Years of Life Lost (YLL) to disease Diabetes in DK as example

Bendix Carstensen  Steno Diabetes Center, Gentofte, Denmark
& Department of Biostatistics, University of Copenhagen
b@bxc.dk
http://BendixCarstensen.com
Chapter 1
Theory and technicalities

This vignette for the Epi package describes the probabilistic and demographic background for and technical implementation of the erl and yll functions that computes the expected residual life time and years of life lost in an illness-death model.

1.1 Years of life lost (YLL)

...to diabetes or any other disease for that matter.

The general concept in calculation of “years lost to...” is the comparison of the expected lifetime between two groups of persons; one with and one without disease (in this example DM). The expected lifetime is the area under the survival curve, so basically the exercise requires that two survival curves that are deemed relevant be available.

The years of life lost is therefore just the area between the survival curves for those “Well”, \( S_W(t) \), and for those “Diseased”, \( S_D(t) \):

\[
YLL = \int_0^\infty S_W(t) - S_D(t) \, dt
\]

The time \( t \) could of course be age, but it could also be “time after age 50” and the survival curves compared would then be survival curves conditional on survival till age 50, and the YLL would be the years of life lost for a 50-year old person with diabetes.

If we are referring to the expected lifetime we will more precisely use the label expected residual lifetime, ERL.

1.2 Constructing the survival curves

YLL can be computed in two different ways, depending on the way the survival curve and hence the expected lifetime of a person without diabetes is computed:

- Assume that the “Well” persons are immune to disease — using only the non-DM mortality rates throughout for calculation of expected life time.

- Assume that the “Well” persons can acquire the disease and thereby see an increased mortality, thus involving all three rates shown in figure 1.1.
The former gives a higher YLL because the comparison is to persons assumed immune to DM (and yet with the same mortality as non-immune prior to diagnosis), the latter gives a more realistic picture of the comparison of group of persons with and without diabetes at a given age that can be interpreted in the real world.

The differences can be illustrated by figure 1.1; the immune approach corresponds to an assumption of $\lambda(t) = 0$ in the calculation of the survival curve for a person in the “Well” state.

Calculation of the survival of a diseased person already in the “DM” state is unaffected by assumptions about $\lambda$.

We can edit the output from boxes

![Diagram of illness-death model](image)

Figure 1.1: Illness-death model describing diabetes incidence and mortality.

1.2.1 Total mortality — a shortcut?

A practical crude shortcut could be to compare the ERL in the diabetic population to the ERL for the entire population (that is use the total mortality ignoring diabetes status).

Note however that this approach also counts the mortality of persons that acquired the disease earlier, thus making the comparison population on average more ill than the population we aim at, namely those well at a given time, which only then become more gradually ill.

How large these effects are can however be empirically explored, as we shall do later.

1.2.2 Disease duration

In the exposition above there is no explicit provision for the effect of disease duration, but if we were able to devise mortality rates for any combination of age and duration, this could be taken into account.
There are however severe limitations in this as we in principle would want to have duration effects as long as the age-effects — in principle for all \((a, d)\) where \(d \leq A\), where \(A\) is the age at which we condition. So even if we were only to compute ERL from age, say, 40 we would still need duration effects up to 60 years (namely to age 100).

The incorporation of duration effects is in principle trivial from a computational point of view, but we would be forced to entertain models predicting duration effects way beyond what is actually observed disease duration in any practical case.

1.2.3 Computing integrals

The practical calculations of survival curves, ERL and YLL involves calculation of (cumulative) integrals of rates and functions of these as we shall see below. This is easy if we have a closed form expression of the function, so its value may be computed at any time point — this will be the case if we model rates by smooth parametric functions.

Computing the (cumulative) integral of a function is done as follows:

- Compute the value of the function (mortality rate for example) at the midpoints of a sequence of narrow equidistant intervals — for example one- or three month intervals of age, say.
- Take the cumulative sum of these values multiplied by the interval length — this will be a very close approximation to the cumulative integral evaluated at the end of each interval.
- If the intervals are really small (like 1/100 year), the distinction between the value at the middle and at the end of each interval becomes irrelevant.

Note that in the above it is assumed that the rates are given in units corresponding to the interval length — or more precisely, as the cumulative rates over the interval.

1.3 Survival functions in the illness-death model

The survival functions for persons in the “Well” state can be computed under two fundamentally different scenarios, depending on whether persons in the “Well” state are assumed to be immune to the disease \((\lambda(a) = 0)\) or not.

1.3.1 Immune approach

In this case both survival functions for person in the two states are the usual simple transformation of the cumulative mortality rates:

\[
S_W(a) = \exp \left( - \int_0^a \mu_W(u) \, du \right), \quad S_D(a) = \exp \left( - \int_0^a \mu_D(u) \, du \right)
\]

Conditional survival functions

If we want the conditional survival functions given survival to age \(A\), say, they are just:

\[
S_W(a|A) = S_W(a)/S_W(A), \quad S_D(a|A) = S_D(a)/S_D(A)
\]
1.3.2 Non-immune approach

For a diseased person, the survival function in this states is the same as above, but the survival function for a person without disease (at age 0) is (see figure 1.1):

\[ S(a) = P\{\text{Well}\}(a) + P\{\text{DM}\}(a) \]

In the appendix of the paper [2] is an indication of how to compute the probability of being in any of the four states shown in figure 1.1, which I shall repeat here:

In terms of the rates, the probability of being in the “Well” box is simply the probability of escaping both death (at a rate of \( \mu_W(a) \)) and diabetes (at a rate of \( \lambda(a) \)):

\[ P\{\text{Well}\}(a) = \exp \left( - \int_0^a \mu_W(u) + \lambda(u) \, du \right) \]

The probability of being alive with diabetes at age \( a \), is computed given that diabetes occurred at age \( s \) \( (s < a) \) and then integrated over \( s \) from 0 to \( a \):

\[ P\{\text{DM}\}(a) = \int_0^a \lambda(s) \exp \left( - \int_0^s \mu_W(u) + \lambda(u) \, du \right) \times \exp \left( - \int_s^a \mu_D(u) \, du \right) \, ds \]

Sometimes we will use a version where the mortality among diabetes patients depend both on age \( a \) and duration of diabetes, \( d \), \( \mu_D(a, d) \), in which case we get:

\[ P\{\text{DM}\}(a) = \int_0^a \lambda(s) \exp \left( - \int_0^s \mu_W(u) + \lambda(u) \, du \right) \times \exp \left( - \int_s^a \mu_D(u, u - s) \, du \right) \, ds \]

because the integration variable \( u \) is the age-scale and the second integral refers to mortality among persons diagnosed at age \( s \), that is, with duration \( u - s \) at age \( u \).

The option of using duration-dependent mortality rates among diseased individuals is not implemented yet.

Conditional survival functions

Unlike the immune approach, the conditional survival function in the more realistic case is not just a ratio of the unconditional to the value at the conditioning age, \( A \), say. This would amount to conditioning on being merely alive at age \( A \), but what we want is to condition on being in the “Well” state at age \( A \).

The formulae for the conditional probabilities of being either in “Well” or “DM”, given being in “Well” at age \( A \) are basically replicates of the unconditional, albeit with changes in
integration limits:

\[ P\{\text{Well}|\text{Well at } A\}(a) = \exp \left( - \int_a^A \mu_W(u) + \lambda(u) \right) \, du \]

\[ P\{\text{DM}|\text{Well at } A\}(a) = \int_a^A \lambda(s) \exp \left( - \int_s^s \mu_W(u) + \lambda(u) \, du \right) \]
\[ \times \exp \left( - \int_s^a \mu_D(u, u - s) \, du \right) \, ds \]

The calculation of these conditional survival functions is implemented but not allowing for duration-dependence. Thus it is only implemented assuming \( \mu_D(a, d) = \mu_D(a) \).
Chapter 2

Analyses for DM in Denmark

The rates we use as basis for the following calculations are derived from the NDR, where we have omitted the blood-glucose criteria, because there is compelling evidence that these have quite a low specificity (particularly in the younger ages among women), and do not substantially contribute to the sensitivity.

As noted above the calculations of YLL requires access to (age-specific) rates of incidence of DM and mortality for persons with and without DM.

2.1 Modeling mortality and incidence data

We read in the dataset of DM and population mortality and incidence, DMepi:

```r
> data(DMepi)
```

The dataset DMepi contains diabetes events, deaths and person-years for persons without diabetes and deaths and person-years for persons with diabetes:

```r
> str(DMepi)
'data.frame': 4200 obs. of 8 variables:
$ sex : Factor w/ 2 levels 'M','F': 1 1 1 1 1 1 1 1 1 1 ...
$ A : num 0 0 0 0 0 0 0 0 0 0 ...
$ P : num 1996 1997 1998 1999 2000 ...
$ D.DM: num 0 0 0 0 0 0 0 0 0 0 ...
$ Y.DM: num 0.484 0.64 1.641 0.552 2.507 ...
$ X : num 1 2 4 4 4 1 1 3 4 1 ...
$ D.nD: num 28 19 20 11 21 16 15 16 16 16 ...
$ Y.nD: num 35469 35085 34240 34056 34002 ...

> head(DMepi)

<table>
<thead>
<tr>
<th>sex</th>
<th>A</th>
<th>P</th>
<th>D.DM</th>
<th>Y.DM</th>
<th>X</th>
<th>D.nD</th>
<th>Y.nD</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.4839151</td>
<td>1</td>
<td>28</td>
<td>35468.92</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.6399726</td>
<td>2</td>
<td>19</td>
<td>35085.18</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.6406571</td>
<td>4</td>
<td>20</td>
<td>34240.14</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5523614</td>
<td>4</td>
<td>11</td>
<td>34055.52</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.5065024</td>
<td>4</td>
<td>21</td>
<td>34002.22</td>
</tr>
<tr>
<td>M</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1184120</td>
<td>1</td>
<td>16</td>
<td>34177.39</td>
</tr>
</tbody>
</table>
```

For each combination of sex, age, period and date of birth in 1 year age groups, we have the person-years in the “Well” (Y.nD) and the “DM” (Y.DM) states, as well as the number of
Analyses for DM in Denmark

2.1 Modeling mortality and incidence data

In order to compute the years of life lost to diabetes and how this has changed over time, we fit models for the mortality and incidence of both groups (and of course, separately for men and women). The models we use will be age-period-cohort models [1] providing estimated mortality rates for ages 0–99 and dates 1.1.1996–1.1.2016.

First we transform the age and period variables to reflect the mean age and period in each of the Lexis triangles. We also compute the total number of deaths and amount of risk time, as we are going to model the total mortality as well. Finally we restrict the dataset to ages over 30 only:

```r
> DMepi <- transform(subset(DMepi, A>30),
+    D.T = D.nD + D.DM,
+    Y.T = Y.nD + Y.DM)
> head(DMepi)
         sex A   P  D.DM Y.DM  X  D.nD Y.nD D.T  Y.T
684      M 31 1996  0305.9671  29  51 44161.83  51  44467.80
685      M 31 1997  2326.2074  31  54 45508.64  56  45834.85
686      M 31 1998  2340.1759  34  46 44954.45  48  45294.63
687      M 31 1999 5330.9918  24  39 41148.97  44  41479.96
688      M 31 2000  1332.0876  41  26 39027.30  27  39359.39
689      M 31 2001  2310.8467  35  33 37975.78  35  38286.63
```

With the correct age and period coding in the Lexis triangles, we fit models for the mortalities and incidences. Note that we for comparative purposes also fit a model for the total mortality, ignoring the

```r
> # Knots used in all models
> (a.kn <- seq(40, 95, , 6))
[1] 40 51 62 73 84 95
> (p.kn <- seq(1997, 2015, , 4))
> (c.kn <- seq(1910, 1976, , 6))
[1] 1910.0 1923.2 1936.4 1949.6 1962.8 1976.0
> # Check the number of events between knots
> ae <- xtabs(cbind(D.nD, D.DM, X) ~ cut(A, c(30, a.kn, Inf)) + sex, data=DMepi)
> ftable(addmargins(ae, 1), col.vars=3:2)

<table>
<thead>
<tr>
<th></th>
<th>D.nD</th>
<th>D.DM</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex</td>
<td>M</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>M</td>
<td>9135</td>
<td>4650</td>
<td>597</td>
</tr>
<tr>
<td>F</td>
<td>25535</td>
<td>15784</td>
<td>2947</td>
</tr>
<tr>
<td>M</td>
<td>59698</td>
<td>40171</td>
<td>10838</td>
</tr>
<tr>
<td>F</td>
<td>106519</td>
<td>81253</td>
<td>26197</td>
</tr>
<tr>
<td>M</td>
<td>158365</td>
<td>156678</td>
<td>34510</td>
</tr>
<tr>
<td>F</td>
<td>100880</td>
<td>179466</td>
<td>16194</td>
</tr>
<tr>
<td>M</td>
<td>6095</td>
<td>21414</td>
<td>640</td>
</tr>
<tr>
<td>F</td>
<td>466227</td>
<td>499416</td>
<td>91923</td>
</tr>
</tbody>
</table>
```

```r
> pe <- xtabs(cbind(D.nD, D.DM, X) ~ cut(P, c(1990, p.kn, Inf)) + sex, data=DMepi)
> ftable(addmargins(pe, 1), col.vars=3:2)
```

2.2 Residual life time and years lost to DM

We now collect the estimated years of life lost classified by method (immune assumption or not), sex, age and calendar time:

```r
> a.ref <- 30:90
> aYLL <- NArray(list(type = c("Imm", "Tot", "Sus"),
                  sex = levels(DMepi$sex),
                  age = a.ref,
                  date = p.ref))
> str(aYLL)
'data.frame': 12 obs. of 4 variables:
  ..$ type: chr [1:3] "Imm" "Tot" "Sus"
  ..$ sex : chr [1:2] "M" "F"
```

```r
2.2 Residual life time and years lost to DM

We now collect the estimated years of life lost classified by method (immune assumption or not), sex, age and calendar time:

```r
> a.ref <- 30:90
> aYLL <- NArray(list(type = c("Imm", "Tot", "Sus"),
                  sex = levels(DMepi$sex),
                  age = a.ref,
                  date = p.ref))
> str(aYLL)
'data.frame': 12 obs. of 4 variables:
  ..$ type: chr [1:3] "Imm" "Tot" "Sus"
  ..$ sex : chr [1:2] "M" "F"
```
Analyses for DM in Denmark

2.2 Residual life time and years lost to DM

> system.time(
+ for(ip in p.ref)
+ {
+ nd <- data.frame(A = seq(30, 90, 0.2)+0.1,
+ P = ip,
+ Y.nD = 1,
+ Y.DM = 1,
+ Y.T = 1)
+ muW.m <- ci.pred(mW.m, nd)[, 1]
+ muD.m <- ci.pred(mD.m, nd)[, 1]
+ muT.m <- ci.pred(mT.m, nd)[, 1]
+ lam.m <- ci.pred(lW.m, nd)[, 1]
+ muW.f <- ci.pred(mW.f, nd)[, 1]
+ muD.f <- ci.pred(mD.f, nd)[, 1]
+ muT.f <- ci.pred(mT.f, nd)[, 1]
+ lam.f <- ci.pred(lW.f, nd)[, 1]
+ aYLL[c("Imm", "M", , paste(ip)) <- yll(int=0.2, muW.m, muD.m, lam=NULL,
+ A=a.ref, age.in=30, note=FALSE)[-1]
+ aYLL[c("Imm", "F", , paste(ip)) <- yll(int=0.2, muW.f, muD.f, lam=NULL,
+ A=a.ref, age.in=30, note=FALSE)[-1]
+ aYLL[c("Tot", "M", , paste(ip)) <- yll(int=0.2, muT.m, muD.m, lam=NULL,
+ A=a.ref, age.in=30, note=FALSE)[-1]
+ aYLL[c("Tot", "F", , paste(ip)) <- yll(int=0.2, muT.f, muD.f, lam=NULL,
+ A=a.ref, age.in=30, note=FALSE)[-1]
+ aYLL[c("Sus", "M", , paste(ip)) <- yll(int=0.2, muW.m, muD.m, lam=lam.m,
+ A=a.ref, age.in=30, note=FALSE)[-1]
+ aYLL[c("Sus", "F", , paste(ip)) <- yll(int=0.2, muW.f, muD.f, lam=lam.f,
+ A=a.ref, age.in=30, note=FALSE)[-1]
+ })
+ user system elapsed
10.390 0.000 10.393

> round(ftable(aYLL[, , seq(1, 61, 10), ], col.vars=c(3, 2)), 1)

<table>
<thead>
<tr>
<th></th>
<th>age 30</th>
<th>age 40</th>
<th>age 50</th>
<th>age 60</th>
<th>age 70</th>
<th>age 80</th>
<th>age 90</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>date</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1996</td>
<td>11.3</td>
<td>10.2</td>
<td>9.4</td>
<td>9.1</td>
<td>7.4</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td>11.1</td>
<td>9.9</td>
<td>9.2</td>
<td>8.9</td>
<td>7.3</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>1998</td>
<td>10.9</td>
<td>9.7</td>
<td>9.1</td>
<td>8.6</td>
<td>7.2</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>1999</td>
<td>10.7</td>
<td>9.4</td>
<td>9.0</td>
<td>8.4</td>
<td>7.1</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>10.5</td>
<td>9.1</td>
<td>8.8</td>
<td>8.2</td>
<td>6.9</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td>2001</td>
<td>10.3</td>
<td>8.9</td>
<td>8.6</td>
<td>7.9</td>
<td>6.8</td>
<td>6.6</td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>10.0</td>
<td>8.6</td>
<td>8.4</td>
<td>7.7</td>
<td>6.6</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>9.7</td>
<td>8.3</td>
<td>8.2</td>
<td>7.4</td>
<td>6.4</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>9.4</td>
<td>8.0</td>
<td>7.9</td>
<td>7.1</td>
<td>6.2</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>9.0</td>
<td>7.6</td>
<td>7.6</td>
<td>6.9</td>
<td>6.0</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td>8.6</td>
<td>7.3</td>
<td>7.4</td>
<td>6.6</td>
<td>5.8</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>2007</td>
<td>8.3</td>
<td>7.0</td>
<td>7.1</td>
<td>6.3</td>
<td>5.5</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>2008</td>
<td>8.0</td>
<td>6.8</td>
<td>6.8</td>
<td>6.1</td>
<td>5.4</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>2009</td>
<td>7.7</td>
<td>6.6</td>
<td>6.6</td>
<td>6.0</td>
<td>5.2</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>7.5</td>
<td>6.4</td>
<td>6.5</td>
<td>5.8</td>
<td>5.1</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>7.4</td>
<td>6.3</td>
<td>6.4</td>
<td>5.8</td>
<td>5.1</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>7.3</td>
<td>6.3</td>
<td>6.4</td>
<td>5.7</td>
<td>5.1</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>7.3</td>
<td>6.2</td>
<td>6.4</td>
<td>5.7</td>
<td>5.1</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>7.3</td>
<td>6.2</td>
<td>6.4</td>
<td>5.7</td>
<td>5.1</td>
<td>4.7</td>
</tr>
</tbody>
</table>
We now have the relevant points for the graph showing YLL to diabetes for men and women by age, and calendar year, both under the immunity and susceptibility models for the calculation of YLL.

```r
> plyll <- function(wh){
+ par(mfrow=c(1, 2), mar=c(3, 3, 1, 1), mgp=c(3, 1, 0)/1.6, bty="n", las=1)
+ matplot(a.ref, aYLL[wh, "M", , ],
+ type="l", lty=1, col="blue", lwd=1:2,
+ ylim=c(0, 12), xlab="Age",
+ ylab="Years lost to DM", yaxs="i")
+ abline(v=50, h=1:10, col=gray(0.7))
```
Figure 2.1: Years of life lost to DM: the difference in expected residual life time at different ages between persons with and without diabetes, assuming the persons without diabetes at a given age remain free from diabetes (immunity assumption — not reasonable). The lines refer to date of evaluation; the top lines refer to 1.1.1996 the bottom ones to 1.1.2016. Blue curves are men, red women.

From figure 2.2 we see that for men aged 50 the years lost to diabetes has decreased from a bit over 8 to a bit less than 6 years, and for women from 8.5 to 5 years; so a greater improvement for women.
Figure 2.2: Years of life lost to DM: the difference in expected residual life time at different ages between persons with and without diabetes, allowing the persons without diabetes at a given to contract diabetes and thus be subject to higher mortality. The lines refer to date of evaluation; the top lines refer to 1.1.1996 the bottom ones to 1.1.2016. Blue curves are men, red women.
Figure 2.3: Years of life lost to DM: the difference in expected residual life time at different ages between persons with and without diabetes. Allowance for susceptibility is approximated by using the total population mortality instead of non-DM mortality. The lines refer to date of evaluation; the top lines refer to 1.1.1996 the bottom ones to 1.1.2016. Blue curves are men, red women.