**Statistical Appendix**

**1. Model specification**

Relative survival at time t since diagnosis SR(t) is defined as the ratio of the survival observed in the patients’ group SO(t) to that expected in the absence of cancer SE(t). The relative survival approach further assumes SE(t)=S\*(t), where the latter is the observed survival in a general population group comparable for age and sex. The standard mixture cure model divides patients into two subgroups: those cured, who will not die of the diagnosed cancer, and the uncured, who risk dying from their cancers and/or other causes. The cumulative relative survival of the whole group of patients is then expressed as:

 SO(t)/S\*(t) = SR(t) =  + (1-) SRc(t) (Model-1)

where is the proportion of cured, and SRc is the cumulative relative survival function for uncured patients. The relative survival of the cured is by definition equal to 1, because their expected mortality rates are assumed to be equal to those of the general population.

Cure Model-2 [9] relaxes this last condition assuming SE(t) = S\*(t), where the parameter  expresses the relative risk of death from other diseases to which patients are exposed, assumed constant with time and equally affecting cured and uncured patients. The cure model can then be rewritten as:

 SO(t)/S\*(t) = SR(t) =  S\*(t)-1 + (1-) SRc(t) S\*(t)-1 (Model-2)

In this analysis we assumed the proportion of cured to be dependent on age at diagnosis, expressed as a categorical variable by five-year age classes. We took SRc(t) to be a Weibull parametric function, also dependent on age at diagnosis, and used the logistic for the proportion of cured to assure  remained bound within 0 and 1. Model-2 was then specified by:

RS(t,Z) = **** S\*(t)-1 + (1-****) [exp(-t)] exp(**** , where:

****exp(-) -exp(****

Model-2, with the constraint  ≡ 1 gives the standard cure Model-1 with the same parametrization of age effects.

**2. Goodness of fit: comparison of Model-2 and the standard cure Model-1**

Fit was assessed by analysis of the residual sum of squares (RSS) and by Q-Q plots to check their normality. Model (2) had a significantly better fit of the data, and residuals closer to the normal than the standard model (1), for all the cancers.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| Analysis of variance |   | RSSA | dfB | Fc |
| Colon & rectum; male | Model-1 | 0.00635688 | 120 | 57.9 |
| Model (2) | 0.00427529 | 119 |
| Colon & rectum; female | Model-1 | 0.00562311 | 120 | 67.6 |
| Model (2) | 0.00358518 | 119 |
| Breast, female | Model-1 | 0.01219606 | 120 | 94.4 |
| Model (2) | 0.00680140 | 119 |
| Lung, male | Model-1 | 0.01571201 | 120 | 66.4 |
| Model (2) | 0.01008248 | 119 |
| Lung, female | Model-1 | 0.00367087 | 120 | 57.6 |
| Model (2) | 0.00247411 | 119 |
| RSS = Residual sum of squares |  |  |  |  |
| Df = Degrees of freedom F = Fischer-Snedecor test comparing nested models  |  |  |  |

**3. Q-Q plots: Quantiles of residuals (y-axis) against quantiles of normal distribution(x-axis)**

 S1: Model-1: Q-Q plot for colon & rectum, male S2: Model-2: Q-Q plot for colon & rectum, male



 S3: Model-1: Q-Q plot for colon & rectum, female S4: Model-2: Q-Q plot for colon & rectum, female



 S5: Model-1: Q-Q plot for breast, female S6: Model-2: Q-Q plot for breast, female



 S7: Model-1: Q-Q plot for lung, male S8: Model-2: Q-Q plot for lung, male



 S9: Model-1: Q-Q plot for lung, female S10: Model-2: Q-Q plot for lung, female