WCE package: weighted cumulative exposure models

Marie-Pierre Sylvestre*, Marie-Eve Beauchamp†, Michal Abrahamowicz‡

January 8, 2015

Abstract

The `WCE` package implements the weighted cumulative exposure model, a flexible method for modeling cumulative effects of time-varying exposures, weighted by recency, in the Cox proportional hazards model (Sylvestre & Abrahamowicz, 2009). The weight function that assigns weights to past exposures (e.g. dose or intensity of exposure), as a function of the time-since-exposure, is estimated using cubic regression splines. This vignette illustrates the use of the functions available in the `WCE` package on a simulated cohort. The current implementation of `WCE` requires that the dataset is in an interval format, using the counting process notation, in which each row corresponds to one time unit. The `WCE` function allows for covariates, whether fixed-in-time or time-dependent. For additional statistical details on the estimation of the WCE model and its validation in simulations, please see (Sylvestre & Abrahamowicz, 2009). For comparisons with alternative time-varying models, please see (Abrahamowicz, Beauchamp & Sylvestre, 2012).

1 Data set

The dataframe `drugdata` is in an interval format with one row corresponding to a day. `Id` identifies individuals. `Start` and `Stop` identify the beginning and the end of each interval, respectively (`Stop` in row `j` = `Start` in row `j+1`). The intervals are closed on the right, i.e. the value of a time-dependent covariate for the interval `(j; j+1]` represents its value at time `(j+1)`. `Event` is a binary indicator for the event of interest, which takes the value of 1 if the event occured in the interval specified by `Start` and `Stop`. For a given subject, `event=1` can only occur in the last interval of his or her follow-up. The last three columns represent two fixed-in-time covariates of interest (`sex` and baseline `age`) as well as the exposure of interest (dose of a `drug`). Note that time-dependent covariates can also be included.

For example, the lines of data for Id 23 are as followed:

```
data(drugdata)
subset(drugdata, drugdata$Id==23)
```

```
 Id Event Start Stop sex age drug
 23     0     0    1   0  39    0
 23     0     1    2   0  39    0
 23     0     2    3   0  39    2
 23     0     3    4   0  39    2
 23     0     4    5   0  39    2
 23     1     5    6   0  39    2
```

It illustrates that subject 23 was unexposed for the first two days of follow-up and then exposed until the end of follow-up, at a dose of 2.

The dataframe `drugdata` has 77,038 observations from 500 different Ids.

*Department Social and Preventive Medicine, Université de Montréal; CHUM Research Centre

†Division of Clinical Epidemiology, McGill University Health Centre

‡Department of Epidemiology, Biostatistics and Occupational Health, McGill University; Division of Clinical Epidemiology, McGill University Health Centre

```
 nrow(drugdata)
```

```
[1] 77038
```

```
 length(unique(drugdata$Id))
```

```
[1] 500
```

The start times for each `Id` corresponds to 0:

```
 table(by(drugdata$Start, drugdata$Id, min))
```

```
  0
500
```

The current implementation of the `WCE` function does not allow the use of cohorts with delayed entry. While the start times do not have to correspond to 0, they have to be the same for all users.

In the dataframe drugdata, the maximum length of follow-up is 365:

```
 max(drugdata$Stop)
```

```
[1] 365
```

There are no missing values in the dataframe:

```
 apply(is.na(drugdata)==1, 2, sum)
```

```
   Id Event Start  Stop   sex   age  drug
    0     0     0     0     0     0     0
```

Note that the current implementation of `WCE` does not allow for missing data in the variables `Id`, `Start`, `Stop`, `Event` and `expos`.

The exposure can take 6 different values and varies over time:

```
 table(drugdata$drug)
```

```
    0   0.5     1   1.5     2   2.5     3
49460  4683  3978  5118  4609  4674  4516
```

2 Estimation of the weighted cumulative exposure model

2.1 Preparing the input for `WCE`

The function `WCEcheck` allows users to verify that the arguments passed to the `WCE` function are correctly specified, to avoid errors:

```
 checkWCE(drugdata, id = "Id", event = "Event", start = "Start",
        stop = "Stop", expos = "drug")
```

```
Data are in the right format for WCE estimation.
```

Arguments include a dataframe `drugdata` and the name of the variables required for `WCE`. `id` corresponds to the variable identifying subjects, `event` corresponds to an event indicator that must be coded 1 = event and 0 = no event, `start` corresponds to the variable identifying the starting time for the interval, `stop` corresponds to the variable identifying the ending time for each interval, and `expos` corresponds to the exposure variable. Note that `start` and `stop` respectively correspond to the `time` and `time2` arguments in function `Surv` in survival package. Covariates are excluded from calls to the `WCEcheck` function.

In cases in which the dataframe `drugdata` is not set up properly, for example if the exposure is mislabeled (e.g. `dose` instead of `drug`), then `checkWCE` returns an (hopefully) informative error message:

```
checkWCE(drugdata, id = "Id", event = "Event", start = "Start",
        stop = "Stop", expos = "dose")
```

```
At least one of id, event, start, stop, or expos is missing from the dataset supplied.
```

It might be advantageous to run `checkWCE` before estimating the WCE model as error messages from `checkWCE` tend to be more informative than those of `WCE` and `WCE` is significantly more computationally intensive than `checkWCE` for large data sets.

2.2 Estimating `WCE`

The function `WCE` estimates the weighted cumulative exposure model described in Sylvestre & Abrahamowicz (2009).

```
wce.obj <- WCE(data = drugdata, analysis = "Cox", nknots = 1:3, cutoff = 90,
  constrained = "R",  int.knots = NULL, aic = FALSE, MatchedSet = NULL,
  id = "Id", event = "Event", start = "Start", stop = "Stop", expos = "drug",
  covariates = c("sex", "age"))
```

The arguments `drugdata`, `id`, `event`, `start` and `stop` are the same as those of the function `WCEcheck` described above. The second argument of `WCE`, `analysis`, needs to be set to `"Cox"` to estimate the WCE within the proportional hazards model, as illustrated in Sylvestre & Abrahamowicz (2009). It is currently the only model implemented in the `WCE` package. Extensions to other models or designs such as the nested case-control are under development.

Arguments `nknots`, `cutoff`, `constrained` and `int.knots` specify features of the weight function to be estimated. `nknots` determines the number of interior knots for the spline function estimating the weight function. Notation `1:3` implies that 3 alternative weight functions with respectively 1,2 and 3 interior knots will be estimated. Based on our experience, it is unlikely that > 3 knots will be useful and often, especially for datasets with < 200 events, 1 or 2 knots are sufficient. Notation `1:1` indicates that only 1 knot is considered. `cutoff` specifies the length in time units for the window to which the weight function corresponds (see section 2.2 of (Sylvestre & Abrahamowicz, 2009) for details). `constrained` indicates whether the weight function should be constrained or not (see section 2.3 of (Sylvestre & Abrahamowicz, 2009) for details). Weight functions can either be constrained to go smoothly to zero on the right, which corresponds to exposures remote in time (`constrained = "R"`) or on the left, which corresponds to recent exposures (`constrained ="L"`). `int.knots` specify whether the user wants the interior knots to be placed at user-specified values (e.g. `int.knots = c(15,30,60)` indicates the values of the 3 interior knots) or selected by the `WCE` function, which places the knots at quantiles across the time window (`int.knots = NULL`).

`aic` is a logical parameter. If set to `TRUE`, then the AIC is used to select the best fitting model among those estimated. If set to `FALSE`, then the BIC is used instead of the AIC. The default corresponds to FALSE (BIC). Note that the BIC implemented in `WCE` is the version suggested by Volinsky and Raftery in Biometrics (2000), which corresponds to $BIC = 2\log(PL) + p\log(d)$ where PL is the model's partial log-likelihood, p is the number of estimated parameters and d is the the number of uncensored events. See Sylvestre and Abrahamowicz (2009) for more details.

`MatchedSet` is an argument strictly related to the nested case-control option of the argument `analysis`, which remains to be implemented. Thus, for the current implementation, it has to be always set to `NULL`. The argument `covariates` is used to indicate the names of the covariates to be included in the models.

The function `WCE` returns a `WCE` object for which `print`, `plot` and `summary` methods exist.

The `print` method associated with the `WCE` object generated by the call to the `WCE` function above returns the following information. Note that for each row of the matrix reporting the estimated WCE functions, the consecutive columns show the estimated values of the weight functions for each day within the time window, from t1 (1 day elapsed) to the cutoff window (here t90 or 90 days in the past).

```
 wce.obj

Estimated right-constrained WCE function(s).

               t1      t2      t3      t4      t5      t6      t7      t8
1 knot(s) 0.03288 0.03157 0.03032 0.02912 0.02797 0.02688 0.02583 0.02483
2 knot(s) 0.02155 0.02286 0.02393 0.02480 0.02546 0.02592 0.02621 0.02633
3 knot(s) 0.03843 0.03479 0.03162 0.02887 0.02653 0.02456 0.02293 0.02162
...
                t89       t90
1 knot(s) 1.853e-05 2.660e-06
2 knot(s) 6.625e-05 9.555e-06
3 knot(s) 1.799e-04 2.604e-05

Number of events:
[1] 383

Partial log-Likelihoods:
     1 knot(s) 2 knot(s) 3 knot(s)
[1,] -1896.541 -1894.843 -1892.303

BIC:
     [,1] [,2] [,3]
[1,] 3823.821 3825.375 3826.242

Matrix of coefficients estimates for the covariates:

             sex     age
1 knot(s) 0.6805 0.01144
2 knot(s) 0.6844 0.01144
3 knot(s) 0.6901 0.01151

Matrix of standard error estimates for the covariates:

             sex      age
1 knot(s) 0.1187 0.003969
2 knot(s) 0.1188 0.003968
3 knot(s) 0.1189 0.003974

If you report these results, please cite
Sylvestre MP, Abrahamowicz M. Flexible Modeling of the Effects of Time-Dependent Exposures
on the Hazard. Statistics in Medicine 2009; 28(27):3437-3453.
```

3 Methods for WCE objects

Summary and plot methods are available for the WCE object produced by the function `WCE`.

3.1 Summary

The default option of the summary method for the WCE object selects the best model estimated by the `WCE` function if the user specified more than 1 value for the number of interior knots using the `nknots` argument. The summary method provides the coefficients estimates for the covariates (if any) included in the WCE model, the values of the partial log-likelihood, the BIC or AIC (depending on what the user specified), and the number of events used in the estimated model(s).

```
summary(wce.obj)

*** Right-constrained estimated WCE function  (Proportional hazards model).***

Estimated coefficients for the covariates:
      coef exp(coef) se(coef)     z     p
sex 0.6805     1.975   0.1187 5.731 0.000
age 0.0114     1.012   0.0040 2.881 0.004

Partial log-likelihood: -1896.542  BIC: 3823.821

Number of events: 383

Use plot(wce.obj) to see the estimated weight function corresponding to this model.

If you report these results, please cite
Sylvestre MP, Abrahamowicz M. Flexible Modeling of the Effects of Time-Dependent Exposures
on the Hazard. Statistics in Medicine 2009; 28(27):3437-3453.
```

Alternatively, by setting `all = TRUE` in the `summary` function, one can obtain the same information for each model (with different numbers of knots) estimated in the `WCE` function.

3.2 Plot

A plot method exists for WCE objects. Figure 1 shows the results of `plot(wce.obj)`, which corresponds to the plot of the best-fitting estimated weight function (here with 1 interior knot, corresponding to the minimum BIC).

Similarly to the summary method, the default setting for `plot` is to plot the best fitting model as determined by BIC or AIC. A single plot of all fitted models can be obtained using `plot(wce.obj, all = TRUE)`, as shown in Figure 2.

4 Obtaining hazard ratios for a WCE model

To facilitate interpretation of the WCE model results, the user may use the `HR.WCE` function to obtain estimates of the hazard ratio (HR), adjusted for the covariates included in the model. A HR can be obtained from a WCE model by comparing any two (arbitrarily selected) exposure vectors or exposure scenarios (e.g. dosing regimens in the case of drug studies). Scenarios are represented by vectors of the same length as the window value selected in the WCE estimation (cutoff) and indicate the value of exposure (e.g. drug dose) at each time units within the window. The first value of the vector represents the exposure today (t0) and subsequent values represent the exposures in the past (e.g. at 1, 2, ... up to 90 days ago, i.e. at t1,...t90). Selected examples are presented here, using the `HR.WCE` function and assuming cut-off= 90 days.

Example 1: Comparing continuous users (with exposure = 1 at each of the past 90 days) vs. non-users (exposure = 0 for the past 90 days):

```
nonusers <- rep(0, 90)
users <- rep(1, 90)
# for all models
HR.WCE(wce.obj, users, nonusers, all = TRUE)

            HR
1 knot(s) 2.496
2 knot(s) 2.620
3 knot(s) 2.825
```

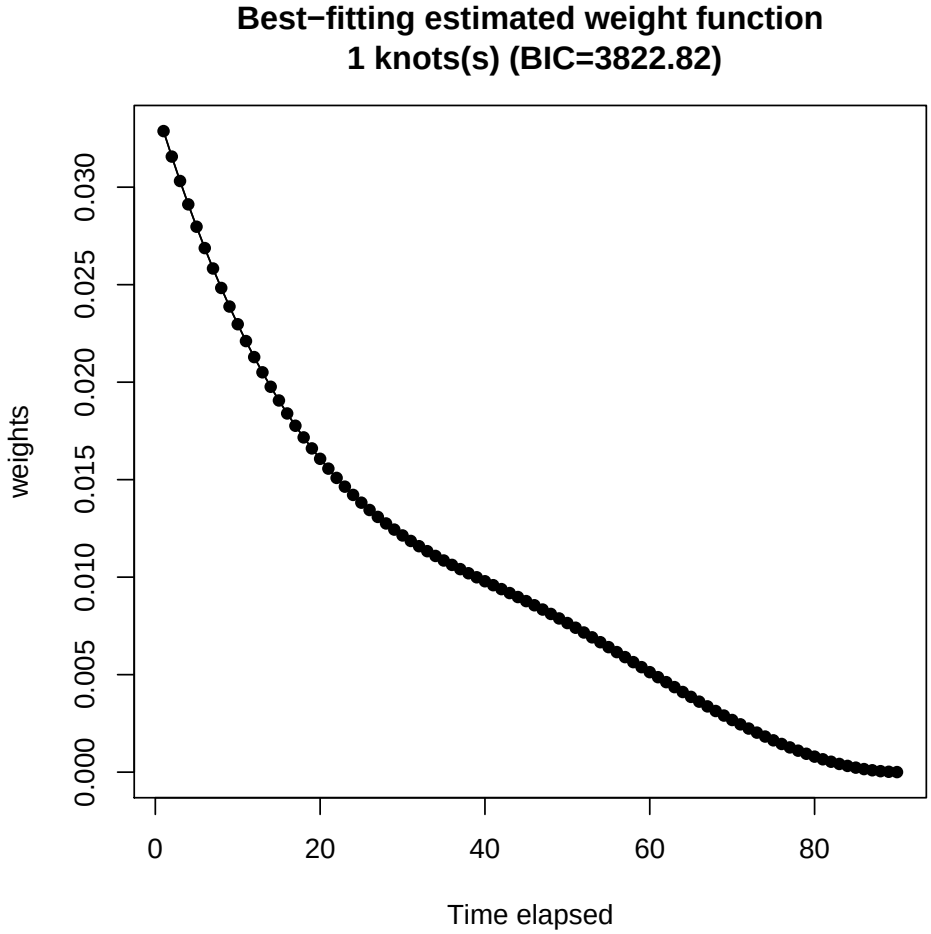

Figure 1: Plot of the best-fitting weight function in a WCE object

```
# for the best model only
HR.WCE(wce.obj, users, nonusers)
```

```
     [,1]
[1,] 2.496
```

The above estimate indicates that subjects exposed to 1 mg dose of the drug across past 90 days have about 2.5 higher hazard than those who did not use the drug at all in the past 90 days.

Example 2: Comparing high exposure (dose = 2, for the past 90 days) to light exposure (dose = 0.5, for the past 90 days):

```
light <- rep(0.5, 90)
heavy <- rep(2, 90)
HR.WCE(wce.obj, heavy, light, all = TRUE)
```

```
            HR
1 knot(s) 3.944
2 knot(s) 4.241
3 knot(s) 4.747
```

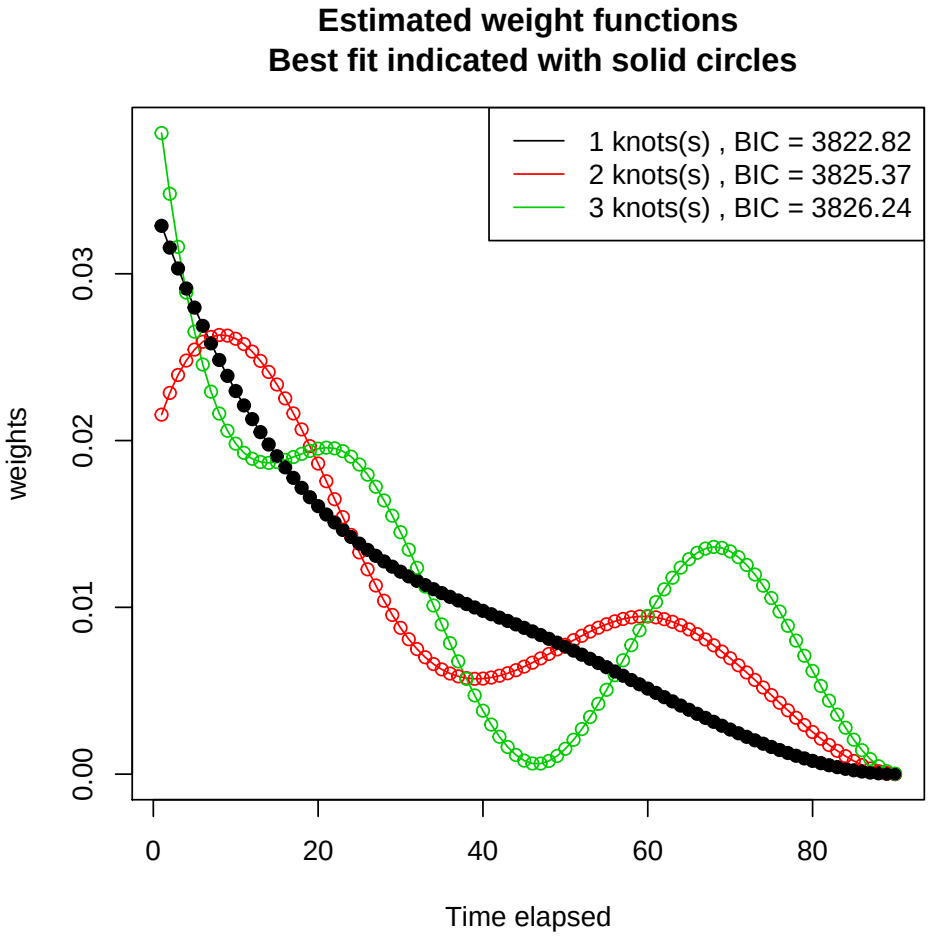

Figure 2: Plot of all of the alternative estimated weight functions (with 1 to 3 interior knots) in a WCE object

Example 3: Comparing current use (dose = 1 for 30 the past days) to past use in a more distant past (dose = 1 for 30 days from but 90 to 60 days ago):

```
past <- c(rep(0, 60), rep(1, 30))
current <- c(rep(1, 30), rep(0, 60))
HR.WCE(wce.obj, current, past, all = TRUE)
```

```
             HR
1 knot(s) 1.727
2 knot(s) 1.597
3 knot(s) 1.479
```

Example 4: Comparing different dose and timings: subject 1 (s1: dose = 2 for most recent 10 days, dose = 0 for 50 days (from 60 to 11 days ago), dose = 1 for 30 days (from 90 to 61 days ago) versus subject 2 (s2: dose = 0 for most recent 30 days, dose = 1.5 for 20 days (from 50 to 31 days ago), dose = 0.5 for 40 days (from 90 to 51 days ago):

```
s1 <- c(rep(2, 10), rep(0, 50), rep(1, 30))
s2 <- c(rep(0, 30), rep(1.5, 20), rep(0.5, 40))
```

```
HR.WCE(wce.obj, s1, s2, all = TRUE)
```

```
            HR
1 knot(s) 1.294
2 knot(s) 1.386
3 knot(s) 1.627
```

5 Bootstrap function to obtain pointwise confidence intervals

Confidence intervals for HRs calculated in the previous section, as well as pointwise confidence bands for the estimated weight function, can be obtained via bootstrap. A simple example is given below. More sophisticated bootsrap routines may be used for optimal results.

Set the number of bootstrap resamples (set to 5 for demonstration purposes only, in actual analyses should be much higher, preferably at least 300, minimum B = 100):

```
B <- 5
```

Obtain the list of Id for sampling:

```
ID <- unique(drugdata$Id)
```

Prepare vectors to extract estimated weight function and (if relevant) HRs for each bootstrap resample:

```
boot.WCE <- matrix(NA, ncol = 90, nrow = B) # To store estimated weight functions
boot.HR <- rep(NA, B) # to store estimated HRs
```

Sample IDs with replacement:

```
for (i in 1:B){
datab <- drugdata[drugdata$Id %in% sample(ID, replace = T),] # select obs. in bootstrap sample
mod <- WCE(data = datab, analysis = "Cox", nknots = 1:3, cutoff = 90,
  constrained = "R",  int.knots = NULL, aic = FALSE, MatchedSet = NULL,
        id = "Id", event = "Event", start = "Start", stop = "Stop", expos = "drug",
        covariates = c("sex", "age"))
# return best WCE estimates and corresponding HR
best <- which.min(mod$info.criterion)
boot.WCE [i,] <- mod$WCEmat[best,]
boot.HR[i] <- HR.WCE(mod, rep(1,90), rep(0, 90))}
```

Summarize bootstrap results using percentile method to obtain a 90% confidence interval:

```
# estimated weight functions
apply(boot.WCE, 2, quantile, p = c(0.05, 0.95))
```

```
       [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]    [,8]    [,9]
5%  0.01907 0.01889 0.01870 0.01851 0.01832 0.01798 0.01748 0.01700 0.01654
95% 0.03848 0.03671 0.03503 0.03343 0.03191 0.03046 0.02909 0.02779 0.02657
      [,10]   [,11]   [,12]   [,13]   [,14]   [,15]   [,16]   [,17]   [,18]
5%  0.01610 0.01567 0.01526 0.01487 0.01444 0.01403 0.01364 0.01325 0.01282
95% 0.02541 0.02432 0.02329 0.02233 0.02142 0.02057 0.01986 0.01923 0.01865
```

...

```
       [,83]     [,84]     [,85]     [,86]     [,87]     [,88]     [,89]     [,90]
5%  0.0003246 0.0002467 0.0001788 0.0001211 7.426e-05 3.852e-05 1.432e-05 2.056e-06
95% 0.0006193 0.0004715 0.0003422 0.0002323 1.426e-04 7.411e-05 2.759e-05 3.962e-06
```

```
# estimated HR
 quantile(boot.HR, p = c(0.05, 0.95))
```

```
   5%   95%
1.991 2.895
```

6 Other methods of interest

Finally, the following three methods may be useful to extract several parameters from a WCE object. First, the `coef` method can be used to obtain the estimated regression coefficients for models estimated by the WCE routine. It provides both the coefficients of the so-called artificial time-dependent variables D in Sylvestre and Abrahamowicz (2009), as well as the estimated coefficients for the covariates (if any).

```
 coef.WCE(wce.obj)
```

```
$WCEest
               D1       D2       D3
1 knots(s) 0.03565 0.012925  0.01089       NA      NA
2 knots(s) 0.01827 0.037901 -0.00639  0.01777      NA
3 knots(s) 0.04713 0.008394  0.03360 -0.01411 0.02696

$covariates
             sex     age
1 knot(s) 0.6805 0.01144
2 knot(s) 0.6844 0.01144
3 knot(s) 0.6901 0.01151
```

Similarly, the `vcov` method returns the variance-covariance matrix of the estimated regression coefficients:

```
 vcov.WCE(wce.obj)
```

```
$`1 knots(s)`
           sex        age         D1         D2         D3
sex  1.410e-02 -1.492e-05  2.170e-05 -7.602e-06  6.563e-05
age -1.492e-05  1.575e-05  8.831e-07 -1.228e-06  1.718e-06
D1   2.170e-05  8.831e-07  1.080e-04 -7.625e-05  4.269e-05
D2  -7.602e-06 -1.228e-06 -7.625e-05  9.258e-05 -6.528e-05
D3   6.563e-05  1.718e-06  4.269e-05 -6.528e-05  7.354e-05

$`2 knots(s)`
           sex        age         D1         D2         D3         D4
sex  1.412e-02 -1.458e-05  6.825e-06  1.769e-05  1.150e-05  4.721e-05
age -1.458e-05  1.574e-05  1.301e-06 -1.051e-06  4.909e-07  7.694e-07
D1   6.825e-06  1.301e-06  2.534e-04 -1.616e-04  8.764e-05 -4.330e-05
D2   1.769e-05 -1.051e-06 -1.616e-04  1.599e-04 -1.055e-04  5.030e-05
D3   1.150e-05  4.909e-07  8.764e-05 -1.055e-04  1.126e-04 -6.393e-05
D4   4.721e-05  7.694e-07 -4.330e-05  5.030e-05 -6.393e-05  6.234e-05

$`3 knots(s)`
           sex        age         D1         D2         D3         D4        D5
```

```
sex  1.414e-02 -1.424e-05  4.577e-05 -2.002e-05  4.662e-05  6.970e-06  4.890e-05
age -1.424e-05  1.579e-05  1.866e-06 -9.165e-07 -1.550e-08  6.185e-07  6.041e-07
D1   4.577e-05  1.866e-06  5.430e-04 -3.194e-04  1.708e-04 -8.106e-05  4.107e-05
D2  -2.002e-05 -9.165e-07 -3.194e-04  2.699e-04 -1.719e-04  8.250e-05 -4.551e-05
D3   4.662e-05 -1.550e-08  1.708e-04 -1.719e-04  1.656e-04 -9.531e-05  5.215e-05
D4   6.970e-06  6.185e-07 -8.106e-05  8.250e-05 -9.531e-05  9.224e-05 -5.698e-05
D5   4.890e-05  6.041e-07  4.107e-05 -4.551e-05  5.215e-05 -5.698e-05  7.097e-05
```

Finally, when the knot placement is not specified by the user but rather selected by the `WCE` routine, it can be obtained from the WCE object by using the `knots` method:

```
knots.WCE(wce.obj)

$`1 knots(s)`
[1] -3 -2 -1  0 46 90 91 92 93

$`2 knots(s)`
 [1] -3 -2 -1  0 31 60 90 91 92 93

$`3 knots(s)`
 [1] -3 -2 -1  0 23 46 68 90 91 92 93
```

7 References

Sylvestre MP, Abrahamowicz M. Flexible Modeling of the Effects of Time-Dependent Exposures on the Hazard. Statistics in Medicine 2009; 28(27):3437-3453.

Abrahamowicz M, Beauchamp ME, Sylvestre MP. Comparison of alternative models for linking drug exposure with adverse effects. Statistics in Medicine 2012; 31(11-12): 1014-1030.

Volinsky CT, Raftery AE. Bayesian information criterion for censored survival models. Biometrics 2000; 56, 256-262.