An introduction to the **cuRe** package

Lasse Hjort Jakobsen

Department of Clinical Medicine, Aalborg University

Contents

1 Key references

Paul C Lambert, John R Thompson, Claire L Weston, and Paul W Dickman. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, 8(3):576–94, 2007

> This was the first comprehensive description of parametric mixture and non-mixture cure models in a relative survival setting. The authors considered parametric modelling using, e.g., a Weibull distribution for modelling the disease-specific survival of the uncured patients and simple generalized linear models for the cure fraction. The models are implemented in the Stata commands **strsmix** and **strsnmix** [7].

P. C. Lambert, P. W. Dickman, C. P. Nelson, and P. Royston. Estimating the crude probability of death due to cancer and other causes using relative survival models. *Statistics in Medicine*, 29(7-8):885–895, mar 2010

> This article described the usage of relative survival to provide crude cause-specific probability of death. From a relative survival model, the excess hazard can be considered a causes-specific hazard function of death from cancer. Using that the relative survival multiplied with the general population survival provides the all-cause survival function, the cumulative incidences of cancer related death can be estimated. Functionalities for computing crude mortality has been implemented in the Stata module **stpm2cm**.

Therese ML Andersson, Paul W Dickman, Sandra Eloranta, and Paul C Lambert. Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. *BMC Medical Research Methodology*, 11(1):96, dec 2011

> This study combined parametric cure models with spline-based survival models. The authors introduced a non-mixture cure model which is assumed to be constant after the last knot of the splines, which forces the excess mortality to be zero from this point. The model was implemented as a part of the **stpm2** function in Stata

Therese M-L Andersson, Paul W. Dickman, Sandra Eloranta, Mats Lambe, and Paul C. Lambert. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. *Statistics in Medicine*, 32(30):5286–5300, dec 2013

> This paper introduces the loss in expectation of life using flexible relative survival models. The loss in expectation of life is computed as the area under the survival function until time infinity, which in general requires extrapolation. This was conducted by obtaining the survival function as the relative survival multiplied to the general population survival. For the relative survivail, three different models were considered, i.e., a regular spline-based relative survival model, a relative survival which is constant after the last knot of the splines, and a relative survival model which has a constant excess hazard after the last knot. Estimation of the loss in expectation of life can be conducted through the **stpm2 postestimation** Stata module.

Sandra Eloranta, Paul C. Lambert, Therese M.-L. Andersson, Magnus Björkholm, and Paul W. Dickman. The Application of Cure Models in the Presence of Competing Risks. *Epidemiology*, 25(5):742–748, sep 2014

> This paper introduced crude analogues of the cure proportion as well as the dynamic probability of cure using a flexible parametric cure model which is constant after 10 years. The crude cure proportion was computed as the cumulative incidence of cancer related death at time 10 years. The dynamic probability of cure analogue, was obtained by using the crude cure proportion, as well as the dumulative incidence of cancer related death and deaths from other causes.

Lasse Hjort Jakobsen, Therese ML Andersson, Jorne Biccler, Tarec Christoffer El-Galaly, and Martin Bgsted. Estimation of the loss of lifetime function using flexible parametric relative survival models. *BMC Medical Research Methodology*, 2017

> This paper considered the loss of lifetime function, a generalization of the loss in expectation of life, which provides the numbers of years lost due to the disease, given survival until a specific time-point. As for the loss in expectation of life, this requires extrapolation of the survival function, which was conducted by fitting a flexible parametric relative survival model and multiplying it to the general population survival. The extrapolation accuracy was assessed using three approach; a regular relative survival model, a relative survival model which is constant after the last knot, and a novel mixture cure model which uses spline to model the survival of the uncured.

2 Quick start

The `cuRe` package has two main functions for estimating parameteric cure models on the relative survival scale, namely `GenFlexCureModel` and `fit.cure.model`. In addition, functionalities for providing loss of lifetime estimates and crude mortality measures are also included.

For this vignette, we consider the `colonDC` dataset, which contain records on 15564 colon cancer patients. The name `colonDC`, indicates that the `colon` dataset of the `rstpm2` package is used after some "data cleaning".

```
library(cuRe)
data("colonDC")
head(colonDC)
```

```
##      sex age     stage    statusCOD               subsite         dx
## 1 female  77   Distant Dead: cancer             Transverse 1977-09-07
## 2 female  78 Localised  Dead: other   Coecum and ascending 1978-10-07
## 3   male  78   Distant Dead: cancer Descending and sigmoid 1978-12-07
## 4   male  76   Distant Dead: cancer Descending and sigmoid 1976-10-07
## 5   male  80 Localised Dead: cancer Descending and sigmoid 1980-04-07
```

```
## 6 female  75 Localised Dead: cancer   Coecum and ascending 1975-11-07
##         exit status   FU    FUyear  agedays
## 1 1979-01-22      1  502 1.3744387 28123.48
## 2 1985-08-22      1 2511 6.8749316 28488.72
## 3 1979-01-22      1   46 0.1259446 28488.72
## 4 1976-11-22      1   46 0.1259446 27758.24
## 5 1980-12-22      1  259 0.7091228 29219.20
## 6 1977-10-22      1  715 1.9576169 27393.00
```

By using the function `general.haz` the general population hazard at the observed follow-up times is extracted from the provided ratetable (in this case `survexp.dk`).

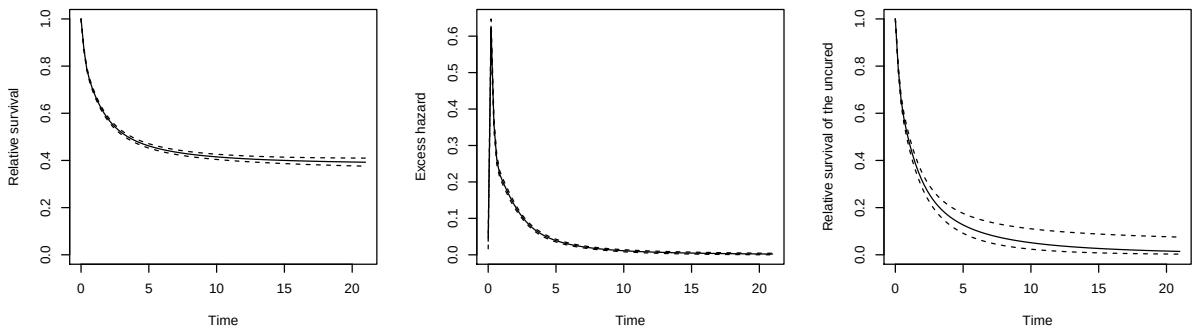
```
colonDC$bhaz <- general.haz(time = "FU", sex = "sex",
                            age = "agedays", year = "dx",
                            data = colonDC,
                            ratetable = survexp.dk)
```

Next, we fit a generalized flexible cure model. By default, a proportional hazard model is assumed for the survival of the uncured and logit link function is assumed for the cure proportion.

```
fit <- GenFlexCureModel(Surv(FUyear, status) ~ 1, data = colonDC, bhazard = "bhaz", df = 4)

## Finding initial values... Completed!
## Fitting the model... Completed!
```

Using `plot`, we may obtain the relative survival, the exess hazard, and the relative survival of the uncured patients.

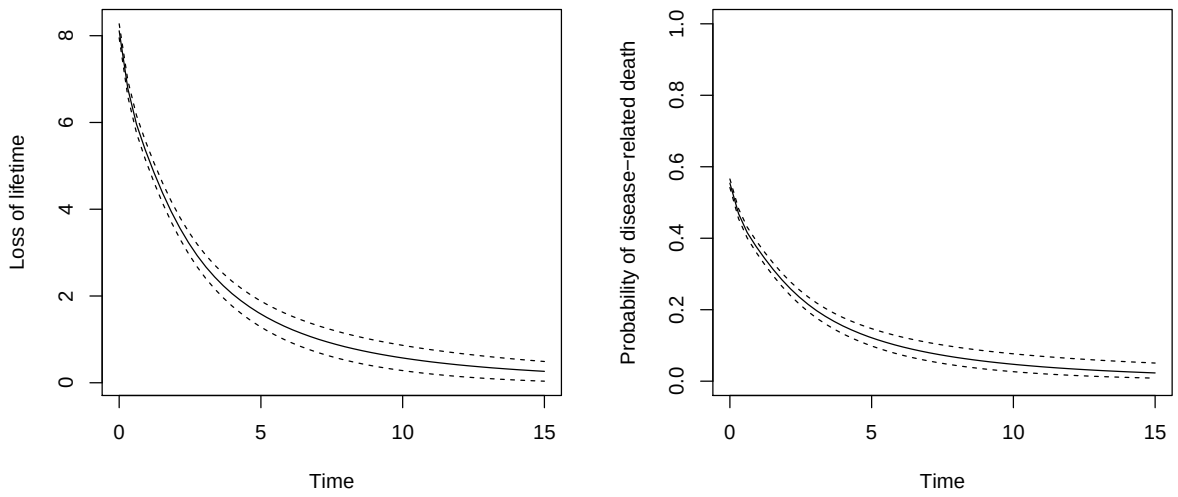
```
par(mfrow = c(1,3))
plot(fit)
plot(fit, type = "hazard")
plot(fit, type = "survuncured")
```



Using the `calc.LL` and `calc.Crude` functions, we may obtain loss of lifetime estimates and the conditional probability of cancer-related death given survival until time t (including pointwise confidence intervals).

```
ll <- calc.LL(fit, time = seq(0, 15, length.out = 50),
              rmap = list(age = agedays, sex = sex, year = dx))

prob <- calc.Crude(fit, time = seq(0, 15, length.out = 50), type = "condother",
```

```
                    rmap = list(age = agedays, sex = sex, year = dx), reverse = T)

par(mfrow = c(1,2))
plot(ll)
plot(prob)
```



3 Parametric cure models

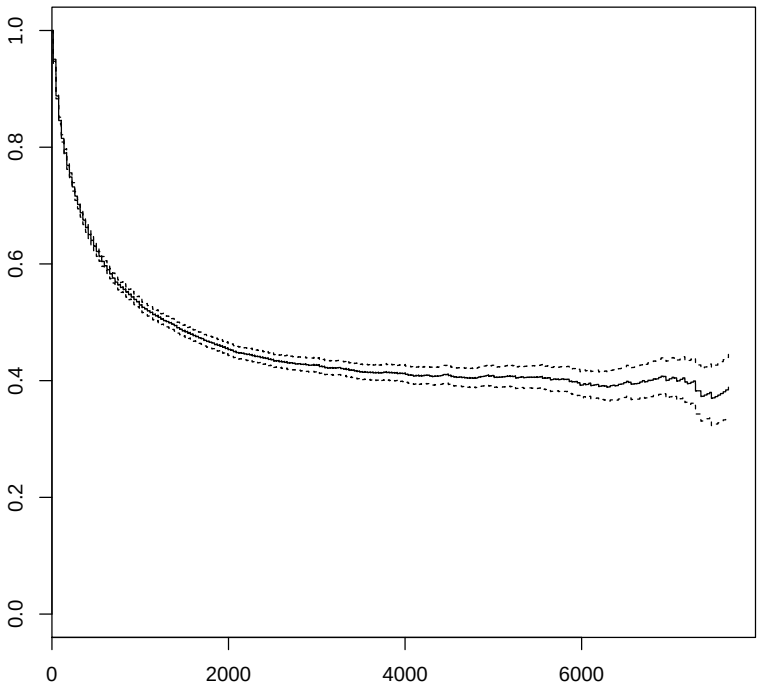
The relative survival given as,

$$R(t) = \frac{S(t)}{S^*(t)} \tag{1}$$

are commonly used to describe the net disease-specific survival in a patient population. Here $S(\cdot)$ and $S^*(\cdot)$ are the all-cause and general population survival functions, respectively. The term "net" refers to the interpretation of the disease-specific survival which is the survival of the patients in a world where it is impossible to die from other causes than the disease. Non-parametric estimators of the relative survival function have been available for many years.

The **relsurv** package contains functions for modelling the relative survival non-parametricaly. For instance, the immediate relative survival of the colon cancer patients may be computed using the function **rs.surv** in the following way.

```
library(relsurv)
fit <- rs.surv(Surv(FU, status) ~ 1 + ratetable(age = agedays,
                                                sex = sex,
                                                year = dx),
               data = colonDC,
               ratetable = survexp.dk,
               method = "ederer2")
plot(fit)
```



During the last 15 years, parametric relative survival cure models have been developed upon to yield new survival information, particularly for cancers where the patient survival approaches that of the general population. The point at which this occurs is termed the cure point and patients surviving beyond the cure point are termed *statistically cured*. The goal of cure models is to estimate the probability of being statistically cured, the survival of those patients, who are not statistically cured, and the conditional probability of being statistically cured given survival until a certain time point.

Commonly two types of cure models are considered, namely mixture and non-mixture cure models. The mixture cure model is given as:

$$R(t|z) = \pi(z) + (1 - \pi(z))S_u(t|z), \tag{2}$$

where $\pi(z)$ is the proportion of statistically cured patients and $S_u(t|z)$ is the net disease specific survival of the uncured patients. The non-mixture cure model is commonly described by:

$$R(t|z) = \pi(z)^{1-\tilde{F}(t|z)}, \tag{3}$$

where $\tilde{F}(t)$ is a proper distribution function, which does not have the same intuitive interpretation as $S_u(\cdot)$.

In the `cuRe` package, the `fit.cure.model` function allows estimation of standard parametric mixture and non-mixture cure models. Using the colon cancer data, we fit a mixture cure model with a Weibull distribution modelling the disease-specific survival of the uncured and a logistic model for the cure fraction. Both include a linear age-effect.

```
fit.cm <- fit.cure.model(Surv(FUyear, status) ~ age,
                         data = colonDC,
                         bhazard = "bhaz",
                         formula.surv = list(~ age, ~ 1))
```

```
predict(fit.cm, newdata = data.frame(age = c(50, 60, 70)),
        type = "curerate")

## [[1]]
##    Estimate     lower     upper
## 1 0.4727256 0.4536441 0.4915529
##
## [[2]]
##    Estimate    lower     upper
## 1 0.4573353 0.443956 0.4706067
##
## [[3]]
##   Estimate     lower     upper
## 1 0.442026 0.4301306 0.4538499
```

By using the `plot` function, the relative survival for three new patients can be plotted.

```
plot(fit.cm, newdata = data.frame(age = c(50, 60, 70)), col = 1:3)

## Warning in max(object$time):  no non-missing arguments to max; returning -Inf
## Error in seq.default(xlim[1], xlim[2], length.out = 100):  'to' must be a finite number
```

Now, replace the Weibull distribution with a log-normal distribution, which does not provide a proportional hazards model for the survival of the uncured.

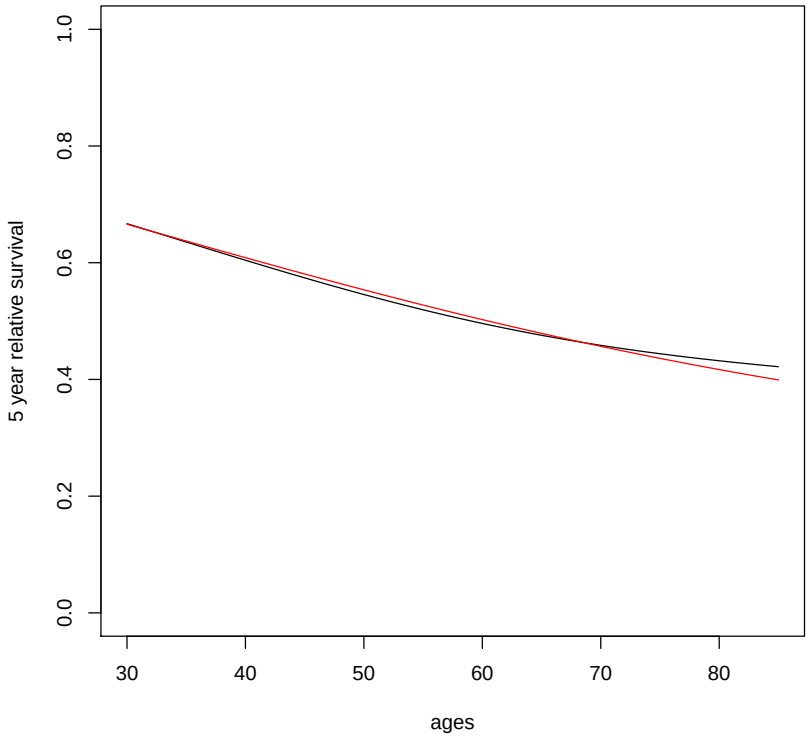
```
fit.cm.ln <- fit.cure.model(Surv(FUyear, status) ~ age, data = colonDC,
                            bhazard = "bhaz", formula.surv = list(~ age, ~ 1),
                            dist = "lognormal")

predict(fit.cm.ln, newdata = data.frame(age = c(50, 60, 70)), type = "curerate")

## [[1]]
##    Estimate     lower     upper
## 1 0.4121228 0.3894115 0.4346752
##
## [[2]]
##    Estimate     lower     upper
## 1 0.3952062 0.3780802 0.4122752
##
## [[3]]
##    Estimate     lower    upper
## 1 0.3785368 0.3628708 0.394184
```

Plot the relative survival of both models for 50, 60, and 70-year-old patients.

```
plot(fit.cm, newdata = data.frame(age = c(50, 60, 70)),
     col = 1, ci = F)

## Warning in max(object$time):  no non-missing arguments to max; returning -Inf
## Error in seq.default(xlim[1], xlim[2], length.out = 100):  'to' must be a finite number

plot(fit.cm.ln, newdata = data.frame(age = c(50, 60, 70)),
     col = 2, ci = F, add = T)

## Warning in max(object$time):  no non-missing arguments to max; returning -Inf
## Error in seq.default(xlim[1], xlim[2], length.out = 100):  'to' must be a finite number
```

Plot the 5-year relative survival against diagnostic age.

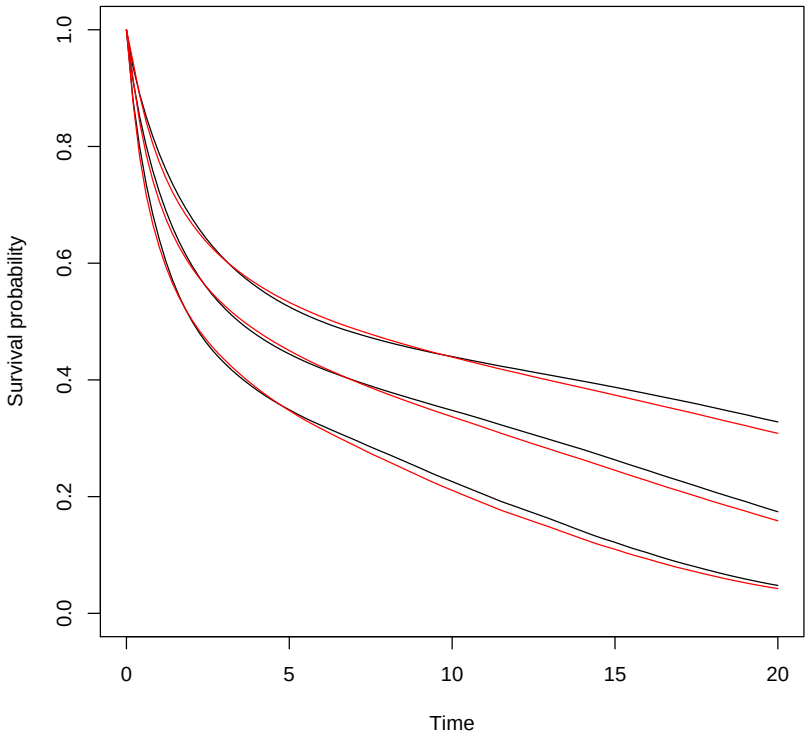
```
ages <- 30:85
rs5 <- predict(fit.cm, time = 5, newdata = data.frame(age = ages))
rs5 <- do.call(rbind, rs5)

rs5.ln <- predict(fit.cm.ln, time = 5, newdata = data.frame(age = ages))
rs5.ln <- do.call(rbind, rs5.ln)

plot(rs5$Estimate ~ ages, type = "l", ylim = c(0, 1), ylab = "5 year relative survival")
lines(rs5.ln$Estimate ~ ages, col = 2)
```



The actual survival curves can be obtained by multiplying the relative survival with the expected survival function, which is done with the `lts` function.

```
lts.wei <- lts(fit.cm,
               newdata = data.frame(age = c(50, 60, 70), sex = "male",
                                    year = 2010,
                                    age_days = c(50, 60, 70) * 365.24),
               time = seq(0, 20, length.out = 100),
               rmap = list(age = age_days))

plot(lts.wei, ci = F)

lts.ln <- lts(fit.cm.ln,
              newdata = data.frame(age = c(50, 60, 70), sex = "male",
                                   year = 2010,
```

```
                                    age_days = c(50, 60, 70) * 365.24),
              time = seq(0, 20, length.out = 100),
              rmap = list(age = age_days))

plot(lts.ln, ci = F, add = T, col = 2)
```



Observe, the all-cause survival curves are very similar despite the relative survival curves being slightly different.

4 Flexible parametric cure models

Royston and Parmar [10] introduced a flexible parametric proportional hazards model using restricted cubic splines to model the baseline hazard function (on the log cumulative hazard scale). This approach was applied to relative survival by Nelson et al. [9] where the log cumulative excess hazard was modelled by restricted cubic splines. Including covariate effects, the relative survival was given by

$$\log(-\log(R(t|\boldsymbol{z})) = s_0(x;\boldsymbol{\gamma}_0) + \boldsymbol{z}^T\boldsymbol{\beta} + \sum_{i=1}^{p} s_i(x;\boldsymbol{\gamma}_i)z_i, \tag{4}$$

where $x = \log(t)$, p is the number of time-varying covariate effects, $s_0(x;\boldsymbol{\gamma}_0)$ is a baseline restricted cubic spline, $\boldsymbol{\beta}$ is a vector of regression coefficients, and $s_i(x;\boldsymbol{\gamma}_i)$ is a spline corresponding to the i^{th} covariate, providing a time varying coefficient. For the i^{th} spline, K_i knots, $k_{i1} < k_{i2} < .. < k_{iK_i}$, are selected on the time scale. The spline is then given as a linear combination of base functions defined through the chosen knots, i.e., $s_i(x;\boldsymbol{\gamma}_i) = \sum_{j=0}^{K_i-1} v_{ij}(x)\gamma_{ij}$, where $\boldsymbol{\gamma}_i$ are model parameters. The base functions are given by $v_{i0}(x) = 1$, $v_{i1}(x) = x$, and

$$v_{ij}(x) = (x - k_{ij})_+^3 - \lambda_{ij}(x - k_{i1})_+^3 - (1 - \lambda_{ij})(x - k_{iK_i})_+^3, \tag{5}$$

for $j = 2, ..., K_i - 1$, where $\lambda_{ij} = \frac{k_{iK_i} - k_{ij}}{k_{iK_i} - k_{i1}}$ and $x_+ = \max(x, 0)$. Generally, the number and placement of the knots in the different spline functions do not need to be the same.

Andersson et al. [2] used (4) to establish a flexible parametric cure model. This model was formulated similarly to (4), but the basis functions of the splines were adjusted to ensure that the relative survival had zero slope after a pre-selected time point which was used as last knot in all spline functions, i.e., $k_K = k_{0K_0} = k_{1K_1} = \cdots = k_{pK_p}$. The cure fraction is estimated by $R(k_K)$. Rewriting (4) we obtain

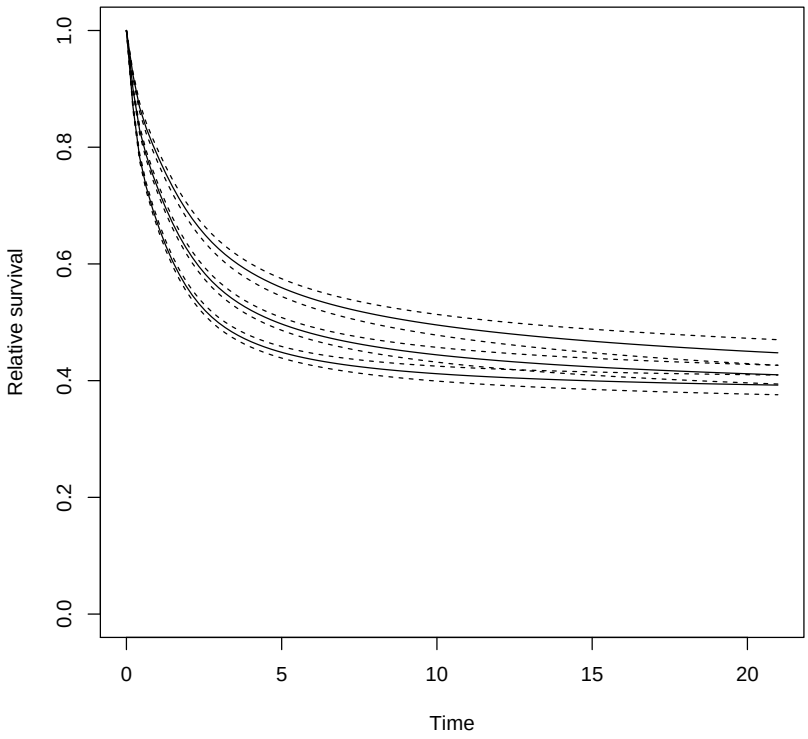
$$R(t|\boldsymbol{z}) = \exp - \exp \gamma_{00} + \boldsymbol{z}^T\boldsymbol{\beta} \exp \sum_{i=1}^{K_0-1} v_i(x)\gamma_i + \sum_{i=1}^{p} s_i(x; \boldsymbol{\gamma}_i)z_i. \tag{6}$$

Hence, the model by Andersson et al. [2] can be viewed as a non-mixture cure model where the cure fraction is modelled through the baseline spline parameter, γ_{00}, and the fixed covariate effects, $\boldsymbol{z}^T\boldsymbol{\beta}$, while the remaining parameters are used to model $1 - S_\zeta(t)$. While this model provides a flexible framework for estimating the cure fraction in cancer studies, the assumption of statistical cure after the last knot is strong. Therefore, we introduce a new flexible parametric cure model which combines regular mixture cure models with flexible parametric survival models. The model is specified by (2) with

$$S_u(t|\boldsymbol{z}) = \exp - \exp s_0(x; \boldsymbol{\gamma}_0) + \boldsymbol{z}^T\boldsymbol{\beta} + \sum_{i=1}^{p} s_i(x; \boldsymbol{\gamma}_i)z_i. \tag{7}$$

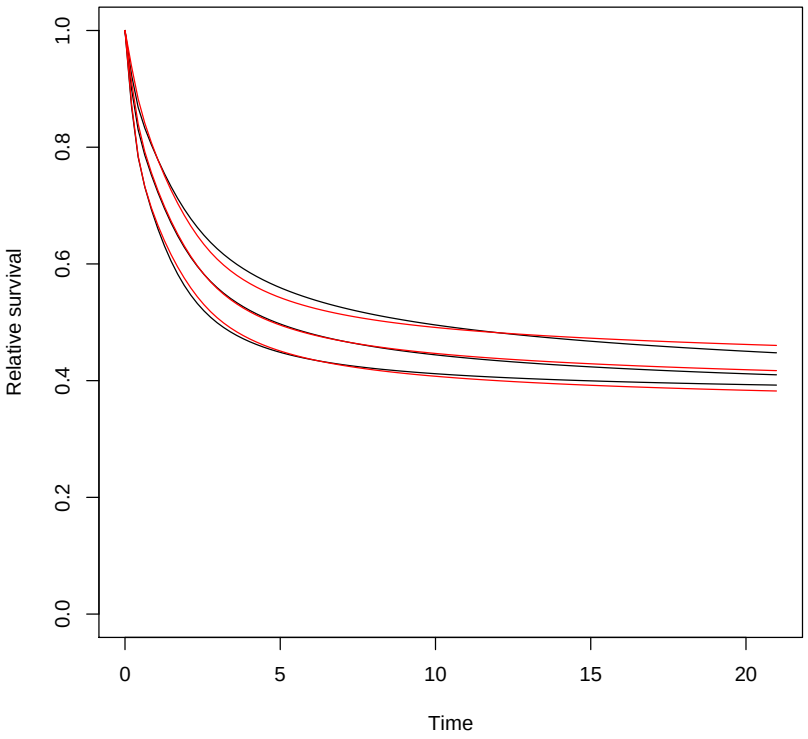
Fit the flexible parametric model using the `FlexCureModel` function.

```
fit.flex <- GenFlexCureModel(Surv(FUyear, status) ~ age, data = colonDC,
                             df = 4, bhazard = "bhaz", cr.formula = ~ age)

## Finding initial values... Completed!
## Fitting the model... Completed!

predict(fit.flex, newdata = data.frame(age = c(50, 60, 70)),
        type = "curerate")

## [[1]]
##    Estimate     lower    upper
## 1 0.3545009 0.3021751 0.407118
##
## [[2]]
##    Estimate     lower     upper
## 1 0.3650184 0.3253261 0.4047484
##
## [[3]]
##    Estimate     lower     upper
## 1 0.3756662 0.3466921 0.4045937

plot(fit.flex, newdata = data.frame(age = c(50, 60, 70)))
```



Fit the flexible parametric model with a time-varying age-effect using the `n.knots.time` argument.

```
fit.flex.time <- GenFlexCureModel(Surv(FUyear, status) ~ age, data = colonDC, df = 4,
                          bhazard = "bhaz", tvc = list(age = 3), cr.formula = ~ age)

## Finding initial values... Completed!
## Fitting the model... Completed!

plot(fit.flex, newdata = data.frame(age = c(50, 60, 70)), ci = F)
plot(fit.flex.time, newdata = data.frame(age = c(50, 60, 70)),
     col = 2, ci = F, add = T)
```




```
predict(fit.flex.time, newdata = data.frame(age = c(50, 60, 70)), type = "curerate")
```

```
## [[1]]
##    Estimate     lower     upper
## 1 0.3716126 0.3016212 0.4415096
##
## [[2]]
##    Estimate     lower     upper
## 1 0.3611471 0.3031027 0.4193812
##
## [[3]]
##    Estimate     lower     upper
## 1 0.3508117 0.3005863 0.4013773
```

```
predict(fit.cm, newdata = data.frame(age = c(50, 60, 70)), type = "curerate")
```

```
## [[1]]
##    Estimate     lower     upper
## 1 0.4727256 0.4536441 0.4915529
##
## [[2]]
##    Estimate    lower     upper
## 1 0.4573353 0.443956 0.4706067
##
## [[3]]
##   Estimate     lower     upper
## 1 0.442026 0.4301306 0.4538499
```

```r
predict(fit.flex, newdata = data.frame(age = c(50, 60, 70)), type = "curerate")
```

```
## [[1]]
##    Estimate     lower    upper
## 1 0.3545009 0.3021751 0.407118
##
## [[2]]
##    Estimate     lower     upper
## 1 0.3650184 0.3253261 0.4047484
##
## [[3]]
##    Estimate     lower     upper
## 1 0.3756662 0.3466921 0.4045937
```

Plot the survival of the uncured from the Weibull cure model, the flexible cure model, and the flexible cure model with a time-varying effect age effect.

```r
plot(fit.cm, type = "survuncured", data.frame(age = 50), ci = F)
```

```
## Warning in max(object$time):  no non-missing arguments to max; returning -Inf
## Error in seq.default(xlim[1], xlim[2], length.out = 100):  'to' must be a finite number
```

```r
plot(fit.flex, type = "survuncured", newdata = data.frame(age = 50),
     col = 2, ci = F, add = T)
```

```
## Error in plot.xy(xy.coords(x, y), type = type, ...):  plot.new has not been called yet
```

```r
plot(fit.flex.time, type = "survuncured", newdata = data.frame(age = 50),
     col = 3, ci = F, add = T)
```

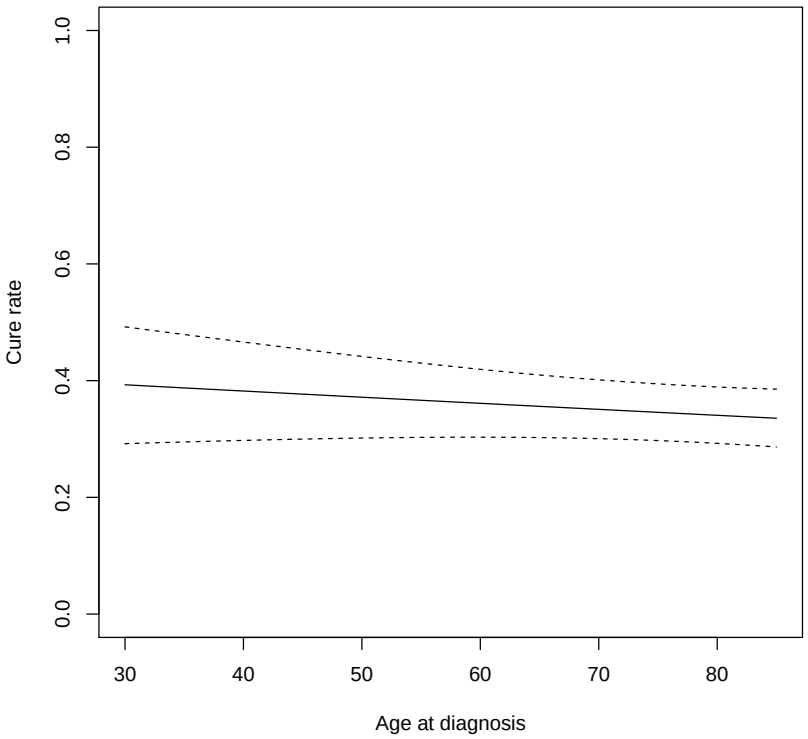
```
## Error in plot.xy(xy.coords(x, y), type = type, ...):  plot.new has not been called yet
```

```r
legend("topright", fill = 1:3, legend = c("Weibull", "Flex", "Flex (time)"))
```

```
## Error in strwidth(legend, units = "user", cex = cex, font = text.font):  plot.new has
not been called yet
```

Compute the cure fraction for a sequence of diagnostic ages.

```r
pred_cure <- predict(fit.flex.time, type = "curerate",
                     newdata = data.frame(age = ages))

pred_cure <- do.call(rbind, pred_cure)

plot(pred_cure$Estimate ~ ages, type = "l", ylim = c(0, 1),
     ylab = "Cure rate", xlab = "Age at diagnosis")

#Add confidence intervals
lines(pred_cure$lower ~ ages, lty = 2)
lines(pred_cure$upper ~ ages, lty = 2)
```



5 Loss of lifetime estimation

The loss of lifetime function is defined as the function,

$$L(t) = \frac{\int_t^\infty S^*(u)du}{S^*(t)} - \frac{\int_t^\infty S(u)du}{S(t)}. \tag{8}$$

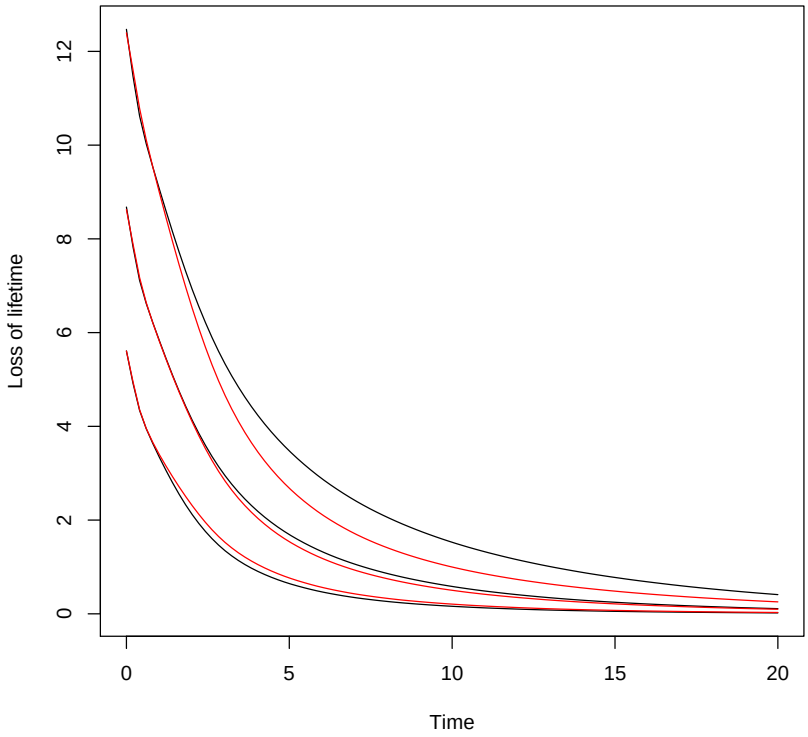
A special case of this function is the loss in expectation of life [1] which is defined as,

$$LL(t) = L(0) = \int_0^\infty S^*(u)du - \int_0^\infty S(u)du. \tag{9}$$

The `cuRe` package contains functions `calc.LL` and `plot.le` for computing and plotting, respectively, the loss of lifetime estimates. Since the function using the `survexp` function from the `survival` package, age, gender, and calendar year normally (depends on the specication of the applied ratetable) has to be specified to compute the loss of lifetime. To compute the loss of lifetime for time point 0 to 20 in the colon cancer data, we use the syntax,

```
ll <- calc.LL(fit.flex,
              newdata = data.frame(age = c(50, 60, 70),
                                   sex = "male",
                                   year = 2010,
                                   age_days = c(50, 60, 70) * 365.24),
              time = seq(0, 20, length.out = 100),
              rmap = list(age = age_days), var.type = "n")
```

```
ll.time <- calc.LL(fit.flex.time,
                   newdata = data.frame(age = c(50, 60, 70),
                                        sex = "male",
                                        year = 2010,
                                        age_days = c(50, 60, 70) * 365.24),
                   time = seq(0, 20, length.out = 100),
                   rmap = list(age = age_days), var.type = "n")

plot(ll, ci = F)
plot(ll.time, ci = F, col = 2, add = T)
```



A small difference is observed between the loss of lifetime computed using a time-varying coefficient.

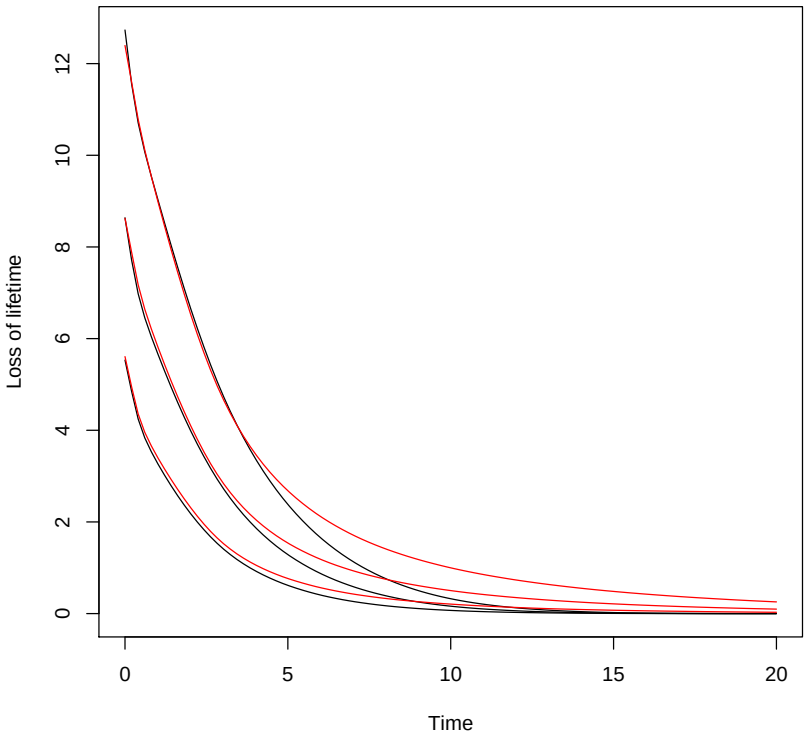
The loss of lifetime function also works with the functionalities of the **rstpm2** package [3]. Consider the following penalized generalized relative survival model with age as time-varying covariate.

```
fit.stpm2 <- stpm2(Surv(FUyear, status) ~ 1, data = colonDC, bhazard = colonDC$bhaz, df = 4,
                   tvc = list(age = 3), cure = TRUE)
```

Now, compute the loss of lifetime and compare the flexible parametric cure model, also with a time varying age effect.

```
ll.time <- calc.LL(fit.flex.time,
                   newdata = data.frame(age = c(50, 60, 70),
                                        sex = "male",
                                        year = 2010,
                                        age_days = c(50, 60, 70) * 365.24),
                   time = seq(0, 20, length.out = 100),
```

```
                    rmap = list(age = age_days), var.type = "n")

ll.stpm2.time <- calc.LL(fit.stpm2,
                         newdata = data.frame(age = c(50, 60, 70),
                                              sex = "male",
                                              year = 2010,
                                              age_days = c(50, 60, 70) * 365.24),
                         time = seq(0, 20, length.out = 100),
                         rmap = list(age = age_days), var.type = "n")

plot(ll.stpm2.time, ci = F, col = 1)
plot(ll.time, ci = F, add = T, col = 2)
```



Also here, a small difference is observed between the loss of lifetime estimates, the largest seen among the 50-year-old patient.

6 Probability of cancer related death

The crude analogue of the conditional probability of cure, as directly derived from the cure models was introduced by [4]. To derive this probability, we need to consider the crude analogue of relative survival, i.e., the cumulative incidence of cancer related death as opposed to the net cancer survival. The cumulative incidene of cancer related death, as described by Lambert et al. [6] can be computed as,

$$\mathrm{P}(T \leq t, D = \text{cancer}) = \int_0^t S^*(u) R(u) \lambda(u) du, \tag{10}$$

where $\lambda(\cdot)$ is the excess mortality. By replacing $\lambda(\cdot)$ by $h^*(\cdot)$, the cumulative incidence of death from other causes than cancer, i.e., $\mathrm{P}(T \leq t, D = \text{other})$ is obtained. The crude analogue of the cure fraction in (2), i.e., the probability of eventually dying from cancer, is computed as,

$$\mathrm{P}(D = \text{cancer}) = \int_0^\infty S^*(u)R(u)\lambda(u)du. \tag{11}$$

Using (11) and (11), dynamic probability of death from other causes than cancer given survival until time t can be computed by,

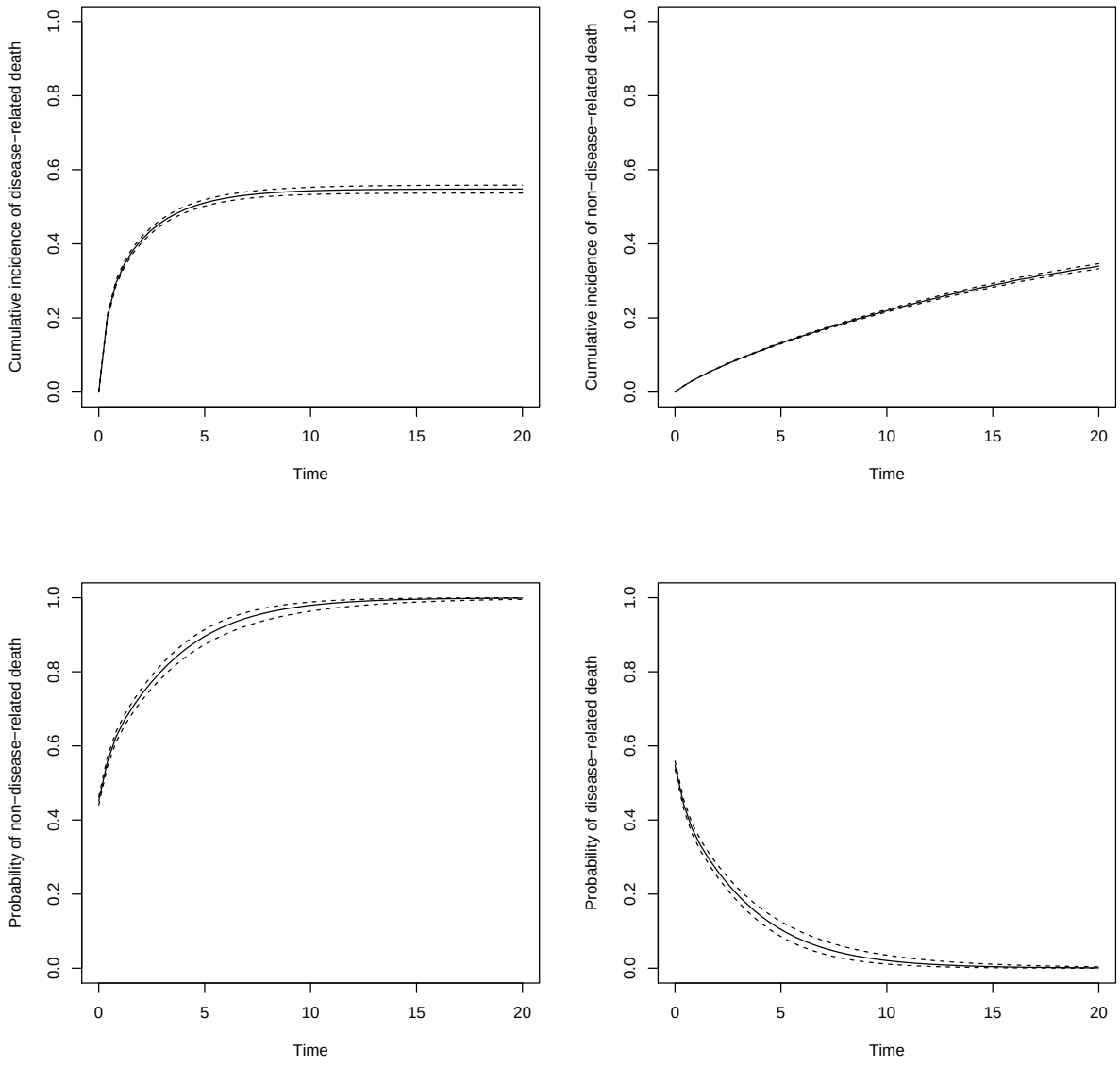
$$\mathrm{P}(D = \text{other}|T > t) = 1 - \frac{\mathrm{P}(D = \text{cancer}, T > t)}{\mathrm{P}(T > t)} \tag{12}$$

$$= \frac{\mathrm{P}(D = \text{cancer}) - \mathrm{P}(T \leq t, D = \text{cancer})}{1 - \mathrm{P}(T \leq t, D = \text{other}) - \mathrm{P}(T \leq t, D = \text{cancer})}. \tag{13}$$

Eloranta et al. [4] used the model by Andersson et al. [2], where a constant relative survival was assumed after the last knot of the splines. This replaces the ∞ in the intgral of (11) by the value of the last knot.

For illustration, we fit a null model, i.e., without any covariates using the `FlexCureModel` function. The crude probabilities are estimated through the `calc.Crude` function. The function has an argument `type` which denotes which type of crude probability to compute; `type = "cancer"` (default) computes (11), `type = "other"` computes (11) with $\lambda(t)$ replaced by $h^*(t)$, and `type = "othertime"` computed (13). If `reverse = TRUE`, $1 - \mathrm{P}(D = \text{other}|T > t) = \mathrm{P}(D = \text{cancer}|T > t)$ is computed.

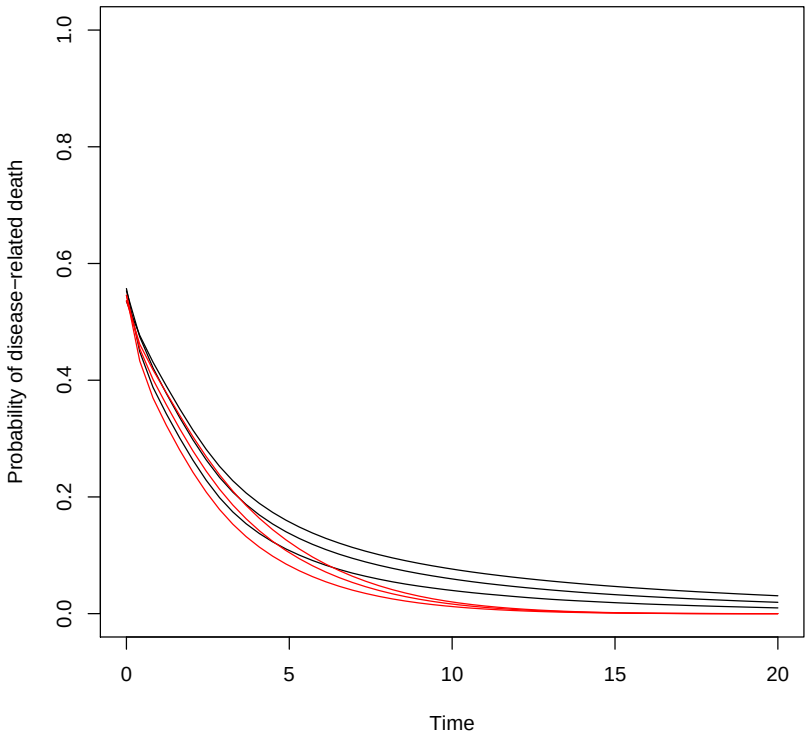
```
fit.null <- GenFlexCureModel(Surv(FUyear, status) ~ 1, data = colonDC, df = 3, bhazard = "bhaz")

## Finding initial values... Completed!
## Fitting the model... Completed!

time.points <- seq(0, 20, length.out = 50)

crude.null.cancer <- calc.Crude(fit.null, time = time.points,
                                var.type = "ci", rmap = list(age = agedays, year = dx))

crude.null.other <- calc.Crude(fit.null, type = "other", time = time.points, var.type = "ci",
                               rmap = list(age = agedays, year = dx))

crude.null.prob <- calc.Crude(fit.null, type = "condother", time = time.points,
                              var.type = "ci", rmap = list(age = agedays, year = dx))

crude.null.prob.reverse <- calc.Crude(fit.null, type = "condother", time = time.points,
                                      reverse = T, var.type = "ci", rmap = list(age = agedays, year = d


par(mfrow = c(2,2))
plot(crude.null.cancer)
plot(crude.null.other)
plot(crude.null.prob)
plot(crude.null.prob.reverse)
```



Using the previously fitted models with a time-varying age effect, we compute the crude probability of cancer related death.

```
newdata <- data.frame(age = c(50, 60, 70),
                      sex = "male",
                      year = 2010,
                      age_days = c(50, 60, 70) * 365.24)

crude.time <- calc.Crude(fit.flex.time, type = "condother", newdata = newdata,
                         time = time.points, reverse = T,
                         var.type = "n", rmap = list(age = age_days))

crude.pen.time <- calc.Crude(fit.stpm2, type = "condother", newdata = newdata,
                             time = time.points, reverse = T,
                             var.type = "n", rmap = list(age = age_days))
```

```r
plot(crude.time)
plot(crude.pen.time, add = T, col = 2)
```



References

[1] Therese M-L Andersson, Paul W. Dickman, Sandra Eloranta, Mats Lambe, and Paul C. Lambert. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. *Statistics in Medicine*, 32(30):5286–5300, dec 2013.

[2] Therese ML Andersson, Paul W Dickman, Sandra Eloranta, and Paul C Lambert. Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. *BMC Medical Research Methodology*, 11(1):96, dec 2011.

[3] Mark Clements and Xing-Rong Liu. *rstpm2: Generalized Survival Models*, 2016. URL `https://CRAN.R-project.org/package=rstpm2`. R package version 1.3.4.

[4] Sandra Eloranta, Paul C. Lambert, Therese M.-L. Andersson, Magnus Björkholm, and Paul W. Dickman. The Application of Cure Models in the Presence of Competing Risks. *Epidemiology*, 25(5): 742–748, sep 2014.

[5] Lasse Hjort Jakobsen, Therese ML Andersson, Jorne Biccler, Tarec Christoffer El-Galaly, and Martin Bgsted. Estimation of the loss of lifetime function using flexible parametric relative survival models. *BMC Medical Research Methodology*, 2017.

[6] P. C. Lambert, P. W. Dickman, C. P. Nelson, and P. Royston. Estimating the crude probability of death due to cancer and other causes using relative survival models. *Statistics in Medicine*, 29(7-8): 885–895, mar 2010.

[7] Paul C. Lambert. Modeling of the cure fraction in survival studies. *Stata Journal*, 7(3):351–375, 2007.

[8] Paul C Lambert, John R Thompson, Claire L Weston, and Paul W Dickman. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, 8(3):576–94, 2007.

[9] Christopher P. Nelson, Paul C. Lambert, Iain B. Squire, and David R. Jones. Flexible parametric models for relative survival, with application in coronary heart disease. *Statistics in Medicine*, 26 (30):5486–5498, dec 2007.

[10] Patrick Royston and Mahesh K. B. Parmar. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21(15):2175–2197, aug 2002.