Validation of a new flexible spline-based relative survival model

Mahboubi A[1,2], Abrahamowicz M[2], Binquet C[1], Giorgi R[3], Bonithon-Kopp C[1], Quantin C[1]

[1] Inserm, U866, Dijon, F-21000, France; Univ de Bourgogne, Dijon, F-21000, France.
[2] Department of Epidemiology and Biostatistics, McGill University, Montreal, Que, Canada H3A 1A2
[3] Université de la méditerranée, Faculté de Médecine, LERTIM, Marseille, F-13000, France.

OUTLINE

- Background
  - Role of relative survival in prognostic studies
  - Limitations of published models
  - Importance of joint modelling of non-linear and time-dependent effects
- New model
  - Validation of our model
    - ✓ Design of simulations
    - ✓ Results
- Conclusions

BACKGROUND

Relative survival

- Useful methodology in situations where the cause of death is unknown, especially in:
  - in population-based cancer registries
  - for rare diseases with long survival
- Permits estimation of the covariate effects on disease-specific mortality

Dickman et al, *Stat-in-Med* 2004;23: 51-64

BACKGROUND

the Estève's et al relative survival model:

- Assumes that the observed hazard for total mortality λ_t, is the sum of two hazards:

$$\lambda_t(t|z,a) = \lambda_e(t+a|z_s) + \lambda_c(t|z)$$

  - λ_e represents the <u>expected hazard function for all-cause mortality</u> in the population at large
  - λ_c is the excess hazard due to disease-specific mortality

- Requires estimation of baseline hazard for disease specific mortality

Estève et al, *Stat-in-Med* 1990;**9**: 529-38

BACKGROUND

- In Estève's et al relative survival model:
  - Disease-related hazard is *a priori* constrained to meet 2 conventional assumptions: $\lambda_c = \sum_k \tau_k I_k(t) * \exp(\beta Z)$
    1. Proportional hazards (PH) → relative risks are constant over time
    2. Log-linearity for continuous covariates (LL) → the logarithm of the hazard increases linearly with increasing value of the covariate

- Several "Crude" and Relative survival models were proposed to relax one or both of these assumptions:
  - Abrahamowicz and MacKenzie (2007)
  - Giorgi et al (2003)
  - Remontet et al (2007)
  - Lambert (2006)

BACKGROUND

- Recently, Abrahamowicz & MacKenzie (2007) demonstrated, in crude survival, the importance of modelling and testing both PH and LL hypotheses jointly in order to :
  - ✓ avoid a spurious non-proportional effects due to not accounting for loglinearity (and vice versa)
  - ✓ avoid residual confounding
  - ✓ increase power for identifying new prognostic factors

Abrahamowicz M, MacKenzie TA. *Stat-in-Med* 2007;**26**: 392-408.

OUR JOINT RELATIVE SURVIVAL MODEL

- We propose a flexible extension of the Estève's relative survival model

$$\lambda_t = \lambda_e + \lambda_c$$

- The effect of covariates on λ_c are modelled with quadratic B-spline functions, which relax PH and LL assumptions

OUR JOINT RELATIVE SURVIVAL MODEL

- Therefore, in our model

$$\lambda_c(t|z) = \exp(\gamma(t)) * \exp(\sum_i \alpha_i(z_i) * \beta_i(t))$$

  - γ(t) represents the baseline hazard
  - α(z) describes how the hazard changes with increasing value of a continuous covariate
  - β(t) represents the pattern of changes over time in the impact of the covariate on the risk of death

OUR JOINT RELATIVE SURVIVAL MODEL

- A 3-step iterative alternative conditional full MLE estimation

1. Baseline hazard: $\gamma(t) = \exp(\Sigma_{k=-2\ldots2}\ \gamma_k B_{k,3}(t))$

2. Non-linear dose response: $\alpha(z) = \Sigma_{m=-2\ldots1}\ \alpha_{im} B_{m,3}(z)$

3. Time-dependent effect: $\beta(t) = \Sigma_{l=-2\ldots2}\ \beta_{il} B_{l,3}(t)$

GENERAL DESIGN OF SIMULATIONS

- We simulated a prognostic study of colon cancer mortality:
- N = 2000 subjects
- Major steps in data generation:

1. Covariate vectors
2. Three times (for each subject):

    (i) Cancer-related death (conditional on covariates)

    (ii) Other-cause death (conditional on age & sex)

    (iii) Right “administrative” censoring (at 5 yrs)

3. Final “Observed” time = min {(i), (ii),(iii)}

COVARIATES GENERATION

- 5 covariates simulated:

1. Sex ~ Bin(1,0.5)
2. Age:
   Age ~ N(69.8,11.1) for men
   Age ~ N(71.4,12.6) for women
3. Year of diagnosis ~ U(1976,2000)
4. Stage: stage II vs stage III
   Stage II ~ Bin (1,0.61)
5. Location: right colon vs left colon
   Left colon ~ Bin(1,0.62)

GENERATION OF TIME TO CANCER-RELATED DEATHS

Using Permutational Algorithm*

- (a) Generate N observed times
  Calculate the expected number of deaths related to cancer per month (independent of covariates)

- (b) Match N covariate vectors with N times generated in step (a) based on the probability derived from the partial likelihood

- Specifically, we assume the following model for hazard of cancer-related death while assuming non-linear and time-dependent effects of age on the hazard of cancer-related mortality:

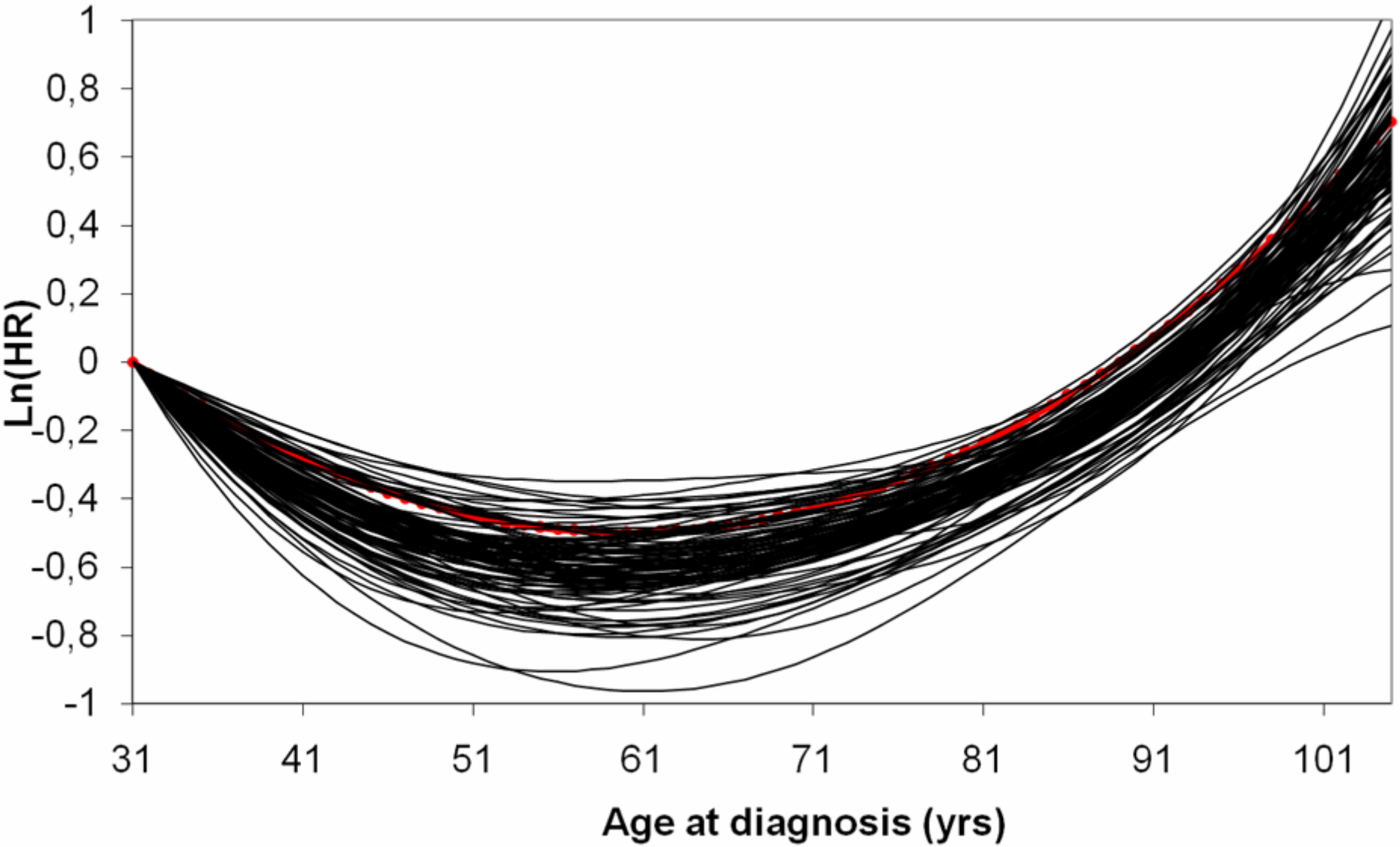
$$HR = \exp\begin{pmatrix} \log(1.25) * sex + \log(1.2) * location + \log(3) * stage + \\ \alpha_{age}(age) * \beta_{age}(time) \end{pmatrix}$$

(*) Sylvestre & Abrahamowicz, Stat-in-Med, 2008, *27(14): 2618-2634*

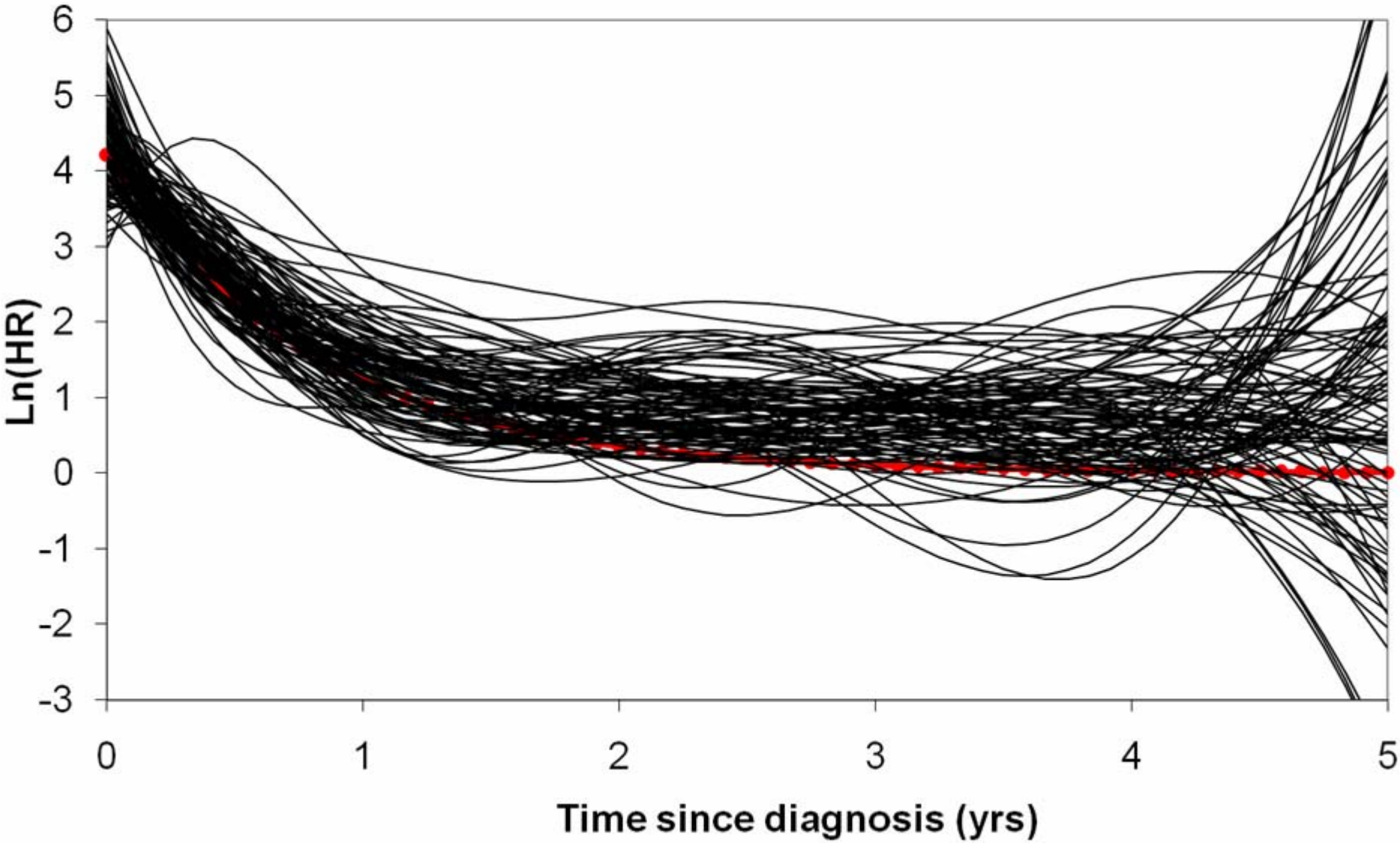
GENERATION OF "NATURAL" (OTHER-CAUSE) DEATHS

- Time to natural death was generated using the French lifetables, stratified by Sex, Age and Calendar Year

NON LOG-LINEAR EFFECTS OF AGE



TIME DEPENDENT EFFECTS OF AGE



CONCLUSION

- We have implemented a new flexible relative survival model
- This model allows for both simultaneous estimation and testing of time-dependent effects and non-log-linear effects for continuous covariates
- The simulations show
  - ✓ The general shape of both the time dependent and non log linear effects is recovered but there is over-fit bias (especially for TD effects)
  - ✓ The LRT tests show that our model is able to detect a true time dependent effect or a non loglinear effect

BIBLIOGRAPHY

1. Dickman PW, Sloggett A, Hills M, Hakulinen T. Regression models for relative survival. *Statistics in Medicine* 2004;23: 51-64
2. Esteve J, Benhamou E, Croasdale M, Raymond L. Relative survival and the estimation of net survival: elements for further discussion. *Statistics in Medicine* 1990;**9**: 529-38.
3. Giorgi R, Abrahamowicz M, Quantin C, Bolard P, Esteve J, Gouvernet J, Faivre J. A relative survival regression model using B-spline functions to model non-proportional hazards. *Statistics in Medicine* 2003;**22**: 2767-84.
4. Remontet L, Bossard N, Belot A, Esteve J. An overall strategy based on regression models to estimate relative survival and model the effects of prognostic factors in cancer survival studies. *Statistics in Medicine* 2007;**26**: 2214-28.
5. Le Teuff G, Abrahamowicz M, Bolard P, Quantin C. Comparison of Cox's and relative survival models when estimating the effects of prognostic factors on disease-specific mortality: a simulation study under proportional excess hazards. *Statistics in Medicine* 2005;**24**: 3887-909.
6. Quantin C, Abrahamowicz M, Moreau T, Bartlett G, MacKenzie T, Tazi MA, Lalonde L, Faivre J. Variation over time of the effects of prognostic factors in a population-based study of colon cancer: comparison of statistical models. American Journal of Epidemiology 1999;150: 1188-200.
7. Abrahamowicz M, MacKenzie TA. Joint estimation of time-dependent and non-linear effects of continuous covariates on survival. *Statistics in Medicine* 2007;**26**: 392-408.
8. Lambert PC, Smith LK, Jones DR, Botha JL. Additive and multiplicative covariate regression models for relative survival incorporating fractional polynomials for time-dependent effects. Statistics in Medicine 2005; 24: 3871-85.
9. Ramsay JO, Abrahamowicz M. Binomial regression with monotone splines: A psychometric approach. *Journal of the American Statistical Association* 1989;**84**: 906-915.
10. Sylvestre M.P., Abrahamowicz M. Comparison of algorithms to generate event times conditional on time-dependent covariates. *Statistics in Medicine* 2008;**27**: 2618-2634