Bayesian approach to predicting cancer incidence for an area without cancer registration by using cancer incidence data from nearby areas

Ramon Clèries

1. Pla Director d'Oncologia de Cataunya – Departament de salut Generalitat de Catalunya

Institut Català d'Oncologia-IDIBELL, Hospital Duran i Reynals. Av. Gran Via de l'Hospitalet, 199-203 – 1ª planta 08908 – L'Hospitalet de Llobregat

2. Department de Ciències Clíniques – Campus de Bellvitge

Universitat de Barcelona

Pla Director d'Oncologia

Generalitat de Catalunya
Departament de Salut

Servei d'Estadística UAB

28 Maig 2014

Motivating Example

Example: Predicting colorectal cancer Incidence in Tarragona for the year 2000 using colorectal cancer incidence in Girona for the year 2000 **(USING INCIDENCE RATES)**

	Colorectal Cancer Incidence in Girona 2000			Colorectal Cancer Incidence in Tarragona 2000			
Age-Group	***Cases***	***PY***	***Incidence Rate***	***PY***	***Expected***	***Observed***	***Observed-Expected***
1	0	13109	0	14142			
2	0	13498	0	14243			
3	0	13760	0	15621			
4	0	16257	0	19138			
5	0	21081	0	24309			
6	1	21571	4.63585E-05	24511			
7	1	21670	4.61467E-05	23914			
8	0	22021	0	23104			
9	5	20458	0.000244403	21753			
10	9	18404	0.000489024	19701			
11	9	16757	0.000537089	19172			
12	13	14009	0.000927975	16206			
13	20	11907	0.001679684	13672			
14	40	13490	0.002965159	14892			
15	45	11564	0.003891387	12978			
16	35	8925	0.003921569	9847			
17	21	4502	0.004664594	5290			
18	17	3268	0.005201958	3834			
TOTAL	216	266251		296327			

Example: Predicting colorectal cancer Incidence in Tarragona for the year 2000 using colorectal cancer incidence in Girona for the year 2000 **(USING INCIDENCE RATES)**

Colorectal Cancer Incidence in Girona 2000

Age-Group	*Cases*	*PY*	*Incidence Rate*
1	0	13109	0
2	0	13498	0
3	0	13760	0
4	0	16257	0
5	0	21081	0
6	1	21571	4.63585E-05
7	1	21670	4.61467E-05
8	0	22021	0
9	5	20458	0.000244403
10	9	18404	0.000489024
11	9	16757	0.000537089
12	13	14009	0.000927975
13	20	11907	0.001679684
14	40	13490	0.002965159
15	45	11564	0.003891387
16	35	8925	0.003921569
17	21	4502	0.004664594
18	17	3268	0.005201958
TOTAL	216	266251	

Colorectal Cancer Incidence in Tarragona 2000

PY	*Expected*	*Observed*	*Observed-Expected*
14142	**0**	0	0
14243	**0**	0	0
15621	**0**	0	0
19138	**0**	0	0
24309	**0**	0	0
24511	**1.1362941**	0	-1.136294099
23914	**1.1035533**	0	-1.103553299
23104	**0**	2	2
21753	**5.3165021**	4	-1.316502102
19701	**9.63426429**	3	-6.63426429
19172	**10.2970699**	12	1.702930119
16206	**15.0387608**	19	3.961239203
13672	**22.9646426**	20	-2.964642647
14892	**44.1571534**	41	-3.157153447
12978	**50.5024213**	43	-7.502421308
9847	**38.6156863**	46	7.384313725
5290	**24.6756997**	27	2.324300311
3834	**19.9443084**	16	-3.944308446
296327	**243.386356**	233	**-10.38635628**

233 - 243.3863563 = -10.3863

Example: Predicting colorectal cancer Incidence in Tarragona for the year 2000 using colorectal cancer incidence and mortality in Girona for the year 2000 and colorectal cancer mortality in Tarragona for 2000 **(MAKING USE OF DIFFERENCE RATES: INCIDENCE-MORTALITY)**

Colorectal Cancer Incidence in Girona 2000

Colorectal Cancer Incidence in Tarragona 2000

Age-Group	*Incidence Cases*	*Mortality Deaths*	*Difference Inc-Mort*	*PY*	*Rate Difference*	*PY*	*Expected Difference*	*Observed Mortality*	*Expected Incidence*	*Observed Incidence*	*O-E*
1	0	0	0	13109	0	14142					
2	0	0	0	13498	0	14243					
3	0	0	0	13760	0	15621					
4	0	0	0	16257	0	19138					
5	0	0	0	21081	0	24309					
6	1	1	0	21571	0	24511					
7	1	0	1	21670	4.61467E-05	23914					
8	0	0	0	22021	0	23104					
9	5	1	4	20458	0.000195523	21753					
10	9	1	8	18404	0.000434688	19701					
11	9	5	4	16757	0.000238706	19172					
12	13	5	8	14009	0.000571061	16206					
13	20	9	11	11907	0.000923826	13672					
14	40	18	22	13490	0.001630838	14892					
15	45	13	32	11564	0.002767209	12978					
16	35	25	10	8925	0.001120448	9847					
17	21	16	5	4502	0.001110618	5290					
18	17	11	6	3268	0.001835985	3834					
TOTAL	216	105	111	266251		296327					

Example: Predicting colorectal cancer Incidence in Tarragona for the year 2000 using colorectal cancer incidence and mortality in Girona for the year 2000 and colorectal cancer mortality in Tarragona for 2000 **(MAKING USE OF DIFFERENCE RATES: INCIDENCE-MORTALITY)**

Colorectal Cancer Incidence in Girona 2000

Age-Group	*Incidence Cases*	*Mortality Deaths*	*Difference Inc-Mort*	*PY*	*Rate Difference*
1	0	0	0	13109	0
2	0	0	0	13498	0
3	0	0	0	13760	0
4	0	0	0	16257	0
5	0	0	0	21081	0
6	1	1	0	21571	0
7	1	0	1	21670	4.61467E-05
8	0	0	0	22021	0
9	5	1	4	20458	0.000195523
10	9	1	8	18404	0.000434688
11	9	5	4	16757	0.000238706
12	13	5	8	14009	0.000571061
13	20	9	11	11907	0.000923826
14	40	18	22	13490	0.001630838
15	45	13	32	11564	0.002767209
16	35	25	10	8925	0.001120448
17	21	16	5	4502	0.001110618
18	17	11	6	3268	0.001835985
TOTAL	216	105	111	266251	

Colorectal Cancer Incidence in Tarragona 2000

PY	*Expected Difference*	*Observed Mortality*	*Expected Incidence*	*Observed Incidence*	*O-E*
14142	0	0	0	0	0
14243	0	0	0	0	0
15621	0	0	0	0	0
19138	0	0	0	0	0
24309	0	0	0	0	0
24511	0	1	1	0	-1
23914	1.1035533	0	1.103553299	0	-1.1035533
23104	0	1	1	2	1
21753	4.25320168	1	5.253201681	4	-1.25320168
19701	8.56379048	4	12.56379048	3	-9.56379048
19172	4.5764755	3	7.576475503	12	4.4235245
16206	9.25462203	4	13.25462203	19	5.74537797
13672	12.6305535	7	19.63055346	20	0.36944654
14892	24.2864344	18	42.2864344	41	-1.2864344
12978	35.9128329	15	50.91283293	43	-7.91283293
9847	11.0330532	21	32.03305322	46	13.9669468
5290	5.87516659	15	20.87516659	27	6.12483341
3834	7.03916769	14	21.03916769	16	-5.03916769
296327	124.528851	104	228.5288513	233	4.47114873

124.528851 + 104 = 228.5288513

233 - 228.5288513= 4.47114873

In our prediction problem based on cancer incidence and mortality rates for 2000

C_G: *Observed Cancer Incidence in Girona (Cases)*

D_G: *Observed Cancer Mortality in Girona (Deaths)*

D_T: *Observed Cancer Mortality in Tarragona (Deaths)*

Y_G: *Person-years in Girona*

Y_T: *Person-years in Tarragona*

C_T^P : *Predicted Cancer Incidence in Tarragona (Cases)*

1) Predicting cancer Incidence in Tarragona using Girona cancer incidence rates

$$C_T^P = Y_T \cdot \frac{C_G}{Y_G}$$

2) Predicting cancer Incidence in Tarragona using Girona cancer incidence and mortality rates and Tarragona cancer mortality rates

$$C_T^P = D_T \cdot \frac{C_G}{D_G}$$

$$C_T^P = Y_T \cdot \left(\frac{C_G - D_G}{Y_G}\right) + D_T$$

Introduction

• Government health services/policy planning managers frequently rely on predictions of future disease burdens by biostatisticians.

• However, it is unclear, which of the available prediction models is best suited to a particular situation or cancer type.

• Cancer incidence data could not be available in every locality, while cancer mortality data usually are.

• Our eventual **goal is to predict future cancer burden in areas without a cancer registry**, extrapolated from cancer incidence data of nearby areas that do have cancer registries (either with or without mortality data).

We assume similar cancer incidence and mortality patterns between areas

• Simple **log-linear and linear extrapolation models** have been proposed to perform predictions of future disease burdens when both incidence and mortality data are available for the area of interest.

• As cancer registries often cover only selected geographical areas in many countries, **incidence estimation methods based on mortality data** have also been proposed.

• In France, for example, *Colonna et al. 1999* applied the **incidence-to-mortality ratios** of breast and colorectal cancer data obtained from regions with cancer registries to national mortality data to estimate the national cancer incidence.

• *Ferlay et al. 2010* made **use of a similar approach** to predict worldwide cancer incidence in 2008

• **Age period cohort (APC) models** have been used increasingly for predicting cancer incidence and mortality.

• Projections based on extrapolating age, period and cohort effects require parametric assumptions in non-Bayesian versions of these models (*Osmond 1985*).

• In Bayesian versions of the APC models, period and cohort effects are smoothed and extrapolated by means of autoregressive priors (*Bashir and Estève, 2001; Knorr-Held 2001, Clèries et al 2006*), where an appropriate degree of smoothing can be learned from the data (*Bray 2000*).

• In situations where rates are low and unstable, Bayesian APC models can achieve sensible predictions (*Clèries et al 2010*) where other methods may fail.

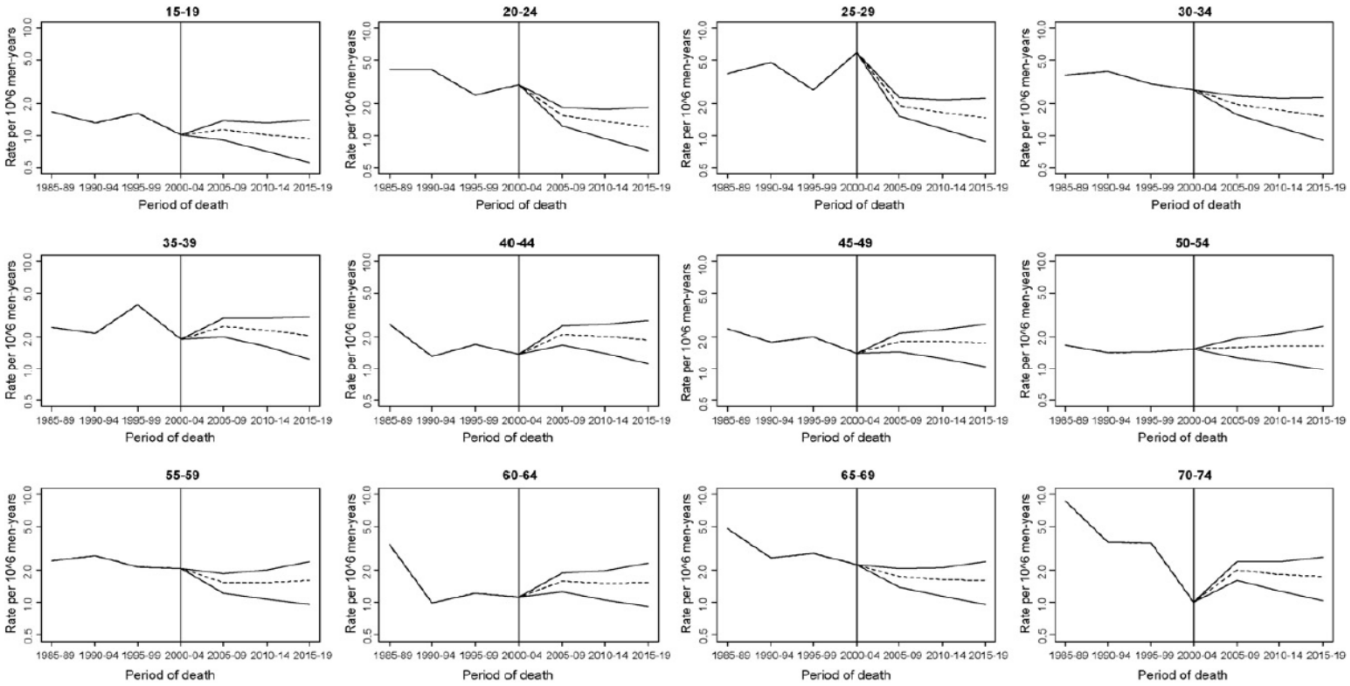


Fig. 2. Projection of the mortality rates (per 1,000,000 men-year) by testicular cancer in Spain for the age groups 15–74 years according to period of death. Projections for periods 2005–2009, 2010–2014, 2015–2019 (dashed line - - -) and their 95% credibility intervals (solid line —).

Ref: Cleries R, Martinez JM, Escriba JM, Esteban L, Pareja L, Borras JM, Ribes J. Monitoring the decreasing trend of testicular cancer mortality in Spain during 2005-2019 through a Bayesian approach. Cancer Epidemiology 2010; 34(3):244-256. DOI:10.1016/j.canep.2010.03.003

However, **APC models require a long period of observed data** as a basis for prediction and may present **interpretation difficulties** in practice with wider credible or prediction intervals than those based on **simple linear or log-linear** models.

Statistical Methods

Statistical Methods (I): Linear and log-linear models

C_{it} : Cancer incident cases in the i-th age-group and year t

D_{it} : Cancer mortality cases in the i-th age-group and year t

Y_{it} : Person-years at risk in the i-th age-group and year t

Assuming $C_{it} \sim Poisson(\mu_{it})$

The linear and log-linear models presented in this study are in the form:

$$h(C_{it}, Y_{it}, D_{it}) = g(t; \alpha_i, \beta_i)$$

α_i : Intercept

β_i : Slope

Dyba and Hakulinen 2000 proposed the following log-linear model to predict cancer incidence

Age-Period-Incidence model (API)

$$\frac{\mu_{it}}{Y_{it}} = e^{\alpha_i + \beta_i (t - t_0)} \qquad (1)$$

Similarly, using incidence-to-mortality ratio

Age-Period-Incidence-to-Mortality Ratio model (API_R)

$$\frac{\mu_{it}}{D_{it}} = e^{\alpha_i + \beta_i (t - t_0)} \qquad (2)$$

We propose a new model using cancer mortality assuming $\mu_{it} \geq D_{it}$

Age-Period-Incidence-Mortality Difference model (API_D)

$$\frac{\mu_{it} - D_{it}}{Y_{it}} = e^{\alpha_i + \beta_i (t - t_0)} \qquad (3)$$

Linear versions of these models have been also used (rising incidence rates):

linAPI Model

$$\frac{\mu_{it}}{Y_{it}} = \alpha_i + \beta_i (t - t_0) \quad (4)$$

linAPI$_R$ model

$$\frac{\mu_{it}}{D_{it}} = \alpha_i + \beta_i (t - t_0) \quad (5)$$

linAPI$_D$ model

$$\frac{\mu_{it} - D_{it}}{Y_{it}} = \alpha_i + \beta_i (t - t_0) \quad (6)$$

Constraints to linear models (preventing prediction of negative rates)

For model (5) $$(\frac{\mu_{it_0}}{D_{it_0}}) \leq (\frac{\mu_{it_q}}{D_{it_q}}) \mid t_q > t_0, \forall t_q$$

For model (6) $$(\mu_{it_0} - D_{it_0}) / Y_{it_0} \leq (\mu_{it_q} - D_{it_q}) / Y_{it_0} \mid t_q > t_0, \forall t_q$$

Models with the same slope parameter for all age groups (age-drift models) can be applied in practice as parsimonious versions of the age-period model

Example: Linear Age-Drift-Incidence-Mortality Difference model ($linADI_D$)

$linADI_D$ model

$$\frac{\mu_{it} - D_{it}}{Y_{it}} = \alpha_i + \beta(t - t_0)$$

We considered these “age-drift” models in addition to models (1)-(6) and, therefore, our analysis included a total of 12 models:
6 Age-Period + 6 Age-Drift

Statistical Methods (II): Model fitting with WinBUGS and R

Model Fitting with WinBUGS 1.3 (I)

(http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml)

- **Prior distributions:**

$$\alpha_i \sim Normal(0, \tau_\alpha) \qquad \tau_\alpha \sim Gamma(\psi, \phi)$$

$$\beta_i \sim Normal(0, \tau_\beta) \qquad \tau_\beta \sim Gamma(\psi, \phi)$$

Gamma definition

$$p(\tau) = \frac{\phi^\psi}{\Gamma(\psi)} \tau^{\psi-1} e^{-\phi\tau}, \tau > 0 \quad \Bigg\} \quad E[\tau] = \frac{\psi}{\phi}, \quad Var[\tau] = \frac{\psi}{\phi^2}$$

Parameter setting

$$\psi = \phi = 0.001$$

Sensibility analysis

$$\psi = 0.5 \quad \phi = 0.0005$$

- Markov Chain Monte-Carlo (MCMC) run of N=10,000 (N=2,000 as “burn-in”) with 3 chains for each model.

- The scale reduction factor (Gelman-Rubin convergence diagnostic) was calculated for each model parameter to assess convergence and adequate mixing of the chains.

Model Fitting with WinBUGS 1.3 (II)

(http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml)

For a linear model, example, $linAPI_{\mathrm{D}}$

$$\frac{\mu_{it} - D_{it}}{Y_{it}} = \alpha_i + \beta_i (t - t_0)$$

- **We assume a constraint to linear models (avoid prediction of negative rates):**

$$\alpha_i + \beta_i \cdot (t - t_0) > 0$$

- **Since 0 is a lower bound for rates, we must solve** $\alpha_i + \beta_i \cdot (t - t_0) = 0$

$$\alpha_i + \beta_i \cdot (t - t_0) = \alpha_i + \beta_i \cdot (t_q) = 0 \mid t_q = (t - t_0) \Leftrightarrow \beta_i = \frac{-\alpha_i}{t_q}$$

1) Sample $\alpha_i \sim Normal(0, \tau_\alpha) I(0.0,)$

2) Let lower bound be $L_i = \frac{-\alpha_i}{t_q}$

3) Sample $\beta_i \sim Normal(0, \tau_\beta) I(L_i,)$

It is a lower bound for β

Statistical Methods (IV):

Predicting future cancer incidence counts in an area without cancer registry making use of the predictive distribution of the number of cases

Prediction of cancer cases: Using data from region A to predict in B

$$C_B^P = Y_B \cdot \frac{C_A}{Y_A}$$

$$C_B^P = D_B \cdot \frac{C_A}{D_A}$$

$$C_B^P = Y_B \cdot \left(\frac{C_A - D_A}{Y_A}\right) + D_B$$

$\boldsymbol{C_A}$: *Observed Cancer Incidence in A*

$\boldsymbol{D_A}$: *Observed Cancer Mortality in A*

$\boldsymbol{D_B}$: *Observed Cancer Mortality in B*

$\boldsymbol{Y_A}$: *Person-years in A*

$\boldsymbol{Y_B}$: *Person-years in B*

$\boldsymbol{C_B^P}$: *Predicted Cancer Incidence in B*

The predictive distribution (example with linADI$_D$ model):

1) Fit the model to *A* data $\mu_{itA} = Y_{itA}(\alpha_i + \beta(t - t_0)) + D_{itA}$

2.1) Use 1) to predict RATES in *B* $\hat{\mu}^P_{itB} = Y_{itB}(\alpha_i + \beta(t - t_0)) + D_{itB}$

2.2) Sample CASES from Poisson distribution $C^P_{itB} \sim Poisson\ (\hat{\mu}^P_{itB})$

Statistical Methods (V):

Cancer Incidence and Mortality Data

Cancer incidence data have been obtained from the Tarragona Cancer Registry (TCR) and Girona Cancer Registry (GCR), which cover 20% of the population of Catalonia:

- Data from TCR were available during 1983-2001, whereas GCR data were available for 1994-2002.

- TCR and GCR data are published in the series of Cancer Incidence in Five Continents since 1983 and 1994, respectively.

- Mortality data were available during 1985-2004.

- Age is categorized into 18 groups (5-year bands): 0-4, 5-9,…, 80-84, 85+.

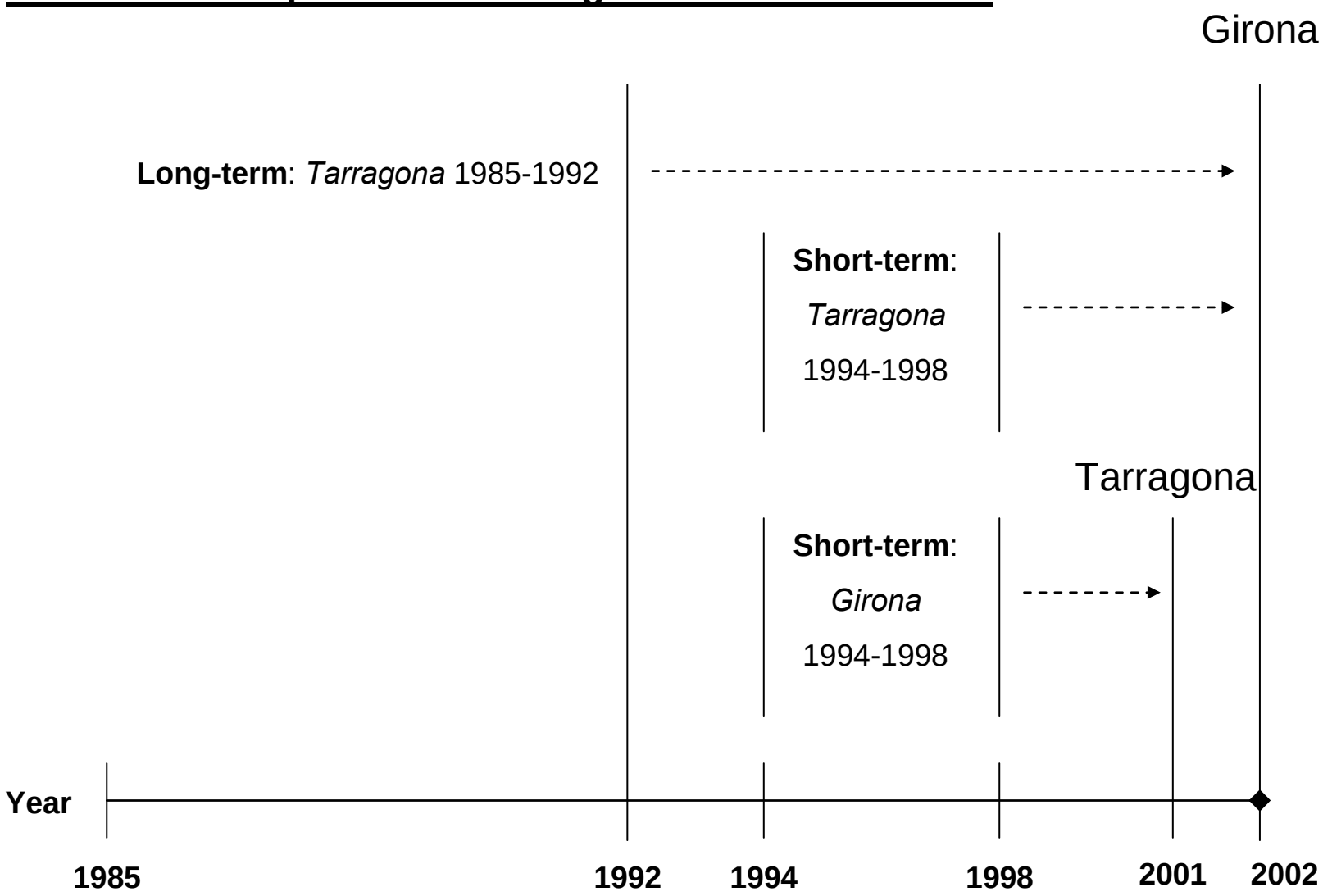
- Cancer sites analyzed were breast (BC), colorectal (CRC), lung (LC) and prostate (PC) cancers, the four most common cancer sites in Catalonia.

Assessment of predictions:

1) Short-term: 3-4 years ahead

2) Long-term: 10 years ahead

Scheme of the predictions: long-term and short-term



1) Long-term: prediction 10 years ahead

2) Short-term: prediction 3 years (Tarragona with base Girona) and 4 years (Girona with base Tarragona) ahead

Scheme of the predictions: long-term and short-term

3.1 Short-term analysis: Projection of cancer incidence for years 2001 and 2002 using cancer incidence and mortality data from 1994-1998

3.1.1 Short-term analysis: Projection of cancer incidence in Tarragona for 2001 using cancer incidence and mortality data from 1994-1998 in Girona

3.1.2 Short-term analysis: Projection of cancer incidence in Girona for 2002 using cancer incidence and mortality data from 1994-1998 in Tarragona

3.2 Long-term analysis: Projection of cancer incidence in Girona for year 2002 using cancer incidence and mortality data in Tarragona from 1985-1992

Statistical Methods (VI):

Model Assessment

In order to compare across the models for cancer incidence prediction, we calculated three scoring rules *(Gneiting T, Raferty AE. 2007)*

Logarithmic Score (LogS)

$-\log(p_x)$ *where* p_x *is the predicted probability mass at the observed count x*

Dawid-Sebastiani Score (DSS)

$(\frac{x-\mu_P}{\sigma_P})^2 + 2\log(\sigma_P)$ *where* μ_P *and* σ_P *mean and standard deviation of the predicted distribution*

Ranked Probability Score

$\sum_{k=0}^{\infty}\{P_k - 1(x \le k)\}^2$ *where* P_k *is the predicted cumulative distribution function.*

For all of the three scores, we took the mean score over the 18 age groups, where **smaller mean score values correspond to better predictions**.

The computation of the ***LogS, DSS and RPS*** was performed using the N=8000 ***MCMC*** samples from the predictive density.

Results (I)

Model Assessment in SHORT-TERM predictions

Scoring Rules: Twelve models predicting Tarragona cancer incidence cases for 2001 using Girona data for the period 1994-1998

Score [ranking]

Men	*Colorectal*				*Lung*				*Prostate*			
	LogS	*RPS*	*DSS*		*LogS*	*RPS*	*DSS*		*LogS*	*RPS*	*DSS*	
API	0.845 [5]	0.991 [4]	3.825 [5]	*	0.868 [4]	1.149 [4]	3.979 [5]		0.850 [6]	0.804 [6]	4.754 [6]	
API$_R$	0.922 [10]	1.591 [10]	6.482 [10]		0.904 [7]	1.303 [7]	6.572 [9]		0.907 [12]	1.533 [12]	8.210 [11]	
API$_D$	0.955 [11]	1.913 [12]	19.751 [12]		0.890 [5]	1.669 [10]	12.093 [12]		0.869 [11]	0.987 [11]	16.454 [12]	
linAPI	0.827 [3]	0.996 [5]	2.850 [3]		0.925 [12]	1.710 [11]	3.728 [4]	*	0.544 [2]	0.547 [4]	1.758 [2]	
linAPI$_R$	0.870 [7]	1.186 [6]	5.769 [8]		0.911 [8]	1.283 [6]	6.297 [8]		0.863 [8]	0.818 [8]	6.091 [9]	
linAPI$_D$	0.859 [6]	1.441 [7]	3.233 [4]		0.918 [10]	1.568 [9]	3.258 [3]	*	**0.537 [1]**	**0.486 [1]**	**1.548 [1]**	
ADI	0.838 [4]	0.966 [3]	4.023 [6]	*	0.851 [3]	**0.978 [1]**	4.160 [6]		0.845 [5]	0.795 [5]	4.115 [5]	
ADI$_R$	0.893 [9]	1.483 [8]	5.804 [9]		0.923 [11]	1.729 [12]	6.573 [10]		0.868 [10]	0.822 [10]	6.012 [8]	
ADI$_D$	0.960 [12]	1.874 [11]	6.694 [11]		0.891 [6]	1.231 [5]	6.694 [11]	*	0.858 [7]	0.816 [7]	5.285 [7]	
linADI	0.813 [2]	0.839 [2]	2.137 [2]	*	0.825 [2]	1.104 [3]	2.811 [2]		0.565 [3]	0.487 [2]	2.038 [3]	
linADI$_R$	0.885 [8]	1.520 [9]	5.047 [7]		0.913 [9]	1.344 [8]	6.223 [7]		0.867 [9]	0.821 [9]	6.380 [10]	*
linADI$_D$	**0.808 [1]**	**0.834 [1]**	**2.104 [1]**		**0.818 [1]**	1.041 [2]	**2.006 [1]**		0.579 [4]	0.503 [3]	2.096 [4]	

Women	*Colorectal*				*Lung*				*Breast*			
	LogS	*RPS*	*DSS*		*LogS*	*RPS*	*DSS*		*LogS*	*RPS*	*DSS*	
API	0.845 [4]	1.031 [2]	4.161 [5]		0.815 [6]	0.792 [5]	2.714 [6]		0.937 [2]	1.671 [2]	4.268 [5]	*
API$_R$	1.024 [12]	3.907 [12]	6.395 [7]		0.871 [10]	0.824 [8]	6.196 [9]		1.059 [11]	6.307 [11]	5.760 [8]	
API$_D$	0.985 [10]	2.415 [11]	27.607 [12]		0.852 [7]	0.879 [11]	6.694 [11]		1.087 [12]	8.046 [12]	16.302 [12]	*
linAPI	0.839 [3]	1.576 [6]	3.442 [4]	*	0.748 [4]	0.939 [12]	2.522 [5]	*	0.981 [7]	2.455 [7]	4.299 [6]	
linAPI$_R$	0.954 [7]	2.001 [7]	6.796 [10]		0.883 [12]	0.836 [10]	6.875 [12]		0.966 [6]	2.157 [6]	5.885 [10]	*
linAPI$_D$	0.830 [2]	1.062 [5]	2.822 [3]	*	0.713 [3]	0.671 [3]	2.139 [3]	*	0.988 [8]	2.600 [9]	4.173 [3]	
ADI	0.856 [5]	1.055 [4]	4.218 [6]		0.755 [5]	0.791 [4]	2.279 [4]		0.943 [3]	1.723 [3]	4.225 [4]	*
ADI$_R$	0.987 [11]	2.283 [10]	6.482 [8]		0.869 [8]	0.823 [6]	5.515 [7]		0.993 [9]	2.580 [8]	5.560 [7]	*
ADI$_D$	0.979 [9]	2.226 [8]	7.302 [11]	*	0.873 [11]	0.826 [9]	6.602 [10]		1.027 [10]	4.217 [10]	5.794 [9]	*
linADI	**0.791 [1]**	**0.871 [1]**	**2.097 [1]**		0.597 [2]	0.552 [2]	1.453 [2]		**0.936 [1]**	**1.631 [1]**	**3.654 [1]**	
linADI$_R$	0.955 [8]	2.248 [9]	6.483 [9]		0.870 [9]	0.824 [7]	5.672 [8]		0.956 [4]	1.930 [4]	5.923 [11]	*
linADI$_D$	0.861 [6]	1.039 [3]	2.789 [2]		**0.569 [1]**	**0.469 [1]**	**1.120 [1]**		0.960 [5]	1.996 [5]	3.973 [2]	

Except prostate cancer, all best models used of the age-drift assumption.

Models *linADI* and *linADI$_D$* performed relatively well, consistently across the cancer sites and sex.

For breast cancer, many models' prediction intervals missed the observed ASR.

Scoring Rules: Twelve models predicting Girona cancer incidence cases for 2002 using Tarragona data for the period 1994-1998

Score [ranking]

Men	*Colorectal*				*Lung*				*Prostate*			
	LogS	*RPS*	*DSS*		*LogS*	*RPS*	*DSS*		*LogS*	*RPS*	*DSS*	
API	0.889 [6]	1.168 [4]	4.590 [6]	*	0.888 [8]	1.383 [11]	4.573 [8]	*	0.910 [7]	1.327 [8]	5.247 [6]	
API$_R$	0.964 [11]	1.994 [11]	5.874 [8]		0.895 [10]	1.305 [7]	4.722 [10]		0.979 [9]	1.812 [10]	6.312 [8]	*
API$_D$	1.069 [12]	3.883 [12]	8.948 [12]		0.898 [11]	1.382 [10]	5.114 [11]		0.904 [6]	1.217 [6]	9.167 [12]	
linAPI	0.880 [5]	1.646 [10]	3.542 [4]		0.864 [4]	1.460 [12]	3.419 [4]		0.676 [2]	0.607 [2]	1.006 [2]	
linAPI$_R$	0.918 [9]	1.450 [8]	6.081 [9]		0.890 [9]	1.143 [4]	4.663 [9]		0.993 [10]	1.734 [9]	7.356 [11]	*
linAPI$_D$	0.859 [3]	1.189 [5]	3.110 [3]		0.877 [7]	1.370 [9]	3.245 [3]		**0.600 [1]**	**0.604 [1]**	**1.003 [1]**	
ADI	0.862 [4]	1.087 [2]	4.319 [5]		0.851 [2]	1.099 [3]	4.540 [7]		0.852 [3]	1.097 [3]	4.721 [5]	
ADI$_R$	0.910 [8]	1.468 [9]	5.435 [7]		0.872 [5]	1.145 [5]	4.341 [5]		1.017 [12]	1.819 [11]	6.438 [10]	*
ADI$_D$	0.902 [7]	1.384 [6]	6.512 [11]	*	0.899 [12]	1.366 [8]	5.115 [12]	*	0.914 [8]	1.317 [7]	6.174 [7]	*
linADI	0.811 [2]	1.143 [3]	2.725 [2]		0.852 [3]	1.234 [6]	2.655 [2]	*	0.871 [4]	1.122 [4]	2.868 [3]	*
linADI$_R$	0.922 [10]	1.438 [7]	6.092 [10]		0.876 [6]	1.027 [2]	4.410 [6]		1.000 [11]	1.820 [12]	6.424 [9]	*
linADI$_D$	**0.733 [1]**	**0.725 [1]**	**1.845 [1]**		**0.838 [1]**	**0.939 [1]**	**2.599 [1]**		0.877 [5]	1.208 [5]	3.006 [4]	*

Women	*Colorectal*				*Lung*				*Breast*			
	LogS	*RPS*	*DSS*		*LogS*	*RPS*	*DSS*		*LogS*	*RPS*	*DSS*	
API	0.841 [4]	1.016 [7]	3.237 [3]		0.840 [5]	0.797 [5]	3.951 [7]		0.917 [3]	1.823 [2]	4.274 [4]	
API$_R$	0.869 [6]	1.220 [9]	5.825 [8]		0.870 [8]	0.824 [7]	4.297 [10]		1.033 [12]	4.623 [12]	5.848 [11]	
API$_D$	0.873 [9]	1.205 [8]	7.033 [12]		0.858 [7]	1.016 [12]	3.737 [5]		1.023 [11]	2.436 [8]	9.350 [12]	*
linAPI	0.853 [5]	1.295 [11]	3.359 [5]	*	0.599 [3]	**0.504 [1]**	1.617 [4]		0.974 [6]	2.351 [7]	4.283 [5]	
linAPI$_R$	0.874 [10]	0.990 [6]	5.997 [9]		0.871 [10]	0.827 [9]	5.291 [11]		0.987 [7]	3.937 [11]	5.036 [8]	
linAPI$_D$	0.883 [12]	1.318 [12]	3.279 [4]	*	0.608 [4]	0.620 [4]	1.553 [3]		0.996 [8]	2.323 [6]	4.305 [6]	
ADI	0.836 [3]	0.811 [2]	3.593 [6]		0.843 [6]	0.798 [6]	4.208 [9]		**0.893 [1]**	**1.287 [1]**	**2.397 [1]**	
ADI$_R$	0.878 [11]	1.232 [10]	6.482 [10]		0.872 [11]	0.831 [10]	6.456 [12]		1.003 [9]	3.275 [9]	5.780 [10]	
ADI$_D$	0.871 [8]	0.921 [4]	6.587 [11]	*	0.873 [12]	0.826 [11]	3.739 [6]		0.963 [5]	2.307 [5]	5.267 [9]	*
linADI	0.796 [2]	0.900 [3]	2.206 [2]		0.597 [2]	0.519 [2]	0.686 [2]		0.905 [2]	1.872 [3]	3.426 [2]	
linADI$_R$	0.870 [7]	0.955 [5]	5.521 [7]		0.872 [9]	0.825 [8]	4.162 [8]		1.017 [10]	3.743 [10]	5.007 [7]	
linADI$_D$	**0.794 [1]**	**0.801 [1]**	**2.184 [1]**		**0.581 [1]**	0.525 [3]	**0.363 [1]**		0.941 [4]	1.945 [4]	3.780 [3]	

Except prostate cancer, all best models used of the age-drift assumption.

Model *linADI$_D$* performed relatively well, consistently across the cancer sites and sex.

For breast cancer, *ADI model performed well*

Results (II)

Model Assessment in LONG-TERM predictions

Scoring Rules: Twelve models predicting Girona cancer incidence cases for 2002 using Tarragona data for the period 1985-1992

Men	*Colorectal*			*Lung*			*Prostate*		
	LogS	*RPS*	*DSS*	*LogS*	*RPS*	*DSS*	*LogS*	*RPS*	*DSS*
API	2.572 [4]	1.993 [9]	4.827 [6]	2.843 [11]	2.068 [10]	5.642 [10]	2.368 [6]	0.826 [8]	5.823 [6]
API$_R$	3.899 [12]	4.205 [12]	7.905 [11]	2.394 [3]	0.922 [2]	4.670 [7]	2.758 [11]	2.093 [11]	7.072 [11]
API$_D$	3.782 [11]	3.605 [11]	8.784 [12]	2.788 [10]	2.878 [12]	5.727 [11]	3.944 [12]	4.498 [12]	8.408 [12]
linAPI	2.778 [5]	2.051 [10]	3.755 [4]	2.955 [12]	2.257 [11]	4.081 [4] *	1.410 [2]	0.397 [2]	0.878 [2]
linAPI$_R$	3.141 [10]	1.909 [7]	6.197 [8]	2.422 [5]	0.948 [4]	4.594 [6]	2.382 [8]	0.823 [7]	6.438 [9] *
linAPI$_D$	2.781 [6]	1.751 [5]	3.362 [3]	2.522 [7]	1.683 [8]	3.622 [3] *	**1.397 [1]**	**0.389 [1]**	**0.868 [1]**
ADI	2.391 [3]	1.405 [3]	3.865 [5]	2.624 [8]	1.494 [7]	5.588 [9]	2.321 [5]	0.815 [5]	4.928 [5]
ADI$_R$	2.957 [8]	1.833 [6]	6.402 [9]	2.397 [4]	**0.864 [1]**	5.271 [8]	2.584 [10]	1.319 [10]	6.424 [8]
ADI$_D$	2.948 [7]	1.631 [4]	6.558 [10] *	2.715 [9]	1.759 [9]	5.727 [11] *	2.487 [9]	1.043 [9]	6.678 [10] *
linADI	**2.117 [1]**	1.079 [2]	**2.449 [1]** *	2.482 [6]	1.146 [6]	3.117 [2]	1.779 [3]	0.666 [3]	1.303 [3] *
linADI$_R$	3.003 [9]	1.952 [8]	6.171 [7]	2.390 [2]	0.942 [3]	4.345 [5]	2.373 [7]	0.821 [6]	6.423 [7]
linADI$_D$	2.154 [2]	**1.072 [1]**	2.686 [2]	**2.326 [1]**	1.067 [5]	**2.764 [1]**	1.874 [4]	0.787 [4]	1.448 [4] *

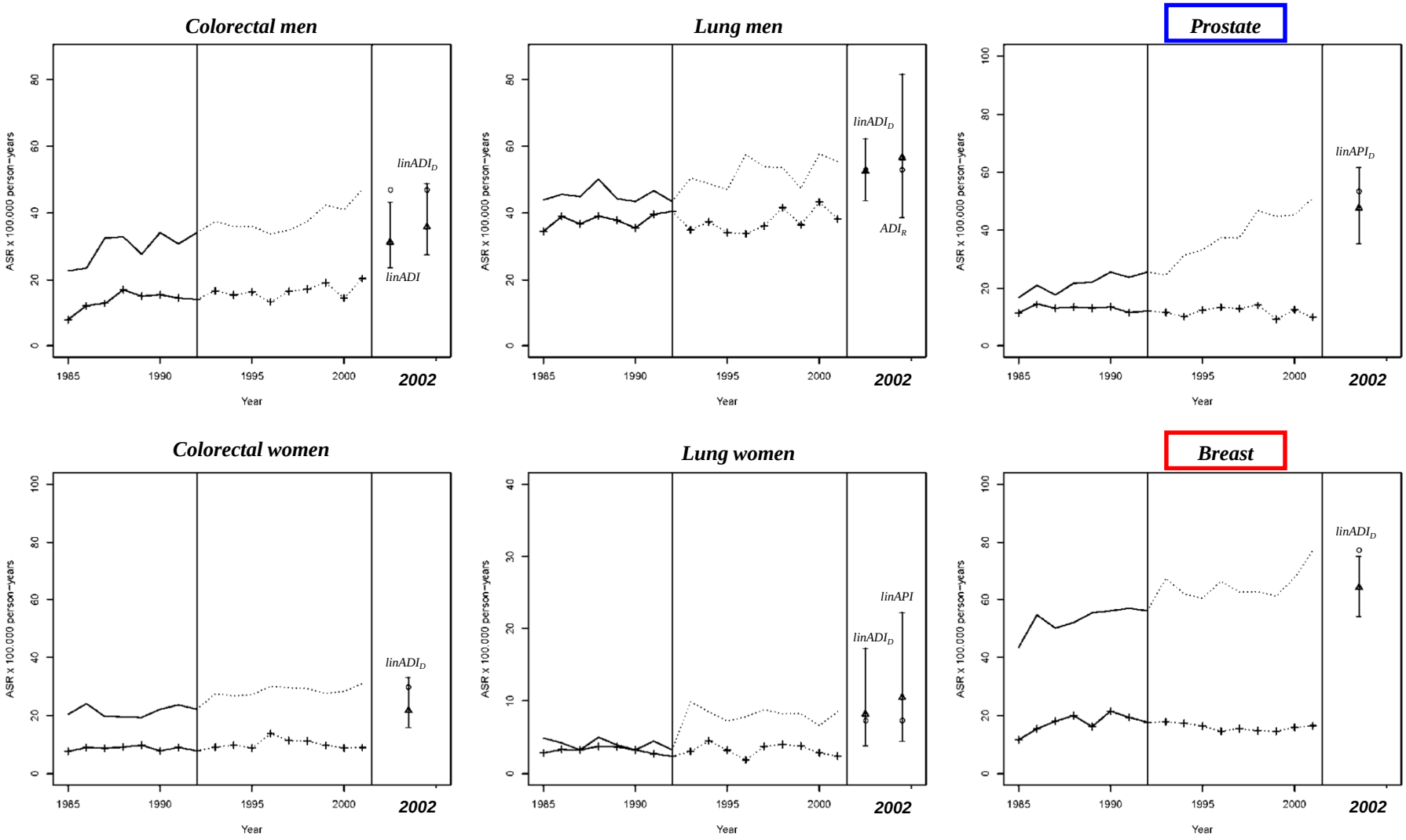
Women	*Colorectal*			*Lung*			*Breast*		
	LogS	*RPS*	*DSS*	*LogS*	*RPS*	*DSS*	*LogS*	*RPS*	*DSS*
API	2.631 [6]	1.602 [6]	4.144 [6] *	2.308 [9]	0.811 [8]	3.227 [9]	2.862 [3]	1.821 [3]	4.631 [3]
API$_R$	3.218 [10]	2.882 [10]	8.732 [10]	1.905 [7]	0.634 [6]	2.283 [7]	4.756 [11]	3.851 [10]	7.975 [11]
API$_D$	5.469 [12]	5.903 [12]	9.433 [12]	1.768 [5]	0.584 [3]	1.581 [4]	5.058 [12]	7.054 [12]	8.425 [12] *
linAPI	2.112 [3]	0.965 [3]	2.817 [4]	1.739 [4]	**0.522 [1]**	1.416 [3]	3.244 [6]	3.359 [7]	5.001 [6]
linAPI$_R$	2.894 [8]	1.989 [8]	6.724 [9]	2.098 [8]	0.827 [12]	2.741 [8]	3.627 [9]	3.606 [8]	6.801 [10] *
linAPI$_D$	2.032 [2]	0.964 [2]	2.808 [3]	1.721 [3]	0.791 [7]	2.105 [6] *	3.037 [5]	2.943 [6]	4.761 [5]
ADI	2.543 [5]	1.445 [5]	3.827 [5]	2.328 [11]	0.816 [9]	5.631 [12]	2.938 [4]	2.235 [4]	4.730 [4]
ADI$_R$	2.955 [9]	2.254 [9]	4.497 [7]	2.324 [10]	0.824 [10]	3.392 [10]	3.487 [7]	2.736 [5]	6.602 [9] *
ADI$_D$	5.266 [11]	3.434 [11]	8.746 [11] *	1.768 [5]	0.584 [3]	1.581 [4]	4.017 [10]	6.821 [11]	6.563 [8] *
linADI	2.299 [4]	1.413 [4]	2.465 [2]	1.678 [2]	0.627 [5]	1.343 [2]	2.660 [2]	1.531 [2]	3.475 [2] *
linADI$_R$	2.785 [7]	1.908 [7]	6.052 [8]	2.379 [12]	0.824 [11]	3.564 [11]	3.531 [8]	3.817 [9]	6.410 [7]
linADI$_D$	**1.975 [1]**	**0.886 [1]**	**1.768 [1]**	**1.619 [1]**	0.553 [2]	**1.119 [1]**	**2.636 [1]**	**1.529 [1]**	**3.379 [1]** *

Scores agreed somewhat less for this long-term prediction than for the short-term predictions.

Model *linADI$_D$*, performed relatively well, consistently across the cancer sites and sex, except for prostate cancer.

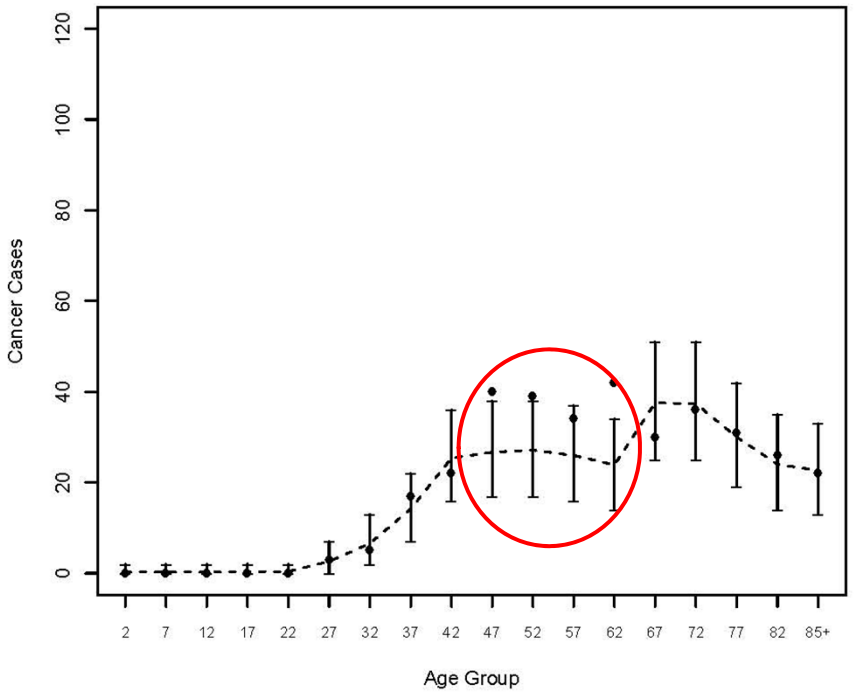
For breast cancer, many models' prediction intervals missed the observed ASR.

Prediction of Age-Standardised rates in Girona for 2002 using Tarragona's data from 1985-1992

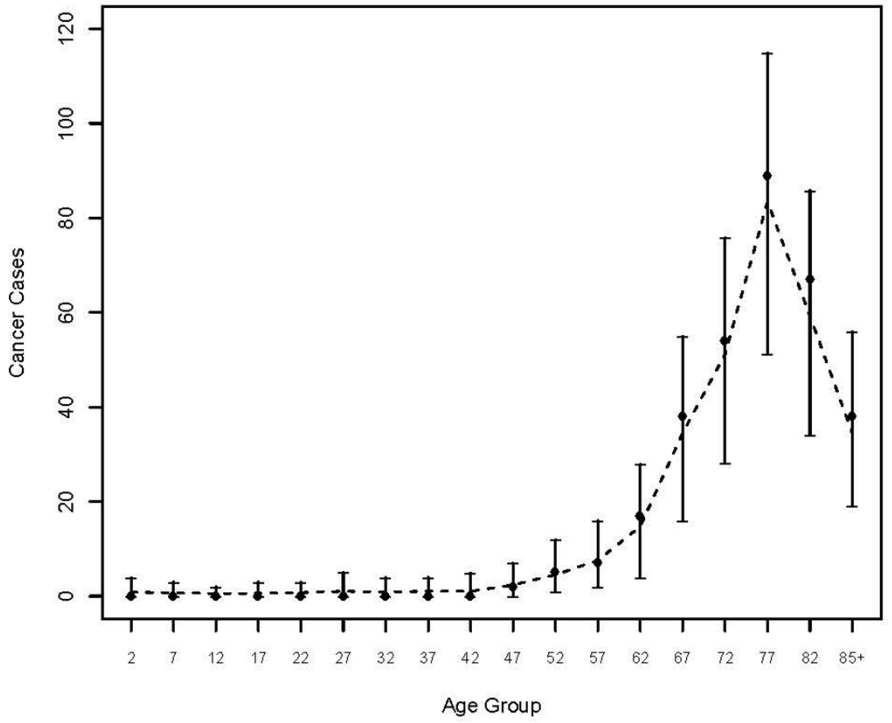


Model Performance: Prediction of the Number of Cases by Age-Group

Breast Cancer



Prostate Cancer



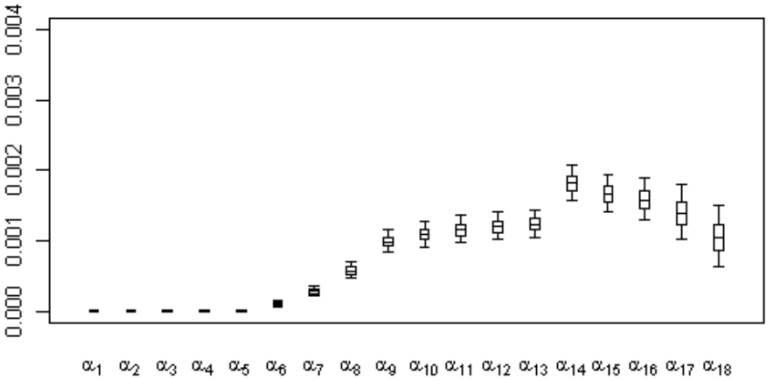
95% Prediction Interval for 2002 in Girona using Tarragona's data from 1985-1992

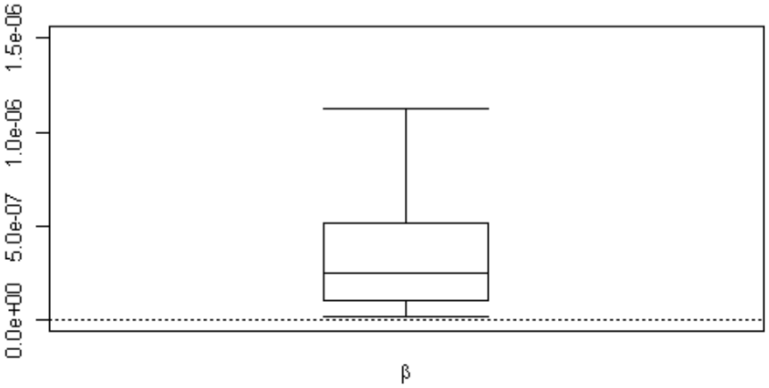
Observed Number of Cases

Posterior distribution of Model Parameters

Breast Cancer

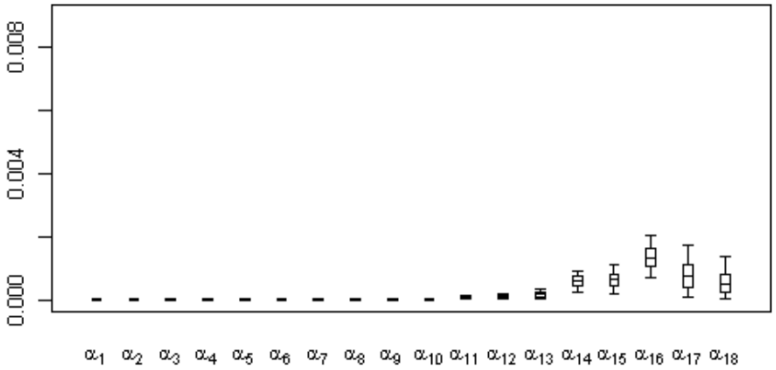
$\boldsymbol{linADI_D}$ $\dfrac{\mu_{it} - D_{it}}{Y_{it}} = \alpha_i + \beta(t - t_0)$

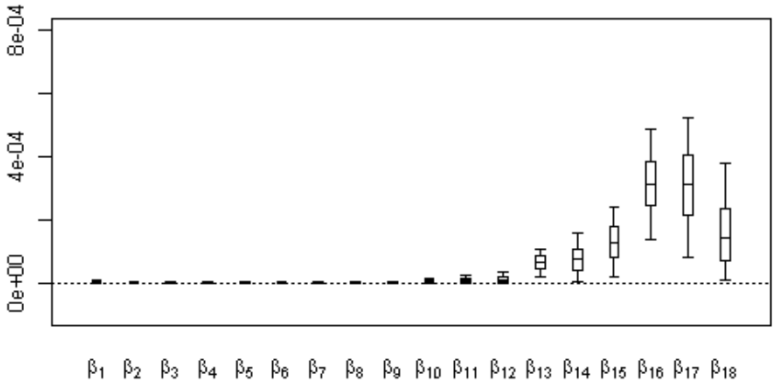




Prostate Cancer

$\boldsymbol{linAPI_D}$ $\dfrac{\mu_{it} - D_{it}}{Y_{it}} = \alpha_i + \beta_i(t - t_0)$





Conclusions

• Of the twelve models we compared, *linADI* and $linADI_D$ showed a relatively better prediction performance consistently across cancer sites and sex, perhaps the latter being slightly better than the former.

• Thus, at least in the settings we considered, the age-drift assumption appeared reasonable.

• An exception to this was prostate cancer for which $linAPI_D$ performed best: use of age-specific slopes for projections may be the reason for better prediction for cancer sites such as prostate.

• Our results suggest that extrapolating time-trends of incidence rates minus mortality rates may have the best predictive performance overall.

• In conclusion, the simple models with Bayesian analysis can be useful for predicting cancer incidence, even in those areas without cancer registries under the assumption that those areas have incidence/mortality patterns similar to nearby areas with cancer registries.

• These methods of population-level disease-incidence prediction are highly relevant to health care planning and policy decisions.

Acknowledgements

This work was supported by *Pla Director d'Oncologia* and:

1) Yamagiwa-Yoshida Memorial International Union Against Cancer Study Grant (YAMAGIWA-YOSHIDA MEMORIAL YY1/09/007)

2) Operating Grant of the Canadian Institute of Health Research entitled "Statistical Methods for Epidemiologic Investigations".

More details in

Research Article

Statistics in Medicine

Received 17 August 2010, Accepted 17 October 2011 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4463

Bayesian approach to predicting cancer incidence for an area without cancer registration by using cancer incidence data from nearby areas‡

Ramon Clèries,[a,b*†] Josepa Ribes,[a,b] Maria Buxo,[c] Alberto Ameijide,[d] Rafael Marcos-Gragera,[c] Jaume Galceran,[d] José Miguel Martínez[e,f,g] and Yutaka Yasui[h]