inclusion of a specific confounder
  function should be monotonic

Plan

Prostate data- effect of cp in models with or without adjustment

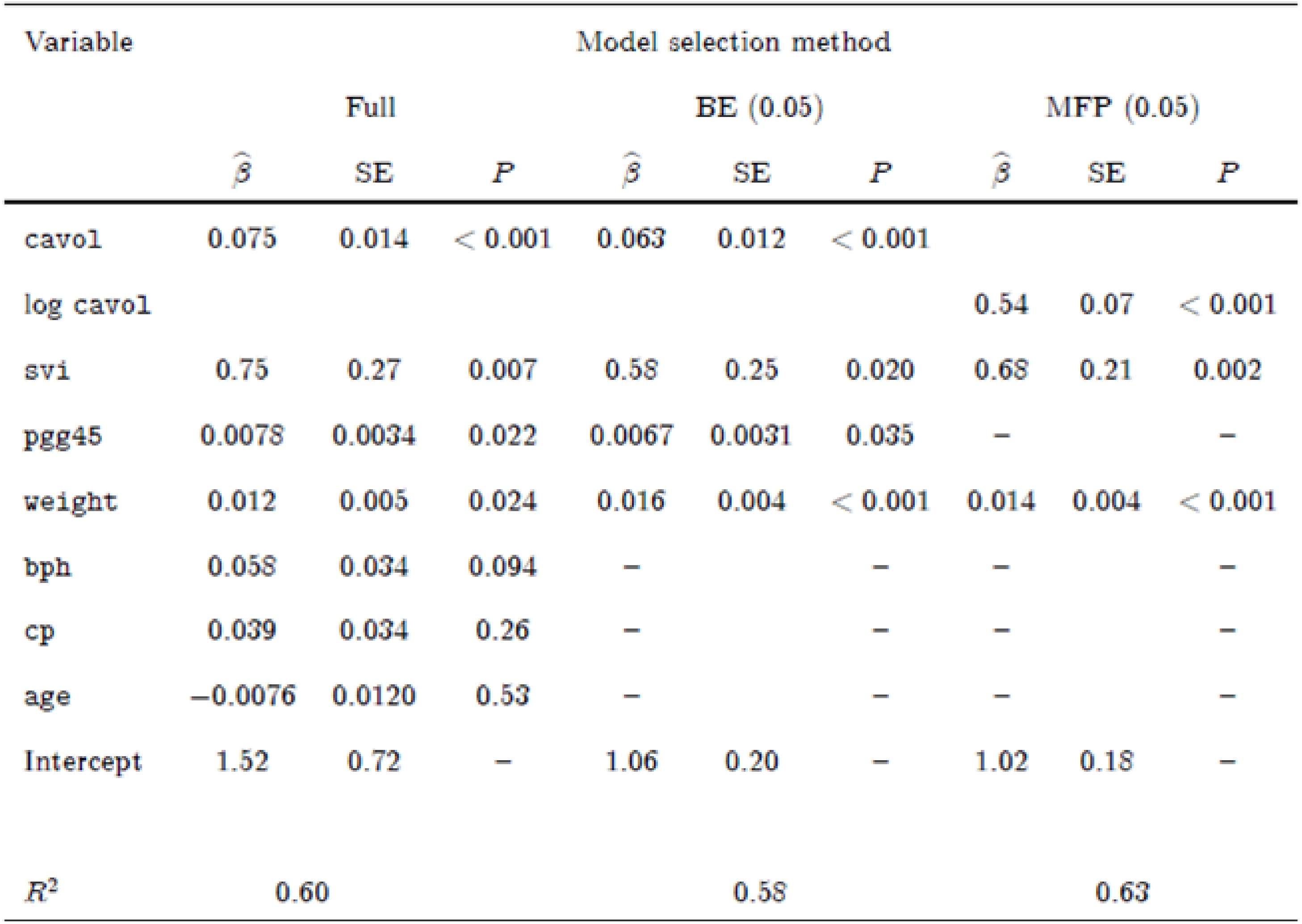


Adjustment model	Model for cp	Dev. diff.*	P	Power(s)
None	Linear	29.7	< 0.001	1
	FP1**	34.8	< 0.001	0
	FP2	37.4	< 0.001	$-0.5, 3$
Linear BE(0.2)	Linear	1.4	0.26	1
	FP1	1.4	0.54	-2
	FP2	5.5	0.29	$-0.5, 0$
MFP(0.2, 0.05)	Linear	0.3	0.62	1
	FP1	0.5	0.80	3
	FP2	4.7	0.36	2, 3

highly significant

not significant

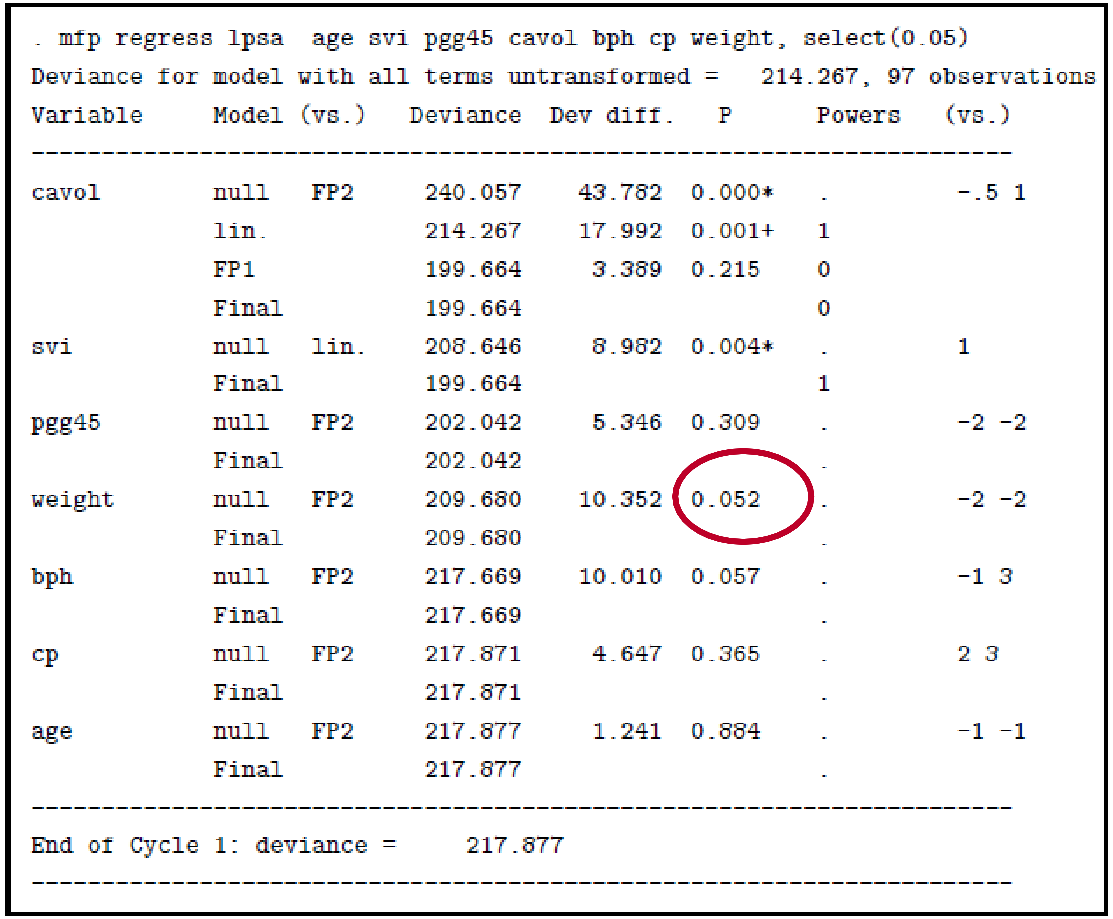
Multivariable analyses selection (no/yes) and non-linearity (no/yes) in prostate data



Variable	Model selection method								
	Full			BE (0.05)			MFP (0.05)		
	$\hat{\beta}$	SE	P	$\hat{\beta}$	SE	P	$\hat{\beta}$	SE	P
cavol	0.075	0.014	< 0.001	0.063	0.012	< 0.001			
log cavol							0.54	0.07	< 0.001
svi	0.75	0.27	0.007	0.58	0.25	0.020	0.68	0.21	0.002
pgg45	0.0078	0.0034	0.022	0.0067	0.0031	0.035	–		–
weight	0.012	0.005	0.024	0.016	0.004	< 0.001	0.014	0.004	< 0.001
bph	0.058	0.034	0.094	–		–	–		–
cp	0.039	0.034	0.26	–		–	–		–
age	−0.0076	0.0120	0.53	–		–	–		–
Intercept	1.52	0.72	–	1.06	0.20	–	1.02	0.18	–
R^2	0.60			0.58			0.63		

R&S 2008, Tab 6.1

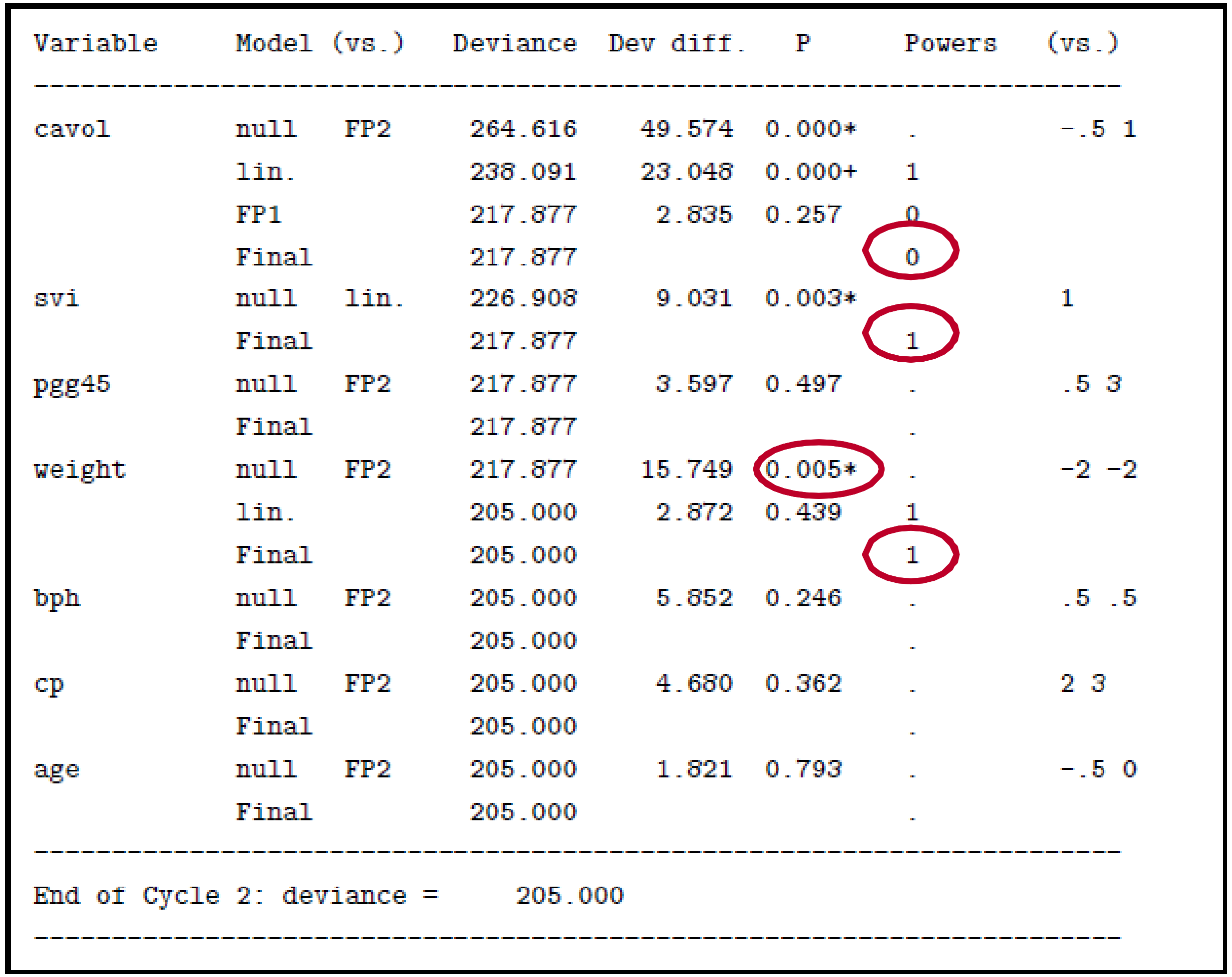
MFP algorithm – cycle 1



```
. mfp regress lpsa  age svi pgg45 cavol bph cp weight, select(0.05)
Deviance for model with all terms untransformed =   214.267, 97 observations
Variable     Model (vs.)   Deviance  Dev diff.   P      Powers   (vs.)
----------------------------------------------------------------------
cavol        null   FP2     240.057    43.782  0.000*   .          -.5 1
             lin.           214.267    17.992  0.001+   1
             FP1            199.664     3.389  0.215    0
             Final          199.664                     0
svi          null   lin.    208.646     8.982  0.004*   .          1
             Final          199.664                     1
pgg45        null   FP2     202.042     5.346  0.309    .          -2 -2
             Final          202.042                     .
weight       null   FP2     209.680    10.352  0.052    .          -2 -2
             Final          209.680                     .
bph          null   FP2     217.669    10.010  0.057    .          -1 3
             Final          217.669                     .
cp           null   FP2     217.871     4.647  0.365    .          2 3
             Final          217.871                     .
age          null   FP2     217.877     1.241  0.884    .          -1 -1
             Final          217.877                     .
----------------------------------------------------------------------
End of Cycle 1: deviance =     217.877
----------------------------------------------------------------------
```

R&S 2008, Box 6.1

MFP algorithm – cycle 2



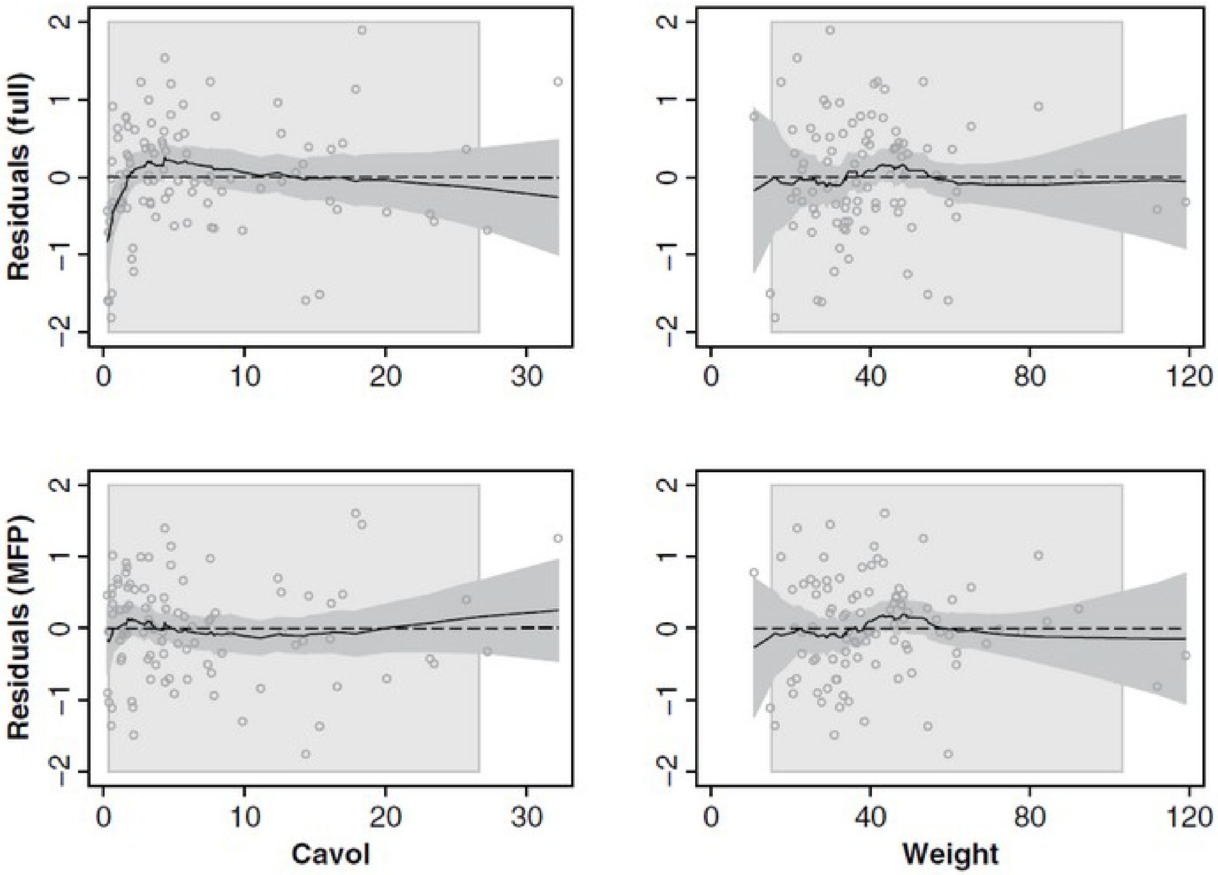
```
Variable     Model  (vs.)   Deviance  Dev diff.   P      Powers   (vs.)
----------------------------------------------------------------------
cavol        null    FP2     264.616    49.574  0.000*    .         -.5 1
             lin.            238.091    23.048  0.000+    1
             FP1             217.877     2.835  0.257     0
             Final           217.877                      0
svi          null    lin.    226.908     9.031  0.003*              1
             Final           217.877                      1
pgg45        null    FP2     217.877     3.597  0.497     .         .5 3
             Final           217.877                      .
weight       null    FP2     217.877    15.749  0.005*    .         -2 -2
             lin.            205.000     2.872  0.439     1
             Final           205.000                      1
bph          null    FP2     205.000     5.852  0.246     .         .5 .5
             Final           205.000                      .
cp           null    FP2     205.000     4.680  0.362     .         2 3
             Final           205.000                      .
age          null    FP2     205.000     1.821  0.793     .         -.5 0
             Final           205.000                      .
----------------------------------------------------------------------
End of Cycle 2: deviance =     205.000
----------------------------------------------------------------------
```

R&S 2008, Box 6.2

Final MFP(0.05,0.05) model for prostate data

	power	
cavol	0	log cavol
svi	1	binary
weight	1	linear

Check residuals

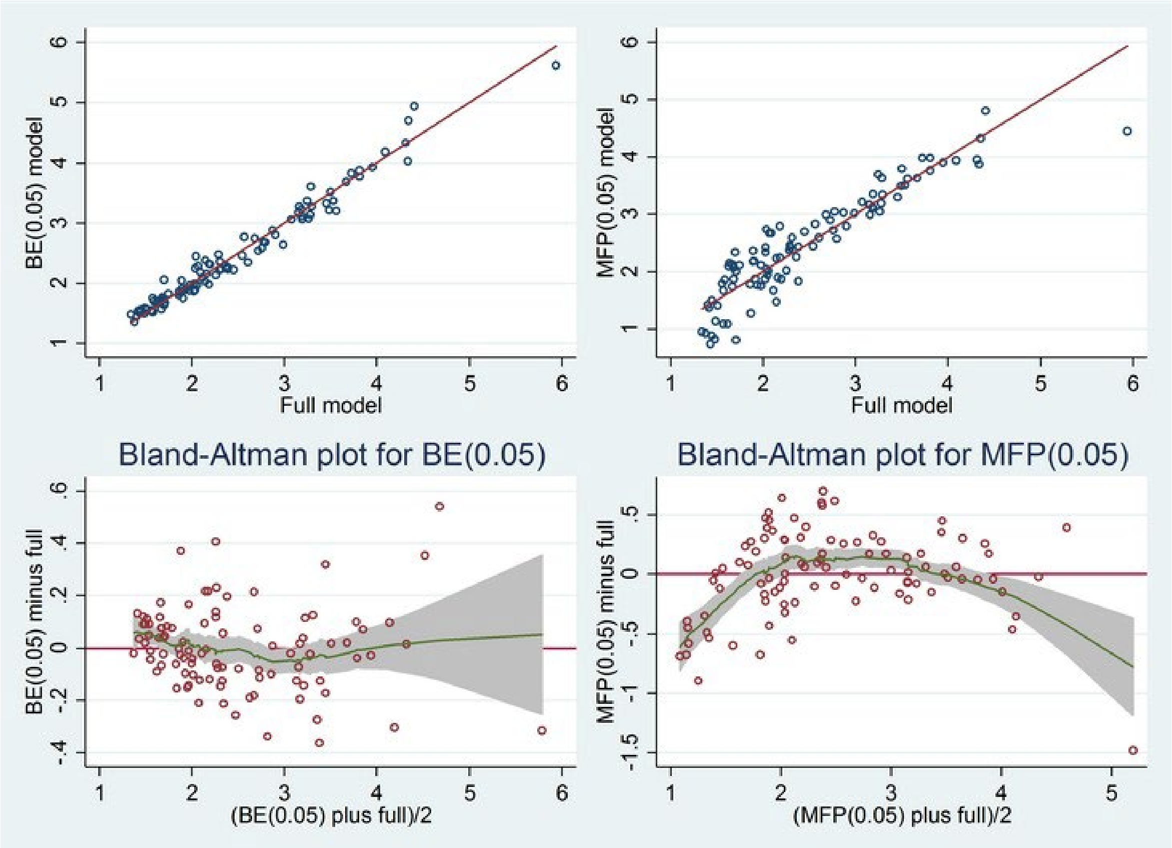


Reproduced from R&S (2008) with permission from John Wiley & Sons Ltd.

Predictions

Full – BE

Full - MFP



Example: Prognostic factors in breast cancer

Aim to develop a prognostic index for risk of tumour recurrence or death
Have 7 prognostic factors
- 5 continuous, 2 categorical

Select variables and functions using 5% significance level
Cox model, n=686, events=299

Randomized trial, hormonal therapy (HT) has an effect on RFS – most models are stratified or adjusted for HT. In the following not mentioned.

Breast cancer: univariate linear analysis

Variable	Name	X^2
X_1	Age	0.58
X_2	Menopausal status	0.28
X_3	Tumour size	15.68
X_{4a}	Grade 2 or 3	19.92
X_{4b}	Grade 3	8.19
X_5	No. of positive lymph nodes	50.02
X_6	Progesterone receptor status	34.04
X_7	Oestrogen receptor status	4.70

Continuous variables – linear or non-linear effect? Univariate FP2 analysis for continuous variables

Variable	Powers	X^2	d.f.	P	Gain
X_1 age	(-2, -0.5)	17.61	4	0.001	17.03
X_3 size	(-1, -3)	19.81	4	0.001	4.13
X_5 nodes	(1, 2)	81.36	4	<0.001	31.34
X_6 PgR	(-0.5, 0)	52.73	4	<0.001	18.69
X_7 ER	(-2, -1)	23.07	4	<0.001	18.37

‘Gain’ assesses non-linearity – corresponds to step 2: chi-square comparing FP2 with linear function, on 3 d.f.

All factors except for X_3 have a stronger non-linear effect

Multivariable FP analysis

Variable	FP etc.	X^2	d.f.	P
X_1 age	(-2, -0.5)	19.33	4	0.001
X_3 size	Out	5.31	4	0.3
X_5 nodes	(-2, -1)	74.14	4	<0.001
X_6 PgR	0.5	32.70	4	<0.001
X_7 ER	Out	2.15	4	0.7
X_2 mens.	Out	0.21	1	0.6
X_{4a} grad 2/3	In	4.59	1	0.03
X_{4b} grad 3	Out	0.15	1	0.7

P is P-to-enter for 'Out' variable,

P-to-remove for 'In' variable

Comments on analysis

Conventional backward elimination at 5% level selects
x4a, x5, x6, and x1 is excluded
FP analysis picks up same variables as backward elimination, and additionally x1
Note considerable non-linearity of x1 and x5
x1 has no linear influence on risk of recurrence

FP model detects more structure in the data than the linear model

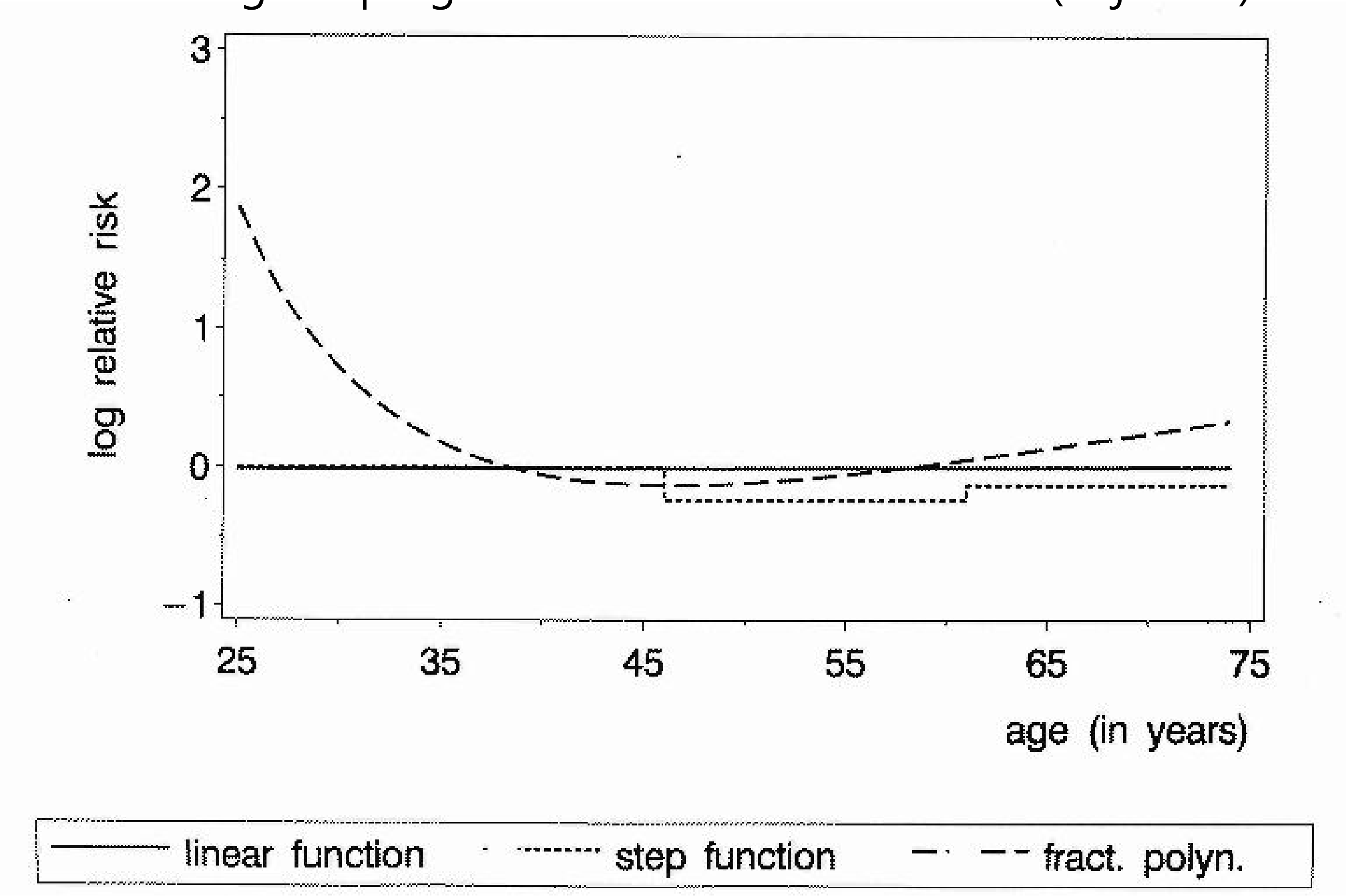
Different analyses- different results

Final models for breast cancer example

	Assumption for continuous variable		
variable	linear	step function	FP function
Grade			
1 vs 2/3	**x**	**x**	**x**
No. of nodes			
linear	**x**		
non- linear			**x**
1-3			
4-9		**x**	
> 10		**x**	
Pg R			
linear	**x**		
non- linear			**x**
≤ vs > 20		**x**	
Age			
non- linear			**x**

Continuous factors
Different analyses may give different results

Age as prognostic factor in breast cancer (adjusted)



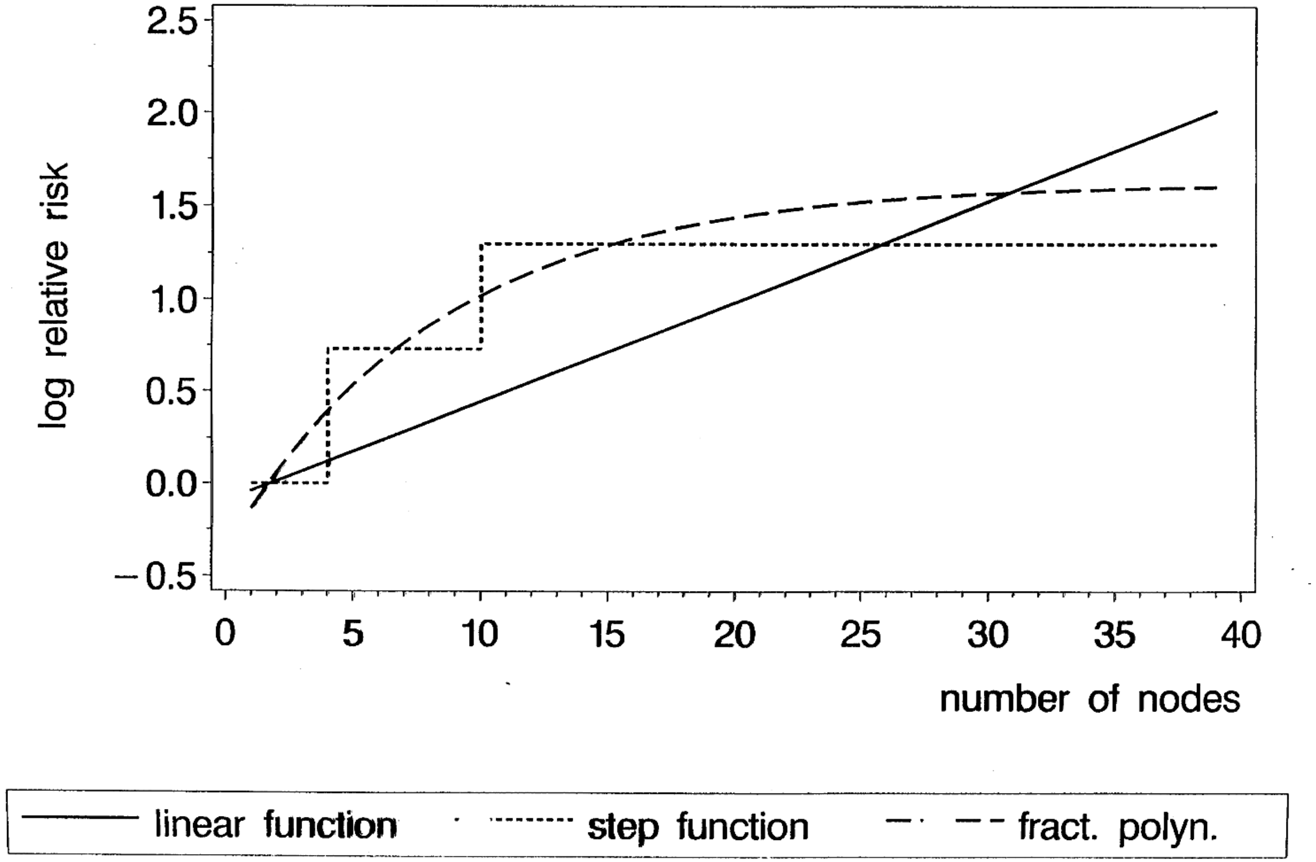
P-value 0.9 0.2 0.001

Sauerbrei + Royston 1999 Fig 1

Results similar or identical?

Nodes as prognostic factor in breast cancer (adjusted)

- medical knowlege: monotonic function (restrict FP class)



P-value	0.001	0.001	0.001

S+R1999 Fig 2

3. Presentation of FP models

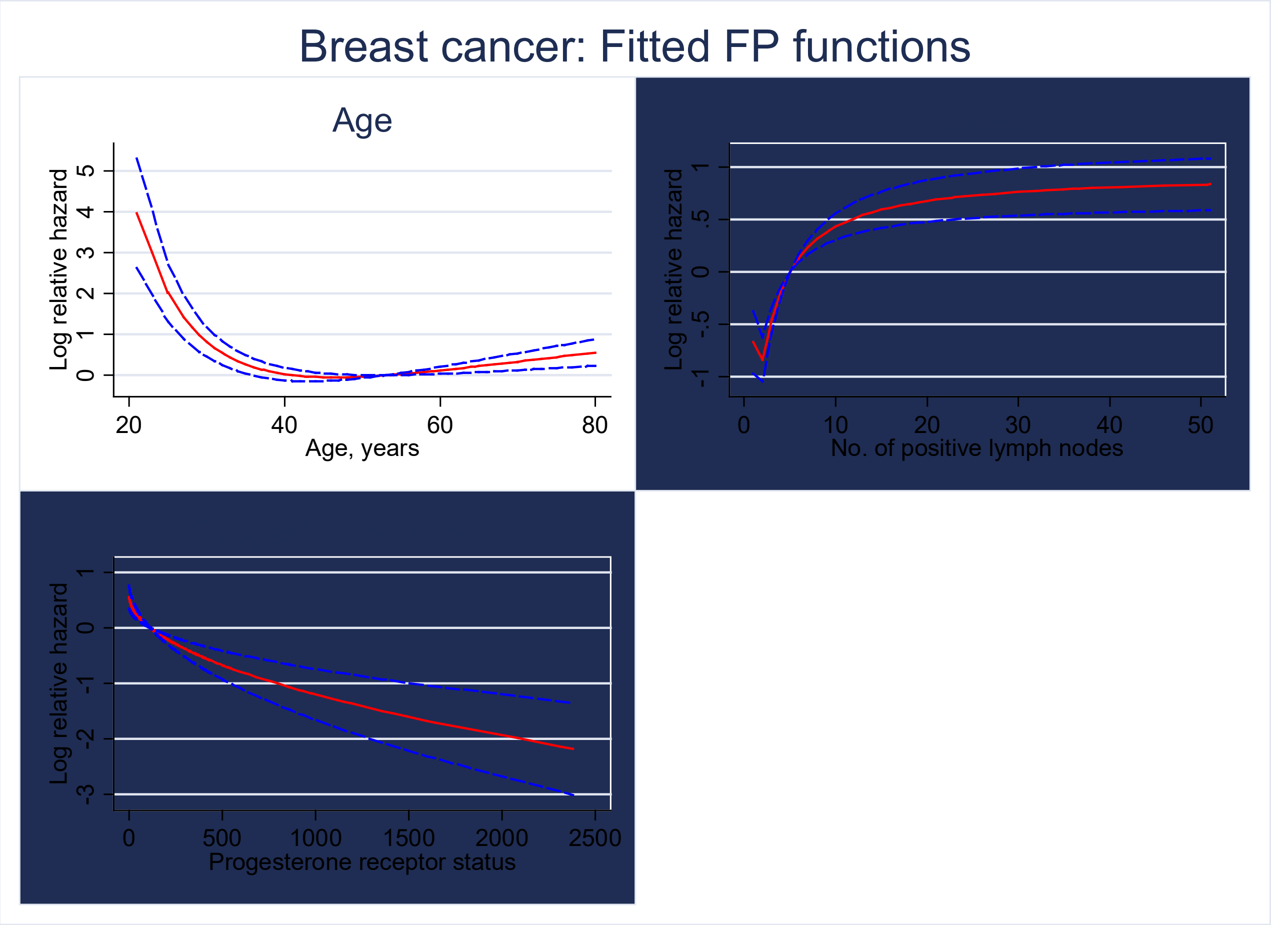
Issues illustrated with further data sets

- Results for two additional data sets

- Whitehall I, Logistic regression
  17260 obs, 1670 events, 10 predicitors
- PBC, Cox model
  418 obs, 161 events, 17 predictors

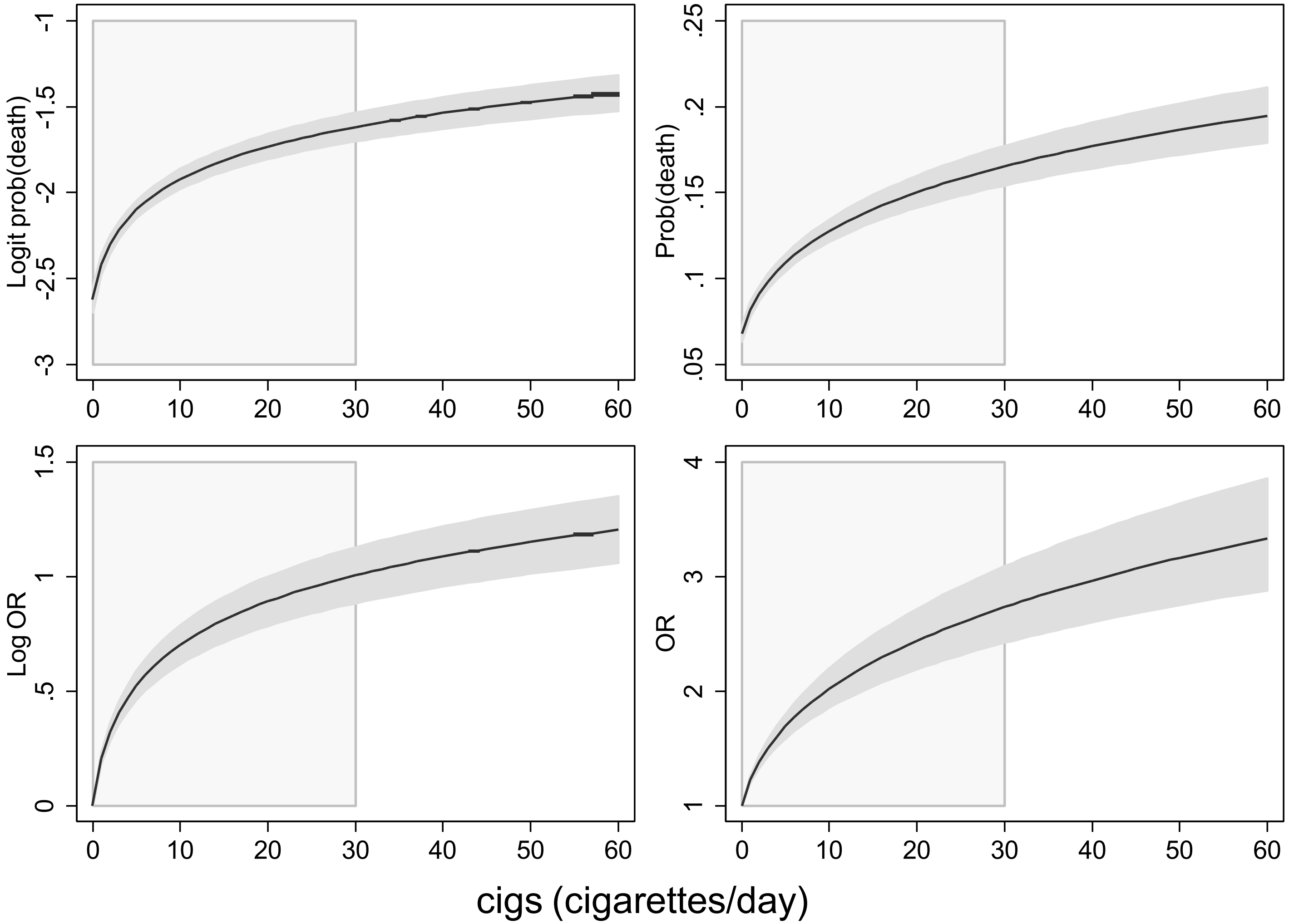
- See the R&S book for details of the data and the analyses

Presentation of FP models: Plots of fitted FP functions



hook for nodes does not agree with medicial knowledge

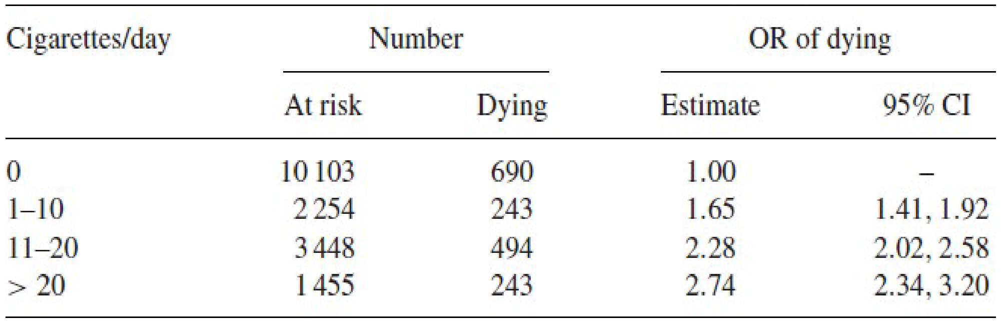
Presentation of FP models: Plots
Whitehall I: Effect of cigarette smoking (power 0)



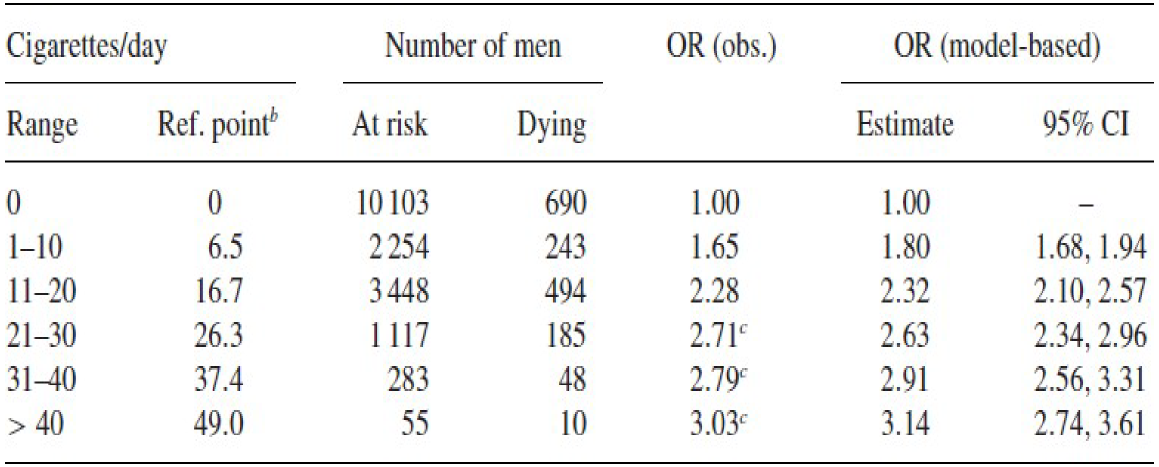
R&S 2008, Fig 4.8

Presentation of FP models: Categories

Whitehall I: OR of dying logOR= -2.62+0.293*log(cigs+1)



Cigarettes/day	Number		OR of dying	
	At risk	Dying	Estimate	95% CI
0	10 103	690	1.00	–
1–10	2 254	243	1.65	1.41, 1.92
11–20	3 448	494	2.28	2.02, 2.58
> 20	1 455	243	2.74	2.34, 3.20



Cigarettes/day		Number of men		OR (obs.)	OR (model-based)	
Range	Ref. point[b]	At risk	Dying		Estimate	95% CI
0	0	10 103	690	1.00	1.00	–
1–10	6.5	2 254	243	1.65	1.80	1.68, 1.94
11–20	16.7	3 448	494	2.28	2.32	2.10, 2.57
21–30	26.3	1 117	185	2.71[c]	2.63	2.34, 2.96
31–40	37.4	283	48	2.79[c]	2.91	2.56, 3.31
> 40	49.0	55	10	3.03[c]	3.14	2.74, 3.61

R&S 2008, Tabs 4.4 and 4.5

4. Tuning parameters for variable and function selection

- **Significance level is the key parameter**

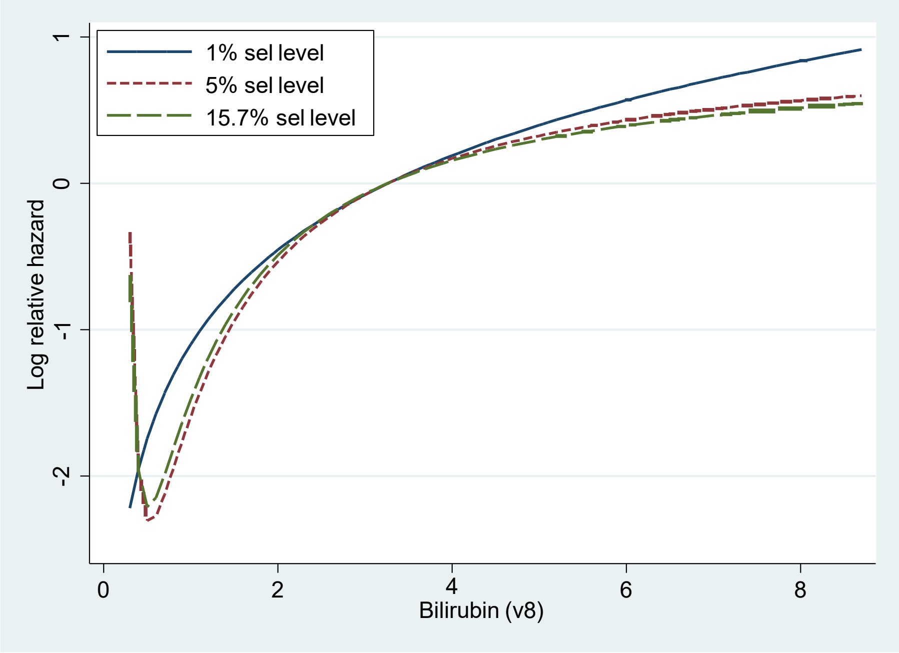
Example: Primary biliary cirrhosis (PBC)

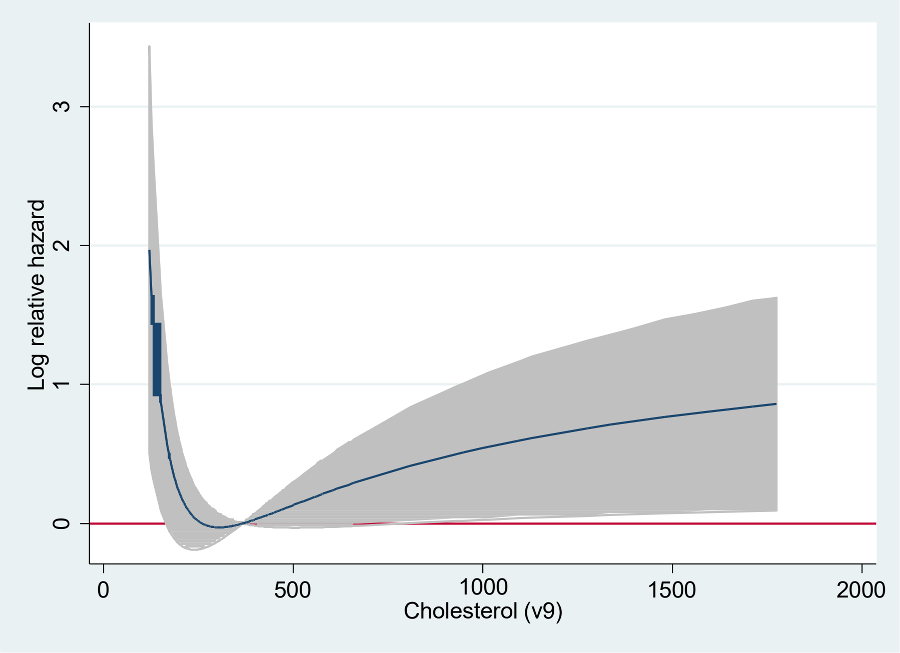
Prognostic factors for overall survival, N=418, Death=161, 17 variables

Significance level

Variable	0.01	0.05	0.157
V2	1	1	(3.3)
V4	-	√	√
V7	√	√	√
V8	0	(-2,-1)	(-2,-1)
V9	-	-	(-2,-0.5)
V10	1	-	-
V11	-	-	1
V17	-	√	√
R^2	0.62	0.65	0.67

PBC data – functions for different significance levels





5. Model criticism

Consistency with subject-matter knowledge
Robustness / influential points
Check of residuals
Complexity of the function

Model criticism 1 –
Consistency with subject- matter knowledge

Breast cancer example showed non-monotonic function for nodes – not medically sensible

Situation can be improved by performing covariate transformation before FP analysis

Sauerbrei & Royston (1999) used

negative exponential transformation of nodes

- exp(–0.12 * number of nodes)

Can be done systematically (Royston & Sauerbrei 2007)

Model criticism 2 – robustness

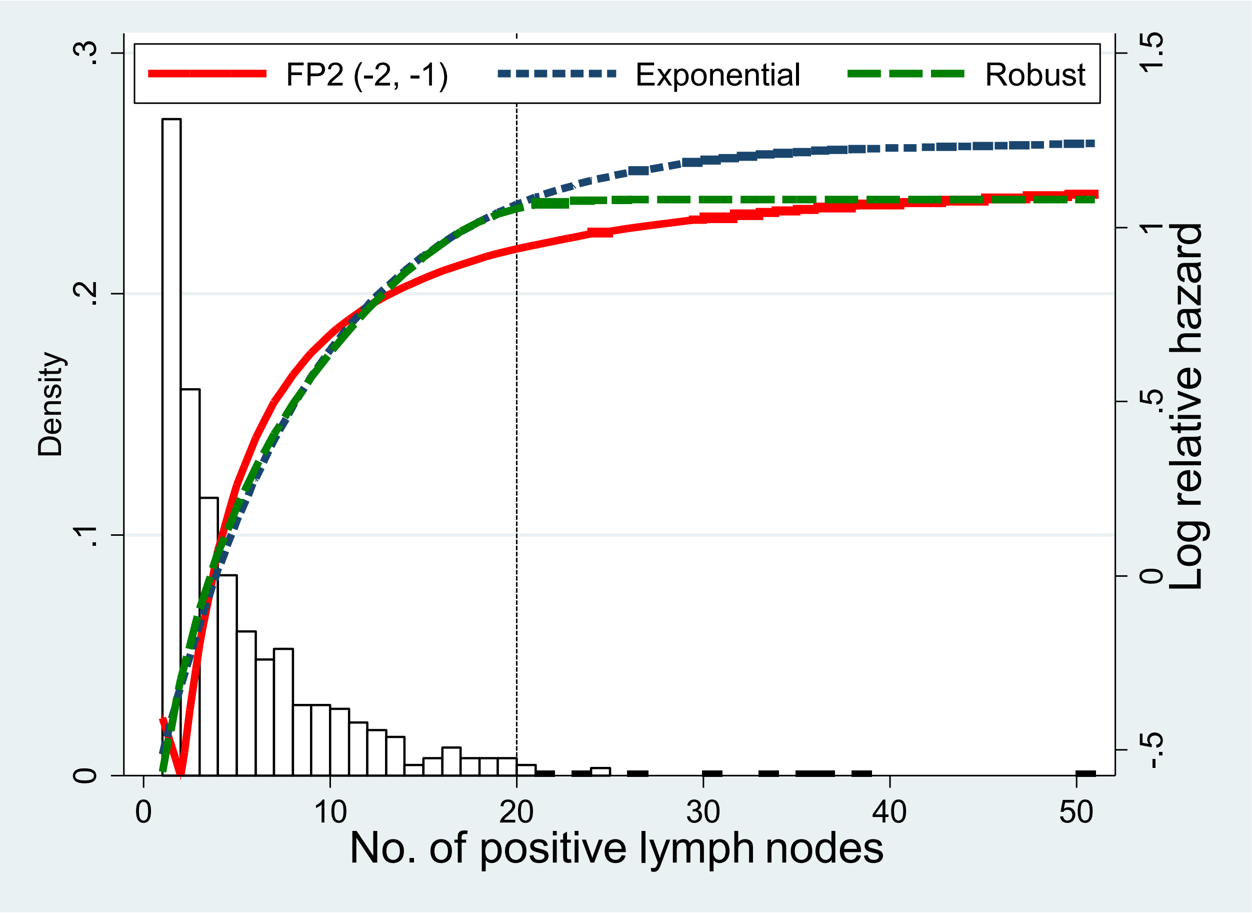
R&S approach to robustification is similar in spirit to double truncation of extreme covariate values

Reduces the leverage of extreme values

- Particularly important after extreme FP transformations – powers -2 or 3

Breast cancer

-making the function for lymph nodes more robust
-include medical knowledge

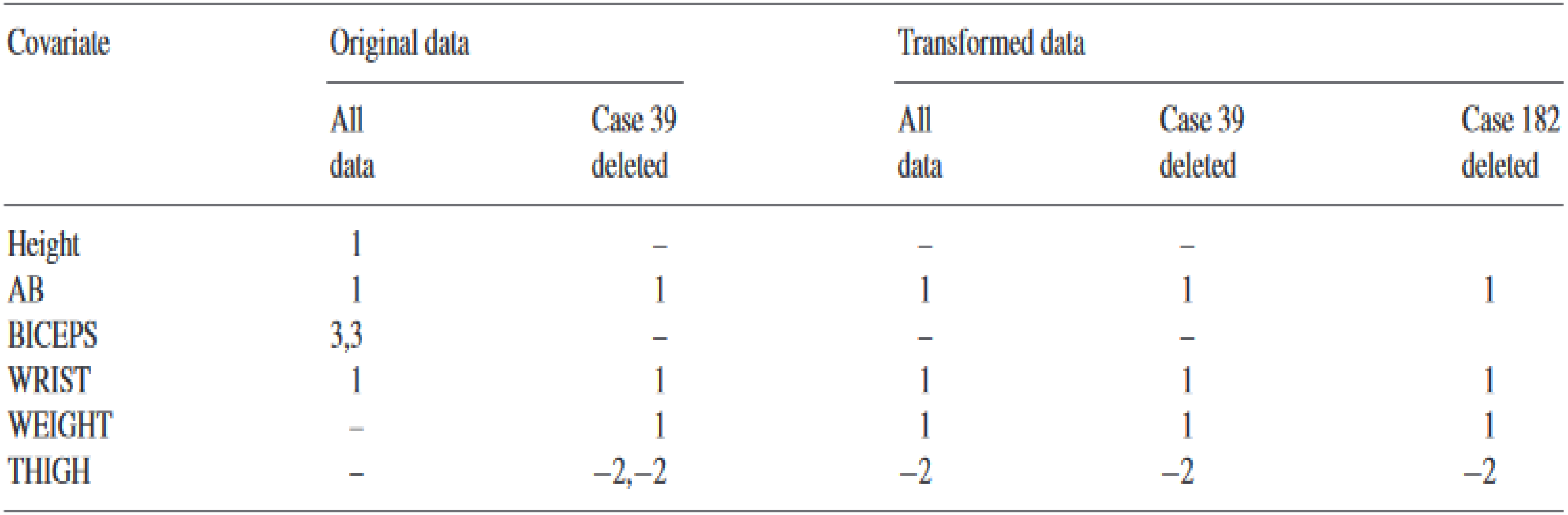


Model criticism 3- influential points

In a multivariable context difficult to detect
Several approaches, but no standard

Example: Body fat data, Outcome continuous, N=252,13 covariates

Result of MFP modelling

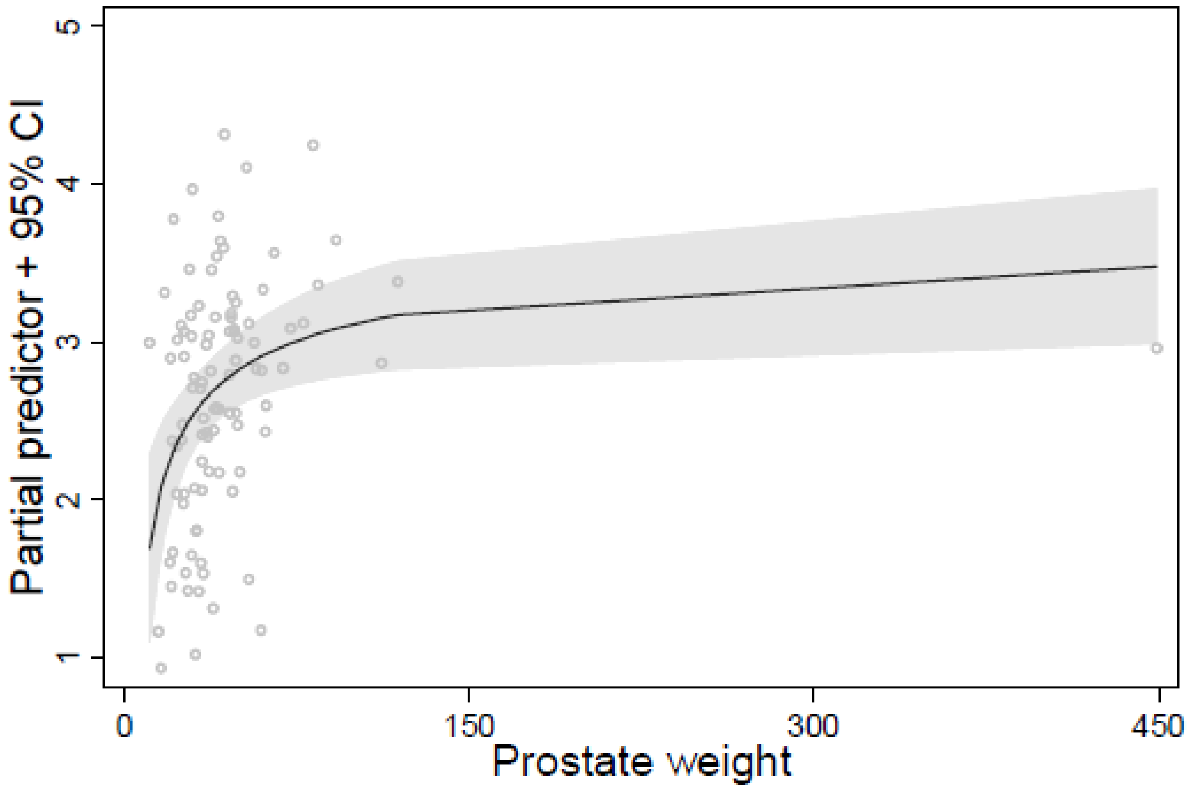


Covariate	Original data		Transformed data		
	All data	Case 39 deleted	All data	Case 39 deleted	Case 182 deleted
Height	1	–	–	–	
AB	1	1	1	1	1
BICEPS	3,3	–	–	–	
WRIST	1	1	1	1	1
WEIGHT	–	1	1	1	1
THIGH	–	–2,–2	–2	–2	–2

R+S 2007 Tab 3

Model criticism 3 - Outliers

- outlier may be the reason for the selection of a non-linear function. Without the outlier weight has a linear effect in the prostate data



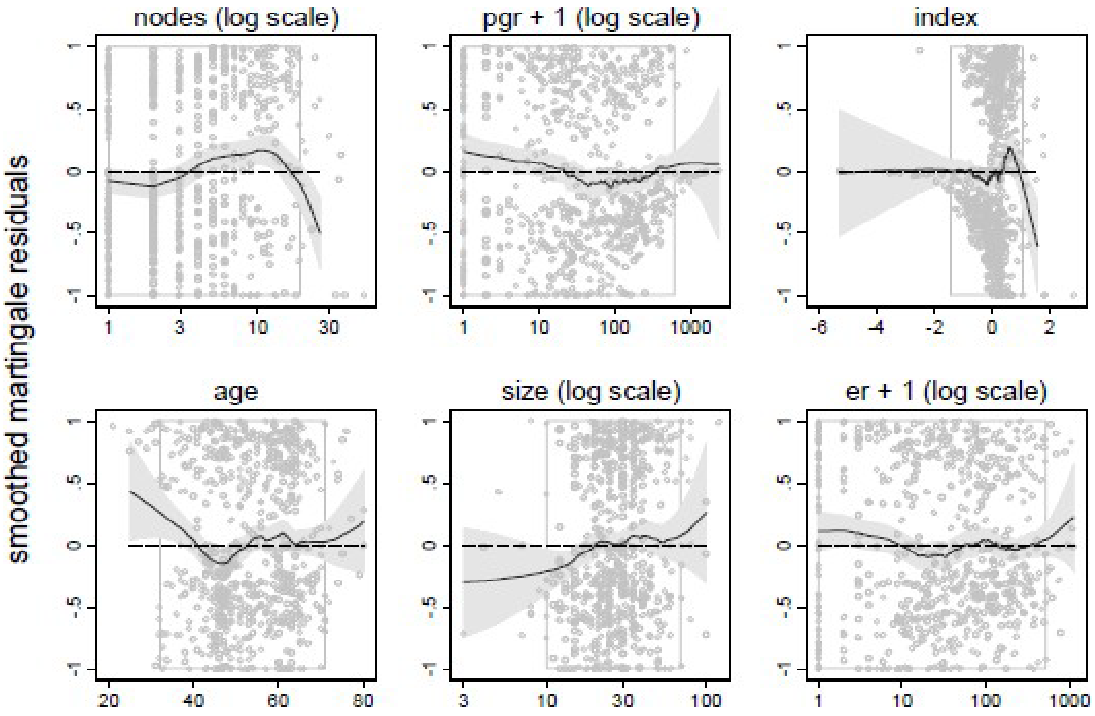
R&S 2008, Fig 6.4

Model criticism 4

Graphical analysis of residuals (all continuous variables)

Breast cancer: BE (0.05) model (assuming all linear)

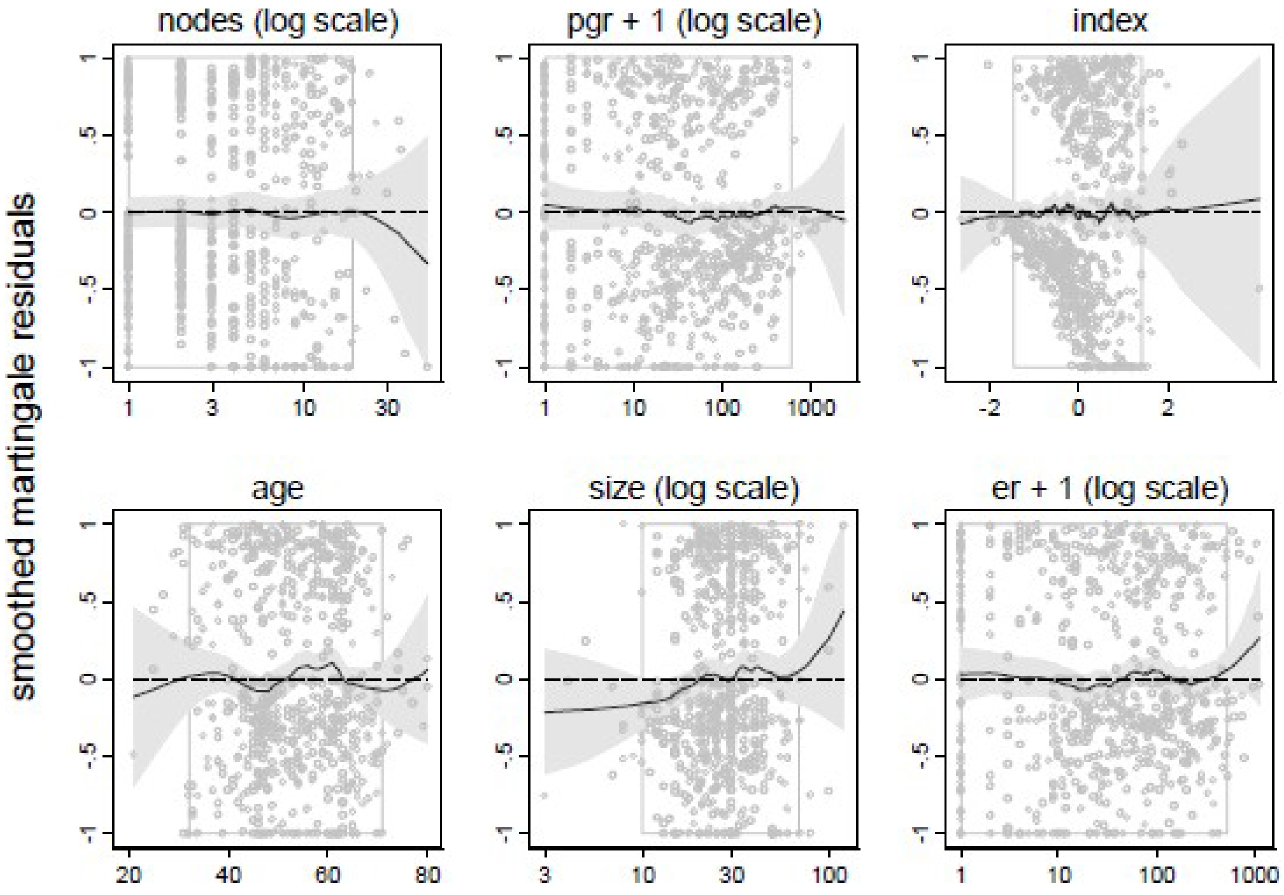
Individual components and index



R&S 2008, Fig 6.6

Indicates severe violations for most variables (+index)

Breast cancer: Residuals from MFP (0.05) model (+ index)



R&S 2008, Fig 6.6

Model criticism 5, …

Do more complex functions fit better?

- relax significance level of FSP part
- extend model class, eg FP3 functions
- add functions not included in FP class
  (check MFP residuals for non-random patterns)

… 6, 7, 8, …

All models are derived under a variety of assumptions. Consider careful whether they are sensible.

Model criticism should receive more attention in practice, eg

bootstrap stability analysis
investigation of interactions
check of the proportional hazards assumption
in the Cox model

Influential points and sample size

Effects of influential points and sample size on the selection and replicability of multivariable fractional polynomial models

Willi Sauerbrei[*†], Edwin Kipruto[†] and James Balmford[^]

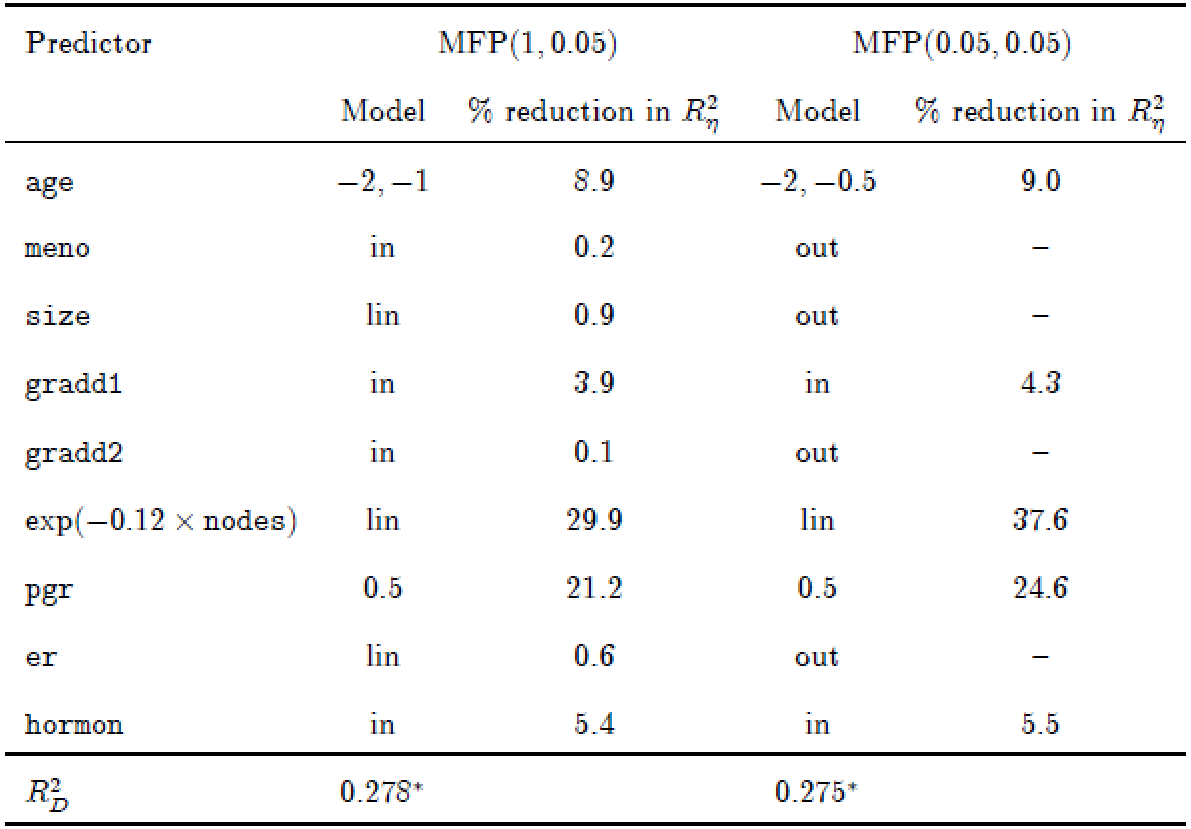
Diagnostic and Prognostic Research (2023) 7:7
doi.org/10.1186/s41512-023-00145-1

MFP+L
check MFP model for overlooked local features

- The MFP approach will pick up global non-linear structure, but might miss local structure
- Model checking is mandatory
- Check for missed local structure: Apply smoothing approach (kernel smoother, splines) to residuals
- MFP+L: automatic procedure that
  - fits local, truncated power basis functions to the residuals of a fit from the MFP approach
  - adds local components to the model if needed
  - is guided by a closed-testing approach, i.e., Type I error for adding of local features is controlled

6. Contributions of individual components

- calculate R2 with and without each variable



Predictor	MFP(1, 0.05)		MFP(0.05, 0.05)	
	Model	% reduction in R_η^2	Model	% reduction in R_η^2
age	$-2, -1$	8.9	$-2, -0.5$	9.0
meno	in	0.2	out	–
size	lin	0.9	out	–
gradd1	in	3.9	in	4.3
gradd2	in	0.1	out	–
$\exp(-0.12 \times \text{nodes})$	lin	29.9	lin	37.6
pgr	0.5	21.2	0.5	24.6
er	lin	0.6	out	–
hormon	in	5.4	in	5.5
R_D^2	0.278*		0.275*	

- Other approaches are available

7. Big Data

- Term ‘Big Data’ is used for many different situations.
  Extremely confusing
- Hand (JRSSA, 2016) distinguishes **two types**
  First type: primarily data manipulation sorting, searching, matching, …

  Examples include online route finders, apps for updated status of bus traffic
  → Mostly addressed by computer scientists & mathematicians

  Second type: uses data to derive models for prediction or understanding of the mechanisms and processes that have generated the collected data
  Achieving these goals will rely primarily on state-of-the-art statistical and machine learning methods

- Aim, design and type of data are key issues
- Data from well designed experiments, systematically collected (eg. registries) or ‘found‘ data?

FPs and Big Data
opportunities and challenges

- **large(r) n**
  - Test-based FP function selection (FSP)
    FSP needs to be adapted, for example replace p-value by improvement of area between curves (Govindarajulu et al., 2007)
  - only monotonic functions – restrict to FP1 class
    non-monotonic functions – best FP2
  - required to investigate for interactions (MFPT, MFPIgen)
  - chances to validate a (MFP) model

- **large p, relatively small n (- omics)**
  - restrict to best FP1 transformation. Much better (Govindarajulu measure?) than linearity?
  - check for influential points (Boulesteix and Sauerbrei, 2011)

Further comments on selection of variables and functional forms in multivariable analysis

- Multivariable model building
  - still many open issues, theoretical knowledge limited
  - software selects a ‚final‘ model, but too often central assumptions are ignored because of various reasons
  - more guidance is urgently required

- Chatfield (2002)

.. literature .. overly concerned with theoretical matters far removed from the day-to-day concern of many working statisticians

With MFP we developed a pragmatic approach

Philosophy of MFP ….

Getting the big picture right is more important than optimising some aspects and ignoring others

strong predictors

strong non-linearity

….. and extensions

strong interactions

strong non-PH in survival model

Towards recommendations for model-building by selection of variables and functional forms for continuous predictors under several assumptions

Issue	Recommendation
Variable selection procedure	Backward elimination; significance level as key tuning parameter, choice depends on the aim of the study
Functional form for continuous covariates	Linear function as the 'default', check improvement in model fit by fractional polynomials. Check derived function for undetected local features
Extreme values or influential points	Check at least univariately for outliers and influential points in continuous variables. A preliminary transformation may improve the model selected. For a proposal see R & S 2007
Sensitivity analysis	Important assumptions should be checked by a sensitivity analysis. Highly context dependent
Check of model stability	The bootstrap is a suitable approach to check for model stability
Complexity of a predictor	A predictor should be 'as parsimonious as possible'

Sauerbrei et al. SiM 2007

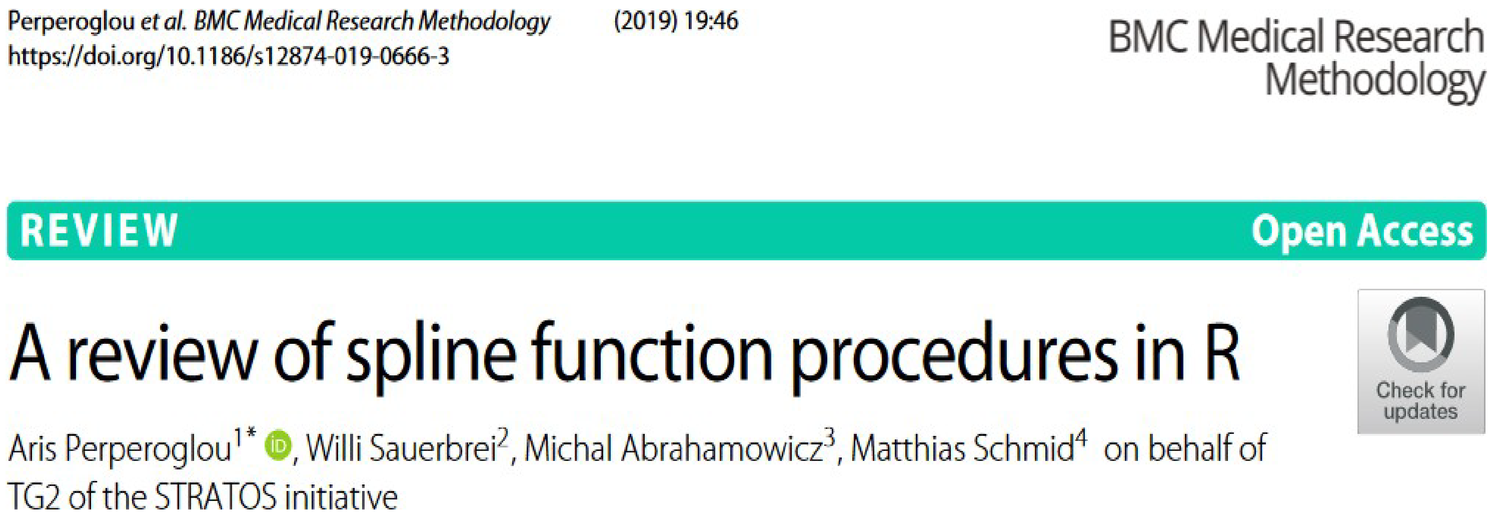
Summary of MFP modelling

- FPs use full information - in contrast to a priori categorisation
- FPs search within flexible class of functions (FP1 and FP(2)-44 models)
- MFP is a well-defined multivariate model-building strategy – combines search for transformations with BE
- Important that model reflects medical knowledge, e.g. monotonic / asymptotic functional forms
- MFP automatically gives a 'reasonable' model
- But, it's essential to check the characteristics of the model
  - May result in model refinement
- Identification of influential observations in a multivariable context is feasible
- Should also look for important interactions between predictors – not covered here

How to work with MFP - Chapter 10 of R&S book

Splines are a more flexible alternative

Many different approaches
Comparisons of splines approaches are difficult and rare



Perperoglou *et al. BMC Medical Research Methodology* (2019) 19:46
https://doi.org/10.1186/s12874-019-0666-3

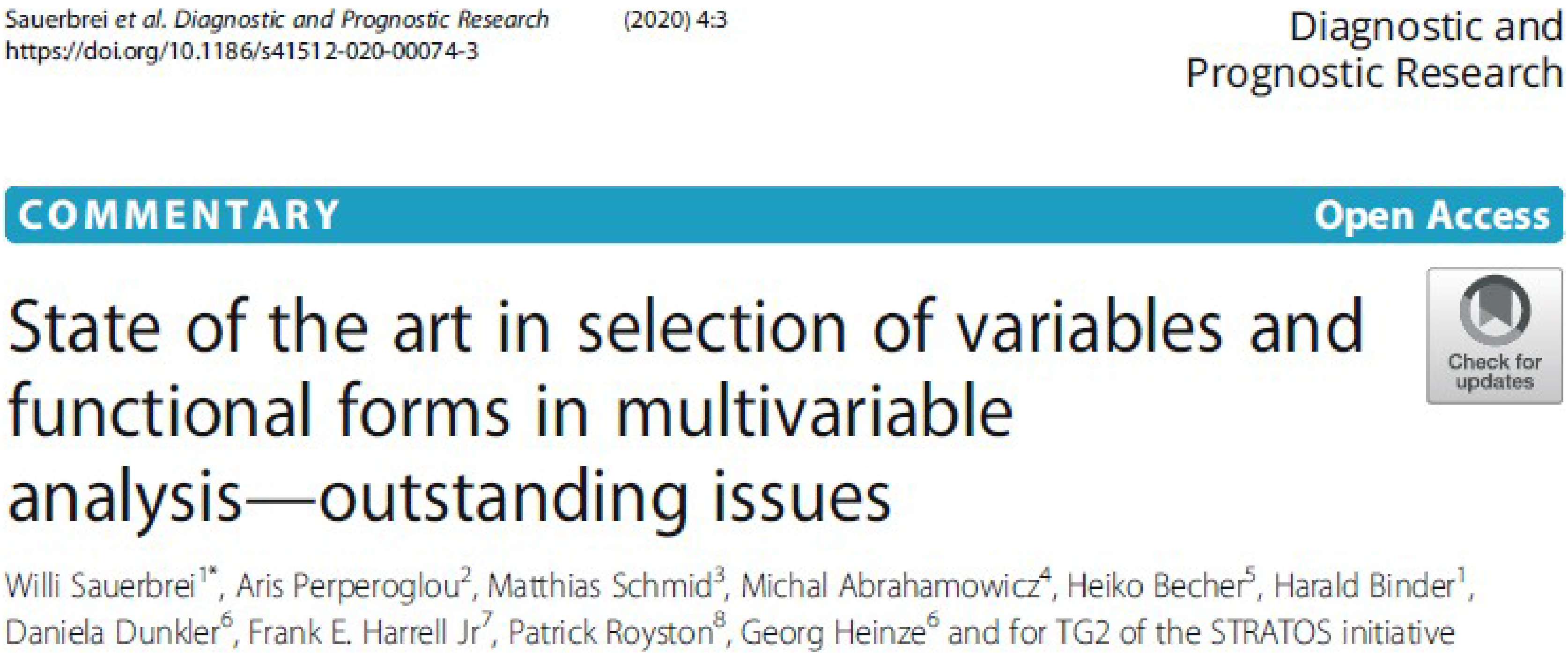
BMC Medical Research Methodology

REVIEW — Open Access

A review of spline function procedures in R

Aris Perperoglou[1]*, Willi Sauerbrei[2], Michal Abrahamowicz[3], Matthias Schmid[4] on behalf of TG2 of the STRATOS initiative

Check for updates



Sauerbrei *et al. Diagnostic and Prognostic Research* (2020) 4:3
https://doi.org/10.1186/s41512-020-00074-3

Diagnostic and
Prognostic Research

COMMENTARY Open Access

State of the art in selection of variables and functional forms in multivariable analysis—outstanding issues

Willi Sauerbrei[1*], Aris Perperoglou[2], Matthias Schmid[3], Michal Abrahamowicz[4], Heiko Becher[5], Harald Binder[1], Daniela Dunkler[6], Frank E. Harrell Jr[7], Patrick Royston[8], Georg Heinze[6] and for TG2 of the STRATOS initiative

State of the art for selection of variables and functional form in multivariable analysis – research required!

– Which strategies for variable selection exist?	What about their properties?
– Data-dependent modelling introduces bias.	What about the role of shrinkage approaches?
– Comparison of spline procedures in a univariate context.	Which criteria are relevant? Can we derive guidance for practice?
– What about variables with a ‘spike-at-zero’?	
– Multivariable procedures	MFP well defined strategy Which of the spline based procedures? Comparison in large simulation studies needed
– Multivariable procedures and correction for selection bias	How relevant? One step or two step approaches? E.g. selection of variables and forms followed by shrinkage
– Big Data	Does it influence properties of procedures and their comparison?
– Role of model validation	

Conclusion

We are far away from ‘state-of-the-art’ on selection of variables and functional forms.

Much research urgently needed!

References

- Binder H., Sauerbrei W. (2010): Adding local components to global functions for continuous covariates in multivariable regression modeling. Statistics in Medicine, 29: 800-817.
- Boulesteix A.-L., Guillemot V., Sauerbrei W. (2011): Use of pretransformation to cope with extreme values in important candidate features. Biometrical Journal, 53(4): 673–688. DOI: 10.1002/bimj.201000189.
- Chatfield, C. (2002): Confessions of a pragmatic statistician, The Statistican 51: 1-20.
- Hand, D.J. (2016): Editorial: 'Big data' and data sharing. Journal of the Royal Statistical; Society, Series A 179, 3: 629–63
- Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M on behalf of TG2 of the STRATOS initiative (2019): A review of spline function procedures in R, BMC Medical Research Methodology, 19:46 doi: 10.1186/s12874-019-0666-3
- Royston, P., Ambler, G., Sauerbrei, W. (1999): 'The use of fractional polynomials to model continuous risk variables in epidemiology' International Journal of Epidemiology, 28:964-974.
- Royston, P., Sauerbrei, W. (2007): Improving the robustness of fractional polynomial models by preliminary covariate transformation: a pragmatic approach. Computational Statistics and Data Analysis, 51: 4240-4253.
- 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, Perperoglou A, Schmid M, Abrahamowicz M, Becher H, Binder H, Dunkler D, Harrell Jr. FE, Royston P, Heinze G for TG2 of the STRATOS initiative (2020). State of the art in selection of variables and functional forms in multivariable analysis - outstanding issues. Diagnostic and Prognostic Research, 4:3, 1-18. DOI: 10.1186/s41512-020-00074-3

References

- Sauerbrei, W. and Royston, P.(2017): The Multivariable Fractional Polynomial Approach, with Thoughts about Opportunities and Challenges in Big Data. In: Hans-Joachim Mucha (Ed.)Big data clustering: Data preprocessing, variable selection, and dimension reduction. WIAS Report 29, Berlin: 36-54. DOI: 10.20347/WIAS.REPORT.29
- Sauerbrei, W. and Royston, P.2016. Multivariable Fractional Polynomial Models. Wiley StatsRef: Statistics Reference Online. 1–8. DOI: 10.1002/9781118445112.stat07861
- Sauerbrei, W., Kipruto, E., Balmford, J. (2023). Effects of Influential Points and Sample Size on the Selection and Replicability of Multivariable Fractional Polynomial Models. Diagnostic and Prognostic Research, 7(1):7.
- Sauerbrei, W., Meier-Hirmer, C., Benner, A., Royston, P.(2006): Multivariable regression model building by using fractional polynomials: description of SAS, STATA and R programs. Computational Statistics and Data Analysis, 50: 3464-3485.
- 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; Corrigendum (2002), 165: 339-400.
- 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., Bojar, H., Schmoor, C., Schumacher, M. and the German Breast Cancer Study Group (GBSG) (1999): 'Modelling the effects of standard prognostic factors in node positive breast cancer', British Journal of Cancer, 79: 1752- 1760.
- Sauerbrei, W., Royston, P., & Kipruto, E. (2025). Multivariable Fractional Polynomial Models and Extensions. In International Encyclopedia of Statistical Science (pp. 1609-1616). Berlin, Heidelberg: Springer Berlin Heidelberg.