International initiative:
STRengthening Analytical Thinking for Observational Studies (STRATOS)

Review of Methods used in recent observational epidemiological studies to Select Variables & their Functional Forms

Michal Abrahamowicz* & Ryan P. Kyle

(McGill University, Montreal, Canada)

for the members of the STRATOS Task Group TG2

Support:

Canadian Institutes for Health Research (CIHR) grant 81275

Members of TG2

- **Chairpersons:**
  - Michal Abrahamowicz (McGill, Montreal, Canada)
  - Willi Sauerbrei (Freiburg, Germany)

- **Additional members so far:**
  - Harald Binder (Mainz, Germany)
  - Frank Harrell (Vanderbilt, Franklin, USA)
  - Gary Collins (Oxford, UK)
  - Patrick Royston (London, UK)

OUTLINE

- Goals of STRATOS TG2
- Overview of Flexible methods for modeling Functional Forms of Continuous Independent Variables (X)
- Impact of X modeling on variables selection
- Literature Review: Objectives & Methods
- Literature Review: summary of Findings
- Real-life & Simulated Examples of Drawbacks of methods currently used in Epidemiological/Clinical research
- Plans for Future TG2 activities

Main issues addressed by TG2

TG2 focuses on **2 inter-related questions**, common to all multivariable explanatory models :

1. Selection of 'relevant' Variables
2. Choice of the Functional Form for the effect of each Continuous variable, i.e. Modeling of the effects of Continuous Independent Variables

FLEXIBLE MODELING of the Functional Forms for Continuous Predictors

- **Flexible Modeling techniques, proposed to estimate Non-linear (NL) effects of Continuous X's, with different Smoothers, include e.g.:**
- Fractional Polynomials (FP) [Royston&Sauerbrei2008;Royston&Altman 1994]
- Regression Splines

  [Ramsay 1988; Abrahamowicz & MacKenzie 2007]
- Restricted Cubic Splines

  [Harrell (2001)]
- Penalized Smoothing Splines

  [Gray JASA 1992, 87: 942-951]
- Generalized Additive Models (GAM)

  [Hastie & Tibshirani , 1990]
- .....+ several other types of (I- , P- ...etc) -Splines

Functional Forms for Continuous Independent Variables

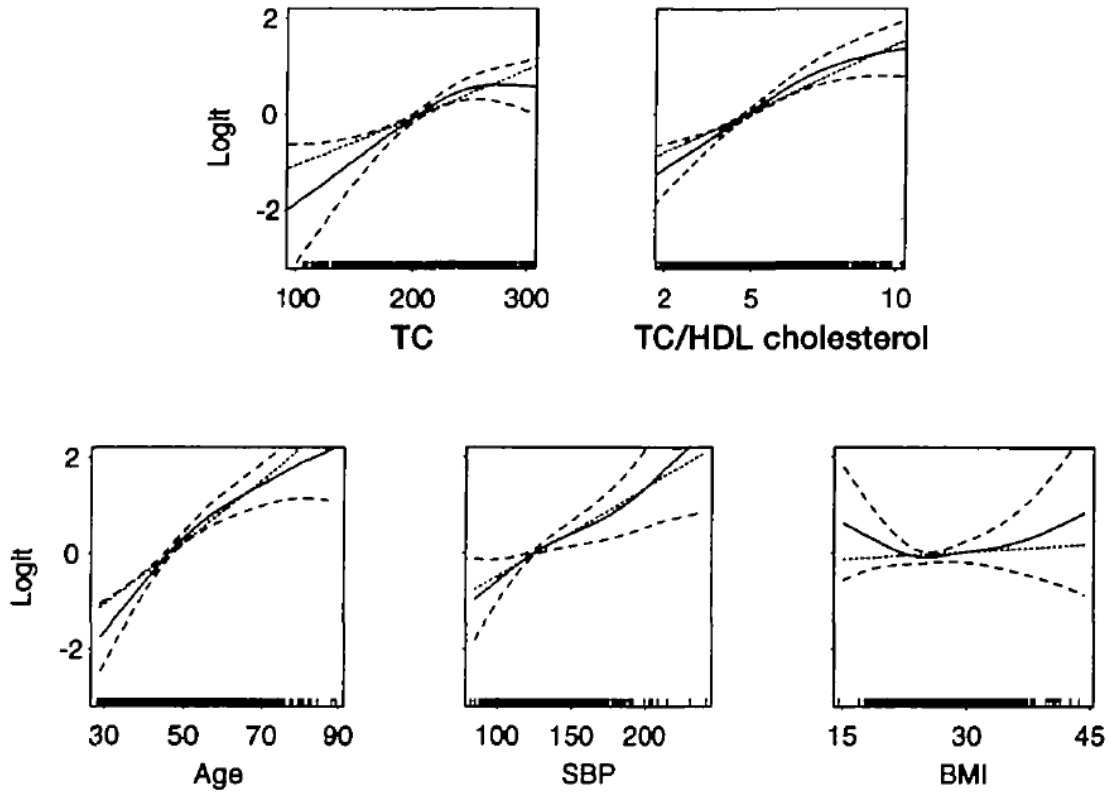
- To understand the role of Continuous Predictor (X) in an Explanatory Model (for a given outcome), **we need to estimate the 'etiologically correct' Dose-Response function *g(x)*** (a continuous, *smooth* transformation of X)
- **Conventional models usually A Priori assume that g(x) is Linear** & include Un-transformed X: **g(x) = βx**
- Linearity assumption is convenient (effect of X summarized by a single β, parsimony = improved power), and often adequate
- **Yet, Linearity should not be imposed *a priori*: numerous examples of Non-Linear or Non-Monotone effects,** e.g.:

(i) **BMI -> all-causes mortality** (both Obese and Too Thin subjects have Increased Risks),

(ii) **Age at diagnosis -> mortality in different cancers** (Youngest subjects have more aggressive disease,

Oldest have increased risk of all-cause mortality)

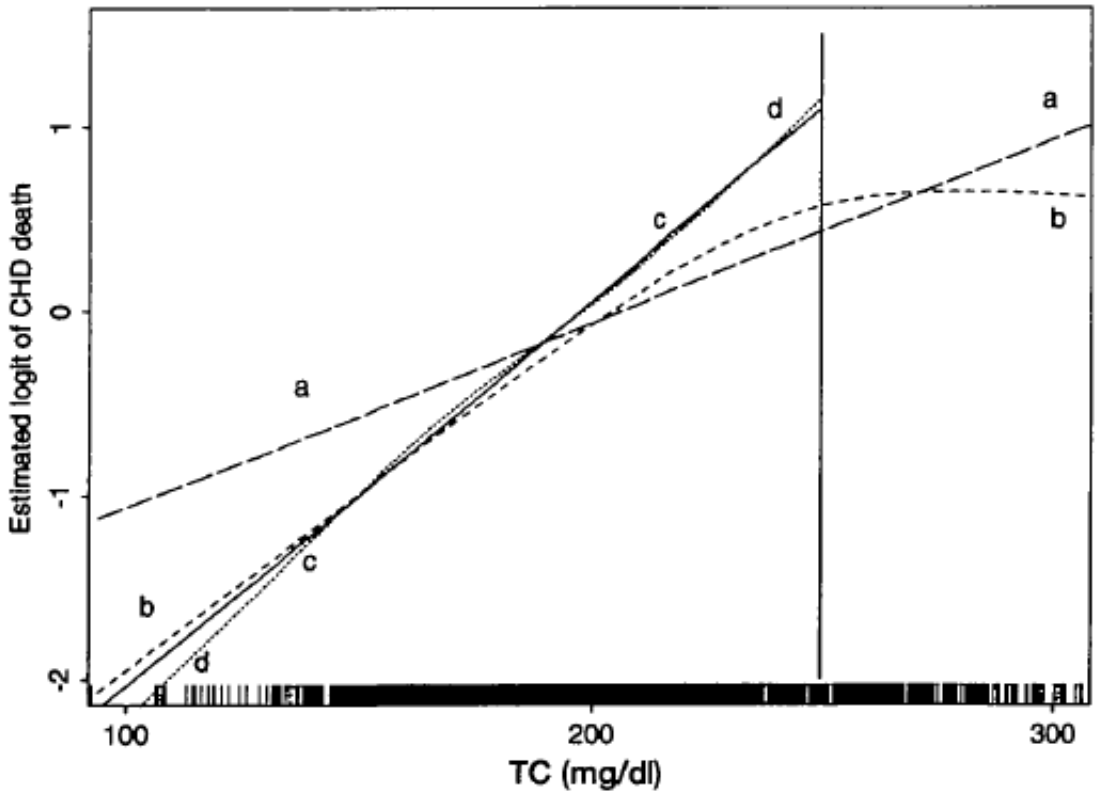
GAM-estimated Non-linear effects of Risk Factors on logit of Coronary Heart Mortality [Abrahamowicz et al, AJE 1997]



Flexible Modeling of Continuous variables avoids 'local biases' of a Linear Function:

Cholesterol (X) vs logit of Cardiovascular Death (Y) [Abrahamowicz et al, *AJE* 1997]

- **(a) & (b): full range of X; (c) & (d) X<250; (a) & (c) linear (βx);**
- **(b) & (d) Smoothing Spline (GAM)**



Inter-Dependence of the Selections of (1) Variables vs (2) Functional Forms

- **An additional CHALLENGE is that the results of**
  (1) Data-dependent selections of Independent Variables ('Predictors')
  may Depend on
  (2) decisions regarding Functional Forms of both:
    (2a) the Predictor of Interest (X) &
    (2b) Other Variables, correlated with X;
  and *vice versa*

[Rosenberg PS, Katki H, Swanson CA, Stat Med 2003, 22: 3369-3381]

Objectives of Literature Review

- **OVERALL:**

  to Demonstrate the Need for STRATOS –initiated efforts to enhance the methodological standards of the analyses reported in current Applied research

- **Specific:**

  > Document the Methods and approaches actually applied in 2013 in Empirical Observational studies published in major Clinical & Epidemiological Journals for:

  (i) select independent variables into a Multivariable Model

  (ii) Model the effects of Continuous Independent Variables

  > Identify the Limitations & Drawbacks of the currently applied methods

Literature Search Methods

- We selected 2 subsets of Journals:
- (A) **5 major Epidemiology journals:**

  American J Epi (AJE), *Epidemiology*, *Epidemiology & Community Health* (JECH), *International J Epi* (IJE), *J Clinical Epi* (JCE)

  (B) **8 major 'general' Clinical journals:**

  Arch Int Med (AIM), *BMJ*, *Circulation*, *JAMA*, *J Infect Dis* (JID), *J Natl Cancer Inst* (JNCI), *Lancet*, *New Engl J Med* (NEJM)

.....

- From Each Subset we selected, by Simple Random Sampling (NOT stratified by journal) 25 Papers, published in the 1st half of 2013, which met our Inclusion/Exclusion Criteria
- We then Reviewed the Methods applied in each paper, focusing on issues most relevant for TG2

Literature Search Methods

- **Inclusion Criteria:**
  > included **Multivariable regression analyses,**
  > **At least 1 Continuous Independent Variables** included in the model(s),
  > electronic or print **Publication Date: 1 January to 30 June 2013.**

- **Exclusion Criteria:**
  > Analyses of **Correlated data** (e.g. GEE, mixed, frailty models),
  > **Experimental** (Non-Observational) studies e.g. **Clinical Trials**,
  > studies with the **Effective Sample Size < 50.**

- **Additional Search Criteria**:
  search strategy also targeted publications with at least 1 of the following "keywords" in the Abstract or Title:
  model, regression, estim*, multiv*, assoc**

Distribution of the 25 Sampled (Eligible) Papers across the pre-Selected Journals

EPIDEMIOLOGY:

AJE	Epidemiology	IJE	JCE	JECH
11 (44%)	2 (8%)	3 (12%)	0 (0%)	9 (36%)

CLINICAL:

AIM	BMJ	Circulation	JAMA	JID	JNCI	Lancet	NEJM
2 (8%)	7 (28%)	4 (16%)	2 (8%)	2 (8%)	3 (12%)	1 (4%)	4 (16%)

Types of Regression Models used

Multivariable Model (*)	EPIDEMIOLOGY journals (**) (% / 25)	CLINICAL journals (**) (% / 25)
Logistic (Binary outcome)	**12 (48%)**	**10 (40%)**
Cox PH	**7 (28%)**	**14 (56%)**
Linear	**9 (36%)**	1 (4%)
Poisson	1 (4%)	1 (4%)
Polytomous logistic	2 (8%)	0 (0%)
Inverse Gaussian (log-linear)	1 (4%)	0 (0%)
Relative risk (log-binomial)	0 (0%)	1 (4%)
TOTAL (*)	**32**	**27**
(*) > 1 model in 11/50 papers	(**) NOT mutually exclusive	(**) NOT mutually exclusive

Criteria/Methods for Selecting Independent Variables into a Multivariable Model

Criteria/Methods	EPIDEMIOLOGY journals (% / 25)	CLINICAL journals (% / 25)
Not Reported explicitly	**11 (44%)**	**13 (52%)**
A priori (based on Substantive knowledge)	**11 (44%)**	**5 (20%)**
A priori: DAG-based (Substantive knowledge)	-	1 (4%)
STAT: P<0.05 for "crude effect" in Bivariate analyses	1 (4%)	**4 (16%)**
STAT: P<0.05 for "Adjusted effect" in Full Multivariable model	1 (4%)	1 (4%)
STAT: Stepwise selection	-	1 (4%)
STAT: > 10% Change in the Estimated 'Exposure' effect	1 (4%)	-

Variables Selection methods in EPI: Comparison with a 2008 Review

- Walter & Tiemeir reviewed methods used for selection of covariates into multivariable models in 300 papers published in 2[illegible] [illegible]gy, Eur[illegible]

	W & T (300 EPI papers in 2008)	Our Review (EPI papers in 2013)
Not described.	105 (33%)	44%
Prior knowledge	87 (27.7%)	44%
Stepwise selection	59 (19.6%)	0%
Change –in-Estimate	44 (14.7%)	4%
Other	9 (3%)	8%

Criteria/Methods for Selecting Independent Variables in studies focusing on building Multivariable Explanatory/Etiologic Models

Criteria/Methods	6 papers in EPIDEMIOLOGY journals (% / 6)	8 papers in CLINICAL journals (% / 8)
Not Reported explicitly	**2 (33.3%)**	**5 (62.5%)**
A priori (based on Substantive knowledge)	**4 (66.7%)**	**-**
A priori: DAG-based (Substantive knowledge)	-	-
STAT: P<0.05 for "crude effect" in Bivariate analyses	-	**2 (25.0%)**
STAT: P<0.05 for "Adjusted effect" in Full Multivariable model	-	**1 (12.5%)**
STAT: Stepwise selection	-	-
STAT: > 10% Change in the Estimated 'Exposure' effect	-	-

Functional Forms for Modeling Continuous Independent Variables

Representation of Continuous Variables (NOT Mutually Exclusive)	EPIDEMIOLOGY journals (% / 25)	CLINICAL journals (% / 25)
Dichotomized	**7 (28%)**	2 (8%)
Categorized (> 2 categories)	**19 (76%)**	**14 (56%)**
Continuous, Un-transformed (**Linear** effect assumed *a priori*)	**19 (76%)**	**22 (88%)**
A priori defined **Parametric Transformation(s)** (e.g. log or polynomial)	**8 (32%)**	1 (4%)
Restricted Cubic Splines	1 (4%)	-
Other **Spline-based** methods	-	-
Fractional Polynomials	-	-

Modeling of Age and BMI (shown Consistently to have Non-linear effects on many health outcomes)

Several articles included in the Multivariable Analyses some “generic” Continuous Risk/Prognostic factors, such as Age & Body Mass Index (BMI)

Modeling	AGE (EPI)	AGE (Clinical)	BMI (EPI)	BMI (Clinical)
Linear Only	1		1	
Linear & Categories	2			
Dichotomi-zed Only			2	
Only > 2 categories				

Ent

Functional Forms: representation of Continuous "Main Exposure" variables

Representation of Continuous Variables (NOT Mutually Exclusive)	12 papers in EPIDEMIOLOGY journals (% / 12)	4 papers in CLINICAL journals (% / 4)
Dichotomized	1 (8.3%)	-
Categorized (> 2 categories)	**6 (50 %)**	**4 (100%)**
Continuous, Un-transformed (**Linear** effect assumed *a priori*)	**6 (50 %)**	**-**
A priori defined **Parametric Transformation(s)** (e.g. log or polynomial)	**3 (25 %)**	-
Restricted Cubic Splines	1 (8.3%)	-
Other **Spline-based** methods	-	-
Fractional Polynomials	-	-

Example of **Categorization of BMI** effects: OR's (95% CI) for **2 different Outcomes** [1]

BMI Category	Mobility problems (Non-Linearity)	Daily Activities problems (Non-Monotonicity)
<18.5	0.91	1.61 *
[18.5; 25) [REF]	1	1
[25; 30)	1.18	1.08
[30; 40)	**2.00 (1.7 – 2.3)** ***	1.48 **
> 40	**5.31 (3.9 – 7.2)** ***	2.14 **

- [1] = [Maheswaran H, et al. Estimating utility values for major behavioral risk factors in England, *JECH* 2013, 67: 172-180]

Comments on Modeling of Age and BMI

- **Yet, both Age & BMI have Non-Linear or even Non-Monotone effects on different outcomes**:

(i) BMI: **Non-Monotone** relationships with risks (**) of e.g.:

(1) CVD mortality [Abrahamowicz et al, AJE 1997],

(2) Both Anxiety & Problems with Daily Activities [Maheswaran et al, JECH 2013]

(**) Both Obese and Too Thin subjects have Increased Risks)

(i) Age at diagnosis -> mortality or recurrence in different cancers (Youngest subjects have more aggressive disease while Oldest have increased risk of all-cause mortality)

Drawbacks of Categorization of Continuous Predictors

- Our Review indicates that **CATEGORIZATION of continuous independent variables is still Very Common in Both Clinical & Epidemiological research**

- Yet, **Several Drawbacks of Categorization were demonstrated [1]**:

(i) Implausibility of the Step-Function effect & 'Local Bias' [2]

(ii) Arbitrary cut-offs for categories often vary wildly across studies of the same predictor-outcome association [3], inducing spurious differences

(iii) 'Bad' *a Priori* selection of cut-offs results in worse fit to data and increased Type II error

(iv) If cut-offs selected *a Posteriori:* standard Inference is Not valid, and increased risk of Type I error and overfit bias [4]

[1] Royston P, Altman DG, Sauerbrei W. *Stat Med* 2006, 25: 127-141.

[2] Sauerbrei W, Royston P, Bojar H, et al. *Br J Cancer* 1999; 79: 1752-1760.

[3] Malats N,Bustos A,Nascimento C et al. *Lancet Oncology* 2005, 6:678-686.

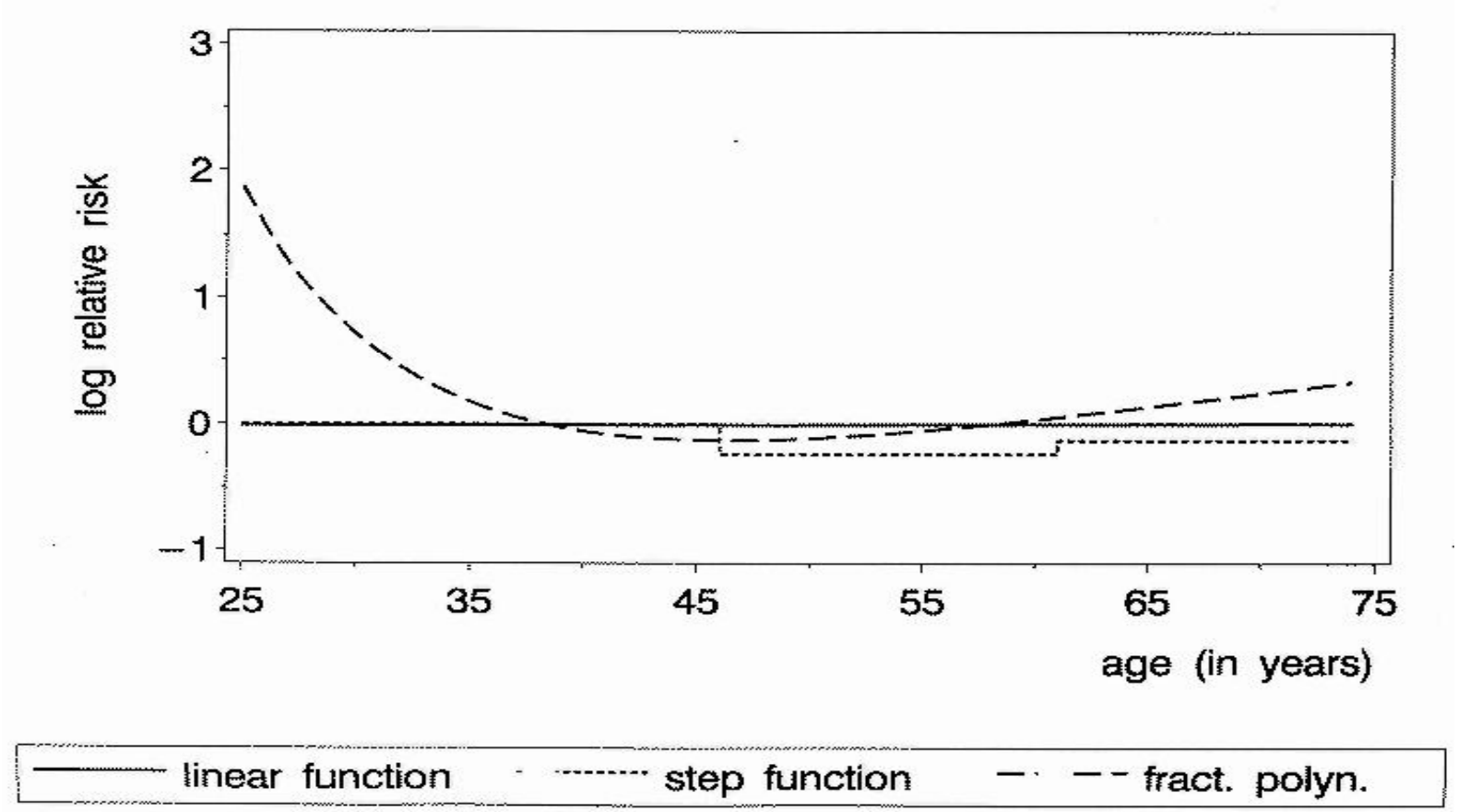
[4] Schulgen G, Lausen B, Olsen JH, Schumacher M. *AJE* 1994, 140(2): 172-184 .

Different Conclusions re: Stat. Significance

(depending on how continuous predictor is modeled)

AGE as predictor of Death or Recurrence in Breast Cancer (adjusted)

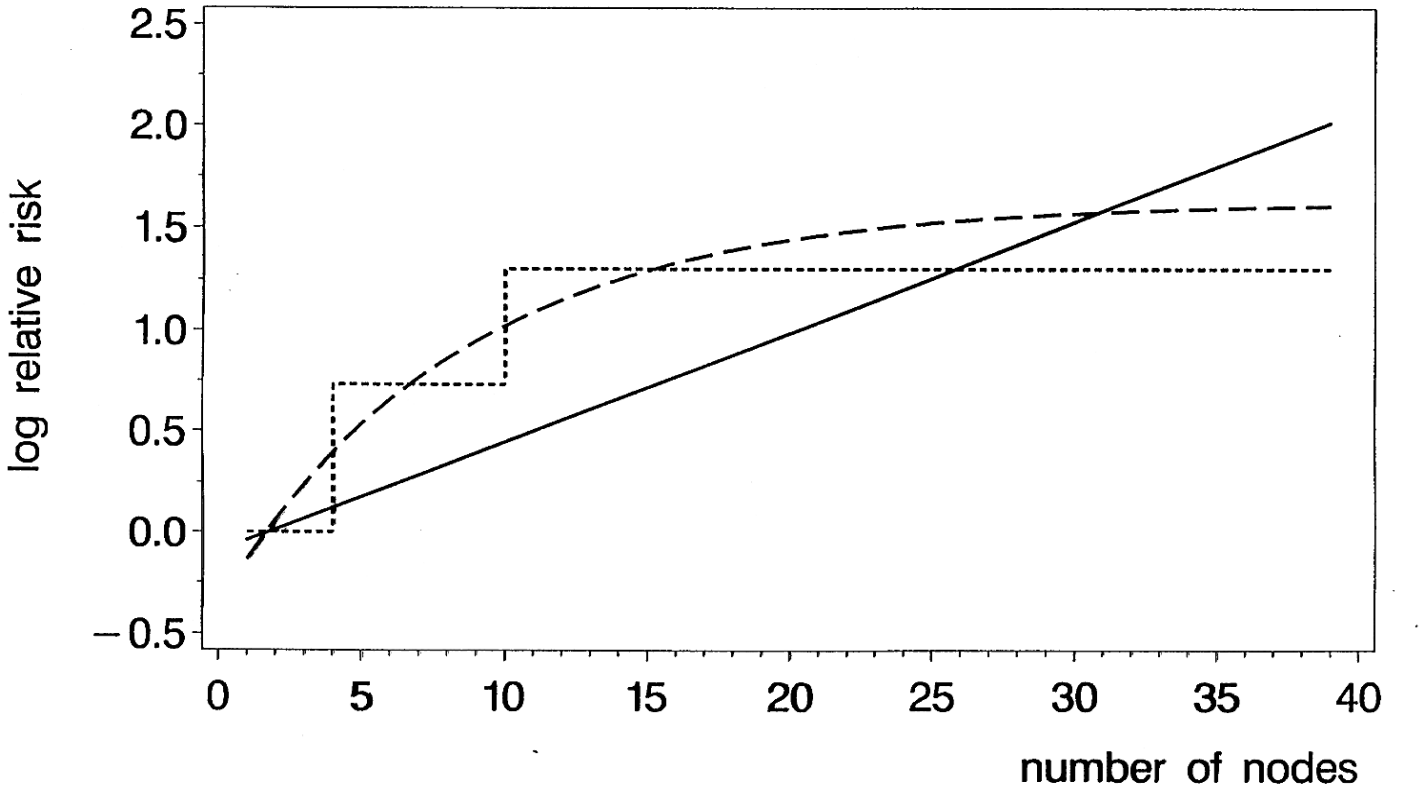
[Sauerbrei et al, Br J Cancer 1999]



P-value 0.9 0.2 0.001

NODES as predictor of Death or Recurrence in Breast Cancer:

Similar P-values but DIFFERENT ESTIMATES [Sauerbrei et al, Br J Cancer 1999]



P-value	0.001	0.001	0.001

Table 3 Results of the multivariable Cox's PH model ($N = 269$)

Variables	HR (95% CI)[a]	*P*-value for test of no association	*P*-value for test of PH	*P*-value for test of linearity
Stage: (IIIB+pleural effusion/4 *vs* IIIA/IIIB)	1.815 (1.268, 2.597)	0.001	0.204	N/A
ECOG[b] performance status: (2 *vs* 0-1)	1.348 (0.958, 1.896)	0.086	0.165	N/A
Smoking status: (ever *vs* never)	2.087 (1.349, 3.230)	0.001	0.135	N/A
Chemotherapy type: (single *vs* double)	1.539 (1.082, 2.188)	0.016	0.067	N/A
Log_2 CRP: (per doubling of CRP values)	1.108 (1.027, 1.196)	0.008	0.039	0.130
Albumin: (per $\downarrow$[c] of 1 $g l^{-1}$)	1.015 (0.974, 1.058)	0.485	<0.001	0.024
Log_2 LDH: (per doubling of LDH values)	2.159 (1.700, 2.742)	<0.001	0.636	0.590
Alkaline phosphatase: (per $\uparrow$[d] of 10 $U l^{-1}$)	1.019 (0.993, 1.047)	0.150	0.075	0.034
Neutrophil counts: (per $\uparrow$ of $1 \times 10^9 l^{-1}$)	1.082 (1.037, 1.129)	<0.001	0.027	0.041
Lymphocytes: (per $\downarrow$ of $1 \times 10^9 l^{-1}$)	1.307 (1.050, 1.626)	0.016	0.550	0.460
Deviance[e]	1902.2			
AIC	1922.2			

Abbreviations: AIC = Akaike's information criterion; CRP = C-reactive protein; LDH = lactate dehydrogenase; PH = proportional hazard. N/A: the test of linearity is not applicable to categorical covariates. [a]Adjusted hazard ratio (HR) and 95% confidence interval (95% CI). [b]Eastern cooperative oncology group. [c]$\downarrow$: decrease. [d]$\uparrow$: increase. [e]Deviance = −2*log-likelihood.

Table 4 Results of the flexible spline-based model ($N = 269$)

Variables	HR (95% CI)[a]	*P*-value for test of no association
Stage: (IIIB+pleural effusion/4 *vs* IIIA/IIIB)	1.859 (1.284, 2.691)	<0.001
ECOG[b] performance status: (2 *vs* 0–1)	1.336 (0.923, 1.935)	0.116
Smoking status: (ever *vs* never)	2.248 (1.419, 3.561)	<0.001
Chemotherapy type: (single *vs* double)	1.462 (0.990, 2.160)	0.041
Log_2 CRP: (per doubling of CRP values)	*	0.003 (overall *P*-value)[#]
Albumin: (per $\downarrow$[c] of 1 $g l^{-1}$)	**	0.001 (overall *P*-value)[#]
Log_2 LDH: (per doubling of LDH values)	2.281 (1.661, 3.142)	<0.001
Alkaline phosphatase: (per $\uparrow$of 10 $U l^{-1}$)	1.012 (0.980, 1.041)	0.366
Neutrophil counts: (per $\uparrow$[d] of $1 \times 10^9 l^{-1}$)	1.072 (1.025, 1.122)	0.001
Lymphocytes: (per $\downarrow$ of $1 \times 10^9 l^{-1}$)	1.313 (1.035, 1.666)	0.012
Deviance[e]	1873.3	
AIC	1909.3	

Abbreviations: AIC = Akaike's information criterion; CRP = C-reactive protein; LDH = lactate dehydrogenase. [a]Adjusted hazard ratio (HR) and 95% confidence interval (95% CI). [b]Eastern cooperative oncology group. [c]$\downarrow$: decrease. [d]$\uparrow$: increase. [e]Deviance = −2*log-likelihood. *Both the time-dependent ($P = 0.033$) and non-linear ($P = 0.015$) effects were significant. The estimated non-linear effects, at selected follow-up times, are shown in Figure 2. **Both the time-dependent ($P = 0.0001$) and non-linear ($P = 0.038$) effects were significant. The estimated non-linear effects, at selected follow-up times, are shown in Figure 3. [#]*P*-value for a likelihood ratio test, with 5 degrees of freedom, of the null hypothesis of no association, obtained by comparing the deviances of (i) a flexible model where both time-dependent and non-linear effects of a given variable are modeled by splines, *vs* (ii) a simpler 'reduced' model, which does not include the variable being tested (see the section on "Statistical analyses" for details of the test).

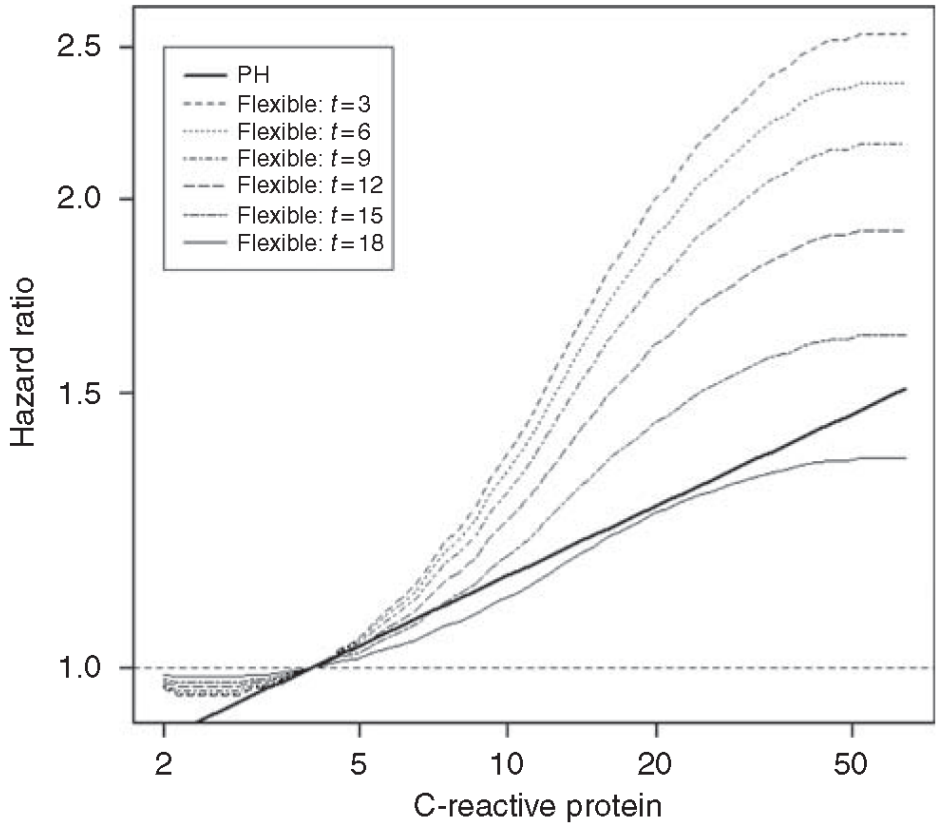


Figure 2 Results of the Cox's PH and flexible spline-based multivariable modeling of the effect of CRP on survival. The bold line represents the linear estimate from the Cox's PH model. The curves correspond to the flexible spline estimates at different times from 3 months ($t=3$) to 18 months ($t=18$) after the initiation of chemotherapy. Each curve shows how the adjusted hazard ratio at the corresponding time, relative to the value of 4 mgl^{-1}, changes with increasing value of C-reactive protein.

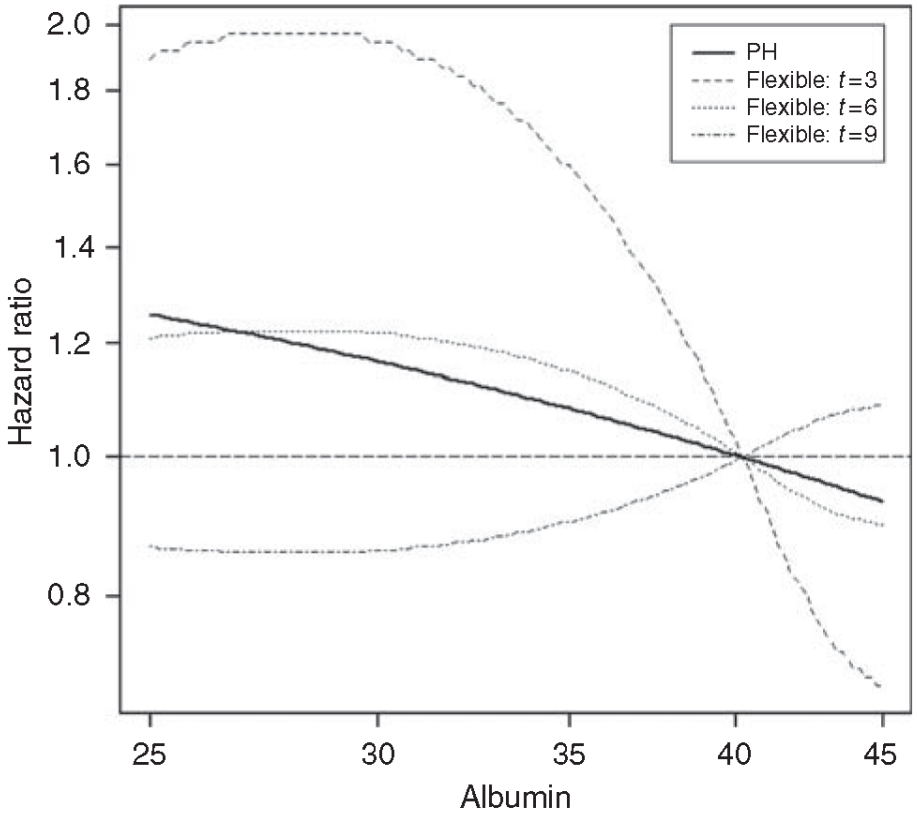
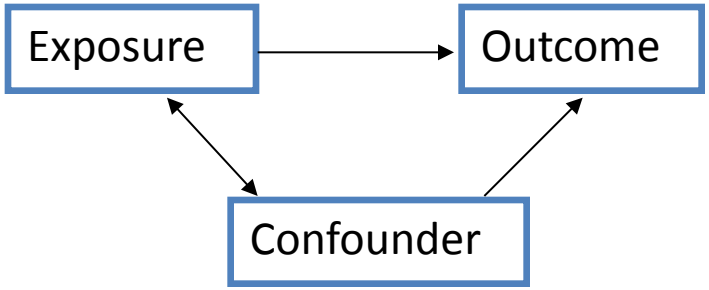


Figure 3 Results of the Cox's PH and flexible spline-based multivariable modeling of the effect of albumin on survival. The bold line represents the linear estimate from the Cox's PH model. The curves correspond to the flexible spline estimates at different times from 3 months ($t=3$) to 9 months ($t=9$) after the initiation of chemotherapy. Each curve shows how the adjusted hazard ratio at the corresponding time, relative to the value of 40 mg l^{-1}, changes with decreasing value of albumin.

Residual Confounding

- Confounder:

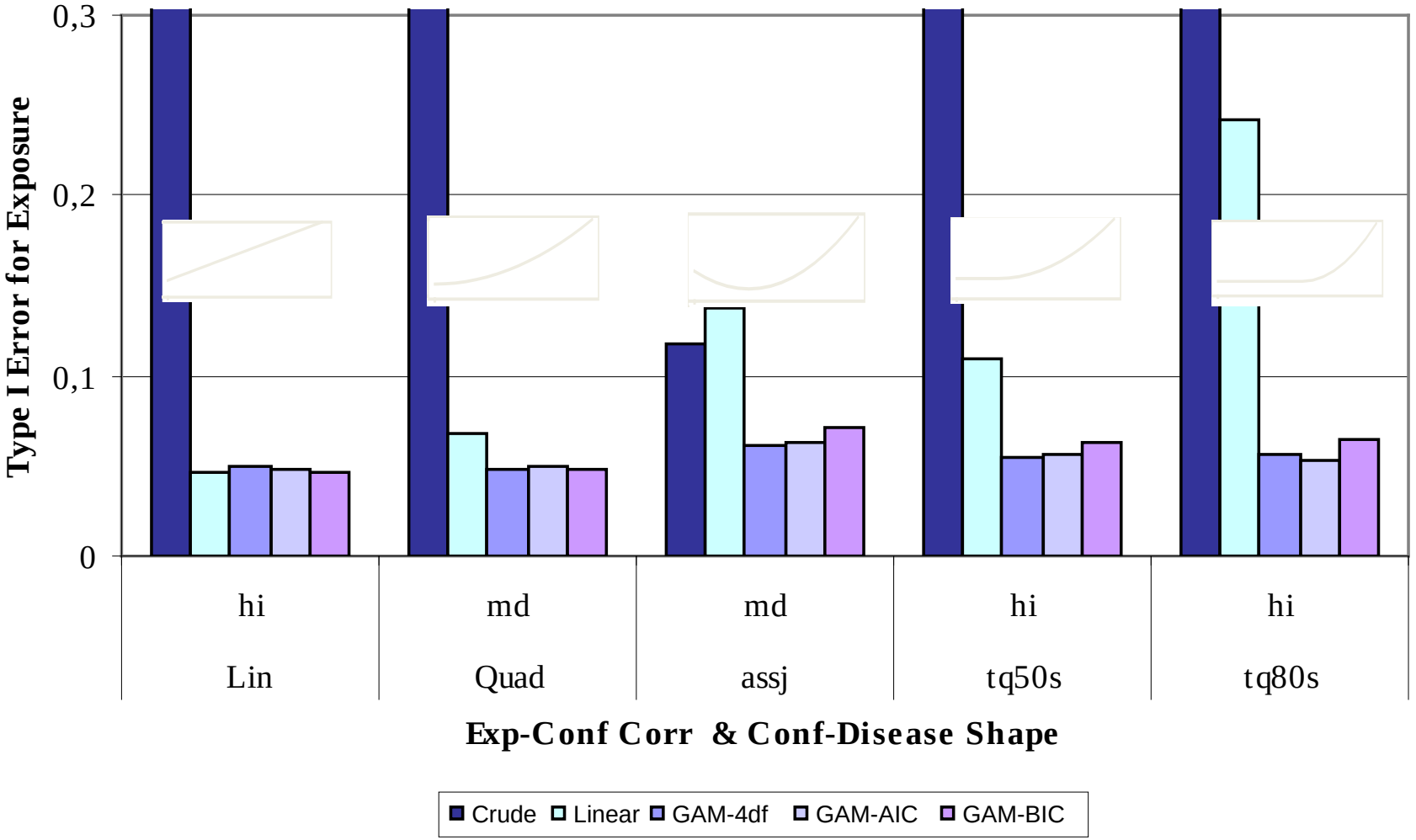


- <u>If not properly controlled for:</u>
  - Residual Confounding
    - "leftover" confounding
    - Biased estimate of the main effect

- 2 possible sources of residual confounding:
  - measurement error
  - mis-modeling of continuous variables

Significance of β_{exp} by Analysis Strategy

N=1000, Dichot Exp, β_{exp}=0



Residual Confounding Example Maternal Smoking and Down Syndrome

- Early investigations showed a **protective** effect.

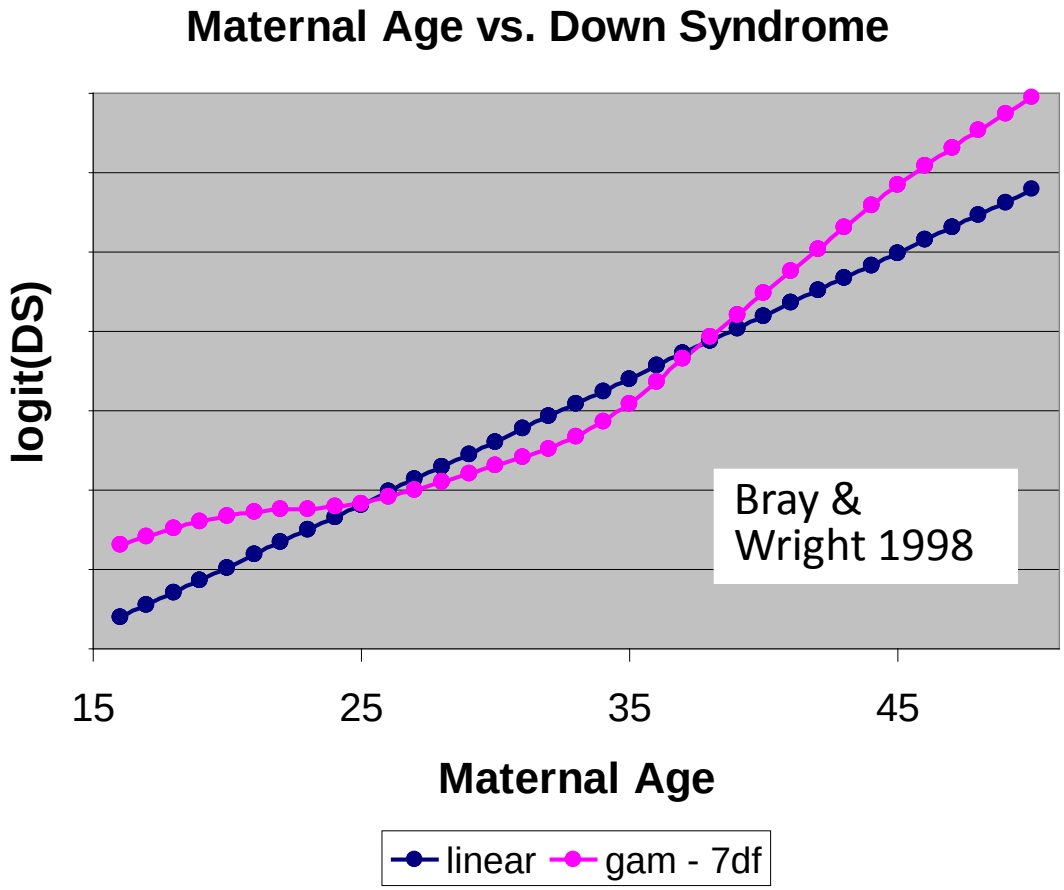
- residual confounding from maternal age?
  - Down syndrome births increase with maternal age.
  - Smoking is more common among younger women.

- Chen et al. looked at smoking and Down syndrome and controlled for maternal age in three ways [Chen et al. 1999]:
  - **Crude association:**
    smoking was statistically significantly protective (OR=0.8, 95% CI: 0.65-0.98)
  - **Adjusting for dichotomous maternal age (<35 years, >35 years) :**
    protective but not stat. sig. (OR=0.87, 95% CI: 0.71-1.07)
  - **Adjusting for continuous maternal age as a linear effect:**
    no effect at all (OR=1.0, 95% CI: 0.82-1.24)

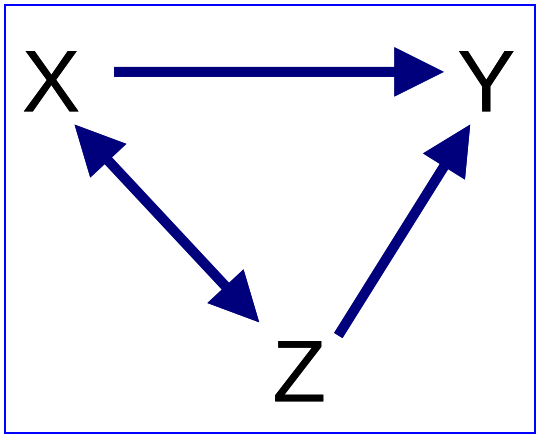
→ if maternal age is not properly adjusted for, it appears that smoking is protective

Maternal smoking & Down Syndrome … continued

- Actually, Down Syndrome risk is probably not linearly associated with maternal age



Modeling Continuous X Non-Parametrically?



If X is continuous:

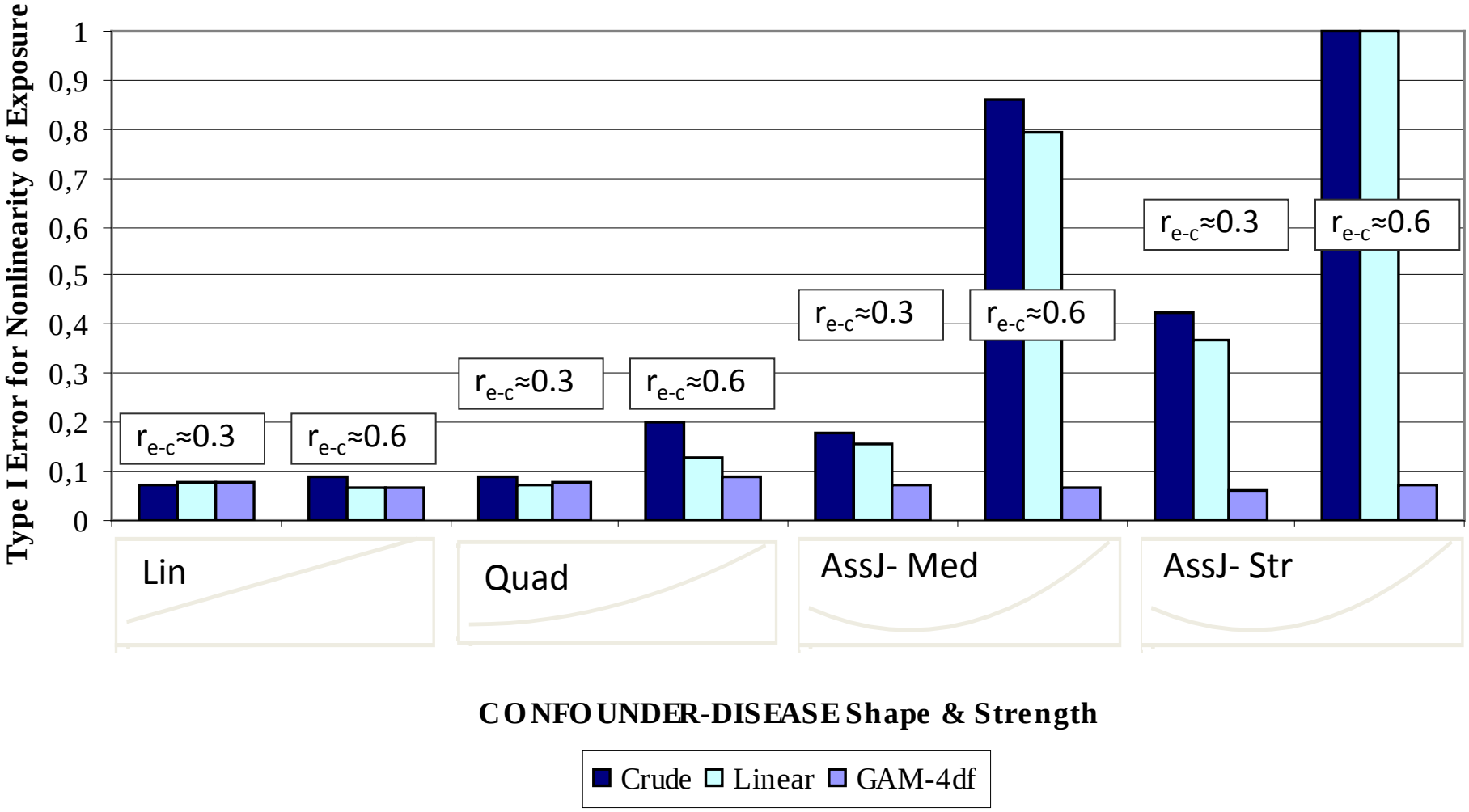
logit(Y)=s(X,4)

logit(Y)=s(X,4)+Z

logit(Y)=s(X,4)+s(Z,4)

Modeling the exposure non-parametrically (n=1000)

Exposure has a linear effect on Disease, df=4



Plans for Future TG2 activities

- To be Filled >>

Suggestions are Welcome

(Selected) literature

- Abrahamowicz M, Berger Rd, Grover SA. Flexible Modeling of the Effects of Cholesterol on Coronary Heart Mortality. *Am J Epi (AJE)* 1997; 145: 714-729.
- Abrahamowicz M, MacKenzie TA. Joint estimation of time-dependent and non-linear effects of continuous covariates on survival. *Stat Med* 2007; 26: 392-408.
- Binder H, Sauerbrei W, Royston P. Comparison between splines and fractional polynomials for multivariable model building with continuous covariates. *Stat Med* 2013; 32: 2262-2277.
- Benedetti A, Abrahamowicz M. Using generalized additive models to reduce residual confounding. *Stat Med* 2004; 23: 3781-3801.
- Gagnon B, Abrahamowicz M, Xiao Y, et al. Flexible modeling improves prognostic value of C-reactive protein. *Br J Cancer* 2010; 102: 1113-1122.
- Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*, 2001.
- Hastie T., Tibshirani R. *Generalized Additive Models.* Chapman & Hall: NY, 1990.
- Miller A. *Subset Selection in Regression.* Taylor & Francis, 2002.
- Ramsay JO. Monotone Regression Splines. *Stat Sciences* 1988; 3: 425-441.
- Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epi (IJE)* 1999; 28: 964-974.
- Royston P, Sauerbrei W. *Multivariable Model Building: A Pragmatic Approach to Regression Anaylsis Based on Fractional Polynomials for Modelling Continuous Variables.* Wiley, 2008.
- Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine* 2006; **25**: 127-141.
- Sauerbrei W, Royston P, Binder H. Selection of important variables and functional form for continuous predictors in multivariable model building. *Stat Med* 2007; 26: 5512-5528.
- Sauerbrei W, Royston P, Bojar H, et al. Modeling the effects of standard prognostic factors in node-positive breast cancer. *Br J Cancer* 1999; 79: 1752-1760.
- Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating.* Springer, 2009.
- Wynant W, Abrahamowicz M. Impact of the model building strategy on the inference about time-dependent & non-linear covariate effects in survival analysis. *Stat Med* 2014;33:3318-37

THANK YOU

- Michal.Abrahamowicz@McGill.ca

for STRATOS TG2

Inter-Dependence of the Selections of (1) Variables vs (2) Functional Forms

- **The CHALLENGE is that the results of Data-dependent selections of (1) 'significant'/relevant Predictors may depend on (2) choices regarding Functional Forms of both, (2a) the Predictor of Interest (X) & (2b) Other Variables, correlated with X, and *vice versa***

[Rosenberg PS, Katki H, Swanson CA, Stat Med 2003, 22: 3369-3381]

Examples of Inter-dependence:

(1) Impact of Inaccurate Modeling on Variable Selection: Incorrect Linearity Assumption increases Type II error for testing the (truly NL) effect of X, resulting in its un-warranted exclusion

[e.g. Abrahamowicz et al 1997; Gagnon et al Br J Cancer 2010]

Impact of **Residual Confounding** (due to Incorrect Modeling of Confounders):

- **Further Examples of Inter-dependence:**

> (2) Failure to adjust for Important Confounders and their NL effects, increases either Type I or Type II error for testing:

- (2a) Linearity of the effect of a continuous X [Binder et al 2013];

- (2b) Association between a binary Z and the outcome [Benedetti & Abrahamowicz 2004] ;

> (3) in Survival analyses, a failure to account for NL effect of X increases type I error for a Time-dependent effect of X [Abrahamowicz & MacKenzie 2007]

Functional Forms for Modeling Continuous Independent Variables

Representation of Continuous Variables (NOT Mutually Exclusive)	EPIDEMIOLOGY journals (% / 25)	CLINICAL journals (% / 25)
Dichotomized	**7 (28%)**	2 (8%)
Categorized (> 2 categories)	**19 (76%)**	**14 (56%)**
Continuous, Un-transformed (**Linear** effect assumed *a priori*)	**19 (76%)**	**22 (88%)**
A priori defined **Parametric Transformation(s)** (e.g. log or polynomial)	**8 (32%)**	1 (4%)
Restricted Cubic Splines	1 (4%)	-
Other **Spline-based** methods	-	-
Fractional Polynomials	-	-

Distribution of the Sampled (Eligible) Papers across the pre-Selected Journals >>SEHAR plse Copy the Text below to Previous slide

- CLINICAL:

				JNCI	NEJM
37%	7%		0		

Types of Regression Models

Multivariable Model	EPIDEMIOLOGY journals	CLINICAL journals
LOGISTIC		
COX PH		
LINEAR		
POISSON		
OTHERS		

Types of Regression Models

I've modified slide 11 (large table) to look like slide 9 or 10. If you need to replace any table, click in the top left or right corner so that a 'frame' surrounds the table and press delete.

4 slides of figures and tables have been added at the end.

Multivariable Model	EPIDEMIOLOGY journals	CLINICAL journals
LOGISTIC		
COX PH		
LINEAR		
POISSON		
OTHERS		

Enter text

Types of Regression Models

Multivariable Model	EPIDEMIOLOGY journals	CLINICAL journals		
Logistic	12			
Cox PH	7			
Linear	9			
Poisson	1			
Polytomous logistic	2			
Inverse Gaussian (log-linear)	1			
GEE (linear)	1			
TOTAL	**33**			