# HCAL calculation and decomposition of differences in HCAL with R code

## Conventional HE and decomposition of differences in HE

Conventional health expectancy (HE) is a health and mortality summary measure that is often calculated with the Sullivan method (1971). It combines the period survival function with age-specific prevalence data:

$$HE(t) = \int_0^{\omega} l(x,t)\pi(x,t)dx, \tag{1}$$

where $l(x,t)$ denotes the period survival function at age $x$ in time $t$ and $\pi(x,t)$ gives the proportion of being healthy at age $x$ in time $x$.

$$\pi(x,t) = \frac{P_{Healthy}(x,t)}{P_{Total}(x,t)}, \tag{2}$$

The age-specific proportions healthy is usually calculated from survey data, with $P_{Healthy}$ being the number of healthy people at age $x$ in time $t$ and $P_{Total}(x,t)$ corresponding total number of individuals.

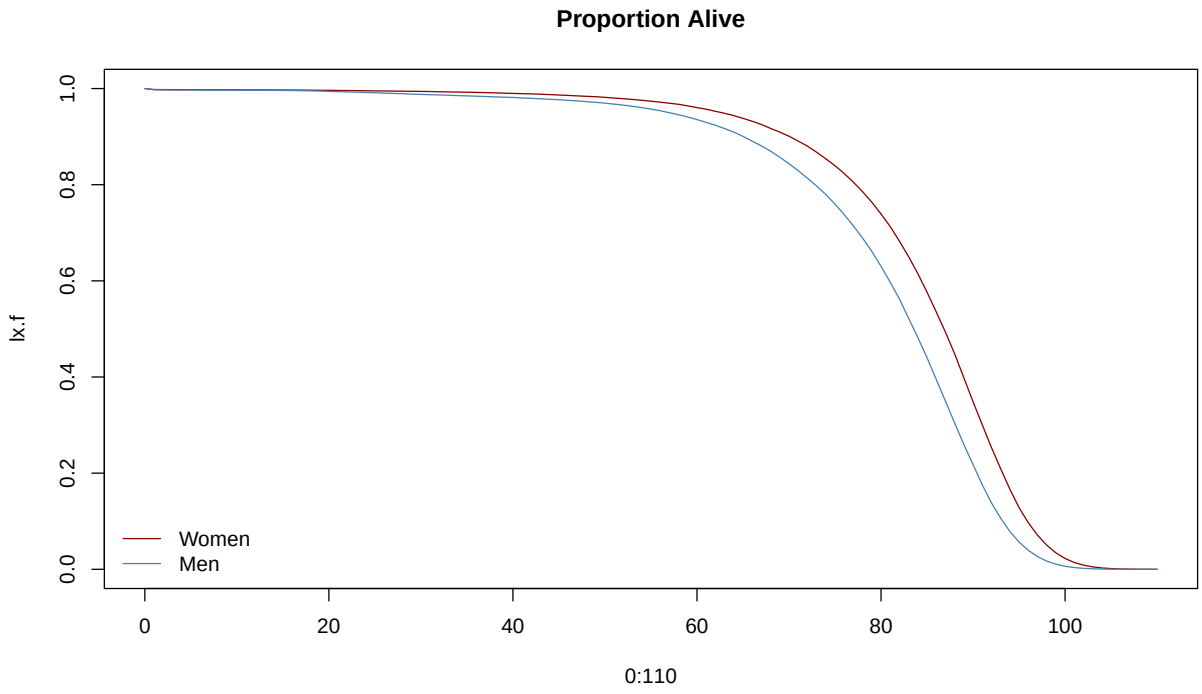
Further, the $l(x,t)$ function comes from the period life table in time $t$ (see e.g., Preston et al. 2001). The Human Mortality Database (HMD) provides period life tables and can be downloaded at www.mortality.org. In this example, we use the period life table for Sweden.

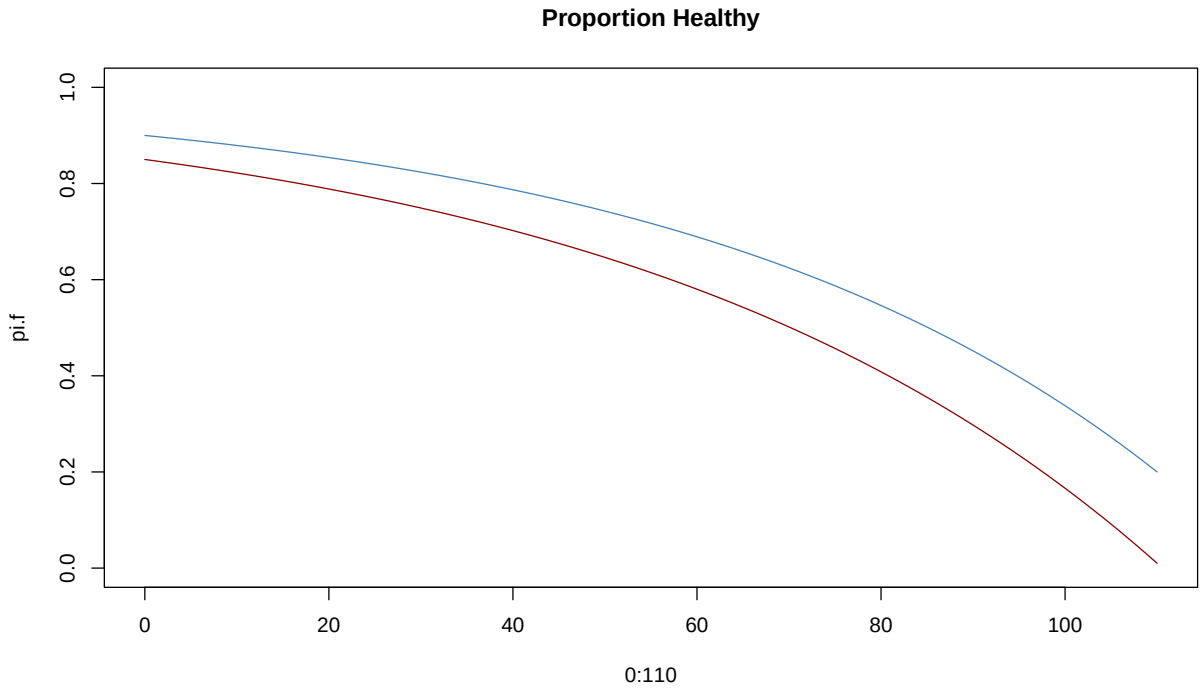
```
setwd("d:/HTCAL/Code/Submission1")

LT.males <- read.table("mltper_1x1.txt", header=TRUE, skip=2)
LT.females <- read.table("fltper_1x1.txt", header=TRUE, skip=2)

lx.m <- LT.males$lx[LT.males$Year==2020]/100000
lx.f <- LT.females$lx[LT.females$Year==2020]/100000
```

For the cross-sectional proportions healthy $\pi(x,t)$, we simply simulate an exponentially decreasing trend. As observed in previous literature and also in our paper, women show a higher prevalence of being unhealthy.

```
pi.m <- 1-exp(seq(log(0.1), log(0.8), length.out = length(lx.m)))
pi.f <- 1-exp(seq(log(0.15), log(0.99), length.out = length(lx.f)))
```

```
par(mfrow=c(2,1))
plot(0:110, lx.f, type="l", main="Proportion Alive", col="darkred")
lines(0:110, lx.m, type="l", col="steelblue")
legend("bottomleft", legend=c("Women","Men"),
       col=c("darkred","steelblue"), lty=1, bty="n")
plot(0:110, pi.f, type="l", main="Proportion Healthy", col="darkred", ylim=c(0,1))
lines(0:110, pi.m, type="l", col="steelblue")
```



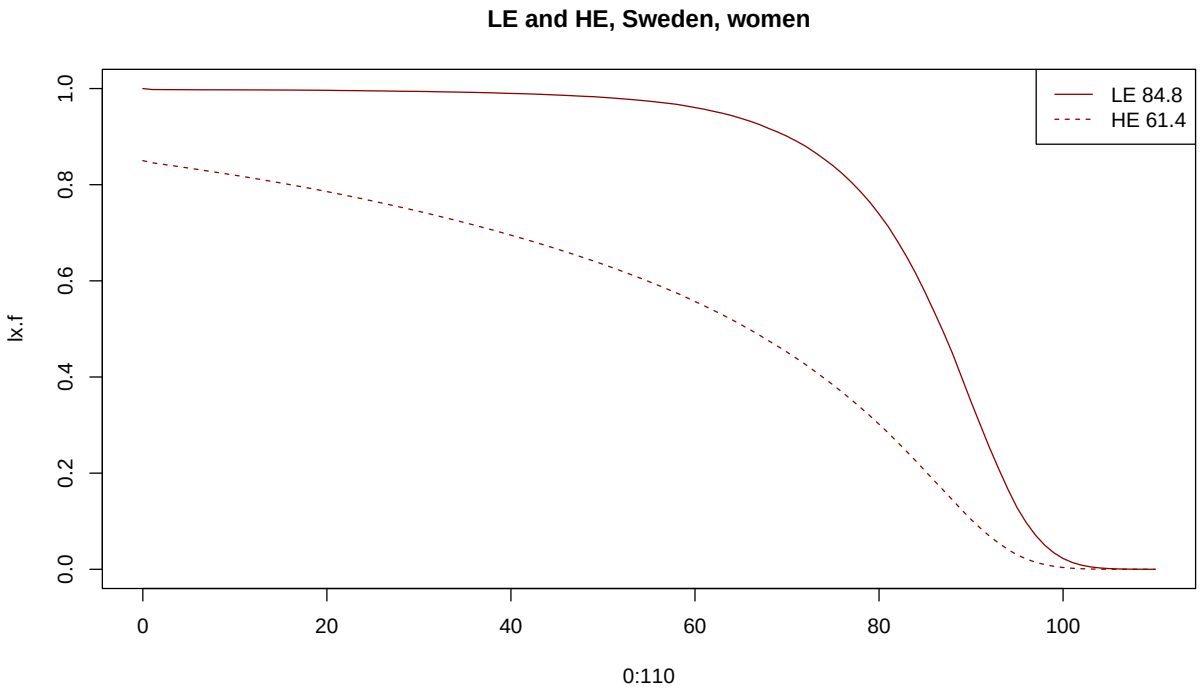

As defined in equation (1), both information, health and mortality, can be summarized in HE.

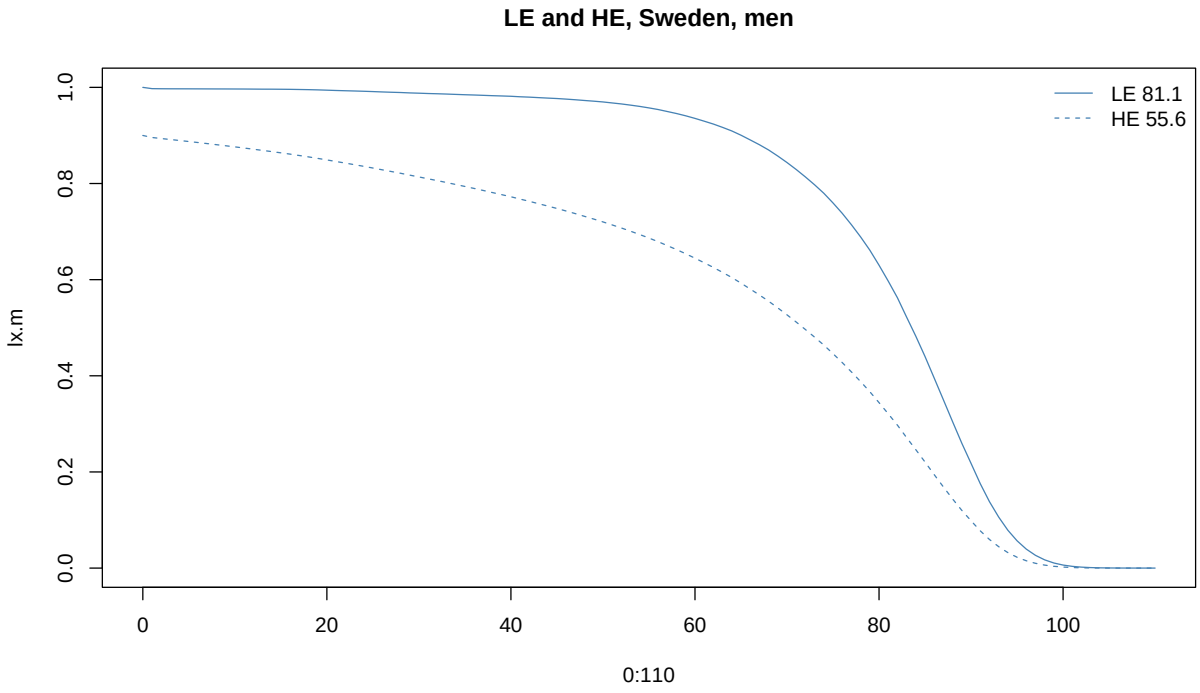
```
HE.m <- sum(lx.f * pi.f) #+0.5
HE.f <- sum(lx.m * pi.m) #+0.5
#We could add a "+0.5" to each HE value
#This is very similar to using the age-specific person-years lived (Lx)
rbind(c("HE.men","HE.women"),round(c(HE.m, HE.f),1))
```

```
##      [,1]     [,2]
## [1,] "HE.men" "HE.women"
## [2,] "55.6"   "61.4"
```

Please note, that the area under the $l(x,t)$ function gives period life expectancy at birth in time $t$ and the area under the health-adjusted function $l^*(x,t) = l(x,t)\pi(x,t)$ defines HE in time $t$.

```
par(mfrow=c(2,1))
plot(0:110, lx.f, type="l", main="LE and HE, Sweden, women", col="darkred")
lines(0:110, lx.f*pi.f, col="darkred", lty=2)
legend("topright", legend=paste(c("LE","HE"),
       round(c(sum(lx.f), HE.f),1)), lty=c(1,2), col="darkred")
plot(0:110, lx.m, type="l", main="LE and HE, Sweden, men", col="steelblue")
lines(0:110, lx.m*pi.m, col="steelblue", lty=2)
legend("topright", legend=paste(c("LE","HE"),
       round(c(sum(lx.m), HE.m),1)), lty=c(1,2), col="steelblue", bty="n")
```




The difference in HE between Swedish women and men is 5.8 years (61.4 − 56.6). Researchers are often interested in investigating to which extent this difference stems from differences in health and mortality. The decomposition introduced by Nusselder and Looman (2004), allows identifying the contributions from each component. Using the notation of a dot on top of the variable to denote derivatives then the difference

between two HE(t) can be calculated as:

$$H\dot{E}(t) = \int_0^{\omega} l(\dot{x},t)\pi(x,t)dx + \int_0^{\omega} l(x,t)\pi(\dot{x},t)dx, \tag{3}$$

where the term on the left side correspond to the changes in HE due to mortality and the right term gives the change in terms of health.

```
delta.lx <- lx.f - lx.m
delta.pi <- pi.f - pi.m
average.lx <- c(lx.f + lx.m) / 2
average.pi <- c(pi.f + pi.m) / 2

Mort.term <- sum(delta.lx * average.pi)
Health.term <- sum(average.lx * delta.pi)
Total <- Mort.term + Health.term

rbind(Mort.term,
      Health.term,
      Total)
```

```
##                   [,1]
## Mort.term    1.755021
## Health.term -7.528754
## Total       -5.773733
```

```
HE.f - HE.m
```

```
## [1] 5.773733
```

The results indicate that men show higher mortality as compared to women but the survival advantage is compensated by lower prevalence rates among men.

## HCAL and previous decomposition of differences in HCAL

In contrast to HE(t), HTCAL(t) does not rely on the period perspective. It does not use the period survival function ($l(x,t)$), but combines prevalence data with the alternative mortality indicator "Cross-Sectional Average Length of Life" (CAL) (Brouard 1986; Guillot 2003). As described by Sauerberg, Guillot, and Luy (2020), it is methodologically more consistent to use CAL because it adjusts the proportions of being healthy with the mortality that individuals have been actually subjected. Let CAL(t) be calculated as:

$$CAL(t) = \int_0^{\omega} l_c(x, t-x)dx, \tag{4}$$

where $l_c(x, t-x)$ gives the probability of surviving to age $x$ for cohorts being born in time $t-x$. The estimation of $l_c(x, t-x)$ requires a long time series of mortality data. This is demonstrated with HMD data for Sweden in the next R code chunk.

```
library(dplyr)
#Select the years 1920 to 2020 and ages 0 to 100
HCAL.f <- filter(LT.females[, c(1:2,4)], Year %in% 1920:2020 & Age %in% 0:100)
HCAL.f$Age <- as.numeric(HCAL.f$Age)
HCAL.f$Cohort <- HCAL.f$Year - HCAL.f$Age
HCAL.f <- arrange(HCAL.f, Cohort)

max.year <- 2020

#Each list element will contain the cohort-specific information
```

```r
#on health and mortality
HCAL.f.list <- list()

  for (x in 1:101) {
        HCAL.f.list[[x]] <-
            data.frame(Age=HCAL.f$Age[HCAL.f$Cohort==max.year+1-x],
                       px=1-c(HCAL.f$qx[HCAL.f$Cohort==c(max.year+1-x)]))

    }
#prevalence is again simulated
for (x in 1:101) {
    HCAL.f.list[[x]]$lx <- c(cumprod(HCAL.f.list[[x]]$px))
    HCAL.f.list[[x]]$pi <- 1-c(0,exp(seq(log(0.1), log(0.99),
                                  length.out = 100)))[1:length(HCAL.f.list[[x]]$lx)]
    HCAL.f.list[[x]]$healthy.lx <- HCAL.f.list[[x]]$lx * c(HCAL.f.list[[x]]$pi)
     HCAL.f.list[[x]]$pi.lx <- c(1,
                                 c(HCAL.f.list[[x]]$pi[-1])/
                                   c(HCAL.f.list[[x]]$pi[-length(HCAL.f.list[[x]]$pi)]))
}

HCAL.f.list[[21]]
```

```
##    Age      px        lx        pi healthy.lx     pi.lx
## 1    0 0.99717 0.9971700 1.0000000  0.9971700 1.0000000
## 2    1 0.99966 0.9968310 0.9000000  0.8971479 0.9000000
## 3    2 0.99984 0.9966715 0.8976573  0.8946694 0.9973970
## 4    3 0.99991 0.9965818 0.8952597  0.8921995 0.9973291
## 5    4 0.99991 0.9964921 0.8928059  0.8896740 0.9972592
## 6    5 0.99996 0.9964522 0.8902947  0.8871361 0.9971872
## 7    6 0.99993 0.9963825 0.8877246  0.8845132 0.9971132
## 8    7 0.99996 0.9963426 0.8850943  0.8818572 0.9970370
## 9    8 0.99991 0.9962529 0.8824024  0.8790960 0.9969586
## 10   9 0.99996 0.9962131 0.8796474  0.8763163 0.9968779
## 11  10 0.99996 0.9961732 0.8768279  0.8734725 0.9967947
## 12  11 0.99989 0.9960637 0.8739423  0.8705022 0.9967091
## 13  12 0.99989 0.9959541 0.8709892  0.8674652 0.9966209
## 14  13 0.99996 0.9959143 0.8679668  0.8644205 0.9965300
## 15  14 0.99994 0.9958545 0.8648737  0.8612883 0.9964363
## 16  15 0.99994 0.9957947 0.8617081  0.8580844 0.9963398
## 17  16 0.99990 0.9956952 0.8584683  0.8547727 0.9962403
## 18  17 0.99986 0.9955558 0.8551526  0.8513521 0.9961377
## 19  18 0.99983 0.9953865 0.8517592  0.8478297 0.9960319
## 20  19 0.99981 0.9951974 0.8482864  0.8442124 0.9959227
## 21  20 0.99970 0.9948988 0.8447322  0.8404231 0.9958101
```

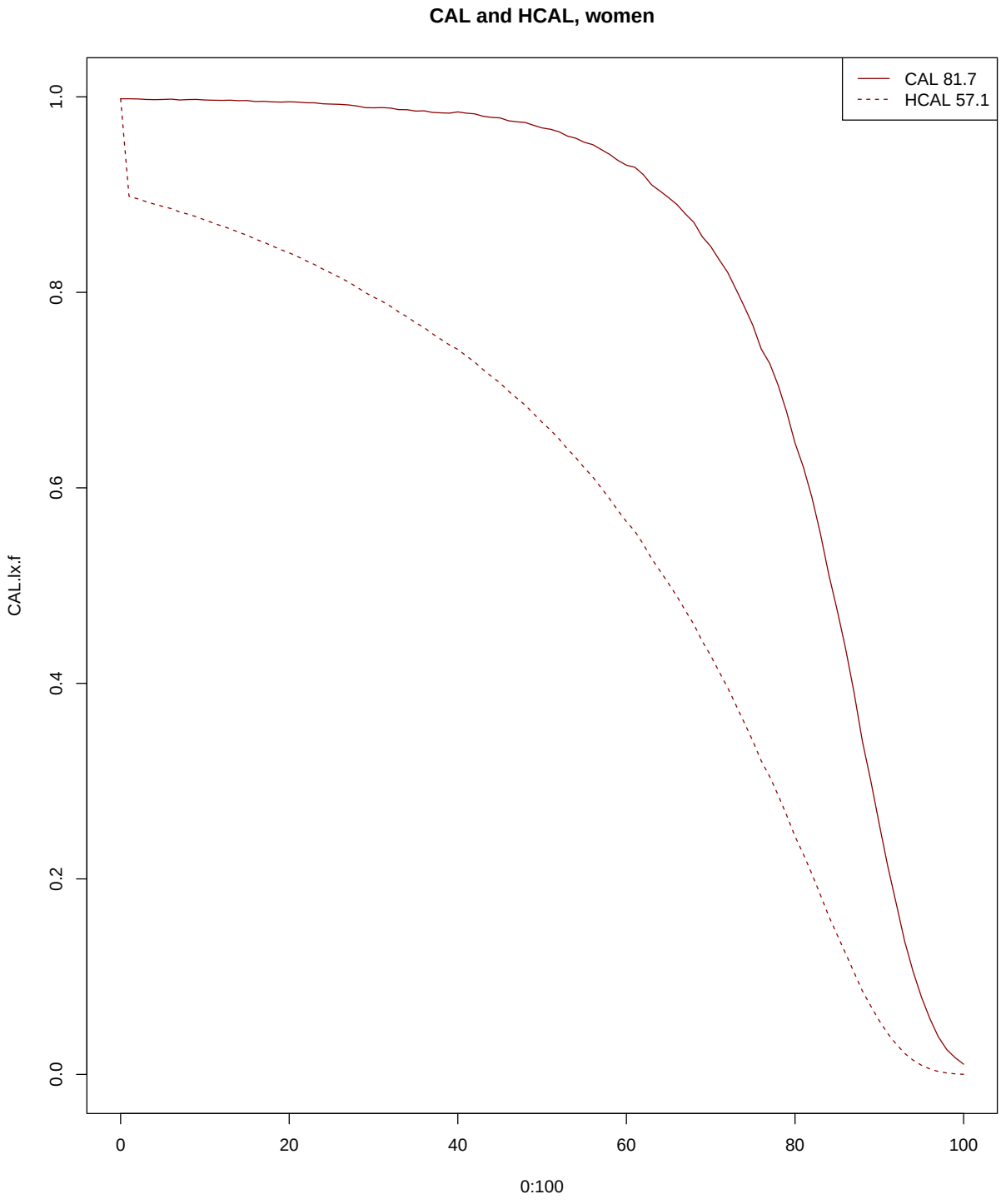
The output shown above, for example, gives the health and mortality information of the cohort born in 1999, reaching age 21 in 2020. The corresponding $l_c(x, t-x)$ value is located in the last row of column $lx$. Please note, this value is a function of the age-specific probabilities of surviving from age 0 to age $x$, $l_c(x, t-x) = {}_1p_0(t){}_1p_1(t)...{}_1p_{x-1}(t)$. Summing them over all cohorts produces CAL (equation 4). Accordingly, HCAL is given by summing only the health-adjusted proportions of cohort survivors,

$$HCAL(t) = \int_0^{\omega} l_c(x, t-x)\pi_c(x, t-x)dx. \tag{5}$$

```r
CAL.lx.f <- c()
pi_c.f <- c()
HCAL.lx.f <- c()

for (x in 1:101) {

  CAL.lx.f[x] <- rev(HCAL.f.list[[x]]$lx)[1]
  pi_c.f[x] <- rev(HCAL.f.list[[x]]$pi)[1]
  HCAL.lx.f[x] <- rev(HCAL.f.list[[x]]$healthy.lx)[1]

}

CAL.f <- sum(CAL.lx.f)
HCAL.f <- sum(CAL.lx.f * pi_c.f)
HCAL.f2 <- sum(HCAL.lx.f)
rbind(round(c(CAL.f, HCAL.f, HCAL.f2),1))
```

```
##      [,1] [,2] [,3]
## [1,] 81.7 57.1 57.1
```

```r
plot(0:100, CAL.lx.f, type="l", main="CAL and HCAL, women", col="darkred", ylim=c(0,1))
lines(0:100, c(CAL.lx.f*pi_c.f), col="darkred", lty=2)
legend("topright", legend=paste(c("CAL","HCAL"),
       round(c(CAL.f, HCAL.f),1)), lty=c(1,2), col="darkred")
```




We can apply the previous shown decomposition technique to HCAL,

$$H\dot{C}AL(t) = \int_0^{\omega} l_c(x, \dot{t} - x)\pi_c(x, t - x)dx + \int_0^{\omega} l_c(x, t - x)\pi_c(x, \dot{t} - x)dx. \tag{6}$$

```
#Copy and paste for men
HCAL.m <- filter(LT.males[, c(1:2,4)], Year %in% 1920:2020 & Age %in% 0:100)
```

```r
HCAL.m$Age <- as.numeric(HCAL.m$Age)
HCAL.m$Cohort <- HCAL.m$Year - HCAL.m$Age
HCAL.m <- arrange(HCAL.m, Cohort)

max.year <- 2020

#Each list element will contain the cohort-specific information
#on health and mortality
HCAL.m.list <- list()

  for (x in 1:101) {
        HCAL.m.list[[x]] <-
            data.frame(Age=HCAL.m$Age[HCAL.m$Cohort==max.year+1-x],
                       px=1-c(HCAL.m$qx[HCAL.m$Cohort==c(max.year+1-x)]))

    }
#prevalence is again simulated
for (x in 1:101) {
    HCAL.m.list[[x]]$lx <- c(cumprod(HCAL.m.list[[x]]$px))
    HCAL.m.list[[x]]$pi <- 1-c(0,exp(seq(log(0.1), log(0.8),
                                  length.out = 100)))[1:length(HCAL.m.list[[x]]$lx)]
    HCAL.m.list[[x]]$healthy.lx <- HCAL.m.list[[x]]$lx * c(HCAL.m.list[[x]]$pi)
     HCAL.m.list[[x]]$pi.lx <- c(1,
                                c(HCAL.m.list[[x]]$pi[-1])/
                                  c(HCAL.m.list[[x]]$pi[-length(HCAL.m.list[[x]]$pi)]))
}

CAL.lx.m <- c()
pi_c.m <- c()
HCAL.lx.m <- c()

for (x in 1:101) {

  CAL.lx.m[x] <- rev(HCAL.m.list[[x]]$lx)[1]
  pi_c.m[x] <- rev(HCAL.m.list[[x]]$pi)[1]
  HCAL.lx.m[x] <- rev(HCAL.m.list[[x]]$healthy.lx)[1]

}

CAL.m <- sum(CAL.lx.m)
HCAL.m <- sum(CAL.lx.m * pi_c.m)
HCAL.m2 <- sum(HCAL.lx.m)
rbind(round(c(CAL.m, HCAL.m, HCAL.m2),1))
```

```
##      [,1] [,2] [,3]
## [1,] 77.5 57.8 57.8
```

```r
#Decomposition
delta.CAL.lx <- CAL.lx.f - CAL.lx.m
delta.pi_c <- pi_c.f - pi_c.m
average.CAL.lx <- c(CAL.lx.f + CAL.lx.m) / 2
average.pi_c <- c(pi_c.f + pi_c.m) / 2

Mort.term.HCAL <- sum(delta.CAL.lx * average.pi_c)
```

```
Health.term.HCAL <- sum(average.CAL.lx * delta.pi_c)
Total.HCAL <- Mort.term.HCAL + Health.term.HCAL

rbind(Mort.term.HCAL,
      Health.term.HCAL,
      Total.HCAL)
```

```
##                        [,1]
## Mort.term.HCAL    1.8689389
## Health.term.HCAL -2.5149683
## Total.HCAL       -0.6460294
```

```
diff(c(HCAL.f, HCAL.m))
```

```
## [1] 0.6460294
```

## New decomposition of differences in HCAL: Age-& cohort-specific contributions

We adapted the same procedure as introduced in Canudas-Romo and Guillot (2015) and decomposed the difference in mortality as,

$$\int_0^{\omega} l_c(x,\dot{t}-x)\pi_c(x,t-x)dx = \int_0^{\omega} l_c(x,t-x)\pi_c(x,t-x)\left[\sum_{i=1}^{x-1}\frac{{}_1\dot{p}_i(t)}{{}_1p_i(t)}\right]dx, \tag{7}$$

where ${}_1p_i(t)$ are the probabilities of surviving from age $i$ to $i+1$. We have demonstrated the relationship between ${}_1p_i(t)$ and $l_c(x,t-x)$ above.

An analogous expression can be found for the healthy proportion element,

$$\int_0^{\omega} l_c(x,t-x)\pi_c(x,\dot{t}-x)dx = \int_0^{\omega} l_c(x,t-x)\pi_c(x,t-x)\left[\sum_{i=1}^{x-1}\frac{{}_1\dot{\rho}_i(t)}{{}_1\rho_i(t)}\right]dx, \tag{8}$$

where ${}_1\rho_i(t)$ change in the proportion of healthy individuals from age $i$ to age $i+1$. As shown in the R example, they can be calculated similar to the life table $lx$ function, ${}_1\rho_i(t) = \frac{\pi_c(i+1,t-x)}{\pi_c(i,t-x)}$. We assume that the proportion of being healthy is equal to 1, or 100% for the first age, $\pi_c(0,t-x)=1$, similar to having a radix for the life table population equal to 1.

```
HCAL.m.list[[20]]
```

```
##    Age      px        lx        pi healthy.lx     pi.lx
## 1    0 0.99601 0.9960100 1.0000000  0.9960100 1.0000000
## 2    1 0.99966 0.9956714 0.9000000  0.8961042 0.9000000
## 3    2 0.99973 0.9954025 0.8978773  0.8937494 0.9976415
## 4    3 0.99994 0.9953428 0.8957096  0.8915381 0.9975857
## 5    4 0.99992 0.9952632 0.8934959  0.8892636 0.9975285
## 6    5 0.99984 0.9951039 0.8912352  0.8868716 0.9974698
## 7    6 0.99994 0.9950442 0.8889265  0.8845211 0.9974095
## 8    7 0.99998 0.9950243 0.8865687  0.8821575 0.9973477
## 9    8 0.99986 0.9948850 0.8841610  0.8796385 0.9972842
## 10   9 0.99994 0.9948253 0.8817021  0.8771396 0.9972190
## 11  10 0.99994 0.9947656 0.8791911  0.8745891 0.9971520
## 12  11 0.99990 0.9946662 0.8766267  0.8719509 0.9970833
## 13  12 0.99990 0.9945667 0.8740079  0.8692591 0.9970126
## 14  13 0.99987 0.9944374 0.8713335  0.8664866 0.9969401
## 15  14 0.99989 0.9943280 0.8686024  0.8636757 0.9968655
## 16  15 0.99983 0.9941590 0.8658132  0.8607560 0.9967889
```

```
## 17  16 0.99987 0.9940297 0.8629649  0.8578128 0.9967102
## 18  17 0.99956 0.9935924 0.8600561  0.8545452 0.9966293
## 19  18 0.99964 0.9932347 0.8570856  0.8512871 0.9965461
## 20  19 0.99938 0.9926189 0.8540520  0.8477481 0.9964606
```

In the R example, the ${}_1\rho_i(t)$ is called "pi.lx" and relates to $\pi_c(x, t-x)$ just like ${}_1p_i(t)$ relates to $l_c(x, t-x)$, $\pi(x, t-x) = {}_1\rho_0(t){}_1\rho_1(t)...{}_1\rho_{x-1}(t)$. The derivatives in equation 7 and 8 can be approximated by taking the log of the ratio of the two functions,

$$\frac{{}_1\dot{p}_i(t)}{{}_1p_i(t)} \approx ln\left[\frac{{}_1p_a(t-x, Men)}{{}_1p_a(t-x, Women)}\right] \tag{9}$$

$$\frac{{}_1\dot{\rho}_i(t)}{{}_1\rho_i(t)} \approx ln\left[\frac{{}_1\rho_a(t-x, Men)}{{}_1\rho_a(t-x, Women)}\right], \tag{10}$$

```
Diff.list <- list()

for (x in 1:101) {
    Diff.list[[x]] <- data.frame(Age = 0:100)

}

for (x in 1:101) {

    Diff.list[[x]]$pxCH <- c(log(HCAL.f.list[[x]]$px/HCAL.m.list[[x]]$px),
                             rep(0,101-x))

    Diff.list[[x]]$pxCH.healthy <- c(log(HCAL.f.list[[x]]$pi.lx/
                                           HCAL.m.list[[x]]$pi.lx),
                                     rep(0,101-x))
}

###getting the matrix px
PxCh <- Diff.list[[1]]$pxCH

for(x in 2:101) {
    PxCh <- cbind(PxCh, Diff.list[[x]]$pxCH)
}

###getting the matrix for pi
PiCh <- Diff.list[[1]]$pxCH.healthy

for(x in 2:101) {
    PiCh <- cbind(PiCh, Diff.list[[x]]$pxCH.healthy)
}

###getting the average functions
CALlx.average <- t(matrix(rep((c(CAL.lx.f * c(pi_c.f)) +
                                 c(CAL.lx.m * c(pi_c.m)))/2,101),101))

Mort.term <- sum(PxCh * CALlx.average)
Dis.term <- sum(CALlx.average * PiCh)
Total <- Mort.term + Dis.term
rbind(Mort.term, Dis.term, Total)
```

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

```
## Mort.term  1.8435365
## Dis.term  -2.4894629
## Total     -0.6459265
```

```
diff(c(HCAL.f, HCAL.m))
```

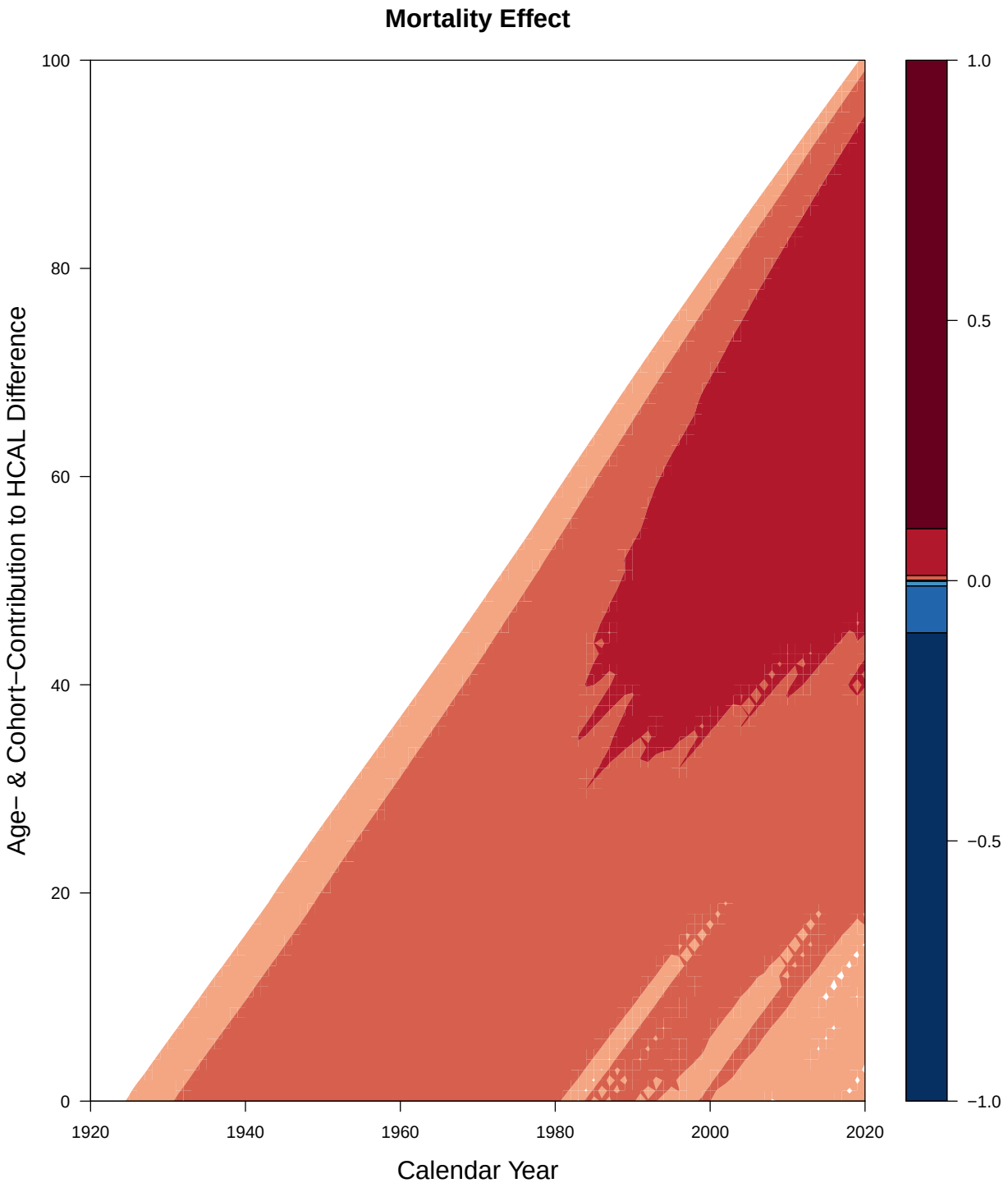
```
## [1] 0.6460294
```

Again, women show lower mortality and higher morbidity rates, resulting in a negative mortality effect and in a positive disability effect. The HCAL approach allows identifying the age- and cohort-specific contributions to each effect.

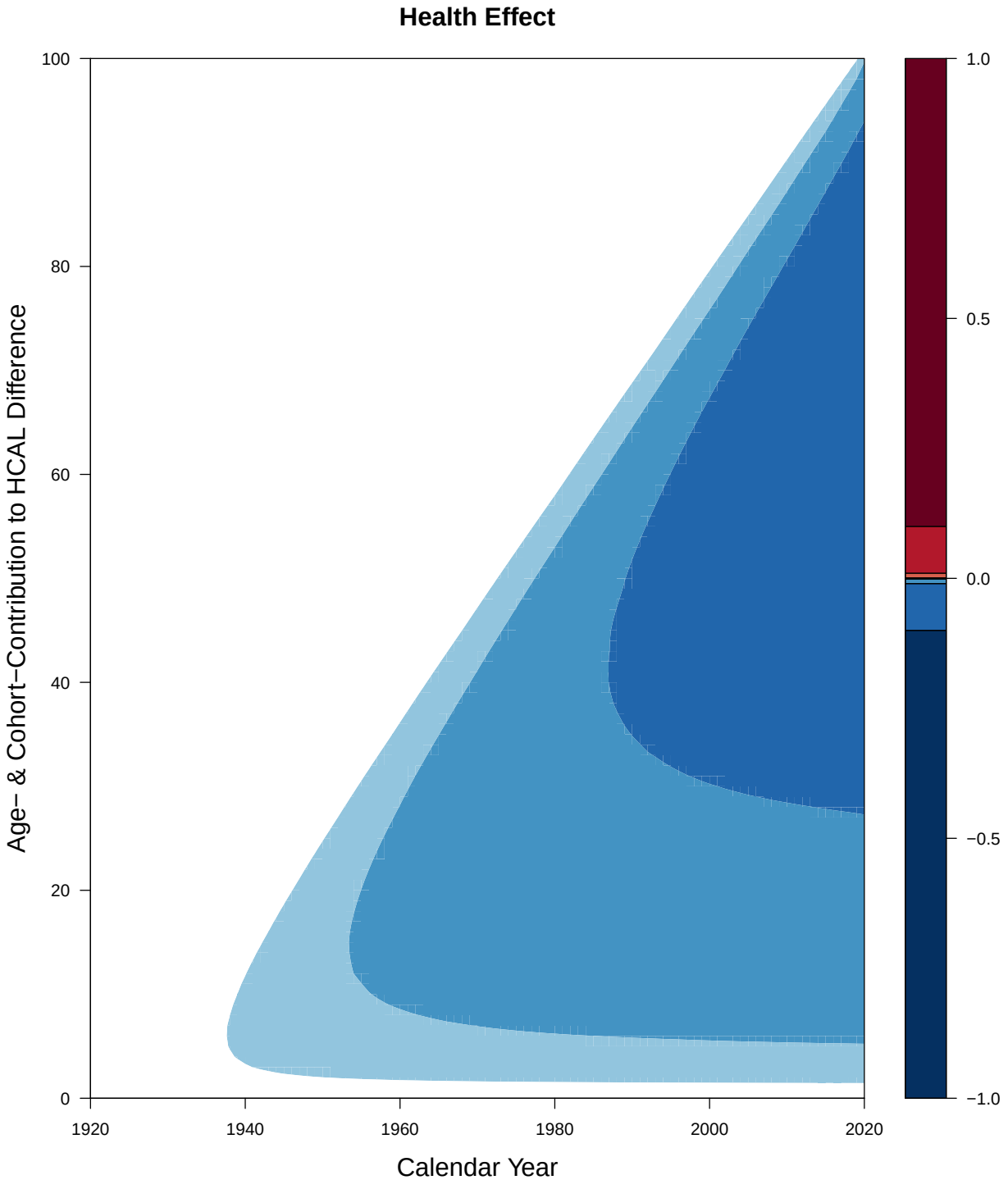
```
###Mort effect
CALlxDecomp <- PxCh * CALlx.average

Age<-c(0:100)

# The correct assignment of contributions and the cummulative changes
CALlxD<-matrix(0,101,101)
CALlxDS<-CALlxD
CALlxD<-CALlxD
CALlxDS<-CALlxDS

YEARS<-c((2020-100):2020)


for (y in 1:101){
        for (x in 1:y){
            CALlxD[x,(101-y+x)]<-CALlxDecomp[x,y]
            CALlxDS[x,(101-y+x)]<-sum(CALlxDecomp[(1:x),y])
        }
}

Mort.matrix <- t(CALlxDS)

###Health effect
CALlxDecomp <- CALlx.average * PiCh


# The correct assignment of contributions and the cummulative changes
CALlxD<-matrix(0,101,101)
CALlxDS<-CALlxD
CALlxD<-CALlxD
CALlxDS<-CALlxDS


for (y in 1:101){
        for (x in 1:y){
            CALlxD[x,(101-y+x)]<-CALlxDecomp[x,y]
            CALlxDS[x,(101-y+x)]<-sum(CALlxDecomp[(1:x),y])
        }
}

Dis.matrix <- t(CALlxDS)

library(RColorBrewer)
```

```
levels<-c(-1,-0.1,-0.01,-0.001,-0.0001,0,.0001,.001,.01,.1,1)
WildColors<-rev(c("#67001f","#b2182b","#d6604d","#f4a582",
                  "white","white",
                  "#92c5de","#4393c3","#2166ac","#053061"))
```

```
filled.contour(xlim=c(1920,2020),YEARS,Age,Mort.matrix,levels=levels,col=WildColors,
               ylab="Age- & Cohort-Contribution to HCAL Difference",xlab="Calendar Year",
               cex.lab=1.5, cex.axis=1.5, cex.main=1.5, cex.sub=1.5,
              main="Mortality Effect")
```




```
filled.contour(xlim=c(1920,2020),YEARS,Age,Dis.matrix,levels=levels,col=WildColors,
                ylab="Age- & Cohort-Contribution to HCAL Difference",xlab="Calendar Year",
                cex.lab=1.5, cex.axis=1.5, cex.main=1.5, cex.sub=1.5,
               main="Health Effect")
```




## The case of a truncated HCAL, THCAL

In our paper, health and mortality data was not available from 1920 to 2020. For this reason, we relied on the truncated version of HCAL, namely THCAL. We used the available information from 1980 to 2019 and set mortality rates as well as morbidity rates before 1980 to zero. This is described in more detail in the supplementary material on data preparation.

## References


- Brouard N. Structure et dynamique des populations: La pyramide des années à vivre, aspects nationaux et exemples régionaux. Espaces Populations Soc. 1986:157–68.
- Canudas-Romo V, Guillot M. Truncated cross-sectional average length of life: a measure for comparing the mortality history of cohorts. Popul Stud. 2015;69:147–59.
- Guillot M. The cross-sectional average length of life (CAL): a cross-sectional mortality measure that reflects the experience of cohorts. Popul Stud. 2003;57:41–54.
- Preston SH, Heuveline P, Guillot M. Demography: Measuring and modeling population processes. 2001. Oxford:Blackwell Publishers.
- Sauerberg, M., Guillot, M. & Luy, M. The cross-sectional average length of healthy life (HCAL): a measure that summarizes the history of cohort health and mortality. Popul Health Metrics 2020;18.