UNIVERSITÀ DEGLI STUDI DI MILANO

Facoltà di Medicina e Chirurgia

Dipartimento di Scienze Cliniche e di Comunità

Sezione di Statistica Medica e Biometria "G. A. Maccacaro"

Corso di Dottorato di Ricerca in

STATISTICA BIOMEDICA - XXVII CICLO

Settore Scientifico Disciplinare MED/01

TESI DI DOTTORATO DI RICERCA

The impact of the widespread screening through Prostate-Specific Antigen (PSA) test on population-based models for estimating prostate cancer burden

Dottorando: Antonella Zucchetto

Tutor: prof. Monica Ferraroni

Coordinatore: prof. Adriano Decarli

A.A. 2013/2014

Essentially, all models are wrong, but some are useful.

— George E.P. Box

INDEX

List of major abbreviations

aaPC – average annual Percent Change

AIC – Akaike Information Criterion

aPC – annual Percent Change

APC – Age-Period-Cohort

CI – Confidence Interval

CRS – Cumulative Relative Survival

ES – Expected survival

FVG – Friuli Venezia Giulia

ICD – International Classification of Diseases

ISTAT – Italian national Institute of Statistics

LRS – likelihood ratio statistics (LRS)

MIAMOD – Mortality and Incidence Analysis Model

ML – Maximum Likelihood

OS – Observed Survival

PCa – Prostate Cancer

PIAMOD – Prevalence and Incidence Analysis Model

PSA – Prostate-Specific Antigen

RS – Relative Survival

RR – Relative risk, Rate Ratio

SSE – Sum of squared errors

1. BACKGROUND AND RATIONALE

1.1   Prostate cancer epidemiology

Prostate cancer (PCa) is the second most common neoplasm among men worldwide [Ferlay et al, 2013]. Out of more than 1 million new diagnoses estimated in 2012 (1,112,000 cases, 15% of cancers diagnosed in men), almost 70% occurs in more developed regions [Ferlay et al, 2013]. The incidence rates greatly vary among countries (25-fold), being highest in Australia/New Zealand and Northern America and in Western and Northern Europe. With an estimated 300 thousands deaths in 2012, PCa is the fifth leading cause of death from cancer in men (6.6% of total deaths). There is less variation in mortality rates worldwide (10-fold) than is observed for incidence, with the number of deaths from PCa being larger in less developed than in more developed regions. The five-year prevalence of PCa was estimated in nearly 4 million men worldwide in 2008 [Bray et al, 2013].

In Italy, as in most developed countries, PCa has become the first cancer diagnosed among men, with nearly 36,000 new cases estimated in 2013 (20% of all cancers among men), and the third leading cause of death from cancer (9,000 deaths among men, 8% of cancer deaths) [AIRTUM-AIOM, 2013]. After a dramatic increase of PCa incidence rates in the period 1998-2003, they were reported to be almost stable thereafter. Conversely, PCa mortality rates were constantly decreasing in the last decade [AIRTUM-AIOM, 2013].

There are only three well-established risk factors for PCa, and they are all not modifiable: older age (PCa is very rare in men younger than 40 years and the risk rapidly increases after age 50), black race/ethnicity (PCa occurs more often in African-American men and Caribbean men of African ancestry), and a family history of the disease (PCa risk is much higher for men with several affected relatives, particularly if they were young at the time of cancer diagnosis) [Leitzmann & Rohrmann, 2012]. Modifiable factors, such as diet (*e.g.*, high intakes of red meat or high-fat dairy products),

obesity, and pattern of sexual behavior might also be involved in the development of PCa [Leitzmann & Rohrmann, 2012].

1.2   Prostate-Specific Antigen test

Prostate-specific antigen (PSA) is a protein produced in the prostate gland. PSA is mostly found in semen, which is produced in the prostate, but small amounts of PSA ordinarily circulate in the blood. High levels of PSA in the blood may indicate the presence of PCa. PSA testing has been shown to increase PCa detection by 81% in comparison with digital rectal examination alone [Catalona et al, 1994]. However, many other conditions, such as an enlarged (*i.e.*, benign prostatic hyperplasia) or inflamed prostate (*i.e.*, prostatitis), can also increase circulating PSA levels.

PSA testing for screening of PCa was introduced in many high-income countries between the mid-1980s and the early 1990s and it represents a unique situation in which a widespread use of the test at a population level has occurred long before of definitive results about its efficacy. As a result, dramatic increases followed by sharp reductions in PCa incidence were observed in the United States (US), Canada, and Australia [Center et al, 2012]. Stabilizing PCa incidence trends for the last decade were primarily observed for these same countries. In other developed countries, such as in northern and western Europe, gradually increasing PCa incidence trends have yet to yield a dramatic peak [Center et al, 2012]. Conversely to incidence trends, PCa mortality rates have been decreasing in most high-resource settings. Improvements in treatments in the 1990s (including radical prostatectomy, radiation therapy, and hormone therapy) coupled with an increased detection of early-stage PCa as a result of PSA testing appear to be reasonable explanations for the declining mortality trends observed in many developed countries [Collin et al, 2008; Etzioni et al, 2008; Baade et al, 2004].

However, the specific role of PSA testing in explaining these favorable recent declines in PCa mortality continues to be debated, particularly given the downward trends observed also in countries where the prevalence of PSA testing was reasonably low (*e.g*., the United Kingdom, UK)

[Gavin et al, 2004] or across areas with very heterogeneous PSA use [Etzioni et al, 2008]. In addition, a recent Cochrane meta-analysis did not find a statistically reduced mortality risk when including the results of five randomized trials [Ilic et al, 2013]. In particular, the findings of the two largest randomized trials on the efficacy of screening with PSA test were not in agreement. The "Prostate, Lung, Colon, and Ovary screening trial" (PLCO) conducted in the US, after 13 years of follow-up, found no evidence of a mortality benefit for organized annual screening compared with opportunistic screening, which is part of usual care (rate ratio, RR=1.09; 95% confidence interval, CI: 0.87-1.36) [Andriole et al, 2012]; conversely, the PLCO study found a relative increase of 12% of cumulative incidence in the intervention arm (RR=1.12; 95% CI: 1.07-1.17). On the other side, the "European Randomized Study of Screening for Prostate Cancer" (ERSPC) found a reduction around 20% in PCa mortality attributable to PSA-testing, confirmed also after 13 years of follow-up by the recently published results (RR=0.79; 95% CI: 0.69-0.91) [Schröder et al, 2014]. However, the corresponding incidence rate increase attributable to screening was approximately 60% (RR=1.57; 95% CI: 1.51-1.62). The ERSPC reported also that the absolute risk reduction of death from PCa was equivalent to one PCa death avoided per 781 (95% CI: 490-1929) men invited for PSA screening, or one per 27 (95% CI: 17-66) additional PCa diagnoses.

Further quantification of harms and benefits of PSA testing use are still needed to decide whether to introduce organized screening at a population level. Actually, even though PSA testing and subsequent treatments do contribute to the observed declining mortality, the harms to benefits ratio remains controversial because of adverse events. PSA testing can detect cancers that may otherwise go undiagnosed during a man's lifetime (*i.e.*, overdiagnosis and, consequently, overtreatment) and treatment of PCa are serious and potentially life-altering with significant risks of sexual, urinary, and bowel-related symptoms. It has been estimated that 23–42% of screen-detected PCa cases in the US result from overdiagnosis due to PSA testing [Draisma et al, 2009]. In the European setting, estimates of overdiagnosis are considerably higher reaching 66% of screen-detected tumors [Draisma et al, 2003]. This result is consistent with a lower baseline incidence of PCa in Europe, a

lower PSA cut-off for biopsy referral (3 ng/mL in several European countries *vs.* 4 ng/mL in the US), and a much higher frequency of compliance with biopsy referral in Europe than in the US [Schröder et al, 2009; Pinsky et al, 2005].

As a consequence of the most recent evidences from the literature, motivated largely by the results of the PLCO and the ERSPC trials, updated clinical guidelines for PSA testing use are periodically released by the American and the European associations of urology. However, these guidelines are not always in agreement. The 2013 American Urological Association guidelines [Carter et al, 2013] do not recommend PSA screening in men below 40 years of age, do not recommend PSA screening in men aged 40-54 years and at average risk, recommend shared decision making for men aged 55-69 years, and do not recommend PSA screening in men >70 years of age or in men with a life expectancy lower than 10-15 years. Conversely, the 2013 update of the European Association of Urology guidelines [Heidenreich et al, 2013], although not recommending widespread mass screening for PCa, strongly recommends screening in men with a life expectancy >10 years (irrespective of age) and a baseline PSA determination at 40-45 years of age.

The response to these recommendations in terms of the clinical practice is evolving. In order to evaluate the impact of the new PSA screening guidelines at a population level, Gulati and colleagues [Gulati et al, 2014] predicted incidence and mortality rates of PCa in the US for men aged 50-84 years in the period 2013-2025. This study used two microsimulation models of PCa natural history (*i.e.*, statistical representations of disease progression, detection, treatment, and survival which simulate, for each man in a population, age at PCa onset, age/stage at diagnosis, etc.), previously developed in the framework of the Cancer Intervention and Surveillance Modeling Network (CISNET) consortium, reconstructed PSA screening patterns in the US, and incidence data from the Surveillance, Epidemiology, and End Results (SEER) program [Etzioni et al, 2008; Tsodikov et al, 2006]. Assuming a survival benefit of PSA screening consistent with the ERSPC trial, the study compared the effects of continuation of recent PSA screening rates *vs.* continuation only for men aged <70 years, or discontinued screening for all men: continuing PSA screening for

10

all men will result in 710,000-1,120,000 overdiagnoses in the US (out of approximately 3,800,000 total cases), but will avoid 36,000-57,000 deaths; whereas, continuing screening only for men aged <70 years will prevent 64%-66% of overdiagnoses but will fail to prevent 36%-39% of avoidable deaths; in contrast, discontinued screening in all ages will eliminate all overdiagnoses but will more than double metastatic cases [Gulati et al, 2014].

As the number of –prevalent– men with PCa increases, additional resources are needed to screen, investigate, biopsy, treat, and follow these patients. To adequately plan heath care resources and cancer control policies, reliable estimates of the number of new PCa diagnoses are required, along with reliable future projections of morbidity and mortality indicators for PCa at a population level. Estimates of PCa based solely on changes in the age distribution of the population or on historical trends do not take into account of changes in screening activities. A Canadian study [Quon et al, 2011] estimated that the number of PCa cases will triple from 2009 to 2021 using a simple additive model incorporating assumption (derived from trials) on population aging, increasing PSA screening, lowered PSA threshold for biopsy, and improved biopsy sensitivity, but not taking into account of other important factors (*e.g.*, cohort effects). Even complex simulation modeling were designed to translate the results of screening trials into population settings in order to estimate the impact of PSA diffusion not only in terms of PCa incidence, but also of overdiagnosis and harm-benefit indicators [Draisma et al, 2009; Draisma et al, 2003; Gulati et al, 2011; Etzioni et al, 2002]; however, screening patterns operating in real populations are quite different from trials' results. In order to capture this higher complexity, Tsodikov et al. [Tsodikov et al, 2006] used a simulation model to predict the effect of PSA screening directly from population databases and cancer registries data but the model was not verified by the data, due to the lack of real information on PSA testing.

Herein, widely used population-based methods for estimating and projecting standard cancer morbidity and mortality indicators will be applied in the population of Friuli Venezia Giulia region (northeastern Italy) to PCa incidence, mortality, and complete prevalence. Taking advantage of the

availability of both a population-based cancer registry and of a digital health archive with complete coverage of the resident population, data on observed PCa cases and data on PSA testing use will be analyzed in order to better understand the impact of PSA diffusion in this real population.

2. THE MIAMOD/PIAMOD METHODS FOR ESTIMATES AND PROJECTIONS OF CANCER BURDEN

The Mortality and Incidence Analysis Model (MIAMOD) and Prevalence and Incidence Analysis Model (PIAMOD), developed by Verdecchia and colleagues (1989 and 2002), are population-based methods for estimating main epidemiological indicators of morbidity and mortality for chronic degenerative diseases. They are currently widely applied in Europe and in the US.

The choice between the two methods depends on data availability: MIAMOD allows to estimate incidence and prevalence using official statistics on disease-specific mortality; PIAMOD allows to estimate mortality and prevalence when incidence data from disease-registries are available.

2.1 The transition rate method

The MIAMOD/PIAMOD methods are based on the assumption that the natural history of chronic irreversible diseases can be considered as a sequence of statistically independent transitions.

Assume for chronic degenerative diseases that the morbid process is irreversible (*i.e.*, an individual who becomes ill at a certain time will remain ill until death). In this setting, transition rate equations allow to link mortality and prevalence to incidence and survival in a unified framework.

Consider the model of Figure 2.1 [Verdecchia et al, 1989], with two live states (*i.e.*, healthy and diagnosed with a specific disease) and two death states (*i.e.*, death from the specific disease or death from all other causes). For a specific age x, $\mu(x)$ represents the disease hazard for healthy people, $\alpha(x)$ the death hazard from all causes together, $\gamma(x)$ the death hazard from the specific disease, $\beta(x,y)$ the all-cause death hazard for people who became ill at age y, and $\delta(x,y)$ the specific-cause death hazard for people who became ill at age y.
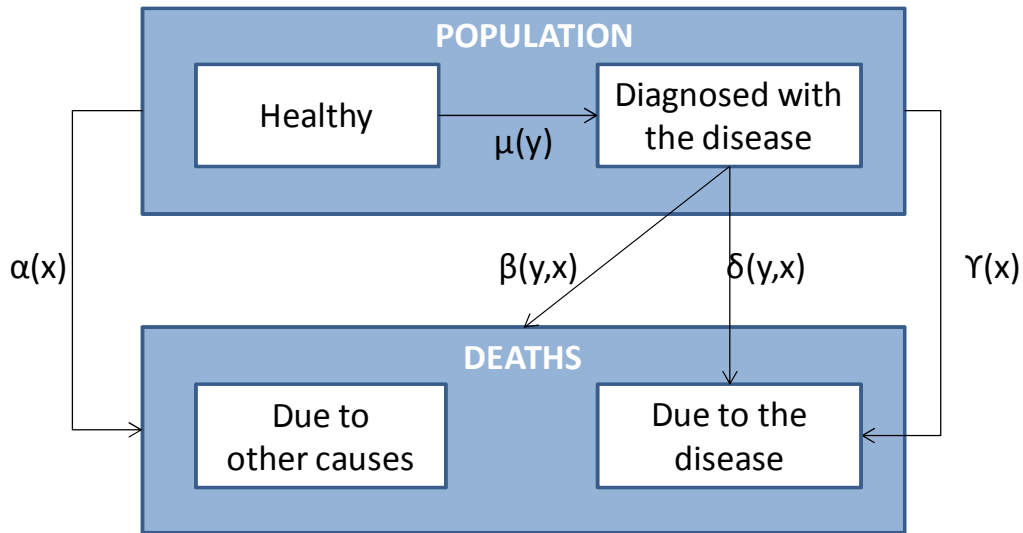
Figure 2.1. A compartmental representation of an irreversible disease-death process (y=age at diagnosis, x=current age)

Usually, $\alpha(x)$ and $\gamma(x)$ are known from official statistics; $\beta(x,y)$ and $\delta(x,y)$ can be also generally derived from epidemiological sources and will be assumed as known in the following; $\mu(x)$ can be known only if a specific-disease registry exists. However, whereas incidence and mortality can be directly derived from collected data, survival and prevalence can only be derived from incidence and mortality data. In particular, prevalence estimates derived from disease registries data are always partial, as they do not include cases occurring before the start of the registration activities. Two models have been developed in this framework, according to data availability.

- MIAMOD is a regression of mortality on observed mortality data (*e.g.*, from official statistics) to back-calculate age-period-cohort incidence model [Verdecchia et al, 1989].

- PIAMOD is a direct regression of age-period-cohort incidence model on observed incidence data (*e.g.*, from Cancer Registries) [Verdecchia et al, 2002].

2.2   Estimating cancer prevalence

For a birth cohort, prevalence at age x, $P(x)$, is the probability of being alive at age x with a past diagnosis of cancer at any previous age (y < x). It is obtained summing up, over age at diagnosis y all the specific-duration prevalence proportions for the same cohort.

Specific-duration prevalence:

$$P(x, x - y) = \frac{ES\,(0,y)\,[1 - P(y)]\,\mu(y)\,OS(y,x)}{ES(0,y)\,ES(y,x)} = [1 - P(y)]\,\mu(y)\,RS(y,x)$$

Complete prevalence: $\qquad P(x) = \sum_{y=0}^{x-1}[1 - P(y)]\,\mu(y)\,RS(y,x)$ (1)

where

$P(0) = 0$, *i.e.,* people are healthy at birth

$1 - P(y)$ is the proportion of healthy people at age y

$\mu(y)$ is the probability of being diagnosed at age y, or incidence at age y

$ES(y,x)$ is the survival of the general population between age y and x

$OS(y,x)$ is the observed survival of the patients between age y at diagnosis and age x

$RS(y,x) = \frac{OS(y,x)}{ES\,(y,x)}$ is the relative survival of the patients between age y at diagnosis and age x.

Equation (1) gives the estimated age-specific prevalence probability for a birth cohort, provided that the disease incidence and patient survival are known. A system of equations (1), including one equation for each birth cohort, allows to reconstruct cross-sectional prevalence series for an entire observation period.

2.3   Estimating cancer mortality

For a birth cohort, mortality for cancer at age x, $M(x)$, is the probability of dying of the specific cancer at age x. It is obtained summing up, over age at diagnosis y, all the specific-duration death probabilities for the same cohort.

Specific-duration death:

$$M(x, x - y) = P(x, x - y)\,\delta(y, x) = [1 - P(y)]\,\mu(y)\,RS(y,x)\,\delta(y,x)$$

Mortality: $\qquad M(x) = \sum_{y=0}^{x}[1 - P(y)]\,\mu(y)\,RS(y,x)\,\delta(y,x)$ (2)

where $\delta(y,x)$ represents the crude probability of dying at age x for the specific cancer, having being diagnosed at age y. The specific mortality is derived from the cumulative relative survival (CRS) curve, under the hypothesis of independent competing risks, as follows:

$\delta(y,x)= CRS(y,x) – CRS(y-1,x)$.

Equation (2) gives the estimated age-specific mortality probability for a birth cohort, provided that the disease incidence, prevalence, and patient survival are known. A system of equations (2), including one equation for each birth cohort, allows to reconstruct cross-sectional mortality series for an entire observation period.

2.4   Modeling cancer incidence with age-period-cohort models

Incidence, as needed to be plugged into equations (1) and (2), can be modeled using age-period-cohort (APC) models. Assume incidence probability (μ) as a polynomial function of age (x), period of diagnosis (t), and birth cohort (c = t – x), throughout a logistic link function Φ:

$$\Phi_{x,t}(\underline{\alpha}) = \alpha_0 + \sum_{k=1}^{A} \alpha_k \, x^k + \sum_{k=1}^{P} \alpha_{1+A+k} t^k + \sum_{k=1}^{C} \alpha_{1+A+P+k} (t-x)^k$$

where

- $\Phi_{x,t}(\underline{\alpha}) = logit\left(\mu_{x,t}(\underline{\alpha})\right)$

- A, P, and C are the degrees of the polynomials for age, period, and cohort, respectively, to be chosen to give the best model fit. Since the cohort term is a linear combination of age and calendar year, the coefficient of the period linear term (t) is suppressed to avoid convergence problems ($\alpha_{1+A+1} = 0$).

- $\underline{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_{1+A+P+C})$ is the parameters' vector to be estimated through statistical regression.

Given a degree for polynomials, parameters $\underline{\alpha}$ are calculated according with MIAMOD/PIAMOD methods as follows:

a) MIAMOD

- $\underline{\alpha}$ parameters are back-calculated as maximum likelihood (ML) estimates assuming Poisson distributed cancer deaths, using a weighted least square iterative procedure as described elsewhere in details [Verdecchia et al, 1989; De Angelis et al, 1994].

16

- For each set of parameters $\underline{\alpha}$, MIAMOD fitting algorithm calculates the estimated cancer deaths and compares them with the observed cancer deaths from official statistics.

b) PIAMOD

- $\underline{\alpha}$ parameters are calculated as ML estimates assuming Poisson distributed incident cancer cases as described elsewhere in details [Verdecchia et al, 2002].

- For each set of parameters $\underline{\alpha}$, PIAMOD calculates the expected incident counts from the APC model and compares them with the incident cancer cases observed by the Cancer Registry.

The degree of the polynomials can be choose by different strategies in order to incorporate available data into the most appropriate model for specific descriptive or explanatory purposes. The choice of the degree of the polynomials is guided by a stepwise procedure based on likelihood ratio statistics (LRS); the significance of inclusion of each additional regression parameter in the incidence model is tested by comparing increasing order nested models as follows:

$$G^2 = (LRS)_k - (LRS)_{k+1} \approx \chi_1^2$$

Standard errors of ML parameters are asymptotically normally distributed and their covariance matrix is obtained by inverting the second-order derivative of log-likelihood function (*i.e.*, Fisher's matrix) [Verdecchia et al, 1989]. The significance of each parameter is evaluated using asymptotic T test. Standardized regression residuals (StRes) can be also calculated using the expected (exp) and observed (obs) counts as follows:

$$StRes_{x,t} = \frac{(exp_{x,t} - obs_{x,t})}{\sqrt{exp_{x,t}}}$$

and plotted for checking regression goodness of fit.

The general criteria for selecting the degree of the polynomials are based on significant reduction in the model's LRS by introducing new parameters, significance of each parameter, substantial improvement of estimates when compared with observed data, and robustness for projections.

2.4.1 Cubic splines for modeling incidence

Cubic spline models can also be used instead of polynomials in modeling incidence function, as they are more flexible for capturing irregular shapes and sudden changes in incidence/mortality rates [Hastie and Tibshirani, 1994]. The model constructs K third-order piecewise continuous polynomials (*i.e.*, cubic splines) that connect K data points (knots) with unit separation. Polynomial coefficients are chosen such that the resulting curve and its first derivative are smooth at the knots.

$$\Phi_t(\underline{a}) = a_0 + a_1 t + \sum_{j=2}^{k-1} a_j (t - t_j)_+^3$$

where $\begin{cases} (t - t_j)_+ = 0 & if\ t \le t_j \\ (t - t_j)_+ = t - t_j & if\ t > t_j \end{cases}$

- t_j is the j-th knot and K is the number of knots

- left and right tails are constrained to be linear (*i.e.*, natural splines)

- the Φ function and its 1st and 2nd derivatives are constrained to be continuous at the knots

- the coefficients of the linear and cubic functions are estimated by the fitting algorithm.

The number and the position of knots have to be fixed in advance; therefore, the model degrees of freedom is artificially low. The best-fitting model is defined as the one minimizing the Akaike Information Criterion (AIC) [Akaike et al, 1973]. The AIC, a variable selection criterion that compromises between a good fit and a simple model, is a penalized likelihood that takes into account the number of parameters estimated in the model.

As the choice of knots is arbitrary, spline models should be used only when strictly necessary.

2.5 Estimating relative survival

Survival used in MIAMOD/PIAMOD is supposed to be net cancer specific survival, that is:

- Cause-specific survival, when the information on the cause of death is available and reliable

- Relative Survival (RS) for a given cancer, defined as the ratio of the proportion of observed survivors (OS) in a cohort of cancer patients to the proportion of expected survivors (ES) in a comparable cohort of people in the general population.

$$RS = \frac{OS}{ES}$$

The cumulative relative survival (CRS) is the product of RS by follow-up interval (i) and includes the survival experience of cancer patients over follow-up time (d)

$$CRS\,(d) = \prod_{i=1}^{d} RS\,(i)$$

RS and CRS are useful when the information on the cause of death is unavailable or unreliable.

There are two different approaches for including RS in MIAMOD/PIAMOD estimates.

- Tabulated RS can be directly derived from incidence and follow-up data and population life-tables using standard methods [Ederer et al, 1961]. The use of tabulated RS requires a registration period long enough to catch main survival dynamics due to improvements in diagnostic procedures and more effective treatments. Moreover, it requires the stationary hypothesis for making projections (*i.e.*, the conservative hypothesis that survival projections are equal to the most recent observable level). These requirements become particularly critical for good prognosis cancer sites (*e.g.*, prostate cancer) with long period dynamics.

- Model-based RS can be derived using mixture cure models. These method allows long-term extrapolation from limited observed survival, survival extrapolation to populations not covered by cancer registration, smoothing of observed trends with high variability, and flexible extrapolation scenarios.

2.5.1 Mixture cure models with power function for modeling relative survival

Mixture cure models are based on the cancer patients heterogeneity assumption and allows for simultaneous estimation of factors associated with proportion of cured patients (*i.e.*, those who will not die for the specific cancer) and factors related with time to death for fatal cases (*i.e.*, those who will die for the specific cancer).

Relative survival can be modeled for MIAMOD/PIAMOD utilizing mixture cure models of the Weibull type with power function [Verdecchia et al, 1998].

$$CRS(d) = \{C + (1 - C)\exp[-(\lambda d)^\gamma]\}^\beta$$

where

- C is the proportion of cured patients (*i.e.*, cure fraction), defined as the limiting value of the cancer survival function CRS, as time to diagnosis *t* approaches to infinity

$$C = \lim_{t \to \infty} CRS(t)$$

- (1 – C) is the proportion of fatal cases

- $\exp[-(\lambda d)^\gamma]$ is a Weibull function $W(\lambda, \gamma, d)$ with:

  λ = scale parameter. It represents the excess death risk of fatal cases and determines the scale of the CRS curve.

  γ = shape parameter. It modulates the excess death risk of fatal cases. The lower is γ (<1) the higher is the risk of death for fatal cases in the short term and decreasing thereafter.

  d = time since diagnosis or follow-up or duration.

- β is the power function that allows to include in the model prognostic covariates.

In principle, all demographical and clinical variables related to survival can be included in the model, playing a different role on cured proportion and on time to failure for fatal cases. However, the availability of such variables is often limited to subsets of patients. Therefore, the principal variables to be included in the survival model are usually sex, age at diagnosis, period of diagnosis, and population (*e.g.*, geographical area).

Explicative covariates are included in the mixture cure models by means of the power function which may depend on them as follows:

$$CRS(\underline{Z}, d) = (C + (1 - C) \, W(\lambda, \gamma, d)^{\exp\,[-\underline{\beta}\underline{Z}]}$$

where $\underline{Z}$ is the vector of covariates and $\underline{\beta}$ includes the corresponding relative risks,

and/or stratifying the baseline function parameters

$$\left\{ C_{\underline{Z}} + (1 - C_{\underline{Z}}) \, W(\lambda_{\underline{Z}}, \gamma_{\underline{Z}}, d) \right\}.$$

The choice of the more appropriate model depends on data characteristics (*e.g.*, number of strata), model hypothesis (*e.g.*, age or population specific trends), and specific aims of survival modeling (*e.g.*, projections, extrapolations to areas not covered by cancer registration).

To summarize the results obtained from these models two indicators are considered: the cure fraction, C, and the mean survival time for fatal cases, T, which is given by

$$T = \frac{1}{\lambda} \, \Gamma \left(1 + \frac{1}{\gamma} \right)$$

where Γ is the Gamma function.

Survival model parameters are estimated by means of a non-linear regression procedure with the inverse of the variances of the observations used as weights, using SAS (PROC NLIN, examples in Appendix A).

2.6   Projections with MIAMOD/PIAMOD

Projections of morbidity and mortality trends in the MIAMOD/PIAMOD applications require several hypotheses concerning incidence, survival, and population evolution patterns.

- APC incidence model projections can be derived by assuming the persistence of both age and cohort effects during the calendar years following the observation period. For cancer disease this assumption is quite reasonable, as cancer risk is generally determined by past exposure to risk factors. Conversely, the period model is not projected into the future, as it is assumed that what happened simultaneously to all age groups in the observation period

cannot happen in the same way in subsequent years; therefore, only a linear drift is retained, based on a defined number of years.

- Survival projections should be derived providing a plausible range of scenarios. For instance, a pessimistic hypothesis consists in assuming patients survival to remain stable in the future, whereas an optimistic one consists in considering survival to continue improving at the same rate observed in recent past years. The following assumptions can be made in the MIAMOD/PIAMOD interface for backward and forward relative survival extrapolations: constant, dynamic with the same slope estimated from the data, dynamic until a given calendar year, or dynamic until al lower bound (for backward only). In MIAMOD, backward survival dynamic modifies the APC incidence model derived from mortality data and determines changes on mortality, incidence, and prevalence whereas forward survival dynamic does not modify the APC incidence model but influences mortality and prevalence projections. In PIAMOD, survival projection does not change the APC incidence model but forward projection influences mortality and prevalence trends and backward projection influences prevalence.

- Minor hypotheses are required for projecting population evolution patterns. The number of new born (*i.e.*, population count at age 0) is assumed to be constant and equal to that of the last available calendar year, also general mortality rates are assumed to be constant and equal to those of the last available calendar year, and no migration is assumed. Population at older age classes is estimated by accounting for the incrementing age of the cohort members and for the expected number of deaths.

Prevalence and mortality projections can be derived by using equations (1) and (2), respectively.

2.7 The MIAMOD/PIAMOD software

MIAMOD software (source code in Fortran 77 for mainframe computers) was developed in 1989 by Italian National Institute of Health (Istituto Superiore di Sanità) [De Angelis et al, 1994].

In 2000 source migrated under Windows Operating System and in 2003, a unique environment to run MIAMOD and PIAMOD was developed in collaboration with the US National Cancer Institute with a graphical user interface written in Visual Basic.

When model-based survival is used, the model parameters estimate should be done externally to the MIAMOD/PIAMOD interface, using SAS software (SAS Institute Inc.).

3. APPLICATION OF THE MIAMOD METHOD IN FRIULI VENEZIA GIULIA

3.1 Comparison between estimated and observed prostate cancer incidence rates

An analysis for estimating and projecting incidence, mortality, and prevalence for major cancer sites (including prostate) was conducted in Friuli Venezia Giulia (FVG) region in the period 1970-2015, using the MIAMOD method, in the framework of a national project coordinated by the Italian National Institute of Health. The results are included in the paper by Zucchetto and colleagues (2013) "*Cancer estimates up to 2015, in Friuli Venezia Giulia*".

Study findings clearly showed the high goodness of fit of estimates with observed data from the FVG Cancer Registry for several cancer sites, except for prostate. Actually, the incidence rates of PCa were clearly underestimated by the MIAMOD method, as explained in the following.

3.1.1 Methods

MIAMOD

The MIAMOD method [Verdecchia et al, 1989; De Angelis et al, 2004], was applied to estimate the absolute number of incident cases, deaths and prevalent cases, crude and age-standardized (using the standard European population) incidence and mortality rates (per 100,000 person-years), and prevalence proportions (per 100,000) for the period 1970-2015. All estimates were carried out up to age 99 years. Mortality data for all cancers, general mortality, and population data by age, calendar year, and geographical area for the period 1970-2002 were obtained from the Italian National Institute of Statistics (ISTAT). Specific-cause mortality data for the years 2003, 2006, and 2007 were used to validate expected mortality projections, as ISTAT had yet to publish data with causes of death for the period 2004-2005. Relative survival estimates were calculated by means of

parametric cure models of the Weibull type at the level of macro area, using data from cancer registries included in the EUROCARE-4 for the period 1985-2002 [Capocaccia et al, 2009]. The survival estimates for the North-East macro area were assigned to FVG.

For PCa, a specific procedure was used to capture recent rapid variations of time trends, as suggested by data from cancer registries [Curado et al, 2007]. Mortality estimation up to the year 2010 was preliminarily performed by means of the PIAMOD method [Verdecchia et al, 2002], using regional mortality data during the period 1970-2007 as input (best model fit was found for an APC model with 2-degree age, 2-degree period, and 2-degree cohort polynomials). This allowed to complete the missing cause-specific mortality time series in the years 2004 and 2005 and to base incidence estimates on more recent mortality data. This longer mortality time series was then used as input for the MIAMOD method (best model fit was found for an APC model with 2-degree age, 3-degree period, and 4-degree cohort polynomials, respectively). The survival time trend was modeled by means of mixture cure models of the Weibull type with power function [Verdecchia et al, 2009] for the period 2003-2005 and then assumed to be constant onwards (*i.e.*, equal to 2005 for the period 2006-2015). The baseline Weibull mixture cure model was stratified by age at diagnosis (age classes: 15-54, 55-64, 65-74, 75-84, 84-99 years) and estimated for each age-stratum at the reference year (1994), including age-stratified period relative risks and an area relative risk equal to that of the North-East for all age strata (Appendix A).

FVG Cancer Registry Data

Since 1995, the population-based Cancer Registry of FVG has been registering incident cancer cases diagnosed in people residing within the whole regional territory [Birri et al, 2011]. The FVG cancer registry, accredited at the Italian Association of Cancer registries (AIRTUM), together with the registries of Umbria and Trento and Bolzano, is one of the very few Italian registries covering a whole region [AIRTUM]. The registry has high completeness and quality, in accordance with the standards required by the International Association of Cancer Registries.

All PCa cases diagnosed between 1995 an 2007 (FVG cancer registry activity period) were considered. Routine indicators for PCa showed that 90.7% of cases were microscopically verified and that only 0.2% were identified on the basis of death certificate only (*i.e.*, an indicator of poor quality of data). Interestingly, 3.5% of cases had been diagnosed solely at autopsy. These cases were excluded from this analysis, in order to increase comparability with MIAMOD estimates, given than usually diagnoses made during autopsy are not reported in death certificates and, hence, do not affect mortality statistics.

As an alternative approach, PIAMOD method was also applied FVG cancer registry data in the period 1995-2007 as input (the best model fit was found for an APC model with 3-degree age, 5-degree period, and 3-degree cohort polynomials, respectively).

Annual PCa incidence rates per 100,000 resident men were calculated using as denominator the mean number of resident males in the corresponding calendar year (i.e., the mean between population at 1st January of the year y, and 1st January of the year y+1), derived from ISTAT [ISTAT. *Popolazione residente*].

3.1.2   Results

The estimates of PCa incidence, mortality, and prevalence for the whole period 1970-2015 using the MIAMOD method are reported in Figure 3.1. The comparisons between observed and MIAMOD/PIAMOD estimated incidence and mortality rates for the common period 1995-2007 are reported in Figures 3.2 and 3.3.

Although both increasing (the annual percent change, aPC, was 5.0; 95% CI: 3.7-6.5 for observed incidence rates; and aPC was 5.2, 95% CI: 4.4-6.0 for expected rates), PCa incidence rates reported by the FVG Cancer Registry were much higher than those estimated using MIAMOD, despite the exclusion of autoptical cases (Figure 3.2). Actually, the huge grow of incidence rates in the same period was not accompanied by a corresponding increase in mortality, which was instead almost

stable (aPC = -0.3, 95% CI: -0.2 to 1.4 for observed mortality; aPC = -0.4, 95% CI: -0.8 to 0.1 for expected mortality).

The application of PIAMOD method, using FVG Cancer Registry data in the period 1995-2007 as input, provided a far better estimation of PCa incidence (Figure 3.3). However, limitations emerged with regard to PCa mortality rates that were clearly underestimated in comparison with observed data up to 1999. As a consequence, complete prevalence estimates resulted to be highly inflated, given the great influence of past mortality rates on prevalence for cancers with high survival, such as PCa (Figure 3.4).

In any case, it is worth noting that lots of Italian areas are not covered by population-based cancer registries or have still few years of observation (*e.g.*, only in the very recent years the population coverage of cancer registration reached 50% in Italy, [AIRTUM]) causing the impossibility of extensively using PIAMOD.
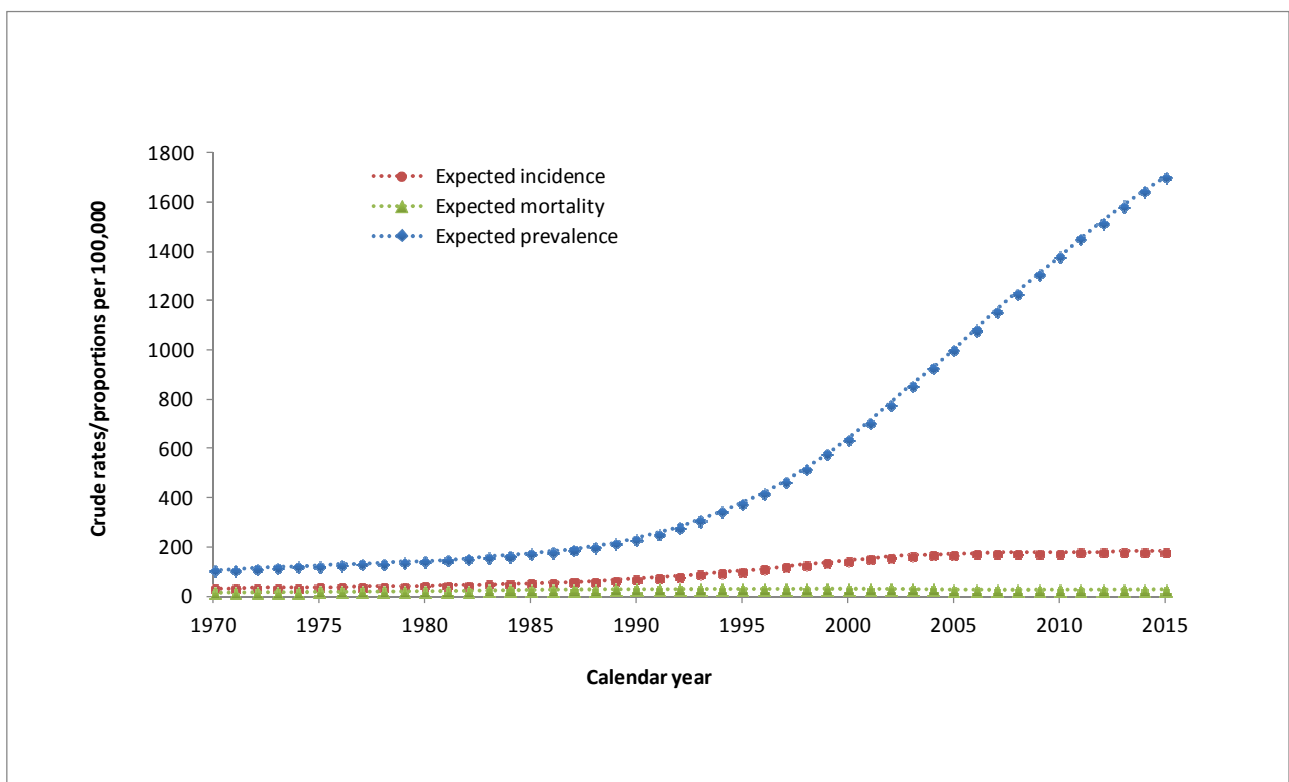
Figure 3.1. Crude prostate cancer incidence and mortality rates (per 100,000 men-year) and prevalence proportions (per 100,000 men) estimated using the MIAMOD method. Friuli Venezia Giulia
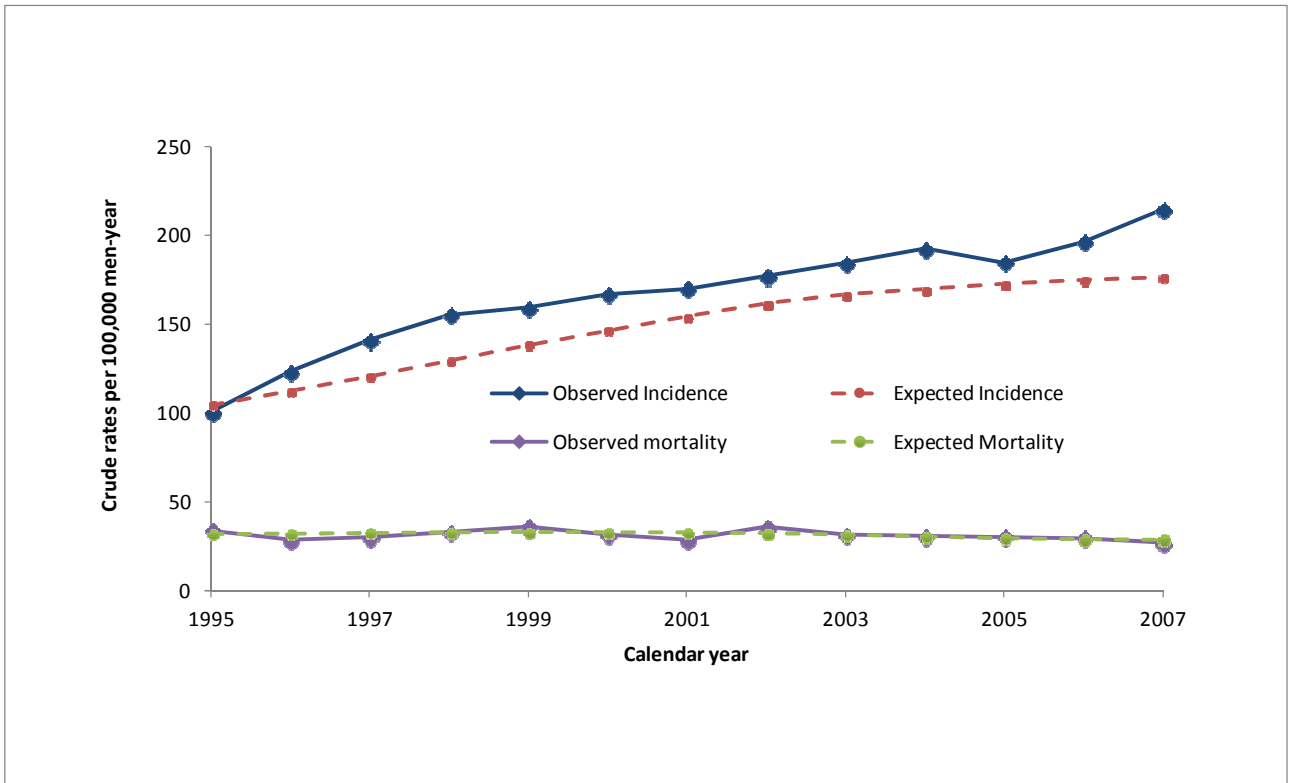
Figure 3.2. Crude prostate cancer incidence and mortality rates (per 100,000 men-year). Observed (continuous lines) and estimated using the MIAMOD method (dashed lines). Friuli Venezia Giulia.
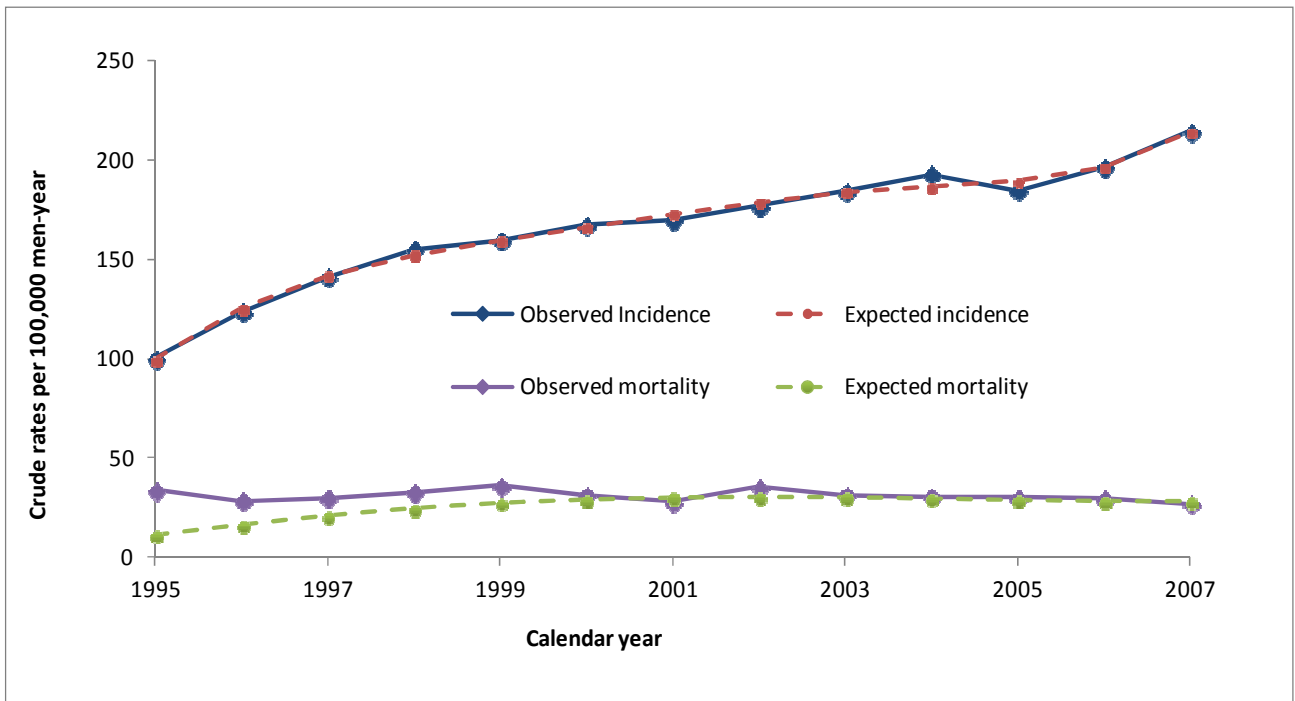
Figure 3.3. Crude prostate cancer incidence and mortality rates (per 100,000 men-year). Observed (continuous lines) and estimated using the PIAMOD method (dashed lines). Friuli Venezia Giulia.
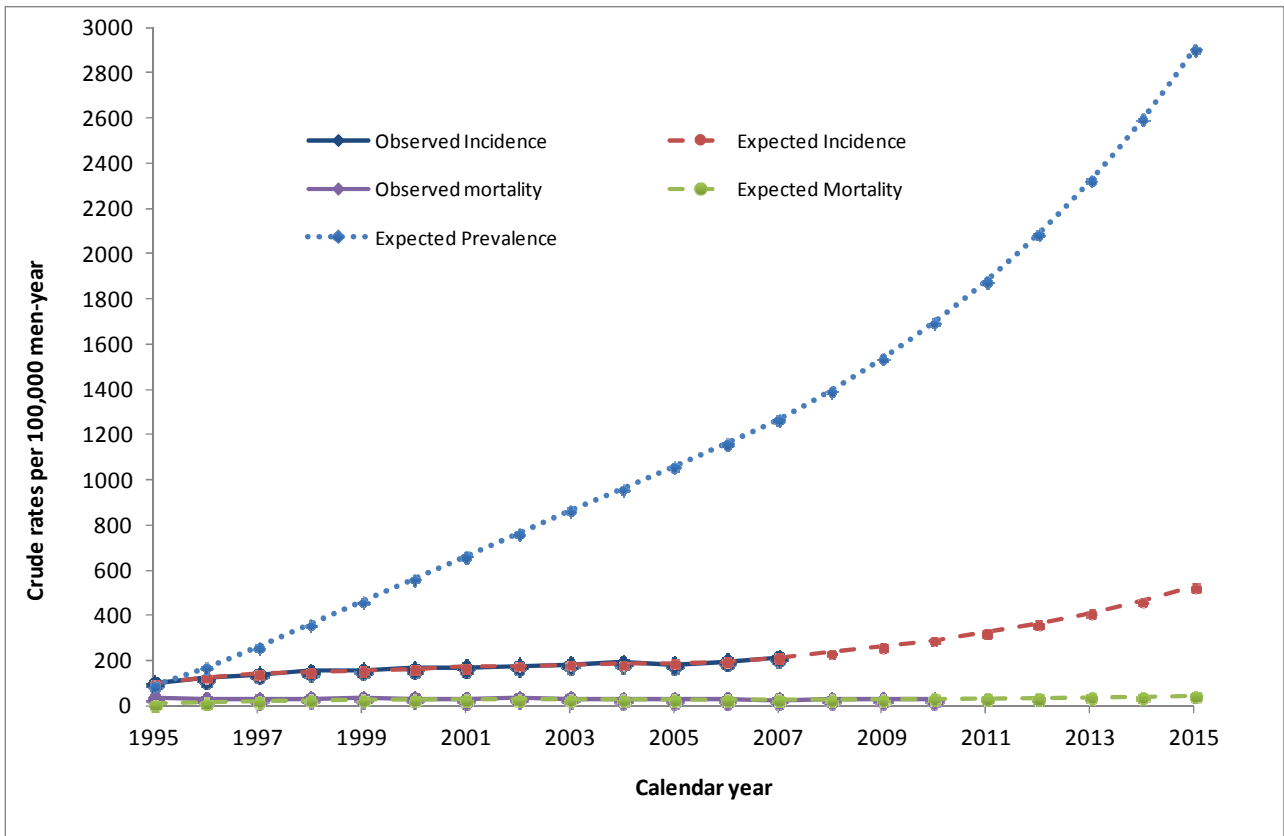
Figure 3.4. Crude prostate cancer incidence and mortality rates and prevalence proportion (per 100,000 men-year). Observed (continuous lines) and estimated using the PIAMOD method (dashed lines) up to 2015. Friuli Venezia Giulia.

4. TRENDS OF PSA TESTING RATES IN FRIULI VENEZIA GIULIA

4.1   PSA testing rates

4.1.1 Methods

Data on PSA testing use, which are collected for administrative purposes at a population-level, are available in several Italian areas, even those not covered by cancer registration. For instance, in FVG, the regional digital health archive (a data-warehouse in SAS, SAS Inc.) collects data on diagnostic procedures, hospital discharges, and pathological archives of the whole region. This archive includes procedures performed in public structures as well as in private structures accredited by the Regional Health System. Data on PSA tests can be extracted from the outpatient clinics database ("*prestazioni ambulatoriali*"), available since 1998.

In the following analysis, both total-PSA and free-PSA tests (FVG regional code 90.56.5 and 90.56.6, respectively) from 1998 up to 2012 (*i.e.*, last available year as of 2013) were considered. Only PSA tests performed on resident men aged 40 years or more were included, as PCa is very rare in men under 40 years and PSA testing at those ages is used in case of acute prostatic conditions (*e.g.*, prostatitis) rather than for PCa screening. Data on PSA testing use include the regional identification code of the tested person, the date of PSA testing and other characteristics, such as test price, but not serum PSA levels. Prescriptions of both types of examinations in the same year to the same man were counted once. The overall PSA testing rates were calculated as the number of PSA tested men in one calendar year over the mean resident male population in the same year [ISTAT. *Popolazione residente*].

4.1.2 Results

PSA testing rates significantly increased from 12,792 per 100,000 men in 1998 up to 30,407 in 2009, and then slightly decreased down to 29,113 in 2012. Similar patterns emerged for all ages, except for men aged ≥70 years, for whom trends did not show any reduction (Figure 4.1). Interestingly, among men aged 65-79 years, PSA testing rates rose up to approximately 50,000 per 100,000, meaning that one out of two men had been tested.
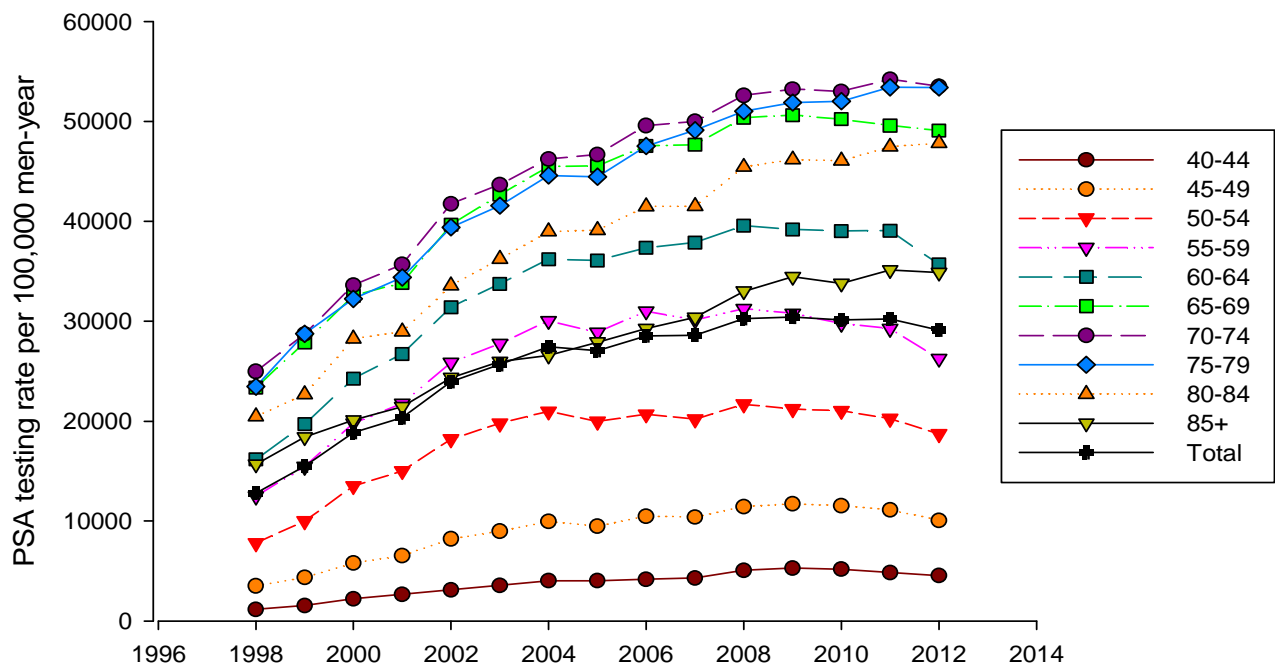
Calendar year	No. of tested men	Mean male population	PSA testing rate	(95% CI)
1998	37,472	292,944	12,791.5	(12662.3-12921.7)
1999	45,728	294,990	15,501.5	(15359.8-15644.3)
2000	56,141	297,675	18,859.8	(18704.1-19016.5)
2001	61,270	300,998	20,355.6	(20194.8-20517.4)
2002	73,239	305,645	23,962.1	(23788.9-24136.3)
2003	79,744	310,707	25,665.3	(25487.5-25844.1)
2004	86,640	315,634	27,449.5	(27267.0-27632.9)
2005	86,778	320,539	27,072.5	(26892.7-27253.3)
2006	92,743	325,043	28,532.5	(28349.2-28716.8)
2007	94,392	330,072	28,597.4	(28415.2-28780.4)
2008	101,513	335,416	30,264.8	(30078.9-30451.6)
2009	103,450	340,221	30,406.7	(30221.7-30592.6)
2010	103,778	344,389	30,133.9	(29950.9-30317.9)
2011	104,103	344,315	30,234.8	(30051.4-30419.0)
2012	99,645	342,276	29,112.5	(28932.0-29293.8)

Figure 4.1. Age-specific PSA testing rates per 100,000 men-year (mean ISTAT pop) by calendar year. Friuli Venezia Giulia, men aged 40+ years, 1998-2012

4.2 Joinpoint analysis of PSA testing rates trends

4.2.1 Methods

Joinpoint regression analysis [Kim et al, 2000] was performed to identify points (knots) where a statistically significant change over time in the log-slope of the age-specific PSA testing rates occurred and to estimated the rates trend within each time span between the knots. Analyses were performed using the Joinpoint Regression Program (Version 3.5 – April 2011; Statistical Methodology and Applications Branch and Data Modeling Branch, Surveillance Research Program National Cancer Institute), using the following options:

- Standard parameterization of Kim *et al.* (2000) of the joinpoint regression model, that is:

$$E[y/x] = \beta_0 + \beta_1 x + \delta_1(x - \tau_1)^+ + ... + \delta_k(x - \tau_k)^+ \tag{1}$$

where τ_k is the unknown joinpoint and $(a)^+ = a$ if $a > 0$ and 0 otherwise.

- Log-linear model for the rate y, that is: $\ln(y) = x'\beta + e.$

- Heteroscedastic random errors with standard deviation specified at each time period. Regression coefficients are estimated by weighted least squares, where weights at each point are: $w = (y^2)/v,$ where y^2 is the square of the response for that point and v is the square of the standard deviation at each time period.

- The maximum number of joinpoints (k) was set at 2, according to the recommendations for $n=15$ observed data points.

Number of Data Points	Maximum Number of Joinpoints
<7	0
7 – 11	1
12 – 16	2
17 – 21	3
22 – 26	4
27+	5

Let be:

- n the total number of data points

- k the number of joinpoints in the model

- p the total number of parameters in the model, including the joinpoint parameters ($p=2k+2$)

- K_{min} and K_{max} the minimum and maximum number of joinpoints

- Q_k the weighted sum of squared errors (SSE) from the model that minimizes weighted SSE with k joinpoints

- $Q_{x,j,k}$ the weighted SSE from the model that minimizes weighted SSE with k joinpoints and with the jth joinpoint occurring at x

- $F^{-1}_{a,b}(p)$ the pth quantile of the F distribution with a and b degrees of freedom.

First, the procedure goes through each of the k-joinpoint models, $K_{min} \le k \le K_{max}$. For each of the models, the program chooses the regression parameters with the smallest weighted SSE. The minimum SSE for a k-joinpoint model is calculated using Lerman's grid-search method [Lerman, 1980] based on standard parametrization (1). The corresponding values for ($\tau_{1,...}$ τ_k) and (β_0, β_1, $\delta_1,..., \delta_k$) are the estimates of joinpoints and regression coefficients, respectively.

If $k > 0$ then the output lists the estimated joinpoints. The associated CIs come from Lerman (1980): the $100(1 - \alpha)$% CI for the jth of k joinpoints includes all values of x from the grid such that $Q_{x,j,k} \le C_2^\alpha$, where

$$C_2^\alpha = Q_k \left(1 + \frac{k}{n - p} \right) F_{k,n-p}^{-1}(1 - \alpha)$$

The sequential permutation test procedure is used to choose the best joinpoint model, as described elsewhere in details [Kim et al., 2000], and here briefly reported.

The procedure tests the hypothesis of no change

$$H_0: E[y/x] = \beta_0 + \beta_1 x$$

against the alternative hypothesis of two joinpoints

$$H_0: \text{there exist } \tau_1 \text{ and } \tau_2, \tau_1 < \tau_2 \text{ such that } E[y/x] = \beta_0 + \beta_1 x + \delta_1 (x - \tau_1)^+ + \delta_2 (x - \tau_2)^+$$

If the null hypothesis is rejected, then the same procedure is applied to test the null hypothesis of one joinpoint against the alternative of two joinpoints, and so on. Because multiple tests are performed, Bonferroni adjustment is used to ensure that the approximate overall type I error is less than the specified significance level α (default 0.05). Each of these permutation test are carried out at significance level of $\alpha_1=\alpha/(K_{max}- K_{min})$. The Bonferroni adjustment is conservative because overall significance level is usually less than the nominal level α.

The annual percentage change (aPC) is computed for each of the trends identified by the knots. For any segment with slope the aPC is $100\{\exp(\beta) – 1\}$. The average annual percentage change (aaPC), which is a weighted mean of the aPC's from the joinpoint models, summarizes the trend over the whole period.

4.2.2   Results

Joinpoint analysis estimated statistically significant changes in PSA testing rates slopes in 2002 and 2008 (Tab. 4.1 and Fig. 4.2): the aPCs in the periods 1998-2002, 2002-2008, and 2008-2012 were 16.9 (95% CI: 12.9 to 21.2), 3.6 (95% CI: 1.7 to 5.5), and -0.7 (95% CI: -3.1 to 1.8), respectively, with an aaPC based on the last 10 years equal to 1.7 (95% CI: 0.4 to 1.9).
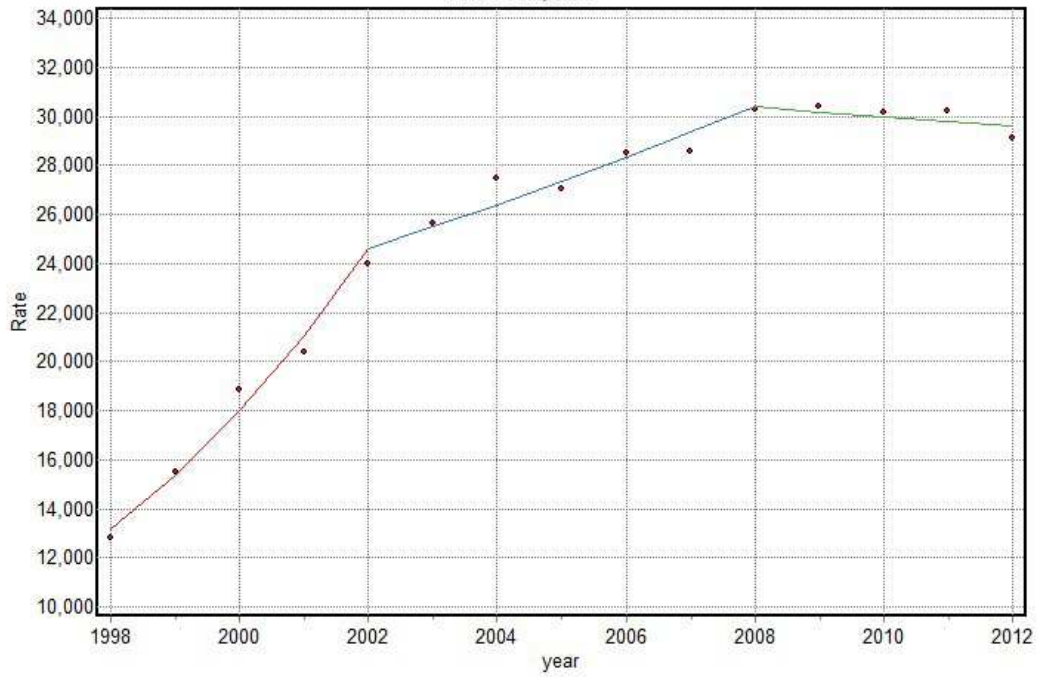
Period	aPC	Lower bound 95% CI	Upper bound 95% CI
1998-2002	16.9	12.9	21.2
2002-2008	3.6	1.7	5.5
2008-2012	-0.7	-3.1	1.8

Figure 4.2. Joinpoint analysis of PSA testing rates (model with 2 joinpoints). Friuli Venezia Giulia
aPC = annual percent change, CI = confidence interval

Table 4.1. Output of Joinpoint analysis of PSA testing rates (model with 2 joinpoints). Friuli Venezia Giulia

Model statistics

| No. of Joinpoints	No. of Observations	No. of Parameters	Degrees of Freedom	Sum of squared errors	Mean Squared Error
2	15	6	9	545.46604	60.60734

Estimated Joinpoints

Joinpoint	Estimate	Lower CI	Upper CI		
1	2002	2000	2004		
2	2008	2003	2010		

Estimated regression coefficients (Beta) Standard Parameterization

Parameter	Parameter Estimate	Standard Error	Z	Prob > |t|	
Intercept 1	-303.231978	30.07501	-10.082523	0.00002	
Slope 1	0.156515	0.01504	10.406767	0.000016	
Slope 2 - Slope 1	-0.121353	0.016977	-7.148073	0.000186	
Slope 3 - Slope 2	-0.041853	0.012974	-3.225834	0.014535	

General parametrization

Parameter	Parameter Estimate	Standard Error	Z	Prob > |t|	
Intercept 1	-303.231978	30.07501	-10.082523	0.00002	
Intercept 2	-60.283413	15.79129	-3.81751	0.006563	
Intercept 3	23.75739	20.729205	1.146083	0.289424	
Slope 1	0.156515	0.01504	10.406767	0.000016	
Slope 2	0.035162	0.007876	4.464653	0.00292	
Slope 3	-0.006691	0.010311	-0.648948	0.537074	

The permutation test found that model with 2 joinpoints was significantly better than model with just 1 (Tab 4.2). However, for comparison, also the results of the model with 1 joinpoint are reported in Figure 4.3. The model found significant changes in PSA testing rates slopes in 2003: the aPCs in the periods 1998-2003 and 2003-2012 were 14.9 (95% CI: 11.7 to 18.3) and 1.4 (95% CI: 0.5 to 2.3), respectively.

Table 4.2. Permutation test output of Joinpoint analysis of PSA testing rates

Null Hypothesis	Alternate Hypothesis	Numerator Degrees of Freedom	Denominator Degrees of Freedom	No. of Permutations	P-Value	Significance Level
0 Joinpoint	2 Joinpoints	4	9	4500	0.000222	0.0250000
1 Joinpoint	2 Joinpoints	2	9	4500	0.020222	0.0500000
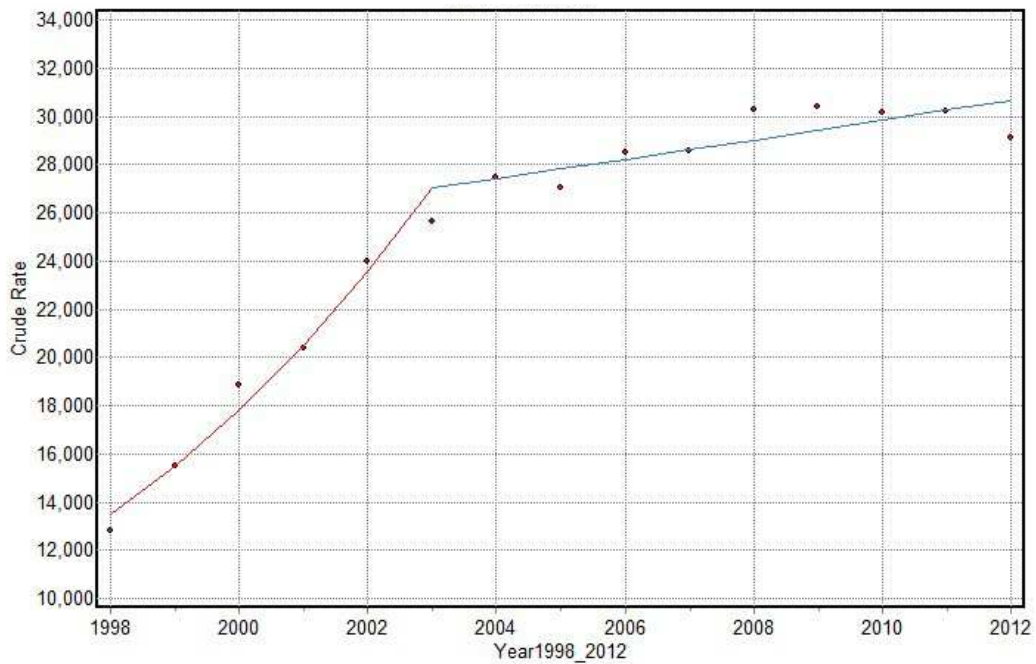
Period	aPC	Lower bound 95% CI	Upper bound 95% CI
1998-2003	14.9	11.7	18.3
2003-2012	1.4	0.5	2.3

Figure 4.3. Joinpoint analysis of PSA testing rates (model with 1 joinpoint). Friuli Venezia Giulia

Stratified analysis according to age of men at PSA testing, found statistically significant changes in the slopes of PSA testing rates for all age groups between 2002-2003 and 2008-2009, except for men aged 80-84 years, for whom only 1 joinpoint in 2003 emerged (Fig. 4.4). For men aged less than 70 years the liner trend of the log-rate in the last period was decreasing, whereas for older men PSA testing rates continue raising, though at a lower extent.
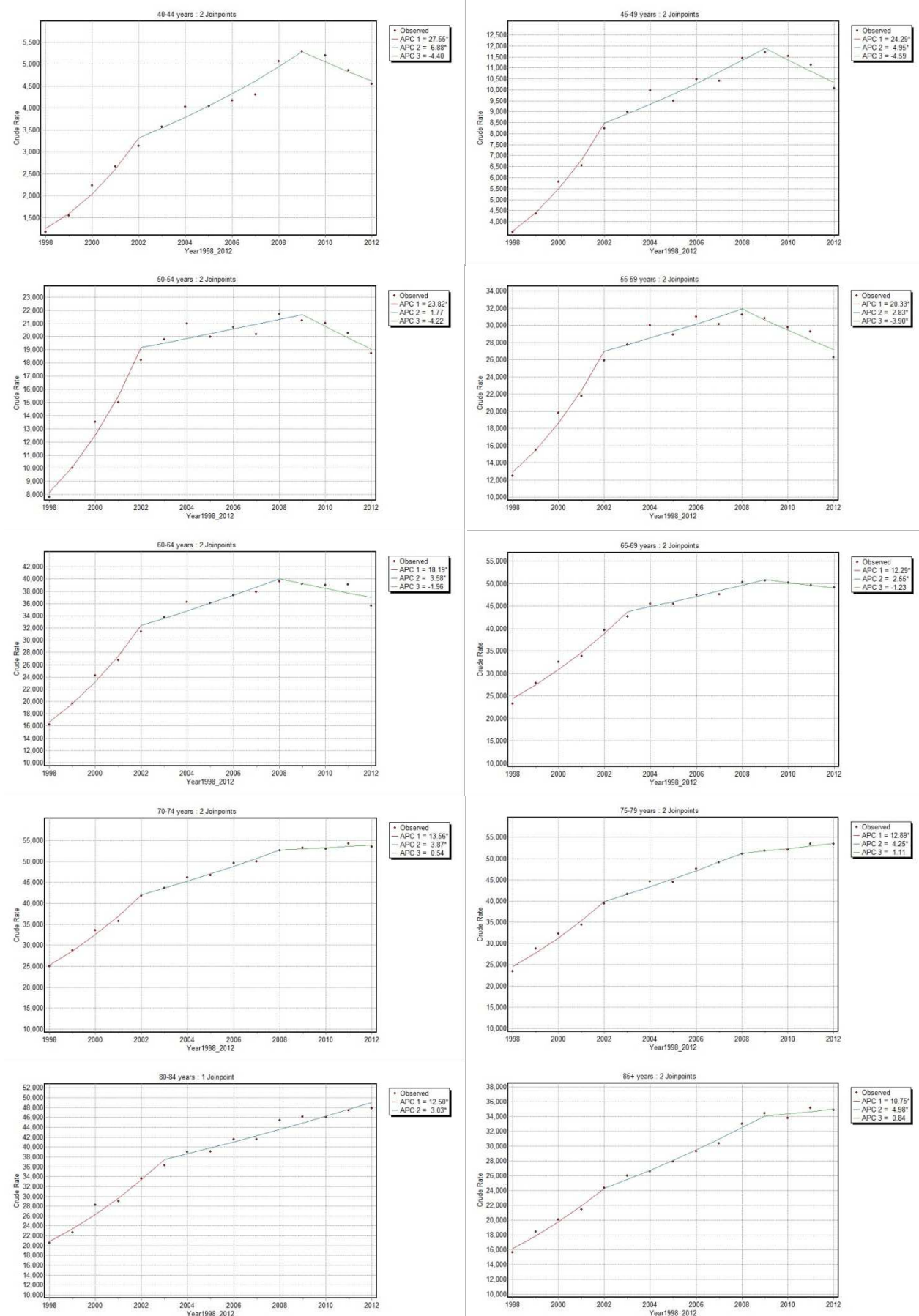
Figure 4.4. Joinpoint analysis of PSA testing rates by age class. APC: annual percent change; [*] Statistically significant

5. AGE-PERIOD-COHORT MODELS

Age-period-cohort models are utilized in order to disentangle the effects on time trends of age at the event, calendar time of the event, and calendar time of birth. Age-period-cohort models are herein described using the approach of Carstensen [Carstensen, 2007].

5.1 Methods

Age-period-cohort models are descriptive tools for rates observed in a Lexis diagram (Fig. 5.1).
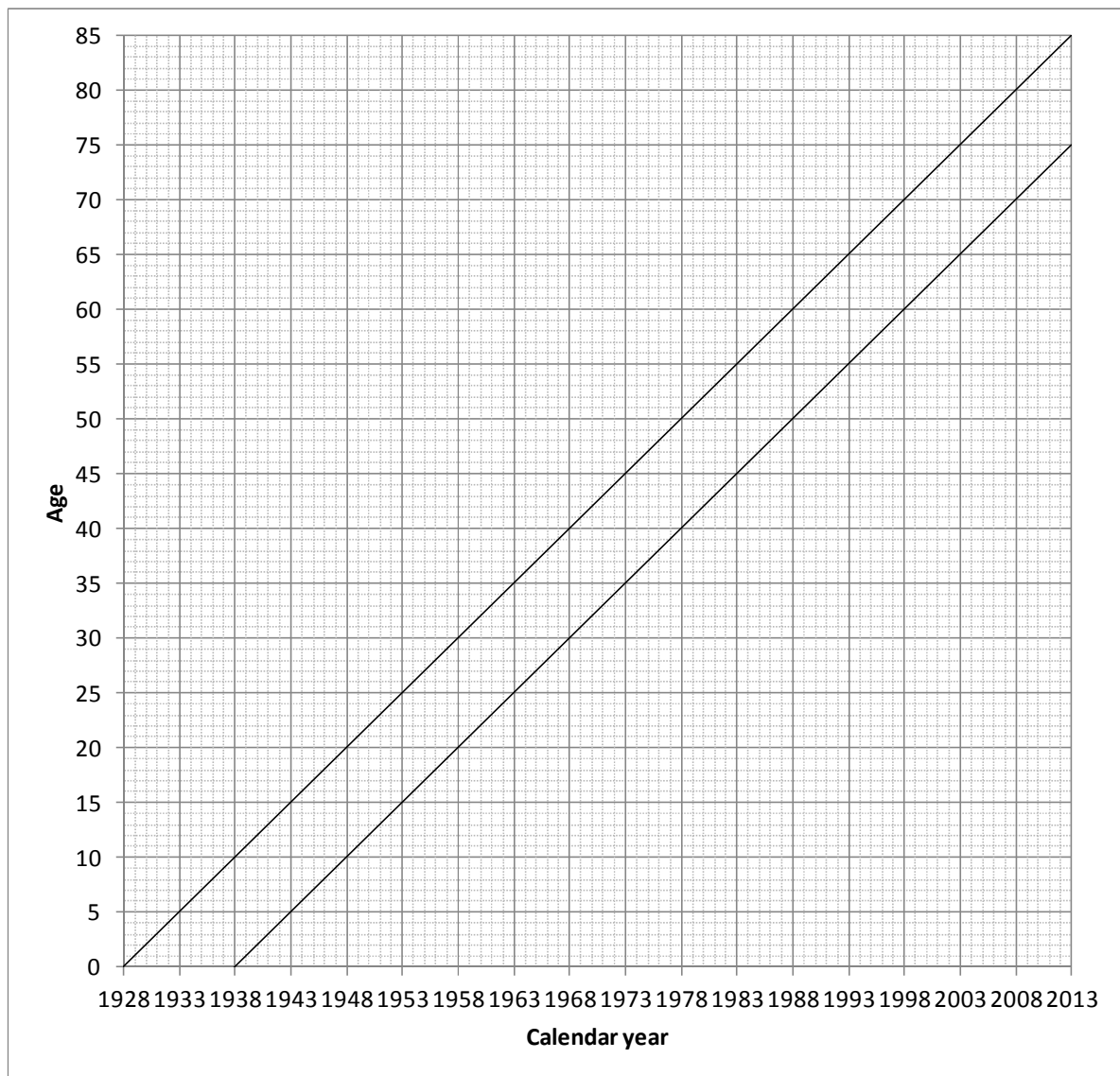
Figure 5.1. Lexis diagram, the birth cohort 1928-1937

Consider λ the rate of an event observed in an arbitrary subset of a Lexis diagram, where D is the number of cases who experimented the event and Y the amount of risk time (*i.e.*, person-years at risk of the event). Assuming that rates are constant within each tabulation category, the log-likelihood contribution from observation of the random quantity (D, Y) in one subset is:

$$l\,(\lambda|D,\,Y) = D\,\log\,(\lambda) - \lambda Y$$

Except for the constant (D log (λ)), this is the same as the log-likelihood for an observation of a random variable D from a *Poisson* distribution with mean λY.

The log-likelihood for the entire table of (D, Y) is the sum of such terms, because individuals are independent and the contributions to different cells from one individual are assumed to be conditionally independent. Hence, models for λ can be fitted using programs for Poisson regression for independent observations that allows for an *offset* term to separate the person-years from the rate.

The rates can be modeled as functions of age class A, period P, and cohort of birth C, by letting D the response, log(Y) the *offset* and A, P, and C categorical explanatory variables in a Poisson model. Since the cohort term is a linear combination of age and calendar year (C = P – A), this produces an unidentifiability problem which requires parametrization constraints to be solved. There are several ways of arriving at a parametrization of an age-period-cohort model, such as:

- to constrain 1 period and 2 cohort parameters to be 0 (or *vice versa*);

- the Holford's residual approach [Holford, 1983] *i.e.*, to regress the age estimates on age, the period estimates on period, and the cohort estimates on cohort, and then to report residuals as age, period, and cohort effects.

- the sequential method, *i.e.*, to fit an age-cohort model, and subsequently a period-alone model using the log-fitted values from the age-cohort model as offset.

5.1.1 The age-period model

The age-period model states that the age-specific rates have the same shape in all periods, eventually with a varying level. The model has one parameter per age class and one per period, but always there is a one-parameter unidentifiability in the formulation.

$$\log[\lambda(a, p)] = f(a) + g(p)$$

In this model only the first derivatives (contrasts) of f and g are identifiable. The natural constraint is to fix one parameter to be 0, $g(p_0) = 0$. For period p_0 we will have:

$$\log[\lambda(a, p_0)] = f(a) + g(p_0) = f(a)$$

Thus, for period p_0, the $f(a)$ are logs of age-specific rates, and the age-specific rates are $\exp[f(a)]$.

Comparing the rates in any age class between period p and p_0 gives

$$\log[\lambda(a, p)/\lambda(a, p_0)] = \log[\lambda(a, p)] - \log[\lambda(a, p_0)] = f(a) + g(p) - f(a) = g(p)$$

Thus, the $g(p)$ are logs of rate-ratios (RR) relative to the period p_0.

5.1.2 The age-cohort model

The age-cohort model states that the age-specific rates have the same shape in all cohorts, but possibly with a varying level. The model has one parameter per age class and one per cohort.

$$\log[\lambda(a, c)] = f(a) + h(c)$$

In this model only the first derivatives (contrasts) of f and h are identifiable. This is traditionally fixed by choosing a reference cohort c_0 and constrain $h(c_0) = 0$. For cohort c_0 we will have:

$$\log[\lambda(a, c_0)] = f(a) + h(c_0) = f(a)$$

the $f(a)$ are logs of age-specific rates, and the age-specific rates are $\exp[f(a)]$.

Comparing the rates in any age class between cohort c and c_0 gives:

$$\log[\lambda(a, c)/\lambda(a, c_0)] = \log[\lambda(a, c)] - \log[\lambda(a, c_0)] = f(a) + h(c) - f(a) = h(c)$$

Thus, the $h(c)$ are logs of rate-ratios (RR) relative to the cohort c_0.

5.1.3 The age-drift model

The age-drift model is a sub-model of both the age-period and the age-cohort model. Inspection of the rate-ratio plots could suggest to replace the period or the cohort parameters by a linear trend in log-rates:

$$\log[\lambda(a, p)] = f(a) + g(p) = f(a) + g(p - p_0)$$

$$\log[\lambda(a, c)] = f^*(a) + h(p) = f^*(a) + h(c - c_0)$$

The two models are analytically the same, given that $p = a + c$. This implies that the rate-ratio display (on the log-scale) would show a straight line.

$$f(a) + g(p - p_0) = f(a) + g(a + c - (a_0 + c_0)) = f(a) + g(a - a_0) + g(c - c_0)$$

So going from the age-period-drift model to the age-cohort-drift model is just to replace the age effect $f(a)$ by $f^*(a) = f(a) + g(a - a_0)$.

The interpretation of this model is that rates increase exponentially by time (period or cohort) at the same pace, $\exp[g(p)] = \exp[h(c)]$ per year for all age classes.

5.1.4 The age-period-cohort model

The general form of a multiplicative age-period-cohort model for rates $\lambda(a, p)$ is:

$$\log[\lambda(a, p)] = f(a) + g(p) + h(c) \qquad (1)$$

where the covariates are mean age a, mean period p, and mean cohort c, and f, g, h are functions of age, period, and cohort, respectively. Given that $a = p - c$, the model can be written as follows:

$$\log[\lambda(a, p)] = f(a) + \delta a - \mu(p) - \mu(c) +$$

$$[g(p) - \delta p + \mu(p)] +$$

$$[h(c) + \delta c \qquad + \mu(c)]$$

for any $\mu(p)$, $\mu(c)$, and δ.

Therefore, the parametrization produced by setting certain period and cohort effects to 0 corresponds to choose values of the three arbitrary parameters $\mu(p)$, $\mu(c)$, and δ.

Any parametrization of the age-period-cohort model fixes 2 levels and a slope among the three functions, but different principles can be used to accomplish this.

One of these principles is based on an extension of the assumptions behind the way the age-cohort model was parametrized:

1. The age-function is interpretable as log of age-specific rates in a cohort c_0 (longitudinal rates), after adjustment for the period effect;

2. The cohort function is interpretable as log-RR relative to the reference cohort c_0;

3. The period function is 0 on average with 0 slope, interpretable as the log-RR relative to the age-cohort prediction (residual log-RR).

Depending on the subject matter, the role of cohort and period could be interchanged, in which case the age-effects would be cross-sectional rates for the reference period (*i.e.*, the period function could be constrained to be 0 at a reference date, p_0, the age-effects at $a_0 = p_0 - c_0$ would equal the fitted rates for period p_0 and cohort c_0, the period effects are interpretable as log-RRs relative to the reference period p_0, and the cohort effects would be residual log-RRs relative to p_0).

A variant of this approach is to extract the drift (δ) and report it as a parameter and then report both cohort and period effects as 'residuals'.

$$\log[\lambda(a, p)] = \widetilde{f_c}(a) + \delta\,(c - c_0) + \tilde{g}(p) + \tilde{h}(c)$$

$$\log[\lambda(a, p)] = \widetilde{f_p}(a) + \delta\,(p - p_0) + \tilde{g}(p) + \tilde{h}(c)$$

where $\tilde{g}(p)$ and $\tilde{h}(c)$ have 0 slope, $\widetilde{f_c}(a)$ are the age-specific rates in the reference cohort c_0 and $\widetilde{f_p}(a)$ are the age-specific rates in the reference period p_0. Hence, age-specific rates can be chosen to refer to either a specific cohort (longitudinal rates) or a specific period (cross-sectional rates).

5.1.5   *Modeling effects*

The usual approach to model effects uses one parameter per distinct value of a, p, c, by defining the variables as 'factors' (*i.e.*, class variables). The classical approach has been to define a tabulation

sufficiently coarse to avoid an excess amount of parameters in the modeling (*i.e.*, typically in 5-year intervals).

Another approach is to model the effects by parametric smoothing functions of the class mean, since the three variables age, period, and cohort are originally continuous variables, such as the followings:

- Splines (*i.e.*, 1st, 2nd, 3rd degree polynomials in predefined intervals, constrained to have identical values and derivatives in interval boundaries called knots)

- Natural splines (*i.e.*, 3rd degree splines constrained to be linear beyond the outermost knots)

- Fractional polynomials (*i.e.*, combination of polynomials of various power, including non-integer powers).

All these models are just generalized linear models. If sufficient data are available there will be little differences between these approaches. If the number of parameters in the terms describing an effect equals the number of categories, then the model will be the same as the factor model (*i.e.*, parametric models are sub-models of the classical factor model). Standard techniques of penalizing the roughness of the effects are available for tuning the number of parameters and the location of knots. However, these methods are not always desirable in describing demographic effects where sudden changes may occur (*e.g.*, due to changes in diagnostic procedures).

Deviance statistics, that are the likelihood-ratio test of each of the following models against the model with a completely freely varying interaction between age and period (or cohort), are usually produced.

Model	$\log[\lambda(a, p)]$
Age	$f(a)$
Age-drift* (cohort)	$f(a) + \delta c$
Age-cohort	$f(a) + h(c)$
Age-period-cohort	$f(a) + g(p) + h(c)$
Age-period	$f(a) + g(p)$
Age-drift* (period)	$f(a) + \delta p$

*The drift models are identical

However, using deviance statistic for choosing the model is not recommended, as deviance statistic depends on the chosen tabulation rather than on the adequacy of the model in describing rates.

5.2   Age-period cohort analysis of PSA testing rates

5.2.1   Basic plots

Four basic plots were performed in order to preliminary evaluate the effects of age, period, and cohort on PSA testing rates time trends.

1) Rates for each age *versus* period (Fig. 5.2).

2) Rates for each age *versus* cohort (Fig. 5.3).

3) Age-specific rates for each period (Fig. 5.4).

4) Age-specific rates for each cohort (Fig. 5.5).

These plots are usually on the log-scale. The plots (1) and (3) indicate whether the major variations in the rates are by period, in which case the curves should be approximately parallel. The plots (2) and (4), are useful for seeing whether the major variations in the rates are by cohort, in which case the curves should be parallel.

The plots (Fig. 5.2-5.5) indicated a more important effect of period rather than of birth cohort on PSA testing rates. The period effect was almost completely due to the change in rates after 1998-2002; whereas, no clear difference across cohorts emerged. Therefore, the period 1998-2002 was selected as reference for the period effect, and cohort effects were constrained to be equal on average. As a consequence:

- The age effects are interpretable as age-specific rates in period 1998-2002 after adjustment for the cohort effects

- The period effects are interpretable as RRs relative to the reference period 1998-2002

- The cohort effects were constrained to be 1 on average with 0 slope, and are interpretable as the RRs relative to the age-period prediction (residual RR).

The age-period-cohort analyses were performed in R, both using classical factor model and natural splines (Appendix B). The analyses were restricted to men 40-84 years old at PSA testing date.

Ten-year birth cohorts were approximated by subtracting the mid-point of the 5-year age group (from 40-44 to 80-84 years) from the corresponding 5-year period (1998-2002, 2003-2007, 2008-2012). For instance, the age class 65-69 years, with mean age 67.5 in the period 1998-2002 with mean date of diagnosis 2000.5 (i.e., 1st July 2000), has mean date of birth 2000.5 – 67.5 = 1933, but comprises men born between 1st January 1928 (1998 – 70 = 1928) and 31st December 1937 (2002 – 65 = 1937) (Fig. 5.1).
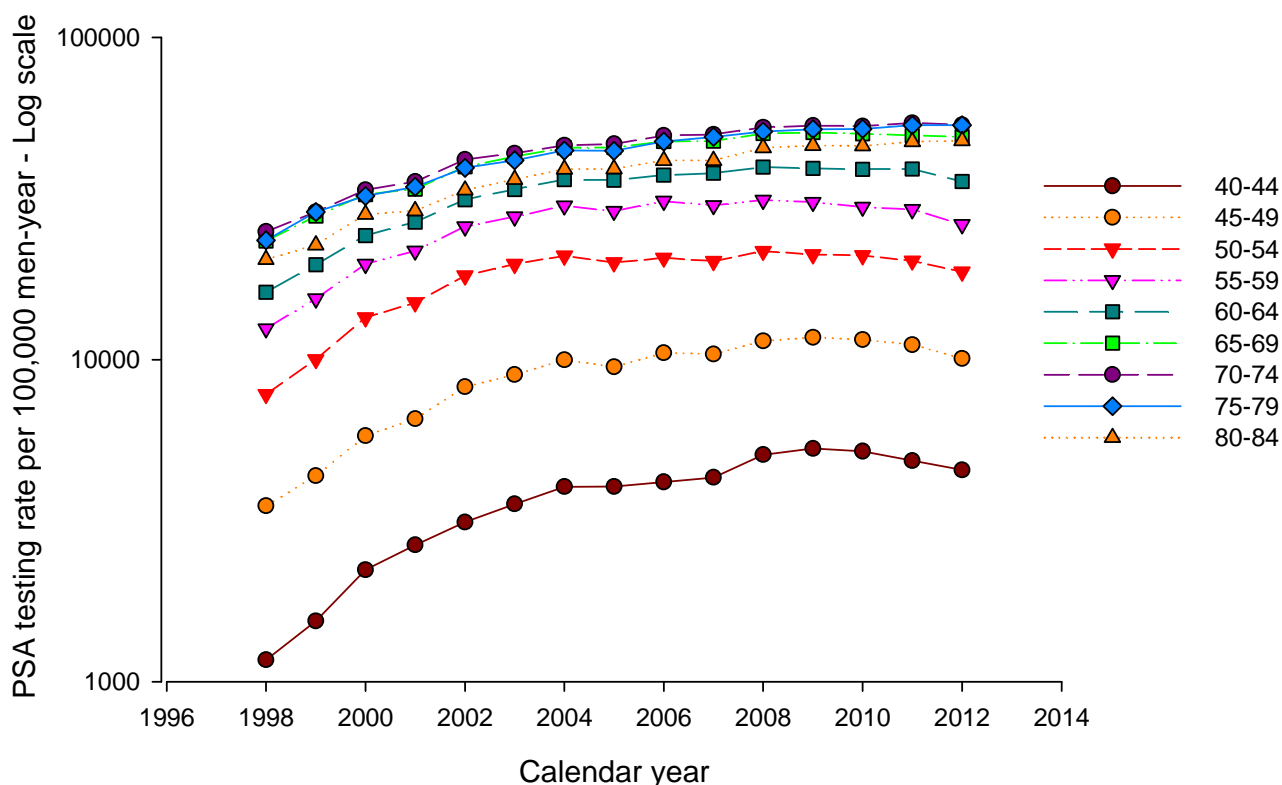
Figure 5.2. PSA testing rates per 100,000 men-year (log-scale) for each age class plotted against calendar year at PSA testing. Friuli Venezia Giulia, 1998-2012
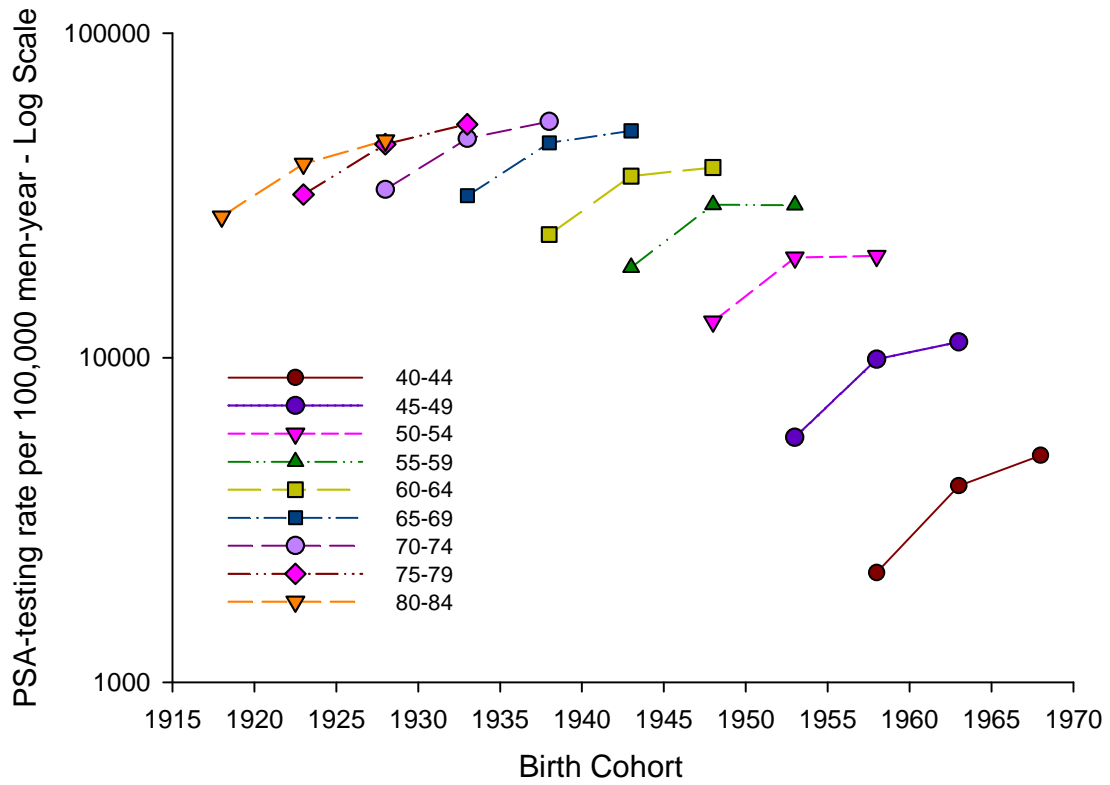
Figure 5.3. PSA testing rates per 100,000 men-year (log-scale) for each age class plotted against birth cohort. Friuli Venezia Giulia, 1998-2012
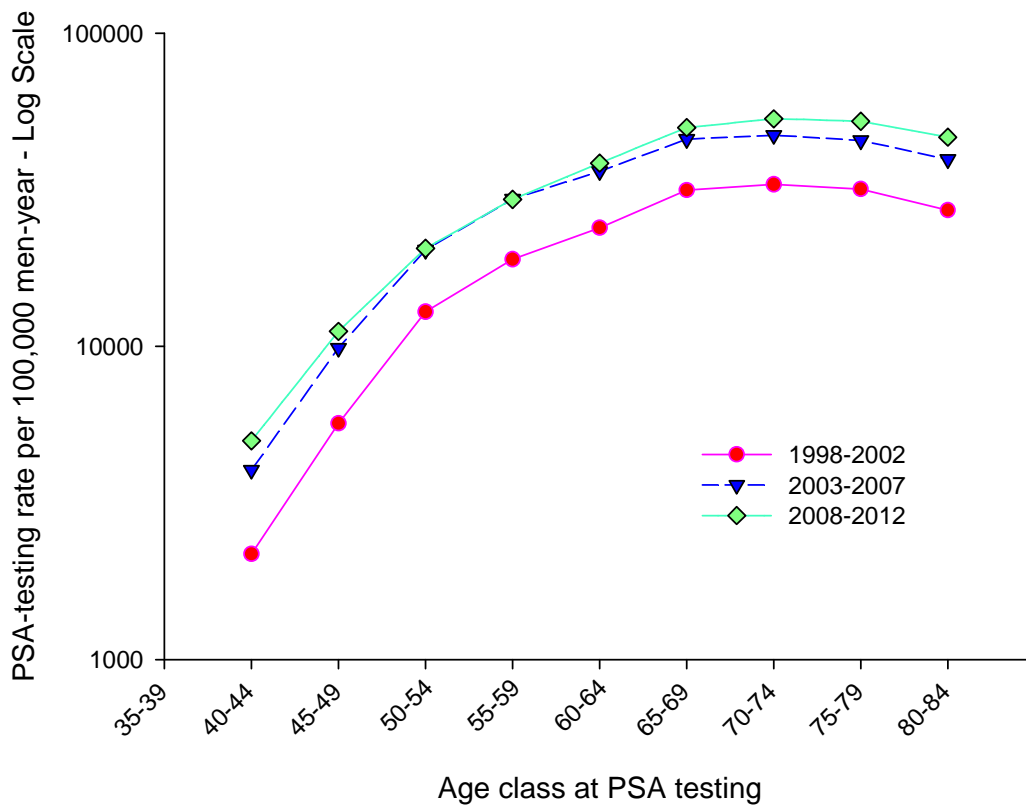
Figure 5.4. Age-specific PSA testing rates per 100,000 men-year (log–scale) by calendar period at PSA testing. Friuli Venezia Giulia, 1998-2012
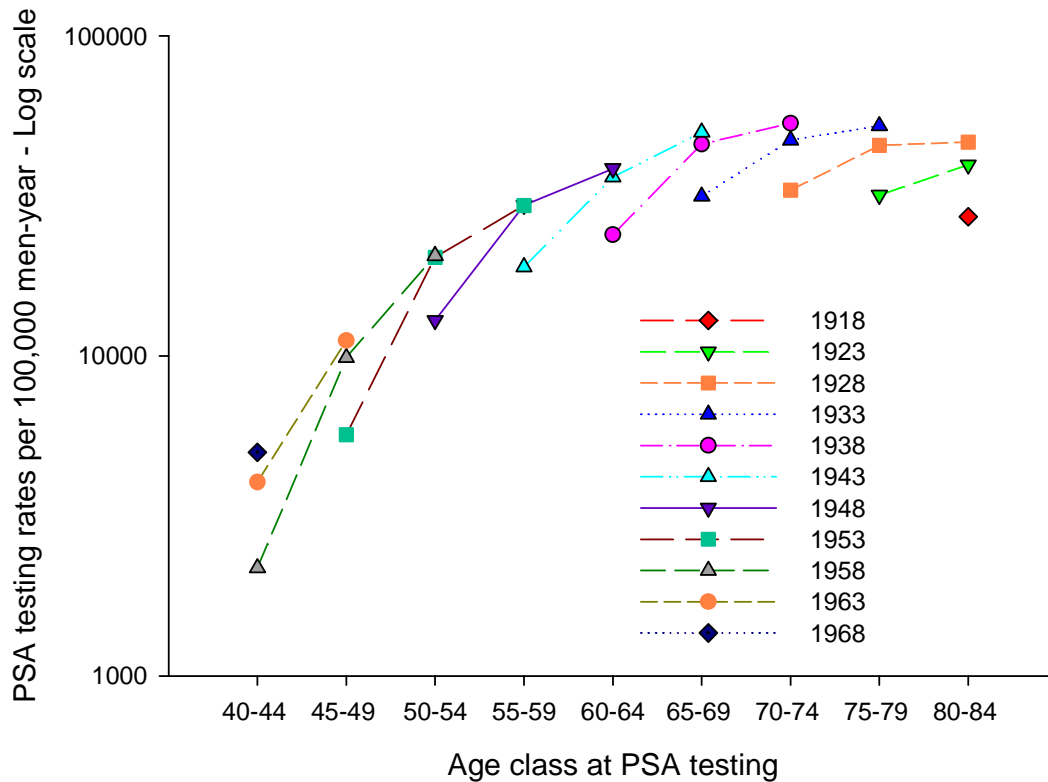
Figure 5.5. Age-specific PSA testing rates per 100,000 men-year (log–scale) by birth cohort. Friuli Venezia Giulia, 1998-2012

5.2.2 *Factor model*

The results of age-period-cohort analysis with factor model and deviance statistics are reported in Table 5.1 (for mean age, mean period, and mean cohort) and plotted in Figure 5.6. Age-specific PSA testing rates estimates were sharply increasing up to the mean age 72.5 (*i.e.*, age class 70-75 years) and slightly reduced thereafter. As compared to the period 1998-2002, PSA testing rates were 1.5-fold (95% CI: 1.50-1.51) and 1.6-fold higher (95% CI: 1.63-1.64) for the periods 2002-2008 and 2008-2012, respectively. No particular cohort effects emerged, except for increasing PSA testing rates for men born after 1960. The estimated drift was 1.047 (95% CI: 1.0465-1.0475), meaning that rates increase at the pace of 4.7% each year (period or cohort) for all age classes.

Table 5.1. Maximum likelihood estimates of age-period-cohort effects for PSA testing rates estimated with factor model and analysis of deviance. Friuli Venezia Giulia, 1998-2012

```
                        Age
        Age     Rate      2.5%      97.5%
        42.5  2427.519  2383.537   2472.312
        47.5  6456.883  6386.006   6528.546
        52.5 13442.762 13340.276  13546.036
        57.5 19221.785 19094.285  19350.136
        62.5 23817.832 23670.292  23966.291
        67.5 30568.037 30388.953  30748.177
        72.5 31999.080 31811.567  32187.698
        77.5 31176.170 30960.684  31393.156
        82.5 27481.201 27205.600  27759.595
                      Period
         Per     P-RR      2.5%      97.5%
        2000.5 1.000000  1.000000  1.000000
        2005.5 1.506169  1.498699  1.513676
        2010.5 1.635976  1.628081  1.643908
                      Cohort
        Coh      C-RR      2.5%       97.5%
        1918 0.9898689 0.9715140  1.0085705
        1923 0.9889873 0.9805391  0.9975082
        1928 1.0074120 1.0019985  1.0128548
        1933 1.0104333 1.0058013  1.0150867
        1938 1.0070713 1.0024257  1.0117384
        1943 1.0007073 0.9957160  1.0057236
        1948 0.9952955 0.9900798  1.0005388
        1953 0.9544327 0.9486290  0.9602719
        1958 0.9556077 0.9480268  0.9632493
        1963 1.0691179 1.0556067  1.0828020
        1968 1.2578845 1.2283499  1.2881292
                       Drift
            exp(Est.)    2.5%      97.5%
        APC  1.046510  1.046015  1.047006
        A-d  1.047002  1.046530  1.047475


                 Analysis of deviance

         Resid. Df Resid. Dev Df  Deviance  Pr(>Chi)
                      Age                 18      48560
Age-drift               17      8081   1   40478 < 2.2e-16
Age-Cohort               8      7489   9     592 < 2.2e-16
Age-Period-Cohort        7       728   1    6761 < 2.2e-16
Age-Period              16      1362  -9    -634 < 2.2e-16
Age-drift               17      8081  -1   -6719 < 2.2e-16
```

Figure 5.6. Estimated effects of age-period-cohort model for PSA testing rates with factor model. Age effects are reported as rates per 100,000 men-year, period and cohort effects are reported as rate ratios (RR). Friuli Venezia Giulia

5.2.3 *Natural splines*

Age, period, and cohort estimates were also modeled using natural spline functions. The best-fitting model, defined as the one minimizing the AIC [Akaike et al, 1973], was found for an APC model with 9 parameters for age, 2 for the period, and 5 for the cohort. The results of age-period-cohort model, deviance statistics, position of knots of this model are reported in Table 5.2. Results were totally comparable to those estimated with the factor model, as expected due to the similar number of parameters.

Table 5.2. Age-period-cohort effects for PSA testing rates, identified through natural splines with 9 knots for age, 2 for period and 5 for cohort, analysis of deviance, and position of knots. Friuli Venezia Giulia 1998-2012

```
                          Age
           Age      Rate      2.5%      97.5%
           42.5  2429.674   2386.235   2473.903
           47.5  6455.204   6387.538   6523.586
           52.5 13443.208  13342.030  13545.154
           57.5 19240.143  19114.384  19366.729
           62.5 23821.204  23674.374  23968.944
           67.5 30538.321  30360.789  30716.891
           72.5 32053.651  31867.793  32240.594
           77.5 31160.371  30946.235  31375.988
           82.5 27488.273  27214.296  27765.007
                        Period
           Per    P-RR       2.5%      97.5%
           2000.5 1.000000  1.000000  1.000000
           2005.5 1.505572  1.498132  1.513050
           2010.5 1.635669  1.627780  1.643597
                        Cohort
           Coh     C-RR       2.5%      97.5%
           1918 0.9748847 0.9607054 0.9892733
           1923 0.9949688 0.9884170 1.0015641
           1928 1.0080206 1.0038048 1.0122540
           1933 1.0083085 1.0054620 1.0111630
           1938 1.0050177 1.0010482 1.0090029
           1943 1.0065176 1.0027764 1.0102728
           1948 0.9914833 0.9871371 0.9958485
           1953 0.9526125 0.9486713 0.9565701
           1958 0.9599768 0.9542127 0.9657757
           1963 1.0676296 1.0576875 1.0776652
           1968 1.2552934 1.2297908 1.2813248
                         Drift
                exp(Est.)    2.5%      97.5%
           APC  1.046498  1.046002  1.046993
           A-d  1.047002  1.046530  1.047475


                  Analysis of deviance
                Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
   Age                18      48560
   Age-drift          17       8081   1    40478 < 2.2e-16 ***
   Age-Cohort         13       7545   4      536 < 2.2e-16 ***
   Age-Period-Cohort  12        746   1     6800 < 2.2e-16 ***
   Age-Period         16       1362  -4     -617 < 2.2e-16 ***
   Age-drift          17       8081  -1    -6719 < 2.2e-16 ***


                  Knots position
                        Age
     11.11111% 22.22222% 33.33333% 44.44444% 55.55556% 66.66667%
  42.50000  46.94444  51.38889  55.83333  60.27778  64.72222  69.16667


                77.77778% 88.88889%
           73.61111  78.05556  82.50000


                        Period
                        50%
                2000.5 2005.5 2010.5


                        Cohort
                 20%   40%   60%   80%
           1918 1929 1938 1948 1957 1968
```
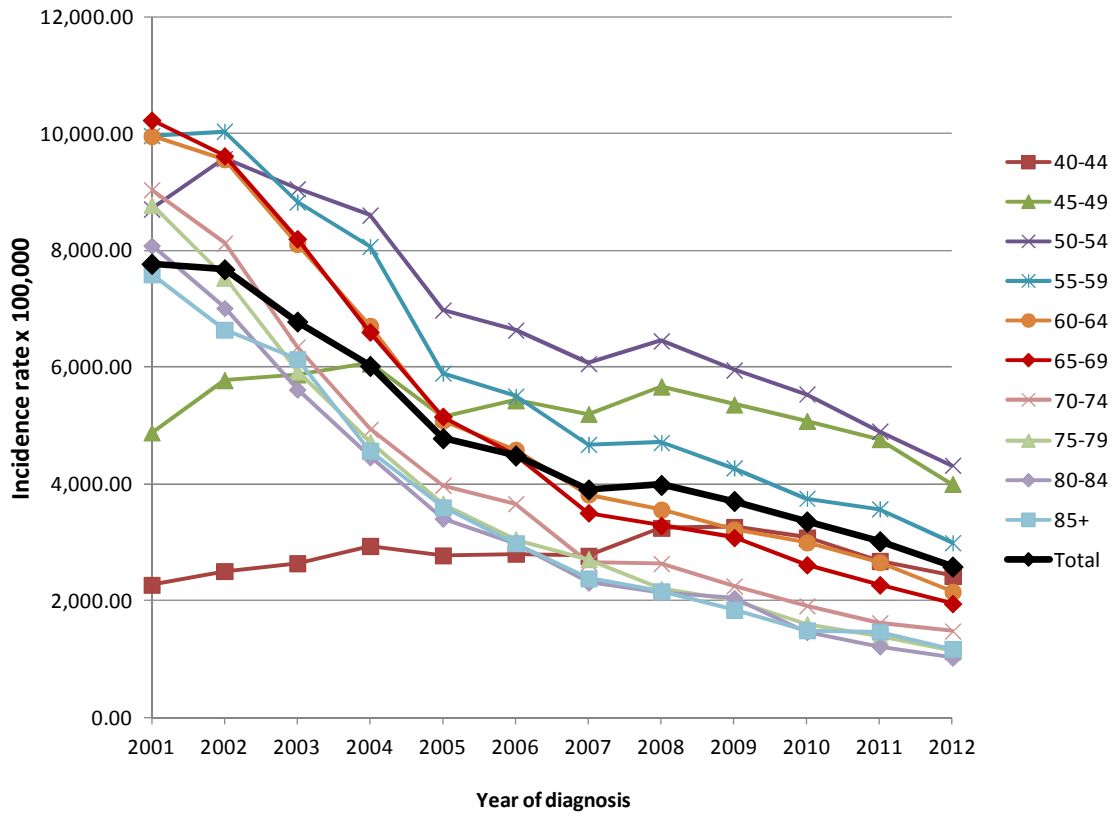
5.3　First PSA test

It should be noted that PSA testing, beyond being a screening and a diagnostic procedure, is also used for monitoring cancer recurrence among patients with PCa during the follow-up subsequent to therapies (prostatectomy included). Therefore, men with PCa are expected to undergo PSA testing for several years also after cancer onset.

In order to have some indications with regard to the amount of new potential cancer diagnoses, an analysis was performed considering only men who have been tested for PSA level for the first time. They could be either men who underwent PSA screening for the first time (without having a PCa) or men newly diagnosed with PCa (assuming that all men with PCa have at least one PSA test at the time of PCa diagnosis).

Because information on PSA test was not available in FVG before 1998, the analysis was restricted to men having had their first PSA test in the period 2001-2012, given that they had not any PSA test in the previous period 1998-2000 (assuming 3 years as a reasonable time interval for considering a man not already 'under surveillance' for PCa). First-PSA-testing rates were calculated as men tested for the first time as respect to the mean resident male population (ISTAT) (Fig. 5.7).

Dramatically decreasing trends were found for men overall (7,770 per 100,000 in 2001 down to 2,590 per 100,000 in 2012, with aPC = -9.6; 95% CI: -10.7 to -8.6) and for all age groups starting from 50-54 years; conversely, men aged less than 50 years showed firstly an increasing and then a decreasing trend.

Figure 5.7. Crude first-PSA testing rates per 100,000 men-year, by age class and calendar year at testing. Friuli Venezia

Giulia, men aged 40+ years, 2001-2012

6. Prostate cancer incidence trends analysis

PCa incidence rates in FVG among men aged 40 years or more were analyzed for the period 1995-2009, taking advantage of the availability in 2014 of –yet unofficial and unpublished (courtesy of the FVG Cancer Registry) – data for the period 2008-2009 from the cancer registry.

A total of 15,107 cases of PCa were diagnosed between 1995 and 2009 in men aged 40 or more years. The overall crude incidence rate of PCa increased from 219.8 per 100,000 men in 1995 up to 385.5 in 2007, and then decreased down to 328.3 in 2009 (Fig. 6.1). As expected, PCa incidence was very low –almost null– before 50 years of age in the whole considered period. Growing PCa incidence rates were observed for all age groups, except men older than 74 years, but the greatest increases were observed for the age classes between 55 and 69 years.
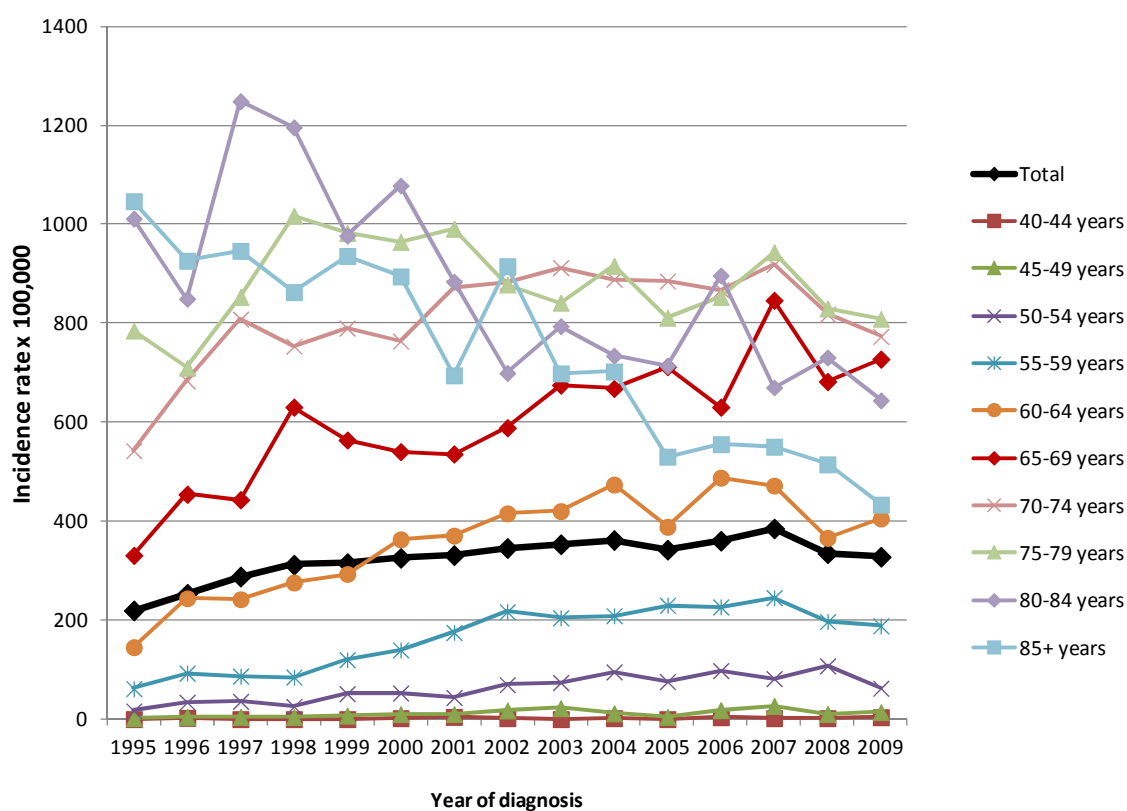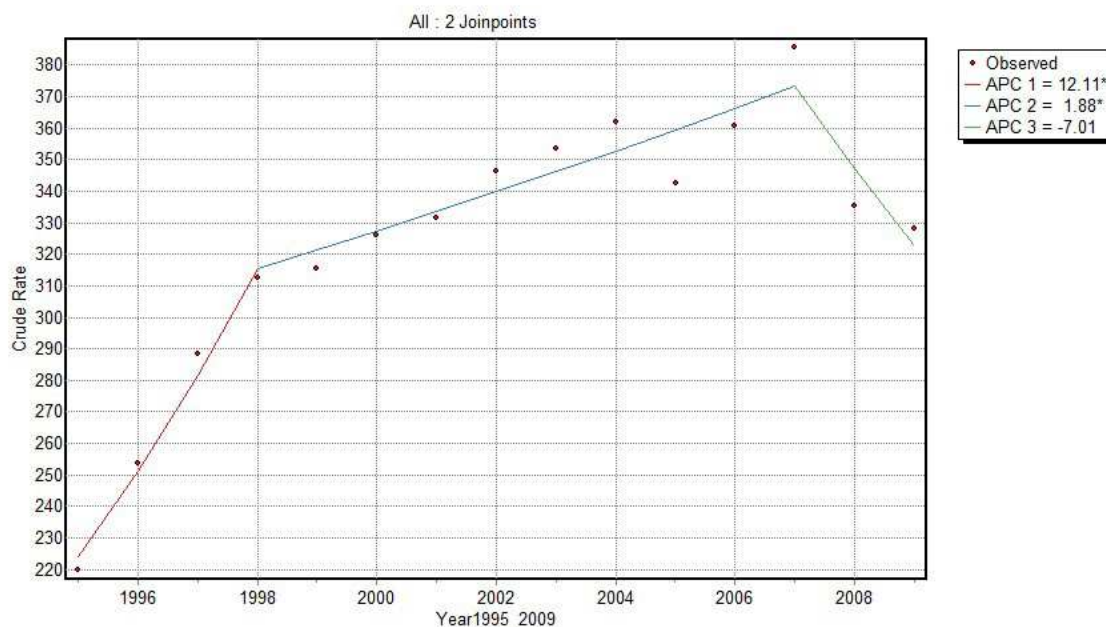

Figure 6.1. Crude prostate cancer incidence rates per 100,000 men-years by calendar year and age class at diagnosis. Friuli Venezia Giulia Cancer Registry, men 40+ years, 1995-2009[*].
[*] Yet unpublished data for the period 2008-2009, courtesy of the FVG Cancer Registry

6.1 Joinpoint analysis of prostate cancer incidence rates

Joinpoint regression analysis was used to identify points where a statistically significant change over time in log-linear slope of the PCa rates occurred (as in the previous analysis of PSA testing rates, the aPCs were computed by means of generalized linear models, assuming that random errors were heteroscedastic).

Statistically significant changes in PCa incidence rates slopes emerged in 1998 and 2007: the aPCs in the periods 1995-1998, 1998-2007, and 2007-2009 were 12.1 (95% CI: 6.6 to 17.9), 1.9 (95% CI: 1.0 to 2.8), and -7.0 (95% CI: -14.3 to 0.9), respectively (Fig 6.2).

Period	aPC	Lower bound 95% CI	Upper bound 95% CI
1995-1998	12.1	6.6	17.9
1998-2007	1.9	1.0	2.8
2007-2009	-7.0	-14.3	0.9

Figure 6.2. Joinpoint analysis of prostate cancer incidence rates (model with 2 joinpoints). Friuli Venezia Giulia, men 40+, 1995-2009[*]. aPC = annual percent change, CI = confidence interval
[*]Yet unpublished data for the period 2008-2009, courtesy of the FVG Cancer Registry.

Stratified analyses by age class, though suffering from low numbers of cases at the youngest ages, highlighted different trends across strata (data not shown): PCa incidence was increasing in the whole period for age classes below 55 years and for the age class 65-69 years; whereas, for the age

classes 55-59 and 60-64 years there was a tendency to reduction or stabilization in the last period; finally, the oldest age groups (70+ years) had decreasing PCa incidence rates in the whole period.

## 6.2   Age-period-cohort analysis of prostate cancer incidence rates

The analyses were restricted to men 40-84 years old at PCa diagnosis. Ten-year birth cohorts were approximated by subtracting the mid-point of the 5-year age group (from 40-44 to 80-84 years) from the corresponding 5-year period (1995-1999, 2000-2004, 2005-2009) (*e.g.*, for the age class 65-69 years −mean age 67.5− in the period 1995-1999 −mean date of diagnosis 1997.5− the mean date of birth was 1930, but comprises men born between 1925 and 1935).

The basic plots (Figures 6.3, 6.4, 6.5, 6.6) suggested, beyond the known age affect, also effects of both birth cohort and period. The period effect was almost completely due to the increase of rates between the first and the second period followed by a stabilization. Conversely, more complex patterns emerged for the cohort effects: the most recently born men had increasing PCa incidence rates between the first and the second period and then stabilized; the 'middle' cohorts showed increasing rates throughout the three periods; the oldest cohorts showed almost stable incidence.

Considering the period 1995-1999 as reference and constraining cohort effects to be equal on average, the age effects are interpretable as age-specific rates for the period 1995-1999, after adjustment for the cohort effects; the cohort effects are interpretable as RRs relative to the reference period 1995-1999; the cohort effects are interpretable as the RRs relative to the age-period prediction (residual RRs).
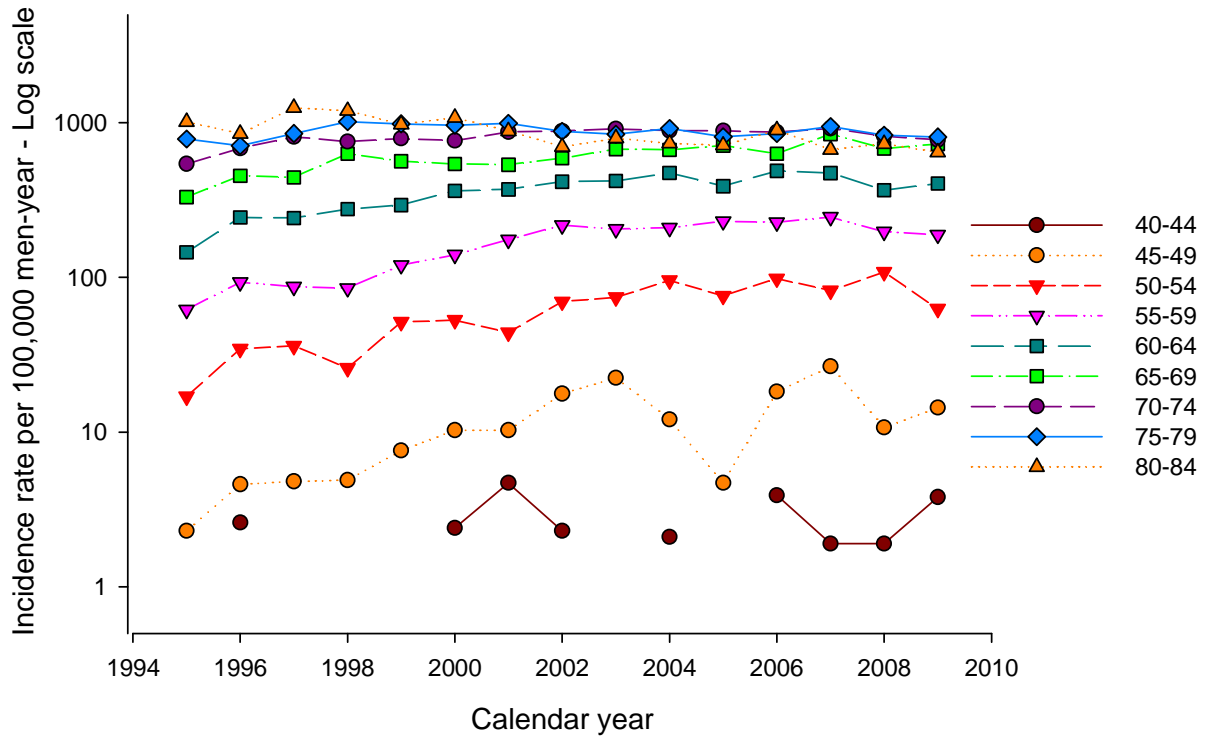
Figure 6.2. Prostate cancer incidence rates per 100,000 men-year (log-scale) for each age class plotted against calendar year at diagnosis. Friuli Venezia Giulia Region, men aged 40-84 years, 1995-2009[*]
[*]Yet unpublished data for the period 2008-2009, courtesy of the FVG Cancer Registry.
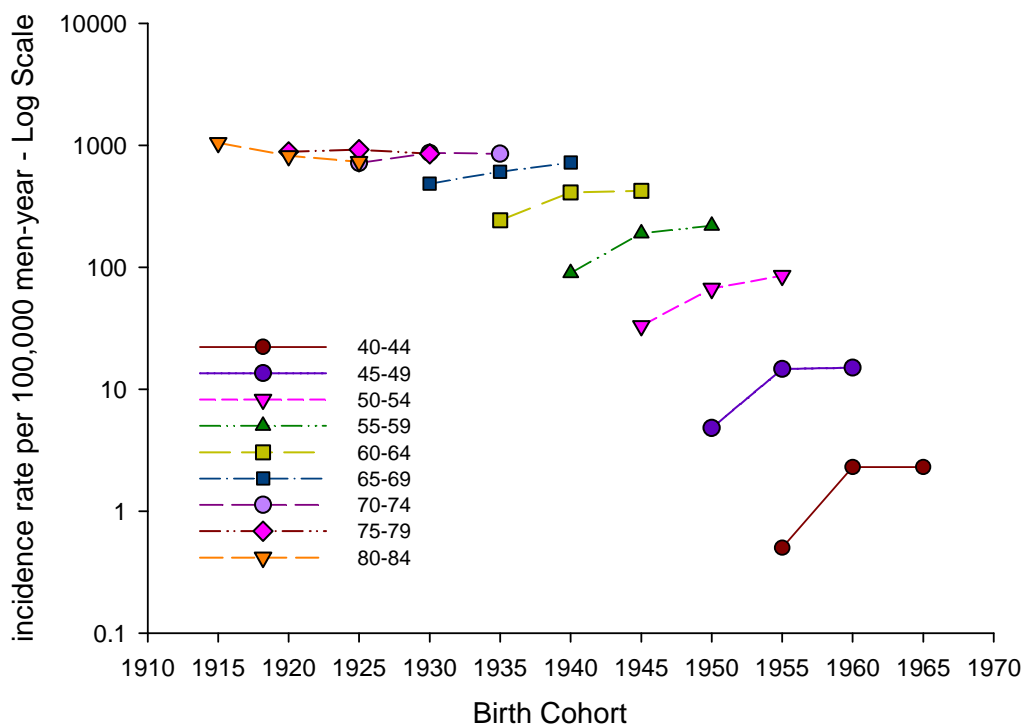
Figure 6.3. Prostate cancer incidence rates per 100,000 men-year (log-scale) for each age class plotted against birth cohort. Friuli Venezia Giulia Region, men aged 40-84 years, 1995-2009[*]
[*]Yet unpublished data for the period 2008-2009, courtesy of the FVG Cancer Registry

60
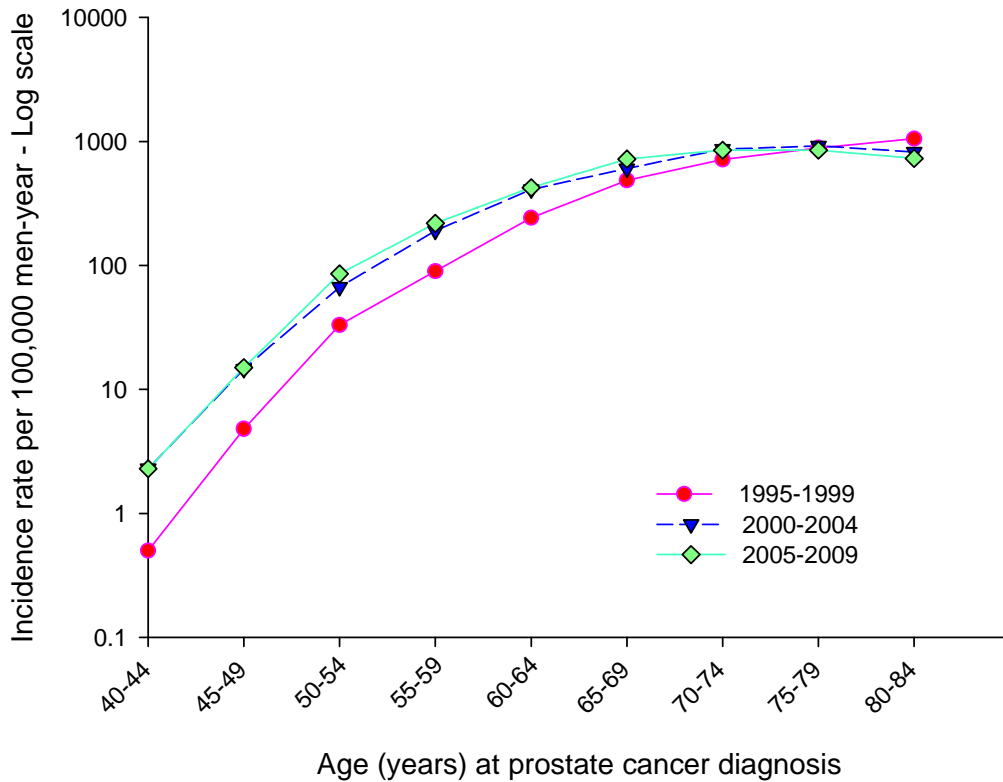
Figure 6.4. Age-specific prostate cancer incidence rates per 100,000 men-year (log –scale), by calendar period at PSA testing. Friuli Venezia Giulia Region, men aged 40-84 years, 1995-2009[*]

[*]Yet unpublished data for the period 2008-2009, courtesy of the FVG Cancer Registry

Figure 6.5. Age-specific prostate cancer incidence rates per 100,000 men-year (log–scale) by birth cohort. Friuli Venezia Giulia Region, men aged 40-84 years, 1995-2009[*]

[*]Yet unpublished data for the period 2008-2009, courtesy of the FVG Cancer Registry
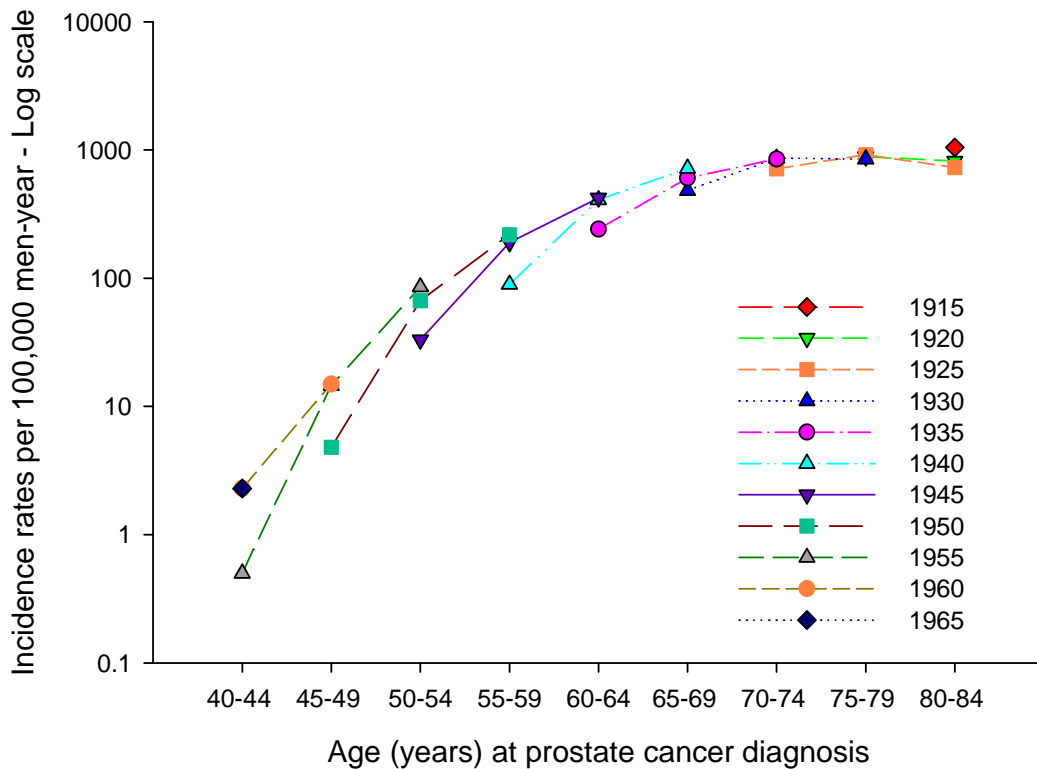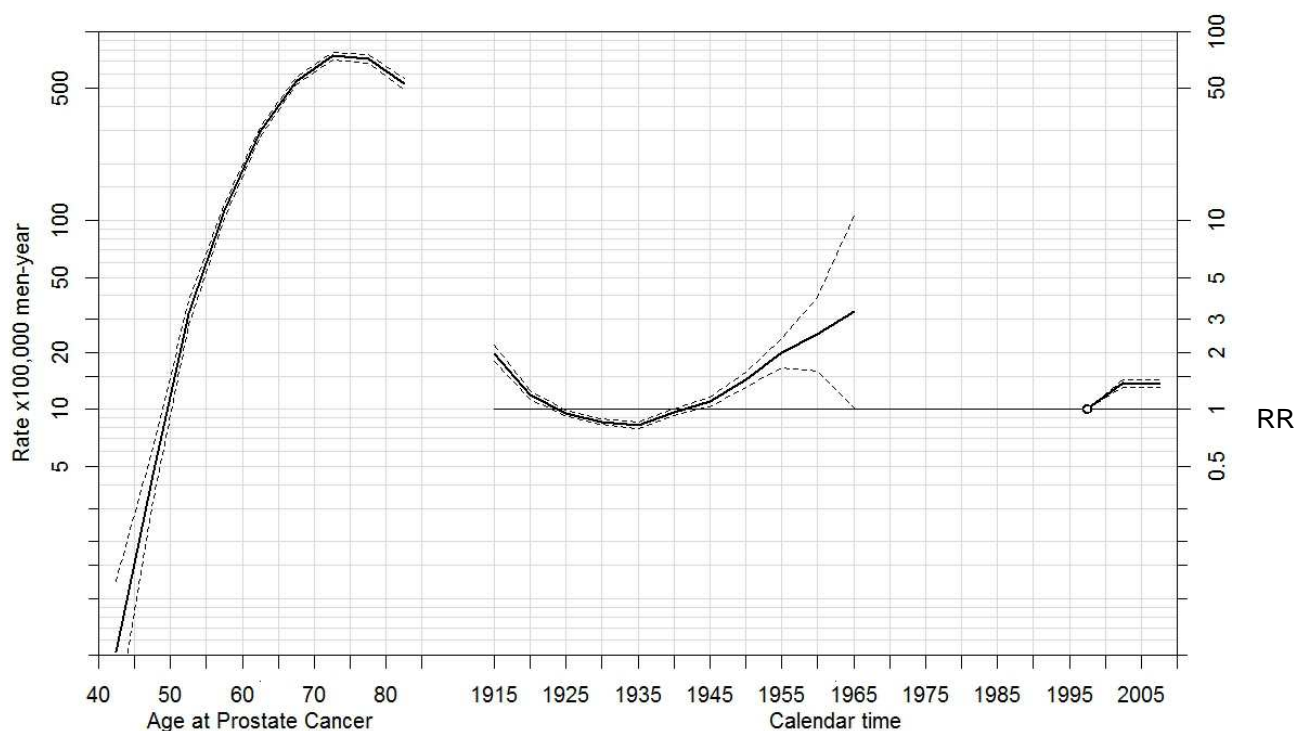
The results of age-period-cohort analysis with factor model are reported in Table 6.1 and plotted in Figure 6.3. Cross-sectional age-specific PCa incidence rates estimates in the period 1995-1999 were sharply increasing up to the mean age class 70-75 years and reduced thereafter. As compared to the period 1995-1999, incidence rates were 1.4-fold higher (95% CI: 1.3-1.4) in the subsequent periods. An U-shaped cohort effect emerged, with more than 2-fold increase of RRs for men born since 1955. The estimated age-drift was 2.3% (95% CI: 1.9%-2.7%) per year.

Table 6.1. Maximum likelihood estimates of age-period-cohort effects with factor model for prostate cancer incidence. Friuli Venezia Giulia, men aged 40-84 years, 1995-2009[*].

```
                              Age
              Age      Rate       2.5%       97.5%
              42.5   0.5193459   0.2199192    1.226451
              47.5   4.4970416   3.2170844    6.286246
              52.5  32.0204454  27.7406472   36.960526
              57.5 112.2174249 103.4575711  121.718984
              62.5 295.2430945 278.4794266  313.015888
              67.5 547.0890697 520.6926803  574.823618
              72.5 744.4386480 711.4365445  778.971652
              77.5 718.7318527 683.8425535  755.401186
              82.5 530.0849055 494.5550783  568.167267
                            Period
               Per    P-RR      2.5%      97.5%
             1997.5 1.000000 1.000000  1.000000
             2002.5 1.372309 1.311457  1.435984
             2007.5 1.368713 1.310946  1.429025
                            Cohort
              Coh    C-RR       2.5%       97.5%
             1915 1.9812000 1.8020217   2.1781943
             1920 1.1889220 1.1270688   1.2541696
             1925 0.9584097 0.9258770   0.9920855
             1930 0.8621196 0.8306825   0.8947465
             1935 0.8211840 0.7883758   0.8553575
             1940 0.9625678 0.9235295   1.0032564
             1945 1.0966895 1.0380661   1.1586235
             1950 1.4348490 1.3093996   1.5723172
             1955 1.9882658 1.6566656   2.3862395
             1960 2.5093354 1.6010833   3.9328148
             1965 3.2817687 1.0193887  10.5651609
                            Drift
                  exp(Est.)    2.5%     97.5%
              APC  1.029528 1.025159 1.033915
              A-d  1.023194 1.019034 1.027372

       Analysis of deviance for Age-Period-Cohort model
                 Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
            Age               18       443.38
  Age-drift          17      320.96  1  122.418 < 2.2e-16 ***
  Age-Cohort          8       98.77  9  222.189 < 2.2e-16 ***
  Age-Period-Cohort   7       25.35  1   73.423 < 2.2e-16 ***
  Age-Period         16      294.67 -9 -269.317 < 2.2e-16 ***
```
[*]Yet unpublished data for the period 2008-2009, courtesy of the FVG Cancer Registry

Figure 6.3. Estimated effects of age-period-cohort analysis for PCa incidence rates with factor model. Age effects are reported as rates per 100,000 men-year, period and cohort effects are reported as rate ratios (RR). Dashed lines represent 95% confidence intervals. Friuli Venezia Giulia, men aged 40-84 years, 1995-2009[*]
[*]Yet unpublished data for the period 2008-2009, courtesy of the FVG Cancer Registry

Age, period, and cohort estimates were also modeled using natural spline functions. The best-fitting model, defined as the one minimizing the AIC, was found for an APC model with 5 parameters for age, 2 for the period, and 3 for the cohort. The results of age-period-cohort model effects are reported in Table 6.2 and plotted in Figure 6.4. Estimates were comparable to those derived from the factor model but with narrower confidence intervals (Fig. 6.4).

Table 6.2. Age-period-cohort effects identified through natural splines (with 5 knots for age, 2 for period, and 3 for cohort) and position of knots. Friuli Venezia Giulia, men aged 40-84 years, 1995-2009[*].

```
                                Age
                 Age      Rate       2.5%        97.5%
                 42.5   1.574411   1.228764     2.017286
                 47.5   6.538197   5.492132     7.783503
                 52.5  27.151762  24.414868    30.195459
                 57.5 106.994066 100.542092   113.860076
                 62.5 307.787904 293.182601   323.120791
                 67.5 543.955944 518.254311   570.932190
                 72.5 751.416460 718.964081   785.333665
                 77.5 713.028205 678.890898   748.882070
                 82.5 538.705232 503.056900   576.879728


                              Period
                Per       P-RR      2.5%      97.5%
                1997.5 1.000000 1.000000 1.000000
                2002.5 1.359145 1.299252 1.421799
                2007.5 1.363957 1.306500 1.423941


                              Cohort
                Coh      C-RR       2.5%      97.5%
                1915 1.7750754 1.6433898 1.9173131
                1920 1.2716607 1.2334780 1.3110252
                1925 0.9682931 0.9477587 0.9892725
                1930 0.8329229 0.8115997 0.8548065
                1935 0.8403689 0.8154174 0.8660839
                1940 0.9472657 0.9221808 0.9730329
                1945 1.1438909 1.1245534 1.1635610
                1950 1.4297263 1.3492651 1.5149856
                1955 1.7972718 1.6149831 2.0001360
                1960 2.2593037 1.9317732 2.6423667
                1965 2.8401121 2.3102844 3.4914473


                              Drift
                    exp(Est.)    2.5%      97.5%
                APC  1.029286 1.024925 1.033665
                A-d  1.023107 1.018946 1.027284


            Analysis of deviance for Age-Period-Cohort model
                          Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
Age                              21     486.44
Age-drift                        20     364.96  1  121.489 < 2.2e-16 ***
Age-Cohort                       18     159.77  2  205.183 < 2.2e-16 ***
Age-Period-Cohort                17      92.55  1   67.225 2.422e-16 ***
Age-Period                       19     338.99 -2 -246.444 < 2.2e-16 ***
Age-drift                        20     364.96 -1  -25.964 3.478e-07 ***


                              Knots
                               Age
                  52.5 62.5 67.5 72.5 77.5 82.5


                             Period
                        1997.5 2002.5 2007.5


                             Cohort
                        1915 1930 1935 1950
```
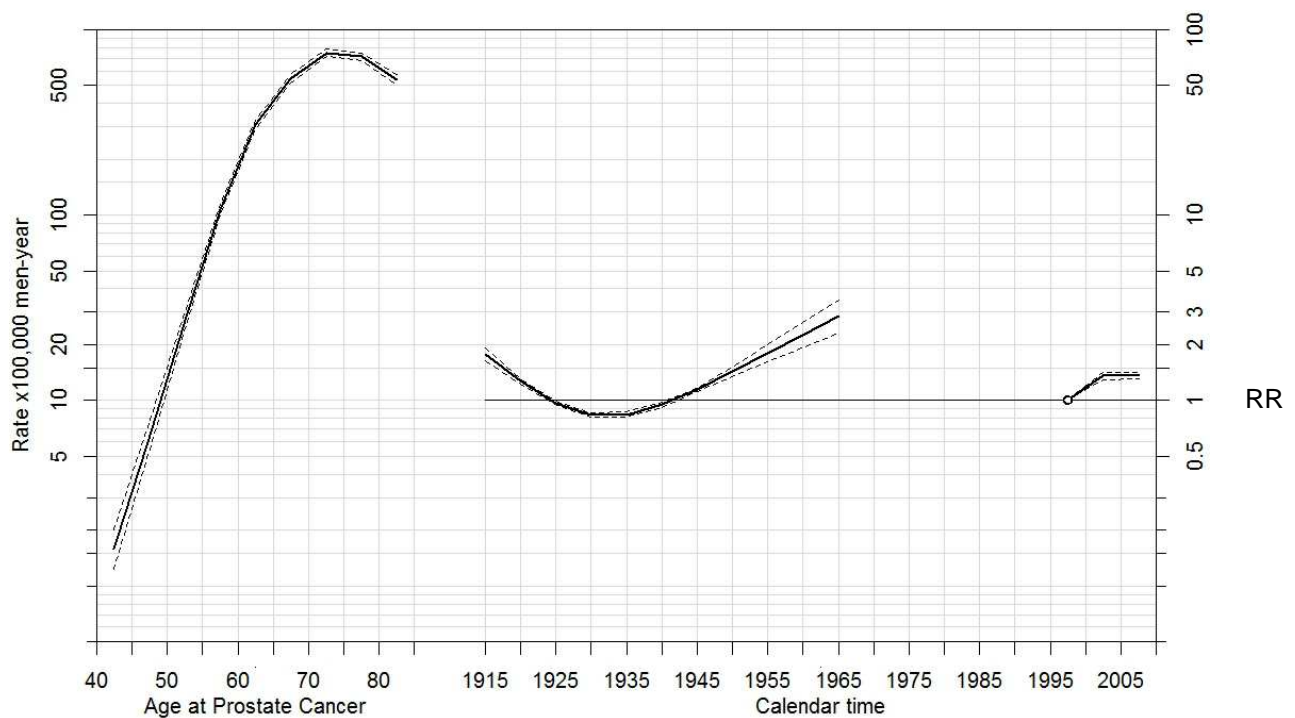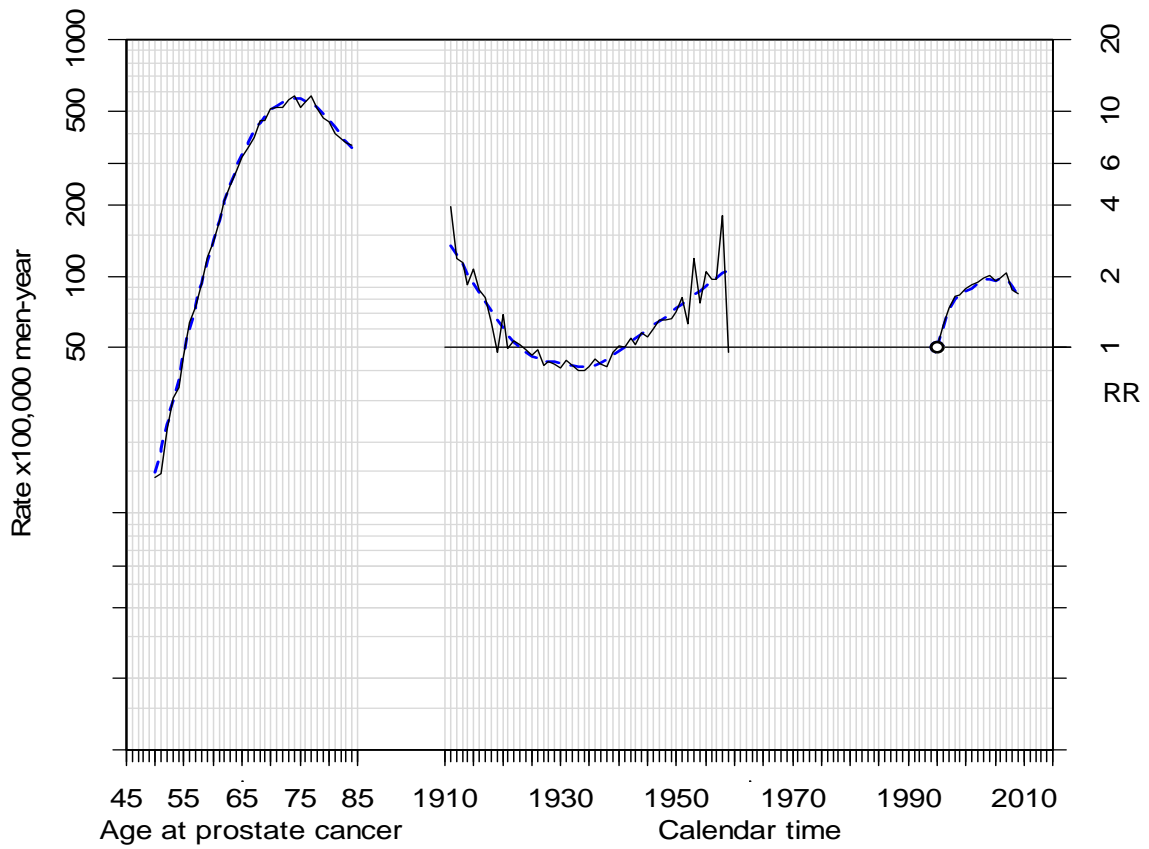
[*]Yet unpublished data for the period 2008-2009, courtesy of the FVG Cancer Registry

Figure 6.4. Estimated effects of age-period-cohort analysis for PCa incidence rates with natural splines model. Age effects are reported as rates per 100,000 men-year, period and cohort effects are reported as rate ratios (RR). Dashed lines represent 95% confidence intervals. Friuli Venezia Giulia, men aged 40-84 years, 1995-2009[*]
[*]Yet unpublished data for the period 2008-2009, courtesy of the FVG Cancer Registry

In order to capture the changes in the trend of PCa incidence rates observed by the joinpoint analysis in the last years, the age-period-cohort analysis was also performed using 1-year intervals for age and calendar time (this analysis was restricted to men 50+ years, due to elevated number of missing counts in each cell of the Lexis diagram at younger ages). The results, using both factor model and natural splines (the lowest AIC was found for a model with 4 parameters for age, 6 for the period, and 5 for the cohort), considering as reference period the year 1995, are reported in Figure 6.5. The estimated age-drift was 2.3% (95% CI: 1.9%-2.7%) per year (equal to the 5-year interval analyses), but the period effects highlighted decreasing RRs in the last 3-4 years, after a higher increase up to 2007.

Figure 6.5. Estimated effects of age-period-cohort analysis for PCa incidence rates with factor model (continuous lines) and natural splines (dashed lines) using 1-year intervals. Age effects are reported as rates per 100,000 men-year, period and cohort effects are reported as rate ratios (RR). Friuli Venezia Giulia, men aged 50-84 years, 1995-2009[*]
[*]Yet unpublished data for the period 2008-2009, courtesy of the FVG Cancer Registry

6.3 Differences between observed and estimated prostate cancer incidence rates up to 2009

Considering all ages, and excluding cases diagnosed solely at autopsy as in the previous comparison (Fig. 3.2), observed PCa incidence data up to 2009 were compared with the expected incidence rates previously estimated using MIAMOD (Fig. 6.6). Observed rates were higher than those expected, with differences ranging between 7% and 18% (median 11%) in the period 1996-2007, but rates seemed to converge after 2007.
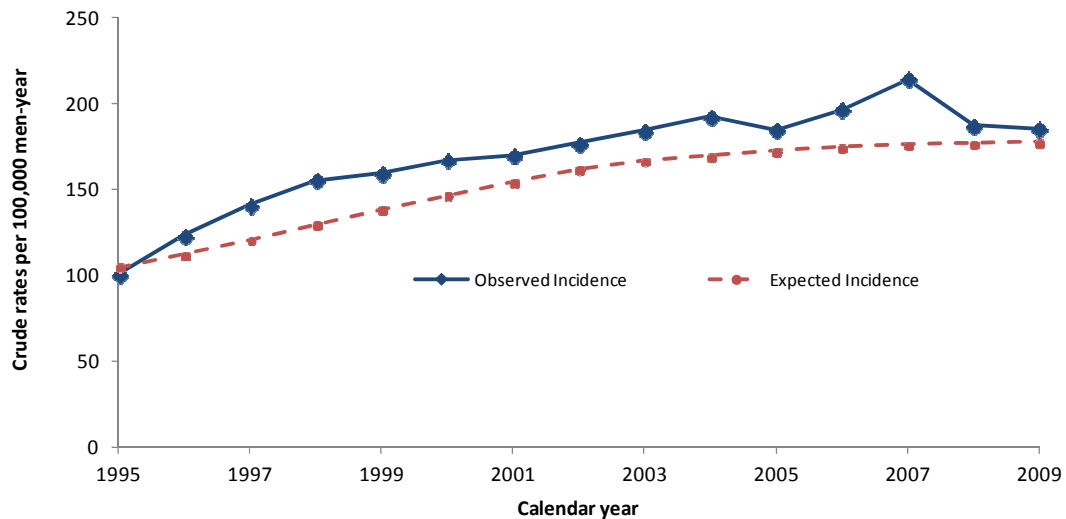
Year	Obs. rate	Exp. rate	Obs./Exp. rate ratio	Difference
1995	100.3	105.0	0.96	-4.7%
1996	123.4	113.0	1.09	8.4%
1997	140.9	121.7	1.16	13.6%
1998	155.2	130.6	1.19	15.9%
1999	159.4	139.3	1.14	12.6%
2000	167.0	147.4	1.13	11.7%
2001	169.9	154.9	1.10	8.8%
2002	176.8	161.4	1.10	8.7%
2003	184.2	166.7	1.11	9.5%
2004	192.5	169.8	1.13	11.8%
2005	184.6	172.4	1.07	6.6%
2006	196.4	174.6	1.12	11.1%
2007	214.4	176.2	1.22	17.8%
2008	187.4*	176.9	1.06	5.6%
2009	185.3*	177.9	1.04	4.0%

Figure 6.6. Crude prostate cancer incidence rates (per 100,000 men-year). Observed (continuous lines) and estimated using the MIAMOD method (dashed lines). Friuli Venezia Giulia, men (all ages), 1995-2009[*]
[*]Yet unpublished data for the period 2008-2009

7. DISCUSSION

Joinpoint and age-period-cohort analyses of PSA testing rates in the period 1998-2012 clearly highlighted a period effect on the use of PSA testing in FVG among men aged 40 years or more. In particular, the PSA testing rates in the period 1998-2002 increased by 17% per year, then by 4% per year up to 2008, and then stabilized. The period effect was cross-sectional and involved all the age groups and birth cohorts at the same time. As compared to the period 1998-2002, the PSA testing rates were found to be 1.5-fold higher in the period 2003-2007 and seemed to stabilize thereafter. Age effect reflects the tendency of men of being tested with PSA during life: this effect was increasing with age up to the class 70-74 years. Cohort effects, which are longitudinal and reflect possible differences in the spread in the use of PSA testing across different birth cohorts, were found to be not so important as the period and age effects, but suggested that the youngest birth cohorts were more prone to be screened through PSA testing.

While an appropriate use of PSA testing among men in the oldest age groups (up to 70 years) cannot be excluded, the use of this test for screening men aged 40-49 years is by far more controversial, given the almost null incidence of PCa before 50 years, except for sub-groups of men at particular high risk (*e.g.*, those with first-degree relatives having had a PCa at young age) [Carter et al, 2013; Heidenreich et al, 2013].

The analysis of first-PSA testing rates between 2001 and 2012 showed dramatic decreasing trends at all ages. Only the youngest men (*i.e.*, under 50 years) showed firstly an increase and then a reduction in the very last years. Beyond confirming a tendency to lessen the use of PSA testing in FVG, this decreasing trend may be also due to an already reached –almost– complete saturation of the potential target population for PSA screening in this region, especially at older ages (*e.g.*, in the period 2008-2012, approximately 1 out of 3 men aged 40+ years has been tested with PSA, and 1 out of 2 men among those aged 65-79 years).

The analysis of PCa incidence rates trends in FVG also highlighted a steep increase of PCa incidence rates between 1995 and 1998 (12% per year), a lower increase up to 2007 (2% per year) followed by stabilization or, probably, a reduction (though not statistically significant) up to 2009. Although the data on PCa incidence derived from the FVG cancer registry for the period 2008-2009 are not official yet, they are eventually overestimated rather than underestimated (*i.e.*, 'potential' new cancer cases are automatically identified through an *ad hoc* algorithm from computerized pathological archives, hospital discharges, and death certificates databases and, in case of disagreement between data sources, they are manually revised by qualified staff and, eventually, refused). The trend of PCa incidence rates, thought the considered period was antecedent (*i.e.*, 1995-2009 *vs*. 1998-2012), resembled to some extent the trend of PSA-testing rates. Conversely to PSA-testing rates, PCa incidence rates in the age-period-cohort analysis resulted to be affected also by cohort effects, indicating a more complex baseline risk according to birth cohort. Several factors could be potentially associated to the observed increased PCa risk for more recently born men, including changes in behavioral and lifestyle factors over time (*e.g.*, dietary and sexual habits, obesity) [Leitzmann et al, 2012].

In Italy, PCa incidence trends started to rise suddenly after 1991, indicating the progressive diffusion of screening with PSA in those years [Crocetti, 2007]. In FVG, the increasing rates of PSA testing observed among all age classes especially from 1998 up to 2002, assuming that this increase started several years before (*e.g.* since 1991, as in Italy), could reasonably explain the gap between the PCa incidence rates estimated by MIAMOD and those reported by the cancer registry, especially in the period 1996-2007. Actually, MIAMOD estimates are based on mortality data which have not been so heavily modified by the introduction of PSA test as PCa incidence rates were. Hence, PCa incidence estimates produced using MIAMOD could be considered as the rates that would be observed in the absence of such a great increase of screening with PSA in FVG. It is reasonably to hypothesize that this difference between observed and estimated incidence rates was attributable to screening with PSA more in terms of overdiagnosis, rather than of early diagnosis

(also considering the concurrent stable mortality). Gathering data on Gleason score at PCa diagnosis could be useful in order to better understand whether an overdiagnosis occurred, as it should be associated to an increase of lower Gleason scores.

The convergence of the two curves observed in the last common period (2008-2009), seems to further support such hypothesis. The potential overdiagnosis proportions (between 7% and 18%) seem to be much lower as respect to those estimated on the basis of trials' results [Etzioni et al, 2013]. However, most of those proportions were referred to screening-detected PCa cases, whereas the present estimate refers to the total PCa cases and it was based on a real population. Draisma and colleagues (2009) reported a range of 9-19% of overdiagnosed cases over total cases (using 3 different simulation models) and Tsodikov and colleagues (2006) estimated an overdiagnosis up to 20% of total detected PCa. These results are in line with the present study estimates.

As suggested by Mistry and colleagues [Mistry et al, 2011], incidence data from a period prior to the extensive use of PSA testing (*e.g*., before 1990 in UK), could be used to fit APC models to predict rates in the absence of PSA testing in the subsequent periods. The predicted rates can be used to calculate age-specific observed to predicted ratios that, in turn, can be used to adjust future predictions. Unfortunately, FVG cancer registry does not cover a period before PSA screening diffusion, but the MIAMOD estimates provided an alternative.

It is worth noting that data on PSA-testing, which are produced for administrative purposes, are available also in areas not covered by cancer registration. Moreover, they are available within very few months from the 'real time' of the event (*e.g.*, in FVG, as of December 2014, PSA testing data were available up to 2013); conversely, cancer registries release incidence data several years (usually 5 or more) after the time of the event, due to the complexity of the required quality checks of data, as previously described (*e.g.*, in FVG, as of December 2014 the last published year was 2007).

7.1   Conclusion

The widespread use of opportunistic PSA testing in the last decades in high-income countries has inflated the incidence of PCa without affecting the overall mortality at the same extent. Given that areas covered by population-based cancer registries are still few –though increasing– also in developed countries, and given that data from cancer registries are usually not up to date, methods applied for estimating the incidence of PCa at a population level are tipically based on mortality data. As a consequence of the PSA-testing diffusion, these methods could produce biased estimates and, furthermore, even more biased projections of incidence and prevalence of PCa. The availability of up-to-date information on PSA-testing in several areas offers the opportunity of evaluating the estimates and projections of PCa incidence taking into account of the observed trends of PSA-testing rates.

REFERENCES

- AIRTUM web site http://www.registri-tumori.it/cms/ [Accessed 22.9.2014]

- AIRTUM-AIOM. I numeri del cancro in Italia 2013. Intermedia Ed., Brescia, 2013. Available at http://www.registri-tumori.it/PDF/AIOM2013/I_numeri_del_cancro_2013.pdf [Accessed 22.9.2014]

- Akaike H. Information theory and an extension of the maximum likelihood principle. In: Second International Symposium on Information Theory. Budapest: Second International Symposium on Information Theory, 1973; p 267–81.

- Andriole GL, Crawford ED, Grubb RL 3rd, Buys SS, Chia D, Church TR, et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. J Natl Cancer Inst 2012;104(2):125-32

- Baade PD, Coory MD, Aitken JF. International trends in prostate cancer mortality: the decrease is continuing and spreading. Cancer Causes Control 2004;15:237–41.

- Birri S, Bidoli E, Zucchetto A, Dal Maso L, Zanier L, Serraino D. Cancer in Friuli Venezia Giulia Incidence, survival, and prevalence data: updates as of 2007. Regione Friuli Venezia Giulia – Centro di Riferimento Oncologico di Aviano, Udine, September 2011. Available at http://www.cro.it/PDF/Tumori%201995-2007%20-rev.pdf [Accessed 6.7.2014].

- Bray F, Ren JS, Masuyer E, Ferlay J. Estimates of global cancer prevalence for 27 sites in the adult population in 2008. Int J Cancer 2013;132(5):1133-45.

- Capocaccia R, Gavin A, Hakulinen T, Lutz JM, Sant M, eds. Survival of Cancer patients in Europe, 1995-2002: the EUROCARE-4 study. Eur J Cancer 2009;45:3119-346.

- Carter HB, Albertsen PC, Barry, MJ et al. Early detection of prostate cancer: AUA guideline. J Urol 2013; 190:419–26

- Carstensen B, 2007. Age-period-cohort model for the Lexis diagram. Statist Med 2007;26:3018-45.

- Catalona WJ, Richie JP, Ahmann FR, et al. Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men. J Urol 1994;151:1283–90.

- Center MM, Jemal A, Lortet-Tieulent J, Ward E, Ferlay J, Brawley O, Bray F. International Variation in Prostate Cancer Incidence and Mortality Rates. European urology 2012;61:1079–92.

- Collin SM, Martin RM, Metcalfe C, et al. Prostate-cancer mortality in the USA and UK in 1975–2004: an ecological study. Lancet Oncol 2008;9:445–52.

- Crocetti E and AIRTUM working group. Tumore della prostate: trend di incidenza e di mortalità. Epidemiologia e prevenzione 2007;31:100.

- Curado MP, Edwards B, Shin HR, Storm H, Ferlay J, Heanue M and Boyle P, eds. Cancer Incidence in Five Continents, Vol. IX IARC Scientific Publications No. 160, Lyon, IARC, 2007; Available at http://ci5.iarc.fr/CI5plus/ci5plus.htm [Accessed 6.7.2014].

- De Angelis G, De Angelis R, Frova L, Verdecchia A. MIAMOD: a computer package to estimate chronic disease morbidity using mortality and survival data. Comput Programs Biomed 1994;44:99-107.

- Draisma G, Boer R, Otto SJ, van der Cruijsen IW, Damhuis RA, Schröder FH, et al. Lead times and overdetection due to prostate specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst 2003;95:868e78.

- Draisma G, Etzioni R, Tsodikov A, Mariotto A, Wever E, Gulati R, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. J Natl Cancer Inst 2009;101:374e83.

- Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. In *End results and mortality trend in cancer*. National Cancer Institute Monograph No. 6 US Government Printing Office: Washington DC, 1961; 101-21.

- Etzioni R, Penson DF, Legler JM, Di Tommaso D, Boer R, Gann PH, Feuer EJ. Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. Journal of the National Cancer Institute 2002; 94:981–90.

- Etzioni R, Tsodikov A, Mariotto A, et al. Quantifying the role of PSA screening in the US prostate cancer mortality decline. Cancer Causes Control 2008;19:175–81.

- Etzioni R, Gulati R, Mallinger L, Mandelbatt J. Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. Ann Intern Med 2013;158:831-838.

- Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer; 2013. Available from: http://globocan.iarc.fr, accessed on day/month/year [Accessed 22.9.2014].

- Gavin A, McCarron P, Middleton RJ, et al. Evidence of prostate cancer screening in a UK region. BJU Int 2004;93:730–4.

- Gulati R, Tsodikov A, Etzioni R, Hunter-Merrill RA, Gore JL, Mariotto AB, Cooperberg MR. Expected population impacts of discontinued prostate-specific antigen screening. Cancer, 2014; 120(22):3519-26.

- Gulati R, Mariotto AB, Chen S, Gore JL, Etzioni R. Long-term projections of the harm-benefit trade-off in prostate cancer screening are more favorable than previous short-term estimates. J Clinical Epidemiol 2011;64: 1412-1417.

- Hastie TJ, Tibshirani RJ. Generalized additive models. London: Chapman & Hall, 1994.

- Heidenreich A1, Abrahamsson PA, Artibani W, Catto J, Montorsi F, Van Poppel H, Wirth M, Mottet N. Early detection of prostate cancer: European Association of Urology recommendation. Eur Urol 2013;64(3):347-54

- Holford TR. The estimation of age, period and cohort effects for vital rates. Biometrics 1983;39:311-24.

- Ilic D, Neuberger MM, Djulbegovic M, Dahm P. Screening for prostate cancer. Cochrane Database Syst Rev 2013; p. CD004720

- ISTAT. Popolazione residente. Available at http://demo.istat.it/ [accessed 6.10.2014].

- ISTAT. Statistiche sulle cause di morte anno 2008. Available at http://www.istat.it/dati/dataset/20110412_00/ [accessed 6.10.2014].

- Kim Hj, Fay MP, Feuer EJ, et al. Permutation tests for joinpoint regression with applications to cancer rates. Stat Med 2000;19:335-51.

- Leitzmann MF, Rohrmann S. Risk factors for the onset of prostatic cancer: age, location, and behavioral correlates. Clin Epidemiol 2012;4:1–11

- Lerman PM. Fitting Segmented Regression Models by Grid Search. Applied Statistics 1980;29:77-84.

- Mistry M, Parkin DM, Ahmad AS, Sasieni P. Cancer incidence in the United Kingdom: projections to the year 2030. Br J Cancer 2011;105:1795-1803

- Pinsky PF, Andriole GL, Kramer BS, Hayes RB, Prorok PC, Gohagan JK. Prostate biopsy following a positive screen in the Prostate, Lung, Colorectal and Ovarian cancer screening trial. J Urol 2005;173:746e50. discussion 50e51.

- Platz EA, Giovannucci E. Prostate cancer. In: Schottenfeld D, Fraumeni JF, editors. Cancer epidemiology and prevention. NewYork, NY: Oxford University Press; 2006. p. 1128–50.

- Quon H, Loblaw A, Nam R. Dramatic increase in prostate cancer cases by 2021. BJU International 2011;08:1734-8.

- Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Zappa M, Nelen V, et al. Screening and prostate cancer mortality in a randomized European study. N Engl J Med 2009;360:1320–8.

- Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Zappa M, Nelen V, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet 2014; 384(9959):2027-35.

- Tsodikov A, Szabo A, Wegelin J. A population model of prostate cancer incidence. Stat Med 2006;25(16):2846-66.

- Verdecchia A, Capocaccia R, Egidi V, Golini A. A method for the estimation of chronic disease morbidity and trends from mortality data. Stat Med 1989;8:01-206.

- Verdecchia A, De Angelis R, Capocaccia R, Sant M, Micheli A, Gatta G, Berrino F. The cure of colon cancer: results from the Eurocare study. Int J Cancer 1998;77:322-9.

- Verdecchia A, De Angelis G, and Capocaccia R. Estimation and Projections of Cancer Prevalence From Cancer Registry Data. Statistics in Medicine 2002;21:3511-26.

- Zucchetto A, Serraino D, Dal Maso L, Birri S, Francisci S, Zigon D, and De Angelis R. Cancer estimates up to 2015 in Friuli Venezia Giulia. Tumori 2013;99:318-26

APPENDIX A – SAS PROC NLIN

Example for modeling survival with mixture cure models

```
/*NON-LINEAR REGRESSION - MIXTURE SURVIVAL MODELS*/

/* PARAMETERs IN THE SAS MODEL:
       A = FATAL CASES (O<=A<=1)
      (1-A)= PROPORTION OF CURED PATIENTS
       GAMA = WEIBULL PARAMETER
       LAMBDA = WEIBULL PARAMETER
       B1= AGE EFFECT
       B2= PERIOD EFFECT
       B3= AREA EFFECT


*/

/* EXAMPLE: MODEL MARGINAL BASELINE, MARGINAL AGE AND PERIOD EFFECTS*/

TITLE 'MARGINAL BASELINE- MARGINAL AGE AND PERIOD EFFECTS';

PROC SORT DATA=prostate;
BY AREA;
RUN;

PROC NLIN method=gauss data=prostate outest=res noITprint  maxiter=500;
     by area;
     parms a=0.3, LAMBDA=1.0, GAMA=1.0, B1=0.0, B2=0.0;
     bounds LAMBDA>0.001, GAMA>0, 0<=a<=1;
     temp1=(LAMBDA*fup)**GAMA;
     temp=exp(-temp1);
     model surv=((1-a)+a*temp)**EXP(B1*(AGE2-AGEMED)+B2*(PER2-PERMED));
     _weight_=1./(survse**2);
     output ouT=NLRES p=pred parms=a LAMBDA GAMA B1 B2;
RUN;
```

APPENDIX B – R code for age-period-cohort analysis

Example of apc.fit for modeling using factor or natural splines

```
library(Epi)
library(splines)

#DEFINITION OF MEAN VALUES FOR PERIOD, AGE, AND COHORT EFFECTS
psa$P[psa$period==1]=2000.5
psa$P[psa$period==2]=2005.5
psa$P[psa$period==3]=2010.5


psa$A[psa$age==9]=42.5
psa$A[psa$age==10]=47.5
psa$A[psa$age==11]=52.5
psa$A[psa$age==12]=57.5
psa$A[psa$age==13]=62.5
psa$A[psa$age==14]=67.5
psa$A[psa$age==15]=72.5
psa$A[psa$age==16]=77.5
psa$A[psa$age==17]=82.5


psa$C[psa$coorte==2]=1920
psa$C[psa$coorte==3]=1925
psa$C[psa$coorte==4]=1930
psa$C[psa$coorte==5]=1935
psa$C[psa$coorte==6]=1940
psa$C[psa$coorte==7]=1945
psa$C[psa$coorte==8]=1950
psa$C[psa$coorte==9]=1955
psa$C[psa$coorte==10]=1960
psa$C[psa$coorte==11]=1965
psa$C[psa$coorte==12]=1970



#FACTOR MODEL
psa.apc<-apc.fit(psa, model="factor", parm="APC", scale=10^5, ref.p=2000.5)
psa.apc

#PLOT
frame=apc.frame(
a.lab=seq(40,90,10),
cp.lab=seq(1920,2020,10),
r.lab=c(2000,5000,10000,20000,40000),
rr.ref=2000,
a.txt = "Age at PSA testing",
cp.txt = "Calendar time",
r.txt = "Rate x100,000 men-year",
rr.txt= "Rate ratio")
```

```
apc.lines(psa.apc, frame.par=frame, lwd = 2, lty = 1, ci = rep( F, 3 )
      )

#NATURAL SPLINE
#AKAIKE INFORMATION CRITERIA (AIC)

AIC=data.frame()

k=1
n=length(psa[,1])
for (i in 1:8) {
      for (j in 1:10) {
            for (l in 1:2) {
psa.APC.ns=apc.fit(psa,model="ns",npar=c(A=i+1,P=l,C=j+1),parm="APC",scale=10^5,
ref.p=2000.5)
      AIC[k,1]=i+1
      AIC[k,2]=l
      AIC[k,3]=j+1
      AIC[k,4]=2*(n-psa.APC.ns$Anova[4,1])+n*(log( psa.APC.ns$Anova[4,2]/n))
            k=k+1
      }
}
}
names(AIC)=c('knA','knP','knC','AIC')
AIC[AIC$AIC==min(AIC$AIC),]


#BEST FIT
psa.apc.ns<-apc.fit(psa, model="ns", npar=c(A=9,P=2,C=5),parm="APC", scale=10^5,
ref.p=2000.5)

psa.apc.ns
```

Acknowledgments