Optimal number of years to make a prediction for mortality or incidence rates

Joan Valls Marsal [1]
Ramon Clèries Soler [2]
[1] Biostatistics Unit. Biomedical Research Institute of Lleida
[2] Catalan Cancer Registry
jvalls@irblleida.cat / r.cleries@iconcologia.net

October 25th 2013

Continguts

Background
Cancer incidence and mortality

- Incidence and mortality cancer rates:
  - usually defined as the number of incident cases or deaths in one single year, per 100,000 persons.
  - recorded by the Incidence and Mortality Cancer Registries, regional or population-based.
  - conventional epidemiological measures for assessing cancer.
  - rate trends and their modelling: key issue to evaluate the impact of new treatments, health programs or other epidemiological related issues.

Background

2008 World wide cancer incidence and mortality rates

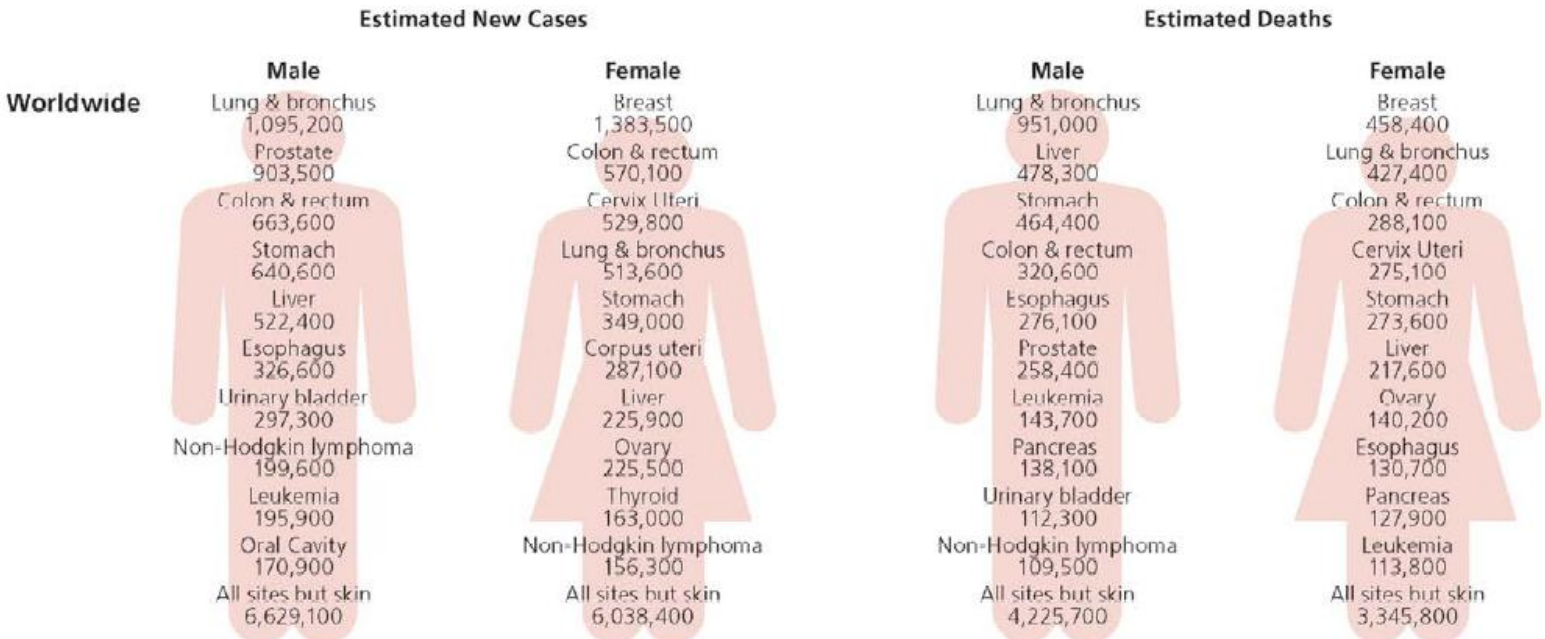
TABLE 1. Incidence and Mortality Rates and Cumulative Probability of Developing Cancer by Age 75 by Sex and Cancer Site for More Developed and Less Developed Areas, 2008

	MORE DEVELOPED AREAS				LESS DEVELOPED AREAS			
	INCIDENCE		MORTALITY		INCIDENCE		MORTALITY	
	ASR	CUMULATIVE RISK (%) [AGE 0-74]	ASR	CUMULATIVE RISK (%) [AGE 0-74]	ASR	CUMULATIVE RISK (%) [AGE 0-74]	ASR	CUMULATIVE RISK (%) [AGE 0-74]
Males								
All cancers* (C00-97, but C44)	300.1	30.1	143.9	15.0	160.3	17.0	119.3	12.7
Bladder (C67)	16.6	1.9	4.6	0.5	5.4	0.6	2.6	0.3
Brain, nervous system (C70-72)	6.0	0.6	3.9	0.4	3.2	0.3	2.6	0.3
Colorectum (C18-21)	37.6	4.4	15.1	1.7	12.1	1.4	6.9	0.8
Esophagus (C15)	6.5	0.8	5.3	0.6	11.8	1.4	10.1	1.2
Gallbladder (C23-24)	2.4	0.3	1.6	0.2	1.4	0.2	1.1	0.1
Hodgkin lymphoma (C81)	2.2	0.2	0.4	0.0	0.9	0.1	0.6	0.1
Kidney (C64-66)	11.8	1.4	4.1	0.5	2.5	0.3	1.3	0.1
Larynx (C32)	5.5	0.7	2.4	0.3	3.5	0.4	2.1	0.3
Leukemia (C91-95)	9.1	0.9	4.8	0.5	4.5	0.4	3.7	0.3
Liver (C22)	8.1	1.0	7.2	0.9	18.9	2.2	17.4	2.0

Jemal et al (2001) CANCER J CLIN ;61:6990)

Background

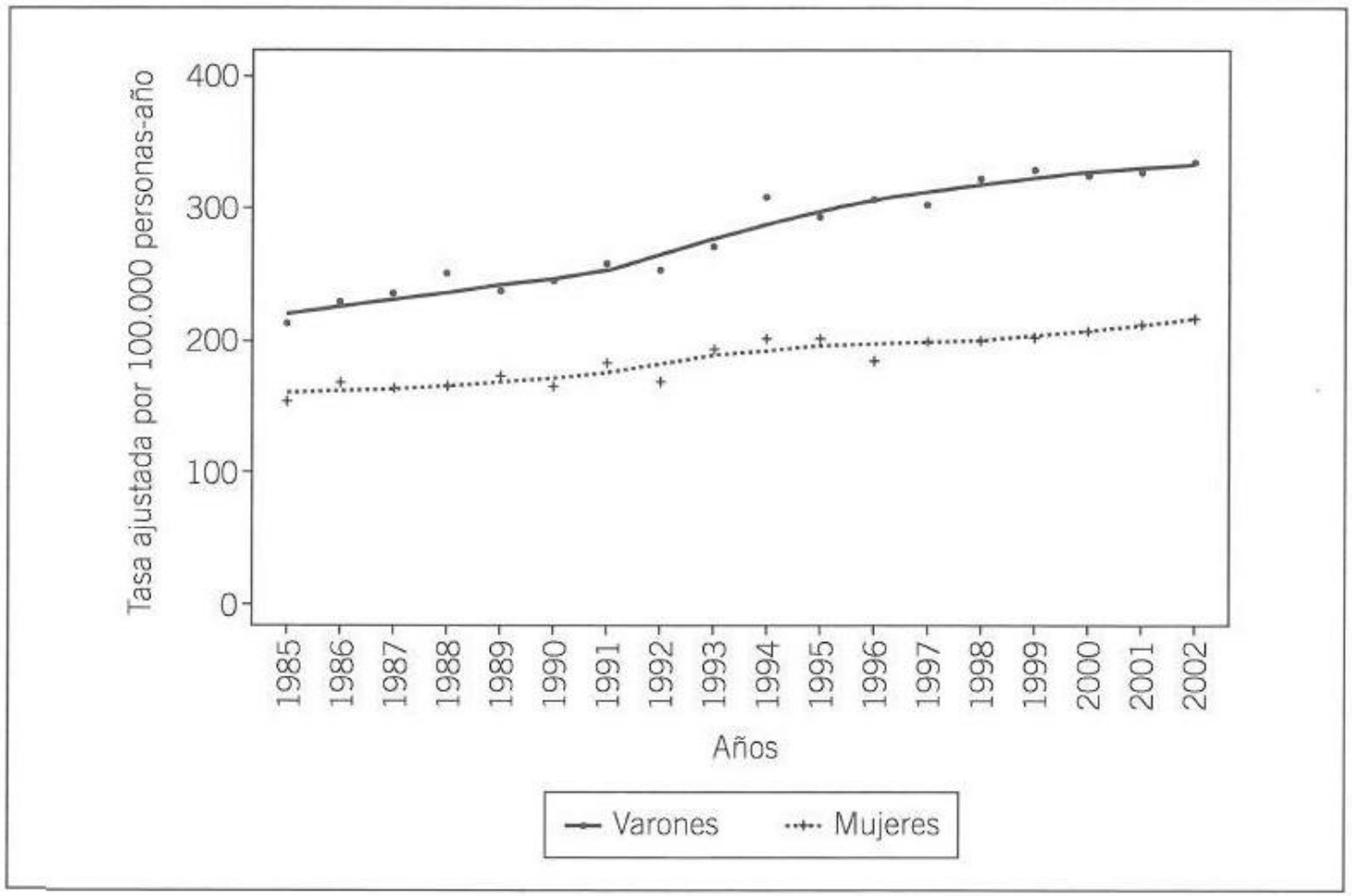
2008 World wide cancer incidence and mortality rates



Jemal et al (2001) CANCER J CLIN ;61:6990)

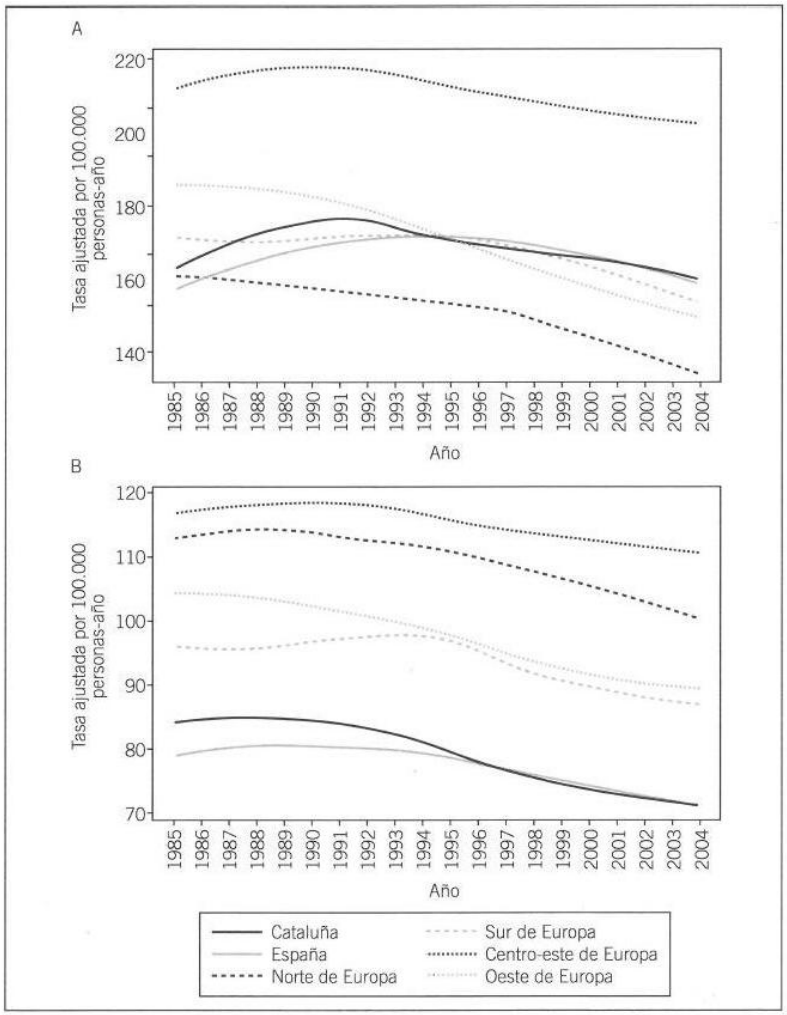
Background

2008 Catalan cancer incidence rates



Borras et al (2008) Med clin

Background
Catalan and European mortality rates



Gispert et al (2008) Med clin

Background
Predictions

- Prediction of incidence or mortality rates in the future.
  - Important both from the epidemiological and health planing point of view and to assess the burden of cancer.
  - Used to evaluate the effect of the new health programs or treatments.
  - Most used models for prediction: Age-period or age-period-cohort models (Poisson distribution)
  - Historical data to *feed* predictions.
  - Prediction out of the range of observed values.

Aims

- To evaluate the optimal number of years to make a prediction in the future for incidence or mortality rates using age-period models.

Methods

Models for rate trends

- c_{it} and n_{it} number of incident cases or deaths and number of persons at risk in the population, respectively, for *ith* age group in the t period (year). Models:
  - Model AAAP (*Additive Age-specific Age-Period*)

$$\mathrm{E}\left(\frac{c_{it}}{n_{it}}\right) = \alpha_i + \beta_i t$$

  - Model MAAP (*Multiplicative Age-specific Age-Period*)

$$\text{Model 'MAAP'} \quad \log\left(\mathrm{E}\left(\frac{c_{it}}{n_{it}}\right)\right) = \alpha_i + \beta_i t$$

  - Model MAP (*Multiplicative Age-specific Period*)

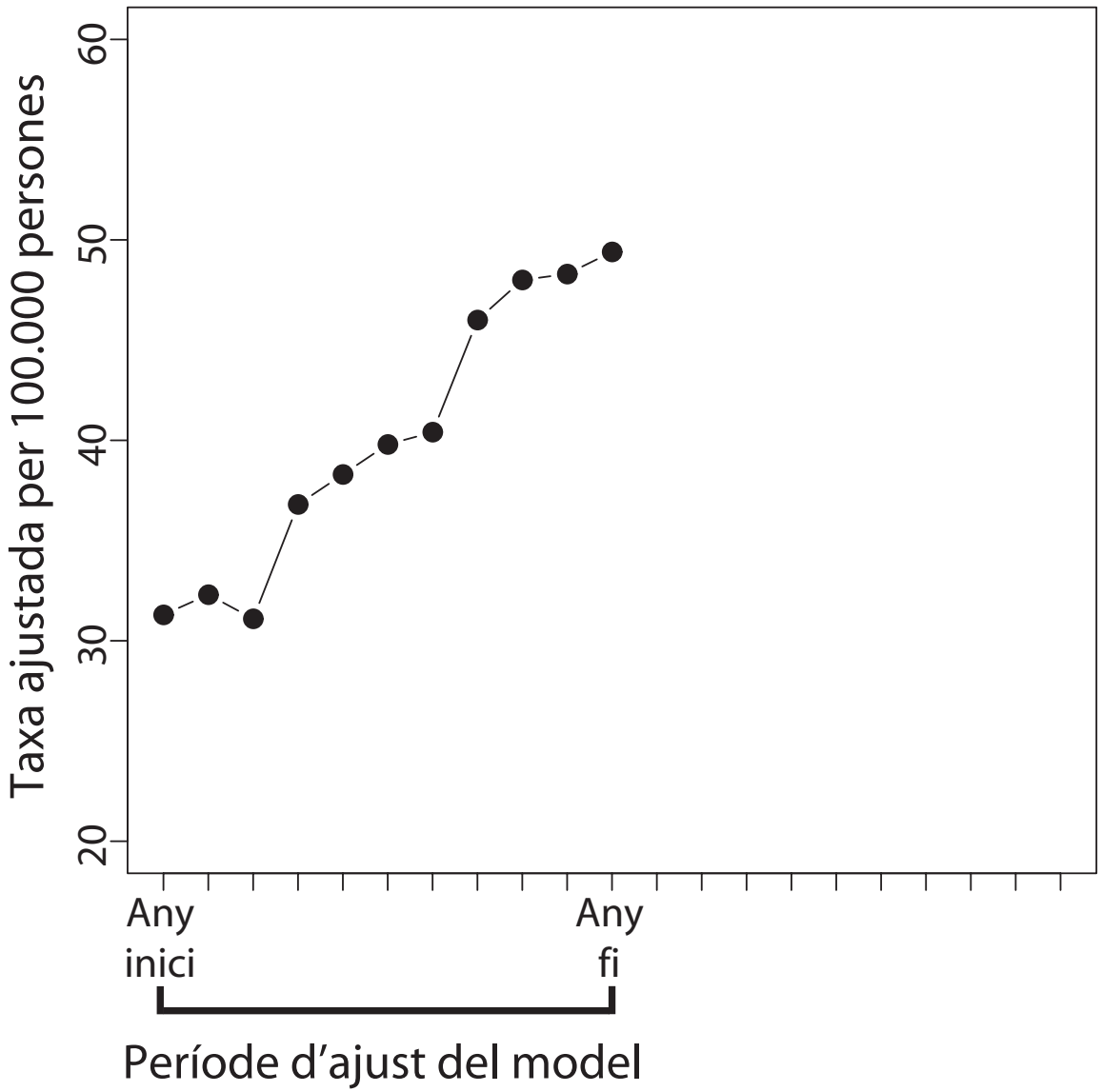
$$\text{Model 'MAP'} \quad \log\left(\mathrm{E}\left(\frac{c_{it}}{n_{it}}\right)\right) = \alpha_i + \beta t$$

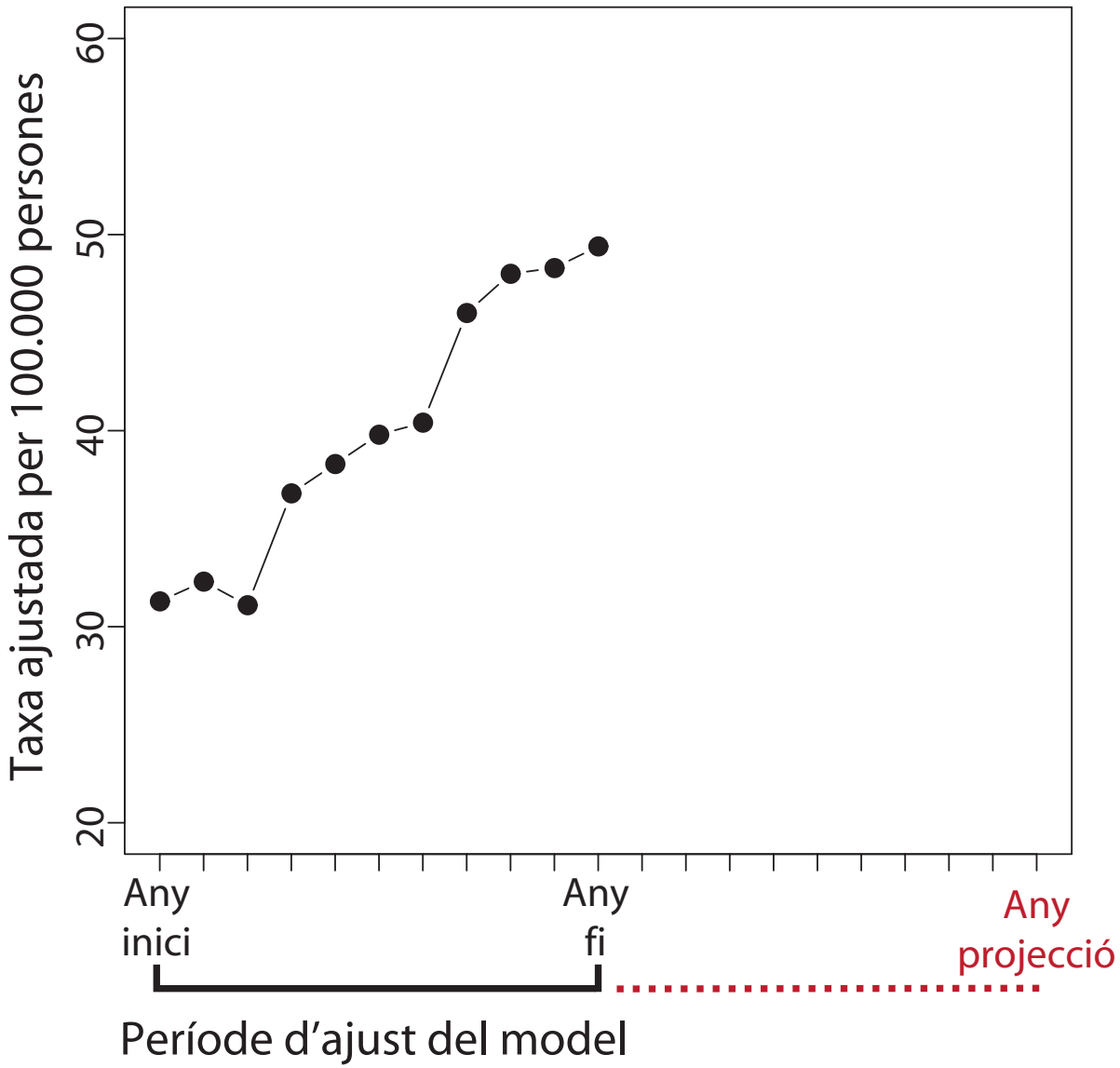
  - c_{it} Poisson distributed and fixed n_{it} .

Dyba i Hakulinen (2000) *Stat Med*, Dyba et al (1997) *Stat Med*

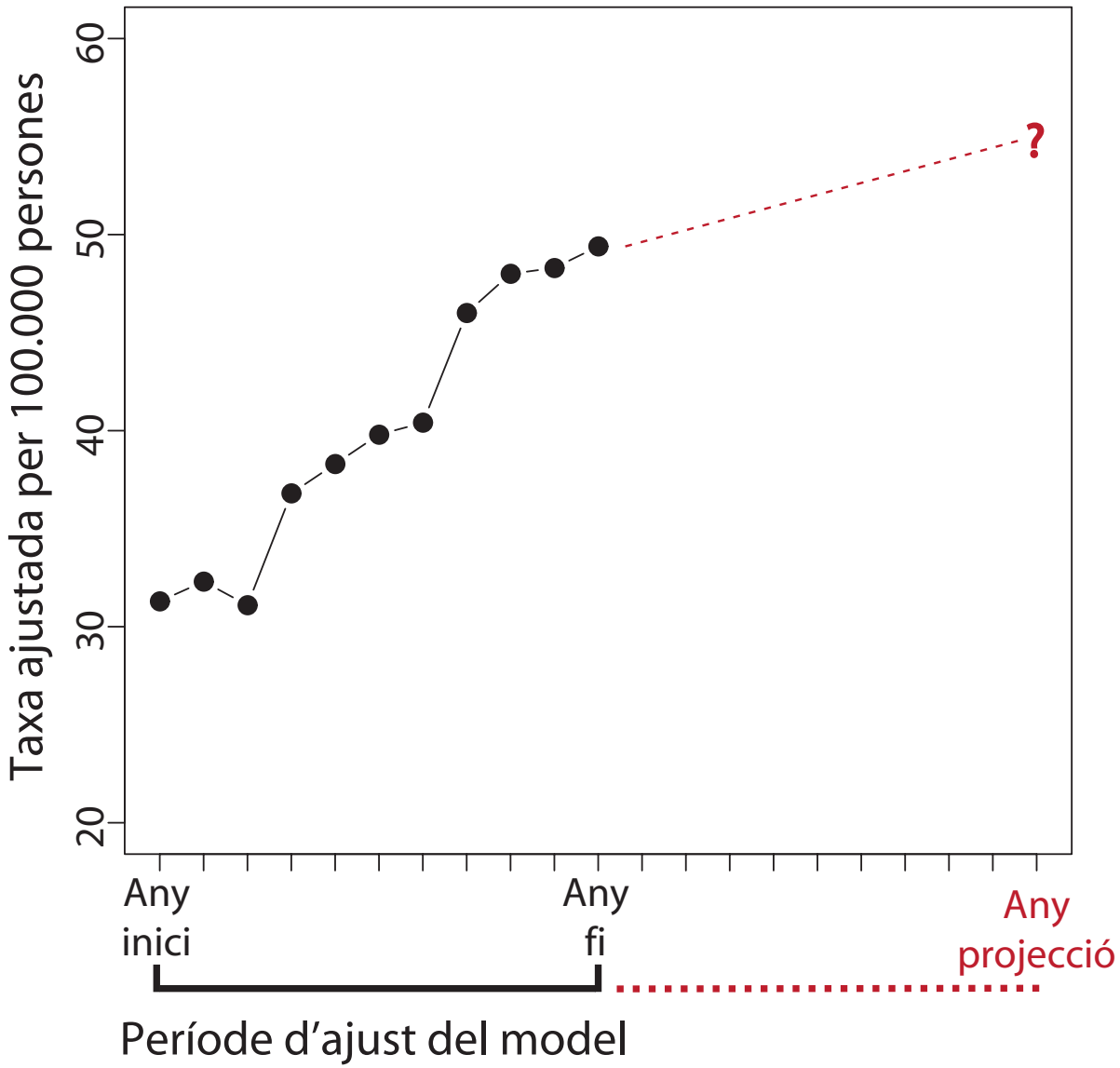
Methods

Predictions from a model



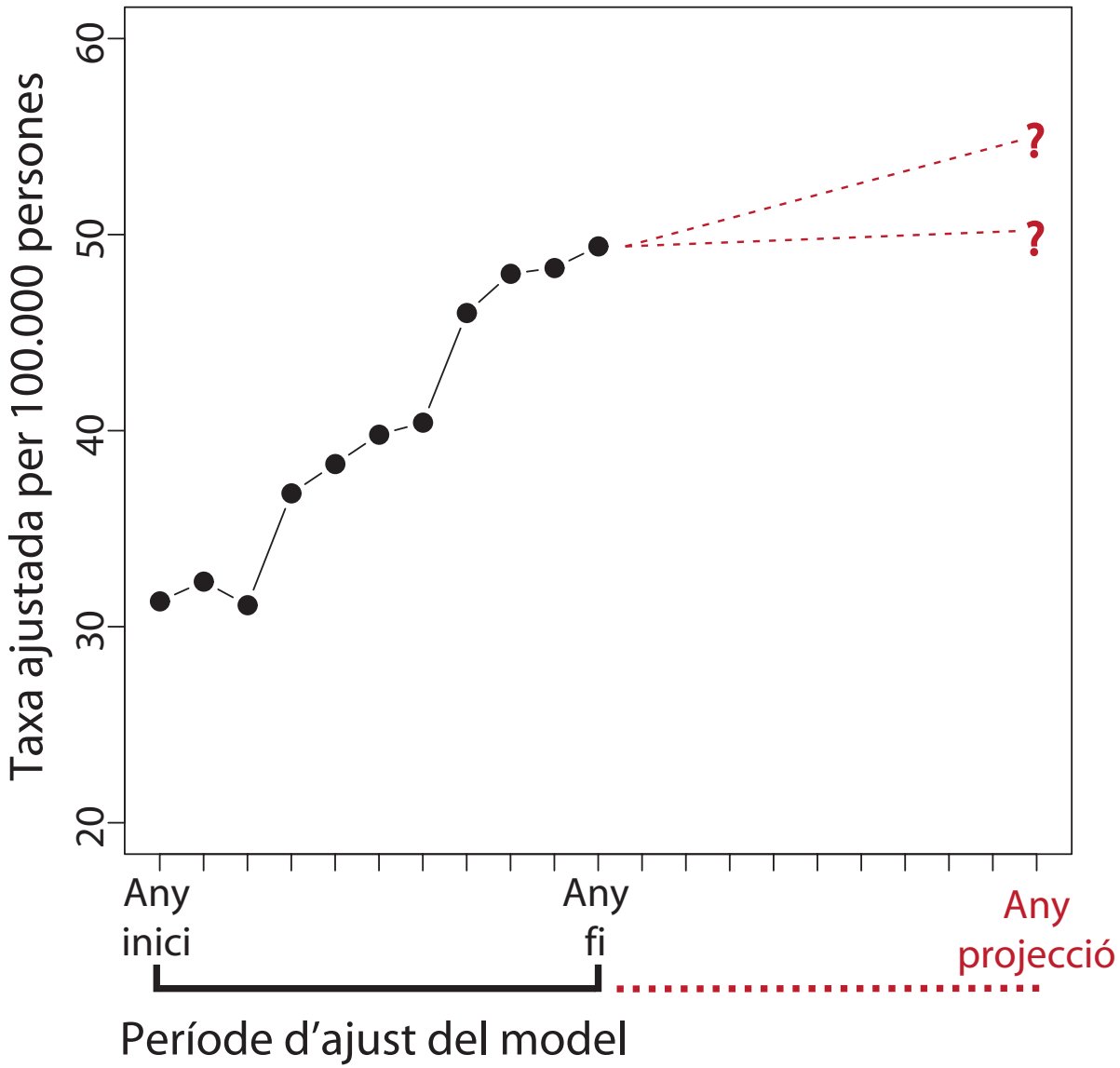


Methods
Predictions from a model



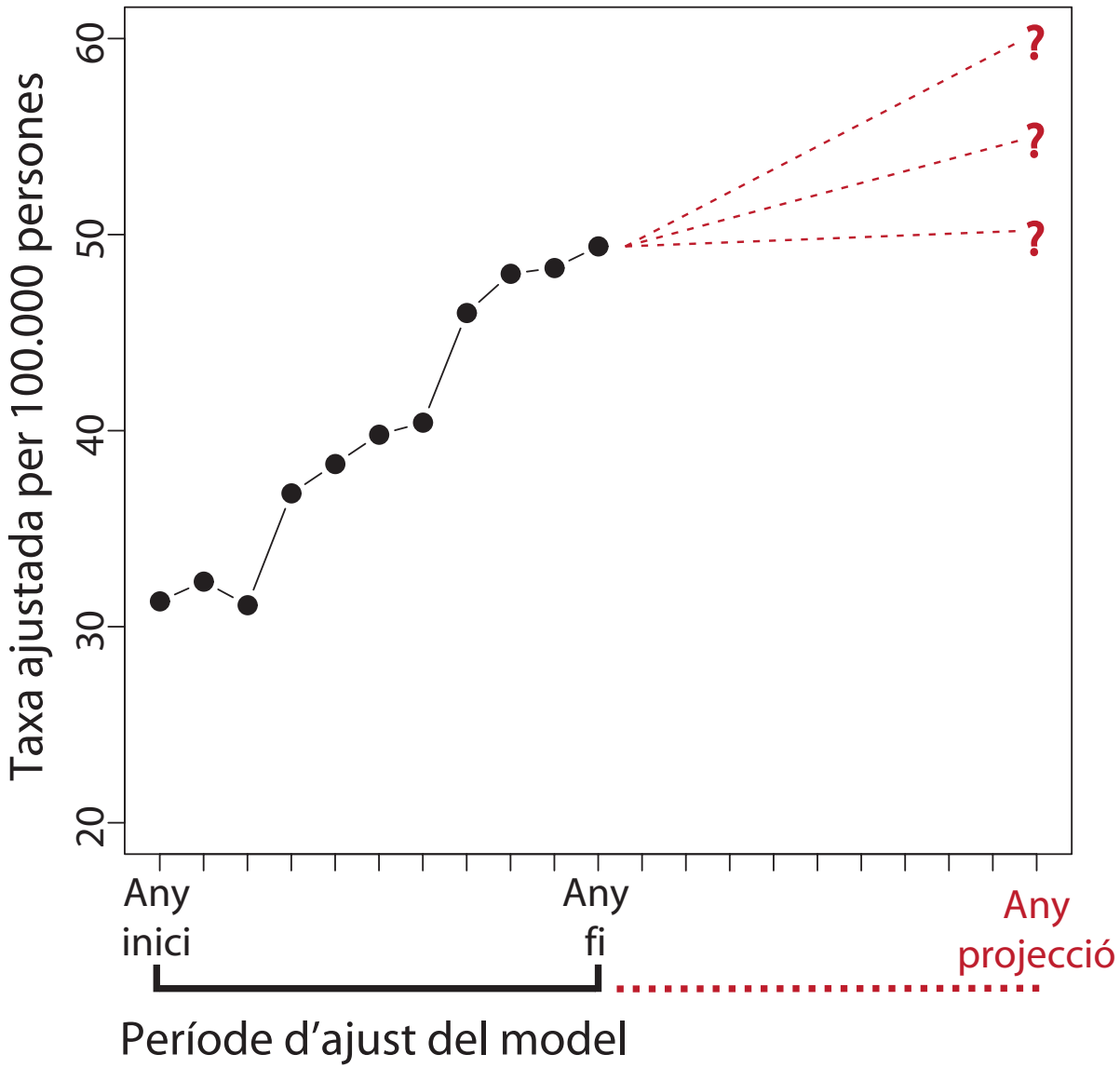
Methods

Predictions from a model



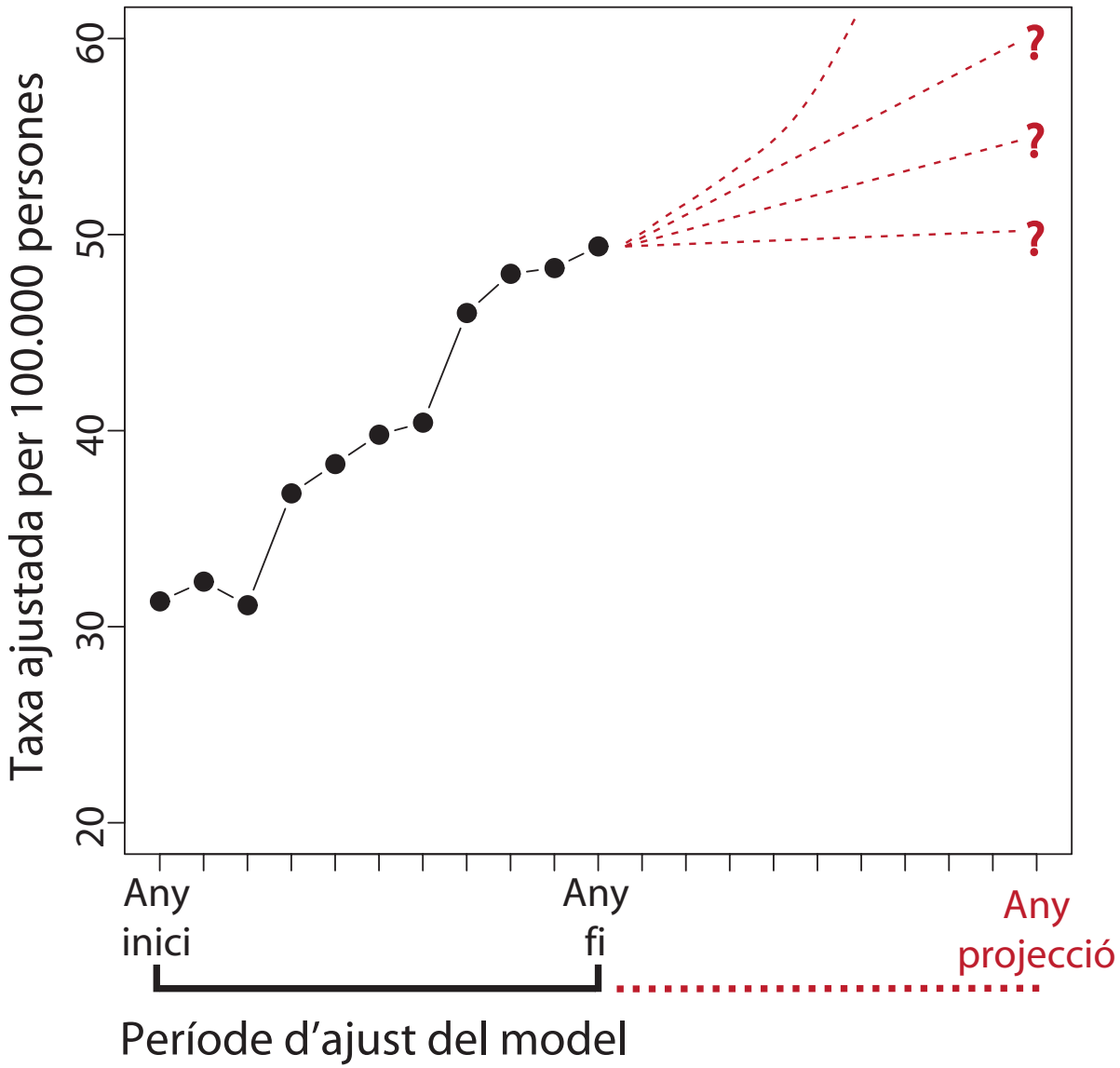
Methods

Predictions from a model

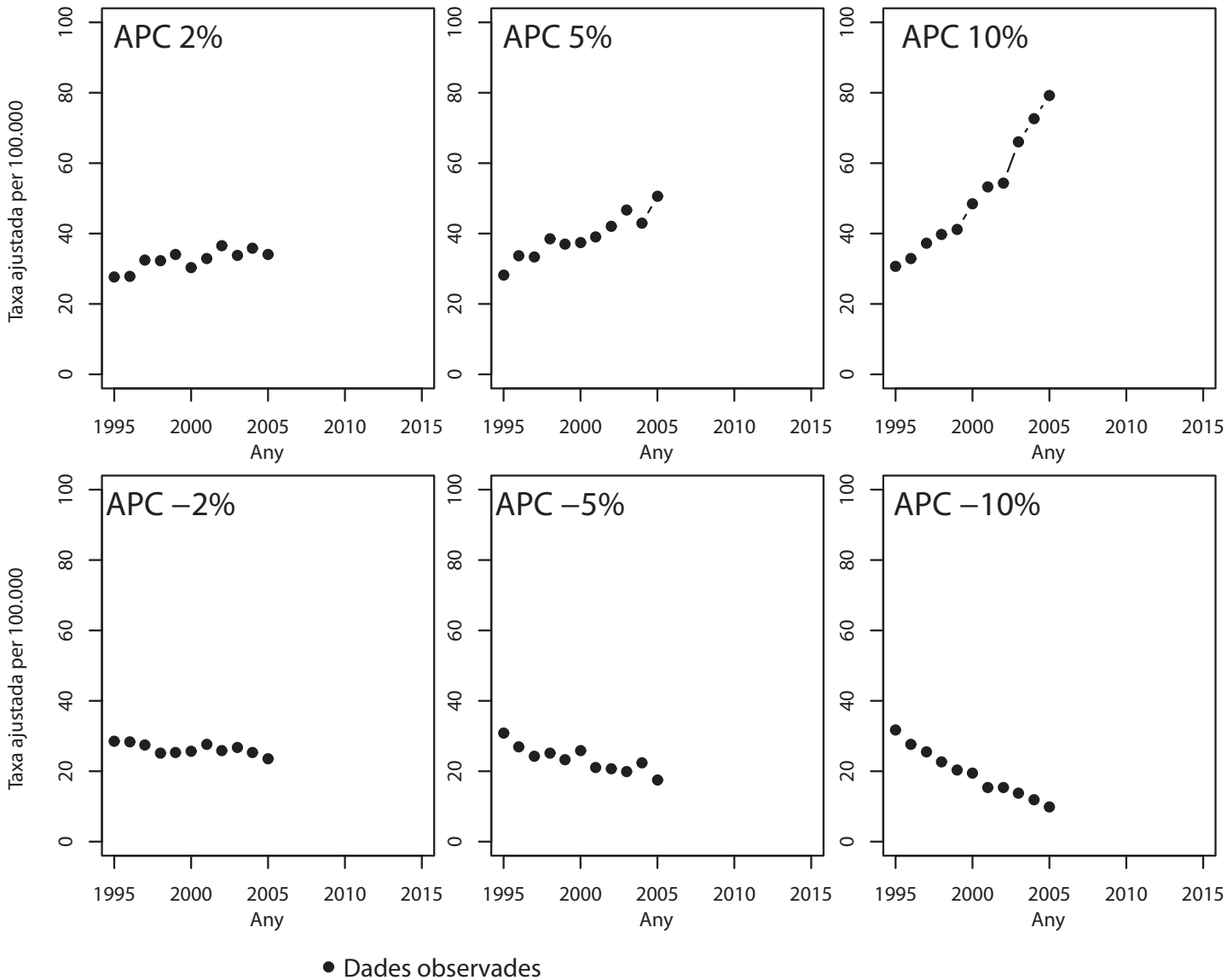


Methods

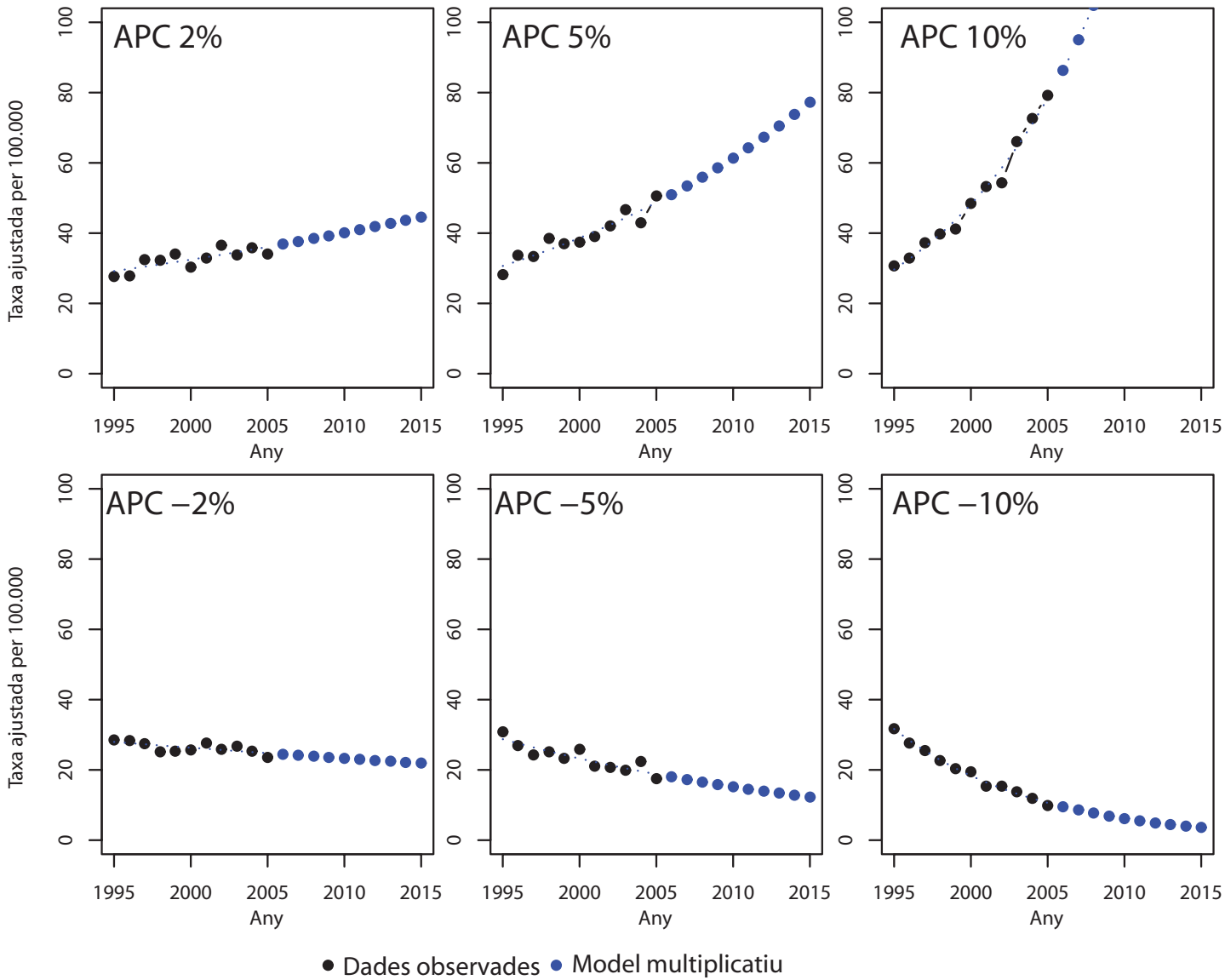
Predictions from a model



Optimal number of years to make a prediction for mortality or incidence rates

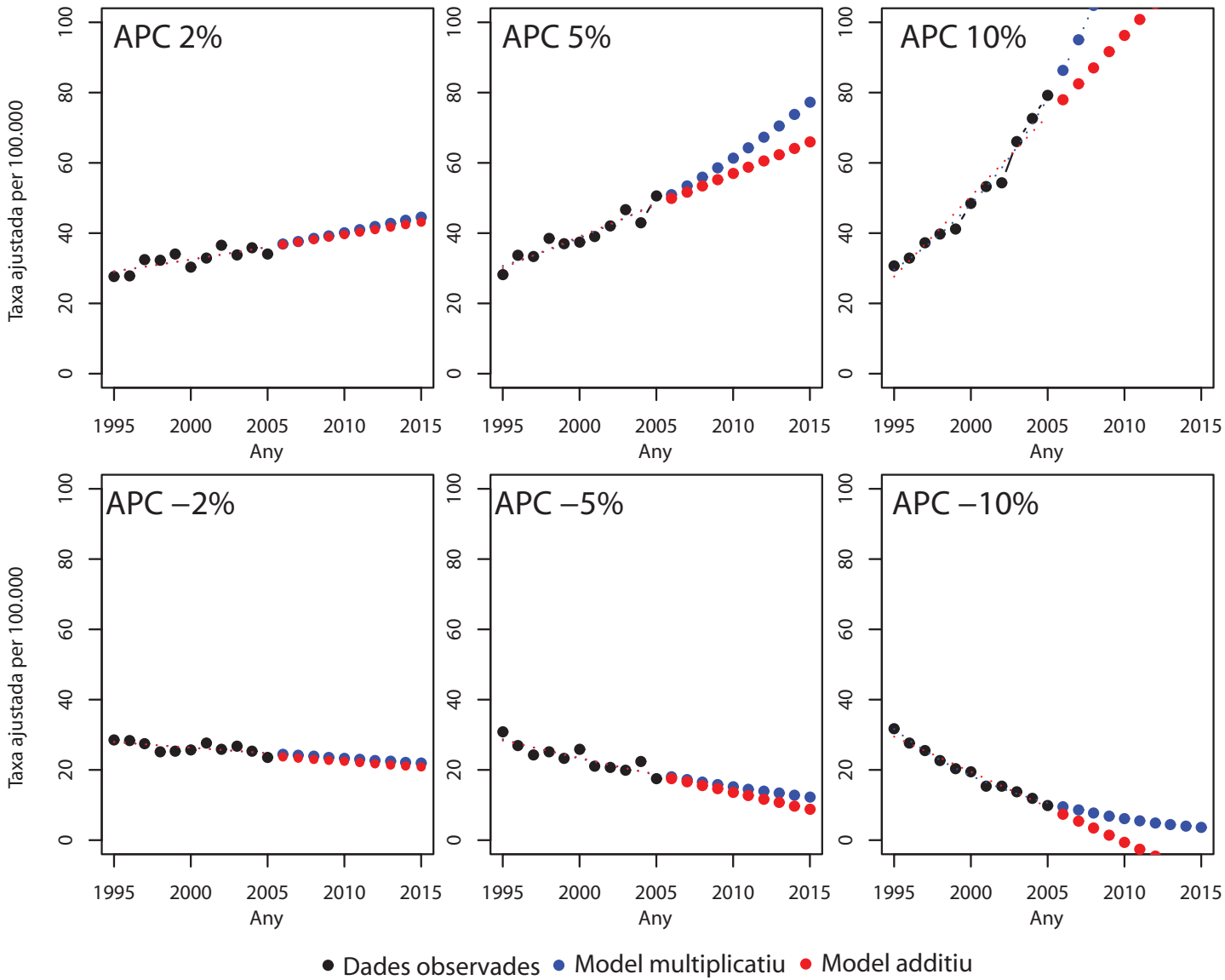


Optimal number of years to make a prediction for mortality or incidence rates



Methods
Model selection

Optimal number of years to make a prediction for mortality or incidence rates



Methods

Pearson goodness of the fit

- Pearson goodness-of-fit statistic used to assess the model

$$X^2 = \sum_{i=1}^{I} \sum_{t=1}^{T} \frac{(c_{it} - \hat{c}_{it})^2}{\mathrm{Var}(\hat{c}_{it})}$$

- Distributed as χ^2 on the residual degrees of freedom for the model

Methods
Analyses performed

- Data of mortality cancer rates in Spain from 1951 to 2011 downloaded from WHO.
- Three cancer sites for women and men:
  - Women: breast, colorectal and lung cancer.
  - Men: lung, colorectal and prostate cancer.
- AAAP model when increasing trends and MAP or MAAP models when decreasing trends.

- Prediction made for each year in 1975 up to 2011
- Prediction using data historical data 5 years before the prediction.
- Basis period for prediction: 1, 2, . . . , 20 last years

Methods
Analyses performed

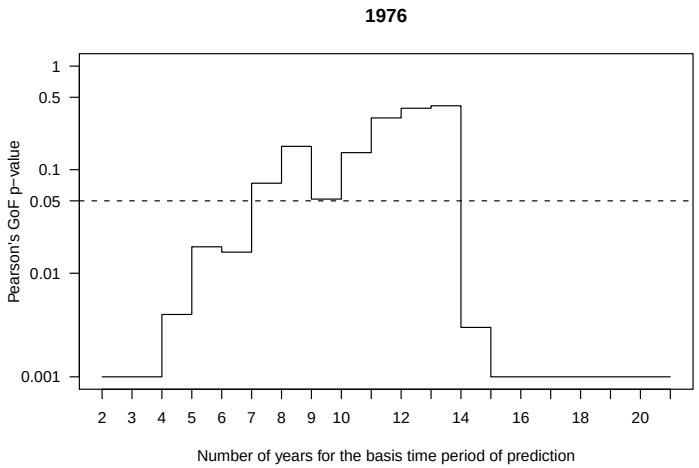
- Different basis period time for the prediction
  - Last year
  - Last 5 years
  - Last 10 years
  - Last 20 years
  - *GoF-optimal* number of years
- *GoF-optimal* number of years: smallest period of time where the model fits, according to Pearson's GoF statistis

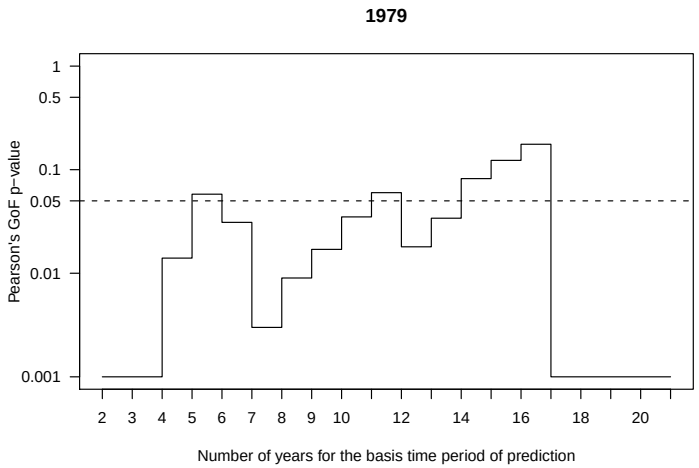
Methods
Analyses performed

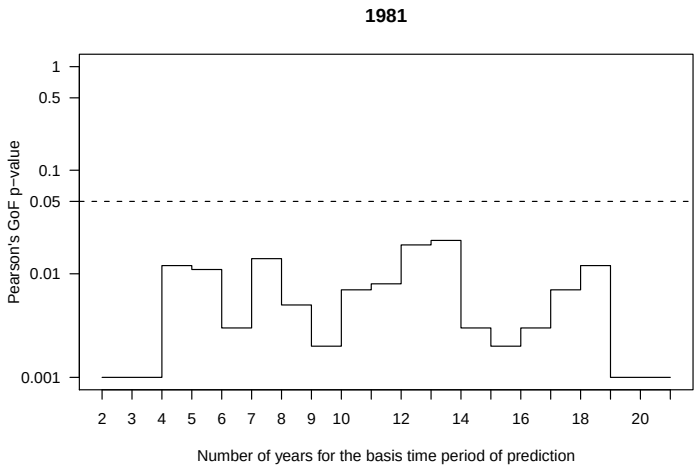
- Comparison of the different methods (1,5,10,20 last years or GoF-optimal):
  - For each year predicted (1975 up to 2011) obtention of 95% confidence interval for the predicted number of deaths.
  - Boolean evaluation of the inclusion of the true observed value respect to the interval.
  - Computation of a whole coverage (% years where the true value is inside the interval).

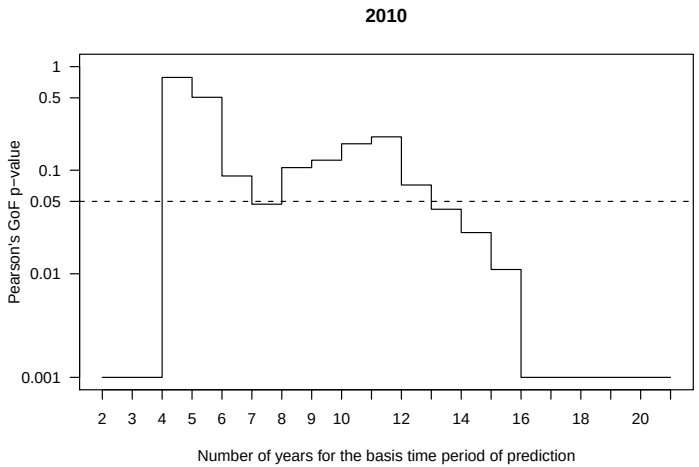
Results

Breast women p-value GoF selection





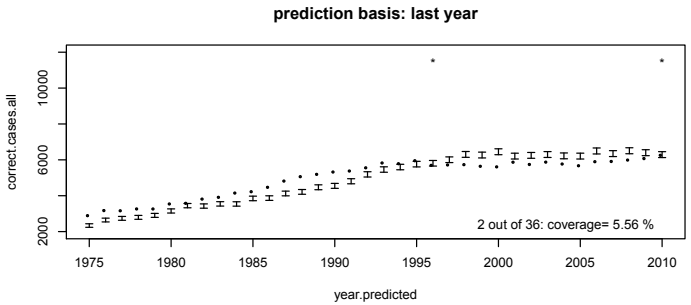


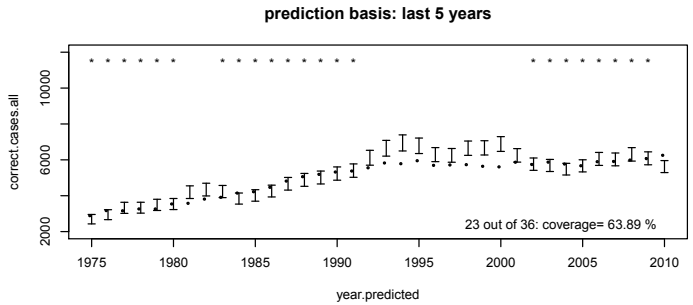


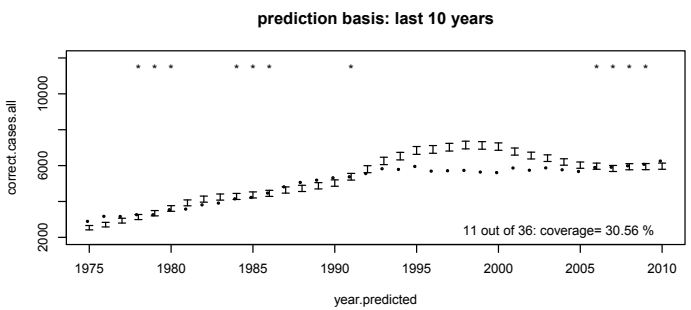
Results

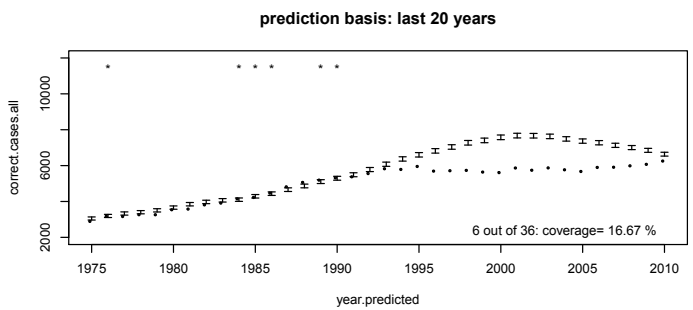
Breast women comparison methods

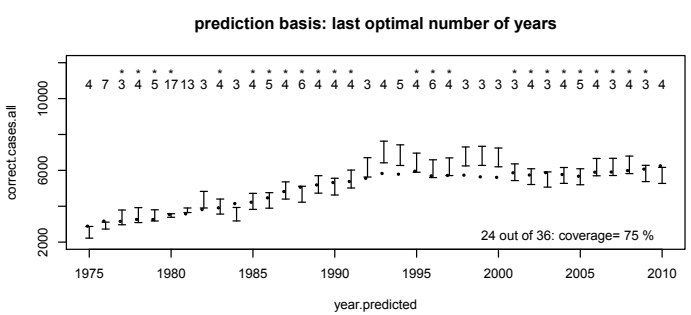
Optimal number of years to make a prediction for mortality or incidence rates







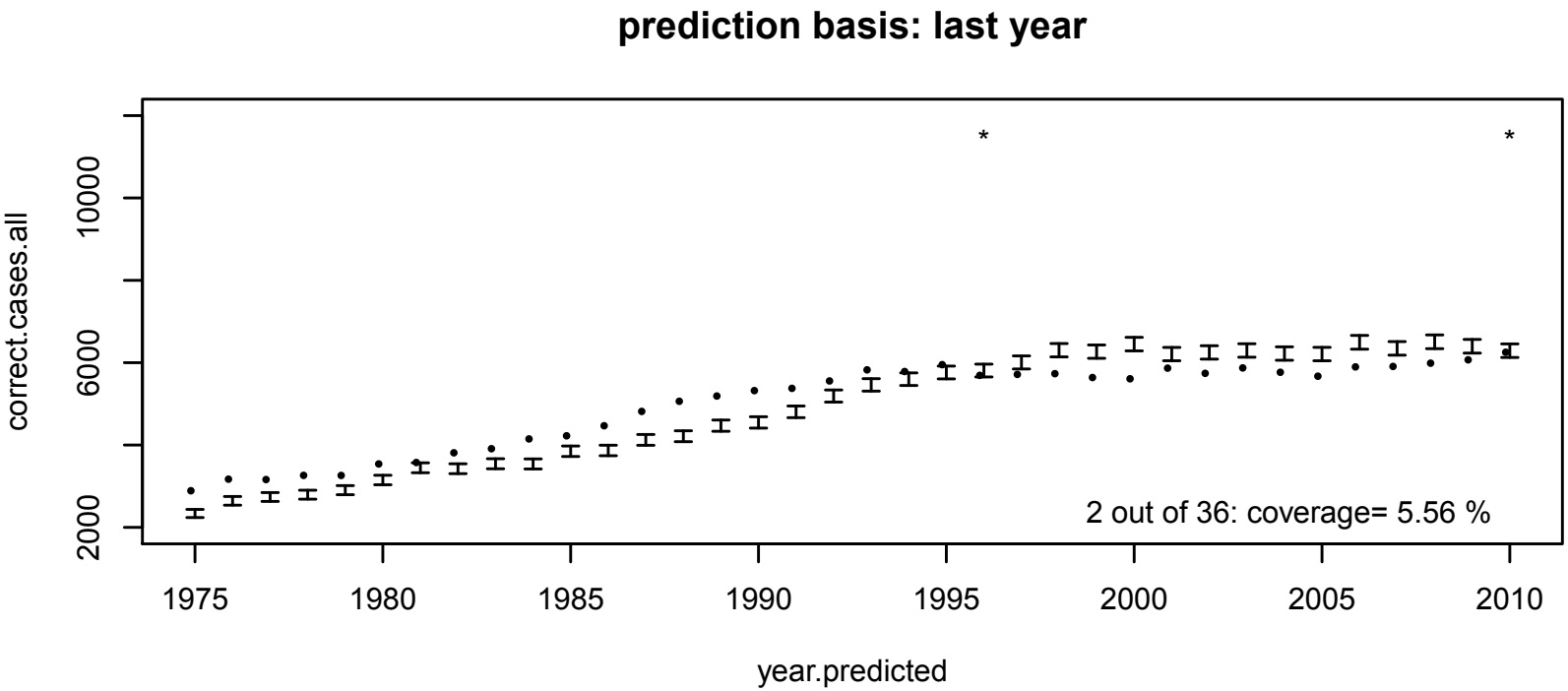




Results

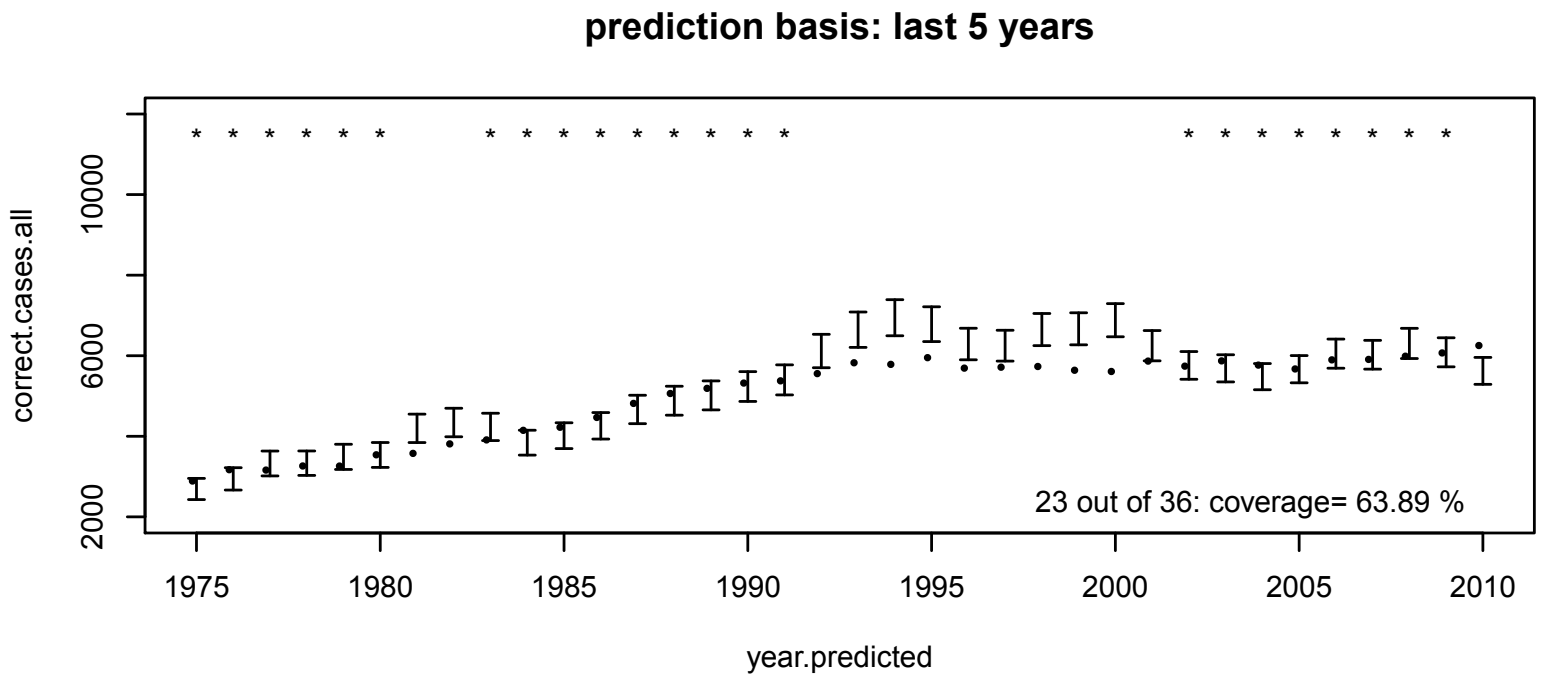
Breast women: last year

Optimal number of years to make a prediction for mortality or incidence rates



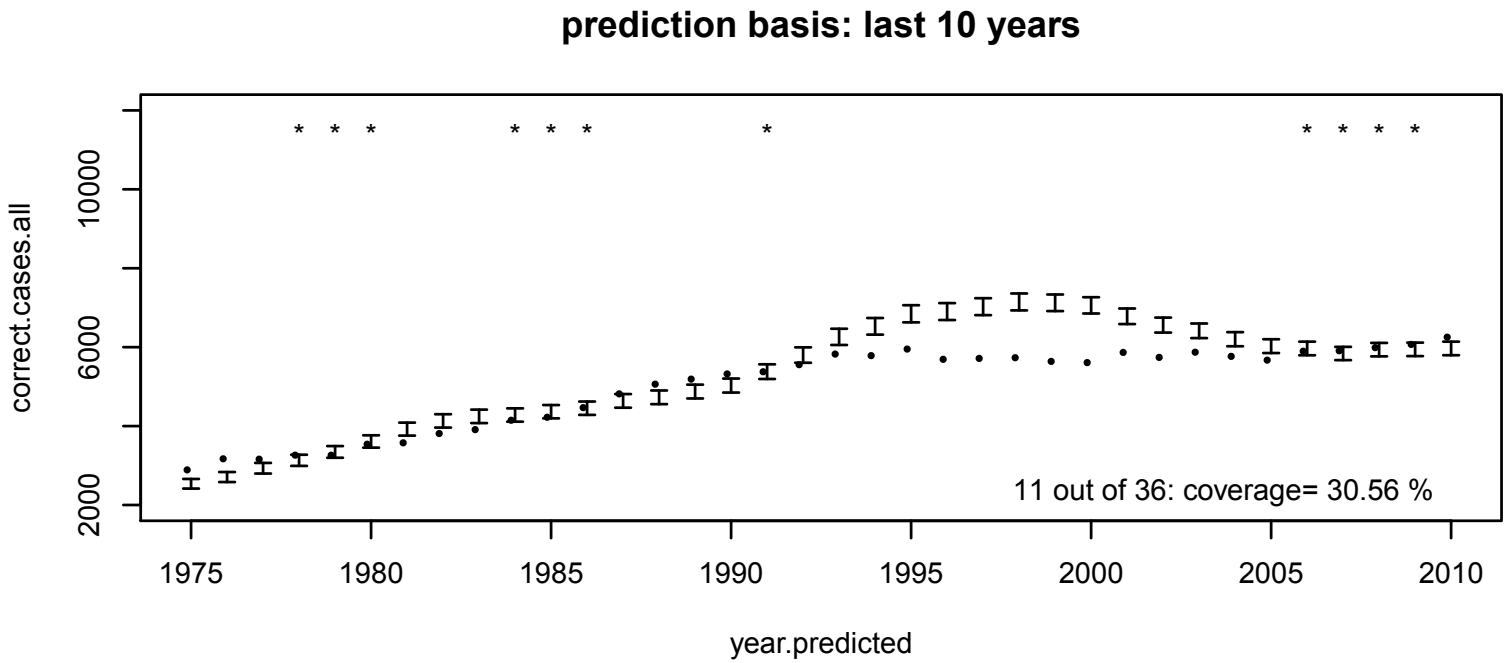
Results

Breast women: last 5 years



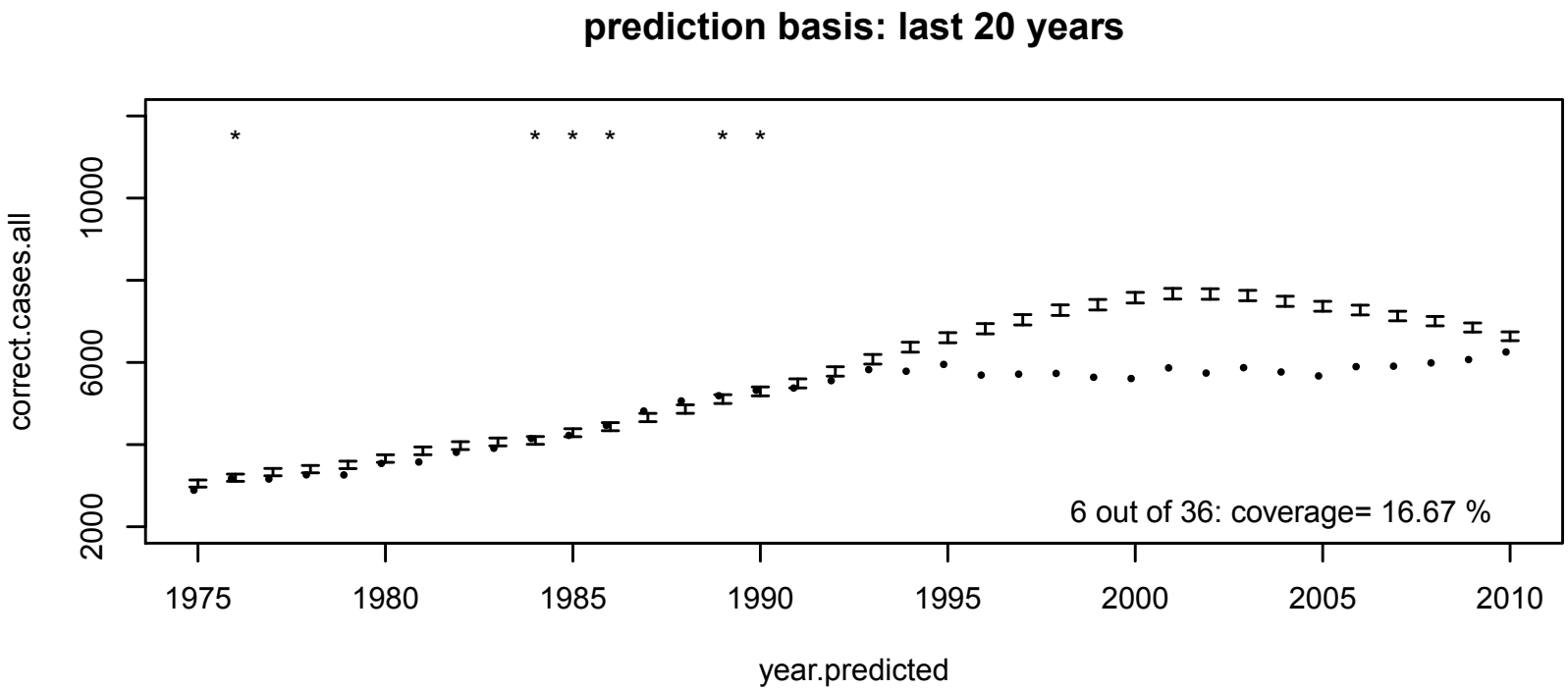
Results

Breast women: last 10 years



Results

Breast women: last 20 years

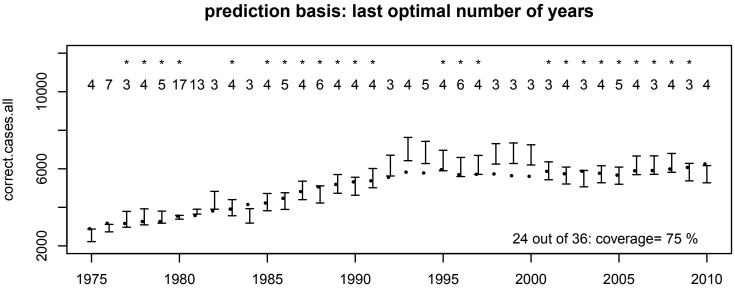


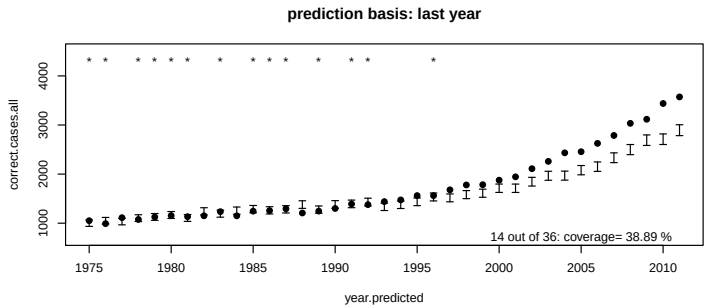
Results

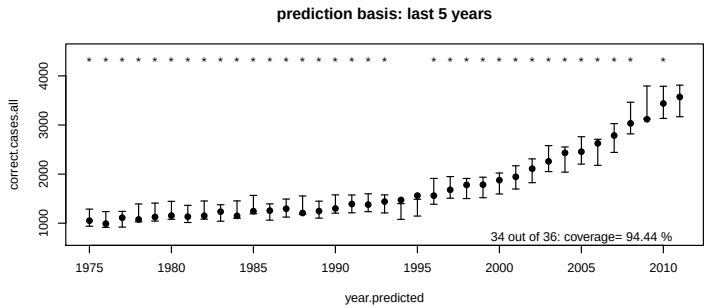
Breast women: GoF-optimal

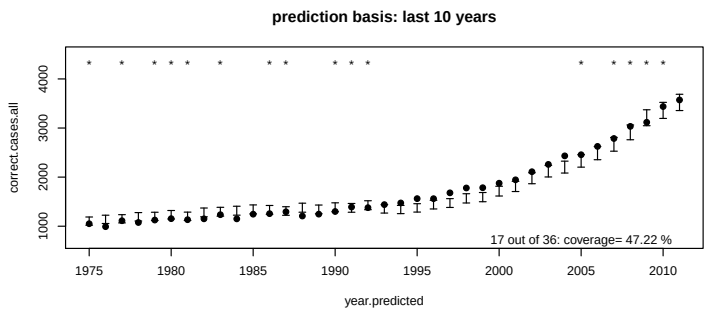
Optimal number of years to make a prediction for mortality or incidence rates

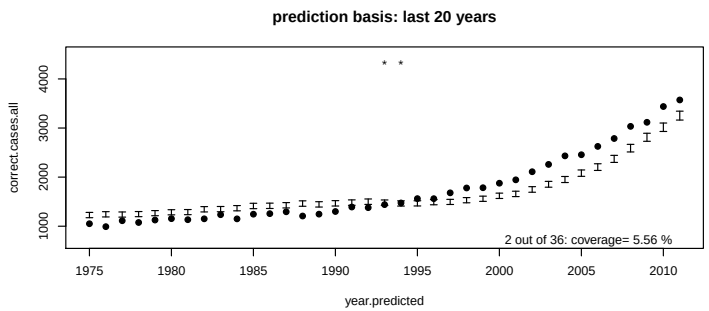
Background
Aims
Methods
Results
Discussion

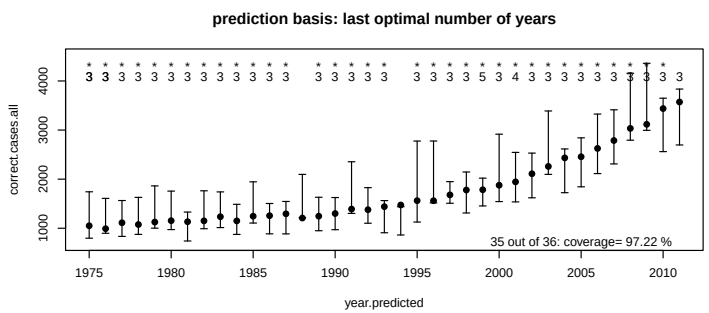












Results
All tumor sites

	Method				
Tumor site	1 year	5 years	10 years	20 years	GoF-optimal
Women					
Breast	5.56 %	63.89 %	30.56 %	16.67 %	75 %
Colorectal	27.78 %	50 %	13.89 %	2.78 %	66.67 %
Lung	38.89 %	94.44 %	47.22 %	5.56 %	97.22%

Discussion

- Simple method with reasonable results and useful at practice.
- Good alternative to choosing a subjective number of years.
- Results supporting that large historical data are not needed for a prediction.
- Limitations:
  - Ongoing results: more results with new tumor sites and countries.
  - Not possible to establish general statistical properties from particular data sets.

Thank you

Thank you for your attention!