GRELL & ENCR

2025 Scientific Meeting
Porto, Portugal

27-30
MAY

GRELL

ENCR
European Network
of Cancer Registries

Detailed trend analyses and up-to-date estimates of cancer incidence and mortality using multidimensional penalized splines : the French example

Z Uhry[1,2], E Dantony [2,3], Fauvernier M[2,3], F Molinié [4,5], E Chatignoux[1], L Remontet[2,3]
and the Francim network

[1] Santé Publique France, [2] Service de Biostatistique des Hospices Civils de Lyon (HCL), [3] UMR CNRS 5558-LBBE, Université de Lyon, [4] Registre des cancers de Loire-Atlantique et Vendée, [5] Réseau des registres de cancers français Francim

Francim – HCL – Santé Publique France – Institut national du cancer

INTRODUCTION AND OBJECTIVES

National cancer incidence and mortality studies in France – Context

- Partnership :

  Francim cancer registries, HCL (biostat.), Santé Publique France, Institut national du cancer

- Study updated every 5 years, with 73 cancer sites or sub-sites analysed

  (with an *in-between* update for 20 main cancer sites)

- National cancer incidence has to be estimated (cancer registries cover 20% of the population):

  => Not the subject today

  => Method will be presented on mortality, to keep focused on how we model trends

These studies have a double objective

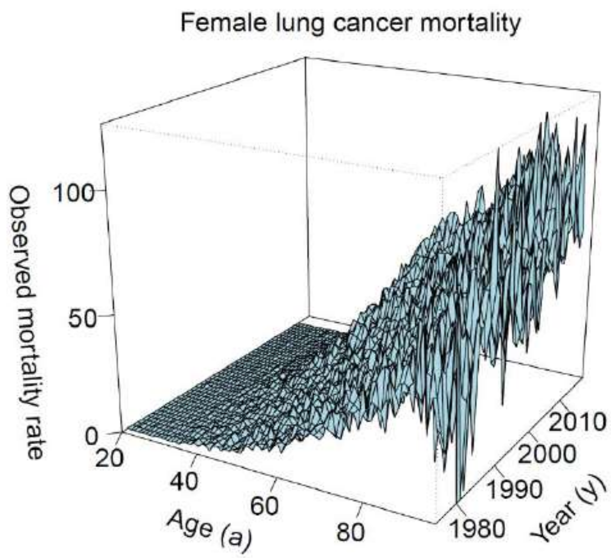
i) **Provide detailed trend analyses**, according to year or birth cohort

ii) **Provide up-to-date incidence and mortality estimate**

$\Rightarrow$ **Need for short-term projections** (3 years in the last main study: 2015 → 2018)

What is our view on this problematic ?

Observed rates



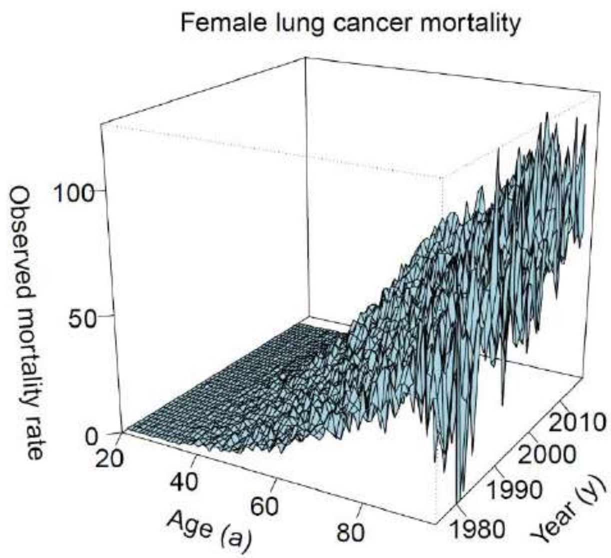
=> Rates: points over the lexis diagram, with random variability

AGE	YEAR	COHORT	DEATH	PY	RATE
60	2000	1940	62	272485.0	22.754
61	2000	1939	70	280991.0	24.912
62	2000	1938	74	280194.0	26.410
63	2000	1937	69	282511.5	24.424

...

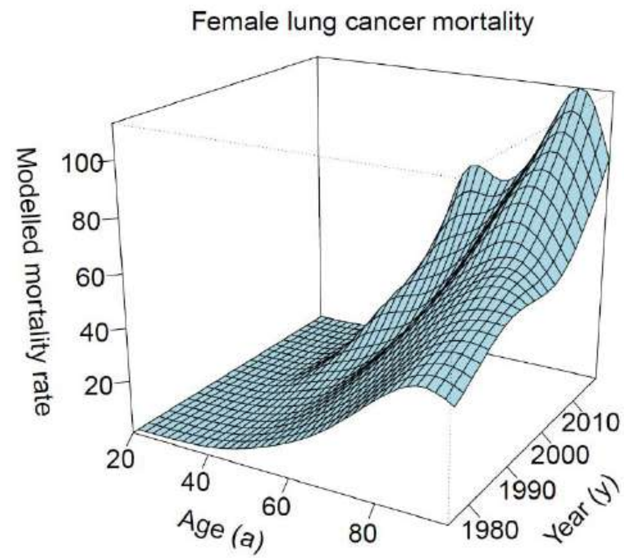
What is our view on this problematic ?

Observed rates



Model

Smooth rates (modelled)



=> Rates: points over the lexis diagram, with random variability

AGE	YEAR	COHORT	DEATH	PY	RATE
60	2000	1940	62	272485.0	22.754
61	2000	1939	70	280991.0	24.912
62	2000	1938	74	280194.0	26.410
63	2000	1937	69	282511.5	24.424

…

- **Smooth surface, modelled by a function of age and year**
- **In particular, trends may vary smoothly with ages…**
- **Projections made by extending the surface over time..**

⇒ **All indicators (and CIs) retrieved from these modelled rates**

⇒ **Rates can be described as well according to birth cohort**

METHOD

Illustrated on mortality

Data

- National mortality data 1975-2015 (in the last main study)
- Aggregated **by annual age and annual year**

Model

- **Flexible** Poisson regression model, using **multidimensional penalized splines (MPS)** as introduced by Wood (Wood 2016 & 2017)

$$D_{a,y} \sim Poisson\,(\,\mu_{a,y}\,.\,PY_{a,y}\,)$$

, *D:* number of deaths, PY: person-years,

a : age at death and y : year of death

$$Log(\mu_{a,y}) = MPS(a,y),$$

$\mu_{a,y}$: mortality rate for age *a* and year y

MPS(a,y) : general function of age and year

where parameters of this spline are estimated by maximizing a **penalized likelihood**

Uhry IJE 2020

- R software, function *gam* of the package *mgcv*

We'll see now …

$\Rightarrow$ **What is a regression spline ?**

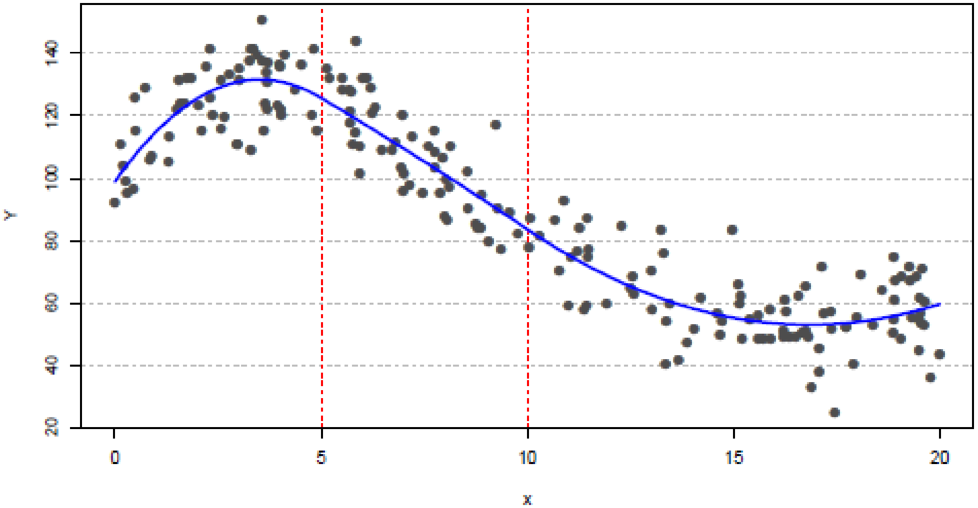
$\Rightarrow$ **Which type of splines is suitable for short-term projections ?**

$\Rightarrow$ **What is a multidimensional spline ?**

$\Rightarrow$ **What is the role of penalization in MPS ?**

1 – What is a regression spline ?

- Splines are piecewise polynomial - usually **cubic** (degree 3); the junction points are called **knots**



$\Rightarrow$ Flexible functions

$\Rightarrow$ Key advantages over polynomials :

- Local regression (no influence of distant point)
- For more flexibility, we add knots (no need to increase the degree of the polynomial)

$\Rightarrow$ Example: a **cubic spline with one knot** placed at year k has **5 parameters :**

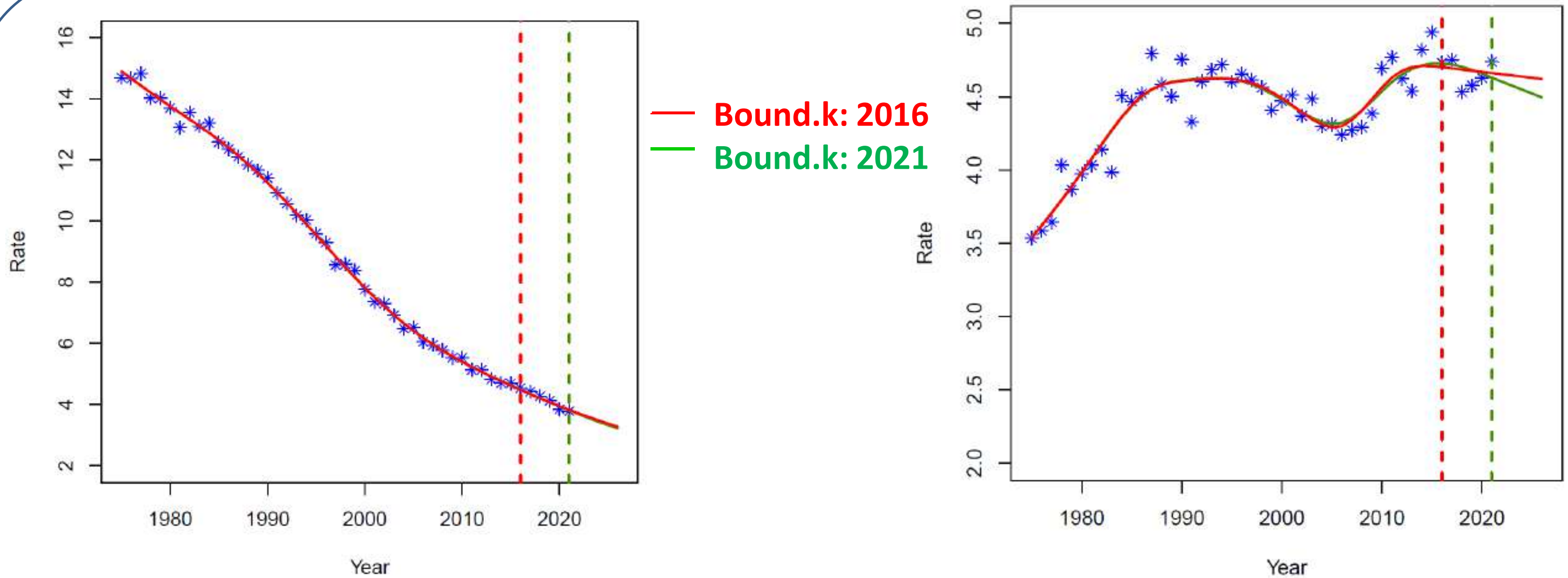
$$f(y) = \beta_0 + \beta_1 y + \beta_2 . y^2 + \beta_3 y^3 + \beta_4 \, I(y \geq k) . (y-k)^3$$

2 – Which type of splines is suitable for short-term projections ?

- **Restricted splines** (also *called natural splines*) are constrained to be **linear beyond their last knot** (called boundary knot)

- **This is an interesting feature for projections :**
    - Projection is linear (here, on the log-scale)
    - Location of the boundary knot define the number of years on which the slope is estimated:
      => user choice

Estimates according to the boundary knot position



=> Similar fit and similar projection

=> Close fit, but projections differ …

3 – What is a multidimensional spline ? (here, bidimensional)

- Simplified example:

  - Consider a linear effect of year (basis for year *h*) :

$$h(y) = \alpha_0 + \alpha_1 y$$

3 – What is a multidimensional spline ? (here, bidimensional)

- Simplified example:

  - Consider a linear effect of year (basis for year *h*) :

$$h(y) = \alpha_0 + \alpha_1 y$$

  - Let now the intercept and slope vary with a quadratic effect of age (basis for age *g*)

$$g(a) = \gamma_0 + \gamma_1 a + \gamma_2 a^2$$

3 – What is a multidimensional spline ? (here, bidimensional)

- Simplified example:

  - Consider a linear effect of year (basis for year *h*) :

$$h(y) = \alpha_0 + \alpha_1 y$$

  - Let now the intercept and slope vary with a quadratic effect of age (basis for age *g*)

$$g(a) = \gamma_0 + \gamma_1 a + \gamma_2 a^2$$

  - The bidimensional effect of age and year is build as the **tensor product** $h(y) \otimes g(a)$ (term-by-term multiplication of h and g bases):

$$MS(a, y) = h(y) \otimes g(a) = \underbrace{(\beta_0 + \beta_1 a + \beta_2 a^2)}_{\alpha_0(a)} + \underbrace{(\beta_3 + \beta_4 a + \beta_5 a^2)}_{\alpha_1(a)}\, y$$

3 – What is a multidimensional spline ? (here, bidimensional)

- Simplified example:

  - Consider a linear effect of year (basis for year *h*) :

  $$h(y) = \alpha_0 + \alpha_1 y$$

  - Let now the intercept and slope vary with a quadratic effect of age (basis for age *g*)

  $$g(a) = \gamma_0 + \gamma_1 a + \gamma_2 a^2$$

  - The bidimensional effect of age and year is build as the **tensor product** $h(y) \otimes g(a)$ (term-by-term multiplication of h and g bases):

  $$MS(a,y) = h(y) \otimes g(a) = \underbrace{(\beta_0 + \beta_1 a + \beta_2 a^2\ )}_{\alpha_0(a)} + \underbrace{(\beta_3 + \beta_4 a + \beta_5 a^2\ )}_{\alpha_1(a)}\ y$$

⇒ **MS accounts for age-year interaction**

⇒ It is a ***varying coefficient model*** (Hastie,1993): flexible but structured model

3 – What is a multidimensional spline ? *Continued*

- In practice, **more complex bases are used** : restricted cubic splines

- **Advantages:** MS are highly flexible, they can model any pattern …

- **Limits:** Many parameters => risk of overfitting (estimates may be erratic)

=> How to benefit from **MS flexibility while avoiding overfitting** ? **Penalization !**

e.g. 7*10=70 parameters if marginal basis of size 10 for age and 7 for year

4 – What is the role of penalization in MPS?

Illustration for a unidimensional spline $f(y)$:

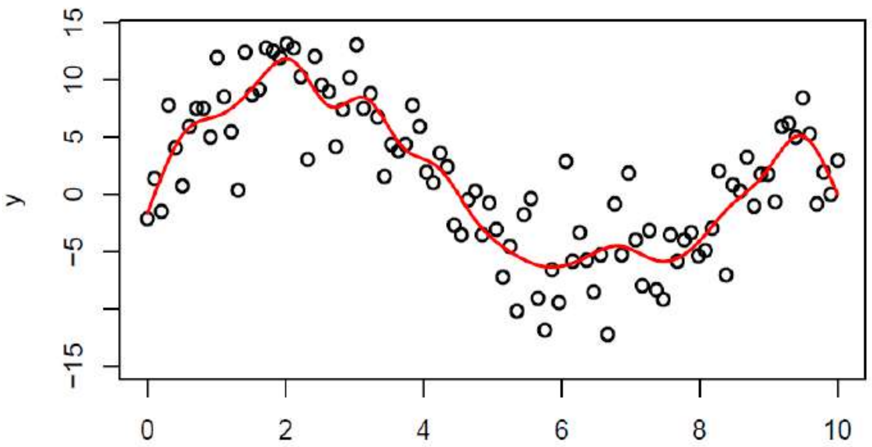
Penalize the likelihood to find a **compromise** between **fit and smoothness** of the estimates :

$$\mathcal{L}_p(\beta|\lambda) = \mathcal{L}(\beta) - \lambda \int f''(u)^2 du$$

- ✓ This compromise is controlled by **smoothing parameter**, noted **λ**
- ✓ The smoothing parameter is **estimated** in order to minimize prediction error

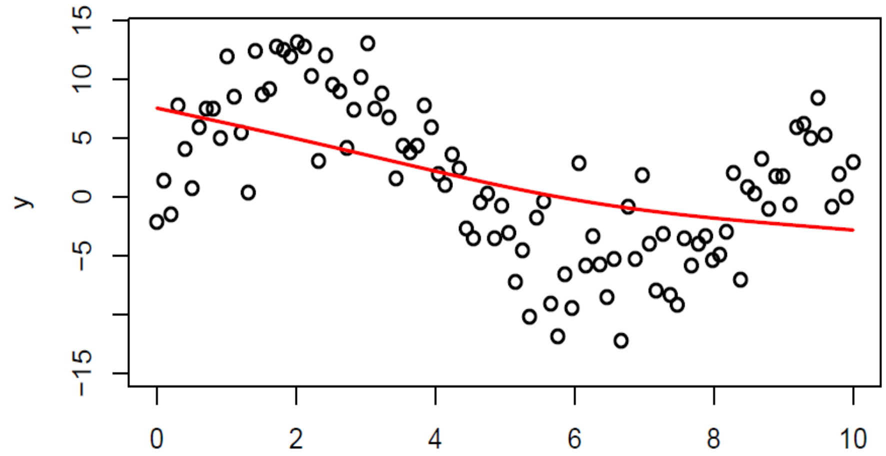
Penalization makes a compromise between fit and smoothness

(a) No penalization (λ=0)



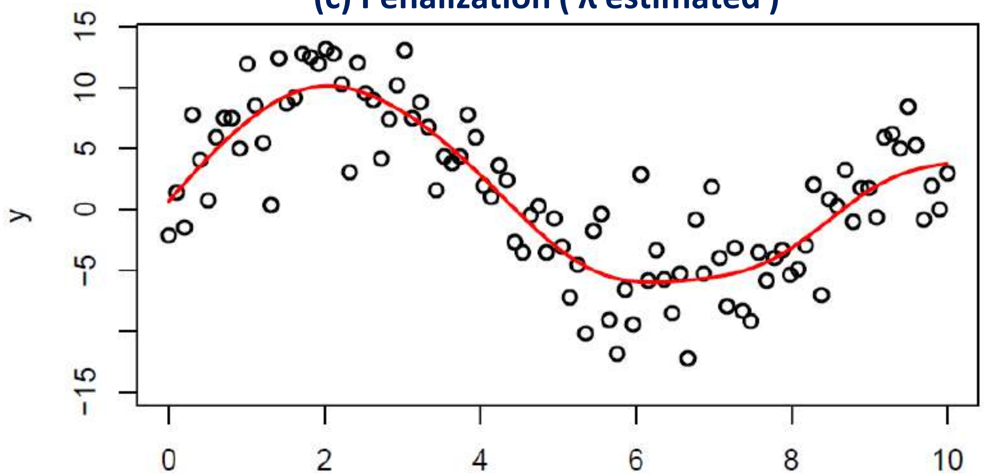
⇒ Overfitting (small insignificant variations)

(b) Too strong penalization (λ force high)



⇒ Simple effect but poor fit

(c) Penalization (λ estimated)

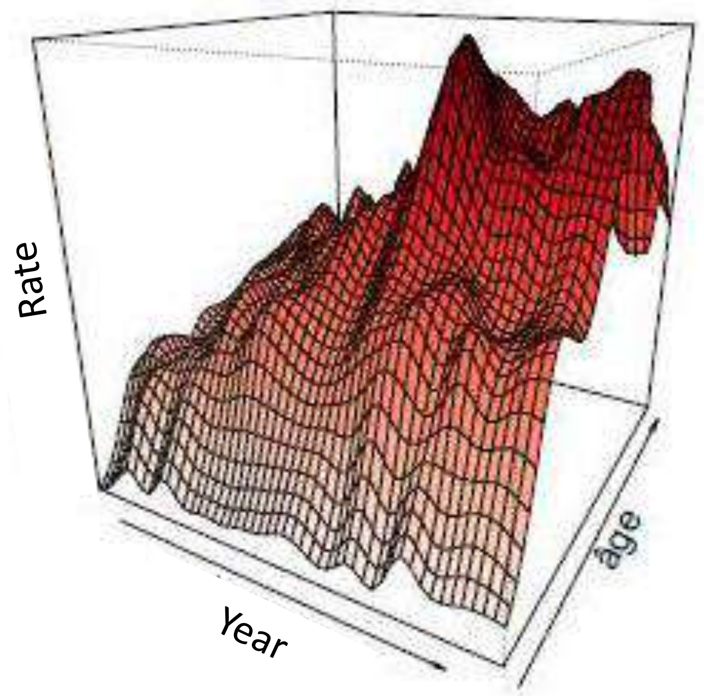


⇒ **Compromise !**

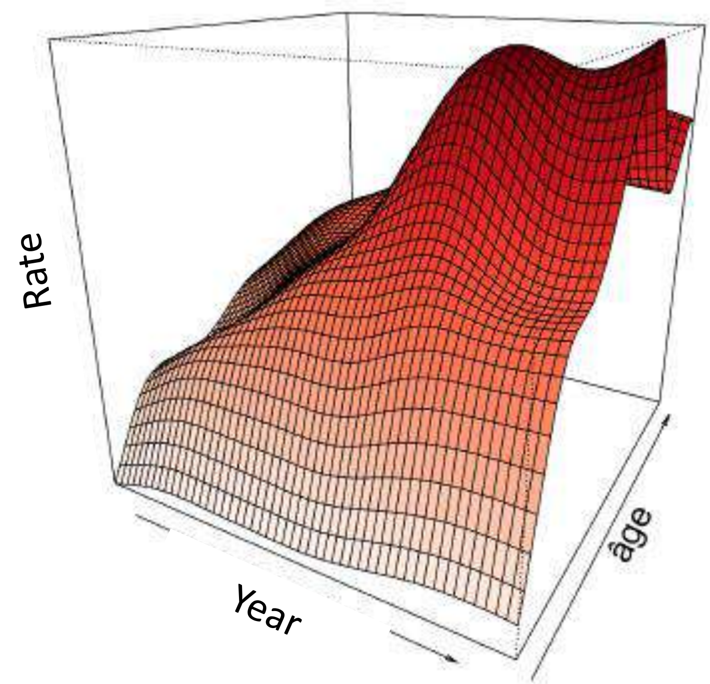
Penalization for bi-dimensional spline

2 smoothing parameters (λ_{age} and λ_{year}), one for each direction

(a) No penalization



(b) Penalization (λ_{age} and λ_{year} estimated)



Penalization **acts like a selection procedure**

=> it adapts the model complexity to the information in the data

RESULTS: ILLUSTRATIONS

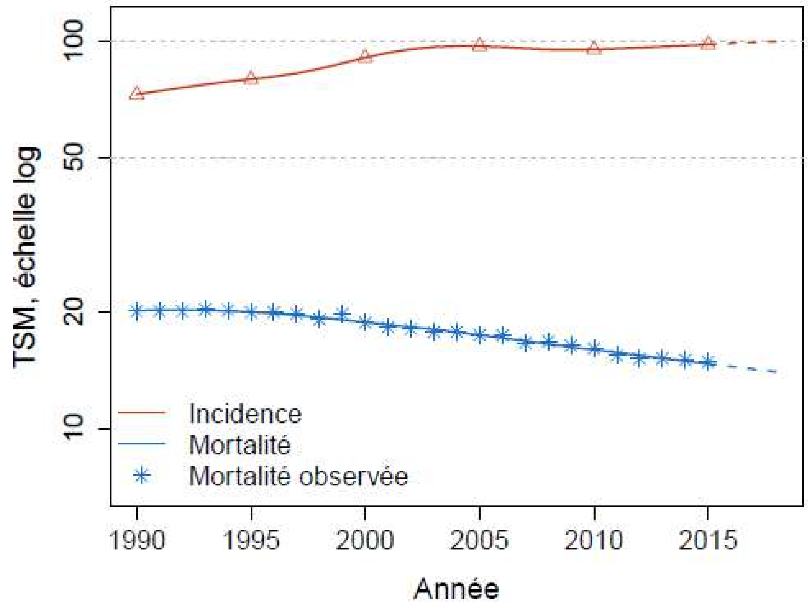
All indicators (and CIs) are retrieved from the modelled rates

(from parameters $\hat{\beta}$ and their variance-covariance matrix $\hat{V}_{\beta}$):

Age-specific rates, age-standardized rates (ASR), AAPC, cumulative risk 0-74 yrs, etc..

Source : Last main study (Defossez 2019)

- **Trends in age-standardized rates (ASR)**

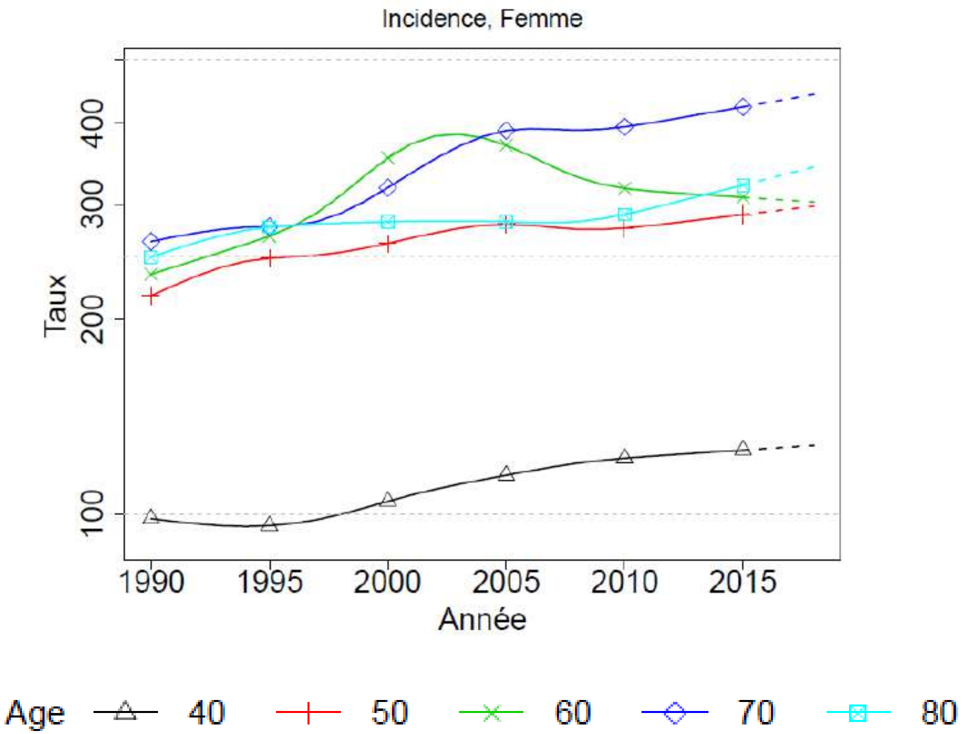


	Année							Variation Annuelle Moyenne(%)	
	1990	1995	2000	2005	2010	2015	2018	De 1990 à 2018	De 2010 à 2018
Incidence	72.8	79.8	90.7	97	95.2	98	99.9	1.1 [1 ;1.2]	0.6 [0.3 ;0.9]
Mortalité	20.2	20	18.8	17.4	16	14.7	14	-1.3 [-1.4 ;-1.2]	-1.6 [-1.8 ;-1.4]
Mortalité observée	*20.1*	*19.9*	*18.7*	*17.4*	*16.1*	*14.8*	-	-	-

=> ASR and their AAPC may be derived from the model

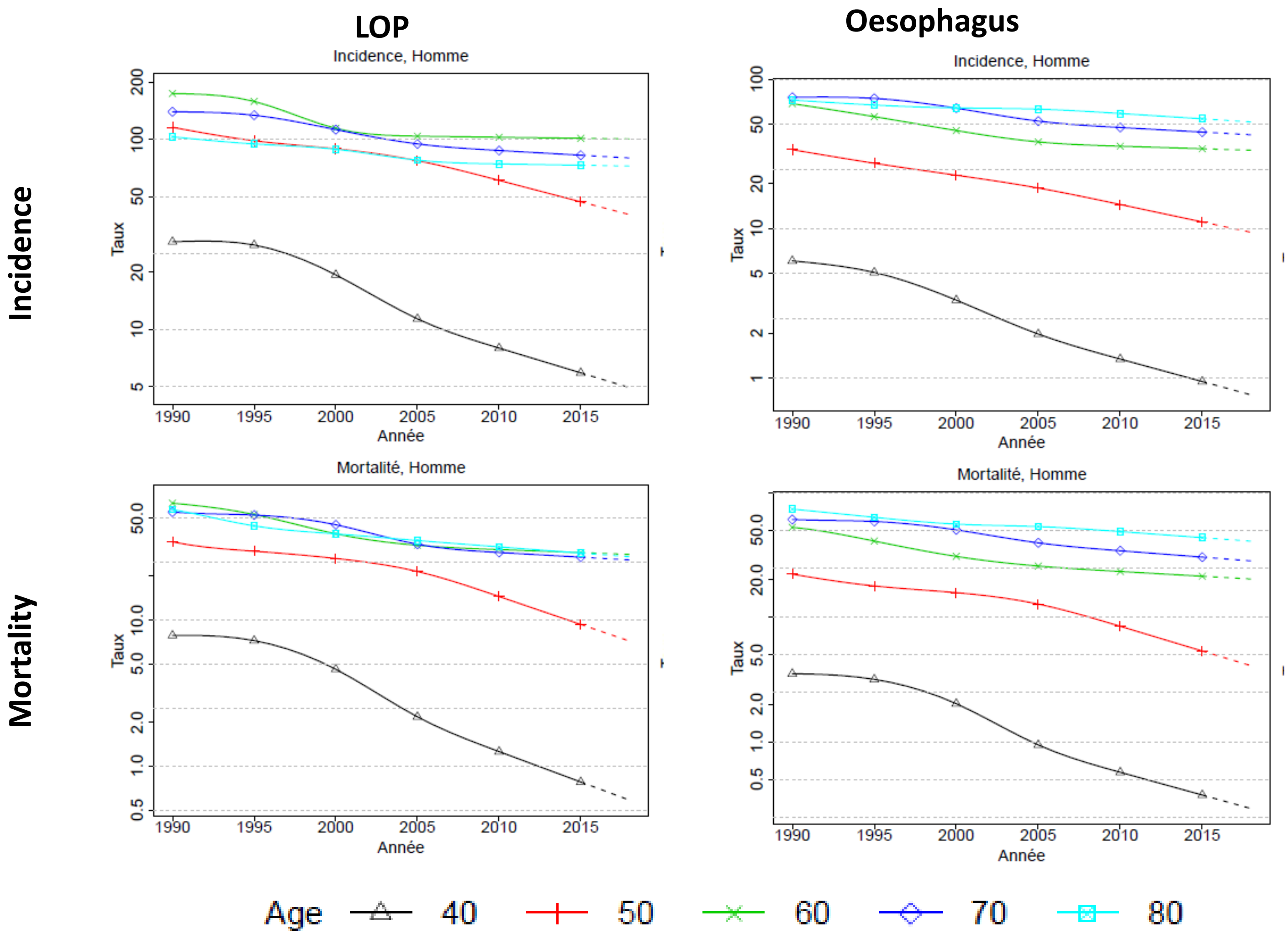
Breast cancer, France, 1990-2018

- **Trends by age**



=> MPS allows to catch complex trends, e.g. trends at age 60 (green curve) very different form other ages

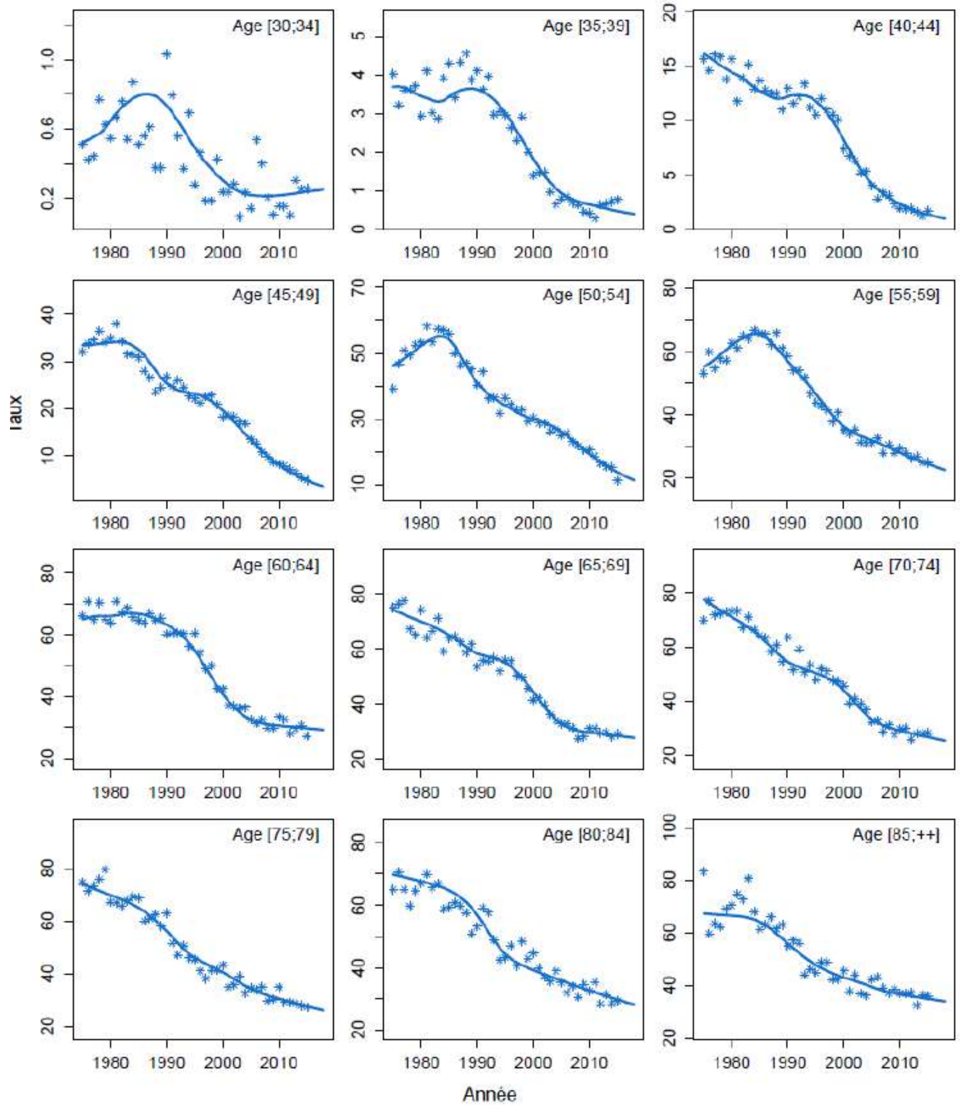
- **Trends by age**



=> Consistent pattern between incidence and mortality, and between cancer sites with common risk factors

- **Graphical fit assessment**

LOP mortality, men, France (1975-2018)



=> Fit assessment remains an important step

DISCUSSION

Discussion

- We use a unique MPS model for our two objectives: trend analyses and projection
- Key principle : MS brings flexibility while penalization avoids overfitting
- Advantages of MPS :

✓ **Age and year** treated as **continuous covariates** (using annual values)

- Avoids loss of information due to categorization
- Avoids stratification by age-class (or a model with one trend coef by age-class)
- Ensures consistency between trends of adjacent ages

✓ **MPS** may **model simple as complex effects** (non-linearity, interactions); in particular, **trends may varies smoothly with age**

✓ Using **restricted splines,** MPS allow to make **linear short-term projections** (here, on the log scale), with a slope dependant on age

- Reactivity to recent change depends on location of the boundary knot => user choice

Discussion

- **Projection**

- CIs do not account properly for uncertainty in projections

- Still, sensitivity analyses on location of the boundary knot may be carried out

- We fixed boundary knot 5 years before last year observed => “conservative” strategy

  (it avoids projecting uncertain recent inflexion in trends, but less reactive if inflexion is real...)

Discussion

- MPS were extended to hazard and **excess hazard models (net survival analyses)** and implemented in the R package **survPen** (Remontet SMMR 2019, Fauvernier 2019a, Fauvernier 2019b)

  $\Rightarrow$ We also use MPS for the French excess hazard and net survival cancer trends studies

  $\Rightarrow$ **Harmonized tool**

- MPS can have more covariates (e.g. age, year and deprivation index) and can be used for spatio-temporal models (Ugarte, 2012)

Conclusion

- **Penalized splines as proposed by Wood are a very interesting and mature tool**

  Wood made a breakthrough in the field of GAM, by proposing parametric penalized splines, together with estimation of the smoothing parameters based on clear criteria

- **Short-term projections are a palliative for up-to-date data… Some will be wrong !**

- **Collective and general demand for "reactivity"**

  … while carcinogenesis is a long process

  … while it takes time to "consolidate recent trends"

- **To summarize: in France, we carry out trend analyses and short-term projections at once,** in a way where : *"forecasting is a natural consequence of the smoothing process"* (Currie 2004), by extending the surface modelled over time

REFERENCES *- Method*

Uhry Z, Chatignoux E, Dantony E et al. Multidimensional penalized splines for incidence and mortality trend analyses and validation process for national cancer incidence trends estimated from local registry data. **Int Jour Epidemiol 2020;***49(4):1294-1306.*

Uhry Z, Belot A, Colonna M et al. National cancer incidence is estimated using the incidence/mortality ratio in countries with local incidence data: is this estimation correct? ***Cancer Epidemiol* 2013** ; 37(3):270-7.

Wood SN. Generalized additive models: an introduction with R. 2nd ed. London: Chapman & Hall/CRC, **2017.**

Wood SN, Pya N, Safken B. Smoothing Parameter and Model Selection for General Smooth Models. **JASA 2016**; 111(516):1548-75.

Hastie T and Tibshirani R. Varying-coefficient models. **J R Stat Soc Series B 1993**; 55: 757–796

Remontet, Uhry Z, Bossard N et al. Flexible and structured survival model for a simultaneous estimation of non-linear and non-proportional effects and complex interactions between continuous variables: Performance of this multidimensional penalized spline approach in net survival trend analysis. **Stat Methods Med Res** 2019;28:2368–84.

Fauvernier M, Roche L, Uhry Z, Tron L, Bossard N, Remontet . Multi-dimensional penalized hazard model with continuous covariates: applications for studying trends and social inequalities in cancer survival. **Journal of the Royal Statistical Society. Series C:** Applied Statistics **2019**a: 68(5): 1233-57.

Fauvernier M, Remontet L, Uhry Z, Bossard N, Roche L . survPen: an R package for hazard and excess hazard modelling with multidimensional penalized splines. **Journal of Open Source Software 2019**b; 4(40), 1434.

Ugarte MD, Goicoa T, Etxeberria J, Militino AF. Projections of cancer mortality risks using spatio-temporal P-spline models. **Stat Methods Med Res. 2012; 21(5**):545-60.

Currie ID, Durban M and Eilers PH. Smoothing and forecasting mortality rates. **Statistical Modelling. 2004**; 4: 279-98.

Clements MS, Armstrong BK, Moolgavkar SH. Lung cancer rate predictions using generalized additive models. **Biostatistics 2005**;6(4):576-89.

REFERENCES - *French cancer incidence and mortality trends studies*

Lapôtre-Ledoux B, Remontet L, Uhry Z, Dantony E, Grosclaude P, Molinié F, Woronoff A-S, Lecoffre-Bernard C, Lafay L, Defossez G, D'Almeida T. Incidence des principaux cancers en France métropolitaine en 2023 et tendances depuis 1990. *BEH,* **2023,** n°. 12-13, p. 188-204

Defossez G, Uhry Z, Delafosse P, ... Remontet L, Monnereau A; French Network of Cancer Registries (FRANCIM). Cancer incidence and mortality trends in France over 1990-2018 for solid tumors: the sex gap is narrowing. **BMC Cancer 2021** Jun 24;21(1):726.

Defossez G, Le Guyader-Peyrou S, Uhry Z, Grosclaude P, Remontet L, Colonna M, et al (2019). Estimations nationales de l'incidence et de la mortalité par cancer en France métropolitaine entre 1990 et 2018. Volume 1 – Tumeurs Solides. 372 p. Saint-Maurice (Fra) : Santé publique France *https://www.santepubliquefrance.fr/maladies-et-traumatismes/cancers/cancer-du-sein/documents/rapport-synthese/estimations-nationales-de-l-incidence-et-de-la-mortalite-par-cancer-en-france-metropolitaine-entre-1990-et-2018-volume-1-tumeurs-solides-etud*

Le Guyader-Peyrou S, Defossez G, Uhry Z, Mounier M, Cornet E, Dantony E, et al (2019). Estimations nationales de l'incidence et de la mortalité par cancer en France métropolitaine entre 1990 et 2018. Volume 2 – Hémopathies malignes. 169 p. Saint-Maurice (Fra) : Santé publique France *https://www.santepubliquefrance.fr/docs/estimations-nationales-de-l-incidence-et-de-la-mortalite-par-cancer-en-france-metropolitaine-entre-1990-et-2018-volume-2-hemopathies-malignes*