

Multiple primary tumours: incidence estimation and age standardization in the presence of competing risks

Stefano Rosso^{*1} , Lea Terracini² , Fulvio Ricceri¹ , Roberto Zanetti¹

¹Piedmont Cancer Registry - CPO, Turin, Italy

²Department of Mathematics, University of Turin, Italy

Email: Stefano Rosso^{*}- stefano.rosso@cpo.it; Lea Terracini - lea.terracini@unito.it; Fulvio Ricceri - fulvio.ricceri@cpo.it; Roberto Zanetti - roberto.zanetti@cpo.it;

^{*}Corresponding author

Abstract

Background: Estimating the risk of developing subsequent primary tumours in a population is difficult since the probability of occurrence is conditioned to the survival probability.

Methods: We proposed to apply Markov models studying the transition intensities from first to second tumour with the Aalen-Johansen (AJ) estimators, as usually done in competing risk models. In a simulation study we applied the proposed method in different settings with constant or varying underlying intensities and applying age standardization. In addition, we illustrated the method with breast cancer data from the Piedmont Cancer Registry.

Results: Results were presented in term of a comparison between classical Standardized Incidence Ratio and standardized indicators computed by using probabilities obtained from the proposed method. The simulation study showed that the classical approach led to an underestimated SIR of -19.42% , while with the AJ estimators the bias was limited to -4.75% . The general underestimation of risk shown by the classical SIR was also present in the case study on the breast cancer from the Piedmont Cancer Registry. After correction with the AJ estimators, the fully-corrected SIR (fcSIR) showed a significant risk increase for subsequent cancer of corpus uteri (fcSIR: 1.91 95% CL 1.47, 2.44).

Conclusions: The increased risk of a cancer of the corpus, also observed in other studies, is usually interpreted as the common shared risk factors such as low parity, early menarche and late onset of menopause. We also

grouped together those cancers possibly associated to a previous local radiotherapy: the cumulative risk at 14 years is still not significant, however the AJ estimators showed a significant risk peak between the eighth and the ninth year. Finally, the proposed approach has been shown to be superior to the classical approach under several aspects and more informative. It allowed for a better estimate of the number of events, conditioning to observed survival.

Introduction

During the last decades, improvements in medical and surgical treatments had substantially increased the chances of surviving from a cancer. Cancer survivors now amounts to more than 3.5% of population in the US [1], and about 3% in Western Europe [2]. Now more cancer survivors face the problem of subsequent cancers related to the late effects of treatments or to the underlying risk factors. As for other epidemics in the past, the challenge is towards a research effort to address, and possibly prevent, those elements that increase the chance of falling for a second tumour. And, as in the past the starting point, is to correctly estimate the incidence of multiple primary tumours on a population basis.

First of all, there is a problem of differential diagnosis, when it comes to distinguish between local and distant metastases, recurrences and the onset of a truly new lesion. Classifications may also vary leading to substantial differences in rates. For example SEER rules [3] differ substantially from those adopted by IARC [4]. Timing of multiple primaries is also important, as they can occur at the same time (synchronous) or after a time lag (metachronous). Usually synchronous tumours are excluded from analyses, in the belief that they rather represent prevalent silent tumours come to evidence during diagnostic procedures.

Secondly, it should be taken into account that incidence rate of multiple primaries is conditional to the probability of surviving the first tumours, having accumulated sufficient time for developing another one. Usually, the studied statistics is the ratio between observed and expected multiple metachronous primary tumours. Expectation is taken computing the person-years observed in the cohort of patients with first tumour, applying general population incidence rates. In this way, a group of 100 patients with a short survival of 1 year is equivalent to 10 patient surviving for 10 years. But we know that rarely these two groups with such a different survival experience can be compared for several aspects, even without difference in age distribution. On the contrary when conditioning on survival probabilities we get the same

number of expected cases only when the overall survival is equal to that of the general population. This assumption holds true only for those pairs where the first tumour has a rather benign course with no substantial influence on the whole survival.

Some of these aspects are not new in the literature, and they were discussed for deriving expected number of deaths (or events) for SMR. Keiding offered a historical perspective of it [5], also showing how this, one of the oldest statistical techniques, was intimately connected to conditional survival probabilities and parametric models. The estimation of the expected number of subsequent cancers adds some complication to the traditional model and should be approached in the framework of competing risks, as many subjects are withdrawn from the population at risk, as time goes by, by death or censorship. Previous works had already shown how the traditional Kaplan-Meier estimator is inappropriate in the presence of competing risks [6], but until now a correct approach taking into consideration competing risks has not yet been applied to the estimation of multiple tumours expected number.

In addition, it is also important to consider the time elapsing dimension, as subsequent tumours are often more frequent in the first years, and then decrease, with a later rise after five to eight years depending on the tumour type [7,8]. Therefore, expecting a constant rate over the whole follow-up period is over-simplistic.

The need to compute the conditional survival probabilities and the time varying rates led us to consider a non-parametric approach based on multi-state models, which can appropriately describe situations where there are several competing outcomes in a time process. Various types of multi-state models have been proposed for analyzing multiple end-points in different situations; from transplants to clinical trials and from pregnancy-birth model to infectious disease epidemics (for a review see: [9]). Indeed, we tried to give the perspective from the underlying stochastic process to the estimation of risk of developing a subsequent cancer, following the enlightening suggestions offered by Aalen and Gjessing in their work [10].

Methods

Subjects

The Piedmont Cancer Registry (RTP) collects all incident tumours in the resident population (about one million inhabitants) of Turin (Italy) since 1985. We selected all first occurrences of breast cancer (following IARC rules, cases occurring in the paired breast gland were excluded). We included all cases diagnosed up to 1998 and then we prolonged the observation period for detecting subsequent tumours up to the end of year 2000, allowing for a reasonable amount of follow-up time also for the last incident cases. We excluded

all cases diagnosed with Death Certificate Only (DCO), skin cancers other than melanoma and all synchronous tumours.

We computed the classical Standardized Incidence Ratio (**SIR**), cumulating person-years at risk for each woman beginning with the first tumour's date of diagnosis and ending with the second tumour's date of diagnosis, death or end of follow-up (December 31, 2000) whichever came first, then applying incidence rates from four different observation periods (1985 – 1987; 1988 – 1992; 1993 – 1997, 1998 – 2000) [Cancer Incidence in Five Continents, volumes VI, VII, VIII [11–13] and <http://www.cpo.it/dationcologici/> for the still unpublished data]. The 95 percent confidence intervals for **SIR** were computed assuming a Poisson distribution with no measurement errors in the expected numbers [14].

Statistical methods

The tumours occurrence in a general population can be depicted as in Figure 1. From demographic sources we know the amount of deaths (m_e) and the general mortality rate (μ_e). From cancer registries we measure the number of first primary tumours (n_1) the incidence rate (λ_1), the number and rates of these patients deceased for other causes ($m_{1e}; \mu_e$) and for the specific cause of death ($m_{1\mu}; \mu_1$). After this first process has taken place we can observe (n_2) second primaries with a rate (λ_2) to be estimated conditioned to the quantities and parameters previously seen.

Since we were interested in estimating rate of occurrence after first primary only, we dealt with the simplified model given in Figure 2. This last one satisfied Markov assumption, since it did not take into consideration the past transitions from health state to first tumour. We applied Markov theory to the process occurring to the first tumour cohort with two different irreversible and reciprocally exclusive outcomes: death and second tumour occurrence. Of course, for computing transition intensities in this model it was necessary to also consider censored observations. We estimated transition intensities by Nelson-Aalen estimators; then we computed occurrence probabilities conditioned to different events (occurrence of a second cancer, death) in each time interval with the Aalen-Johansen [15] method (**AJ**) in the framework of a Markov process (for details see Appendix).

The proposed model is a simple version of the competing risk model on which a vast literature already exists (see, for example, Satten and Datta for marginal estimation of multi-state models with right-censored data [16]). In our case, simplification occurs since there are not reversible states. As shown in the following, the novelty of our approach was to use the **AJ** estimator for correcting the classical Standardized Incidence Ratio (**SIR**) in the case of multiple tumours.

To directly compare classical **SIR** results and results from **AJ** estimators we computed the expected cumulative number of transitions (to second tumour) at the end of the fourteen years period, as:

$$trans_{AJ} = \hat{P}_{12}(0, 14) \cdot N.$$

We compared it with the number of expected transitions under the assumption of a Markov model with constant transition intensities equal to those expected from the general population.

$$E(trans_{AJ}) = \hat{P}_g(0, 14) \cdot N.$$

So we can derive three different types of standardized incidence ratios:

$$\begin{aligned} \text{Standardized Incidence Ratio (SIR)} &= \frac{O}{E}; \\ \text{corrected SIR (cSIR)} &= \frac{trans_{AJ}}{E}; \\ \text{fully corrected SIR (fcSIR)} &= \frac{trans_{AJ}}{E(trans_{AJ})}. \end{aligned}$$

It must be noted that imposing the same censorship mechanism in computing expected cases resulted in a less biased estimator, since the same bias originated by the censorship mechanism was at work both in the numerator and in the **fcSIR** denominator. We computed 95 percent confidence limits for classical **SIR** following [14], and using the **AJ** variance in formula 7 for **cSIR** and **fcSIR**.

Age standardization

Since rates of first and second tumours strongly depend on age, analysis must be done in age strata or a standardization procedure must be defined. We pursued both strategies grouping ages in five classes: 0 – 44, 45 – 54, 55 – 64, 65 – 74, 75+. A standardized **AJ** estimator for the whole population can be obtained as follows:

- For each age class k we compute the **AJ** estimator $\hat{P}_{ijk}(s, t)$: let N_k be the number of subjects in class k at time 0 and set a weight $W_k = \frac{N_k}{N}$, where N equals the sum of the N_k 's;
- define $\hat{P}_{12}^{\text{stand}}(s, t) = \sum_k W_k \cdot \hat{P}_{12k}(s, t)$;
- under the assumption that weights are deterministic variables,

$$var(\hat{P}_{12}^{\text{stand}}(s, t)) = \sum_k W_k^2 \cdot var(\hat{P}_{12k}(s, t)).$$

Simulation study

We also wanted to compare, from a methodological point of view, the estimated number of second tumour using the **AJ** estimator with that obtained with the classical Person-Years approach. This comparison can be better performed simulating the process of second tumour occurrence taking under control biasing factors such as censoring. We then simulated different dynamics of second tumour occurrence. We considered a simulated cohort of 10,000 patients with a first primary and with same age and period of incidence, followed up for 10 years. We imposed a survival exponential law with a constant mortality rate of 0.2. Firstly, occurrence of a second primary was kept constant for the whole follow-up period. We compared the simulated number of second tumours to the number estimated both by the Person-Year, and by the **AJ** approach. We let the second tumour incidence rate vary from 0.00025 to 0.004, corresponding to a rate ratio of 0.25, 0.5, 1, 2 and 4. Secondly, effect of standardization by age was investigated repeating the simulation for the five age classes, varying the occurrence rates, but always keeping them constant for the whole period. Thirdly, the simulation was extended to situations where also occurrence rates varied in time: at a constant decreasing or increasing trend, or in a bimodal way. Age standardization was then applied on bimodal rates simulation.

The simulation engine was based on random chains of multinomial probabilities $M(P_\alpha, P_\beta, P_\gamma)$ at each time-click t , with the constraint that $P_\alpha + P_\beta + P_\gamma = 1$. P_α and P_β are normally distributed hyper-parameters in a three class model, respectively representing the probability of transition from steady state to second tumour and transition from steady state to death.

Results

Risk of a second tumour following breast cancer

We identified 9,233 women with breast cancer in Turin from 1985 to 1998, 249 cases were excluded as they were identified only from death certificate (DCO), and 58 cases were excluded since they were synchronous cancers, leaving 8,926 cases for analysis. From this cohort, 353 second (metachronous) primary tumours (excluding skin cancers) developed during the prolonged follow-up period (1985 – 2000). The median interval time between the two tumours was four years, but with an highly positive skewed distribution. The completeness of clinical documentation was rather high, considering that registries work on a population basis, with a 94.5 percent of microscopic confirmation for first tumours that reached 99.3 percent for second tumours. In Table I we presented the observed number of second tumours by site and then the different values of **SIRs**, assuming the traditional approach with an homogeneous conditional

survival **SIR**; assuming observed probabilities conditioned to survival (**cSIR**), and finally, assuming both observed and expected probabilities fully conditioned to survival (**fcSIR**).

The classical **SIR** showed a risk decrease for overall tumours and for a number of specific cancer sites (stomach, colon-rectum, gallbladder, lung, bladder, cervix and other and unspecified sites). A significant risk increase was observed only for corpus uteri cancer. The evidence of a sort of protection against the occurrence of a second cancer is attenuated using **cSIR** and **fcSIR**. The only cancer site where the attenuated risk persisted statistically significant was other and unspecified. On the other hand, the significant association with second primaries occurring to corpus uteri is enhanced. In addition, **fcSIR** showed a suggestive increase of risk also for cancers of oesophagus, stomach and pancreas, and for non-Hodgkin lymphoma, although confidence limits still included unity.

Another interesting feature of the **AJ** estimator is the possibility of studying the dynamic of second primary occurrence over time. In Figures 3, 4 and 5 we presented the time trend of the **AJ** operator, $\hat{P}_{12}(0, t)$, which is the estimated cumulative probability of a second primary until time t for all cancers (Figure 3), for cancer of corpus (Figure 4), and for cancers related to breast cancer radiotherapy (Figure 5). In particular, Figure 3 showed that there was a relative increase of probability of developing a second primary from the fourth to the tenth year after diagnosis of a breast cancer, even if the cumulative **fcSIR** for all cancers was not significant. For corpus uteri the Figure 4 showed a persistent increase of risk across all the observation period; Figure 5 showed an increased risk after 5 years of follow-up.

Simulation study

With a sample size of 10,000 subjects replicated for 1,000 times we had an average of 55.1 cases with a second tumour in ten years given a constant annual death rate of 0.2 and a constant annual second tumour rate of 0.001. These were estimated as 53.00 cases by the **AJ** estimator and as 44.4 cases using the person-year approach (Table II). Varying the base rate from 0.00025 to 0.004 the percentage of bias in the **AJ** estimators stayed between 3.81% to 5.80%, while the bias in the number of estimated cases using the classical amount of person-years ranged between 19.06% to 19.73%. Age-standardization averaged the observed bias with 4.75% for the **AJ** estimator and 19.42% for the classical approach. In Table III we presented the effect of varying the base incidence rate by follow-up time. We observed a wider bias, that, however, when averaged by age-standardization, was kept reasonably low (4.67%) for the **AJ** estimator and 4.93% for the person-year approach. The bias was the lowest (+1.25%) in the case of constantly decreasing rates. Indeed, the cumulative effects of competing mortality are larger as time goes by and the bias reached

a 14.7% for the **AJ** estimators, and a 26.3% computing the estimated cases from the average person-years, in the case of a constantly increasing rate. However, the most frequent case when dealing with subsequent primary malignancies is the situation where the rate is constantly decreasing or bimodal. The occurrence of a constantly increasing rate is quite uncommon.

Discussion

The occurrence of subsequent primary tumours can be due to several factors. Subsequent malignancies can initially result from intense clinical surveillance after the first tumour; they can occur later on as therapies for the first primary can induce carcinogenesis. Finally, they can also be due to shared risk factors, including environment, life styles and inherited genes predisposing to higher susceptibility.

However, the high fatality of several cancers or the competing risk for other causes hinder the possibility of observing subsequent events, even if their probability is sensibly increased. Following the suggestions of Hougaard [9], we applied a simple Markov model for competing risks and we studied the transition probabilities from first to second tumour varying in time. The observed time trend of second primary occurrence is often not constant with two or more waves of increased risks during the observed period. For this reason we resorted to a non-parametric approach, directly computing **AJ** estimators. Some other possible parametric approaches could be based on a piecewise constant hazard function [17], or, on stratification by age or other covariates when proportional hazard model cannot be used [18].

Simulation showed how **AJ** estimators led to less biased estimates than classical person-year method. This is essentially due to the fact that the **AJ** estimators are built up taking into consideration in numerator and in denominator the exact amount of transitions and person-time at risk at each time interval. On the contrary, the person-year method computes denominators only at the end of the observation period. Moreover, not only **AJ** estimators can give a more precise result at the end of the period, but they also describe the full probability trend over time. When keeping the incidence rate constant over the period, the person-year approach led to a larger bias, underestimating the number of events. Also the **AJ** estimator had some limitations, however, the bias was within the 5% error probability as shown in Table II by the simulation.

The situation is even more complicate with varying rates. The largest bias was seen with constantly increasing rates. In this case, the person-year approach gave rise to a larger than simulated number of events, while the **AJ** estimator underestimated the overall number of events, although to a lesser extent. This limit is due to the unavoidable introduction of discrete time intervals in the analysis of an intrinsically

continuous dimension.

The discrepancy between the two methods is even more evident when analyzing a real situation. Indeed, results from the Piedmont Cancer Registry, with a follow-up of 14 years, showed that the person-year-based **SIR** computed on large time intervals, measured at the end of follow-up and assuming a constant rate led to a substantial risk underestimation with the exception of thyroid cancer. We learned that the results for thyroid cancer can be due to a rate with an increasing trend, as it emerged from an analysis of $AJ_{12}(s, t)$ estimators for this cancer, showing a substantial and significant risk increase after six-seven years of follow-up (results not shown).

The increased risk of a cancer of the corpus was also observed in other studies [19–23], although with lower values more similar to the person-year-based **SIR**, and it is usually interpreted as the common shared risk factors such as low parity, early menarche and late onset of menopause. The study by Evans et al. [21], based on the larger population of the Thames Cancer Registry, with a follow-up of 34 years, observed several associations of breast cancer with other tumours. In particular the increased risk of esophagus, stomach, lung and thyroid cancer was suggestive of a late effect of local radiotherapy of the breast tumour. Indeed, we observed an increased risk, although not significant, for esophagus and stomach cancer when fully correcting **SIRs**. Grouping together those cancers possibly associated to a previous local radiotherapy, the cumulative risk at 14 years is still not significant, however the $AJ_{12}(s, t)$ estimators showed a significant risk peak between the eighth and the ninth year (Figure 5).

In conclusion the proposed approach has been shown to be superior to the classical approach under several aspects and more informative. It allowed for a better estimate of the number of events, conditioning to observed survival. Finally, we must note that, given a known, although complex, hazard function for the incidence of second primaries, it is challenging to study specific survival. This could be done both using the method proposed by Heinävaara et al. [24], or directly using the **AJ** estimators.

Appendix

Markov models and Aalen-Johansen estimators

Markov models deal with situations where individuals can belong to a finite set of states and move to one state to some others with a probability, possibly depending on time. The main hypothesis (the Markov assumption) is that the probability of moving from state i to state j at time t depends only on i, j and t and not on the previous states.

For every possible move $i \rightarrow j$ it is defined a *transition intensity map* from state i to state j , $\alpha_{ij}(t)$, and a

cumulative intensity map $A_{ij}(t) = \int_0^t \alpha_{ij}(s) ds$ Then we define the *probability transition maps* and the *probability transition matrix* in the interval $[s, t]$

$$\begin{aligned} P_{ij}(s, t) &= \text{probability for an individual being in state } i \text{ at time } s \\ &\quad \text{to be in state } j \text{ at time } t \\ \mathbf{P}(s, t) &= \text{matrix of the } P_{ij}(s, t)\text{'s} \end{aligned}$$

The key mathematical ingredient for estimating probability transitions maps is the following formula which is a consequence of Chapman-Kolmogorov equation for Markov models:

$$\mathbf{P}(s, t) = \lim \prod_k (\mathbf{I} + \mathbf{A}(u_{k+1}) - \mathbf{A}(u_k)) \quad (1)$$

where $u_1 < \dots < u_r$ is a partition of $[s, t]$ and the limit is taken as the partition is refined; $\mathbf{A} = \{a_{ij}\}$ is the upper triangular matrix defined by

$$a_{ij} = \begin{cases} A_{ij} & \text{if there is a move } i \rightarrow j \\ 0 & \text{if } i \neq j \text{ and there is not a move } i \rightarrow j \\ -\sum_{h>i} A_{ih} & \text{if } i = j \text{ and } i \text{ is not an absorbing state} \\ 0 & \text{if } i = j \text{ and } i \text{ is an absorbing state} \end{cases} \quad (2)$$

In the presence of right censoring, the cumulative intensity maps can be estimated by using the Nelson-Aalen estimators (\mathbf{NA}), as follows. For every time t , let $N_{ij}(t)$ be the number of transitions from state i to state j in the time interval $[0, t]$, and $Y_i(t)$ be the number of individuals which are in state i at time t . Then the *Nelson-Aalen estimator*, giving an estimation for the cumulative intensity, is

$$\hat{A}_{ij}(t) = \int_0^t \frac{dN_{ij}(u)}{Y_i(u)}. \quad (3)$$

The estimation of the probability transition matrix can be obtained by using the values of Nelson-Aalen in formula (1). We get in this way the Aalen-Johansen (\mathbf{AJ}) estimator for the probability transition matrix:

$$\hat{\mathbf{P}}(s, t) = \lim \prod_k (\mathbf{I} + \hat{\mathbf{A}}(u_{k+1}) - \hat{\mathbf{A}}(u_k)). \quad (4)$$

For a complete reference see the book on statistical models based on counting processes by Andersen et al [15].

Application to multiple primary tumours

We assumed that the starting time, 0, is the time of diagnosis of the first tumour for each individual.

We construct a simple model with three states

- 1 first tumour
- 2 second tumour
- 3 death after a first (but not a second) tumour

where 2 and 3 are absorbing states and the possible moves are

$$1 \rightarrow 2, \quad 1 \rightarrow 3.$$

In order to fit our situation in a Markov model, we need to make sure that every individual goes through at most one move in every time unity. This is the mathematical reason for eliminating all synchronous situations.

According to definitions 3 and 4 we set for every move $i \rightarrow j$:

$$\hat{A}_{ij}(t) = \sum_{u=1}^t \frac{N_{ij}(u) - N_{ij}(u-1)}{Y_i(u-1)} \quad (5)$$

The estimation of the probability transition matrix can be obtained by approximating the integral product in formula 4 via time discretisation. We get in this way the **AJ** estimator for the probability transition matrix:

$$\hat{\mathbf{P}}(s, t) = \prod_{s < u \leq t} (\mathbf{I} + \hat{\mathbf{A}}(u) - \hat{\mathbf{A}}(u-1)). \quad (6)$$

where $\hat{\mathbf{A}}$ is the matrix obtained by formula 2 with the estimated values.

The **AJ** estimators are consistent and valid also with right censoring and when the underlying process is non-Markovian [25].

Following [15] we estimated the variance:

$$\begin{aligned} \widehat{\text{var}}(\hat{P}_{12}(s, t)) &= \\ &= \sum_{u=s}^t \left[\left(\hat{P}_{11}(s, u) - \hat{P}_{12}(s, t) + \hat{P}_{12}(s, u) \right)^2 Y_1(u)^{-2} (N_{12}(u) - N_{12}(u-1)) + \right. \\ &\quad \left. + \left(\hat{P}_{12}(s, t) - \hat{P}_{12}(s, u) \right)^2 Y_1(u)^{-2} (N_{13}(u) - N_{13}(u-1)) \right] \end{aligned} \quad (7)$$

Authors contributions

Stefano Rosso conceived the idea for the study. Stefano Rosso, Fulvio Ricceri and Lea Terracini planned and designed the research. Lea Terracini developed the statistical models. Fulvio Ricceri analysed the data. Stefano Rosso and Roberto Zanetti wrote the first draft of the manuscript. Roberto Zanetti coordinated the project. All authors edited and approved the final version of the manuscript.

Acknowledgements

This research was partly founded by Regione Piemonte - Ricerca Sanitaria Finalizzata (years 2004 and 2006).

We express thanks to the researchers and professors of the Me.Ri.Ma. group of the University of Turin (Department of Mathematics) who shared their ideas with us and gave us their time and comments.

We also thank Federica and Simona Gallo for their editorial assistance.

The authors declare that they have no competing economic or financial interests.

Figures

Figure 1 - The multi-state model

Figure 2 - The simplified Markov model

Figure 3 - All cancers (excluding skin cancer) as second tumour.

\hat{P}_{12} is the cumulative observed probability of a second tumour in the cohort of patients with a primary tumour, with its 95% upper (95% ul) and lower (95% ll) confidence limits. \hat{P}_g is the cumulative estimated probability of a second tumour assuming a constant intensity taken from the general population.

Figure 4 - Corpus Uteri as second tumour.

\hat{P}_{12} is the cumulative observed probability of a second tumour in the cohort of patients with a primary tumour, with its 95% upper (95% ul) and lower (95% ll) confidence limits. \hat{P}_g is the cumulative estimated probability of a second tumour assuming a constant intensity taken from the general population.

Figure 5 - Cancers related to Radiotherapy (Oesophagus, Stomach, Lung and Thyroid gland) as second tumour.

\hat{P}_{12} is the cumulative observed probability of a second tumour in the cohort of patients with a primary tumour, with its 95% upper (95% ul) and lower (95% ll) confidence limits. \hat{P}_g is the cumulative estimated probability of a second tumour assuming a constant intensity taken from the general population.

Tables

Table 1: Number of observed second tumours in a cohort of women with breast cancer in Turin (Italy), SIR, corrected SIR (cSIR) and fully corrected SIR (fcSIR)

| Cancer Site | Exp. cases | SIR (95% C.I.) | cSIR (95% C.I.) | fcSIR (95% C.I.) |
|----------------------------------|------------|-------------------|--------------------|---------------------|
| Mouth Pharynx | 7 | 0.65 (0.26-1.35) | 0.96 (0.47-1.76) | 0.80 (0.39-1.47) |
| Oesophagus | 5 | 1.16 (0.37-2.2) | 1.54 (0.60-3.24) | 2.38 (0.92-5.01) |
| Stomach | 29 | 0.65 (0.43-0.93) | 0.79 (0.55-1.10) | 1.38 (0.97-1.93) |
| Colon-Rectum | 66 | 0.57 (0.44-0.73) | 0.64 (0.50-0.81) | 0.87 (0.69-1.10) |
| Liver | 7 | 0.52 (0.21-1.07) | 0.47 (0.18-0.99) | 0.61 (0.23-1.30) |
| Gallbladder | 8 | 0.37 (0.16-0.73) | 0.50 (0.25-0.90) | 0.78 (0.38-1.40) |
| Pancreas | 20 | 0.79 (0.48-1.22) | 0.95 (0.61-1.41) | 1.39 (0.89-2.07) |
| Lung | 24 | 0.53 (0.34-0.79) | 0.64 (0.43-0.92) | 0.80 (0.54-1.15) |
| Melanoma | 14 | 1.11 (0.61-1.86) | 1.22 (0.69-2.00) | 1.15 (0.65-1.88) |
| Cervix uteri | 9 | 0.52 (0.24-0.98) | 0.63 (0.32-1.13) | 0.68 (0.34-1.22) |
| Corpus uteri | 54 | 1.62 (1.22-2.12) | 1.89 (1.45-2.42) | 1.91 (1.47-2.44) |
| Ovary | 24 | 0.87 (0.56-1.29) | 0.96 (0.63-1.41) | 1.12 (0.74-1.64) |
| Bladder | 14 | 0.56 (0.30-0.94) | 0.61 (0.35-1.01) | 0.74 (0.42-1.21) |
| Brain & CNS | 4 | 0.42 (0.11-1.09) | 1.16 (0.54-2.19) | 0.56 (0.18-1.31) |
| Thyroid | 11 | 1.39 (0.70-2.50) | 1.16 (0.53-2.19) | 1.00 (0.46-1.89) |
| NHL | 21 | 1.02 (0.64-1.57) | 1.21 (0.79-1.80) | 1.32 (0.85-1.94) |
| Leukaemias | 9 | 0.55 (0.25-1.05) | 0.61 (0.29-1.12) | 0.81 (0.39-1.49) |
| Other & unsp | 27 | 0.28 (0.17-0.41) | 0.34 (0.23-0.48) | 0.48 (0.33-0.68) |
| Total (breast and skin excluded) | 353 | 0.65 (0.58-0.72) | 0.76 (0.69-0.84) | 0.99 (0.91-1.10) |

Table 2: Simulation Study, I

| Averages over 1000 simulation runs | 0.00025 | 0.0005 | 0.001 | 0.002 | 0.004 | Age-stand. Indicators |
|---|----------|----------|----------|----------|----------|-----------------------|
| Number of simulated cases | 13.8 | 27.6 | 55.1 | 109.34 | 218.5 | 82.5 |
| \hat{P}_{12} | 0.0013 | 0.0026 | 0.0053 | 0.0104 | 0.0208 | 0.0079 |
| Person-Year at risk | 44575.68 | 44526.58 | 44442.83 | 44252.52 | 43849.93 | 44339.19 |
| Number of estimated cases(Aalen-Johansen) | 13.0 | 26.0 | 53.0 | 104.0 | 208.0 | 78.6 |
| <i>Bias %</i> | 5.80 | 5.80 | 3.81 | 4.88 | 4.81 | 4.75 |
| Number of estimated cases (Person-Year) | 11.1 | 22.3 | 44.4 | 88.5 | 175.4 | 66.5 |
| <i>Bias %</i> | 19.25 | 19.34 | 19.34 | 19.06 | 19.73 | 19.42 |

Table 3: Simulation Study, II (¹rates = 0.002; 0.0018; 0.00165; 0.0015; 0.00135; 0.0012; 0.001; 0.0008; 0.00065; 0.0005, ²rates = 0.0005; 0.00065; 0.0008; 0.001; 0.0012; 0.00135; 0.0015; 0.00165; 0.0018; 0.002, ³rates = 0.001; 0.002; 0.001; 0.0005; 0.0005; 0.0005; 0.001; 0.002; 0.001; 0.0005, ⁴rates for age groups = annual rates by increasing Relative Risks {0.25; 0.5; 1; 2; 4})

| Averages over 1000 simulation runs | Costantly decreasing ¹ | Costantly increasing ² | Bimodal ³ | Age-stand. Indicators ⁴ |
|---|-----------------------------------|-----------------------------------|----------------------|------------------------------------|
| Number of simulated cases | 83.95 | 52.81 | 59.41 | 88.56 |
| \hat{Q}_{12} | 0.0015 | 0.0015 | 0.0012 | 0.0019 |
| \hat{P}_{12} | 0.0085 | 0.0045 | 0.0057 | 0.0084 |
| Person-Year at risk | 44309.59 | 44479.55 | 44426.39 | 44309.51 |
| Number of estimated cases(Aalen-Johansen) | 85.00 | 45.00 | 57.00 | 84.42 |
| <i>Bias %</i> | 1.25 | 14.79 | 4.06 | 4.67 |
| Number of estimated cases (Person-Year) | 66.46 | 66.72 | 53.31 | 84.19 |
| <i>Bias %</i> | 20.83 | 26.34 | 10.26 | 4.93 |

References

1. P. Ganz, "A teachable moment for oncologists: cancer survivors, 10 million strong and growing.," *J. Clin. Oncology*, vol. 23, pp. 5468–60, 2005.
2. J. Ferlay, F. Bray, P. Pisani, and D. Parkin, *LOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide*. IARC CancerBase No. 5. version 2.0, Lyon: IARCPress, Lippincott-Raven, 2004.
3. C. Jonhson, *The SEER coding and staging manual 2004. Fourth Edition*. NIH Pub No 04-5581, Bethesda, MD: National Cancer Institute, 2004.
4. IARC/IACR, *International rules for multiple primary cancer (ICD-O Third edition)*. Internal Report n. 2004/02, Lyon: IARC, 2004.
5. N. Keiding, "The method of expected number of deaths 1786-1886-1986," *Int. Statist. Rew.*, vol. 55, pp. 1–20, 1987.
6. T. Gooley, W. Leinsenring, J. Crowley, and B. Storer, "Estimation of failure probabilities in the presence of competing risks: new representations of old estimators," *Statistics in Medicine*, vol. 18, pp. 695–706, 1999.

7. E. Crocetti, E. Buiatti, P. Falini, and the Italian Multiple Primary Cancer Working Group, "Multiple primary cancer incidence in Italy," *Eur. J. Cancer*, vol. 37, pp. 2449–2456, 2001.
8. K. Filali, G. Hédèlin, P. Shaffer, J. Estève, P. Arveux, C. Bouchardy, C. Exbrayat, J. Faivre, F. Levi, J. Macé-Lesech, D. Pottier, and J. Torhorst, "Multiple primary cancers and estimation of the incidence rates and trends," *Eur. J. Cancer*, vol. 32A, pp. 683–690, 1996.
9. P. Hougaard, "Multi-state models: a review," *Lifetime Data Anal.*, vol. 5, pp. 239–264, 1999.
10. O. Aalen and H. Gjessing, "Understanding the shape of the hazard rate: a process point of view," *Stat. Sci.*, vol. 16, pp. 1–14, 2001.
11. D. Parkin, C. Muir, S. Whelan, Y. Gao, J. Ferlay, and J. Powell, *Cancer incidence in five continents. Volume VI*. IARC Scientific Publications (120), IARC Press, 1992.
12. D. Parkin, S. Whelan, J. Ferlay, L. Raymond, and J. Young, *Cancer incidence in five continents. Volume VII*. IARC Scientific Publications (143), IARC Press, 1997.
13. D. Parkin, S. Whelan, J. Ferlay, L. Teppo, and D. Thomas, *Cancer incidence in five continents. Volume VIII*. IARC Scientific Publications (155), IARC Press, 2002.
14. F. Liddell, "Simple exact analysis of the standardised mortality ratio," *J. Epidemiol. Community Health*, vol. 38, pp. 85–88, 1984.
15. P. Andersen, O. Borgan, R. Gill, and N. Keiding, *Statistical models based on counting processes*. Springer Series in Statistics, Springer-Verlag, 1993.
16. G. Satten and S. Datta, "Marginal estimation for multi-stage models: waiting time distributions and competing risks analyses," *Statistics in Medicine*, vol. 21, pp. 3–19, 2002.
17. I. Gijbels and G. U., "Estimation of a change point in a hazard function based on censored data.," *Lifetime Data Anal.*, vol. 9, pp. 395–411, 1996.
18. R. Prentice, B. Williams, and A. Peterson, "On the regression analysis of multivariate failure time data," *Biometrika*, vol. 68, pp. 373–379, 1981.
19. L. Teppo, E. Pukkala, and E. Saxen, "Multiple cancers - an epidemiological exercise in Finland," *J. Natl. Cancer Inst.*, vol. 75, pp. 207–217, 1985.

20. N. Volk and V. Pompe-Kirn, "Secondary primary cancers in breast cancer patients in Slovenia," *Cancer Causes Control*, vol. 8, pp. 764–770, 1997.
21. H. Evans, C. Lewis, D. Robinson, C. Bell, H. Møller, and H. S.V., "Incidence of multiple primary cancers in a cohort of women diagnosed with breast cancer in southeast England," *Br. J. Cancer*, vol. 84, pp. 435–440, 2001.
22. I. Soerjomataram, E. Louwman, WJ abd de Vries, V. Lemmens, W. Klokman, and J. Coebergh, "Primary malignancy after primary female breast cancer in the south of the Netherlands, 1972-2001," *Breast Cancer Res Treat.*, vol. 93, pp. 91–95, 2005.
23. D. Curtis, RE andFreedman, E. Ron, L. Ries, D. Hacker, B. Edwards, M. Tucker, and J. Fraumeni, *New Malignancies Among Cancer Survivors: SEER Cancer Registries, 1973-2000*, vol. 05-5302 of *National Cancer Institute*. Bethesda, MD: NIH Publ., 2006.
24. S. Heinävaara, L. Teppo, and T. Hakulinen, "Cancer-specific survival of patients with multiple cancers: an application to patients with multiple breast cancers," *Statistics in Medicine*, vol. 21, pp. 3183–3195, 2002.
25. S. Datta and G. Satten, "Validity of the aalen-johansen estimators of stage occupation probabilities and nelson-aalen estimators of integrated transition hazards for non-markov models," *Statist. Probab. Lett.*, vol. 55, pp. 403–411, 2001.

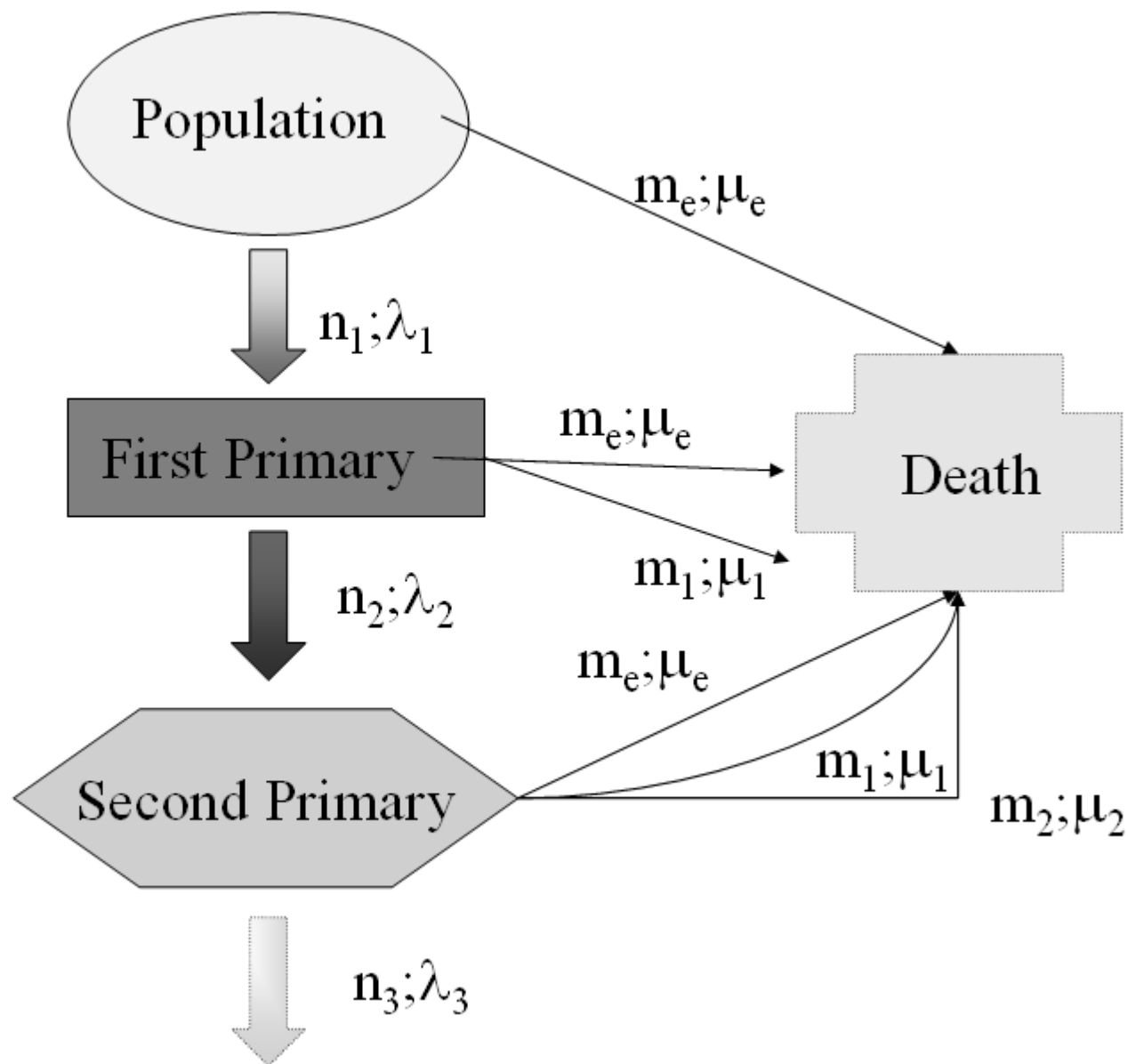


Figure 1

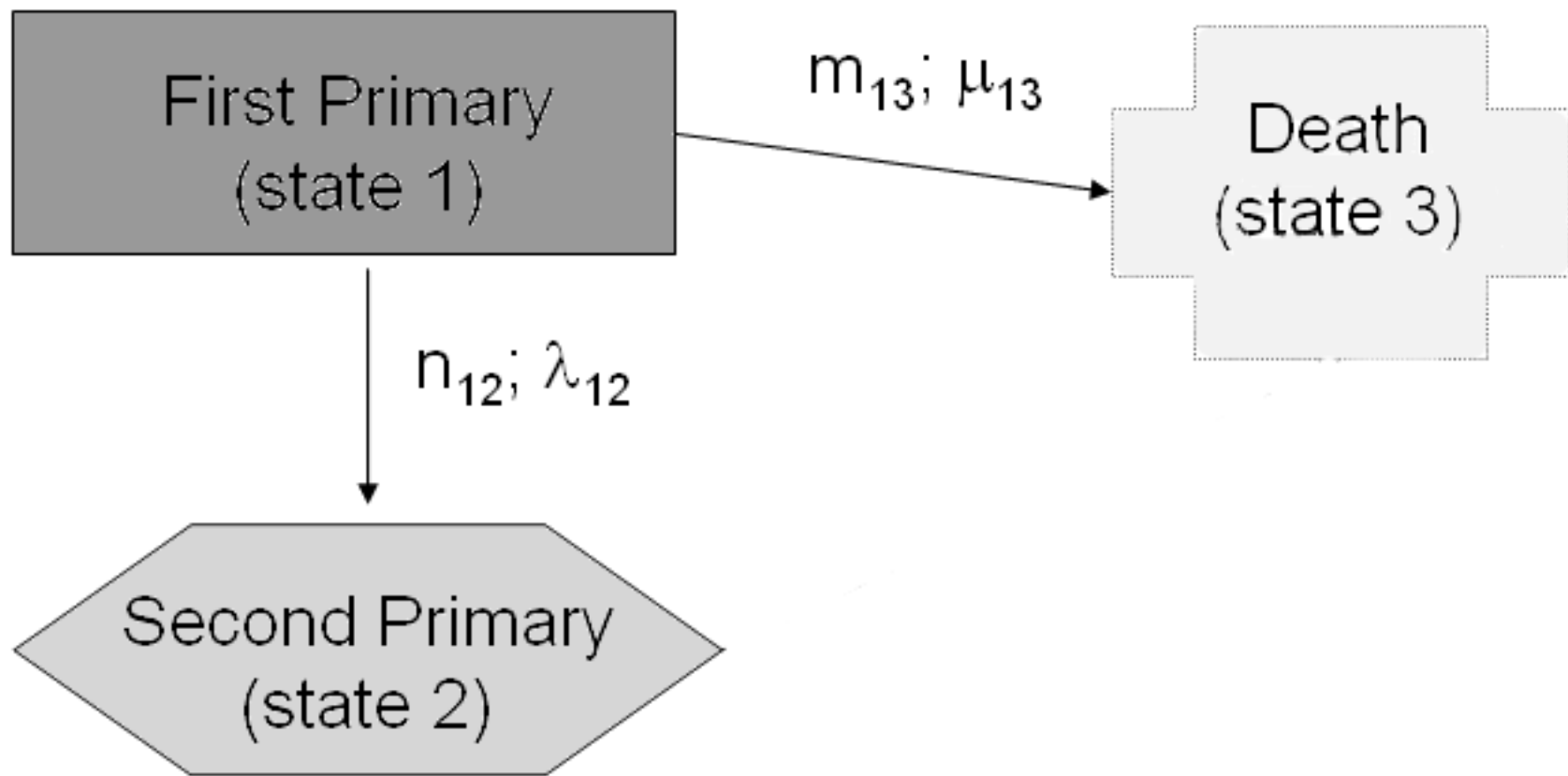


Figure 2

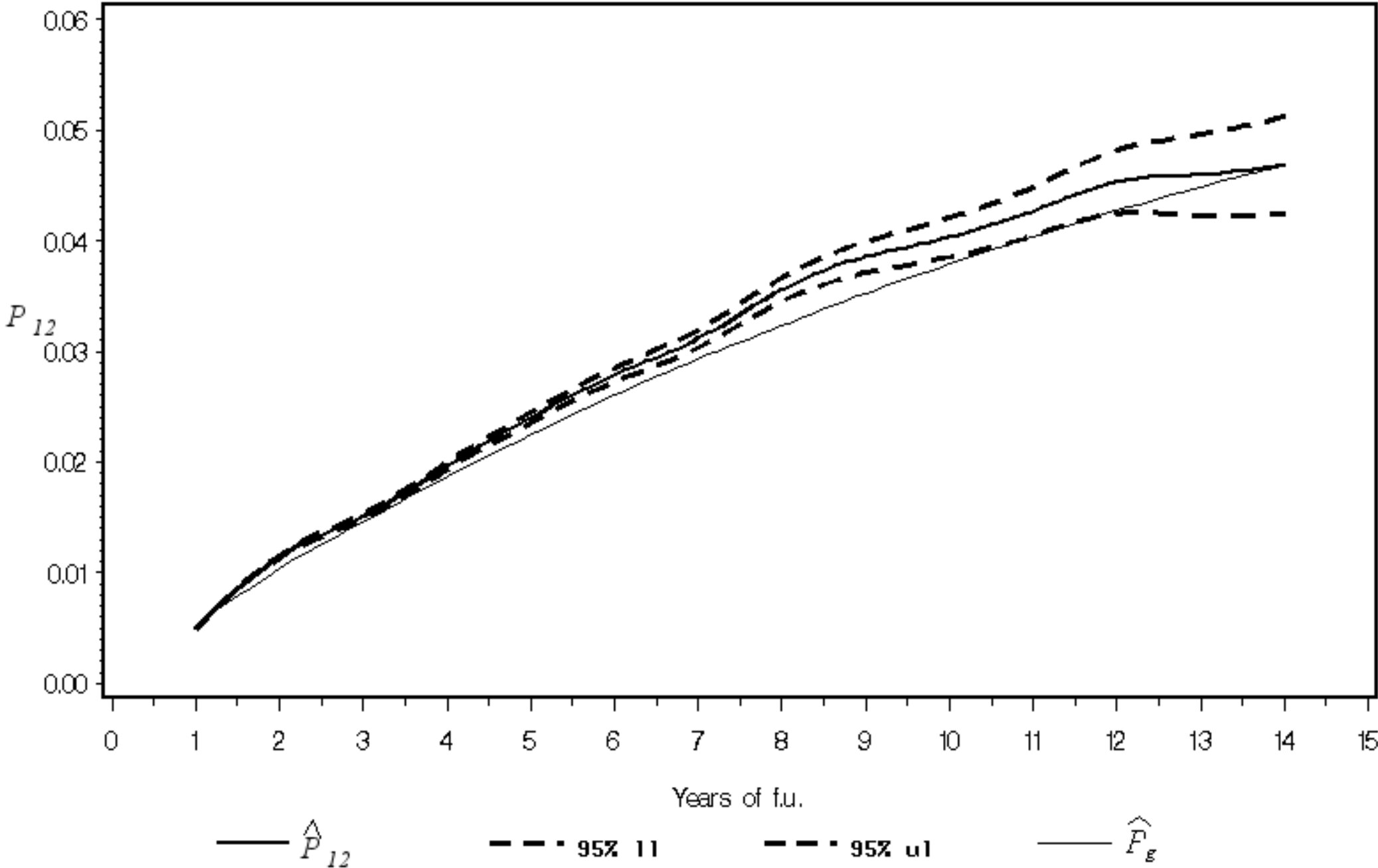


Figure 3

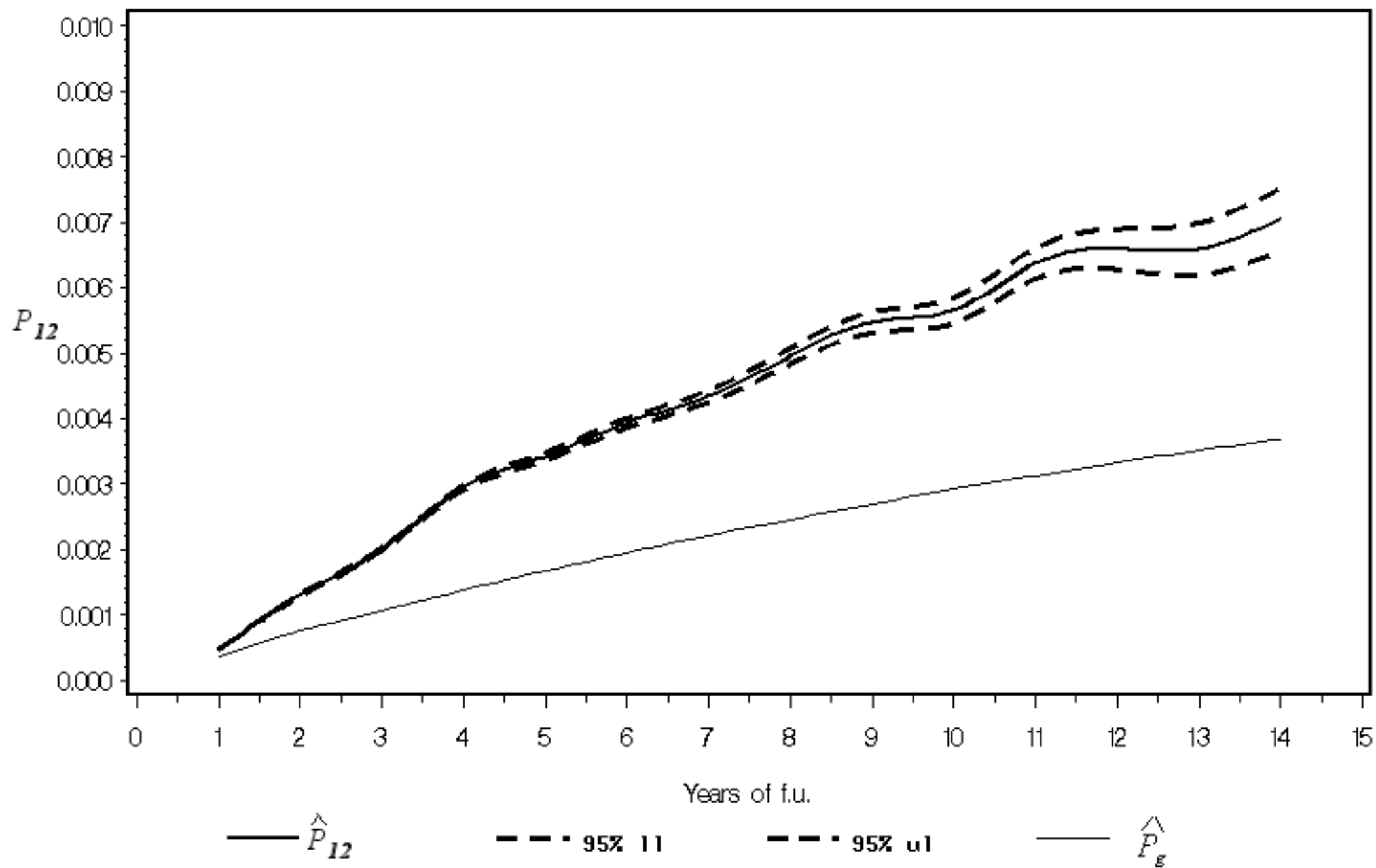


Figure 4

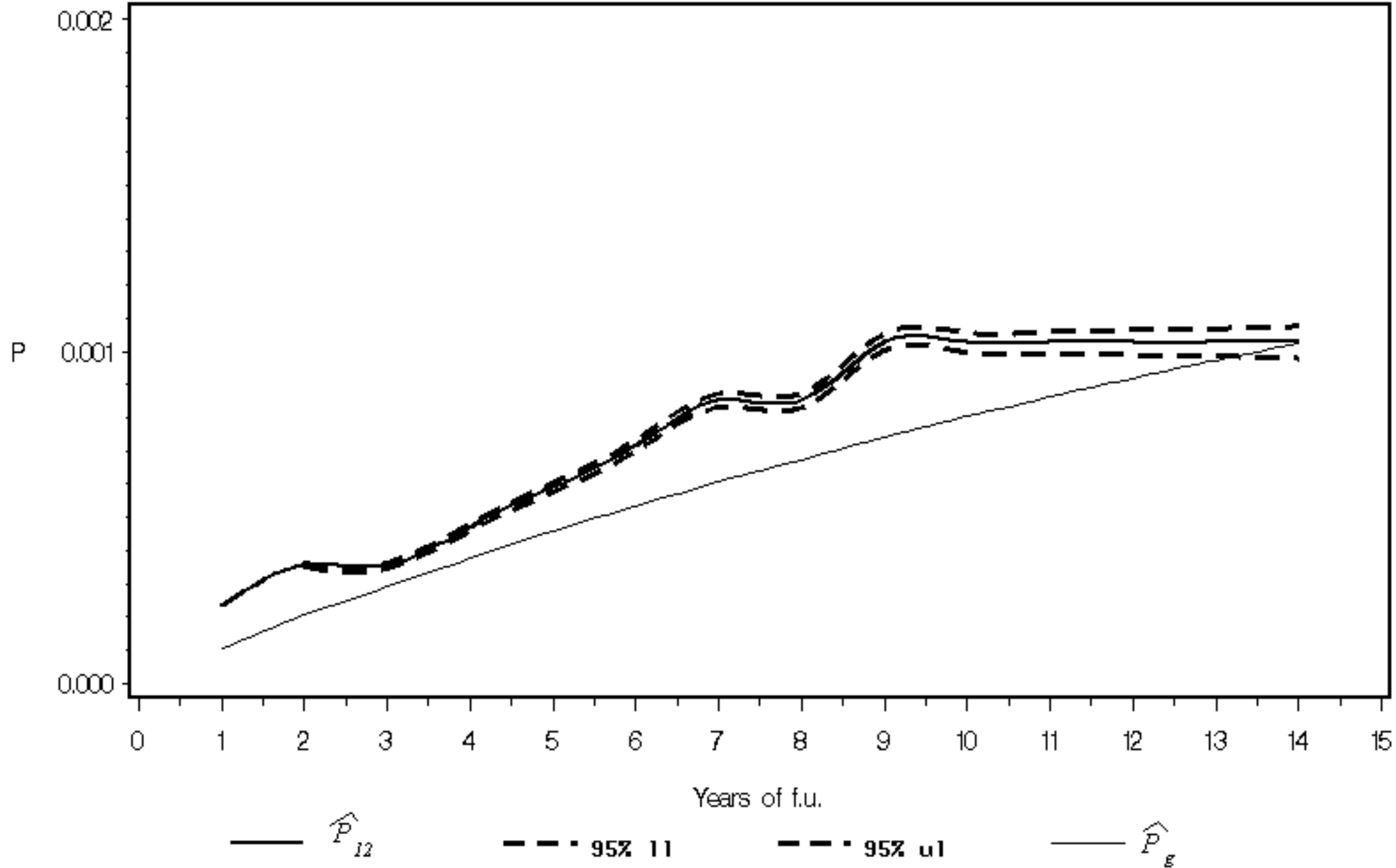


Figure 5