

Appendix for the paper entitled “Summarising and communicating on survival data according to the audience: A tutorial on different measures illustrated with population-based cancer registry data”

Results of the different measures by age-group

Authors : Aurélien Belot, Aminata Ndiaye, Miguel-Angel Luque-Fernandez, Dimitra-Kleio Kipourou, Camille Maringe, Francisco Javier Rubio, Bernard Rachet

Cancer Survival Group, Department of Non-Communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom.

Table S1: 10-year overall survival in men by deprivation and age group

	Dep 1	Dep 2	Dep 3	Dep 4	Dep 5
[15;44]	0.53 (0.45 - 0.63)	0.53 (0.44 - 0.63)	0.5 (0.42 - 0.6)	0.55 (0.47 - 0.65)	0.46 (0.38 - 0.56)
[45;54]	0.49 (0.44 - 0.55)	0.46 (0.41 - 0.51)	0.49 (0.44 - 0.55)	0.4 (0.34 - 0.46)	0.43 (0.38 - 0.49)
[55;64]	0.45 (0.42 - 0.48)	0.43 (0.4 - 0.46)	0.42 (0.39 - 0.45)	0.4 (0.37 - 0.43)	0.33 (0.3 - 0.36)
[65;74]	0.35 (0.33 - 0.37)	0.33 (0.31 - 0.35)	0.33 (0.31 - 0.36)	0.29 (0.27 - 0.31)	0.23 (0.21 - 0.25)
[75;80]	0.21 (0.18 - 0.24)	0.19 (0.16 - 0.21)	0.2 (0.18 - 0.23)	0.18 (0.15 - 0.2)	0.13 (0.11 - 0.16)

Table S2: 5-year Conditional Survival given that the patients already survived 5 years in men by deprivation and age group

	Dep 1	Dep 2	Dep 3	Dep 4	Dep 5
[15;44]	0.96 (0.91 - 1)	0.9 (0.82 - 0.97)	0.94 (0.88 - 1)	0.91 (0.84 - 0.97)	0.91 (0.84 - 0.98)
[45;54]	0.9 (0.86 - 0.95)	0.94 (0.9 - 0.97)	0.92 (0.87 - 0.96)	0.85 (0.79 - 0.91)	0.87 (0.81 - 0.92)
[55;64]	0.84 (0.81 - 0.87)	0.84 (0.81 - 0.88)	0.83 (0.8 - 0.87)	0.8 (0.77 - 0.84)	0.78 (0.74 - 0.83)
[65;74]	0.73 (0.7 - 0.76)	0.73 (0.69 - 0.76)	0.73 (0.7 - 0.77)	0.7 (0.67 - 0.74)	0.65 (0.61 - 0.69)
[75;80]	0.58 (0.52 - 0.63)	0.54 (0.48 - 0.59)	0.58 (0.53 - 0.63)	0.58 (0.53 - 0.63)	0.46 (0.4 - 0.52)

Table S3: 10-year Restricted Mean Survival Time in men by deprivation and age group

	Dep 1	Dep 2	Dep 3	Dep 4	Dep 5
[15;44]	6.06 (5.29 - 6.84)	6.34 (5.59 - 7.1)	5.91 (5.13 - 6.69)	6.51 (5.78 - 7.25)	5.59 (4.84 - 6.35)
[45;54]	5.99 (5.54 - 6.45)	5.56 (5.1 - 6.03)	5.77 (5.27 - 6.28)	5.3 (4.82 - 5.79)	5.49 (4.98 - 6)
[55;64]	5.76 (5.5 - 6.02)	5.62 (5.36 - 5.89)	5.53 (5.26 - 5.81)	5.37 (5.09 - 5.64)	4.73 (4.43 - 5.03)
[65;74]	5.15 (4.95 - 5.34)	4.94 (4.74 - 5.13)	4.98 (4.78 - 5.18)	4.55 (4.35 - 4.74)	4.04 (3.83 - 4.24)
[75;80]	4 (3.75 - 4.26)	3.9 (3.67 - 4.14)	3.96 (3.72 - 4.2)	3.55 (3.32 - 3.77)	3.22 (2.97 - 3.47)

Table S4: 10-year net survival in men by deprivation and age group

	Dep 1	Dep 2	Dep 3	Dep 4	Dep 5
[15;44]	0.54 (0.46 - 0.64)	0.51 (0.46 - 0.57)	0.49 (0.46 - 0.53)	0.48 (0.45 - 0.51)	0.41 (0.36 - 0.47)
[45;54]	0.53 (0.45 - 0.63)	0.48 (0.42 - 0.54)	0.49 (0.45 - 0.52)	0.47 (0.44 - 0.5)	0.41 (0.36 - 0.47)
[55;64]	0.51 (0.43 - 0.61)	0.52 (0.46 - 0.58)	0.49 (0.45 - 0.53)	0.51 (0.47 - 0.54)	0.49 (0.43 - 0.55)
[65;74]	0.56 (0.48 - 0.66)	0.43 (0.37 - 0.49)	0.48 (0.44 - 0.52)	0.46 (0.43 - 0.5)	0.44 (0.39 - 0.5)
[75;80]	0.47 (0.39 - 0.57)	0.48 (0.42 - 0.55)	0.42 (0.38 - 0.46)	0.4 (0.37 - 0.44)	0.36 (0.3 - 0.43)

Table S5: 5-year Conditional Net Survival given that the patients already survived 5 years in men by deprivation and age group

	Dep 1	Dep 2	Dep 3	Dep 4	Dep 5
[15;44]	0.96 (0.9 - 1.01)	0.9 (0.82 - 0.96)	0.95 (0.88 - 0.99)	0.92 (0.85 - 0.97)	0.92 (0.84 - 0.98)
[45;54]	0.92 (0.88 - 0.96)	0.96 (0.92 - 0.99)	0.95 (0.9 - 0.99)	0.89 (0.82 - 0.95)	0.92 (0.86 - 0.97)
[55;64]	0.89 (0.86 - 0.92)	0.91 (0.87 - 0.94)	0.91 (0.87 - 0.95)	0.9 (0.86 - 0.94)	0.9 (0.85 - 0.95)
[65;74]	0.89 (0.85 - 0.93)	0.89 (0.86 - 0.93)	0.94 (0.89 - 0.98)	0.93 (0.88 - 0.97)	0.89 (0.84 - 0.95)
[75;80]	0.87 (0.8 - 0.96)	0.86 (0.77 - 0.95)	0.97 (0.88 - 1.05)	0.98 (0.9 - 1.08)	0.82 (0.72 - 0.95)

Table S6: 10-year Restricted Mean Net Survival time in men by deprivation and age group

	Dep 1	Dep 2	Dep 3	Dep 4	Dep 5
[15;44]	6.09 (5.28 - 6.93)	6.38 (5.52 - 7.13)	5.95 (5.19 - 6.68)	6.57 (5.8 - 7.3)	5.66 (4.86 - 6.33)
[45;54]	6.08 (5.65 - 6.52)	5.65 (5.14 - 6.11)	5.9 (5.41 - 6.47)	5.45 (4.99 - 6)	5.71 (5.14 - 6.17)
[55;64]	5.99 (5.71 - 6.28)	5.9 (5.63 - 6.14)	5.88 (5.59 - 6.17)	5.78 (5.48 - 6.09)	5.2 (4.88 - 5.53)
[65;74]	5.81 (5.6 - 6.05)	5.65 (5.43 - 5.88)	5.85 (5.59 - 6.09)	5.44 (5.19 - 5.69)	4.96 (4.69 - 5.23)
[75;80]	5.14 (4.73 - 5.5)	5.16 (4.83 - 5.5)	5.43 (5.03 - 5.82)	4.95 (4.61 - 5.31)	4.61 (4.21 - 5.02)

Table S7: 10-year crude probability of death due to cancer and due to other causes in men, by deprivation and age group

		Dep 1	Dep 2	Dep 3	Dep 4	Dep 5
[15;44]	Cancer	0.46 (0.37 - 0.55)	0.47 (0.37 - 0.56)	0.49 (0.4 - 0.58)	0.43 (0.34 - 0.52)	0.52 (0.43 - 0.61)
	Other	0.01 (0.01 - 0.01)	0.01 (0.01 - 0.01)	0.01 (0.01 - 0.01)	0.02 (0.01 - 0.02)	0.02 (0.02 - 0.02)
[45;54]	Cancer	0.48 (0.43 - 0.54)	0.52 (0.46 - 0.58)	0.48 (0.42 - 0.54)	0.56 (0.5 - 0.62)	0.51 (0.45 - 0.58)
	Other	0.02 (0.02 - 0.02)	0.02 (0.02 - 0.03)	0.03 (0.03 - 0.04)	0.04 (0.03 - 0.04)	0.05 (0.05 - 0.06)
[55;64]	Cancer	0.5 (0.46 - 0.53)	0.5 (0.47 - 0.54)	0.5 (0.46 - 0.53)	0.51 (0.47 - 0.54)	0.56 (0.52 - 0.6)
	Other	0.06 (0.05 - 0.06)	0.07 (0.06 - 0.07)	0.08 (0.08 - 0.09)	0.1 (0.09 - 0.1)	0.11 (0.1 - 0.12)
[65;74]	Cancer	0.5 (0.47 - 0.53)	0.51 (0.48 - 0.53)	0.47 (0.44 - 0.5)	0.51 (0.48 - 0.54)	0.56 (0.53 - 0.59)
	Other	0.15 (0.15 - 0.16)	0.17 (0.16 - 0.17)	0.2 (0.19 - 0.21)	0.2 (0.19 - 0.21)	0.21 (0.2 - 0.22)
[75;80]	Cancer	0.53 (0.49 - 0.58)	0.53 (0.49 - 0.57)	0.48 (0.44 - 0.52)	0.52 (0.48 - 0.56)	0.57 (0.52 - 0.61)
	Other	0.26 (0.24 - 0.28)	0.28 (0.26 - 0.3)	0.32 (0.29 - 0.34)	0.3 (0.28 - 0.32)	0.3 (0.28 - 0.33)

Table S8: 10-year Number of Life Years Lost due to cancer and due to other causes in men, by deprivation and age group

		Dep 1	Dep 2	Dep 3	Dep 4	Dep 5
[15;44]	Cancer	3.9 (3.21 - 4.69)	3.61 (2.89 - 4.33)	4.03 (3.22 - 4.9)	3.41 (2.67 - 4.14)	4.32 (3.48 - 5.11)
	Other	0.03 (0.03 - 0.04)	0.04 (0.04 - 0.05)	0.06 (0.05 - 0.06)	0.08 (0.07 - 0.08)	0.09 (0.07 - 0.1)
[45;54]	Cancer	3.9 (3.47 - 4.38)	4.32 (3.84 - 4.83)	4.07 (3.61 - 4.61)	4.5 (3.97 - 5.02)	4.24 (3.66 - 4.88)
	Other	0.11 (0.1 - 0.12)	0.12 (0.11 - 0.13)	0.16 (0.14 - 0.17)	0.2 (0.18 - 0.21)	0.27 (0.25 - 0.3)
[55;64]	Cancer	3.96 (3.72 - 4.25)	4.04 (3.75 - 4.28)	4.06 (3.78 - 4.37)	4.13 (3.84 - 4.45)	4.69 (4.36 - 5.01)
	Other	0.28 (0.27 - 0.3)	0.34 (0.32 - 0.36)	0.41 (0.39 - 0.43)	0.5 (0.48 - 0.53)	0.58 (0.55 - 0.62)
[65;74]	Cancer	4.07 (3.81 - 4.3)	4.2 (3.95 - 4.41)	4 (3.76 - 4.25)	4.38 (4.14 - 4.62)	4.79 (4.53 - 5.06)
	Other	0.79 (0.76 - 0.82)	0.86 (0.83 - 0.9)	1.02 (0.98 - 1.06)	1.08 (1.03 - 1.12)	1.18 (1.12 - 1.24)
[75;80]	Cancer	4.6 (4.27 - 4.93)	4.55 (4.22 - 4.85)	4.33 (4.01 - 4.68)	4.74 (4.39 - 5.08)	5.03 (4.64 - 5.41)
	Other	1.4 (1.32 - 1.48)	1.55 (1.47 - 1.63)	1.71 (1.62 - 1.8)	1.72 (1.62 - 1.82)	1.75 (1.62 - 1.88)

Figure S1: 10-year overall survival in men by deprivation and age group

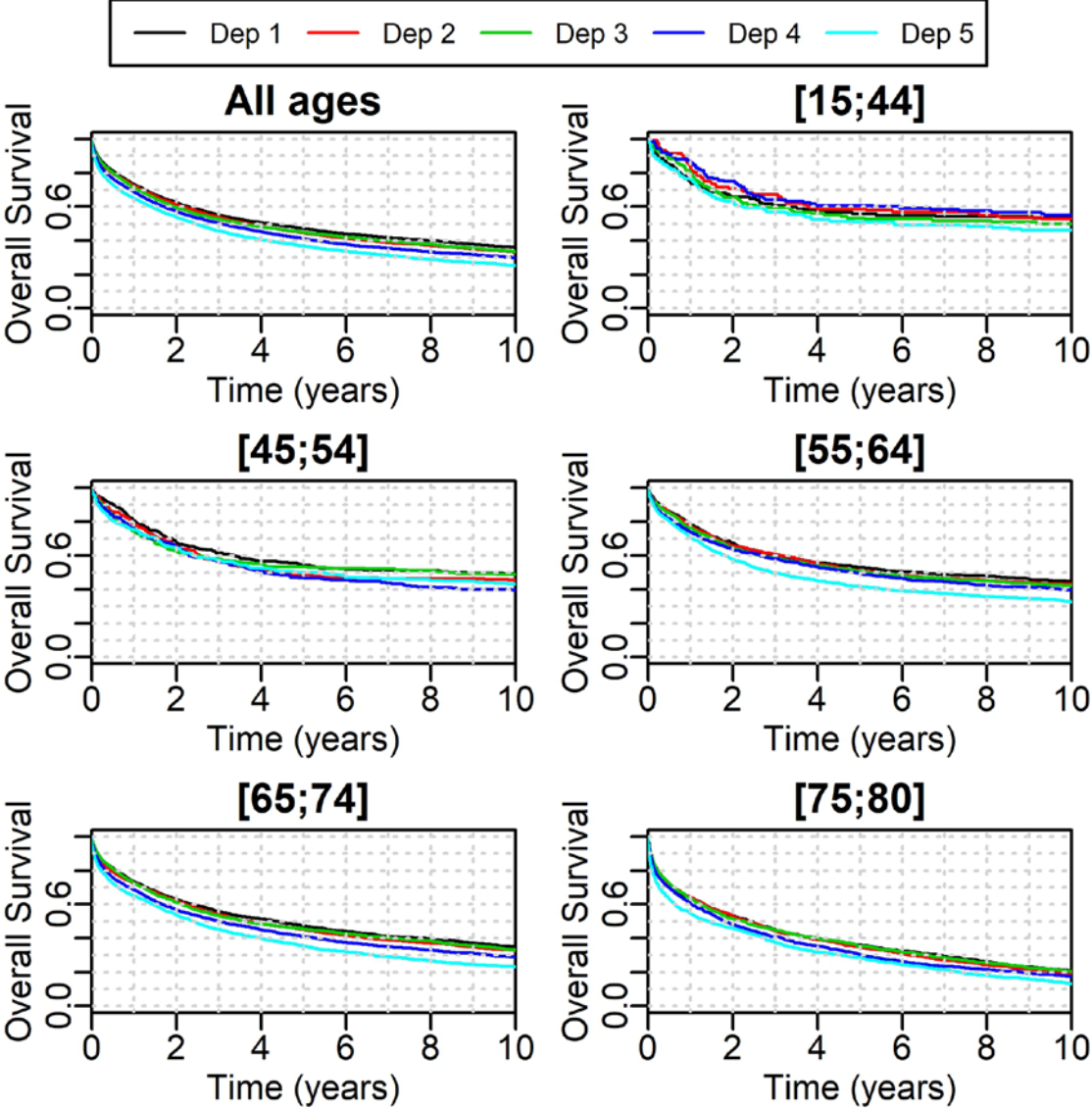


Figure S2: 5-year Conditional Survival in men by deprivation and age group, given that patients survived y years (y varying from 1 to 5 on the x-axis). Notice that the points are not all aligned on the exact year y to make the graph more readable.

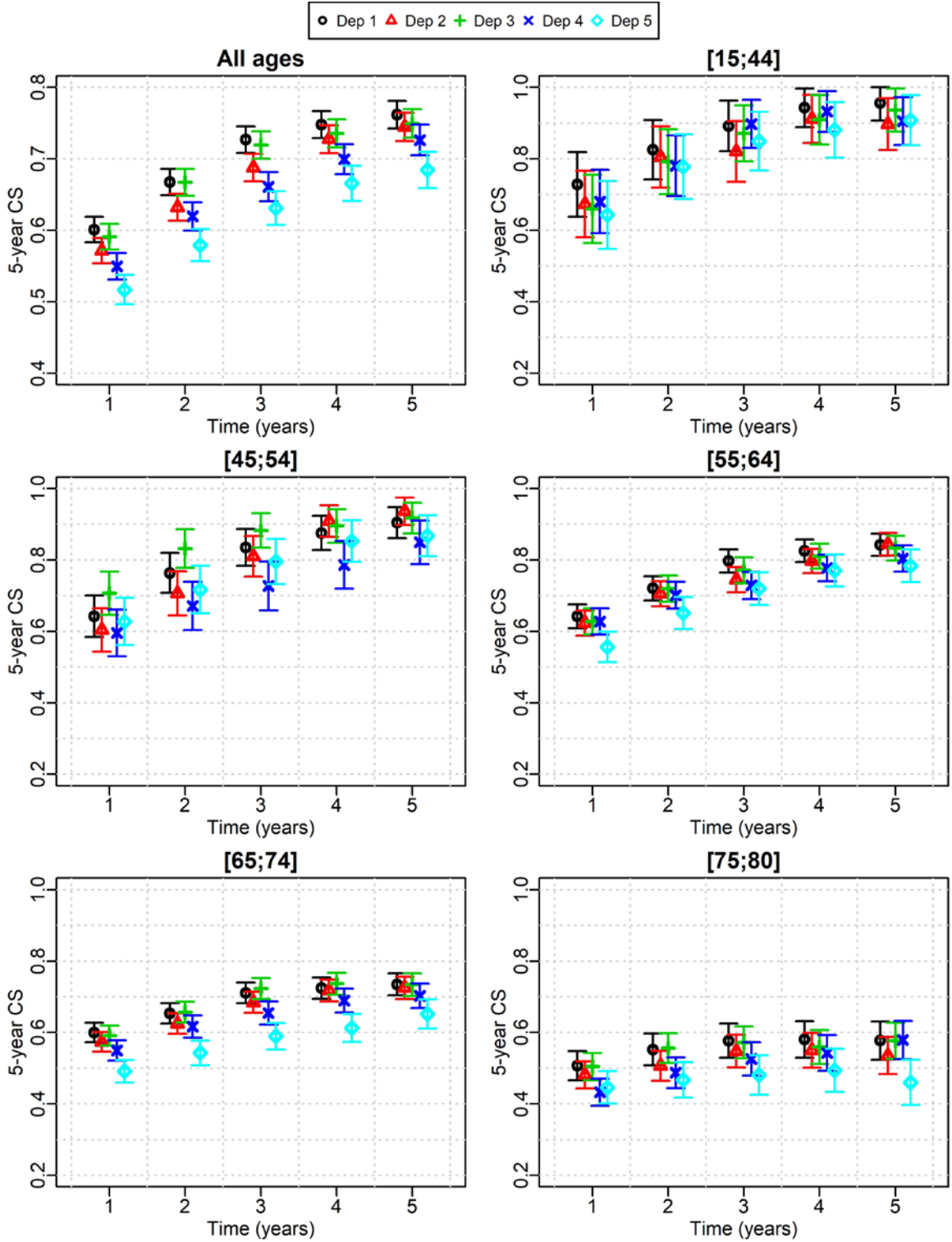


Figure S3: 10-year Restricted Mean Survival Time in men by deprivation and age group

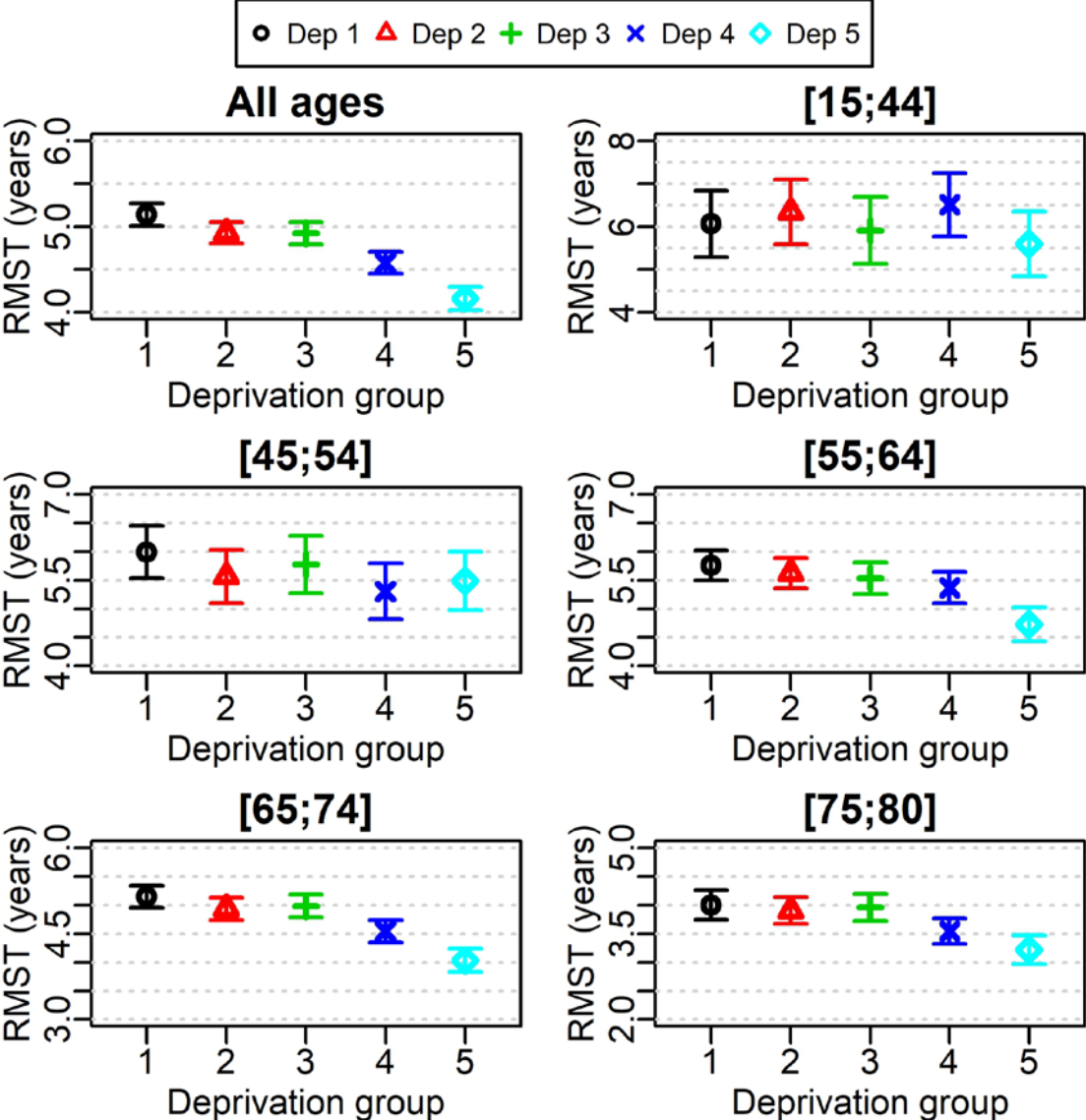


Figure S4: 10-year Net Survival for all men and by age-group according to the deprivation group.

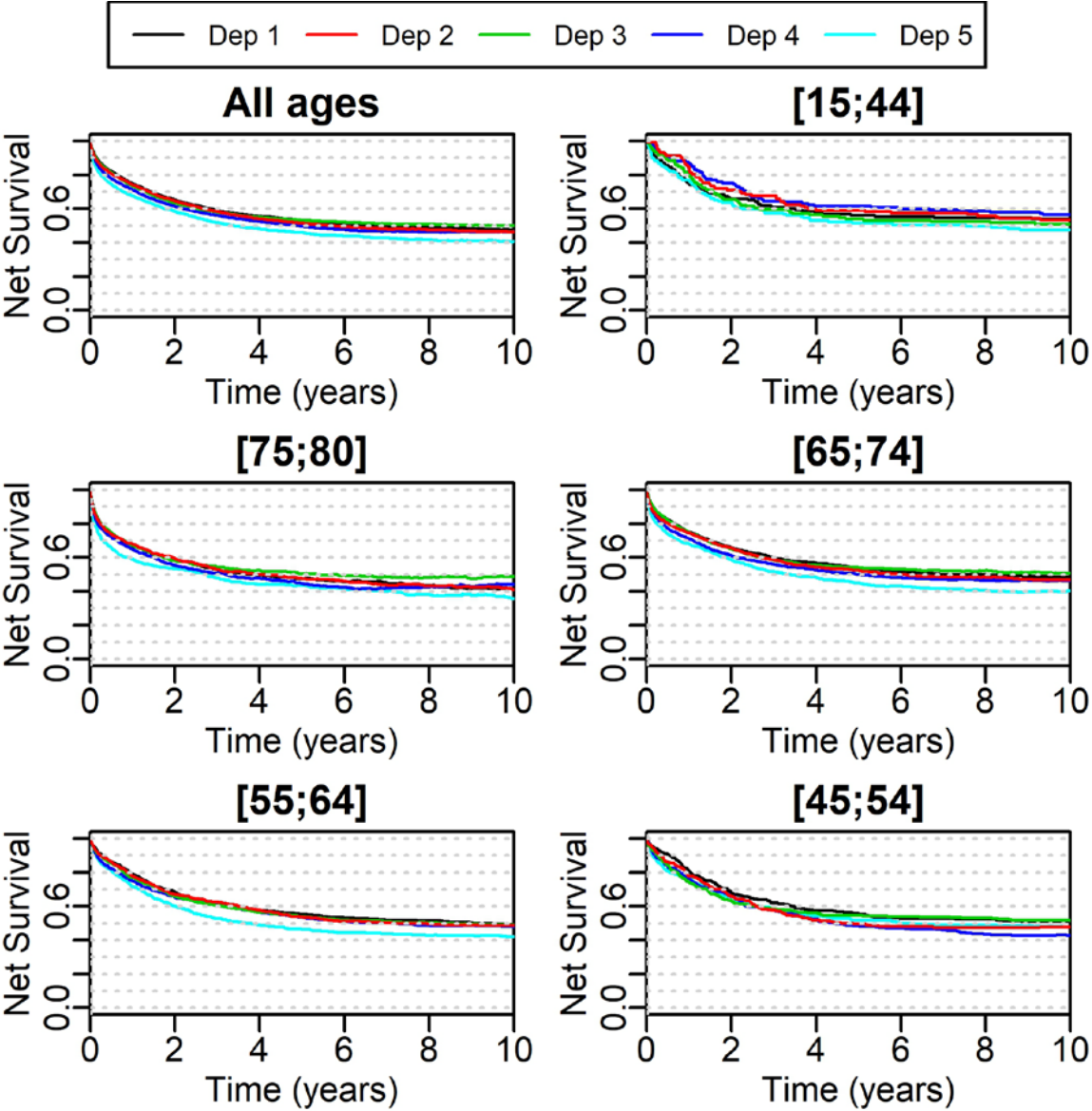


Figure S5: 5-year Conditional Net Survival for men by deprivation and age group, given that they survived y years (y being varied from 1 to 5 on the x-axis). Notice that the points are not all aligned on the exact year y to make the graph more readable.

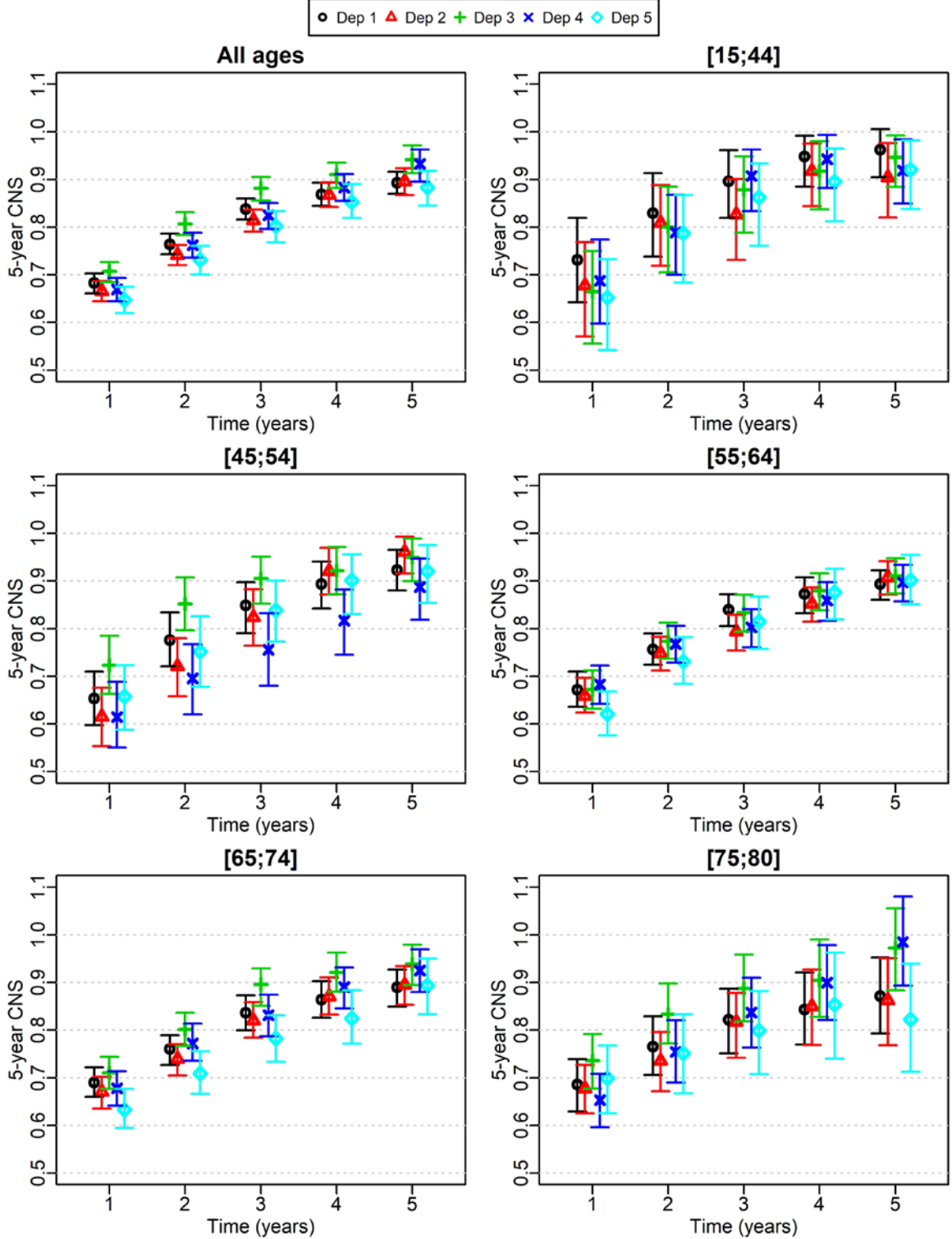


Figure S6: 10-year Restricted Mean Net Survival time in men by deprivation and age group

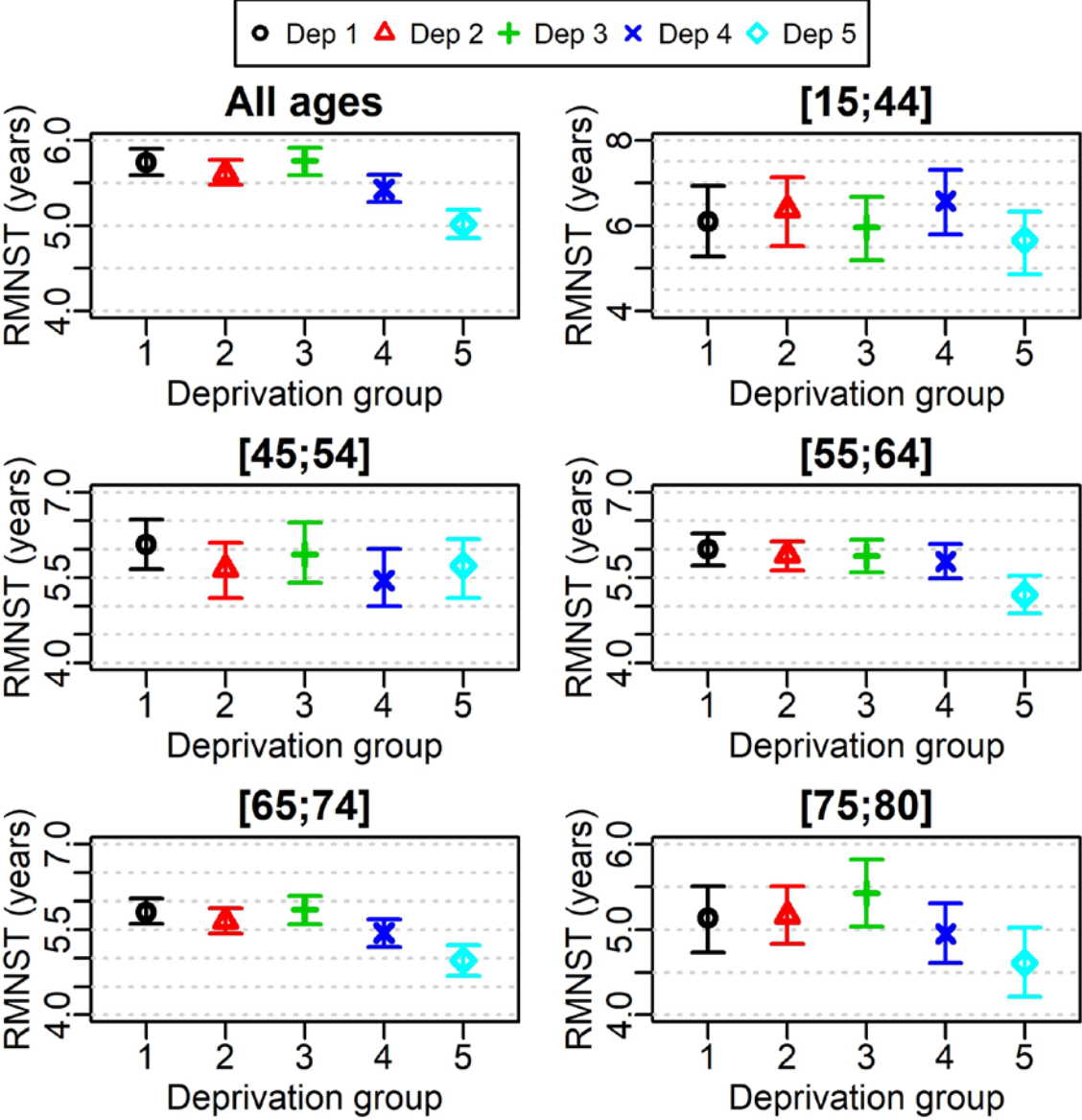


Figure S7: Crude probability of death from cancer and from other causes for the least and the most deprived groups, for all men and by age-group. The two upper curves are those of death from cancer (solid for deprivation 1 group and dashed for deprivation 5 group), while the two lower are death due to other causes (solid for deprivation 1 group and dashed for deprivation 5 group).

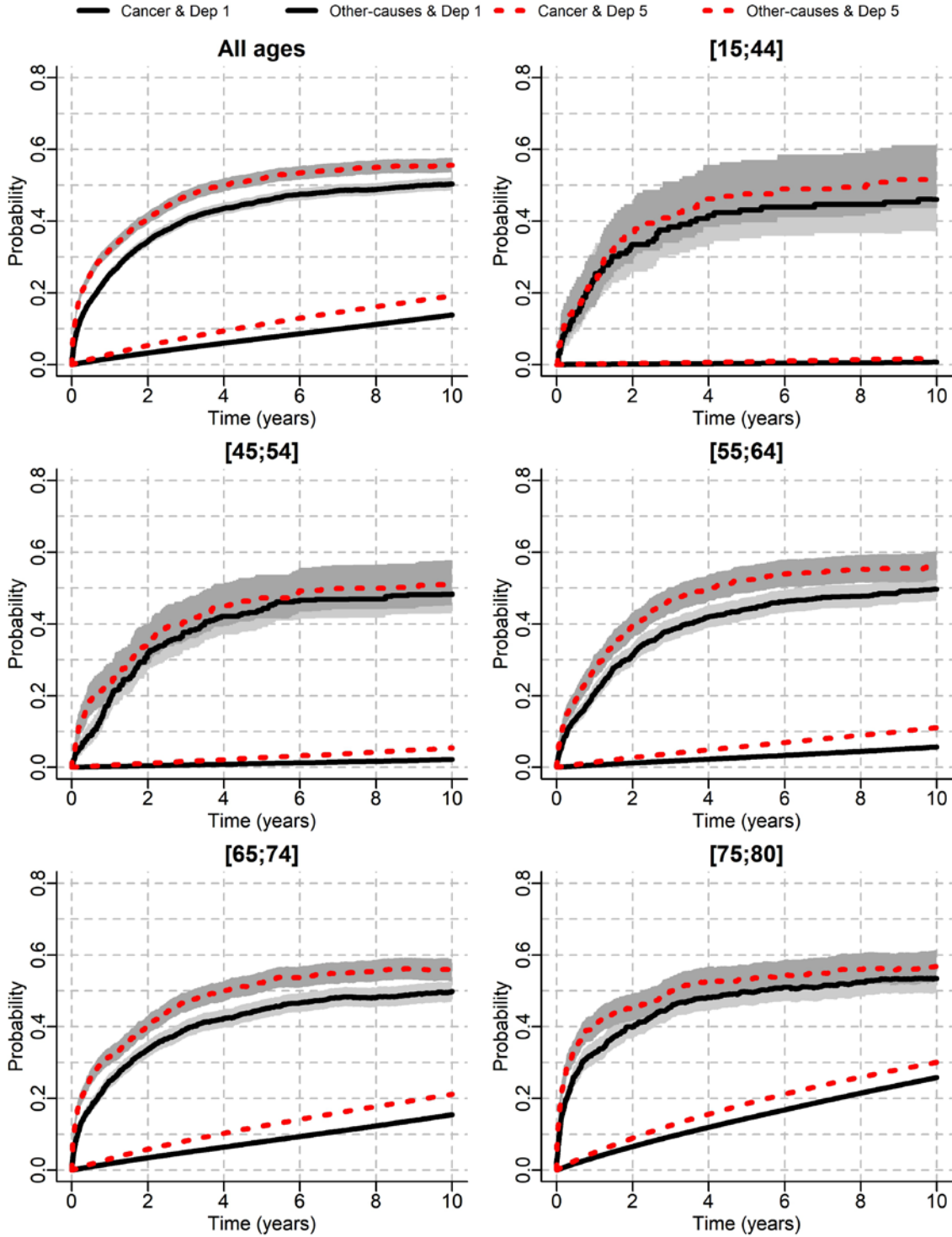
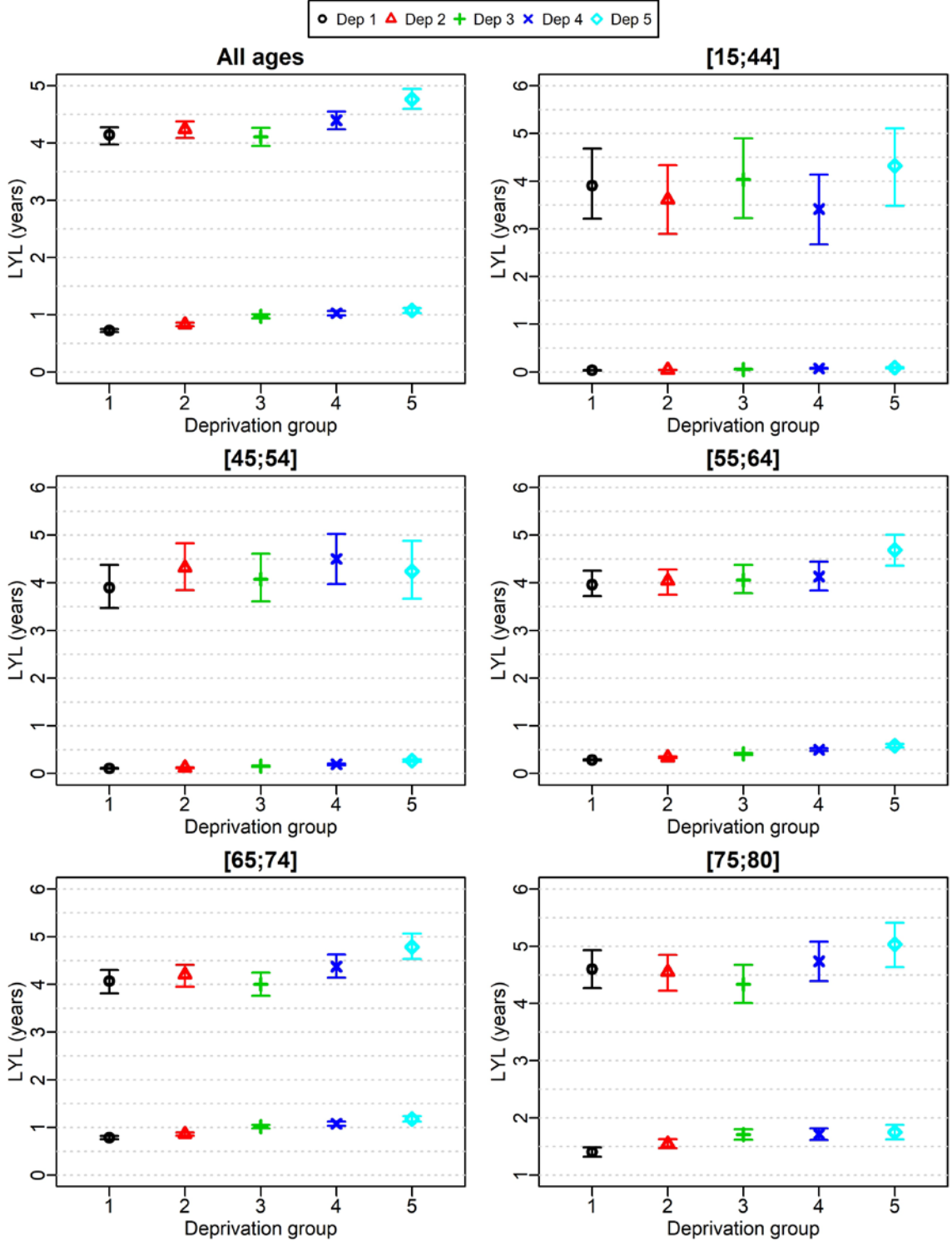


Figure S8: 10-year Number of Life Years Lost (LYL) due to cancer and to other causes in men, by deprivation and age group. For a given deprivation group, the higher point corresponds to the LYL due to cancer, the lower to the LYL due to other causes



Appendix 2: Definitions of the different measures and details of the R-code for their estimation

From the article entitled Summarising and communicating on survival data according to the audience: A tutorial on different measures illustrated with population-based cancer registry data.

Aurélien Belot

Aminata Ndiaye

Miguel-Angel Luque-Fernandez

Dimitra-Kleio Kipourou

Camille Maringe

Francisco Javier Rubio

Bernard Rachet

18 September 2018

In this appendix, we detail in the first part how we can derive the different measures from a theoretical point of view, and in the second part we illustrate how those measures can be estimated with the free R software using simulated data. Notice that the data used here are NOT the data used in the application in the main paper.

Definition of the different measures

Overall survival setting

We consider a time-to-event random variable T , with a cumulative distribution function $F(t)$:

$$F(t) = \mathbb{P}(T \leq t) = \int_0^t f(u)du$$

where $f(t)$ is the probability density function of T .

If the event studied is death, then $F(t)$ is the cumulative probability of death between 0 and t , and the **survival probability** is

$$S(t) = \mathbb{P}(T > t) = 1 - F(t)$$

To define equivalently the time-to-event distribution, we could also use the hazard function, which is the instantaneous rate of failure:

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{P}(t \leq T < t + dt \mid T \geq t)}{dt}$$

From the definition of the hazard, we can derive some useful relationships:

$$\begin{aligned} \lambda(t) &= \lim_{dt \rightarrow 0} \frac{\mathbb{P}(t \leq T < t + dt \mid T \geq t)}{dt} \\ &= \left(\lim_{dt \rightarrow 0} \frac{\mathbb{P}(t \leq T < t + dt)}{dt} \right) \frac{1}{S(t)} \\ &= \frac{dF(t)}{dt} \frac{1}{S(t)} = \frac{f(t)}{S(t)} \\ &= -\frac{dS(t)}{dt} \frac{1}{S(t)} \end{aligned}$$

From this last equality, we have

$$\int_0^t \lambda(u)du = \Lambda(t) = -\log(S(t)) \iff S(t) = \exp\left(-\int_0^t \lambda(u)du\right)$$

Moreover, we notice that we can write the cumulative probability of death between 0 and t , $F(t)$, as

$$F(t) = \int_0^t \lambda(u)S(u)du$$

Conditional Survival

The **Conditional Survival** $CS(t \mid s)$ is the probability to survive after time $s + t$ given that the patient has survived after time s :

$$CS(t \mid s) = \mathbb{P}(T > s + t \mid T > s) = \frac{\mathbb{P}(T > s + t)}{\mathbb{P}(T > s)} = \frac{S(s + t)}{S(s)}$$

Mean Survival Time and Restricted Mean Survival Time

The **Mean Survival Time** is defined as $MST = E[T] = \int_0^\infty uf(u)du$, which simplifies (after integrating by parts using $f(t) = -S'(t)$, and given that $\lim_{u \rightarrow +\infty} S(u) = 0$) as

$$MST = E[T] = \int_0^\infty S(u)du$$

The **τ -Restricted Mean Survival Time** is defined as

$$RMST(\tau) = E[\min(T, \tau)] = \int_0^\tau S(u)du$$

From the above definition of the $RMST(\tau)$, we can see that it corresponds to the area under the survival curve between 0 and τ .

On the individual and population (marginal) mortality hazards and survival functions

Consider a population of N heterogeneous patients, each patient $i = 1, \dots, N$ being exposed to a mortality hazard $\lambda_{O_i}(t)$. The **Survival function for one individual** is defined as $S_i(t) = \exp(-\int_0^t \lambda_{O_i}(u)du)$, and the **overall survival on the whole population** (i.e. the marginal survival) could be obtained from the individual survivals:

$$S(t) = \frac{1}{N} \sum_{i=1}^N S_i(t)$$

From this last equation, we could derive the **marginal hazard** $\lambda(t)$ as a function of the individual hazards $\lambda_{O_i}(t)$.

Indeed, starting from

$$\Lambda(t) = -\log(S(t)) = -\log\left(\frac{1}{N} \sum_{i=1}^N S_i(t)\right) \quad (1)$$

we obtain

$$\lambda(t) = \frac{d\Lambda(t)}{dt} = -\frac{1/N \sum_{i=1}^N \frac{dS_i(t)}{dt}}{1/N \sum_{i=1}^N S_i(t)} = \frac{\sum_{i=1}^N \lambda_{O_i}(t) S_i(t)}{\sum_{i=1}^N S_i(t)} \quad (2)$$

Thus we can write the **marginal hazard** $\lambda(t)$ as a weighted average of the N **individual hazards** $\lambda_{O_i}(t)$:

$$\lambda(t) = \sum_{i=1}^N \frac{S_i(t)}{\sum_{i=1}^N S_i(t)} \lambda_{O_i}(t) \quad (3)$$

Relative survival setting

Definition of excess mortality hazard

For a patient i , we decompose the observed mortality hazard $\lambda_{O_i}(t)$ as the sum of an excess mortality hazard $\lambda_{E_i}(t)$ and an expected mortality hazard (usually obtained from the general population lifetable) $\lambda_{P_i}(t)$:

$$\lambda_{O_i}(t) = \lambda_{E_i}(t) + \lambda_{P_i}(t) \quad (4)$$

$$\lambda_{E_i}(t) = \lambda_{O_i}(t) - \lambda_{P_i}(t) \quad (5)$$

The excess mortality hazard $\lambda_{E_i}(t)$ is the hazard associated with the individual net survival:

$$S_{N_i}(t) = \exp\left(-\int_0^t \lambda_{E_i}(u) du\right) \quad (6)$$

On the individual and population (marginal) mortality hazards and survival functions

By plugging in the hazard decomposition defined at individual level (equation 4) into equation 3, we decompose the overall mortality hazard (defined at population level) as the sum of a cancer-specific mortality hazard and an expected mortality hazard:

$$\lambda(t) = \lambda_C(t) + \lambda_P(t) \quad (7)$$

where

$$\lambda_C(t) = \sum_{i=1}^N \frac{S_i(t)}{\sum_{i=1}^N S_i(t)} \lambda_{E_i}(t)$$

and

$$\lambda_P(t) = \sum_{i=1}^N \frac{S_i(t)}{\sum_{i=1}^N S_i(t)} \lambda_{P_i}(t)$$

One important thing to notice is that $\lambda_C(t)$ **is not the net hazard associated to the net survival**, because it depends on the population mortality through the overall survival (acting as weights in the formula above).

The marginal net hazard $\lambda_N(t)$ could be derived from the marginal net survival following the same reasoning done in equations 1 and 2, leading to:

$$\lambda_N(t) = \sum_{i=1}^N \frac{S_{N_i}(t)}{\sum_{i=1}^N S_{N_i}(t)} \lambda_{E_i}(t)$$

where $S_{N_i}(t)$ is defined in equation 6.

Crude probability of death

Given $F(t) = \mathbb{P}(T \leq t) = \int_0^t \lambda(u)S(u)du$ and equation 7, we can see that the **Cumulative Probability of Death** could be split into two components, the **Cumulative Probability of Death due to Cancer** and the **Cumulative Probability of Death due to Other Causes**:

$$F(t) = \int_0^t \lambda(u)S(u)du = \int_0^t \lambda_C(u)S(u)du + \int_0^t \lambda_P(u)S(u)du = F_C(t) + F_P(t)$$

Number of Life Years Lost

Finally it was shown that $\int_0^\tau F_C(u)du$ (resp. $\int_0^\tau F_P(u)du$) could be interpreted as the **Number of Life Years Lost due to cancer** (resp. **due to other causes**) between 0 and τ (see Andersen PK. *Decomposition of number of life years lost according to causes of death*. Stat Med. 2013;32: 5278-5285).

Estimation of the different measures using R

In our sample, we observed $t_i, i = \{1, \dots, N\}$ associated with a vital status variable δ_i , equals to 0 when t_i was a censoring time, and 1 for a failure time. From these observed data, we could derive non-parametric estimates of all the quantities defined above. Using simulated data of men diagnosed with a given cancer, we illustrated below **how those measures can be estimated using the R-software**. In the part corresponding to the estimation of the Restricted Mean Net Survival Time, we provide an example of using the `boot` R-package for calculating the Bootstrap Confidence Intervals.

R setup and data management

```
library(date)
library(survival)
library(relsurv)

## Loading required package: splines

library(boot)

##
## Attaching package: 'boot'

## The following object is masked from 'package:survival':
##
##      aml

library(knitr)

sessionInfo()

## R version 3.5.1 (2018-07-02)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 16299)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=French_France.1252 LC_CTYPE=French_France.1252
## [3] LC_MONETARY=French_France.1252 LC_NUMERIC=C
## [5] LC_TIME=French_France.1252
##
## attached base packages:
## [1] splines stats graphics grDevices utils datasets methods
## [8] base
##
## other attached packages:
## [1] knitr_1.20 boot_1.3-20 relsurv_2.2-1 survival_2.42-6
## [5] date_1.2-38
##
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.18 lattice_0.20-35 digest_0.6.17 rprojroot_1.3-2
## [5] grid_3.5.1 backports_1.1.2 magrittr_1.5 evaluate_0.11
## [9] stringi_1.1.7 Matrix_1.2-14 rmarkdown_1.10 tools_3.5.1
## [13] stringr_1.3.1 yaml_2.2.0 compiler_3.5.1 htmltools_0.3.6
```

```

# DATA Importation

(download.file(url='https://www.dropbox.com/s/y9yebnda05tx16d/mydat.RData?raw=true',
               destfile = "./mydat.RData", mode = "wb"))

## [1] 0
load("./mydat.RData")

head(mydat)

##      sex dep  datebirth   datediag  datevstat  vstat
## 7678   1  4 1923-09-07 2003-02-06 2005-04-04     1
## 7106   1  2 1928-11-29 2004-09-14 2004-12-14     1
## 18641  1  5 1931-03-07 2002-10-09 2011-01-28     1
## 7632   1  2 1925-01-29 2004-05-17 2005-03-06     1
## 17017  1  3 1934-05-19 2004-07-24 2006-12-08     1
## 5992   1  4 1948-05-16 2005-01-01 2006-12-27     1

mydat$fu.timed <- as.numeric(mydat$datevstat-mydat$datediag)
mydat$fu.timey <- mydat$fu.timed/365.25

# Administrative censoring at 10 years
cens <- rep(10,times=nrow(mydat))
mydat$surv10y <- pmin(mydat$fu.timey, cens)
mydat$status10y <- ifelse(mydat$fu.timey<=cens, mydat$vstat, 0)

#cross-tabulation of variables sex * deprivation
table(mydat$sex, mydat$dep)

##
##      1      2      3      4      5
## 1  989 1080 1047 1027  857

# ==> Only men in this simulated dataset

# Number of events
table(mydat$status10y)

##
##      0      1
## 1625 3375

summary(mydat$surv10y)

##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## 0.002738 2.554415 4.579055 4.832667 7.094456 10.000000

```

For each patient, the variables `datebirth`, `datediag` and `datevstat` correspond to the date of birth, the date of cancer diagnosis and the date of last known vital status, respectively. We simulated data of men only. We observed 3375 deaths during the first 10 years of follow-up.

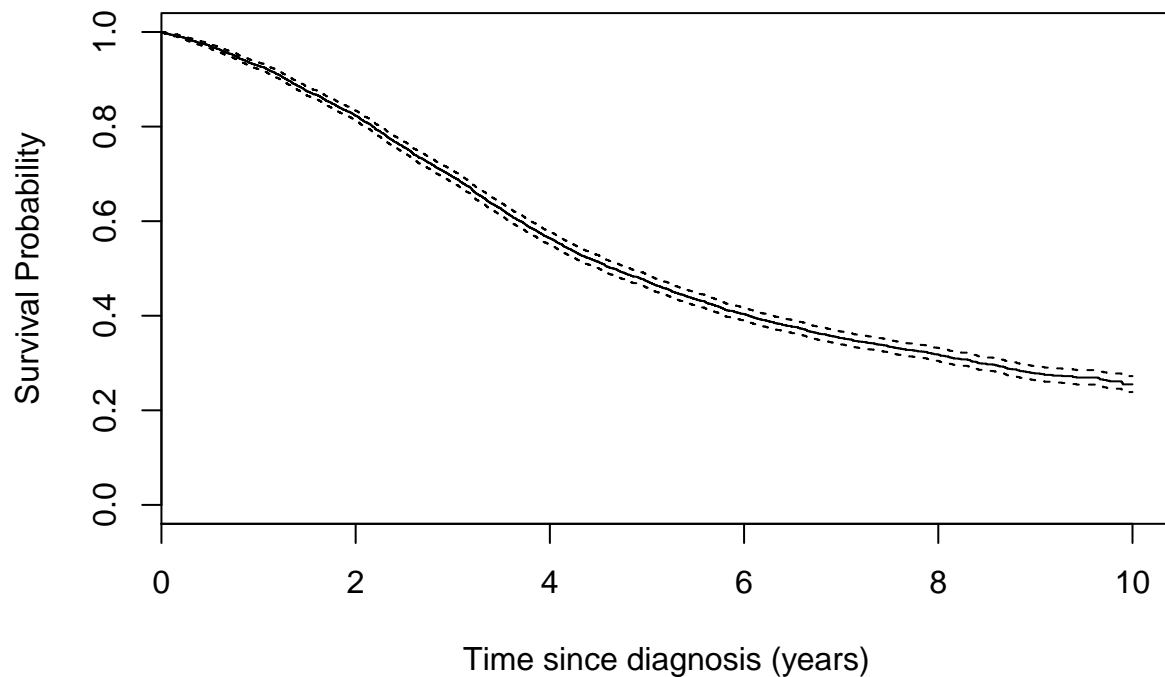
Overall survival setting

Overall survival estimates

```
KM1 <- survfit(Surv(surv10y, status10y)~1, data=mydat)
summary(KM1, times=c(1,3,5,10))
```

```
## Call: survfit(formula = Surv(surv10y, status10y) ~ 1, data = mydat)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    1   4644    356   0.929 0.00364   0.922   0.936
##    3   3470   1174   0.694 0.00652   0.681   0.707
##    5   2303   1101   0.473 0.00707   0.459   0.487
##   10    141    744   0.255 0.00861   0.239   0.272
```

```
plot(KM1, xlab="Time since diagnosis (years)", ylab="Survival Probability")
```



The estimation of the 5-year overall survival was 47.3%, 95% confidence interval (45.9, 48.7).

Conditional Survival

Estimation of the probability to survive 5 more years given that a patient has survived 5 years

```
# Function slightly adapted from the R-code  
# https://github.com/zabore/condsurv/blob/master/R/condKMest.R  
  
CondSurv <- function(kmfit, s, splust) {  
  cs <- summary(kmfit, times = splust)$surv/summary(kmfit, times = s)$surv  
  cs.sq <- cs^2  
  d <- kmfit$n.event[kmfit$time >= s & kmfit$time <= splust &  
                    kmfit$n.event > 0]  
  r <- kmfit$n.risk[kmfit$time >= s & kmfit$time <= splust &  
                  kmfit$n.event > 0]  
  dr <- d / (r * (r - d))  
  var.cs <- cs.sq * sum(dr)  
  ci <- cs + c(-1, 1) * (qnorm(0.975) * sqrt(var.cs))  
  ci.cs <- round(ci, 3)  
  if(ci.cs[1] < 0) {ci.cs[1] <- 0}  
  if(ci.cs[2] > 1) {ci.cs[2] <- 1}  
  return(list(cs = round(cs, 5), cilow = ci.cs[1], cihigh = ci.cs[2]))  
}  
  
CondSurv(kmfit=KM1, s=5, splust=10)  
  
## $cs  
## [1] 0.53898  
##  
## $cilow  
## [1] 0.507  
##  
## $cihigh  
## [1] 0.571
```

The estimation of the conditional survival at 10 years, given that a patient has survived 5 years was 53.9%, 95% confidence interval (50.7, 57.1).

Restricted Mean Survival Time

Three (equivalent) ways could be used to obtain the RMST at 10 years

```
# Way 1
fit.tableM <- summary(KM1, rmean=10)$table
RMST <- fit.tableM["*rmean"] # Estimated RMST
RMST.se <- fit.tableM["*se(rmean)"]
RMST

##      *rmean
## 5.455861

(RMST.CI <- RMST + c(-1, 1) * (qnorm(0.975)*RMST.se))

## [1] 5.360789 5.550933

# Way 2
print(KM1, print.rmean=TRUE, rmean=10)

## Call: survfit(formula = Surv(surv10y, status10y) ~ 1, data = mydat)
##
##          n      events      *rmean *se(rmean)      median      0.95LCL
## 5.00e+03 3.38e+03 5.46e+00 4.85e-02 4.65e+00 4.50e+00
## 0.95UCL
## 4.83e+00
## * restricted mean with upper limit = 10

# Way 3
survival:::survmean(KM1, rmean=10)

## $matrix
##      records      n.max      n.start      events      *rmean
## 5.000000e+03 5.000000e+03 5.000000e+03 3.375000e+03 5.455861e+00
## *se(rmean)      median      0.95LCL      0.95UCL
## 4.850702e-02 4.651608e+00 4.501027e+00 4.826831e+00
##
## $end.time
## [1] 10
```

The estimation of the Restricted Mean Survival Time at 10 years was 5.46 years, 95% confidence interval (5.36, 5.55).

Relative survival setting

Data management

For this setting, we have to use expected mortality rates, built in the *ratetable* format, for being able to use the R package *reلسurv*. Moreover, in the R package *reلسurv* we have to use

- the follow-up time in days,
- the date of diagnosis as the number of days since the first of January 1960
- the age at diagnosis should also be coded in days

These data-management steps are detailed below.

```
## Importation of an EXPECTED MORTALITY ratetable
# (by sex, year, age and deprivation in our example)

exrates.RT=dget(file="https://www.dropbox.com/s/v8gea0b2xa5h85d/exrates.RT.dat?dl=1")
summary(exrates.RT)

## Rate table with 4 dimensions:
## AGE.RT ranges from 0 to 36159.75; with 100 categories
## SEX.RT has levels of: 1 2
## DEPRIVQ.RT has levels of: 1 2 3 4 5 9
## YEAR.RT ranges from 7671 to 20089; with 35 categories

# Variables useful to use the ratetable
mydat$agediagindays <- as.numeric(mydat$datediag-mydat$datebirth)
summary(mydat$agediagindays)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 13720  23351  26098   25722  28471  35356

mydat$diagdays1960 <- as.numeric(mydat$datediag-as.Date("01/01/1960", format="%d/%m/%Y"))
summary(mydat$diagdays1960)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 15027  15587  15842   15847  16111  16690
```

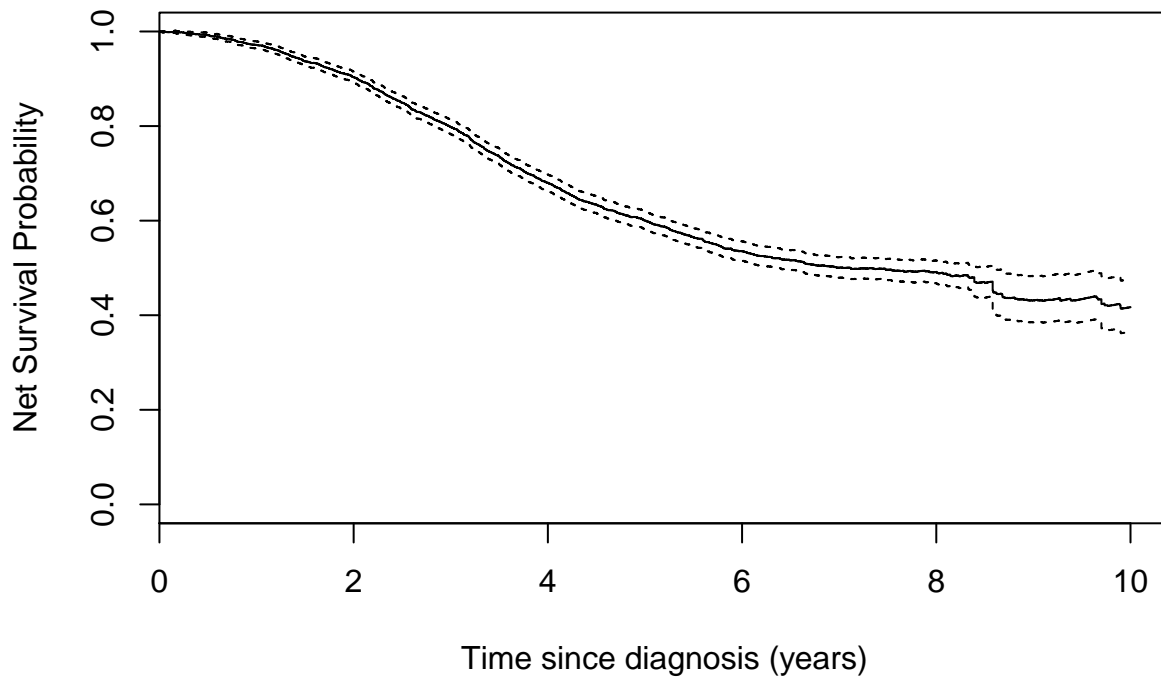
Net survival

```
NS <- rs.surv(Surv(surv10y*365.25, status10y)~1+ratetable(AGE.RT=agediagindays, SEX.RT=sex,
                                                         DEPRIVQ.RT=dep, YEAR.RT=diagdays1960),
              data=mydat, ratetable=exrates.RT, method = "pohar-perme")
summary(NS, times=c(1,3,5,10)*365.25)
```

```
## Call: rs.surv(formula = Surv(surv10y * 365.25, status10y) ~ 1 + ratetable(AGE.RT = agediagindays,
##   SEX.RT = sex, DEPRIVQ.RT = dep, YEAR.RT = diagdays1960),
##   data = mydat, ratetable = exrates.RT, method = "pohar-perme")
##
```

```
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
## 365 4644 356 0.972 0.00383 0.964 0.979
## 1096 3470 1174 0.798 0.00777 0.783 0.813
## 1826 2303 1101 0.600 0.00957 0.582 0.619
## 3652 141 744 0.417 0.02848 0.365 0.477
```

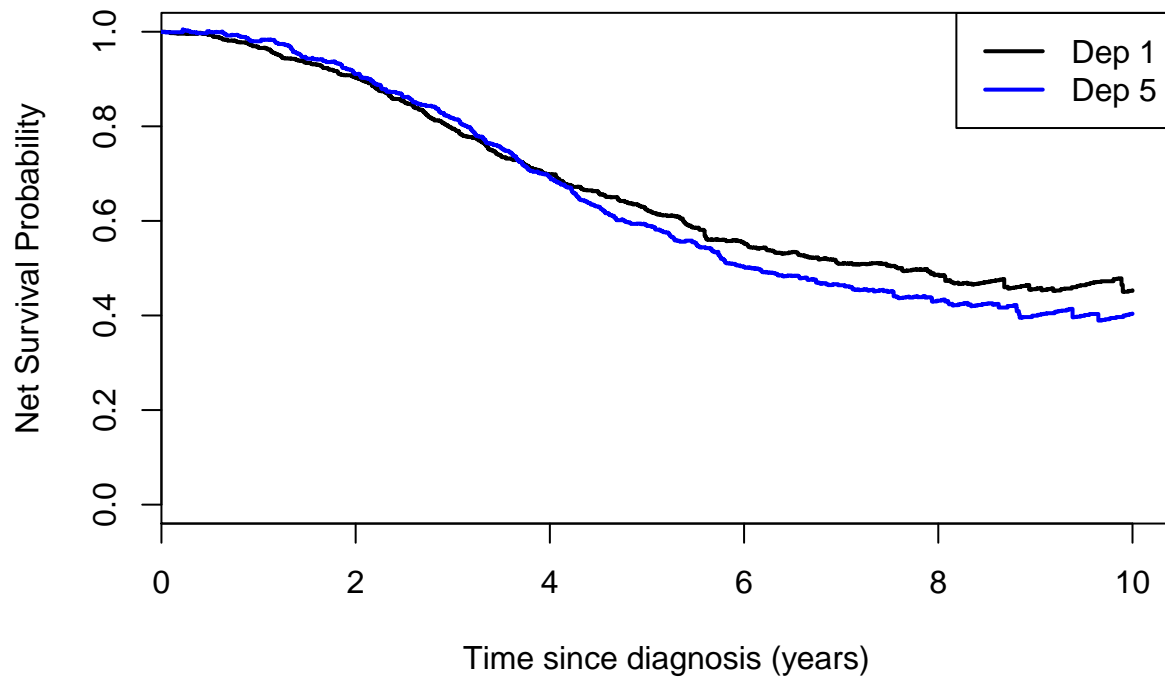
```
plot(NS, xscale = 365.25, xlab="Time since diagnosis (years)", ylab="Net Survival Probability")
```



```
NSdep1 <- rs.surv(Surv(surv10y*365.25, status10y)~1+ratetable(AGE.RT=agediagindays, SEX.RT=sex,
                                                             DEPRIVQ.RT=dep, YEAR.RT=diagdays1960),
                  data=mydat[mydat$dep==1,], ratetable=exrates.RT, method = "pohar-perme")
NSdep5 <- rs.surv(Surv(surv10y*365.25, status10y)~1+ratetable(AGE.RT=agediagindays, SEX.RT=sex,
                                                             DEPRIVQ.RT=dep, YEAR.RT=diagdays1960),
                  data=mydat[mydat$dep==5,], ratetable=exrates.RT, method = "pohar-perme")
plot(NSdep1, xscale = 365.25, xlab="Time since diagnosis (years)", ylab="Net Survival Probability",
     conf.int = F, lwd=2)
```



```
lines(NSdep5, col="blue", conf.int = F, lwd=2)
legend("topright", c("Dep 1", "Dep 5"), lty=1, col=c(1,"blue"), lwd=2)
```



The estimation of the 5-year net survival was 60.0%, 95% confidence interval (58.2, 61.9).

Conditional Net Survival

```
# s and splust should be in years (the recoding in days is done inside the function)
condNSest <- function(nsfit, s, splust) {
  cns <- summary(nsfit, times = splust*365.25)$surv/summary(nsfit, times = s*365.25)$surv
  return(cns = round(cns, 5))
}

condNSest(nsfit=NS, s=5, splust=10)
```

```
## [1] 0.69509
```

The estimation of the conditional net survival at 10 years, given that a patient has survived 5 years was 69.5%.

Restricted Mean Net Survival Time

```
ForRMNST <- survival::survmean(NS, rmean=10*365.25)

# RMNST:
survival::survmean(NS, rmean=10*365.25)[[1]][ "*rmean" ]/365.25

## *rmean
## 6.617598

myfunc <- function(dat, i){
  myns <- rs.surv(Surv(surv10y*365.25, status10y)~1+ratetable(AGE.RT=agediagindays,
                                                             SEX.RT=sex,
                                                             DEPRIVQ.RT=dep,
                                                             YEAR.RT=diagdays1960),
                 data=dat[i,], ratetable=exrates.RT, method = "pohar-perme")
  RMNST <- survival::survmean(myns, rmean=10*365.25)[[1]][ "*rmean" ]/365.25
  return(as.numeric(RMNST))
}
# myfunc(mydat, i=seq(1, nrow(mydat)))

# Bootstrap Confidence intervals
NBOOT=10
myboot <- boot(data=mydat, statistic = myfunc, R=NBOOT,
               parallel= "multicore", ncpus=6)
(myCIboot <- boot.ci(boot.out = myboot, type = "perc", index=1))

## Warning in norm.inter(t, alpha): extreme order statistics used as endpoints

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 10 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = myboot, type = "perc", index = 1)
##
## Intervals :
## Level      Percentile
## 95%      ( 6.469, 6.708 )
## Calculations and Intervals on Original Scale
## Warning : Percentile Intervals used Extreme Quantiles
## Some percentile intervals may be unstable
```

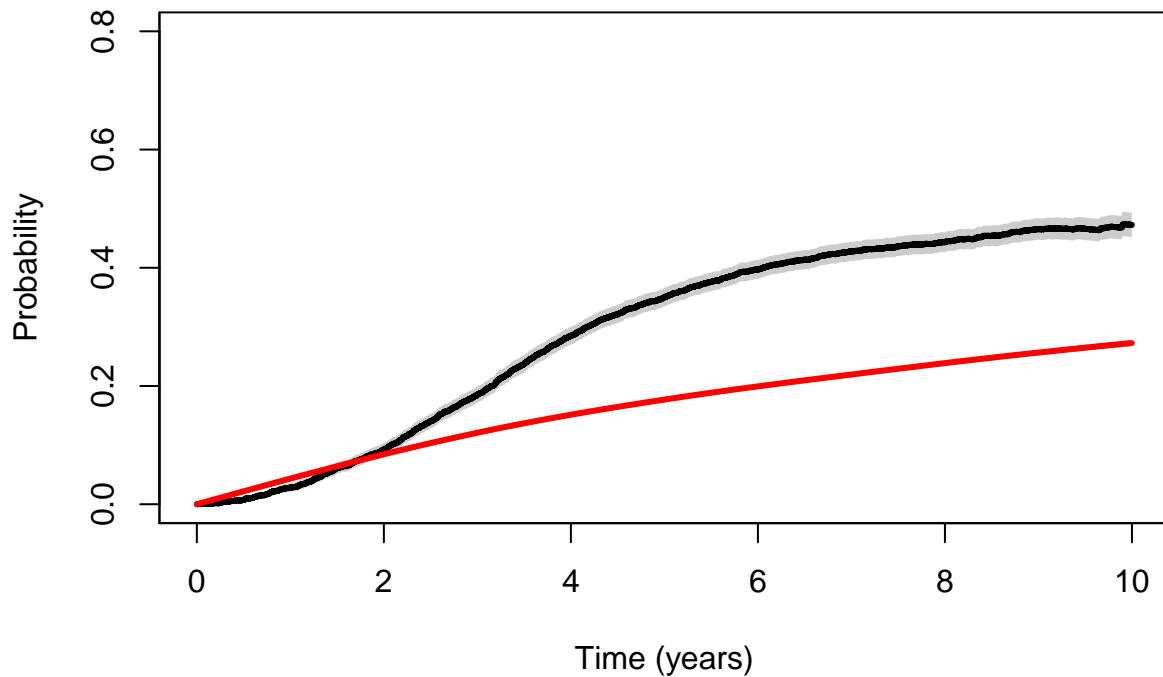
The estimation of the Restricted Mean Net Survival Time at 10 years was 6.62 years.

Crude Probability of Death

```
CPD <- cmp.rel(Surv(surv10y*365.25, status10y)~1+ratetable(AGE.RT=agediagindays,
  SEX.RT=sex,
  DEPRIVQ.RT=dep,
  YEAR.RT=diagdays1960),
  data=mydat, ratetable=exprates.RT, tau=10*365.25)
summary(CPD, times = c(1,3,5,10), xscale = 365.25)

## $est
##           1           3           5           10
## causeSpec 0.02788501 0.1851964 0.3500203 0.4724758
## population 0.04331499 0.1208036 0.1769559 0.2725736
##
## $var
##           1           3           5           10
## causeSpec 1.377008e-05 4.795215e-05 6.263708e-05 1.101885e-04
## population 7.753599e-09 2.519390e-07 1.080574e-06 6.712050e-06

plot(CPD, col = 1:2, lwd = 3, xscale = 365.25, curves=c(1,2), lty=rep(c(1,3), each=2),
  ylim=c(0,0.8), xlab = "Time (years)", conf.int = c(1,2), wh=c(0,-5))
box()
```



The estimation of the Crude Probability of Death due to cancer at 10 years was 47.2%, and the Crude Probability of Death due to other causes at 10 years was 27.3%.

Number of Life Years Lost due to cancer and due to other causes

```
(LYLC <- CPD$causeSpec$area)
```

```
## [1] 2.916029
```

```
(LYLO <- CPD$population$area)
```

```
## [1] 1.630098
```

The estimation of the Number of Life Years Lost due to cancer over a 10-year time window was 2.92 years and the Number of Life Years Lost due to other causes over a 10-year time window was 1.63 years.

Some checks

Given equation (8) from the main paper, the sum of the mean number of life years lost due to cancer and due to other causes up to 10 years should be equal to 10 minus the Restricted Mean Survival Time:

$$\int_0^{10} (1 - S(u))du = NLYC_C(10) + NLYL_P(10)$$

Thus we have:

$$NLYC_C(10) + NLYL_P(10) + RMST(10) = 10$$

```
(LYLC + LYLO + RMST)
```

```
## *rmean
## 10.00199
```

Moreover, we have seen above that

$$F(t) = 1 - S(t) = F_C(t) + F_P(t), \forall t \in [0, 10]$$

, or equivalently that

$$S(t) + F_C(t) + F_P(t) = 1, \forall t \in [0, 10]$$

```
# At time t=1 year:
```

```
sum(summary(KM1,times=1)$surv,
     summary(CPD, times = 1, xscale = 365.25)$est["causeSpec",],
     summary(CPD, times = 1, xscale = 365.25)$est["population",])
```

```
## [1] 1
```

```
sum(summary(KM1,times=4.85)$surv,
     summary(CPD, times = 4.85, xscale = 365.25)$est["causeSpec",],
     summary(CPD, times = 4.85, xscale = 365.25)$est["population",])
```

```
## [1] 1
```

```
sum(summary(KM1,times=10)$surv,
     summary(CPD, times = 10, xscale = 365.25)$est["causeSpec",],
     summary(CPD, times = 10, xscale = 365.25)$est["population",])
```

```
## [1] 1
```

Graphical summary of the different measures

