Appendix

Temporal trends in the proportion of "cure" in children, adolescents and young adults diagnosed with chronic myeloid leukemia in England: a population-based study

Daniel Drozdov[1,2], Audrey Bonaventure[1], Kayo Nakata[1,3], Meinolf Suttorp[4] and Aurélien Belot[1]

Affiliations

[1] Cancer Survival Group, Department of Non-Communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK

[2] Division of Pediatric Oncology, Department of Pediatrics, Cantonal Hospital of Aarau, Aarau, Switzerland

[3] Cancer Control Centre, Osaka International Cancer Institute, Osaka, Japan

[4] Division of Pediatric Hematology-Oncology-Blood and Marrow Transplantation, University Hospital Carl Gustav Carus, Dresden, Germany

Statistical details of the parametrizations for the univariable and the multivariable mixture cure models

Mixture cure models

A mixture cure model expresses the survival $S(t)$ as a mixture of two sub-populations, those who will not experience the event (the "cured") and those who will (the "uncured")

$$S(t) = \pi + (1 - \pi)S_u(t) \tag{1}$$

where π is the proportion of "cured" and $S_u(t)$ is the survival function for the "uncured".

When the event of interest is death (instead of disease recurrence), we need to take account of the fact that patients can die from other causes than cancer. Therefore, if using data on death from all causes, mortality due to other diseases clearly needs to be considered:

$$S(t) = S_0(t)\pi + (1 - \pi)S_0(t)S_u(t) \tag{2}$$

where $S_0(t)$ is the survival from the general population (that is the "expected" survival).

We here analyzed cancer data among children and young adults in England. In this population, deaths from causes other than cancer are very rare. Therefore, we have assumed expected mortality to be negligible and therefore $S_0(t) = 1$.

So, the models used in our work were based on equation (1).

Univariable cure model

We assumed a parametric Weibull distribution for the survival of the "uncured" $S_u(t)$, meaning that the hazard function is defined as $h_u(t) = \lambda\kappa t^{\kappa-1}$, and that the survival function is $S_u(t) = exp\{-\lambda t^{\kappa}\}$, where λ is the scale parameter and κ is the shape parameter.

To constraint the proportion of "cured" π to be between 0 and 1, we used a logistic transformation $logit(\pi) = log\left(\frac{\pi}{1-\pi}\right) = \alpha_0$

Multivariable cure models

For the multivariable cure models, we expressed the quantities detailed above as functions of covariables (the vector $\boldsymbol{x}$):

$$h_u(t, \boldsymbol{x}) = \lambda\kappa t^{\kappa-1} e^{\boldsymbol{\beta}'\boldsymbol{x}}$$

$$logit(\pi(\boldsymbol{x})) = log\left(\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right) = \boldsymbol{\alpha}'\boldsymbol{x}$$

In our application, we used age at diagnosis (centered at 15 years, $= (agediag - 15)$) and year at diagnosis (centered at 1990, $ydiagc = (ydiag - 1990)$) as covariables. We modelled their association with the "cured" proportion and the hazard of death for the "uncured" through quadratic regression splines, each spline having one knot located at 0 (thus corresponding of 15 years for age, and 1990 for year). The splines were expressed in a truncated power basis, and could be written as

$f(agec) = \beta_1 agec + \beta_2 agec^2 + \beta_3 (agec)_+^2$ for age and
$g(ydiagc) = \alpha_1 ydiagc + \alpha_2 ydiagc^2 + \alpha_3 (ydiagc)_+^2$ for year,

where $(u)_+ = 0$ if $u \leq 0$, and $(u)_+ = u$ if $u > 0$

Finally, the multivariable cure model could be written in its most complex formulation as

$$h_u(t, \boldsymbol{x}) = \lambda\kappa t^{\kappa-1} e^{(\beta_1 ydiagc + \beta_2 ydiagc^2 + \beta_3 (ydiagc)_+^2 + \beta_4 agec + \beta_5 agec^2 + \beta_6 (agec)_+^2)} \quad (3)$$

$$logit(\pi(\boldsymbol{x})) = \alpha_0 + \alpha_1 ydiagc + \alpha_2 ydiagc^2 + \alpha_3 (ydiagc)_+^2 + \alpha_4 agec + \alpha_5 agec^2 + \alpha_6 (agec)_+^2 \quad (4)$$

Other models can be obtained from equations (3) and /or (4). For example, when we constraint the association between $ydiagc$ and the (log-odds of the) cure proportion to be linear, the terms $\alpha_2 ydiagc^2 + \alpha_3 (ydiagc)_+^2$ are removed from equation (4).

Assessing the fit of cure models

From equation (1), one could assess the quality of the fit of a cure models by comparing a non-parametric overall survival estimate to the overall survival derived from the model. This was done in Figure S1 for the univariable model, and in figure S2 for the multivariable cure model for specific subgroups of age at diagnosis and year of diagnosis.

In addition, one could also compare the model-based survival estimates of the uncured $\widehat{S_u}(t)$ to the non-parametric survival estimates after removing the components of cure, i.e. $(\hat{S}(t) - \pi)/(1 - \pi)$. This comparison was done below for the final model (Figure S3), as explained in the method and results sections of the main paper.

Figure S1 Comparison of the non-parametric overall survival estimates to the ones obtained from the univariable mixture cure model, by period of diagnosis. Children and young adults diagnosed with chronic myeloid leukemia in England, 1980-2005

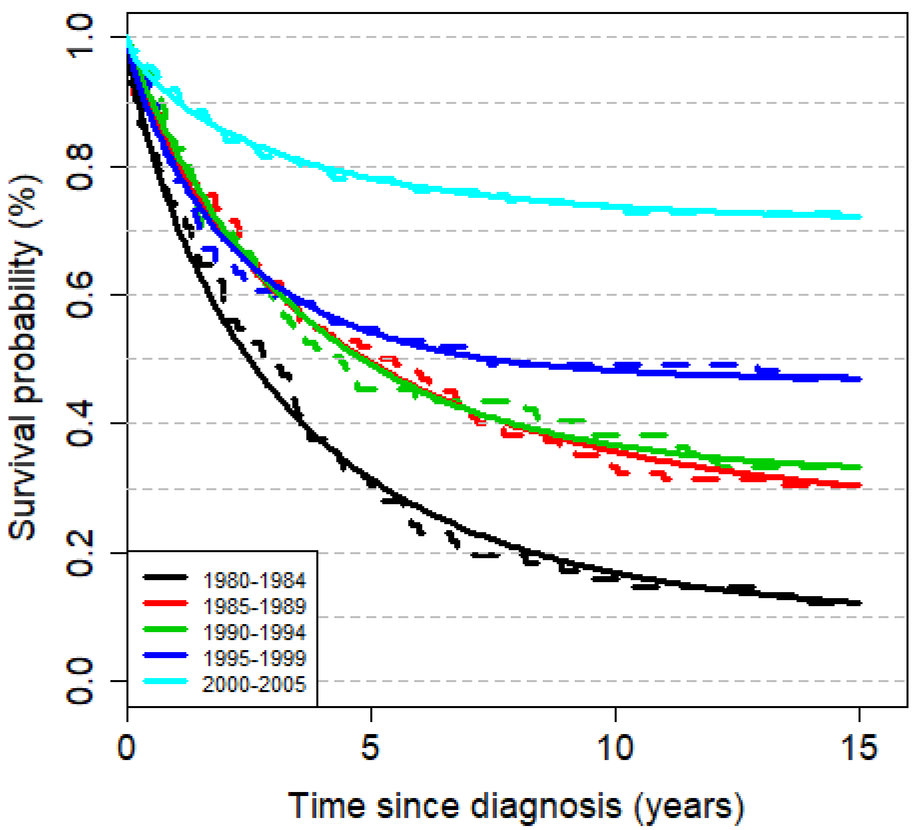

Figure S2 Comparisons of the non-parametric overall survival estimates to the ones obtained from the multivariable mixture cure models (**using the final model**), for pre-specified subgroups defined by age at diagnosis a and period of diagnosis y. The non-parametric estimates were obtained from subgroups of patients aged [a-2; a+2] and diagnosed in the period [y-2; y+2]

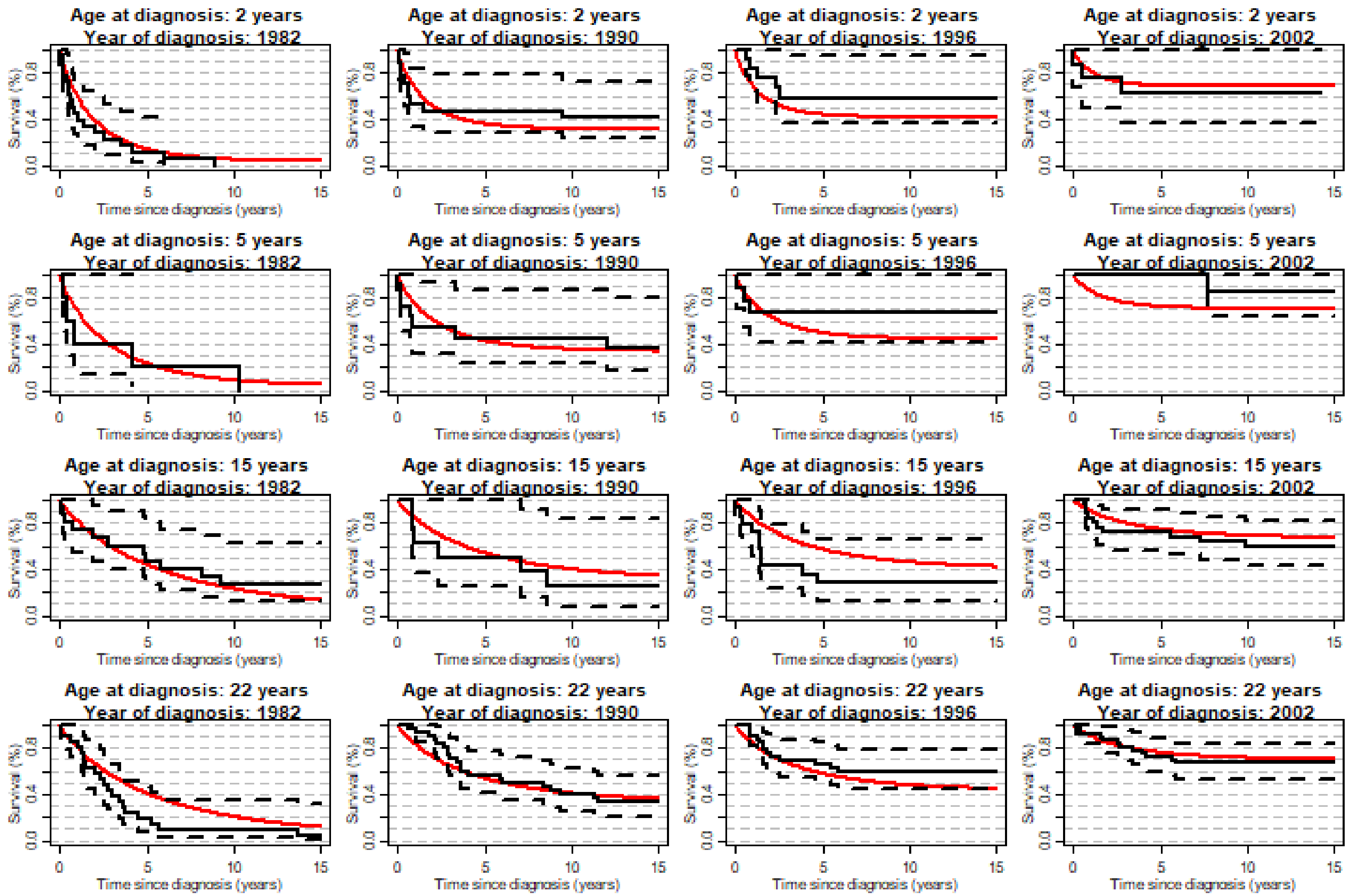

Figure S3 Comparisons of model-based survival of the “uncured” (**using the final model**) to the non-parametric overall survival (once removed from the latter the estimated component due to the “cured”) for pre-specified subgroups defined by age at diagnosis a and period of diagnosis y. The non-parametric estimates were obtained from subgroups of patients aged [a-2; a+2] and diagnosed in the period [y-2; y+2]

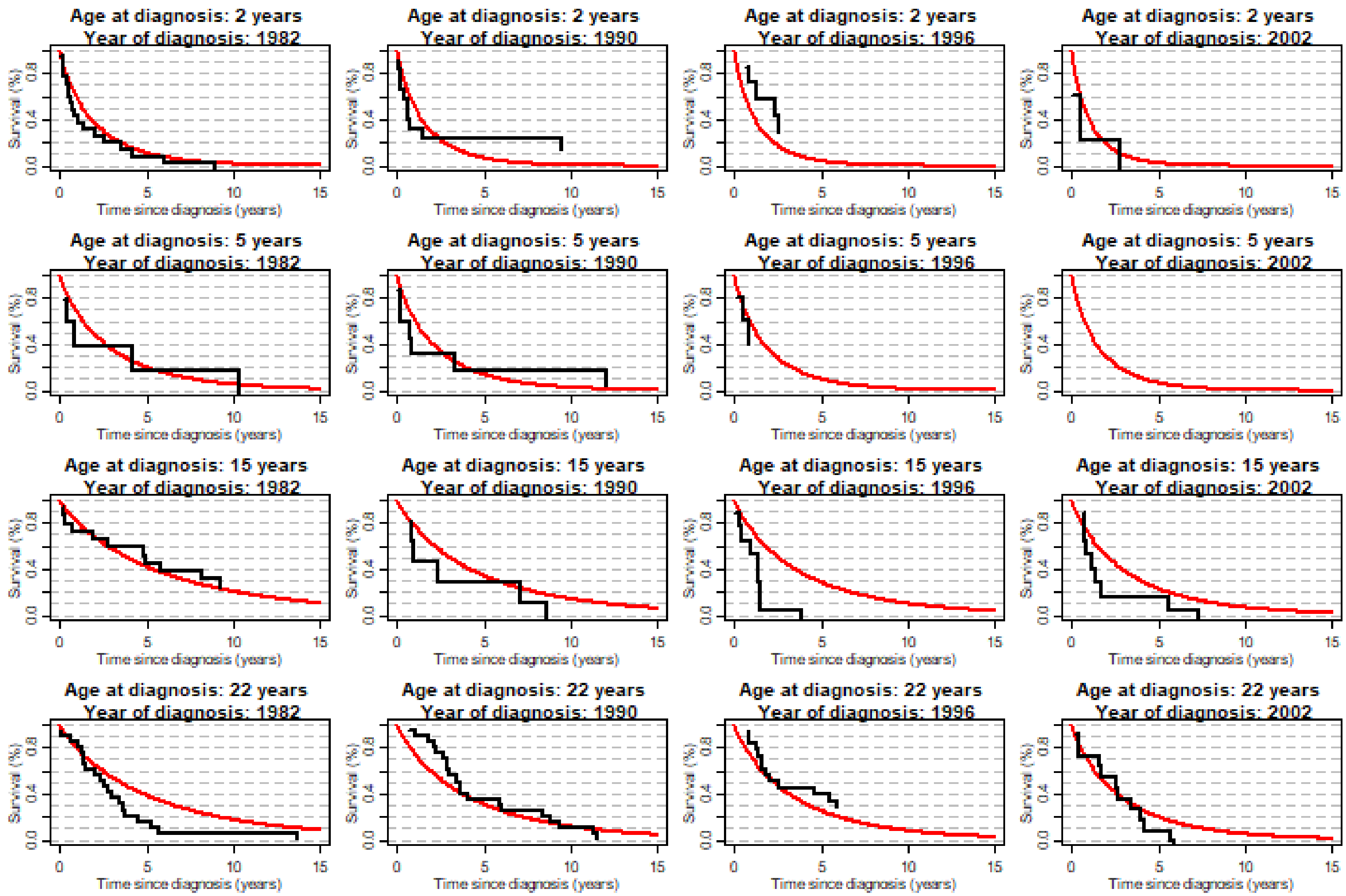

Figure S4 Temporal trends of (i) “cure” proportion (solid line) and of (ii) median survival time (in years) among the “uncured” (dashed line) with their corresponding 95% confidence intervals, estimated with the **most complicated multivariable mixture cure models** (i.e. assuming non-linear functional form for both year of diagnosis and age at diagnosis, on both the survival of the “uncured” and on the “cure” proportion). Children, adolescents and young adults diagnosed with chronic myeloid leukemia in England, 1980-2005

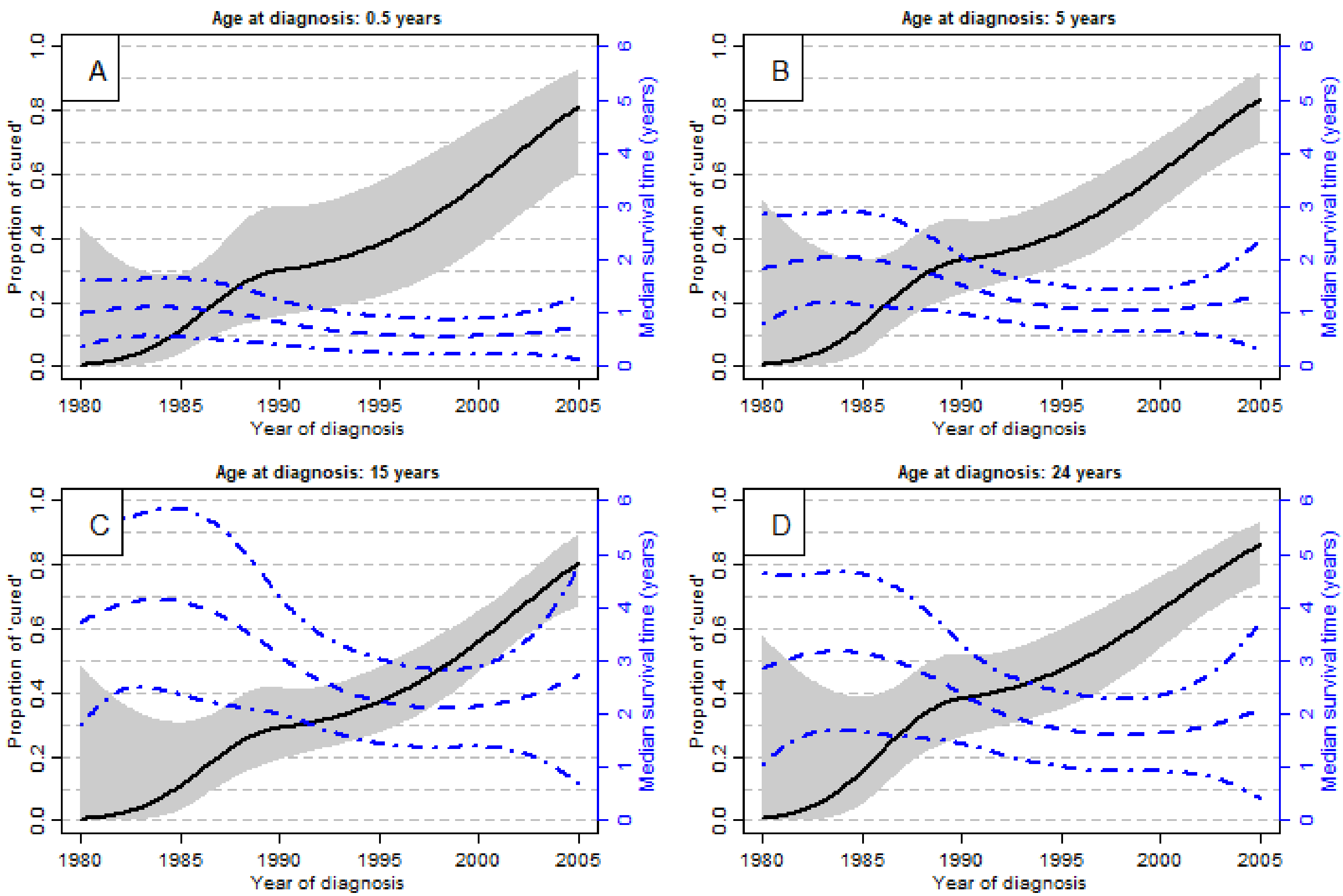

Table S1 Results from the 4 different multivariable mixture cure models: "cure" proportion and median survival time of the "uncured" (in years) with their 95% confidence intervals for specific age at diagnosis and year of diagnosis

		Both Survival of "uncured" and Proportion of "cured" with linear effect of year of diagnosis AIC = 1911.392		Survival of "uncured" with non-linear effect of year, Proportion of "cured" with linear effect of year AIC = 1911.41		Survival of "uncured" with linear effect of year, Proportion of "cured" with non-linear effect of year (final model) AIC = 1910.75		Both Survival of "uncured" and Proportion of "cured" with non-linear effect of year (most complicated model) AIC = 1912.64	
Age at diagnosis (years)	**Year of diagnosis**	**Proportion of "cured" (95% CI)**	**Median survival time [years]of "uncured" (95% CI)**	**Proportion of "cured" (95% CI)**	**Median survival time [years]of "uncured" (95% CI)**	**Proportion of "cured" (95% CI)**	**Median survival time [years]of "uncured" (95% CI)**	**Proportion of "cured" (95% CI)**	**Median survival time [years]of "uncured" (95% CI)**
0.5	1980	0.06 (0.03-0.13)	1.13 (0.58-1.67)	0.06 (0.03-0.13)	0.83 (0.31-1.35)	0.01 (0-0.22)	1.13 (0.58-1.68)	0 (0-0.44)	0.99 (0.37-1.61)
0.5	1990	0.23 (0.13-0.38)	0.87 (0.47-1.27)	0.23 (0.13-0.38)	0.96 (0.47-1.45)	0.29 (0.16-0.48)	0.83 (0.44-1.22)	0.3 (0.16-0.5)	0.83 (0.41-1.24)
0.5	2000	0.57 (0.38-0.74)	0.68 (0.31-1.04)	0.58 (0.38-0.75)	0.63 (0.28-0.98)	0.56 (0.37-0.73)	0.61 (0.27-0.95)	0.57 (0.37-0.75)	0.57 (0.24-0.90)
0.5	2005	0.74 (0.54-0.87)	0.60 (0.23-0.96)	0.74 (0.55-0.87)	0.85 (0.12-1.58)	0.81 (0.61-0.92)	0.52 (0.2-0.85)	0.81 (0.6-0.93)	0.73 (0.12-1.34)
5	1980	0.08 (0.04-0.15)	1.85 (1.13-2.58)	0.08 (0.04-0.15)	1.36 (0.59-2.13)	0.01 (0-0.27)	2 (1.18-2.82)	0.01 (0-0.52)	1.83 (0.80-2.86)
5	1990	0.28 (0.2-0.38)	1.44 (1.01-1.87)	0.28 (0.2-0.37)	1.59 (1.03-2.14)	0.34 (0.24-0.46)	1.47 (1.02-1.93)	0.33 (0.23-0.46)	1.52 (0.99-2.06)
5	2000	0.63 (0.53-0.73)	1.11 (0.71-1.51)	0.64 (0.53-0.73)	1.04 (0.64-1.44)	0.61 (0.49-0.71)	1.09 (0.69-1.48)	0.61 (0.49-0.71)	1.05 (0.64-1.46)
5	2005	0.78 (0.68-0.86)	0.98 (0.55-1.41)	0.79 (0.68-0.87)	1.4 (0.32-2.47)	0.84 (0.71-0.92)	0.93 (0.53-1.34)	0.83 (0.7-0.92)	1.34 (0.30-2.38)
15	1980	0.07 (0.03-0.14)	3.75 (2.26-5.24)	0.06 (0.03-0.14)	2.79 (1.28-4.3)	0.01 (0-0.25)	4.08 (2.45-5.72)	0 (0-0.49)	3.72 (1.78-5.66)
15	1990	0.25 (0.17-0.35)	2.91 (2.07-3.74)	0.24 (0.16-0.34)	3.25 (2.07-4.44)	0.3 (0.2-0.42)	3.01 (2.15-3.87)	0.29 (0.19-0.42)	3.10 (1.99-4.20)
15	2000	0.60 (0.51-0.68)	2.25 (1.49-3.01)	0.59 (0.5-0.68)	2.13 (1.37-2.89)	0.56 (0.47-0.65)	2.21 (1.47-2.95)	0.56 (0.46-0.65)	2.14 (1.38-2.90)
15	2005	0.76 (0.67-0.83)	1.98 (1.17-2.8)	0.76 (0.66-0.83)	2.87 (0.72-5.01)	0.82 (0.69-0.9)	1.9 (1.12-2.68)	0.81 (0.67-0.89)	2.73 (0.69-4.76)
24	1980	0.10 (0.05-0.18)	2.87 (1.62-4.12)	0.10 (0.05-0.18)	2.02 (0.75-3.3)	0.01 (0-0.32)	3.14 (1.73-4.56)	0.01 (0-0.58)	2.84 (1.06-4.63)
24	1990	0.33 (0.24-0.44)	2.23 (1.44-3.01)	0.33 (0.23-0.44)	2.36 (1.47-3.24)	0.39 (0.27-0.52)	2.32 (1.49-3.15)	0.38 (0.26-0.52)	2.37 (1.44-3.30)
24	2000	0.69 (0.58-0.78)	1.73 (1.03-2.42)	0.69 (0.58-0.78)	1.55 (0.87-2.22)	0.65 (0.54-0.76)	1.71 (1.02-2.39)	0.66 (0.54-0.76)	1.64 (0.92-2.35)
24	2005	0.82 (0.72-0.89)	1.52 (0.8-2.24)	0.83 (0.73-0.9)	2.08 (0.43-3.73)	0.87 (0.75-0.93)	1.46 (0.78-2.15)	0.86 (0.74-0.93)	2.09 (0.41-3.76)

AIC: Akaike Information Criterion