

How to estimate a new mixture cure model for increased risk of non cancer death

Juste Goungounga, Laura Botta, Riccardo Capocaccia, Gaelle Romain, Marc Colonna, Gemma Gatta, Olayidé Boussari, and Valérie Jooste

Introduction

`curesurv` is an R-package to fit a variety of cure models using excess hazard modelling methodology. It can be a mixture cure model with the survival of uncured patients following a Weibull (Botta et al. 2023, BMC Medical Research Methodology). The package also implements non-mixture cure models such as the time-to-null excess hazard model proposed by Boussari et al (2021) (Boussari et al. 2021). If the modelling assumption of comparability between the expected hazard in the study cohort and the general population doesn't hold, an extra effect (due to life table mismatch) can be estimated for these two classes of cure models. In the following we will only be interested by mixture cure models implemented in the `curesurv` R-package.

For more details regarding mixture cure models in net survival setting please read the methods section in the article untitled: "A new cure model that corrects for increased risk of non-cancer death. Analysis of reliability and robustness, and application to real-life data" in BMC medical research methodology.

Installation

The latest version of `curesurv` can be installed using the tar.gz version of the R-package uploaded to the journal website using the following r script:

```
install.packages("curesurv_0.1.0.tar.gz",  
                 repos = NULL,
```

```
type = "source")
```

`curesurv` depends on the `stringr` and `survival` R packages, which can be installed directly from CRAN.

It also uses other R packages that can be imported, such as: `numDeriv`, `stats`, `randtoolbox`, `bbmle`, `optimx`, `Formula`, `Deriv`, `statmod`, and `xhaz`.

Once these other packages are installed, load the `curesurv` R package.

```
library(xhaz)
#> Warning: le package 'xhaz' a été compilé avec la version R 4.2.2
#> Le chargement a nécessité le package : statmod
#> Le chargement a nécessité le package : survival
library(survexp.fr)
library(curesurv)
#> Le chargement a nécessité le package : stringr
```

Fitting a classical mixture cure model in excess hazard modelling setting (M0)

We illustrate the fitting of mixture cure models using a simulated dataset named `testiscancer`. This dataset is available in the `curesurv` package. It consists of 2,000 patients with information on their age, follow-up time, and vital status. The patients are assumed to be male and diagnosed in 2000-01-01 for a testis cancer. The French life table in the R package `survexp.fr` can be used as the mortality table of the French general population for illustration purposes.

Importing the dataset

```
# We had these information in the dataset
data("testiscancer", package = "curesurv")

head(testiscancer)
#>      age      age_cr age_classe  time_obs event  cumehazard      ehazard
#> 1 30.81199 -1.0669519      <40 15.0000000      0 4.360425e-03 1.060500e-03
#> 2 30.36516 -1.0905122      <40  0.6849198      1 1.332385e-05 2.169431e-05
#> 3 24.87704 -1.3798868      <40 10.7766145      0 2.748548e-04 8.644893e-05
#> 4 32.19556 -0.9939996      <40  8.6270943      0 1.151121e-03 3.347585e-04
```

```

#> 5 30.36346 -1.0906016      <40  0.9690360      1 1.976659e-05 2.375032e-05
#> 6 30.52206 -1.0822390      <40 15.0000000      0 4.068062e-03 9.952639e-04
#>   weisurvpop
#> 1  0.9956491
#> 2  0.9999867
#> 3  0.9997252
#> 4  0.9988495
#> 5  0.9999802
#> 6  0.9959402

dim(testiscancer)
#> [1] 2000      8
testiscancer$sex <- "male"
levels(testiscancer$sex) <- c("male", "female")
testiscancer$year <- as.Date("2000-01-01")

```

Extracting the population background mortality from the French life table

Here we show how to extract background mortality information from `survexp.fr` (2022), the life table for France available in the eponymous R package.

```

attributes(survexp.fr)
#> $dim
#> [1] 100  2  43
#>
#> $dimnames
#> $dimnames[[1]]
#>  [1] "0"  "1"  "2"  "3"  "4"  "5"  "6"  "7"  "8"  "9"  "10" "11" "12" "13" "14"
#> [16] "15" "16" "17" "18" "19" "20" "21" "22" "23" "24" "25" "26" "27" "28" "29"
#> [31] "30" "31" "32" "33" "34" "35" "36" "37" "38" "39" "40" "41" "42" "43" "44"
#> [46] "45" "46" "47" "48" "49" "50" "51" "52" "53" "54" "55" "56" "57" "58" "59"
#> [61] "60" "61" "62" "63" "64" "65" "66" "67" "68" "69" "70" "71" "72" "73" "74"
#> [76] "75" "76" "77" "78" "79" "80" "81" "82" "83" "84" "85" "86" "87" "88" "89"
#> [91] "90" "91" "92" "93" "94" "95" "96" "97" "98" "99"
#>
#> $dimnames[[2]]
#> [1] "male"  "female"
#>
#> $dimnames[[3]]
#>  [1] "1977" "1978" "1979" "1980" "1981" "1982" "1983" "1984" "1985" "1986"

```

```

#> [11] "1987" "1988" "1989" "1990" "1991" "1992" "1993" "1994" "1995" "1996"
#> [21] "1997" "1998" "1999" "2000" "2001" "2002" "2003" "2004" "2005" "2006"
#> [31] "2007" "2008" "2009" "2010" "2011" "2012" "2013" "2014" "2015" "2016"
#> [41] "2017" "2018" "2019"
#>
#>
#> $dimid
#> [1] "age" "sex" "year"
#>
#> $factor
#> [1] 0 1 0
#>
#> $cutpoints
#> $cutpoints[[1]]
#> [1] 0.000 365.241 730.482 1095.723 1460.964 1826.205 2191.446
#> [8] 2556.687 2921.928 3287.169 3652.410 4017.651 4382.892 4748.133
#> [15] 5113.374 5478.615 5843.856 6209.097 6574.338 6939.579 7304.820
#> [22] 7670.061 8035.302 8400.543 8765.784 9131.025 9496.266 9861.507
#> [29] 10226.748 10591.989 10957.230 11322.471 11687.712 12052.953 12418.194
#> [36] 12783.435 13148.676 13513.917 13879.158 14244.399 14609.640 14974.881
#> [43] 15340.122 15705.363 16070.604 16435.845 16801.086 17166.327 17531.568
#> [50] 17896.809 18262.050 18627.291 18992.532 19357.773 19723.014 20088.255
#> [57] 20453.496 20818.737 21183.978 21549.219 21914.460 22279.701 22644.942
#> [64] 23010.183 23375.424 23740.665 24105.906 24471.147 24836.388 25201.629
#> [71] 25566.870 25932.111 26297.352 26662.593 27027.834 27393.075 27758.316
#> [78] 28123.557 28488.798 28854.039 29219.280 29584.521 29949.762 30315.003
#> [85] 30680.244 31045.485 31410.726 31775.967 32141.208 32506.449 32871.690
#> [92] 33236.931 33602.172 33967.413 34332.654 34697.895 35063.136 35428.377
#> [99] 35793.618 36158.859
#>
#> $cutpoints[[2]]
#> NULL
#>
#> $cutpoints[[3]]
#> [1] "1977-01-01" "1978-01-01" "1979-01-01" "1980-01-01" "1981-01-01"
#> [6] "1982-01-01" "1983-01-01" "1984-01-01" "1985-01-01" "1986-01-01"
#> [11] "1987-01-01" "1988-01-01" "1989-01-01" "1990-01-01" "1991-01-01"
#> [16] "1992-01-01" "1993-01-01" "1994-01-01" "1995-01-01" "1996-01-01"
#> [21] "1997-01-01" "1998-01-01" "1999-01-01" "2000-01-01" "2001-01-01"
#> [26] "2002-01-01" "2003-01-01" "2004-01-01" "2005-01-01" "2006-01-01"
#> [31] "2007-01-01" "2008-01-01" "2009-01-01" "2010-01-01" "2011-01-01"

```

```

#> [36] "2012-01-01" "2013-01-01" "2014-01-01" "2015-01-01" "2016-01-01"
#> [41] "2017-01-01" "2018-01-01" "2019-01-01"
#>
#>
#> $class
#> [1] "ratetable"

fit.haz <- exp haz(
  formula = Surv(time_obs, event) ~ 1,
  data = testiscancer,
  ratetable = survexp.fr,
  only_ehazard = FALSE,
  rmap = list(age = 'age',
              sex = 'sex',
              year = 'year'))
#instantaneous population hazard
testiscancer$haz <- testiscancer$ehazard
#cumulative population hazard
testiscancer$cumhaz <- testiscancer$cumehazard

```

Fitting a conventional mixture cure model with the uncured survival following a Weibull distribution

We fit a mixture cure model in the excess hazard modeling setting. We assume that reduced and centered age affect the cure rate through a logistic link function and uncured survival proportionally (De Angelis et al. 1999).

```

fit_mod0 <- curesurv(Surv(time_obs, event) ~ age_cr | age_cr,
  pophaz = "haz",
  cumpophaz = "cumhaz",
  model = "mixture", dist = "weib",
  data = testiscancer,
  method_opt = "L-BFGS-B")

fit_mod0
#> Call:
#> curesurv(formula = Surv(time_obs, event) ~ age_cr | age_cr, data = testiscancer,
#>   pophaz = "haz", cumpophaz = "cumhaz", model = "mixture",
#>   dist = "weib", method_opt = "L-BFGS-B")

```

```

#>
#>               coef se(coef)      z      p
#> beta0          1.24636  0.08549  14.579 < 2e-16
#> beta_1_age_cr -1.20017  0.10120 -11.859 < 2e-16
#> lambda        -1.28957  0.10792 -11.949 < 2e-16
#> gamma          0.52508  0.07693   6.825 8.79e-12
#> delta_1_age_cr -0.29320  0.08734  -3.357 0.000788
#>
#> Baseline hazard estimates and their 95% CI
#>
#> after back-transformation
#>      exp(coef)    LCI    UCI
#> lambda      0.2754 0.2171 0.334
#> gamma       1.6906 1.4357 1.946
#>
#> Cured proportion (Z) = ((1+exp(- 0- Z))^-1 and its 95% CI
#> (For each Z of (age_cr) the others are at reference level)
#>      estimates    LCI    UCI
#> 0              0.7767 0.7476 0.806
#> _1_age_cr      0.5115 0.4620 0.561
#>
#> log-likelihood: -2561.118 (for 5 degree(s) of freedom)
#> AIC: 5132.237
#>
#> n= 2000 , number of events= 949

```

We found that the effect of reduced and centered age on survival of uncured patients was 0.29 with a standard error of 0.08. The cure rates and their 95% confidence intervals for patients 51.05 and 70.01 years of age (representing individuals with the average age in the study cohort and that of individuals whose age corresponds to an increase of one unit of reduced centered age) were 77.67% [74.76; 80.60] and 51.15% [46.20; 56.10], respectively.

Fitting a new mixture cure model with the uncured survival (M1) allowing the background mortality to be corrected

We fit a new mixture cure model in the excess hazard modelling setting proposed by Botta et al. 2023. It allows the background mortality to be corrected using the argument `pophaz.alpha = TRUE`, in order to account for increased non-cancer mortality as in (Phillips, Coldman, and McBride 2002). As previously, the new model assumes that reduced and centered age affects the cure rate through a logistic link function and the uncured survival proportionally. This model was presented at the 15th Francophone Conference on Clinical Epidemiology (EPICLIN)

and at the 28th Journées des statisticiens des centres anticancéreux (CLCC), in Marseille, France (Botta et al. 2021).

```
fit_mod1 <- curesurv(Surv(time_obs, event) ~ age_cr | age_cr,
                    pophaz = "haz",
                    cumpophaz = "cumhaz",
                    model = "mixture", dist = "weib",
                    data = testiscancer,
                    pophaz.alpha = TRUE,
                    method_opt = "L-BFGS-B")

fit_mod1
#> Call:
#> curesurv(formula = Surv(time_obs, event) ~ age_cr | age_cr, data = testiscancer,
#>      pophaz = "haz", cumpophaz = "cumhaz", pophaz.alpha = TRUE,
#>      model = "mixture", dist = "weib", method_opt = "L-BFGS-B")
#>
#>               coef se(coef)      z      p
#> beta0           1.87764  0.13503 13.905 < 2e-16
#> beta_1_age_cr  -0.38887  0.14957 -2.600 0.00933
#> lambda         -1.53696  0.18253 -8.420 < 2e-16
#> gamma           0.78748  0.07703 10.223 < 2e-16
#> delta_1_age_cr -0.35511  0.13654 -2.601 0.00930
#> pophaz.alpha0   0.67301  0.04690 14.350 < 2e-16
#>
#> Baseline hazard estimates and their 95% CI
#>
#> after back-transformation
#>           exp(coef)    LCI    UCI
#> lambda           0.2150 0.1381 0.292
#> gamma            2.1979 1.8660 2.530
#> pophaz.alpha0    1.9601 1.7799 2.140
#>
#> Cured proportion (Z) = ((1+exp(- 0- Z))^-1 and its 95% CI
#> (For each Z of (age_cr) the others are at reference level)
#>           estimates    LCI    UCI
#> 0              0.8673 0.8369 0.898
#> _1_age_cr      0.8159 0.7719 0.860
#>
#> log-likelihood: -2496.391 (for 6 degree(s) of freedom)
#> AIC: 5004.781
#>
```

```
#> n= 2000 , number of events= 949
```

We found that the effect of reduced and centered age on survival of uncured patients was higher in the new model and was 0.35 with a standard error of 0.13. The cure rates and their 95% confidence intervals for patients 51.05 and 70.01 years of age (representing individuals with the average age in the study cohort and that of individuals whose age corresponds to an increase of one unit of reduced centered age) were also higher with the new mixed cure model, with estimates of 86.73% [83.69; 89.80]% and 81.59% [77.19; 86.00]%, respectively. The new mixture cure model also estimated the increased risk of non-cancer death to be 1.96, with 95% confidence intervals [1.77; 2.14] not including 1. These results support the hypothesis of increased non-cancer mortality in the simulated testicular cancer patients.

How to select the best model

We can compare the output of these two models using Akaike information criteria (AIC).

```
fit_mod0$AIC
#> [1] 5132.237
fit_mod1$AIC
#> [1] 5004.781
```

The best model was the new mixture cure model accounting for increased risk of non-cancer death.

Plot of net survival, excess hazard and probability of being cured for the two models

Predictions by varying age values

We can be interested in the prediction of excess hazard and net survival. We propose to calculate these predictions of excess hazard and net survival at equal time 2 years since diagnosis, as a function of centered and reduced age ranging from -1.39 to 1.5, which corresponds to age values ranging from 24.69 to 83.48. We also added the curves of $P_i(t)$, the probability of being cured at a given time t after diagnosis knowing that he/she was alive until the time 2 years since diagnosis (Boussari et al. 2018), which can be obtained from the estimates of these models.


```

val_age <- seq(-1.39, 1.8, 0.1)
newage <- round(val_age * sd(testiscancer$age) + mean(testiscancer$age), 2)

newdata3 <- with(testiscancer,
  expand.grid(
    event = 0,
    age_cr = val_age,
    time_obs = 2))

pred_age_mod0 <- sapply(1:(nrow(newdata3)),
  function(i)
    predict(object = fit_mod0,
      newdata = newdata3[i, ]))

pred_age_mod1 <- sapply(1:(nrow(newdata3)),
  function(i)
    predict(object = fit_mod1,
      newdata = newdata3[i, ]))

par(mfrow=c(2,2))
plot(newage,
  unlist(pred_age_mod0["ex_haz",]), type = "l",
  lty=1, lwd=2,
  xlab = "age",
  ylab = "excess hazard")

lines(newage,
  unlist(pred_age_mod1["ex_haz",]), type = "l",
  lty=1, lwd=2, col = "blue")

legend("topleft",
  legend = c("M0",
    "M1"),
  lty = c(1,1),
  lwd = c(2,2),
  col = c("black", "blue"))
grid()

```

```

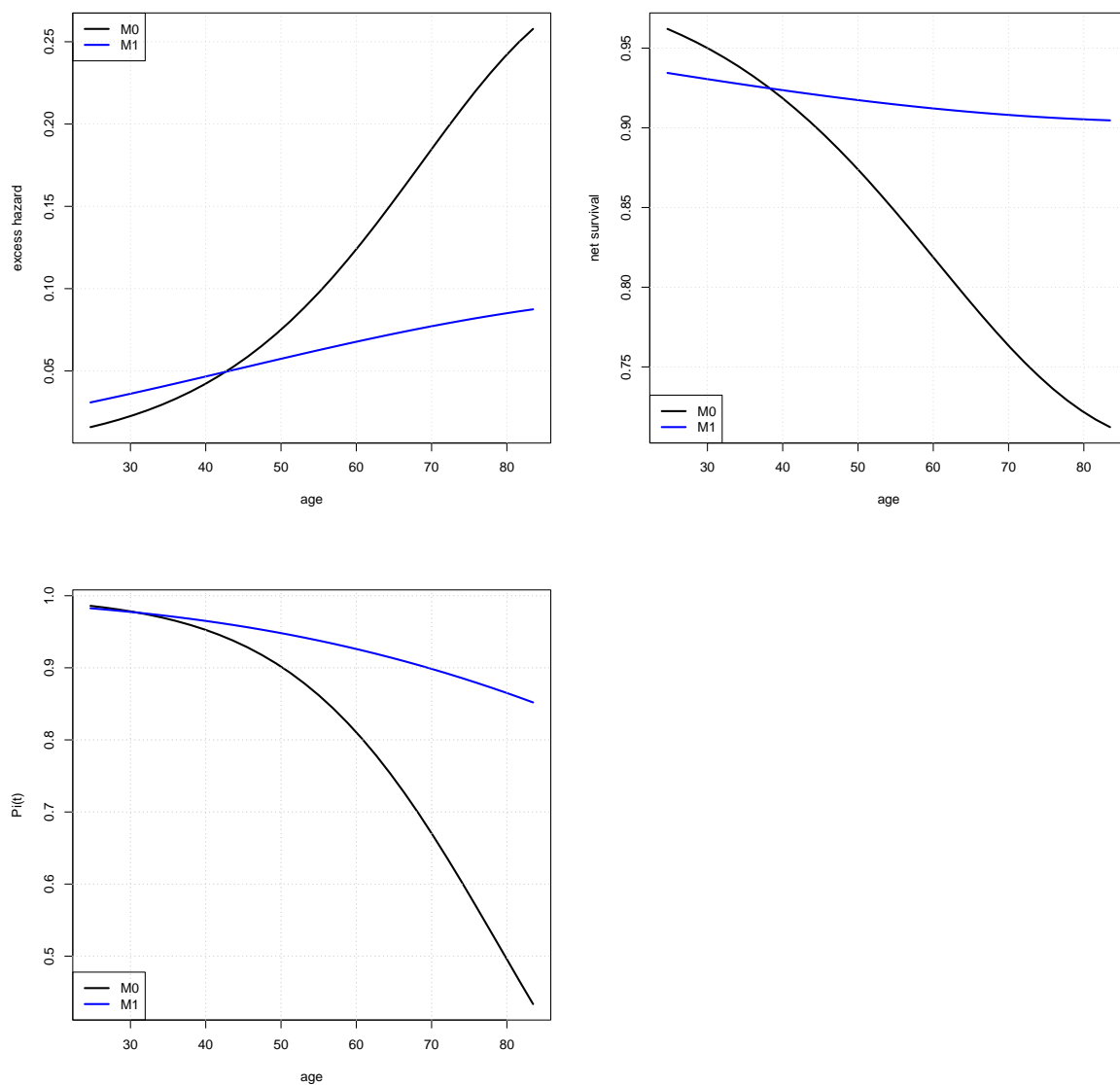
plot(newage,
     unlist(pred_age_mod0["netsurv",]) , type = "l", lty=1,
     lwd=2, xlab = "age", ylab = "net survival")

lines(newage,
     unlist(pred_age_mod1["netsurv",]) , type = "l", lty=1,
     lwd=2, col = "blue")
grid()
legend("bottomleft",
     legend = c("M0",
                 "M1"),
     lty = c(1,1),
     lwd = c(2,2),
     col = c("black", "blue"))

plot(newage,
     unlist(pred_age_mod0["pt_cure",]), type = "l", lty=1, lwd=2,
     xlab = "age",
     ylab = "Pi(t)")
grid()

lines(newage,
     unlist(pred_age_mod1["pt_cure",]), type = "l", lty=1, lwd=2,
     xlab = "age", col = "blue")
grid()
legend("bottomleft",
     legend = c("M0",
                 "M1"),
     lty = c(1,1),
     lwd = c(2,2),
     col = c("black", "blue"))
par(mfrow=c(1,1))

```



Predictions by varying time since diagnosis and for age 50 and 70

```
age50 <- c(50)
agecr50 <- (age50 - mean(testiscancer$age))/sd(testiscancer$age)

age70 <- c(70)
```

```

agecr70 <- (age70 - mean(testiscancer$age))/sd(testiscancer$age)

time <- seq(0.1, 15, 0.1)

newdata_age50 <- with(testiscancer,
  expand.grid(
    event = 0,
    age_cr = agecr50,
    time_obs = time))

newdata_age70 <- with(testiscancer,
  expand.grid(
    event = 0,
    age_cr = agecr70,
    time_obs = time))

pred_age50_mod0 <- sapply(1:(nrow(newdata_age50)),
  function(i)
    predict(object = fit_mod0,
            newdata = newdata_age50[i, ]))

pred_age70_mod0 <- sapply(1:(nrow(newdata_age70)),
  function(i)
    predict(object = fit_mod0,
            newdata = newdata_age70[i, ]))

pred_age50_mod1 <- sapply(1:(nrow(newdata_age50)),
  function(i)
    predict(object = fit_mod1,
            newdata = newdata_age50[i, ]))

pred_age70_mod1 <- sapply(1:(nrow(newdata_age70)),
  function(i)
    predict(object = fit_mod1,
            newdata = newdata_age70[i, ]))

par(mfrow=c(2,2))

```

```

plot(
  time,
  unlist(pred_age50_mod0["ex_haz", ]),
  type = "l",
  lty = 1,
  lwd = 2,
  xlab = "time since diagnosis (years)",
  ylab = "excess hazard", ylim = c(0,0.20))

lines(
  time,
  unlist(pred_age50_mod1["ex_haz", ]),
  type = "l", col = "blue",
  lty = 1,
  lwd = 2)

lines(
  time,
  unlist(pred_age70_mod0["ex_haz", ]),
  type = "l",
  lty = 2,
  lwd = 2)

lines(
  time,
  unlist(pred_age70_mod1["ex_haz", ]),
  type = "l", col = "blue",
  lty = 2,
  lwd = 2)
grid()

legend("topright",
  legend = c("M0 - age = 50",
             "M1 - age = 50",
             "M0- age = 70",
             "M1- age = 70"),
  lty = c(1,1,2,2),
  lwd = c(2,2,2,2),
  col = c("black", "blue", "black", "blue"))
plot(time,
  unlist(pred_age50_mod0["netsurv",]) , type = "l", lty=1,

```

```

    lwd=2,
    xlab = "time since diagnosis (years)",
    ylab = "net survival",
    ylim = c(0,1))

lines(time,
      unlist(pred_age50_mod1["netsurv",]) , type = "l", lty=1,
      lwd=2, col = "blue")

lines(time,
      unlist(pred_age70_mod0["netsurv",]) , type = "l", lty=2,
      lwd=2)

lines(time,
      unlist(pred_age70_mod1["netsurv",]) , type = "l", lty=2,
      lwd=2, col = "blue")
grid()

legend("bottomleft",
      legend = c("M0 - age = 50",
                  "M1 - age = 50",
                  "M0- age = 70",
                  "M1- age = 70"),
      lty = c(1,1,2,2),
      lwd = c(2,2,2,2),
      col = c("black", "blue", "black", "blue"))

plot(time,
      unlist(pred_age50_mod0["pt_cure",]),
      type = "l", lty=1, lwd=2,
      xlab = "time since diagnosis (years)",
      ylab = "Pi(t)",
      ylim = c(0,1))
grid()

lines(time,
      unlist(pred_age50_mod1["pt_cure",]),
      type = "l", lty=1, lwd=2,

```

```

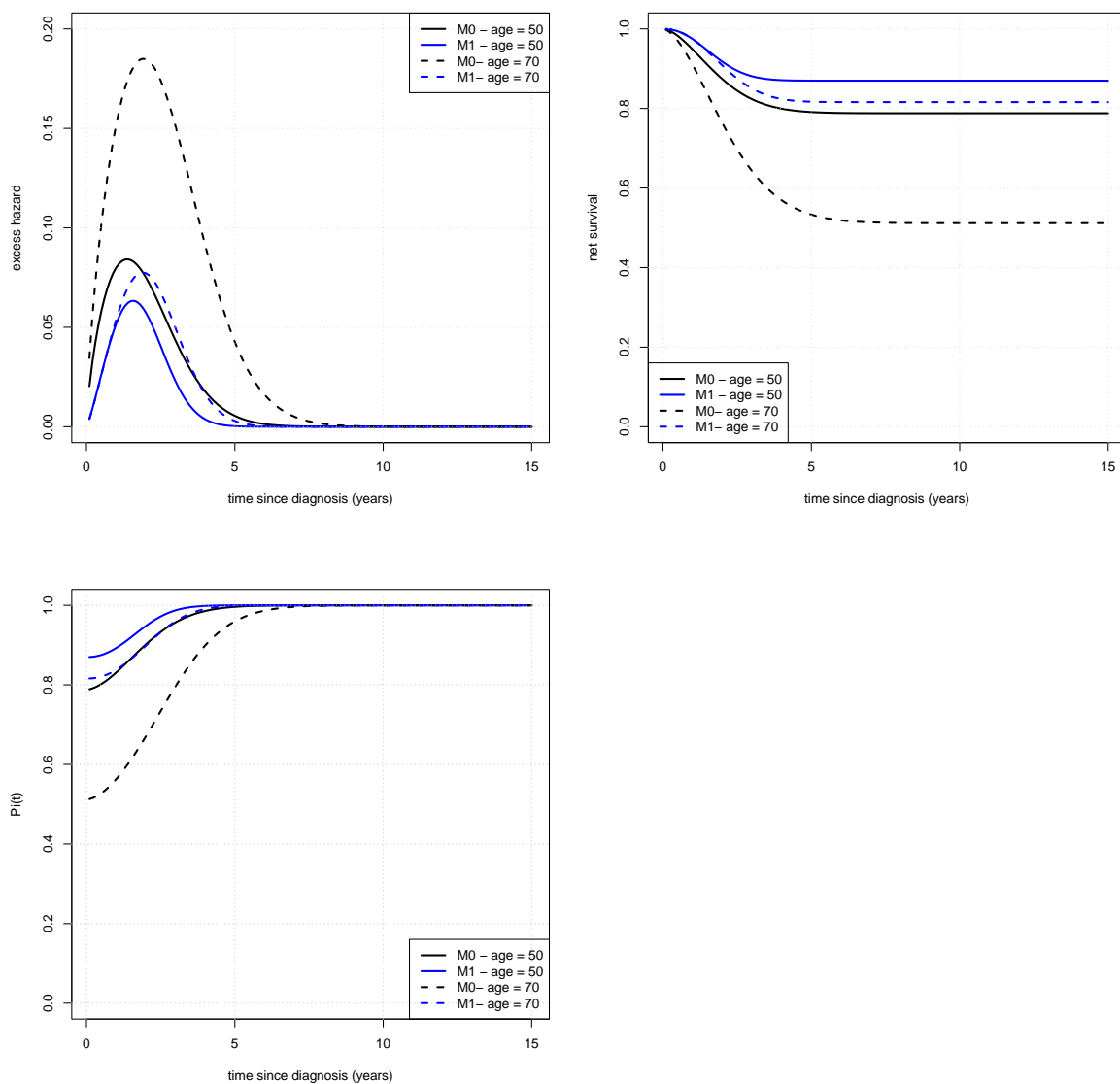
      xlab = "age", col = "blue")

lines(time,
      unlist(pred_age70_mod0["pt_cure",]),
      type = "l", lty=2, lwd=2)

lines(time,
      unlist(pred_age70_mod1["pt_cure",]),
      type = "l", lty=2, lwd=2,
      xlab = "age", col = "blue")
grid()
legend("bottomright",
      legend = c("M0 - age = 50",
                  "M1 - age = 50",
                  "M0- age = 70",
                  "M1- age = 70"),
      lty = c(1,1,2,2),
      lwd = c(2,2,2,2),
      col = c("black", "blue", "black", "blue"))

par(mfrow=c(1,1))

```



Session info

```
sessionInfo()
#> R version 4.2.1 (2022-06-23 ucrt)
#> Platform: x86_64-w64-mingw32/x64 (64-bit)
#> Running under: Windows 10 x64 (build 19044)
```



```

#>
#> Matrix products: default
#>
#> locale:
#> [1] LC_COLLATE=French_France.utf8 LC_CTYPE=French_France.utf8
#> [3] LC_MONETARY=French_France.utf8 LC_NUMERIC=C
#> [5] LC_TIME=French_France.utf8
#>
#> attached base packages:
#> [1] stats      graphics  grDevices  utils      datasets  methods    base
#>
#> other attached packages:
#> [1] curesurv_0.1.0 stringr_1.4.1 survexp.fr_1.1 xhaz_2.0.1      survival_3.3-1
#> [6] statmod_1.4.37
#>
#> loaded via a namespace (and not attached):
#> [1] Formula_1.2-4      rstudioapi_0.14     knitr_1.40
#> [4] magrittr_2.0.3      splines_4.2.1       lattice_0.20-45
#> [7] rlang_1.0.6         fastmap_1.1.0       tools_4.2.1
#> [10] optimParallel_1.0-2 parallel_4.2.1       grid_4.2.1
#> [13] xfun_0.33           cli_3.4.1           htmltools_0.5.4
#> [16] yaml_2.3.5          digest_0.6.29       numDeriv_2016.8-1.1
#> [19] Matrix_1.5-1        WriteXLS_6.4.0      evaluate_0.16
#> [22] rmarkdown_2.16      stringi_1.7.8       compiler_4.2.1
#> [25] jsonlite_1.8.2

```

License

GPL 3.0, for academic use.

Acknowledgments

We are grateful to the “Fondation ARC pour la recherche sur le cancer” for its support of this work.

References

- Botta, L., J. Goungounga, R. Capocaccia, G. Romain, M. Colonna, G. Gatta, O. Boussari, and V. Jooste. 2021. “A New Cure Model Accounting for Extra Non-Cancer Mortality: Validation and Application to Real Data.” *Revue d’Épidémiologie Et de Santé Publique* 69: S24. <https://doi.org/https://doi.org/10.1016/j.respe.2021.04.039>.
- Boussari, Olayidé, Laurent Bordes, Gaëlle Romain, Marc Colonna, Nadine Bossard, Laurent Remontet, and Valérie Jooste. 2021. “Modeling Excess Hazard with Time-to-Cure as a Parameter.” *Biometrics* 77 (4): 1289–1302.
- Boussari, Olayidé, Gaëlle Romain, Laurent Remontet, Nadine Bossard, Morgane Mounier, Anne-Marie Bouvier, Christine Binquet, Marc Colonna, and Valérie Jooste. 2018. “A New Approach to Estimate Time-to-Cure from Cancer Registries Data.” *Cancer Epidemiology* 53: 72–80. <https://doi.org/https://doi.org/10.1016/j.canep.2018.01.013>.
- De Angelis, R, R Capocaccia, T Hakulinen, B Soderman, and A Verdecchia. 1999. “Mixture Models for Cancer Survival Analysis: Application to Population-Based Data with Covariates.” *Statistics in Medicine* 18 (4): 441–54.
- Jais, Jean-Philippe, Hugo Varet, Maintainer Hugo Varet, and LazyData Yes. 2022. “Package ‘Survexp. Fr’.”
- Phillips, Norman, Andrew Coldman, and Mary L McBride. 2002. “Estimating Cancer Prevalence Using Mixture Models for Cancer Survival.” *Statistics in Medicine* 21 (9): 1257–70.