Statistical Methods in Medical Research

Online supplementary material for:

Flexible and structured survival model for a simultaneous estimation of non-linear, non-proportional effects and complex interactions between continuous variables: performance of this multidimensional penalized splines approach in net survival trend analysis.

Laurent Remontet, Zoé Uhry, Nadine Bossard, Jean Iwaz, Aurélien Belot, Coraline Danieli, Hadrien Charvat, Laurent Roche

Corresponding author:
Laurent Remontet
Email: Laurent.remontet@chu-lyon.fr

Contents

Supplementary Table S1. Description of the models used for generation of the data **2**

Supplementary Figure S1. 3D-plots of the theoretical excess mortality hazard for cervix uteri **3**

Supplementary Figure S2. Standardized net survival as a function of the year with 2000 cases **4**

Supplementary Figures S3-S6. Excess mortality hazard by age as a function of time, by method and sample size **5**
Figure S3. MPS approach with 2000 cases 5
Figure S4. MPS approach with 10000 cases 6
Figure S5. PH model with 2000 cases 7
Figure S6. PH model with 10000 cases 8

Supplementary Figures S7-S10. Age-specific net survival as a function of year, by method and sample size **9**
Figure S7. MPS approach with 2000 cases 9
Figure S8. MPS approach with 10000 cases 10
Figure S9. PH model with 2000 cases 11
Figure S10. PH model with 10000 cases 12

Case study: trends in net survival and in the dynamics of excess hazard from cervical cancer, in France **13**

Supplementary Table S1. Model used for $\log[h_{\mathrm{E}}(t,a,y)]$ to generate the data in the simulation study (FP: Fractional Polynomial).

Scenario	Cancer site	Type of model used for $\log[h_{\mathrm{E}}(t,a,y)]$ [(1)]	Detail of the model used for $\log[h_{\mathrm{E}}(t,a,y)]$ [(2)]	Characteristics
1	Esophagus	$\sim f(t) + h(t) \times a + g(a)$	$\sim FP_0(t) + FP_1(t) \times a + \beta_7 a + \beta_8 \left(1/(\sqrt{a})\right) + \beta_9 \log(a)$	No effect of y
2	Stomach	$\sim f(t) + h(t) \times a + g(a) +$ $k(y) + y{:}\,\gamma(t)$	$\sim FP_0(t) + FP_1(t) \times a + \beta_7 a + \beta_8 (1/a^2) + \beta_9 a^3 +$ $\beta_{10}\left(\frac{1}{y^2}\right) + \beta_{11} y^2 + y{:}\left[\beta_{12}\left(\frac{1}{(t+1)}\right) + \beta_{13}\log(t+1) + \beta_{14} t\right]$	NLIN-NPH effect of y and no a-y interaction
3	Breast	$\sim f(t) + h(t) \times a + g(a) +$ $k(y) + n(a){:}\,k(y)$	$\sim FP_0(t) + FP_1(t) \times a + \beta_7 a + \beta_8 a^3 + \beta_9 [a^3 \log(a)] +$ $\beta_{10} y^2 + \beta_{11} y^3 + \beta_{12} a^2 y^2 + \beta_{13} a^2 y^3 + \beta_{14} a^3 y^2 + \beta_{15} a^3 y^3$	NLIN-**PH** effect of y with a-y interaction
4	Cervix uteri	$\sim f(t) + h(t) \times a + g(a) +$ $y + y{:}\,t +$ $n(a){:}\,(y + y{:}\,t)$	$\sim FP_0(t) + FP_1(t) \times a + \beta_7 a + \beta_8 a^2 + \beta_9 \log(a) +$ $\beta_{10} y + \beta_{11} \mathrm{yt} +$ $\beta_{12} a^2 y + \beta_{13} a^2 \mathrm{yt} + \beta_{14} a^3 \log(a)\, y + \beta_{15} a^3 \log(a)\, \mathrm{yt}$	**LIN**-NPH effect of y with **high** a-y interaction and **triple** t-a-y interaction
5	Ovary	$\sim f(t) + h(t) \times a + g(a) +$ $k(y) + y{:}\,t +$ $n(a){:}\,[k(y) + y{:}\,t]$	$\sim FP_0(t) + FP_1(t) \times a + \beta_7 a + \beta_8 \sqrt{a} + \beta_9 a^2 +$ $\beta_{10} y^2 + \beta_{11} y^2 \log(y) + \beta_{12} y{:}\, t +$ $\beta_{13} y^2 \sqrt{a} + \beta_{14} y^2 \log(y) \sqrt{a} + \beta_{15} y t \sqrt{a} + \beta_{16} y^2 log(a) \sqrt{a} +$ $\beta_{17} log(a) \sqrt{a} y^2 \log(y) + \beta_{18} log(a) \sqrt{a} yt$	**NLIN-NPH** effect of y and **complex triple** t-a-y interaction

(1) Choice of scenarios (cancer site combined with type of model) were guided by the French results of SUDCAN study. Note that for esophagus (scenario 1 with no year-effet), a slight linear effect of y was actually observed in SUDCAN and that for cervix uteri, the type of model is based on an updated analysis of French data (in SUDCAN, where period of diagnosis ended in 2004, no year effect was actually found)

(2) $FP_0(t) = \beta_0 + \beta_1 \left(\frac{1}{t+1}\right) + \beta_2\, log(t+1) + \beta_3 t$ and $FP_1(t) = \beta_4 \left(\frac{1}{t+1}\right) + \beta_5\, log(t+1) + \beta_6 t$

Figure S1. Theoretical excess mortality hazard used for the generation of cervix uteri data in scenario 4. First row: 3d-plot of the excess hazard as a function of time and age, at years 1990, 2000, and 2010; second row: 3d-plot of the excess hazard as a function of year and age, at 0.5, 1, and 5 years; third row: 3d-plot of the excess hazard as a function of year and time, at 3 ages (10^{th}, 50^{th}, and 90^{th} percentiles of the age distribution of the cases).

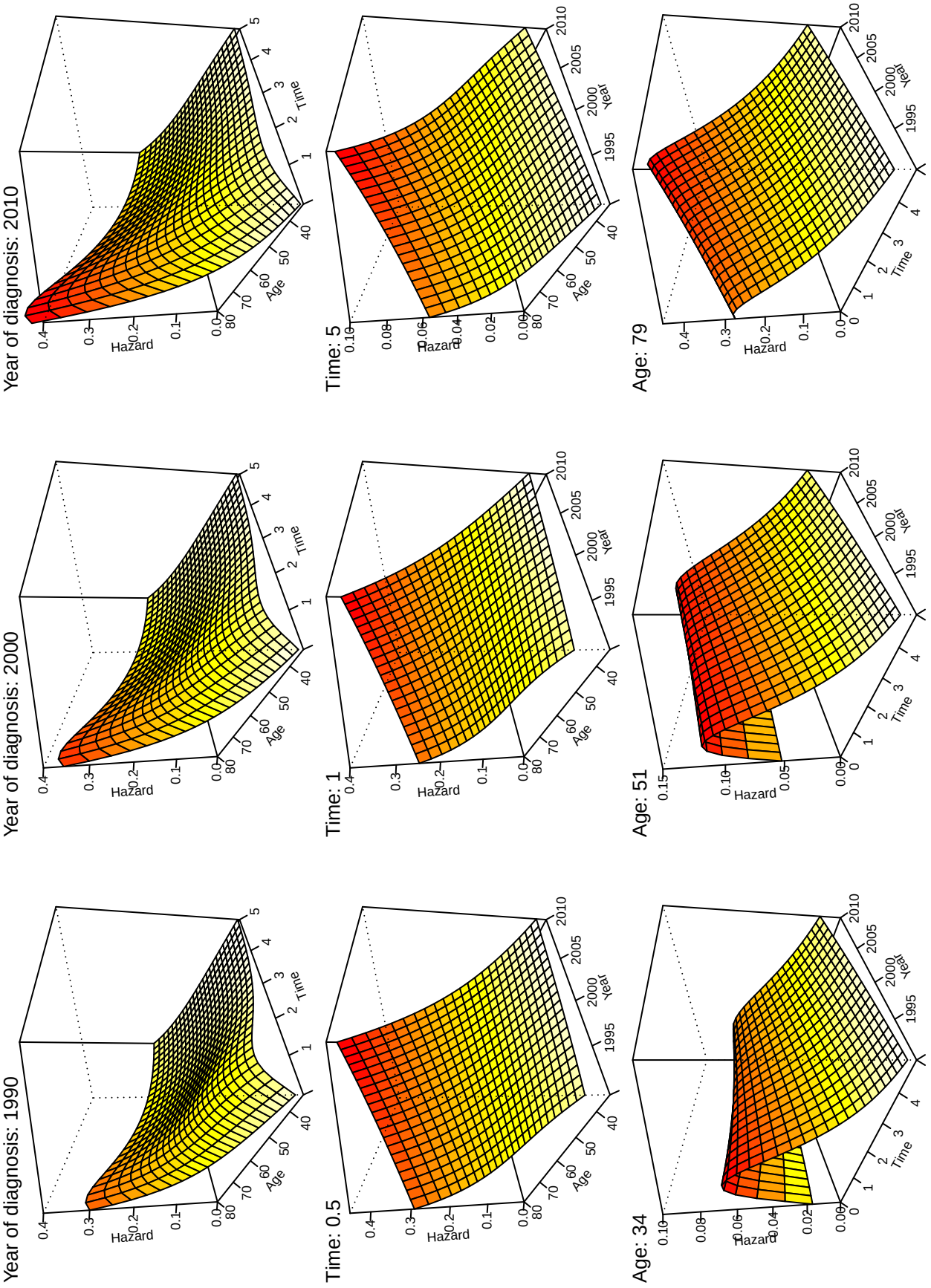


Figure S2. Standardized net survival at 1 and 5 years as a function of the year of diagnosis in the five scenarios, with 2000 cases. Black solid curve: Theoretical standardized net survival; blue double-dashed curve: Mean of the standardized net survival estimated using the Proportional Hazard model (PH); red dashed curve: Mean of the standardized net survival estimated using the multidimensional penalized splines approach (MPS).

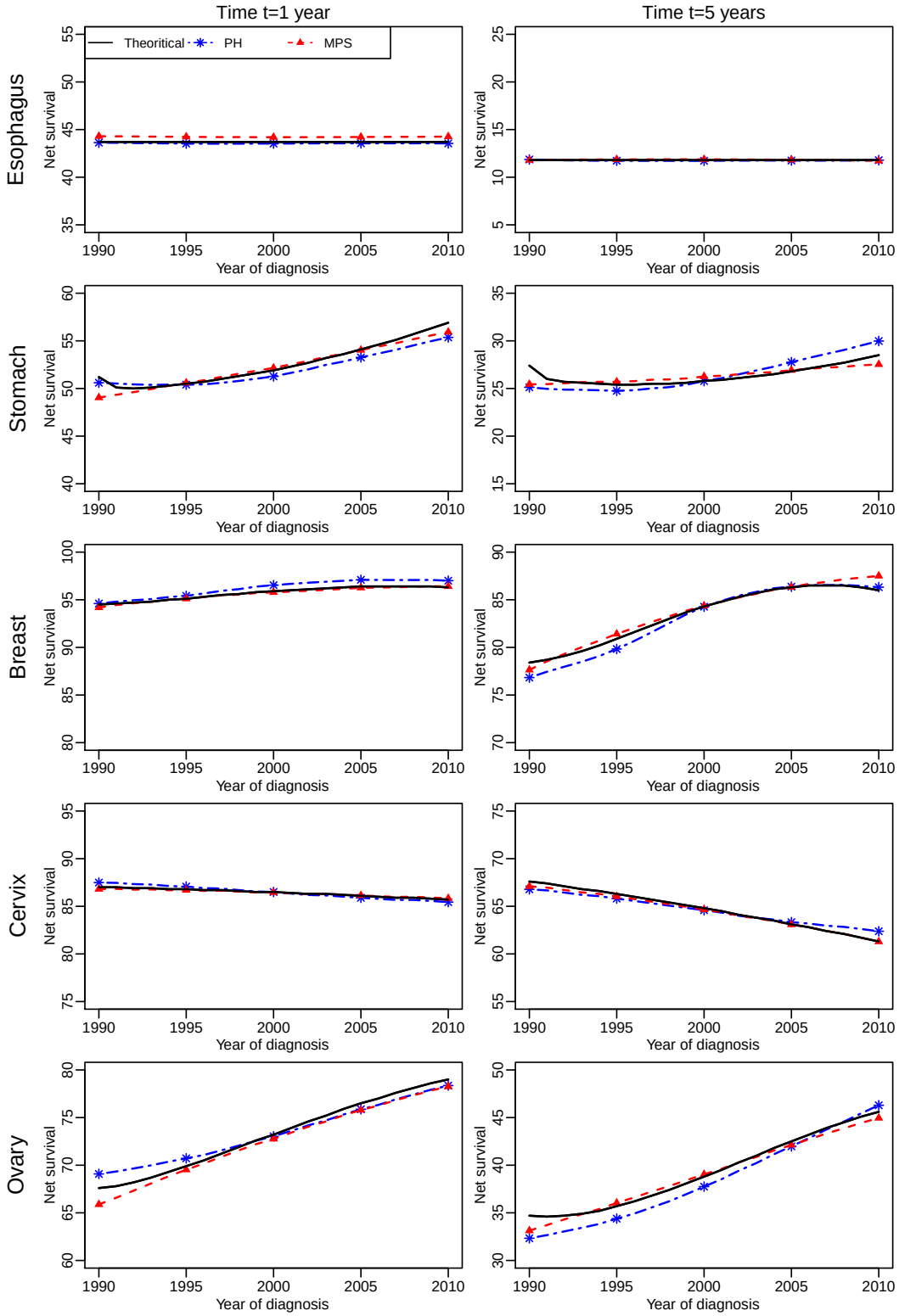


Figure S3. Excess mortality hazard as a function of time since diagnosis in the five scenarios **with 2000 cases**, at 3 ages (10^{th}, 50^{th}, and 90^{th} percentiles of the age distribution of the cases). Solid curve: Theoretical excess mortality hazard; dashed curve: Mean of the excess mortality hazard using the **multidimensional penalized splines approach**.

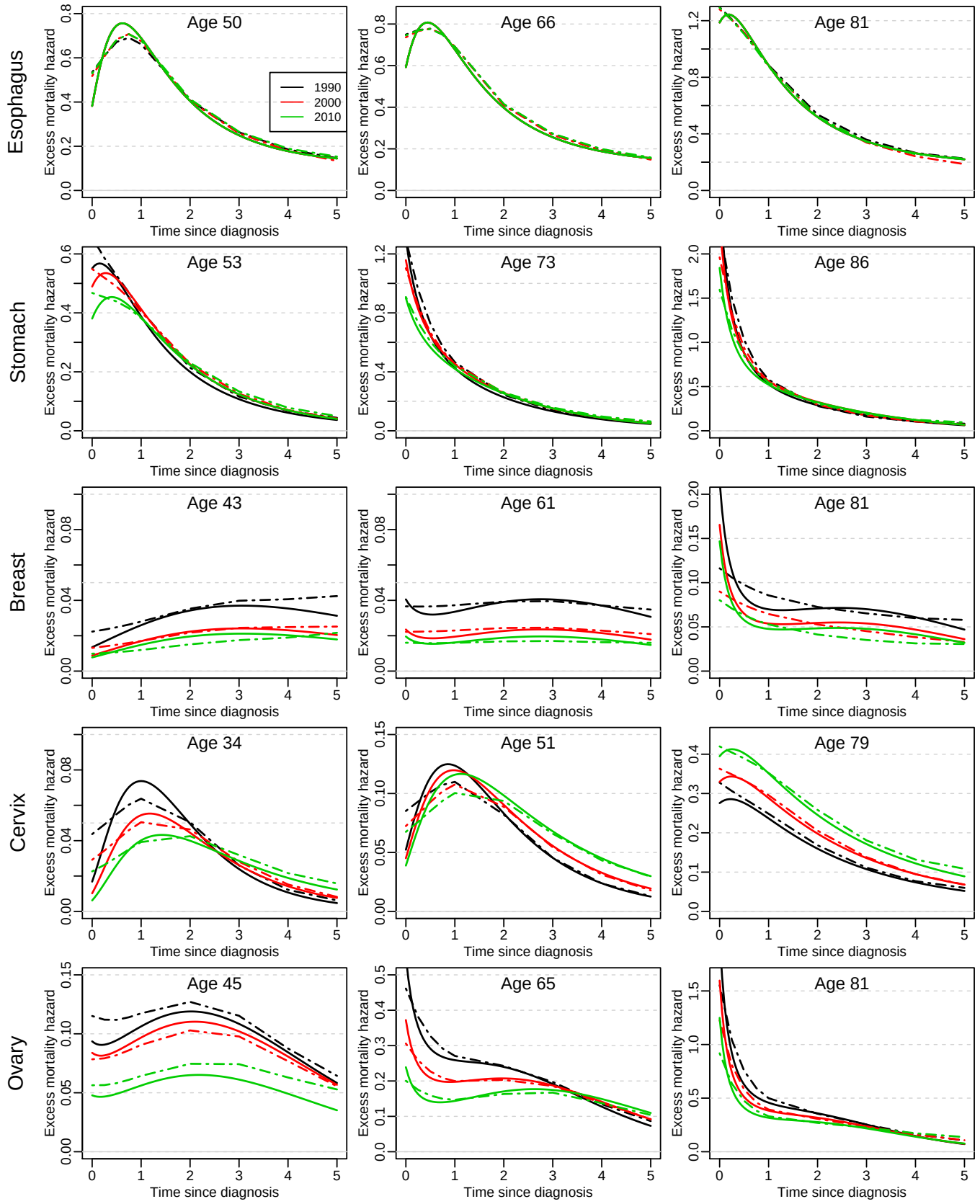


Figure S4. Excess mortality hazard as a function of time since diagnosis in the five scenarios **with 10000 cases**, at 3 ages (10^{th}, 50^{th}, and 90^{th} percentiles of the age distribution of the cases). Solid curve: Theoretical excess mortality hazard; dashed curve: Mean of the excess mortality hazard using the **multidimensional penalized splines approach**.

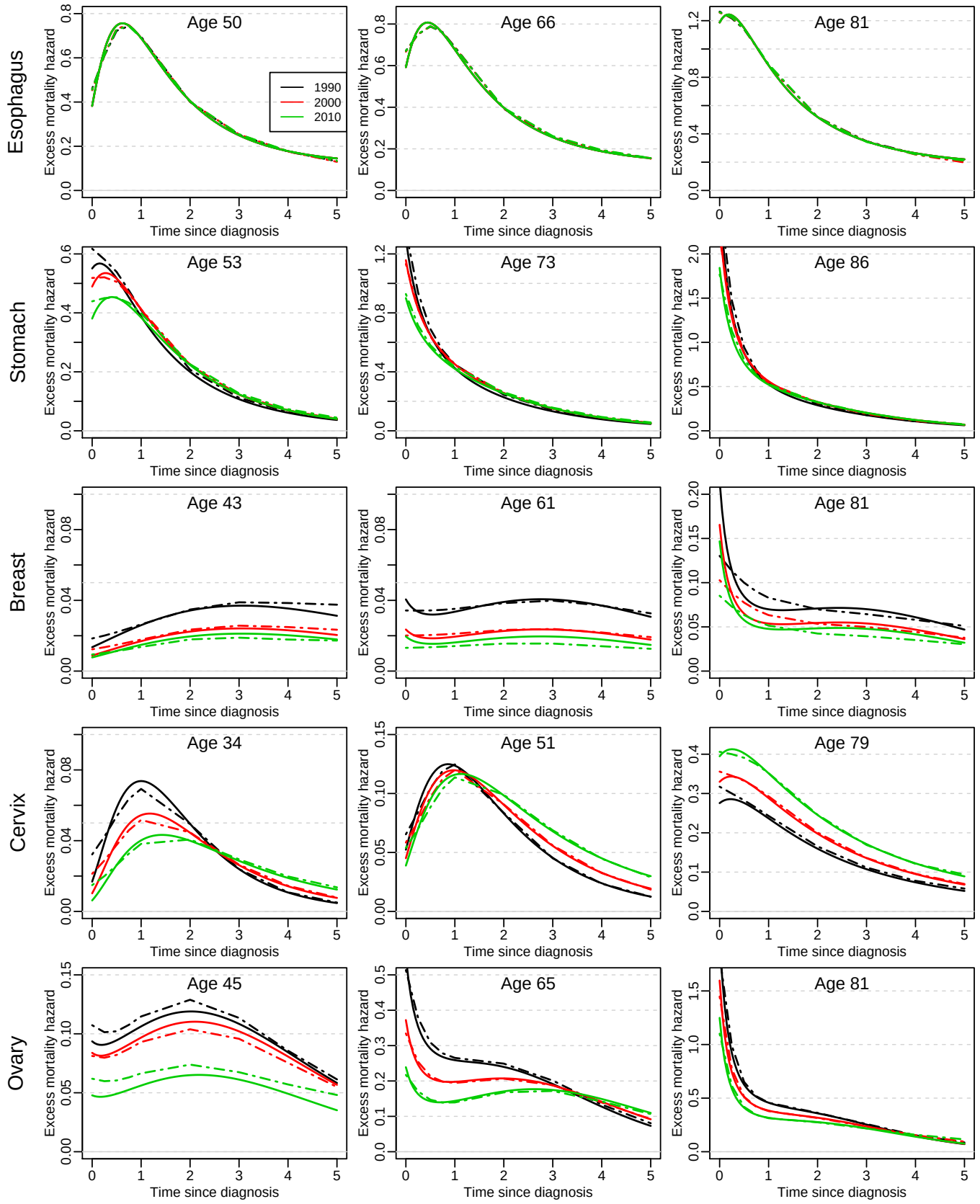


Figure S5. Excess mortality hazard as a function of time since diagnosis in the five scenarios **with 2000 cases**, at 3 ages (10^{th}, 50^{th}, and 90^{th} percentiles of the age distribution of the cases). Solid curve: Theoretical excess mortality hazard; dashed curve: Mean of the excess mortality hazard using the **Proportional Hazard model**.

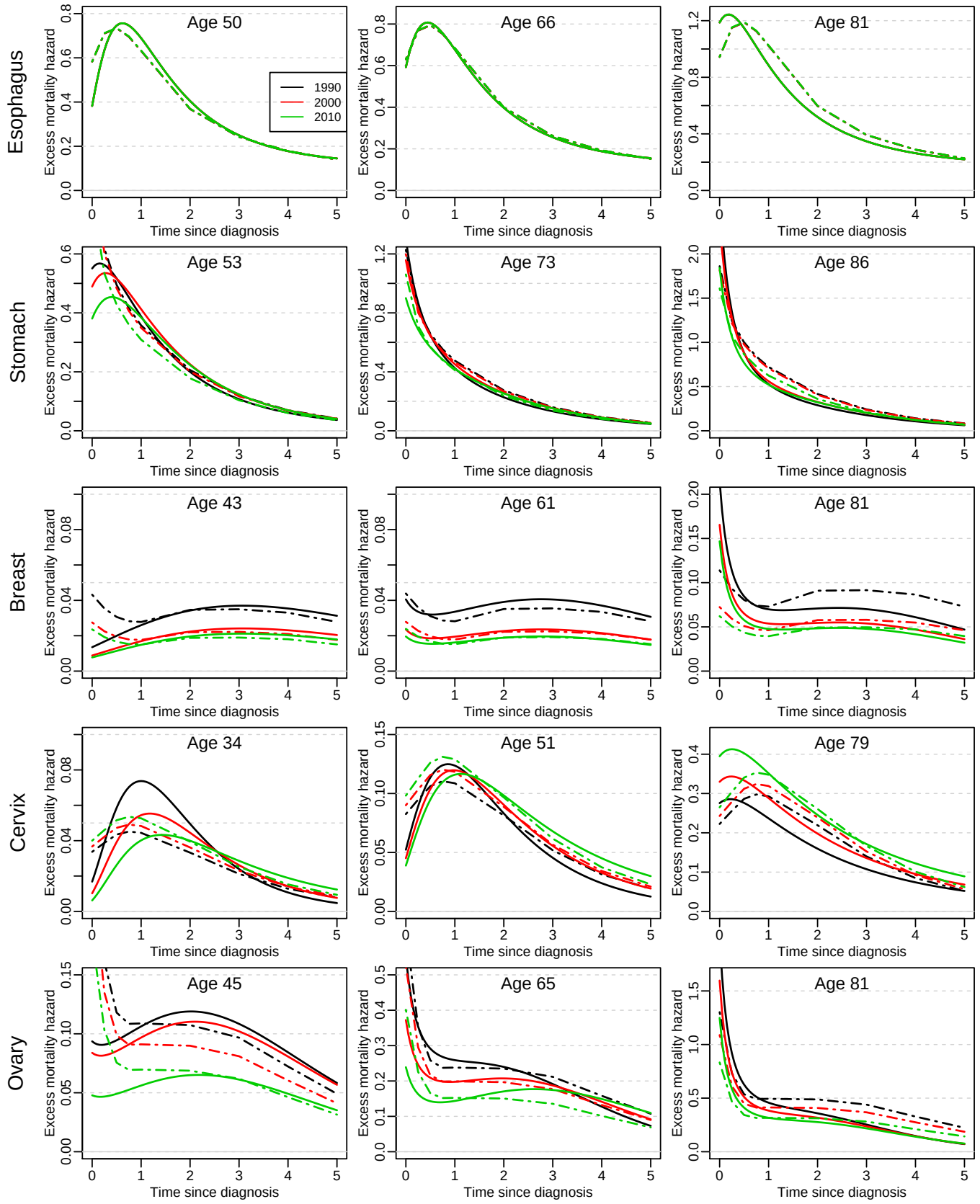


Figure S6. Excess mortality hazard as a function of time since diagnosis in the five scenarios **with 10000 cases**, at 3 ages (10^{th}, 50^{th}, and 90^{th} percentiles of the age distribution of the cases). Solid curve: Theoretical excess mortality hazard; dashed curve: Mean of the excess mortality hazard using the **Proportional Hazard model**.

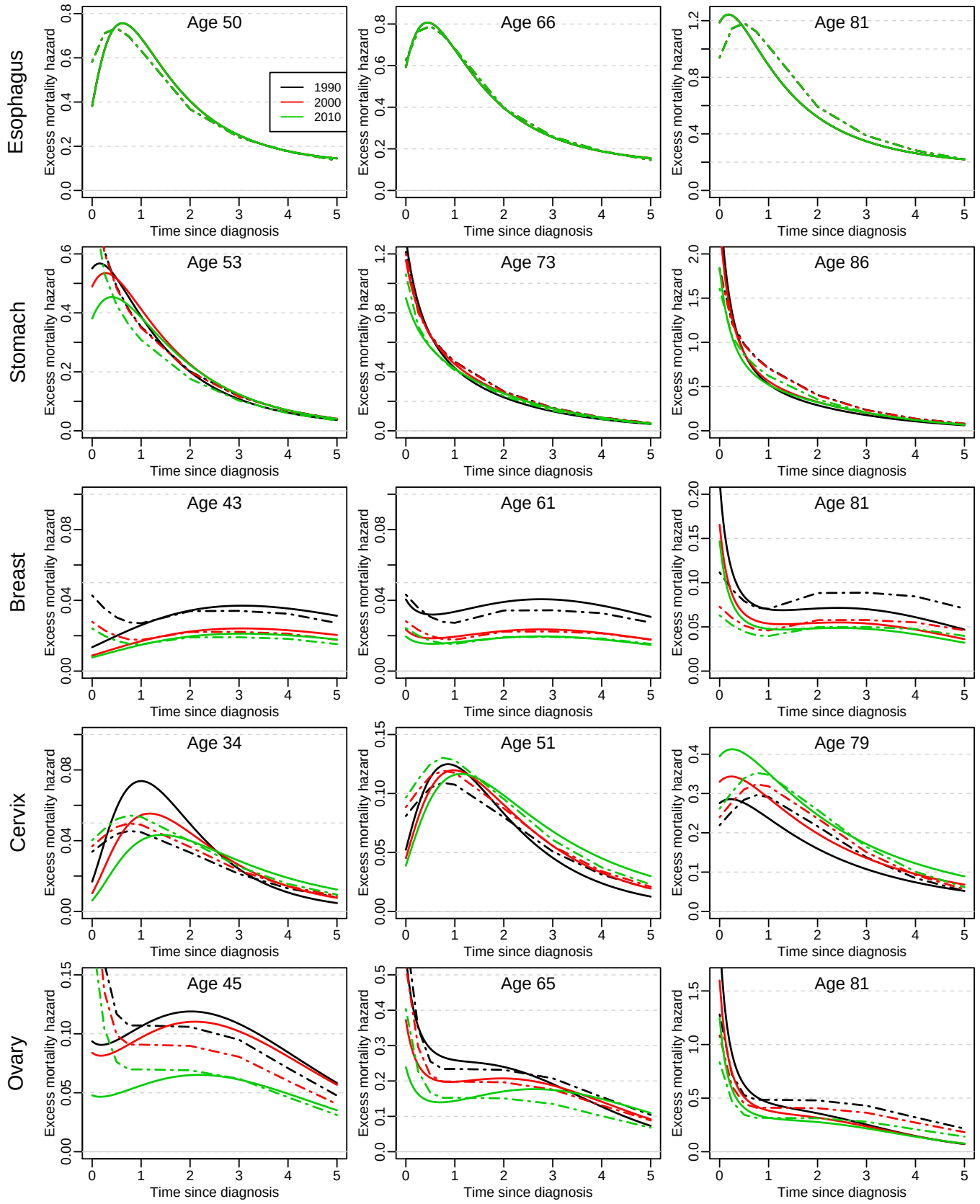


Figure S7. Net survival at 1 and 5 years as a function of the year of diagnosis in the five scenarios **with 2000 cases**, at 3 ages (10th, 50th, and 90th percentiles of the age distribution of the cases). Solid curve: Theoretical age-specific net survival; dashed curve: Mean of the age-specific net survival estimated using the **multidimensional penalized splines approach**.

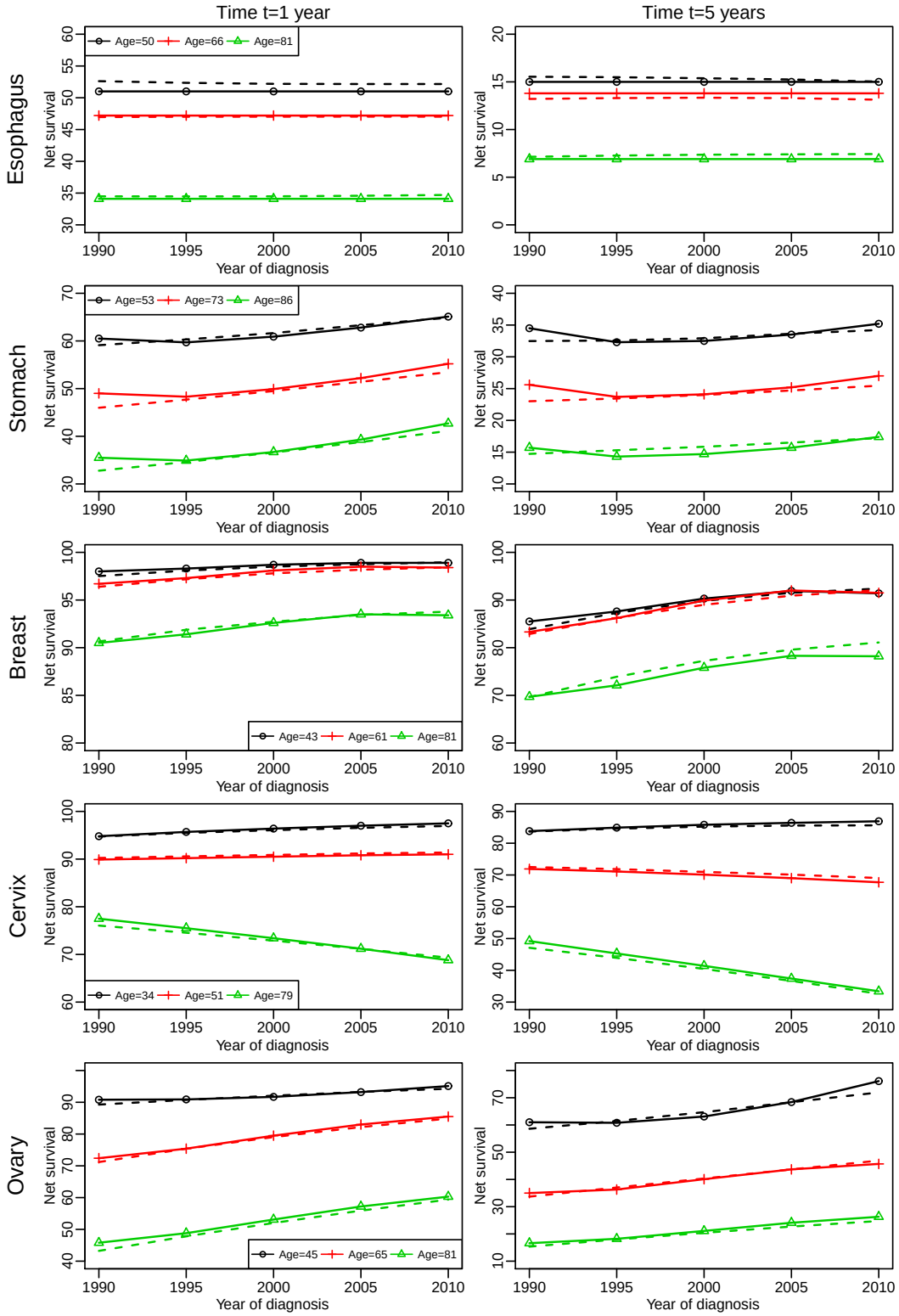


Figure S8. Net survival at 1 and 5 years as a function of the year of diagnosis in the five scenarios **with 10000 cases**, at 3 ages (10th, 50th, and 90th percentiles of the age distribution of the cases). Solid curve: Theoretical age-specific net survival; dashed curve: Mean of the age-specific net survival estimated using the **multidimensional penalized splines approach**.

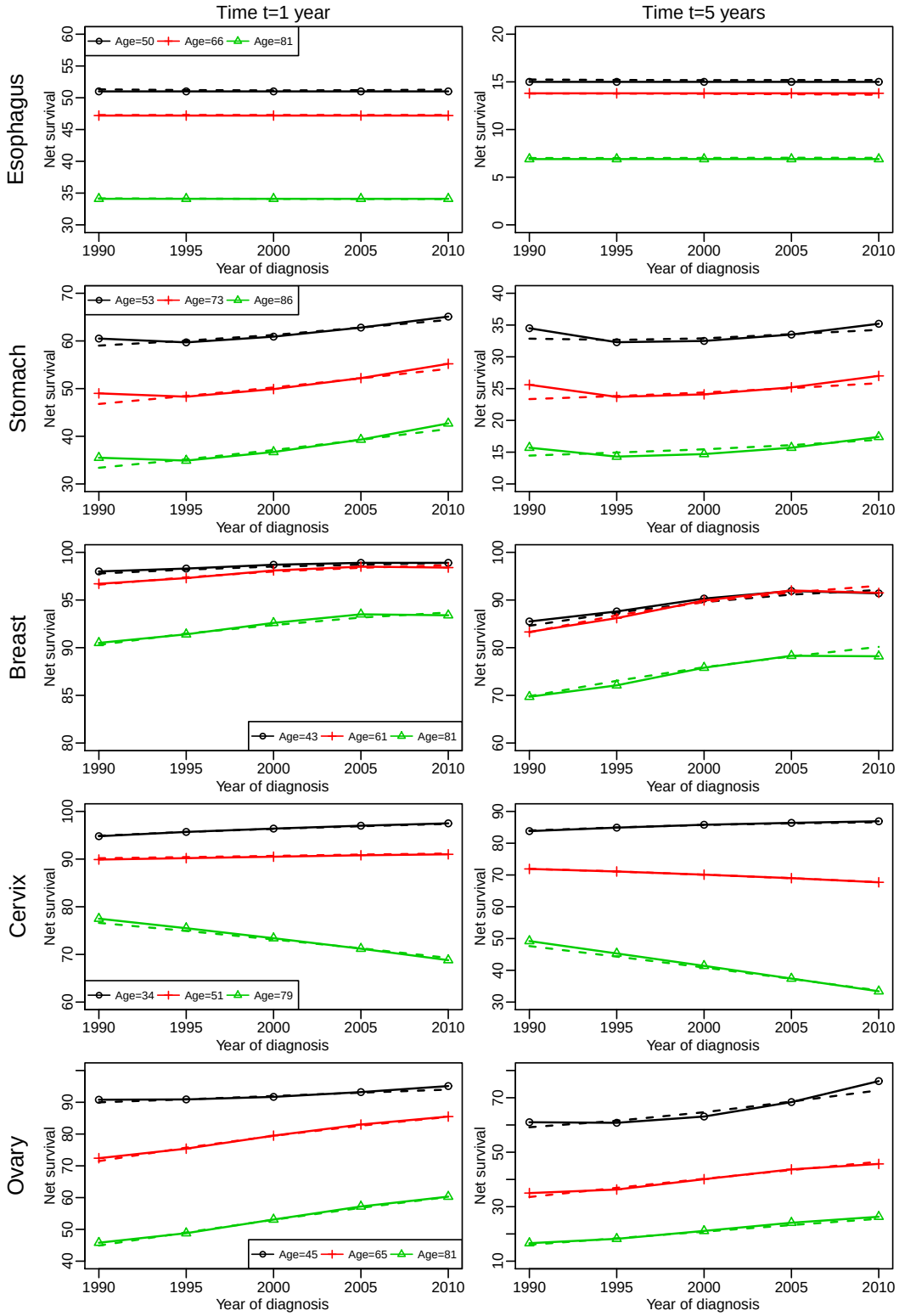


Figure S9. Net survival at 1 and 5 years as a function of the year of diagnosis in the five scenarios **with 2000 cases**, at 3 ages (10th, 50th, and 90th percentiles of the age distribution of the cases). Solid curve: Theoretical age-specific net survival; dashed curve: Mean of the age-specific net survival estimated using the **Proportional Hazard model**.

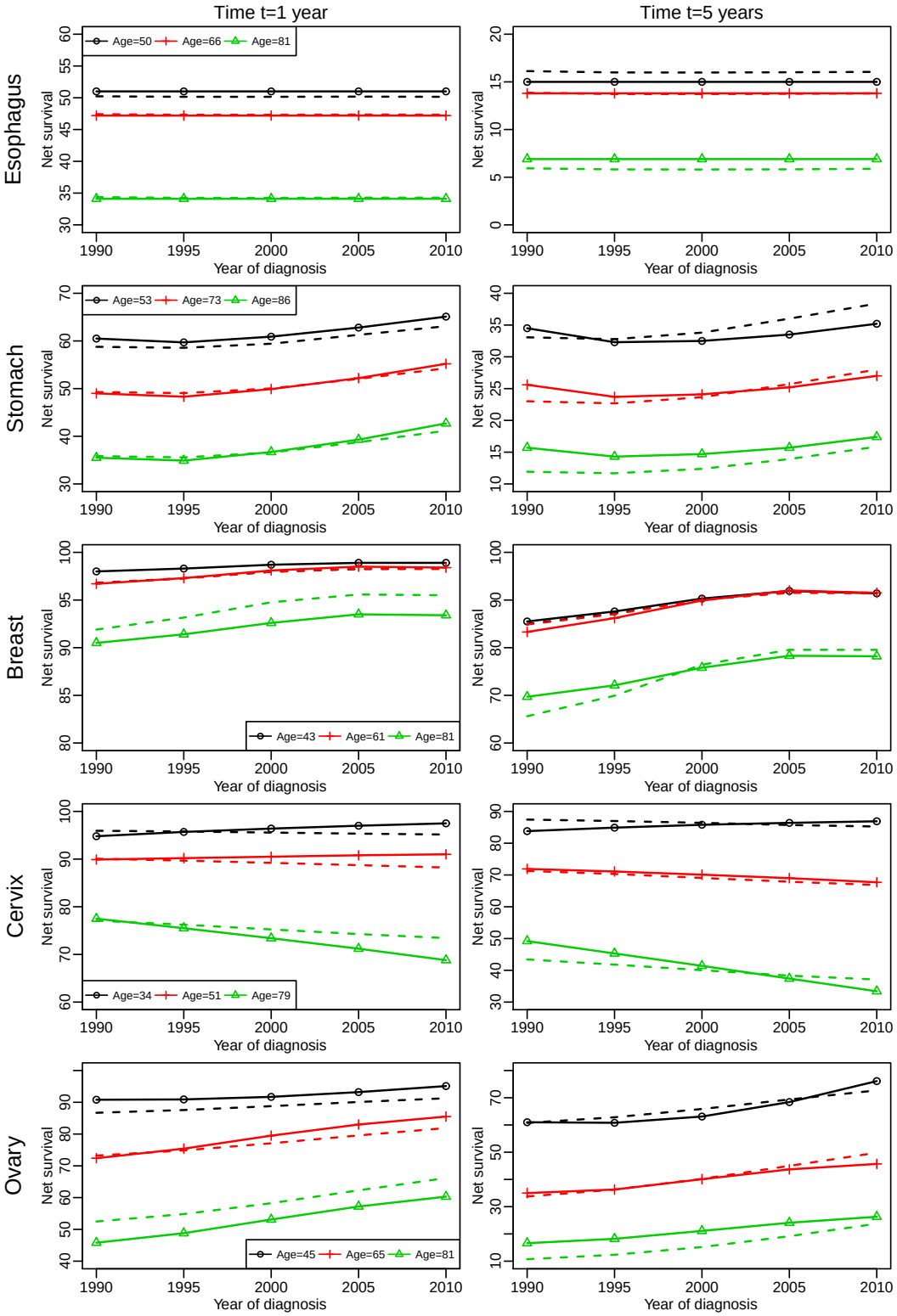
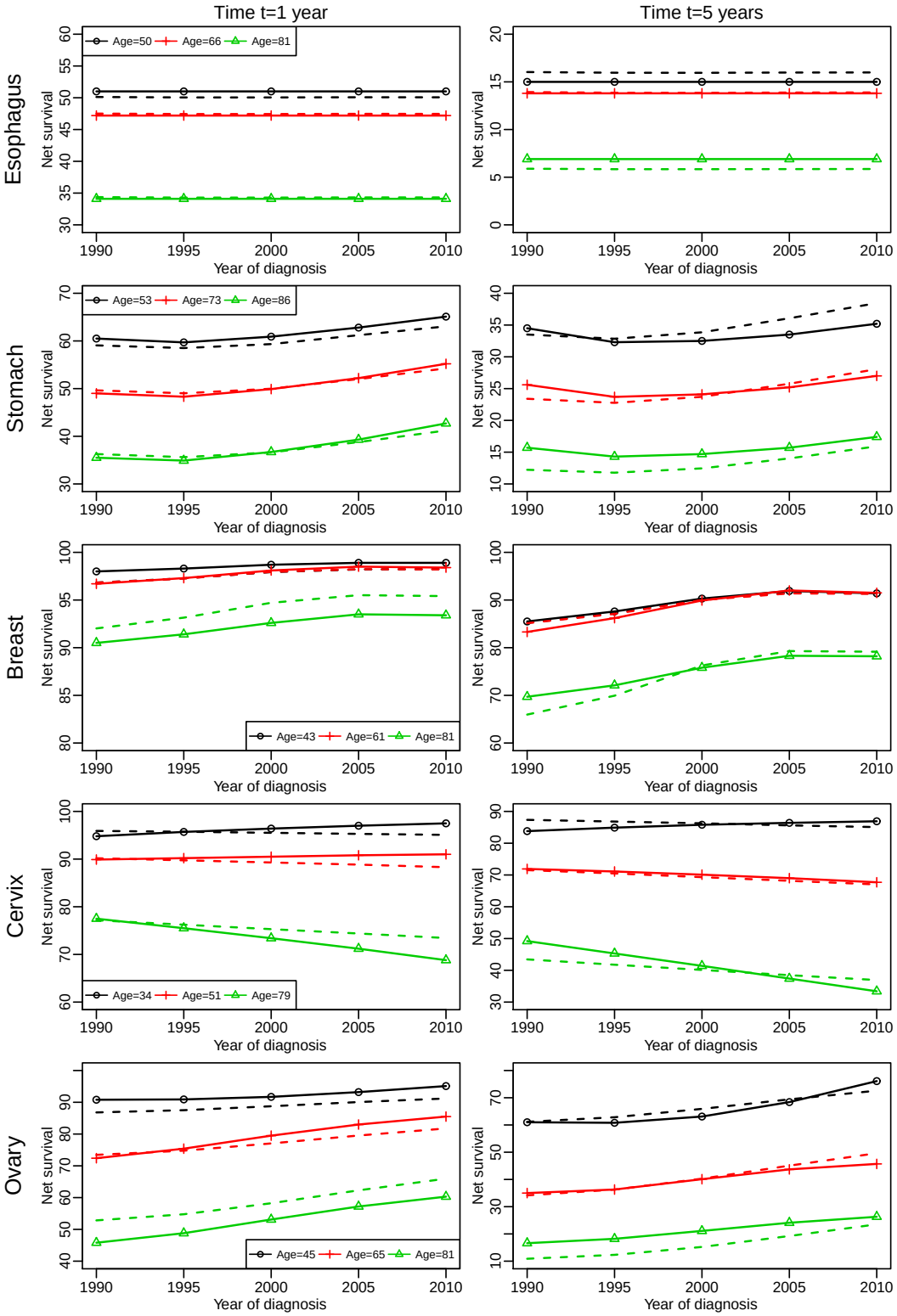


Figure S10. Net survival at 1 and 5 years as a function of the year of diagnosis in the five scenarios **with 10000 cases**, at 3 ages (10^{th}, 50^{th}, and 90^{th} percentiles of the age distribution of the cases). Solid curve: Theoretical age-specific net survival; dashed curve: Mean of the age-specific net survival estimated using the **Proportional Hazard model**.



Case study: trends in net survival and in the dynamics of excess hazard from cervical cancer, in France.

This section is an illustration of a survival trends population-based study, as performed by the Multidimensional Penalized Splines approach (MPS) and the Proportional Hazard model (PH).

Here, we studied trends in net survival (NS) and in the excess hazard for cervical cancer in France; this study included all incident cases of primary invasive cervical cancer (ICD-03 code C53) diagnosed between January 1, 1989 and December 31, 2010 in the area covered by 7 registries of the French Network of Cancer registries (FRANCIM). The end of follow-up was June 30, 2013. This dataset was the one used to determine the theoretical parameters in the cervix uteri scenario (see section 3 of the paper). It included 5977 cervical cancer cases and 2139 (35.8%) deaths were observed within 5 years from diagnosis. Age at diagnosis ranged from 18 to 100 years (median: 49). More information about this dataset can be found in the works of Cowppli-bony and al.[1, 2]

The MPS and PH approaches were identical to those described in the simulation study (see sections 2.3.3, 2.3.4 of the paper). The age-standardized NS for a given year of diagnosis was also calculated as in the paper. We just recall that, for the MPS approach, the log-excess hazard was modelled as a function of time t, age a, and year of diagnosis y using a tensor product smooth which basis was built using restricted cubic splines of dimension 6, 5, and 4, respectively. The knot location of these splines was based on the empirical percentiles observed in the population of patients who died. The smoothing parameters were estimated using the REML criterion. For the PH approach, the excess hazard was modelled as

$$\log(h_E(t,a,y)) = f_t(t) + f_a(a) + f_y(y), \text{where}$$

f_t, f_a, and f_y were restricted cubic splines with the same features as the marginal bases of the MPS approach (same number and location of the knots). The 13 parameters of this PH model were obtained using maximum likelihood method (without any penalization).

We also replicated the analysis performed in Cowppli-bony and al,[1, 2] which is very typical of what has been done up-to-now in survival trends studies. In this study, NS was estimated using the non-parametric estimator of Pohar-Perme[3] (PP) and analysis was stratified by age-class (5 strata), and period of diagnosis (4 strata).

The resulting trends in age-standardized NS at 1 and 5 years are depicted in Figure S11. The MPS estimates are reasonably concordant with the PP estimates, whereas an unobserved increase in standardized NS at 5 years after year 2005 was obtained with the PH approach.

Figure S11. Standardized net survival at 1 and 5 years as a function of the year of diagnosis in Cervical cancer. Red solid curve: Multidimensional Penalized Splines approach (MPS); blue dashed curve: Proportional Hazard model (PH); gray segment: non-parametric estimation using the Pohar-Perme method with 95% CI (vertical bar).

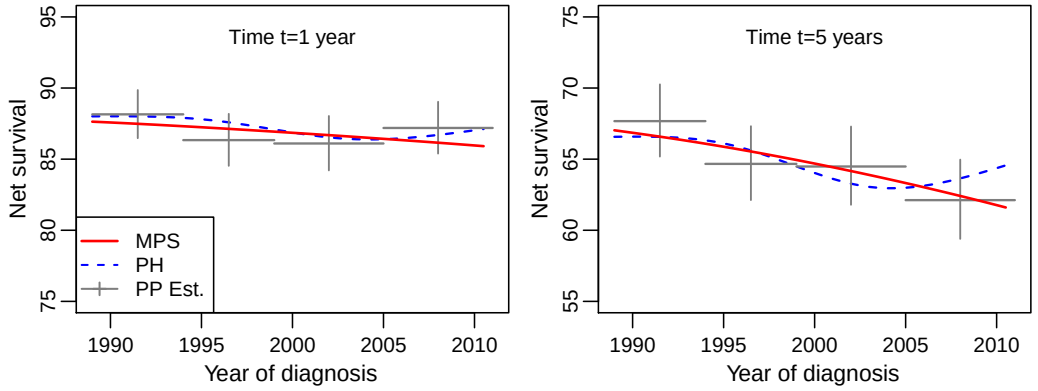


Figure S12. Net survival (NS) at 1 and 5 years as a function of the year of diagnosis in Cervical cancer, by age. The gray segments correspond to the estimates by period and age-class obtained with the Pohar-Perme method with 95% CI (vertical bar). Using the Multidimensional Penalized Splines approach (MPS; red solid curve) and the Proportional Hazard approach (PH; blue dashed curve), NS was estimated at 5 ages, each age corresponding to the median of age within each of the 5 age-classes.

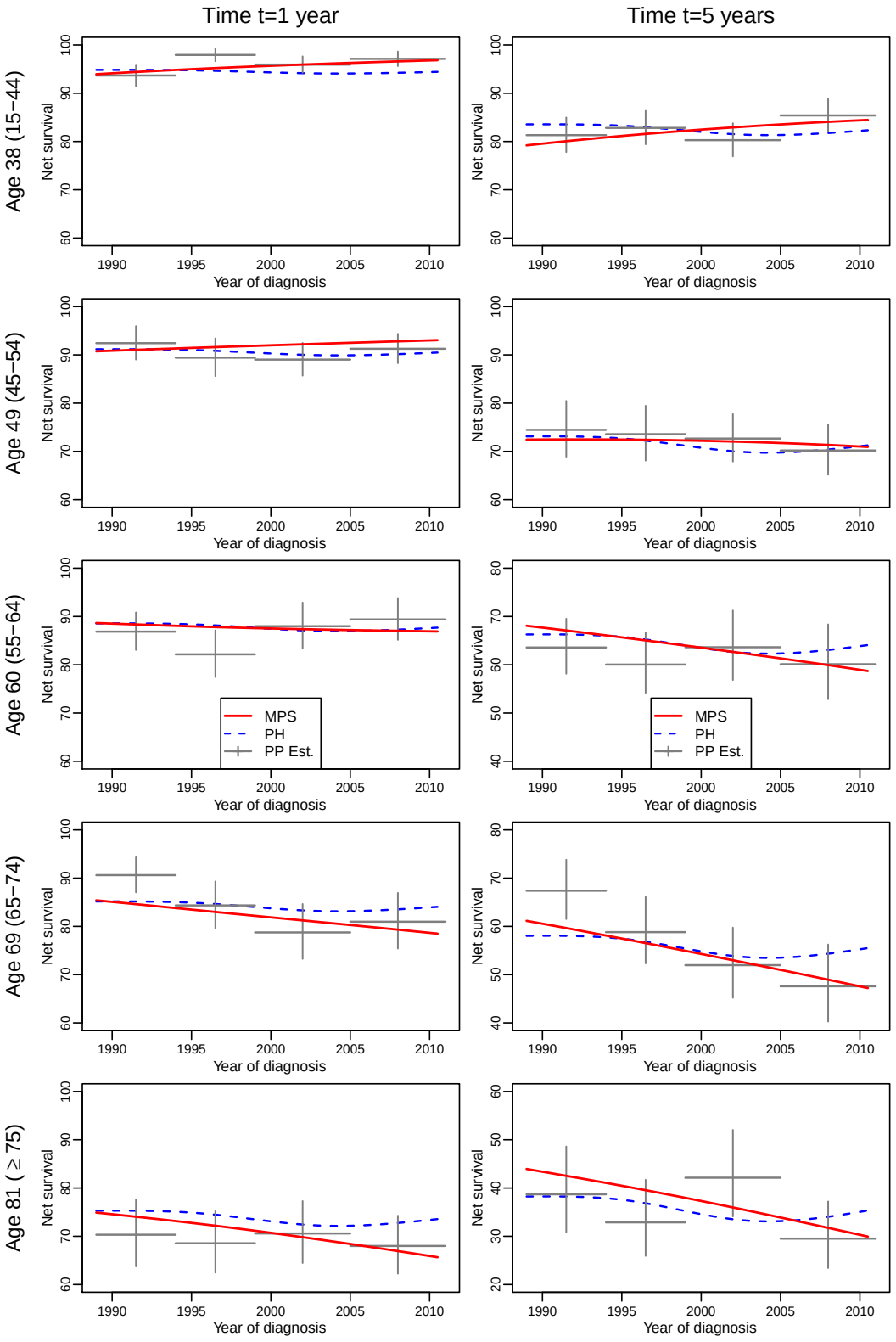
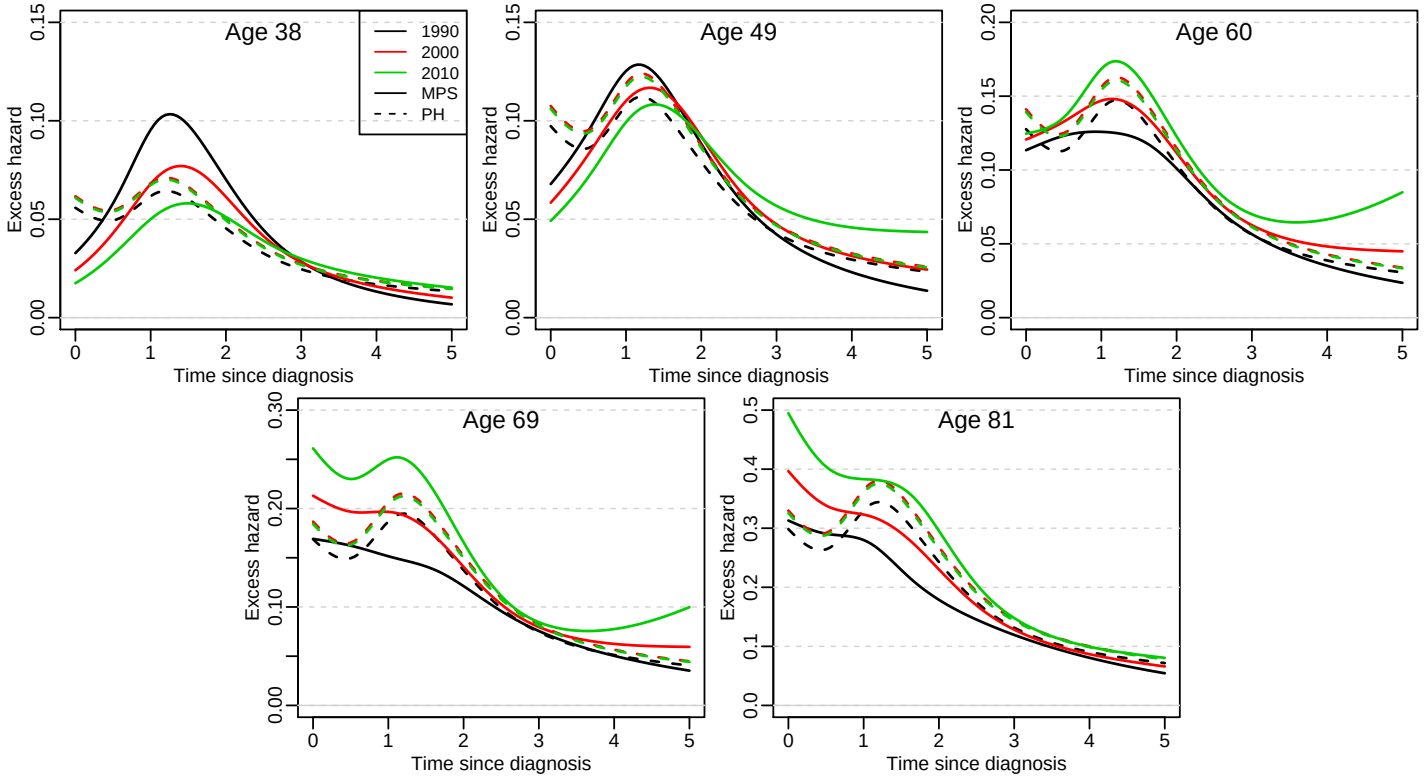


Figure S12 shows the corresponding trends by age-class (PP) or at the median age determined within each class (MPS and PH approaches). The MPS approach (red solid line) showed well distinct trends in survival at 1 and 5 years across ages, with an improvement observed in younger women and deterioration in older women. This pattern was overall confirmed by the PP estimates, although variability of these estimates led to somewhat erratic behaviors. As for the PH approach (blue dashed line), the pattern of trends in survival was inevitably similar whatever the time and the age because of the constraints induced by this model: survival decreased between years 1989 and circa 2004, then increased afterwards.

Figure S13 shows the dynamics of the excess hazard by age and year of diagnosis. The PH assumption and the absence of interaction (dashed curves) can clearly be seen in this graph; for example, the resulting excess hazard for y=2000 was higher than for y=1990 whatever the time and age. Conversely, the MPS approach provided a more complex picture of the dynamics of the excess hazard, exhibiting strong time-age-year interactions. So the dynamics were different according to age; excess hazard decreased regularly with time at older ages whereas it peaked around 1.5 years from diagnosis at younger ages. Furthermore, excess hazard increased with year of diagnosis for women aged 60 and over throughout the follow-up, while, in younger ages, it mainly decreased with years of diagnosis (this led to the different NS trends according to age seen in figure S12).

Figure S13 thus provides fundamentals medical results and this kind of figure is indispensable for clinicians and epidemiologists to help them understand the way medical practises have changed patient mortality over the year of diagnosis.

Figure S13. Excess mortality hazard as a function of time since diagnosis in Cervical cancer, at 5 ages. Solid curve: excess mortality hazard using the Multidimensional Penalized Splines approach; dashed curve: excess mortality hazard using the Proportional Hazard model.



In our view, this example in cervical cancer illustrates the advantages of an efficient modelling approach, such as the MPS one, to study trends in survival and hazard. On one hand, both the degree of details and interpretation of the results are limited with stratified analyses based on PP estimator. On the other hand, the PH approach cannot describe properly the trends in survival or hazard whenever interactions are present. The MPS approach is an appealing alternative to us, as it is able to catch complex trends, but still provides smooth estimates.

The R-code to reproduce this analysis is available on the GitHub repository **https://github.com/RocheLHCL/SMMR_Remontet2018** (Cf. the readme.pdf for explanations of the contents). However, due to copyright issues, we cannot provide the original real dataset. So, we provided one of the simulated dataset used in the simulation study on cervix uteri cancer data on 10,000 patients. The results may thus differ, to some extent, from those presented in the article.

References

1. Cowppli-Bony A, Uhry Z, Remontet L, et al. Survival of solid cancer patients in France, 1989-2013: a population-based study. *Eur J Cancer Prev*. 2017; 26: 461-8.
2. Cowppli-Bony A, Uhry Z, Remontet L, et al. Survie des personnes atteintes de cancer en France métropolitaine, 1989-2013. Partie 1 - Tumeurs solides. Saint-Maurice: Institut de veille sanitaire, 2016. *http://invs.santepubliquefrance.fr/fr../layout/set/print/Publications-et-outils/Rapports-et-syntheses/Maladies-chroniques-et-traumatismes/2016/Survie-des-personnes-atteintes-de-cancer-en-France-metropolitaine-1989-2013-Partie-1-tumeurs-solides*
3. Perme MP, Stare J and Esteve J. On estimation in relative survival. *Biometrics*. 2012; 68: 113-20.