Continuous Predictors in Regression Analyses

Goal

- To discuss ways to incorporate continuous covariates into regression models, with a focus on restricted cubic splines and fractional polynomials.

Outline

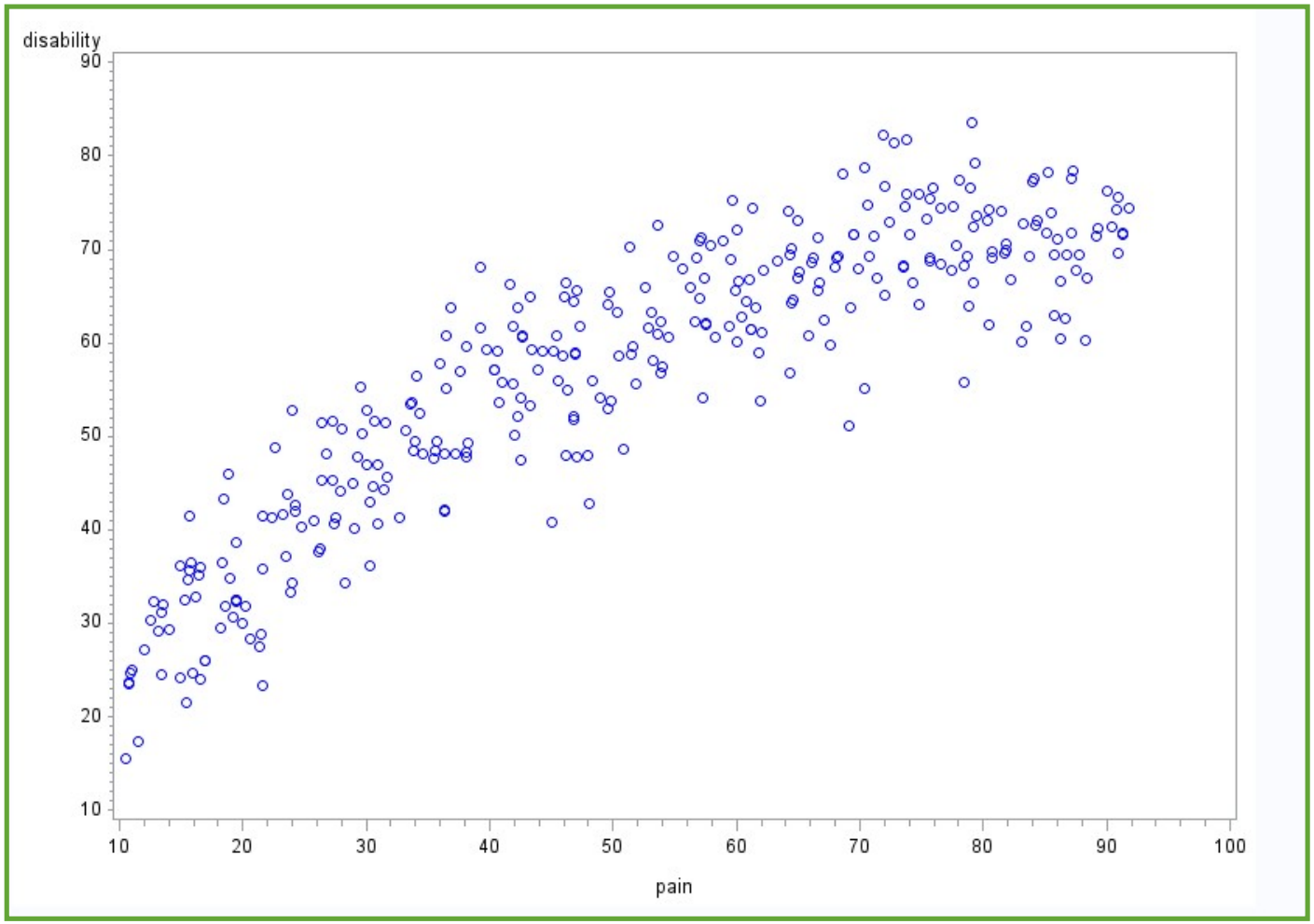
- An overview of ways to include continuous predictors in regression models
- Restricted Cubic Splines
- Fractional Polynomials

No real data sets were harmed in the production of this presentation.

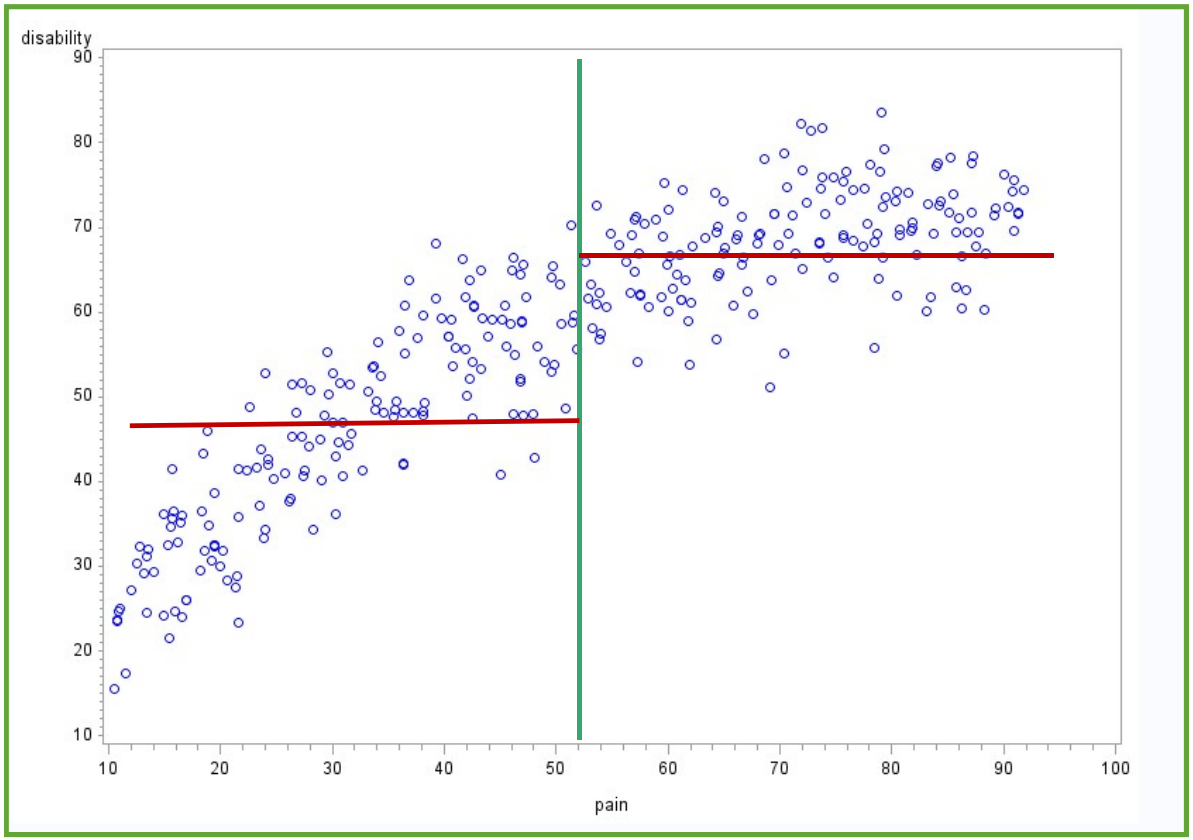
Unless otherwise indicated, all data are simulated.

Overview

Procedure	Characteristics	Recommendation
Dichotomization	Simple, easy interpretation	Bad idea
More categories	Categories capture prognostic information better, but are not smooth; sensitive to choice of cut-points and hence instable	Primarily for illustration, comparison with published evidence
Linear	Simple	Often reasonable as a start
Transformations	Log, square root, inverse, exponent, etc.	May provide robust summaries of non-linearity
Restricted cubic splines	Flexible functions with robust behaviour at the tails of predictor distributions	Flexible descriptions of non-linearity
Fractional polynomials	Flexible combinations of polynomials; behaviour in the tails may be unstable	Flexible descriptions of non-linearity
Adapted from Steyerberg (2009), Clinical Prediction Models		


disability
90
80
70
60
50
40
30
20
10
10
20
30
40
50
60
70
80
90
100
pain

Dichotomization



high_pain = (pain score > 51)

Model: disability = $\beta_0 + \beta_1$high_pain

Dichotomization

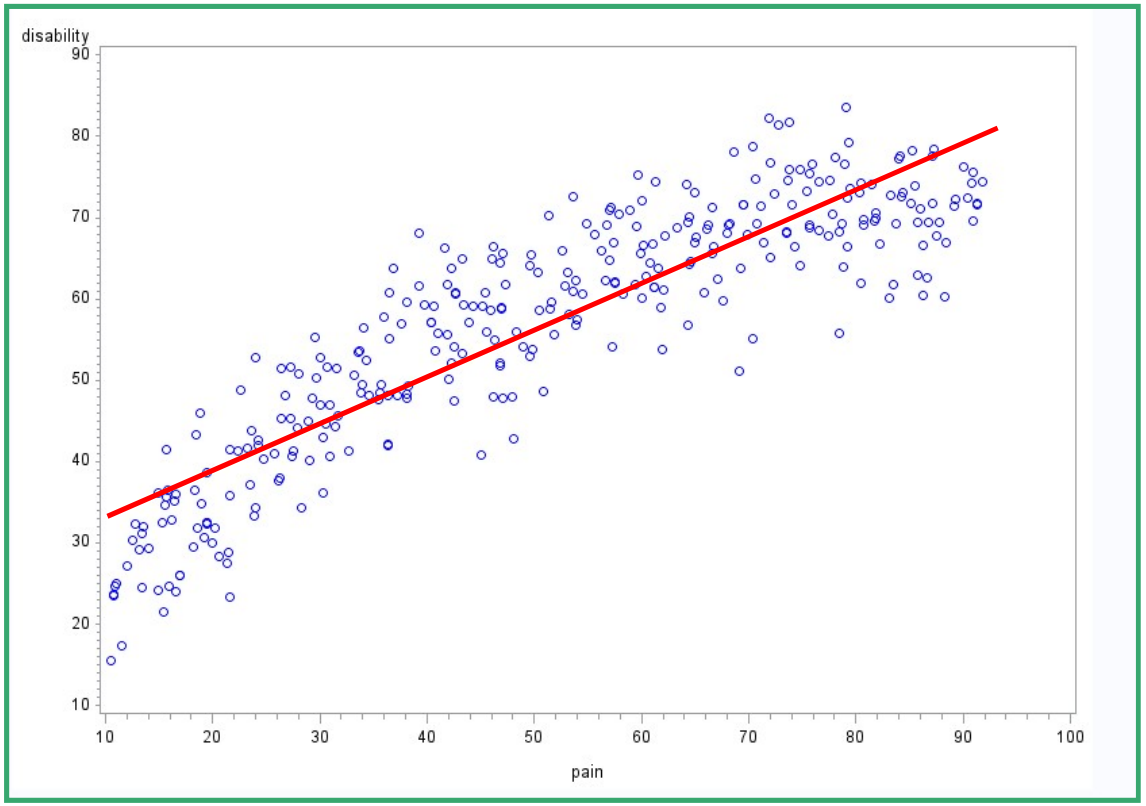
- Makes the analysis and interpretation of results simple
- Clinical decision-making often requires two classes.
- Loss of power (equivalent to losing a third to half the data)
- Conversely, it may increase the probability of false positives.
- Unrealistic model
- Impossible to detect non-linearity
- Choice of cut point is extremely problematic
- When > 1variable is dichotomized, results likely to be misleading
- Leads to residual confounding
- Etc.

Dichotomization

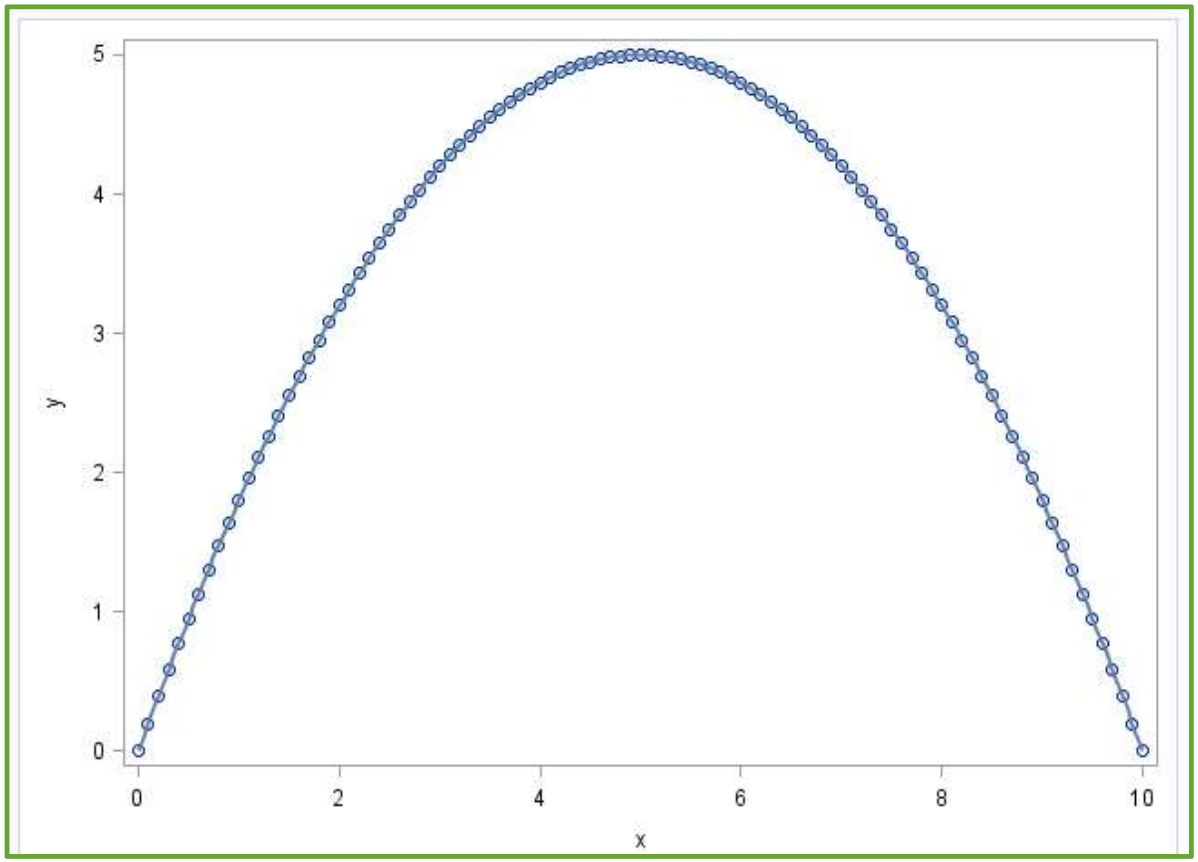
Royston et al. Dichotomizing continuous predictors in multiple regression: a bad idea. Statist Med. 2006; 25: 127-141.

Jiming Fang et al. Test for linearity between continuous confounder and binary outcome first, run a multivariate regression analysis second. http://support.sas.com/resources/papers/proceedings09/252-2009.pdf

Linear fit


disability
90
80
70
60
50
40
30
20
10
10
20
30
40
50
60
70
80
90
100
pain

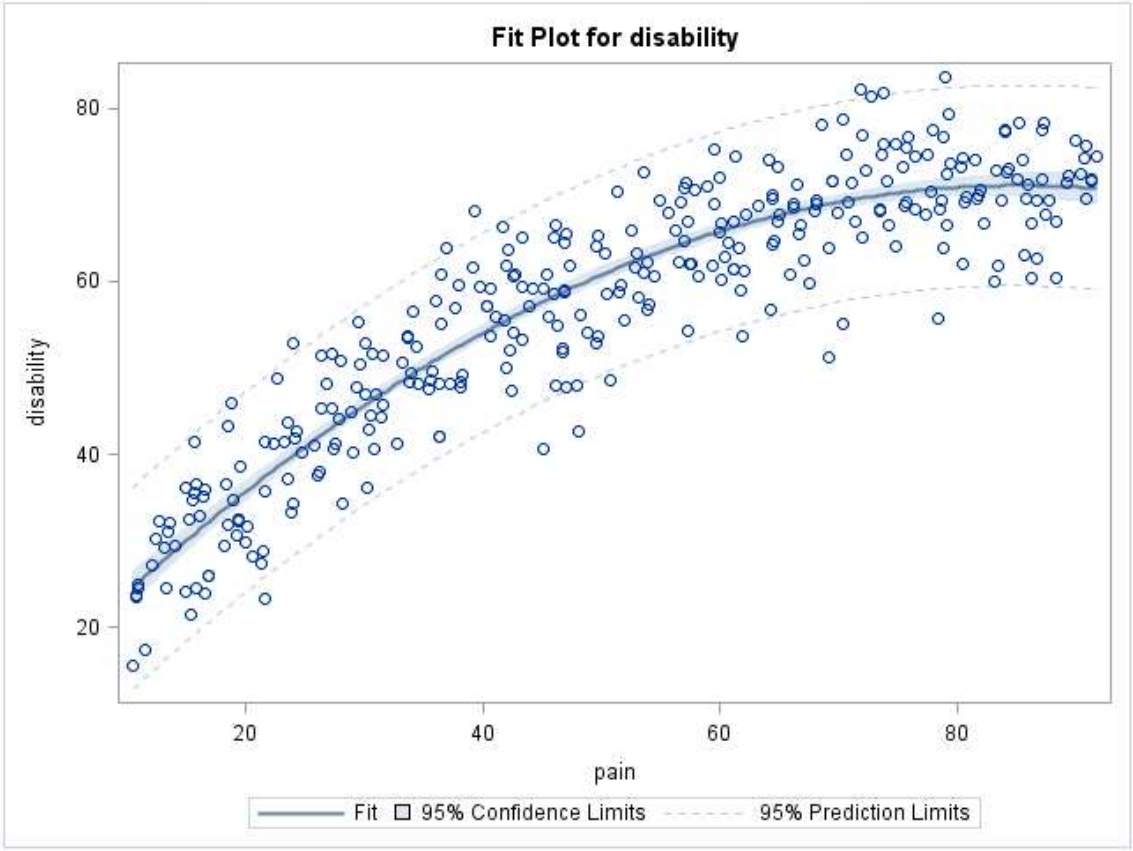
Polynomials



Polynomials do not have asymtotes.

They often fit badly at the extremes of the data.

Polynomials



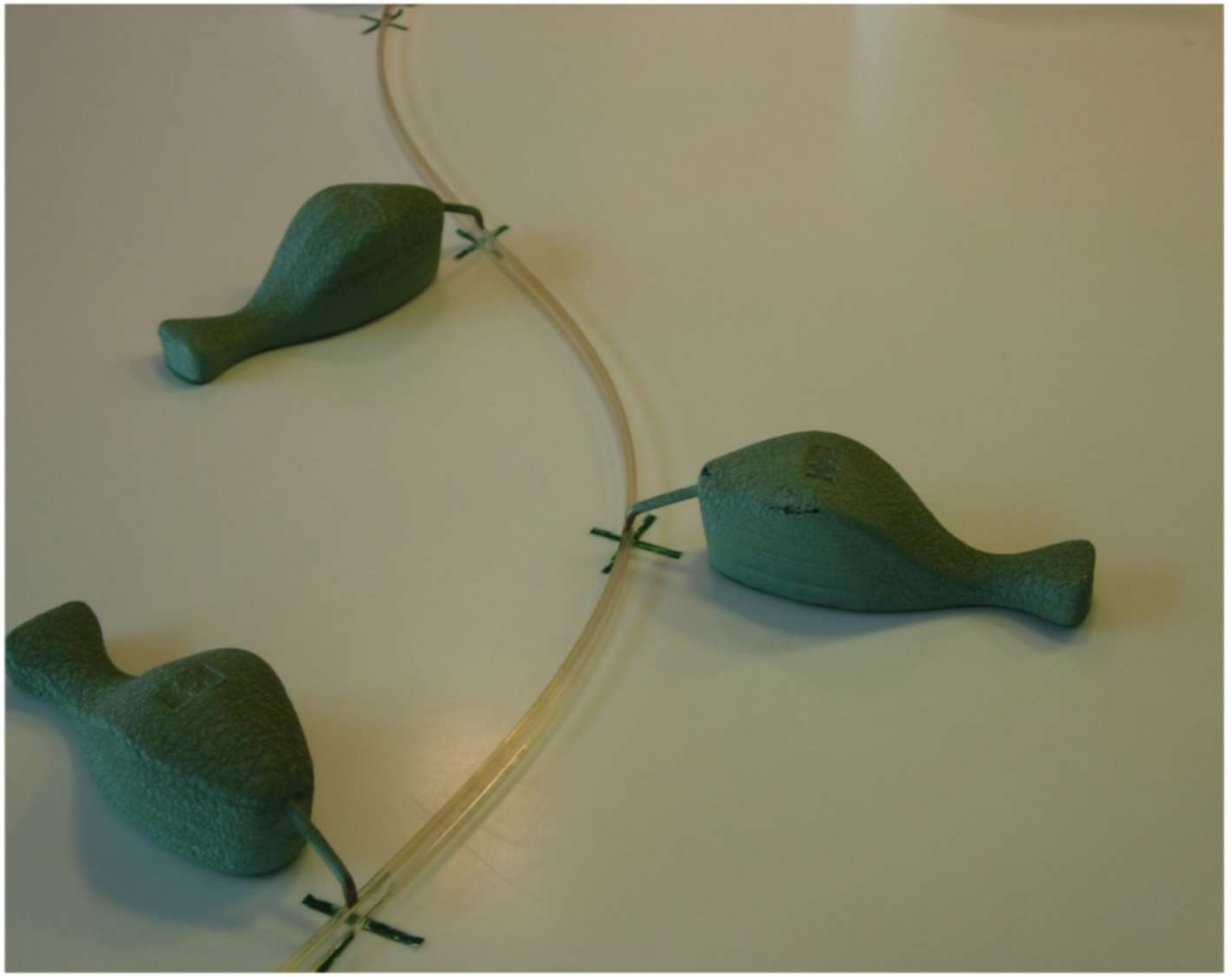
Log-Transformation

- P-value, predicted values not affected by choice of base (ln or $\log_{10}$)
- The model becomes: Disability = $\beta_0 + \beta_1 \log_{10}$ (pain score)
  - disability(pain2) – disability (pain1) = β_1 (log(pain2) – log(pain1))
  - = β_1 log(pain2 / pain1)

- This means that the change in disability depends on the relative increase in pain, not on the absolute increase in pain
- Moving from a pain score of 10 to a pain score of 20 will have the same impact as moving from a pain score of 20 to a pain score of 40.
- Parameter predicts change in disability for a 10-fold increase in pain; multiply by 0.301 to get change for a 2-fold increase.

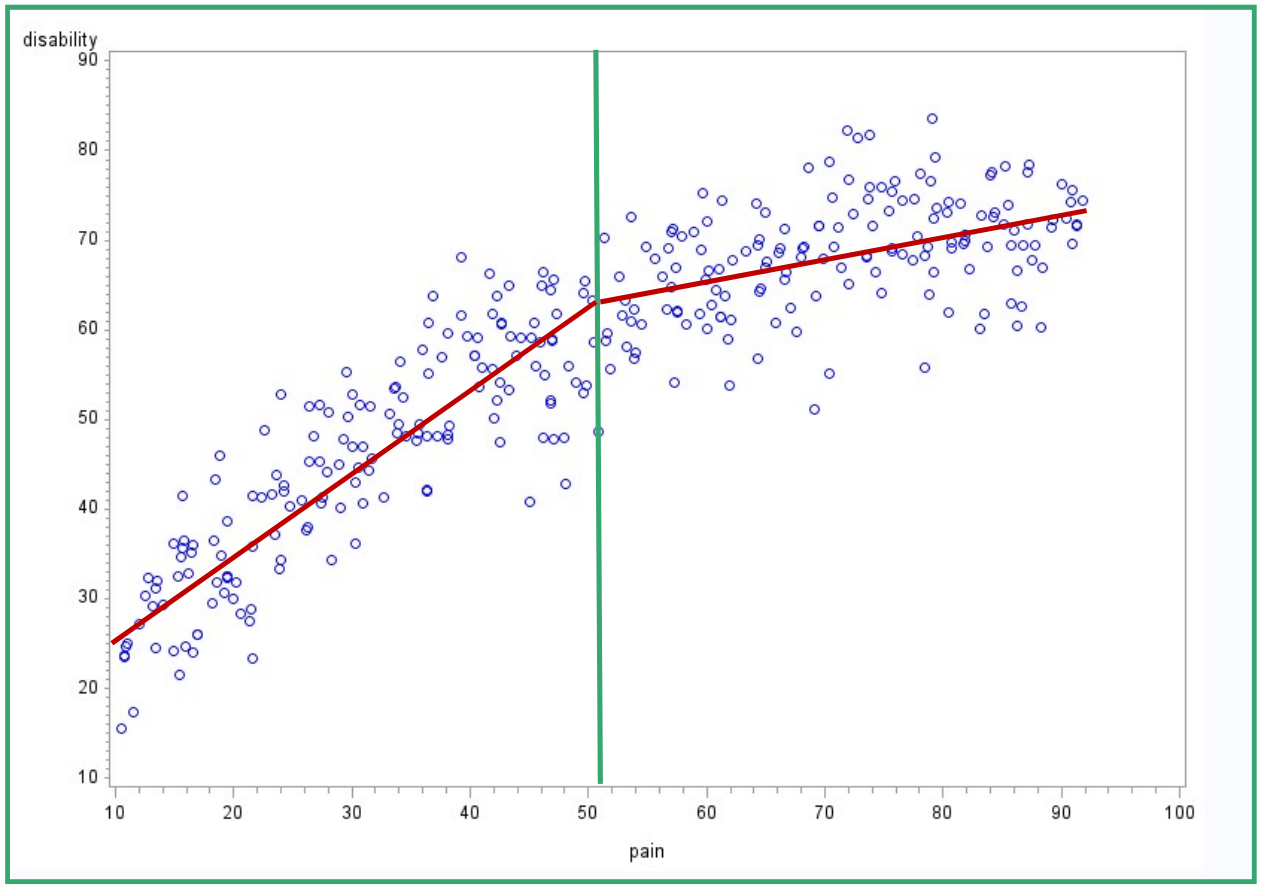
Restricted Cubic Splines

A Spline



A draftsman’s spline, with ducks. Used with permission. See pages.cs.wisc.edu/~deboor/draftspline.html

Linear Splines (aka hockey stick, broken stick, piecewise regression)



Linear Splines for a continuous predictor, X

- Knot (change point)
- For a single knot, requires 1 new variable

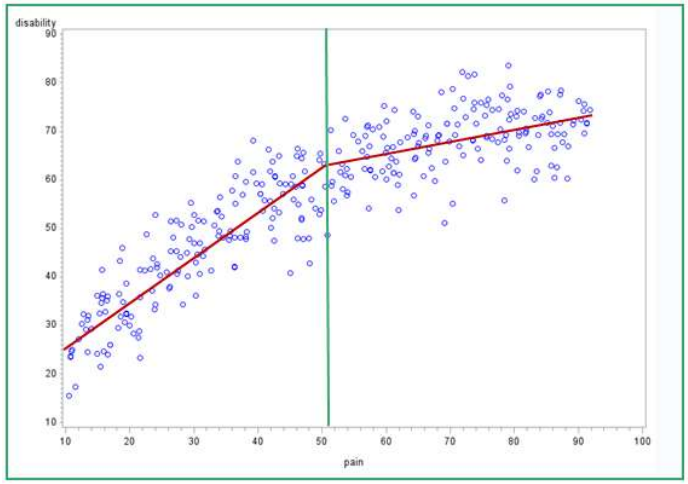
  $V1 = X$

  $V2 = (X - \text{knot})_+$

$$\text{Where } (X - C)_+ = \begin{cases} 0 & \text{If } X < C \\ (X - C) & \text{If } X \geq C \end{cases}$$

$$Y = \beta_0 + \beta_1 V1 + \beta_2 V2 + \varepsilon$$

This produces two line segments which join at the change point C (aka the knot)



Linear Splines for a continuous predictor, X

- Knot (change point)
- You may be more familiar with a different way of expressing the same thing:

  I is a dummy variable which is equal to 0 if $X < C$ and 1 if $X \geq C$

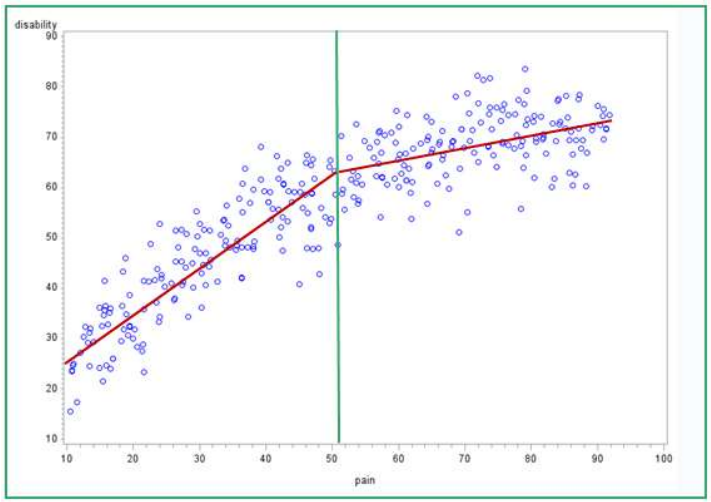
  $V1 = X$

  $V2 = I\,(X - C)$

$$Y = \beta_0 + \beta_1 V1 + \beta_2\, V2 + \varepsilon$$

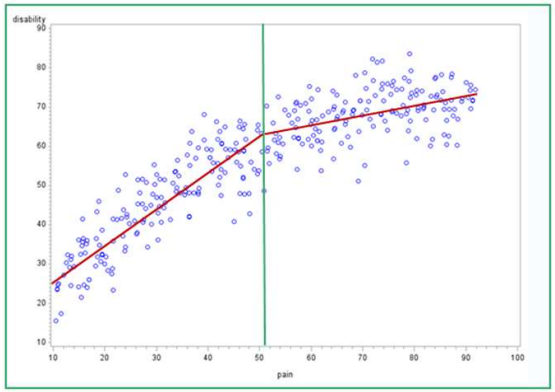
or

$$Y = \beta_0 + \beta_1 X + \beta_2\, (X-C)I + \varepsilon$$



A Few Points About Linear Splines

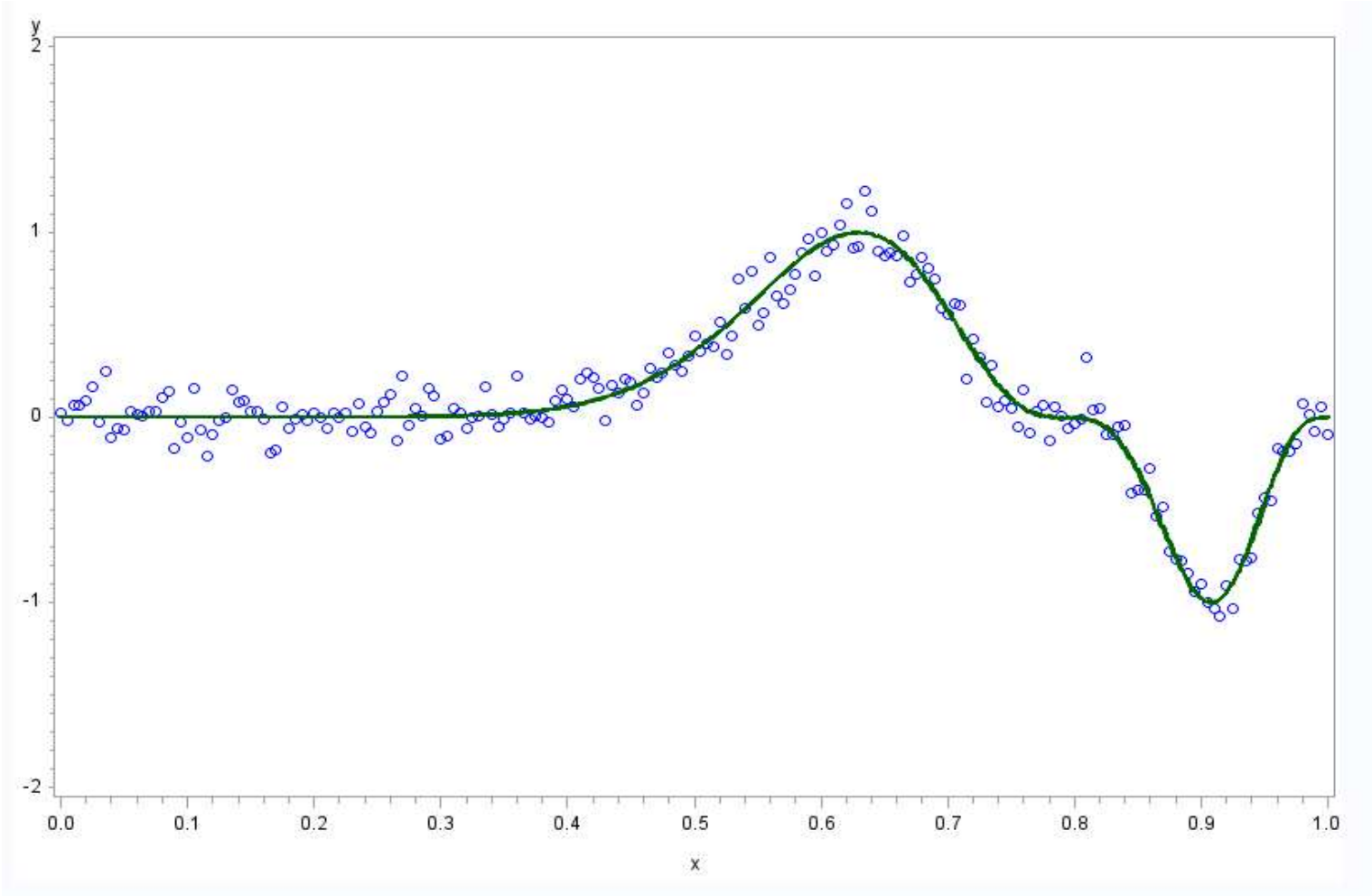
- The continuous predictor is broken into segments, with end points specified by knots.
- New variables were needed, capturing the information in X for just a particular segment.
- Separate regression lines were fit to each segment
- The individual lines were free to fit their portion of the data, with the constraint that they have to meet at the knot.



A Regression Spline

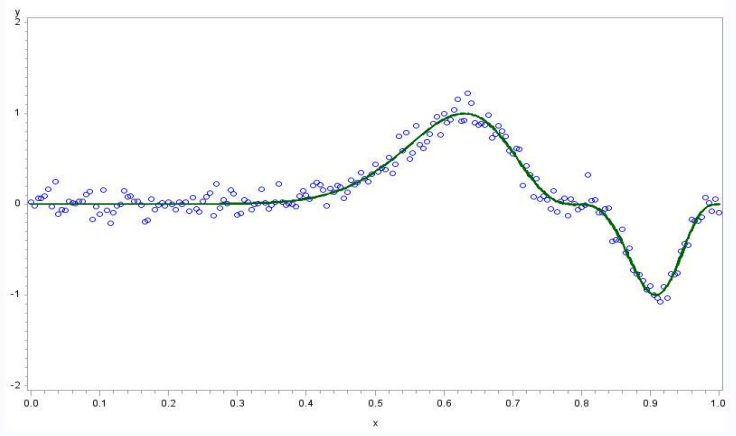
- A flexible way of modeling the continuous variable which
  - Allows curvature
  - Allows different curves for each segment of X
  - Forces the curves to join “smoothly”

My Data



Approach

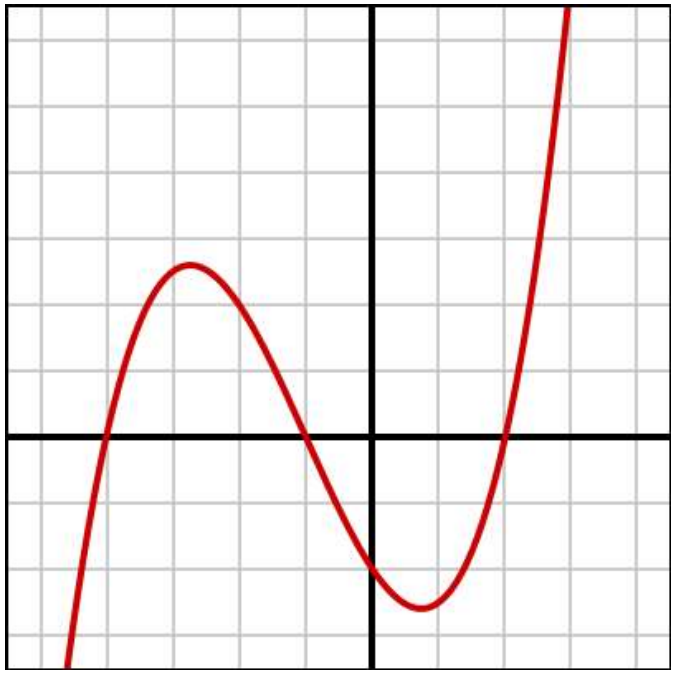
- Analogous to the photograph of the spline
- Range of X-values will be subdivided using a set of knots.
- Separate regression curves will be fit between the knots.
- Decisions
  - What degree of polynomial for the curves
  - Where to place the knots.
- We require that the pieces join “smoothly”.
  What does this mea?



What Degree of Polynomial for the Curves?

- Using *k* knots and a polynomial of degree *n* requires *k + n + 1* regression parameters.

- Cubic splines require *k + 3* parameters compared to 1 for a linear fit.
- Smallest degree of polynomial allowing for inflection.



Location and Number of Knots

- Stone (1986): 5 knots are sufficient
- Harrell (2001): 3 knots if N < 30, 5 knots if N ≥ 100, 4 otherwise
- Location of the knots

Number of knots, K	Knot locations expressed in quantiles of the x variable						
3	0.1	0.5	0.9				
4	0.05	0.35	0.65	0.95			
5	0.05	0.275	0.5	0.725	0.96		
6	0.05	0.23	0.41	0.59	0.77	0.95	
7	0.025	0.1833	0.3417	0.5	0.6583	0.8167	0.975

Restricted Splines

- Linear fit below the first knot and above the last knot
- Uses k - 1 degrees of freedom
  - the original linear predictor X
  - k – 2 piecewise cubic variables

New Variables: transformation of X

- For a restricted cubic spline with k knots, (k-1) variables are required
- One is the original X, so (k-2) new variables
- The knots are positioned at t_1, t_2, ..., t_k
- A "smooth" fit means that both the first and second derivatives (the slope and the rate of change in the slope) are continuous at the knots.

$$x_i = (x - t_i)_+^3 - (x - t_{k-1})_+^3 \frac{t_k - t_i}{t_k - t_{k-1}} + (x - t_k)_+^3 \frac{t_{k-1} - t_i}{t_k - t_{k-1}}$$

$$for\ i = 1, \ldots, k-2$$

New Variables: There's a SAS macro.

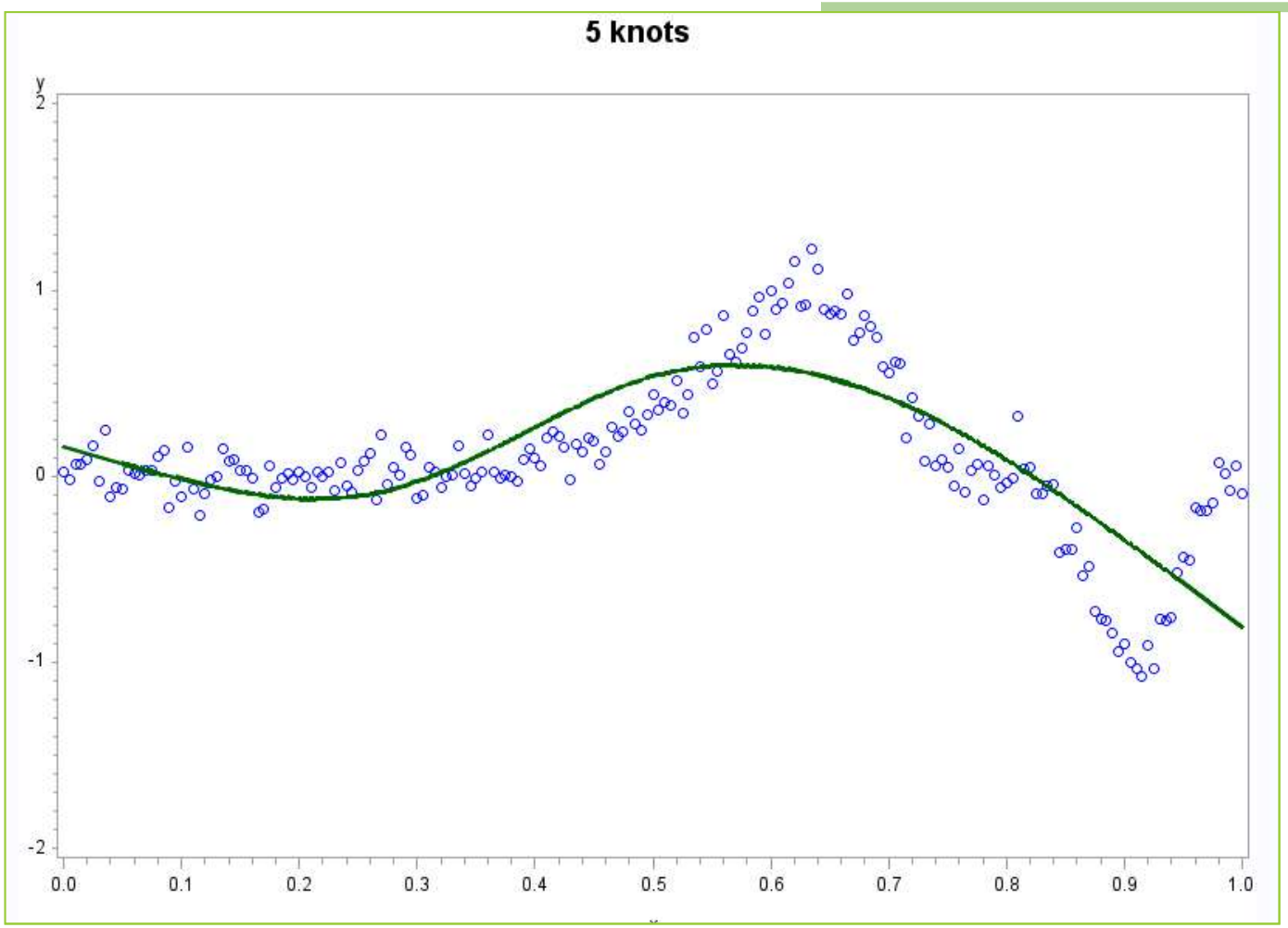
```
proc univariate data = messy;
  var x;
  output out=percentiles pctlpts=5 27.5 50 72.5 95
  pctlpre=p;
run;
data messy;
  if _N_ = 1 then set percentiles;
  set messy;
  %rcspline (x, p5, p27_5, p50, p72_5, p95);
run;
proc glm data = messy;
  model y = x x1 x2 x3; /* 5 knots, 3 new variables */
run;
```

http://biostat.mc.vanderbilt.edu/wiki/pub/Main/SASMacros/survrisk.txt

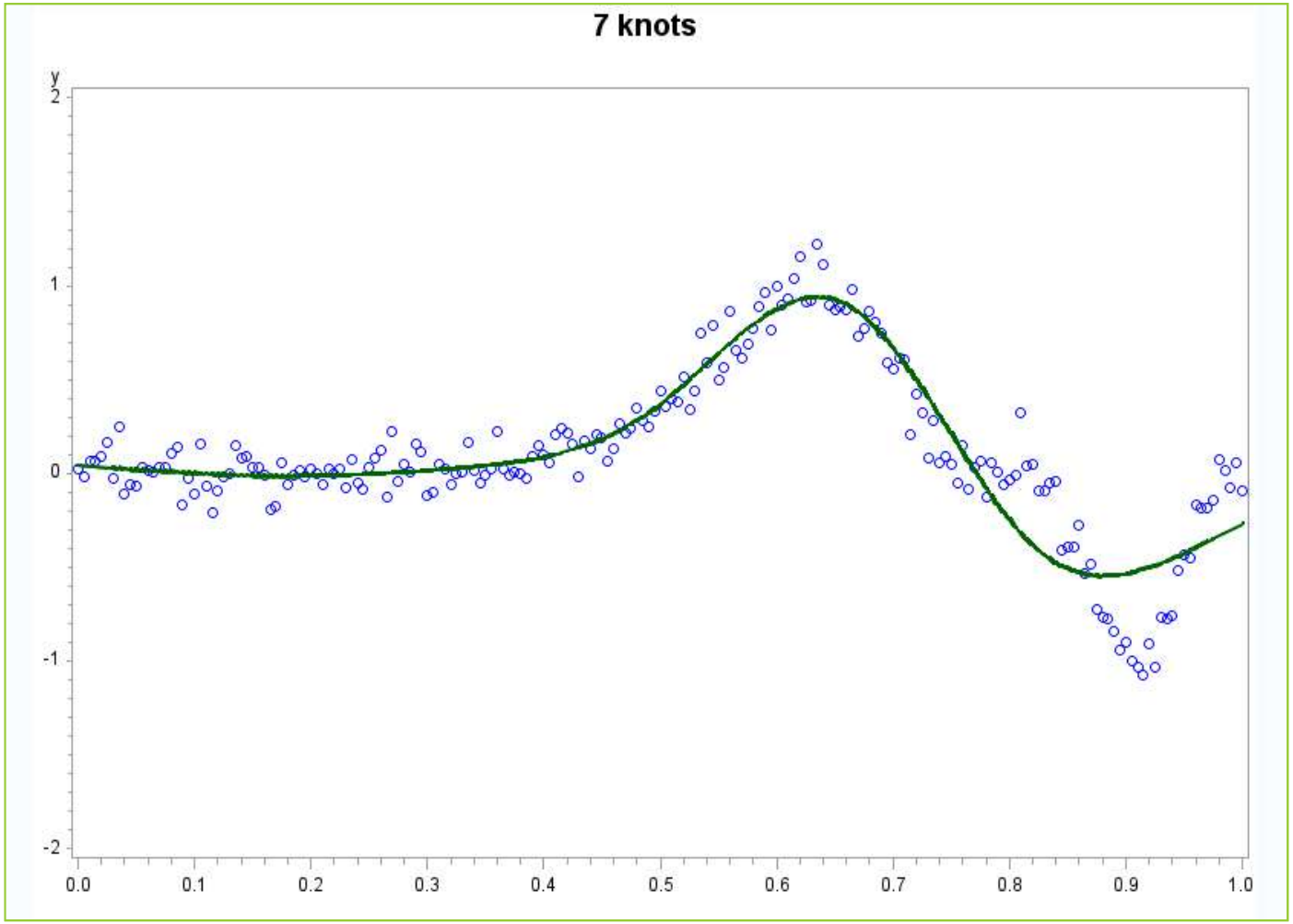
New Variables:

- Or you can use the ‘rcs’ function in R.

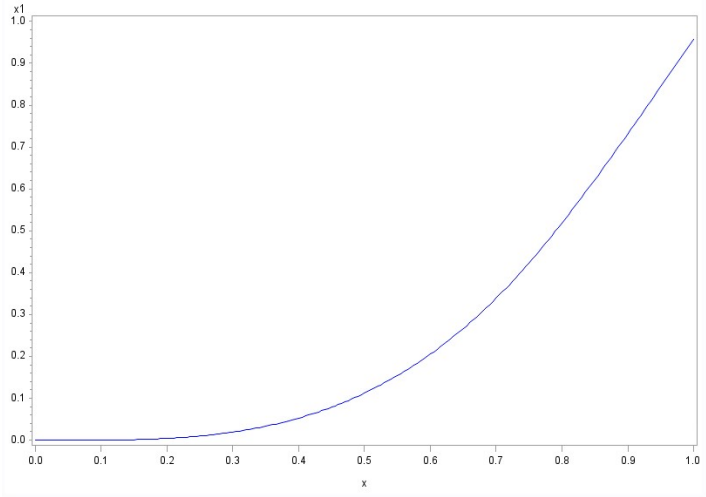
$R^2 = 0.61$

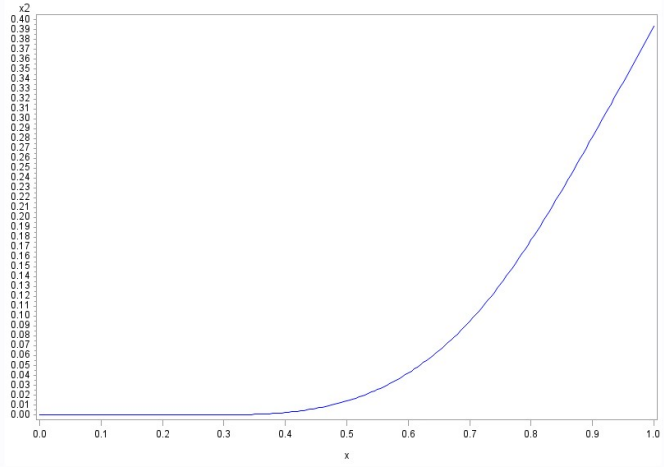


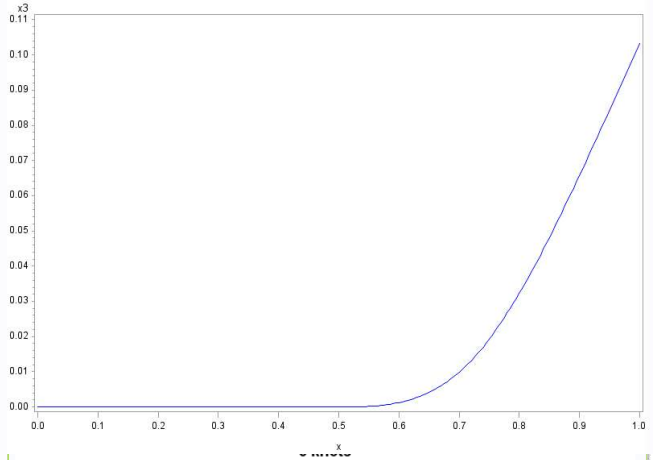
$R^2 = 0.85$

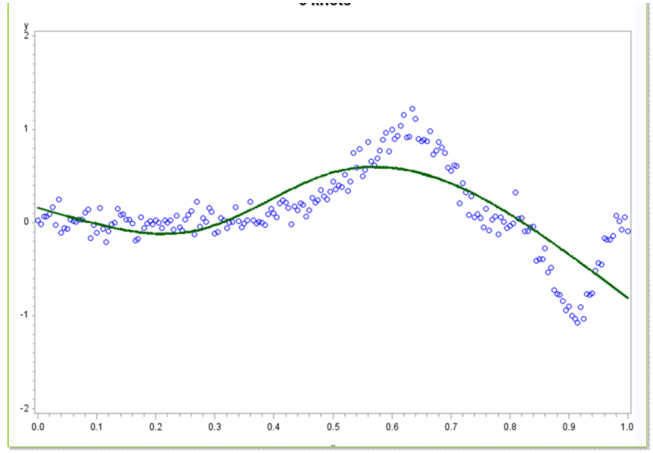


Percentiles (knots) are at X = 15, 33, 51, 69, and 87







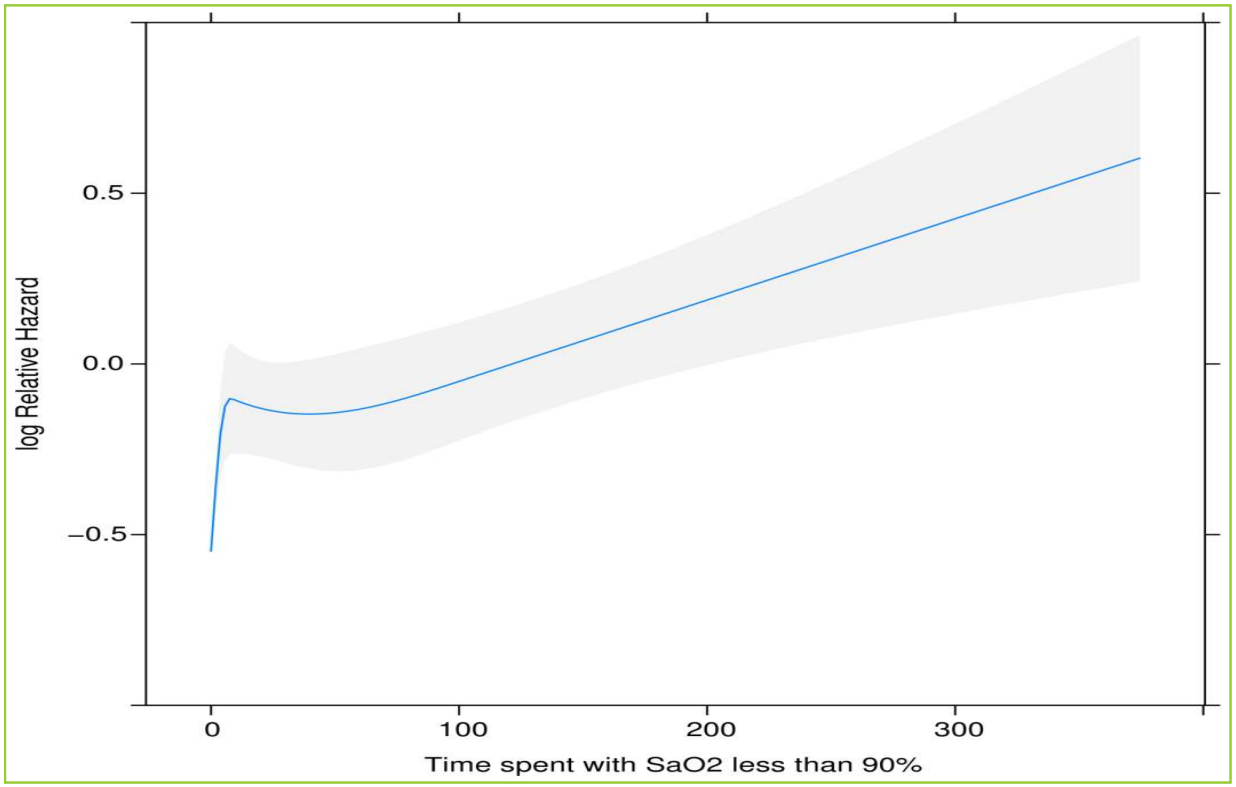


Can be Used with Any Type of Regression

- The original continuous predictor has been replaced by a set of variables, each of which is linear in the regression coefficients.
- The model can be fit using the usual regression procedures.
- Inferences can be drawn as usual.
- Can test for non-linearity
- Since the new variables are simply a restatement of the predictor, restricted cubic splines can be used in any type of regression (OLS, logistic, survival)

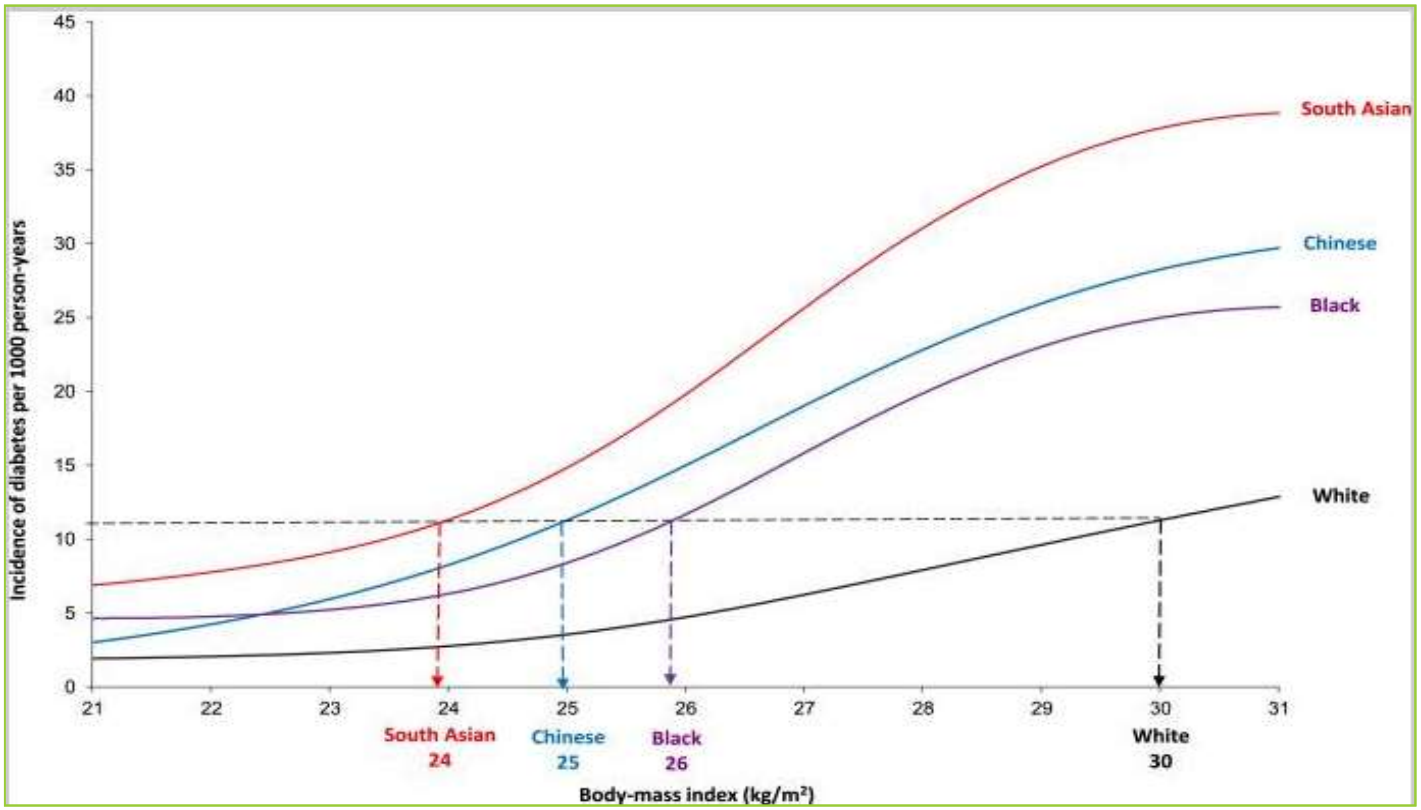
Splines at ICES

Example



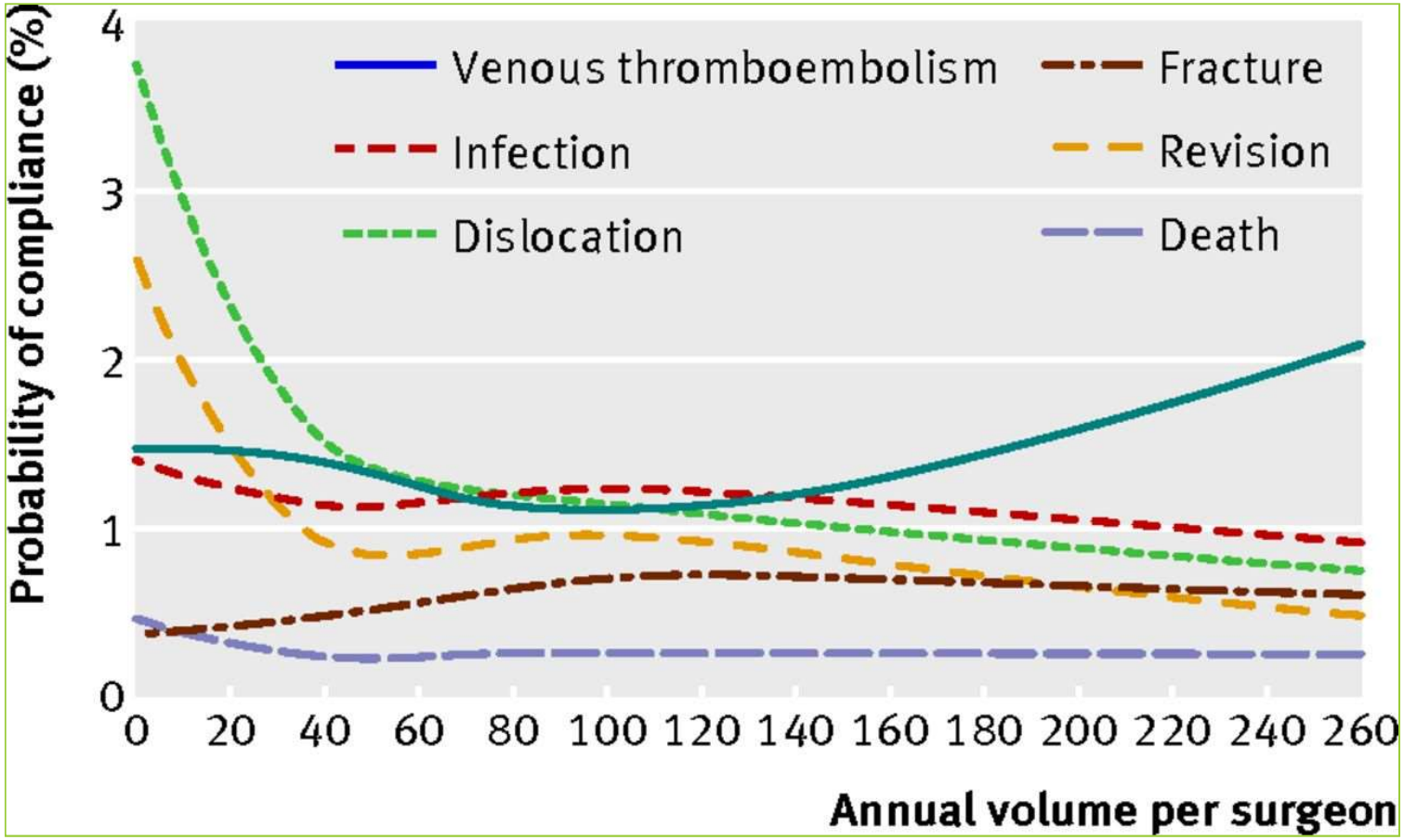
Kendzerska et al. Obstructive sleep apnea and risk of cardiovascular events and all-cause mortality. PLOS Medicine, 2014

Example



Chiu M et al. Deriving ethnic-specific BMI cutoff points for assessing diabetes risk. Diabetes Care, 2011

Example



Ravi B et al. Relation between surgeon volume and risk of complications after total hip arthroplasty. BMJ 2014.

Drawbacks

- Addition of new predictors – a problem for small datasets
- Difficult in presenting and interpreting the results.
- Graphical presentation is helpful.

Fractional Polynomials

Fractional Polynomials

- Another flexible method for modelling curved relationships using few parameters
- Proposed by Royston and Altman in 1994*
  - Noted that polynomials lack flexibility, often fit badly at the extremes, cannot model asymptotes
- Similar to conventional polynomials
- Non-integer and negative powers are also allowed

*Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. Applied Statistics. 1994; 43(3): 429-467.

Fractional Polynomials

- *m* powers are chosen from a small, predefined set of values
- "We have so far found that models with degree higher than 2 are rarely required in practice." In practice, m is usually 2.
- For m = 2: S = {-2, -1, -0.5, 0, 0.5, 1, 2, 3}

  $X^{(0)} = \ln(X)$

- X has to be positive
- Can create a new variable: $X + \delta$ if X is not positive. (For $X \geq 0$, add 1.) Note that this changes the variable.

Fractional Polynomials

- The degree m of an FP model is the number terms in powers of X in the model:

  Example of FP1 (*m* = 1): $Y = b_0 + b_1X^{-1}$

  Example of FP2 (*m* = 2): $Y = b_0 + b_1X^{-1} + b_2X^2$

- As in the second example, the model does not have to be hierarchical.

- Each additional term in the model adds 2 d.f. (one for the power and one for the regression coefficient)

Fractional Polynomials

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

  $X^{(0)} = \ln(X)$

- How can we limit our search for the “best” model to so few choices of power?
  - Royston and Altman argue that because the likelihood surface is usually nearly flat near the maximum, there is no need to precisely estimate the powers.
  - Powers not included in the set can be approximated using the log plus one of the powers in the set.

How Many Models are There?

To find the best models with degrees 1 and 2, we fit 44 models:

- 8 choices for *m* = 1 (FP1) (one for each of the powers in S)
- 28 choices for m = 2 and $p_1 \neq p_2$
- 8 choices for m = 2 and $p_1 = p_2$
  - If $p_1 = p_2 \neq 0$, the model contains $X^{(p1)}$ and $X^{(p1)} \ln(X)$.
  - If $p_1 = p_2 = 0$ then the mode contains $\ln(X)$ and $\ln(X)^2$.
- The best model for a given *m* is the one with the smallest deviance

Comparing Models

- The “best” model of a given size is that with the highest likelihood (lowest deviance D = -2 log-likelihood).
- To compare the deviances of 2 models, the difference between the deviances is compared to a χ^2 distribution.

Choice of Model (the RA2 algorithm)

1. Compare the best FP2 model for X against the null model using a test with 4 degrees of freedom and p-value α. If the test is not significant, then stop the process and conclude that the effect of X is not significant at the α level.

2. Test for the best FP2 for X against a linear relationship at the α level using a test with 3 d.f. If the test is not statistically significant, stop the process and conclude that the best model is a linear model. Otherwise, continue.

3. Test the best FP2 for X against the best FP1 at the α level using a test with 2 d.f. If the test is not significant, the final model is FP1; otherwise the final model is FP2.

Choice of Model (the RA2 algorithm)

This algorithm is actually usually a bit conservative, meaning that the type I error rate is usually a bit smaller than α.

Ambler G, Royston P. J Stat Comput Simul. 2001; 69: 89-108

Choice of Model (the RA2 algorithm)

When several models fit equally well, the final choice may depend on

- Parsimony
- Visual comparison of the curves in relation to the data
- If appropriate, the plausibility of the curves beyond the observed range of X
- Non-statistical considerations ("the science of the problem").

Royston P, Altman DG. Applied Statistics. 1994; 43(3): 429-467.

Models with More than 1 Continuous Covariate

With a lot of hand waving:

- Uses backward selection
- Fits fractional polynomials to one covariate at a time, starting with the one which is most significant when backward selection is applied to the raw data. The functional forms of the remaining covariates are fixed.
- Cycle through the covariates multiple times.
- The algorithm terminates when no more excluded and when the functional forms of the remaining continuous covariates do not change.

Patrick Royston and Willi Sauerbrei. *Multivariable Model-Building. A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables.* Chapter 6.

1. Nominal P-values α_1 and α_2 are chosen. Typical values are $\alpha_1 = \alpha_2 = 0.05$. Values may differ among variables. Taking $\alpha_1 = 1$ for a given variable 'forces' it into the model (no variable selection). Taking $\alpha_2 = 1$ for a continuous variable forces the most complex permitted FP function to be fitted for it (no function selection).
2. Maximum permitted d.f. for FP functions are chosen; for example, four, two and one d.f. mean FP2, FP1 and linear functions respectively. Our suggested default is four d.f.
3. The full linear model is fitted. The 'visiting order' of the predictors is determined according to the P-value for omitting each predictor from the model. The most significant predictor is visited first and the least last. Assume that the variables $x_1, \ldots, x_k$ have been arranged in this order, which is retained in all cycles of the procedure.
4. Let $c = 0$, to initialize the cycle counter.
5. Let $j = 1$, to initialize the variables counter within each cycle.
6. If x_j is continuous, go to step 7. Otherwise, x_j is categorical or binary. The joint significance of its dummy variable(s) is tested at the α_1 level. All other variables currently in the model are included as adjustment terms. If x_j is significant, then it is retained; otherwise it is dropped. Go to step 8.
7. Step 1 of the FSP is applied to x_j at the $\alpha = \alpha_1$ level (see Section 4.10.2). If x_j is not significant it is dropped. Otherwise, steps 2 and 3 of the FSP are applied at the $\alpha = \alpha_2$ level, to choose an FP or linear function. All other variables currently in the model are included as adjustment terms. In the case of an FP with powers $\mathbf{p}$ being chosen, x_j is represented by transformed variables $x_j^{\mathbf{p}}$ in subsequent steps in which other variables are considered.
8. Including or dropping x_j applies until x_j is reconsidered in the next cycle.
9. Let $j = j + 1$. If $j \leq k$, return to step 6 to process the next predictor. Otherwise, continue to step 10.
10. Let $c = c + 1$. The cth cycle is complete. If $c > c_{\max}$, stop (a practical value is $c_{\max} = 5$). Report that the algorithm has failed to converge in $c_{\max}$ cycles. Otherwise, check whether included variables and FP transformations have changed from cycle $c - 1$ to cycle c. If so, return to step 5 to start a new cycle. If not, stop and report the current model estimates. End of procedure.

Fractional Polynomials at ICES

Example: deriving a prognostic index

Derivation of the LACE+ index. We first derived a logistic regression model for 30-day death or urgent readmission using data for a randomly selected group of 250 000 patients. We entered all candidate covariates into an initial multivariable model and then performed variable selection (with a significance level of $\alpha = 0.05$) using methods described by Sauerbrei and Royston. These methods combined backward selection with a systematic process of identifying the optimal first-degree fractional polynomial transformation for continuous covariates.

Van Walraven C et al. Open Medicine, 6(3) (2012)
http://www.openmedicine.ca/article/view/498/468

and a number of similar articles deriving a prognostic index

Table 2
Final risk prediction model

Covariate*	Parameter estimate (SE)	Adjusted odds ratio (95% CI)†	*p* value
Male	0.10422 (0.01690)	1.11 (1.07–1.15)	<.0001
Urgent admission	0.60273 (0.02333)	1.83 (1.75–1.91)	<.0001
Discharge institution‡			0.0067§
Teaching vs small non-teaching hospital	−0.01328 (0.02600)	0.99 (0.94–1.04)	
Large vs small non-teaching hospital	−0.06150 (0.02384)	0.94 (0.90–0.99)	
Age^2	0.00032 (0.00002)	NA	<.0001
Log (length of stay)	0.28249 (0.01147)	NA	<.0001
Charlson $score^{0.5}$	1.32586 (0.07319)	NA	<.0001
Log (number of ED visits in previous 6 months)	0.37177 (0.01622)	NA	<.0001
Number of urgent admissions in previous $year^{0.5}$	1.81390 (0.09126)	NA	<.0001
Number of elective admissions in previous $year^{-1}$	−0.50616 (0.05751)	NA	<.0001
CMG $score^2$	0.01393 (0.00031)	NA	<.0001
Number of days on ALC status	−0.01033 (0.00209)	0.99 (0.99–0.99)	<.0001
Age^2 × Charlson $score^{0.5}$	−0.00005 (0.00001)	NA	<.0001
Age^2 × Number of urgent admissions in previous $year^{0.5}$	−0.00011 (0.00001)	NA	<.0001
Charlson $score^{0.5}$ × Number of urgent admissions in previous $year^{0.5}$	−0.31468 (0.04856)	NA	<.0001

Fractional Polynomials at ICES

Austin PC, Park-Wyllie LY, Juurlink DN. Using fractional polynomials to model the effect of cumulative duration of exposure on outcomes: applications to cohort and nested case-control designs. Pharmacoepidemiology and Drug Safety 2014; 23: 819-829.

and additional articles looking at the effect of cumulative drug use

Advantages and Disadvantages

- Offer flexibility, “a remarkable range of curves can be created”
- Straightforward to fit using standard methods
- Difficult in presenting and interpreting the results.
- Graphical presentation is helpful.

Which Acronym: RCSs or FPs?

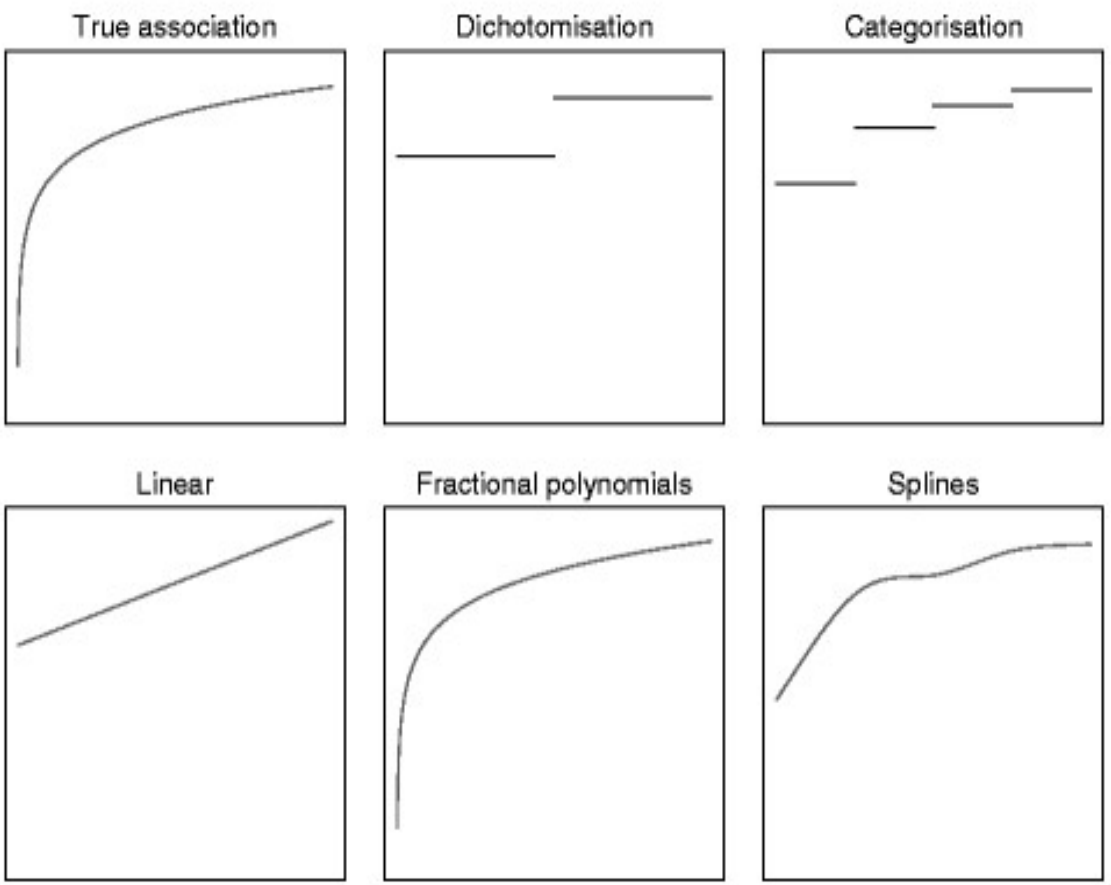
RCSs or FPs?

- The performance of restricted cubic splines vs. fractional polynomials depends on the nature of the data (which of course is not known outside of a simulation study)
- On average, it appears that fractional polynomials perform better (less bias and/or more precision) than restricted cubic splines, though the differences don't seem to be large.

- RCS's cannot be used to model time-dependent covariates in survival analysis (e.g., to model cumulative exposure to a drug), since the location of the knots would constantly change.

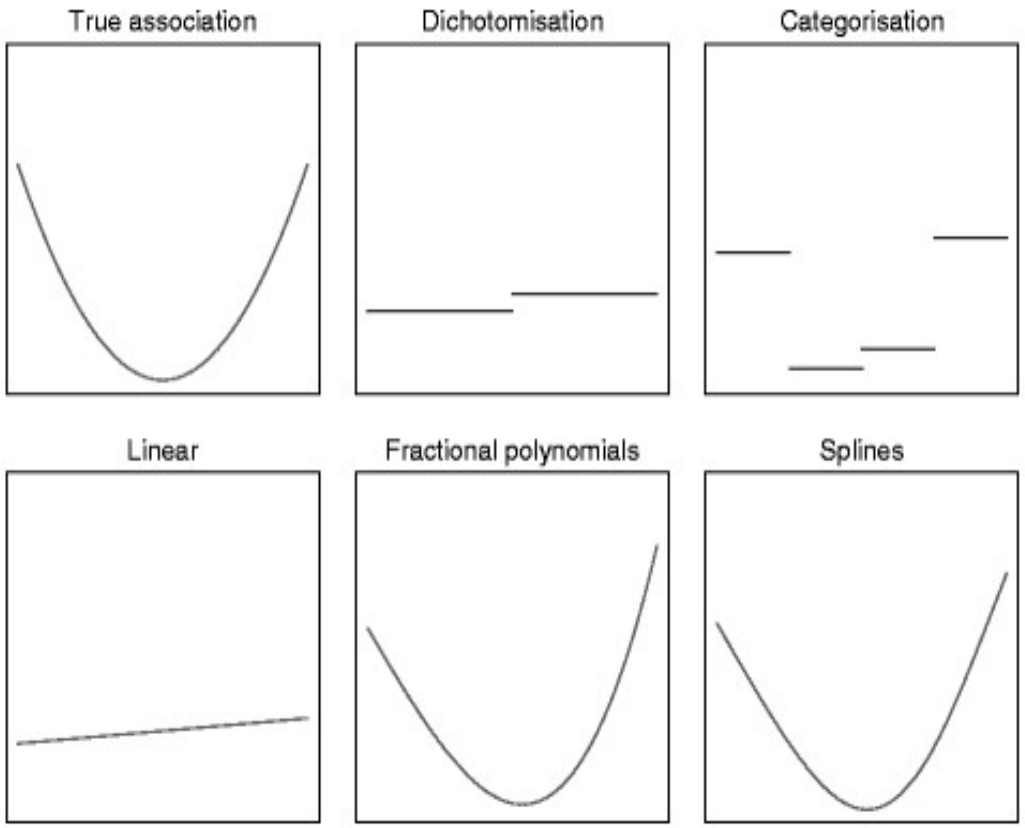
RCSs or FPs?

True association: Y = log(X)



RCSs or FPs?

True association: $Y = X^2$

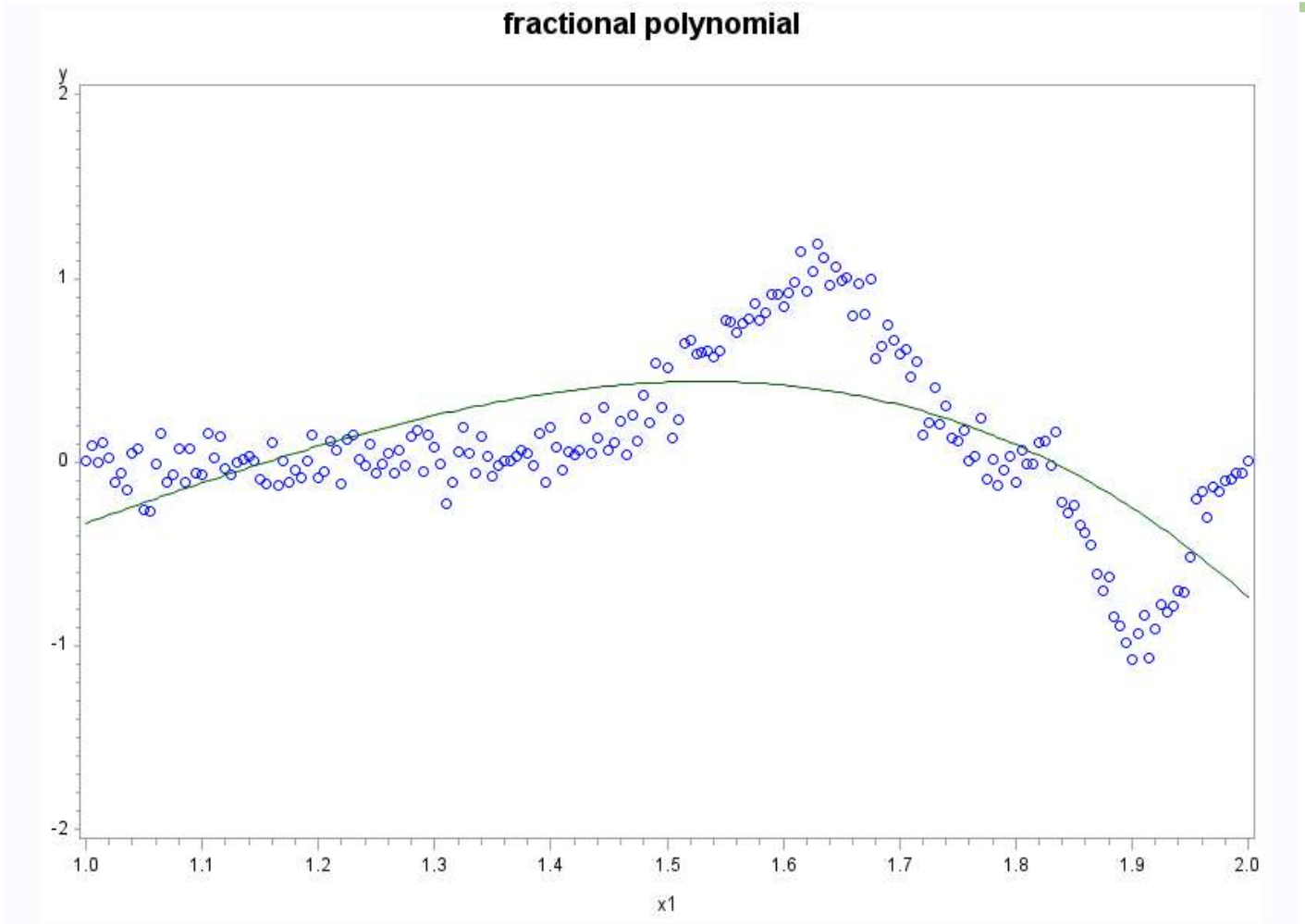


Kahan BC. et al. A comparison of methods to adjust for continuous covariates in the analysis of randomised trials. BMC Medical Research Methodology 2016; 16: 42
https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-016-0141-3

And what about my “messy” equation?

- Had to add 1 to the X values, because the minimum value was 0
- Best FP2 model was 3, 3 (i.e. $((X+1)^3 + \ln(X+1))$; deviance = 22.47
- Best FP1 model was 3 (i.e. $(X+1)^3$) ; deviance = 38.68
- Linear model: deviance = 39.95
- Null model: deviance = 40.36

- Compare best FP2 model with null model: $\chi^2_4 = 17.9$, p = 0.001
- Compare best FP2 model with linear model: $\chi^2_3 = 17.48$, p = 0.001
- Compare best FP2 model with best FP1 model: $\chi^2_2 = 16.21$, p = 0.000


fractional polynomial
y
2
1
0
-1
-2
1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
2.0
x1

Summary

- Failure to identify nonlinearity and include it in a model can result in an overestimated or underestimated relationship or a relationship that is missed altogether
- Methods which keep covariates as continuous typically have higher power than methods that use categorization
- When the true association is non-linear, categorization or the assumption of a linear association lead to large reductions in power.

Summary

- Restricted cubic splines and fractional polynomials are ways to test whether a relationship is non-linear, using standard techniques.

- Both methods allow non-linear relationships to be modeled well:
  - Reducing model misspecification
  - Providing insight
  - Allow for great flexibility in the form of the relationship between predictor and outcome

- Can be used in OLS, logistic, and survival analysis.

Questions?