Expected Residual Life time and Years of Life Lost

Bendix Carstensen Steno Diabetes Center Copenhagen, Denmark
& Dept. of Biostatistics, University of Copenhagen

VicBiostats, Melbourne,

23 February 2017

http://BendixCarstensen.com

Life lost to disease

- Persons with disease live shorter than persons without
- The difference is the life lost to disease — years of life lost
- Possibly depends on:
  - sex
  - age
  - duration of disease
  - definition of persons with/out disease
- Conditional or population averaged?
- . . . the latter gives a seductively comfortable single number
- . . . the former confusingly relevant insights

Expected life time — the formals:

. . . the age at death integrated w.r.t. the distribution of age at death:

$$\mathrm{EL} = \int_0^\infty a\, f(a)\, \mathrm{d}a$$

The relation between the density f and the survival function S is $f(a) = -S'(a)$, so integration by parts gives:

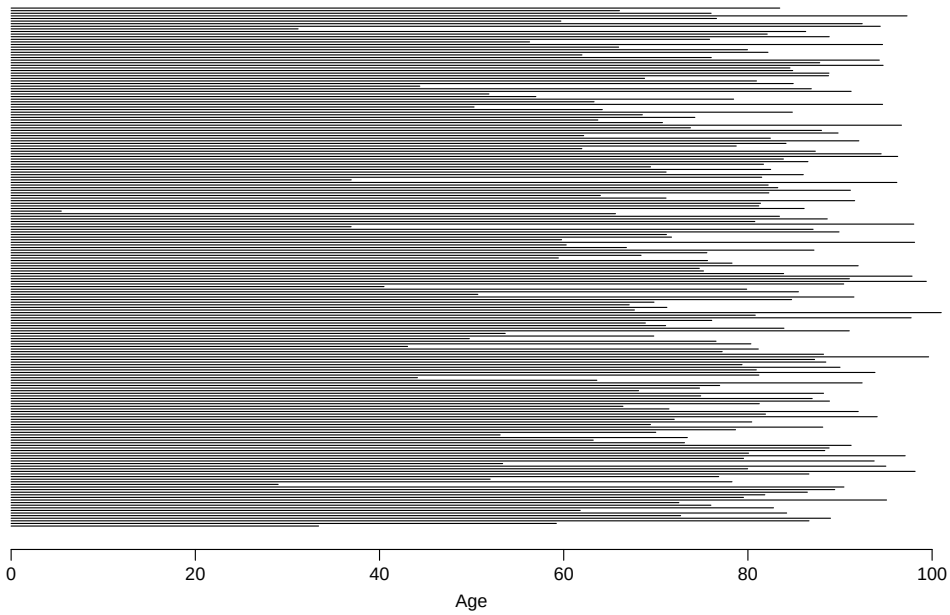
$$\mathrm{EL} = \int_0^\infty a\big(-S'(a)\big)\, \mathrm{d}a = -\Big[aS(a)\Big]_0^\infty + \int_0^\infty S(a)\, \mathrm{d}a$$

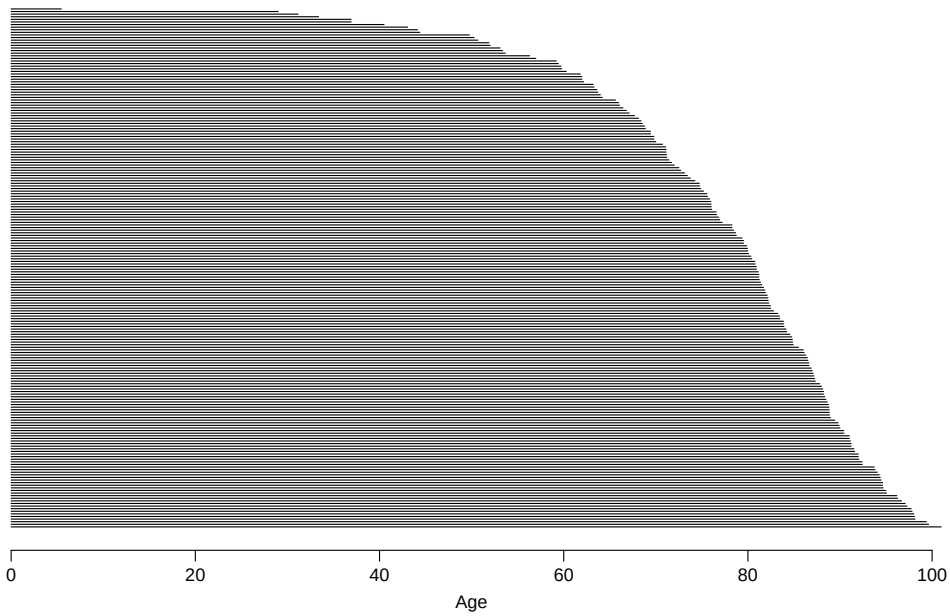
The first term is 0 so:

$$\mathrm{EL} = \int_0^\infty S(a)\, \mathrm{d}a$$

Expected life time illustrated

- Take, say 200, persons
- follow till all are dead
- compute the mean age at death (life time)
- — that is the **life expectancy** (at birth)





Expected life time and years lost

- ERL (**E**xpected **R**esidual **L**ifetime):
  Area under the survival curve
- YLL (**Y**ears of **L**ife **L**ost) (to diabetes, say):
  $\mathrm{ERL}_{\mathrm{pop}} - \mathrm{ERL}_{\mathrm{DM}}$
- **difference** between areas under the survival curves
- $\Rightarrow$ area **between** the curves
- . . . all the way till all are dead

Wikipedia: PYLL

Potential **Y**ears of **L**ife **L**ost

- Fix a threshold, T, (the population EL, or say 75)
- A person dead in age $a < T$ contributes $T - a$
- A person dead in age $a > T$ contributes 0

. . . seems to assume that the expected age at death is T regardless of attained age ?

WHO — Years of Life Lost

Rationale for use
Years of life are lost (YLL) take into account the age at which deaths occur by giving greater weight to deaths at younger age and lower weight to deaths at older age. The years of life lost (percentage of total) indicator measures the YLL due to a cause as a proportion of the total YLL lost in the population due to premature mortality.

Definition
YLL are calculated from the number of deaths multiplied by a standard life expectancy at the age at which death occurs. The standard life expectancy used for YLL at each age is the same for deaths in all regions of the world (. . .)

`www.who.int/whosis/whostat2006YearsOfLifeLost.pdf`

$\Rightarrow$ a person dying in age a contributes $\mathrm{ERL}(a)$. . .

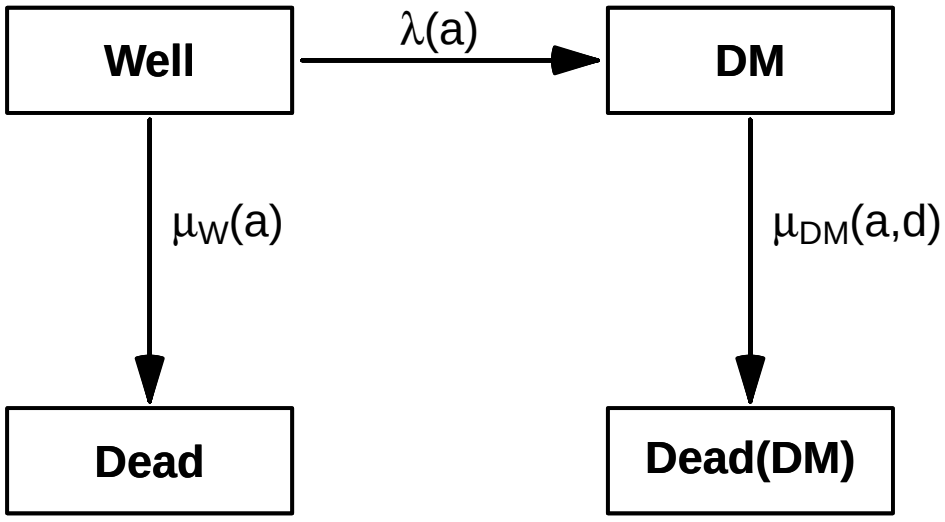
Comparing men and women

- When a man dies age a, say,
  - YLL is $\mathrm{ERL}_w(a) > 0$
  - — the expected residual life time of a woman aged a.
- When a woman dies age a, say,
  - YLL is $\mathrm{ERL}_m(a) > 0$
  - — the expected residual life time of a man aged a.
- . . . so both sexes lose years relative to the other !

Healthy lifestyle and excercise. . .

- Any one who dies before age 75 (PYLL)
- Any one who dies (WHO YLL)
- . . . contribute a **positive** number to YLL
- $\Rightarrow$ **any** subgroup of the population have positive years of life lost when compared to the general population!
- . . . well, indeed compared to **any** population (men vs. women)
- No shortcuts:
  - no unfounded algorithms
  - the YLL is a difference of expectations
  - use a statistical model (specify $f(a)$, that is)
  - diabetes in Denmark as an example

How the world looks



Comparing DM and well

$$\mathrm{YLL} = \int_0^\infty S_W(a) - S_D(a)\,\mathrm{d}a$$

the **conditional** YLL given attained age A, just use:

$$S_W(a|A) = S_W(a)/S_W(A), \qquad S_D(a|A) = S_D(a)/S_D(A)$$

The survival functions we need are:

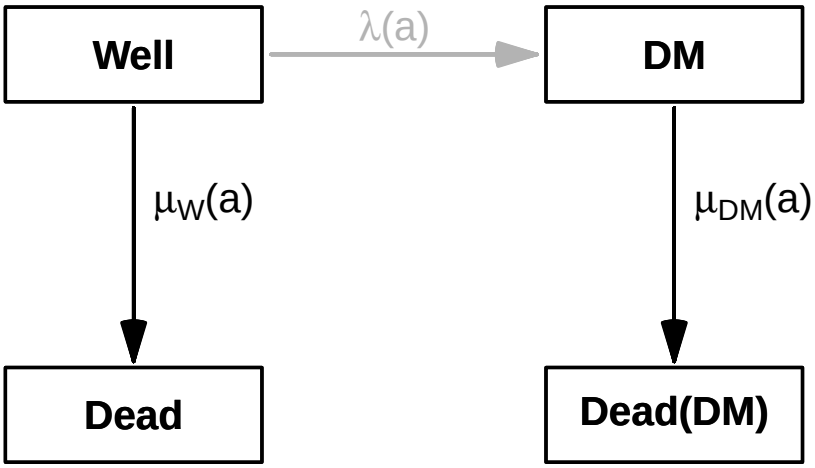
$$S_W(a) = \exp\left(-\int_0^a \mu_W(u)\,\mathrm{d}u\right), \qquad S_D(a) = \exp\left(-\int_0^a \mu_D(u)\,\mathrm{d}u\right)$$

. . . or is it?

Assumes that persons in "Well" cannot contract "DM"

The immunity assumption — which is widely used in the literature

How the world looks



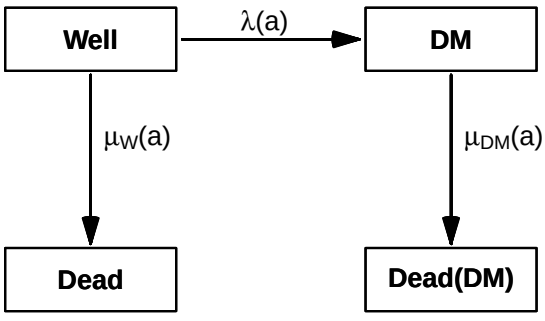
. . . with immunity to diabetes

Comparing DM and Well in the real world

$$\mathrm{YLL} = \int_0^\infty S_W(t) - S_D(t)\,\mathrm{d}t$$

still the same, but $S_W(t)$ should be:

$$S_W(a) = \mathrm{P}\{\mathsf{Well}\}(a) + \mathrm{P}\{\mathsf{DM}\}(a)$$



Comparing DM and well in the real world

The survival function $S_W(a)$ is the sum of:

$$\mathrm{P}\left\{\text{Well}\right\}(a) = \exp\left(-\int_0^a \mu_W(u) + \lambda(u)\right) \mathrm{d}u$$

and

$$\begin{aligned}\mathrm{P}\left\{\text{DM}\right\}(a) = \int_0^a & \mathrm{P}\left\{\text{survive to } s, \text{ DM diagnosed at } s\right\} \\ & \times \mathrm{P}\left\{\text{survive with DM from } s \text{ to } a\right\} \mathrm{d}s \\ = \int_0^a & \lambda(s) \exp\left(-\int_0^s \mu_W(u) + \lambda(u)\,\mathrm{d}u\right) \\ & \times \exp\left(-\int_s^a \mu_D(u)\,\mathrm{d}u\right) \mathrm{d}s\end{aligned}$$

Comparing DM and well in the real world

The **conditional** survival function given **Well at** A is the sum of

$$\mathrm{P}\left\{\text{Well}|\text{Well at } A\right\}(a) = \exp\left(-\int_A^a \mu_W(u) + \lambda(u)\right) \mathrm{d}u$$

$$\begin{aligned}\mathrm{P}\left\{\text{DM}|\text{Well at } A\right\}(a) = \int_A^a & \lambda(s) \exp\left(-\int_A^s \mu_W(u) + \lambda(u)\,\mathrm{d}u\right) \\ & \times \exp\left(-\int_s^a \mu_D(u)\,\mathrm{d}u\right) \mathrm{d}s\end{aligned}$$

Note: This is **not** $S_W(a)/S_W(A)$ because we are not conditioning on being alive, but conditioning on being alive and well at A

A brutal shortcut

. . . sooo hairy, so why don't we not just use the **total** population mortality, μ_T, and instead compare:

$$S_T(a) = \exp\left(-\int_0^a \mu_T(u)\,\mathrm{d}u\right), \qquad S_D(a) = \exp\left(-\int_0^a \mu_D(u)\,\mathrm{d}u\right)$$

There is no simple inequality between S_T and the correctly computed S_W so there is no guarantee that it will be useful, nor the direction of bias

The comparison will be between a random person with diabetes and a random person (with or without diabetes)

Empirical question whether this is a reasonable approximation

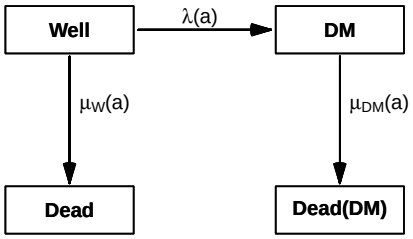
From probability theory to statistics:

- get data on diabetes and death events by diabetes status
- get data on risk time by diabetes status
- fit models for the rates
- get expressions for $\mu_W(a)$, $\lambda(a)$ and $\mu_D(a)$
- compute the integrals for say $A = 50, 60, \ldots$

From probability theory to statistics: data

```
> library( Epi )
> data( DMepi )
> head( DMepi )
```

```
  sex A    P X D.nD     Y.nD D.DM       Y.DM
1   M 0 1996 1   28 35453.65    0  0.4757016
2   F 0 1996 9   19 33094.86    0  3.8767967
3   M 1 1996 4   23 36450.73    0  4.9199179
4   F 1 1996 7   19 34789.99    0  7.2484600
5   M 2 1996 7    7 35328.92    0 12.4743326
6   F 2 1996 2    8 33673.43    0  8.0951403
```



From probability theory to statistics: models

```
> # knots used for splines in all models
> a.kn <- seq(40,95,,6)
> p.kn <- seq(1996,2011,,4)
> c.kn <- seq(1910,1970,,6)
> #
> #
> # APC-model for death for non-DM men
> mW.m <- glm( D.nD ~ Ns(  A,knots=a.kn) +
+                     Ns(P  ,knots=p.kn) +
+                     Ns(P-A,knots=c.kn),
+           offset = log(Y.nD),
+           family = poisson,
+             data = subset( DMepi, sex=="M" & A>29 ) )
```

. . . estimates mortality (and incidence) rates over the grid:

- age: $30 - 99$
- calendar time: $1996 - 2015$

From probability theory to statistics: predictions

Mortality rates for men in ages $30 - 100$ using rates from 2012:

```
> nd <- data.frame( A = seq(30,99.8,0.2)+0.1,
+                   P = 2012,
+                 Y.nD = 1,
+                 Y.DM = 1,
+                 Y.T  = 1 )
> muW.m <- ci.pred( mW.m, nd )[,1]
> cbind( nd$A, muW.m )[200+0:5,]

              muW.m
200 69.9 0.02017309
201 70.1 0.02056253
202 70.3 0.02096210
203 70.5 0.02137211
204 70.7 0.02179289
205 70.9 0.02222479
```

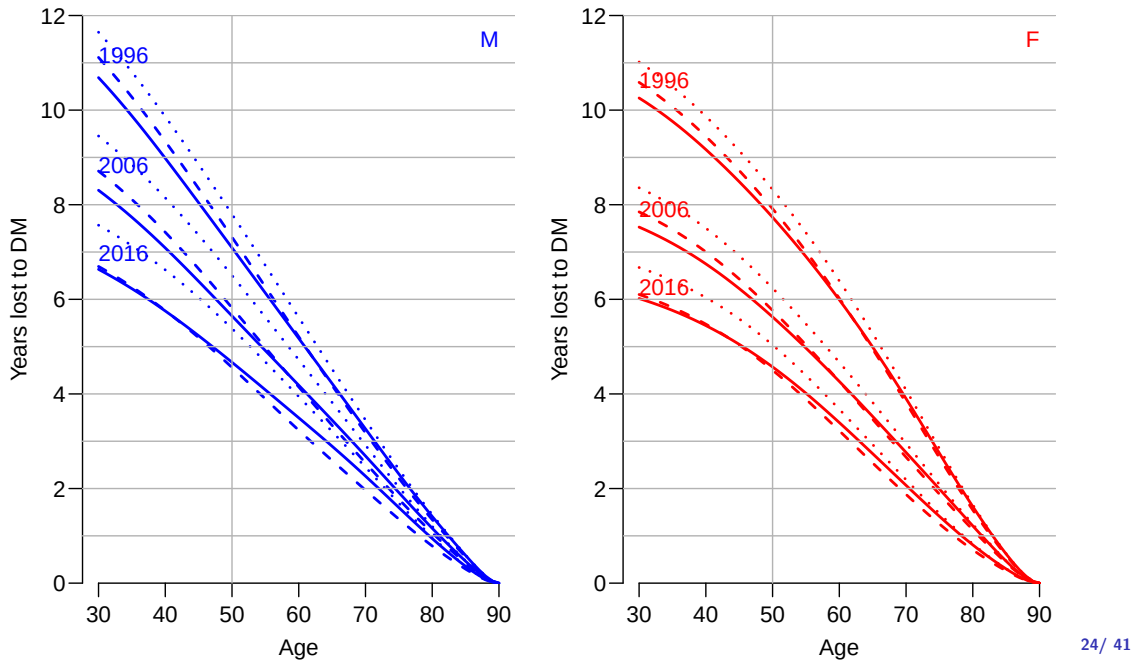
Rates representation when used in computing integrals:
Compute the function value in small equidistant intervals

From probability theory to statistics: YLL calculation

Epi package for **R** contains the dataset `DMepi` as well as the functions `erl` and `yll` that implements the formulae:

```
> YLL.m.60 <- yll( int=0.2,
+                  muW=muW.m, muD=muD.m, lam=lam.m,
+                  A=60, age.in=30 )
```

This is then done for different conditioning ages (A), men/women and based on predicted rates from 1996, 2006 and 2016.



Years of life lost to disease: Conclusion

- Use a model
- for **all** your rates
- use your probability theory
- credible models for rates requires:
  smooth parametric function of age and calendar time
- continuous time formulation simplifies concepts and computing
- using non-DM mortality overestimates YLL
- If you cannot do it correctly for want of data:
  compare with the **total** population mortality
- Note: **Conditional** YLLs — given date, age and sex.

Years of life lost to disease: Generalization

- YLL is really a generalization
- from a multistate model
- of the expected sojourn time in a given state
- . . . well, differences of these
- here is an example

Diabetologia
DOI 10.1007/s00125-016-4065-6

CrossMark

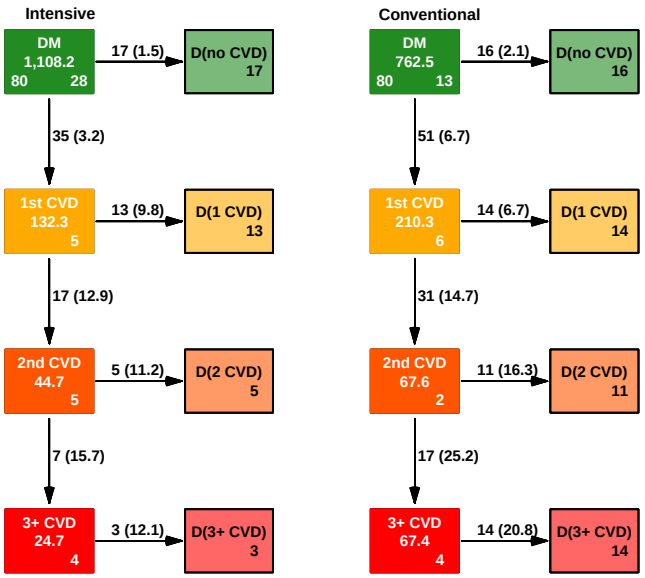
ARTICLE

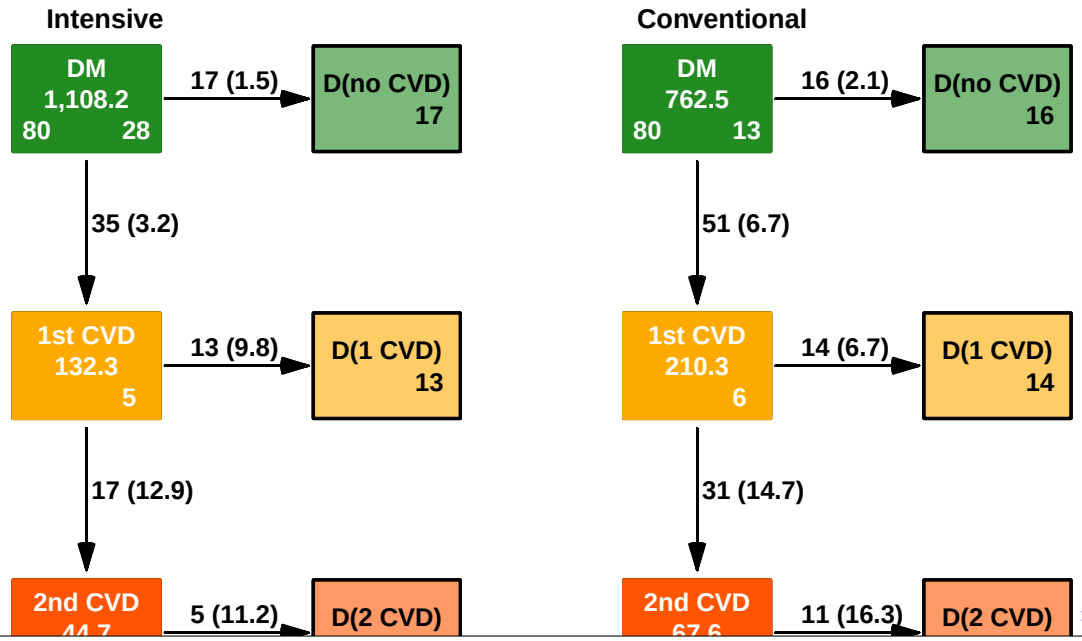
Years of life gained by multifactorial intervention in patients with type 2 diabetes mellitus and microalbuminuria: 21 years follow-up on the Steno-2 randomised trial

Peter Gæde[1,2] · Jens Oellgaard[1,2,3] · Bendix Carstensen[3] · Peter Rossing[3,4,5] · Henrik Lund-Andersen[3,5,6] · Hans-Henrik Parving[5,7] · Oluf Pedersen[8]

Received: 7 April 2016 / Accepted: 1 July 2016

Abstract
Aims/hypothesis The aim of this work was to study the potential long-term impact of a 7.8 years intensified multifactorial pharmacological approaches. After 7.8 years the study continued as an observational follow-up with all patients receiving treatment as for the original intensive-therapy group. The pri-





Hazard ratios

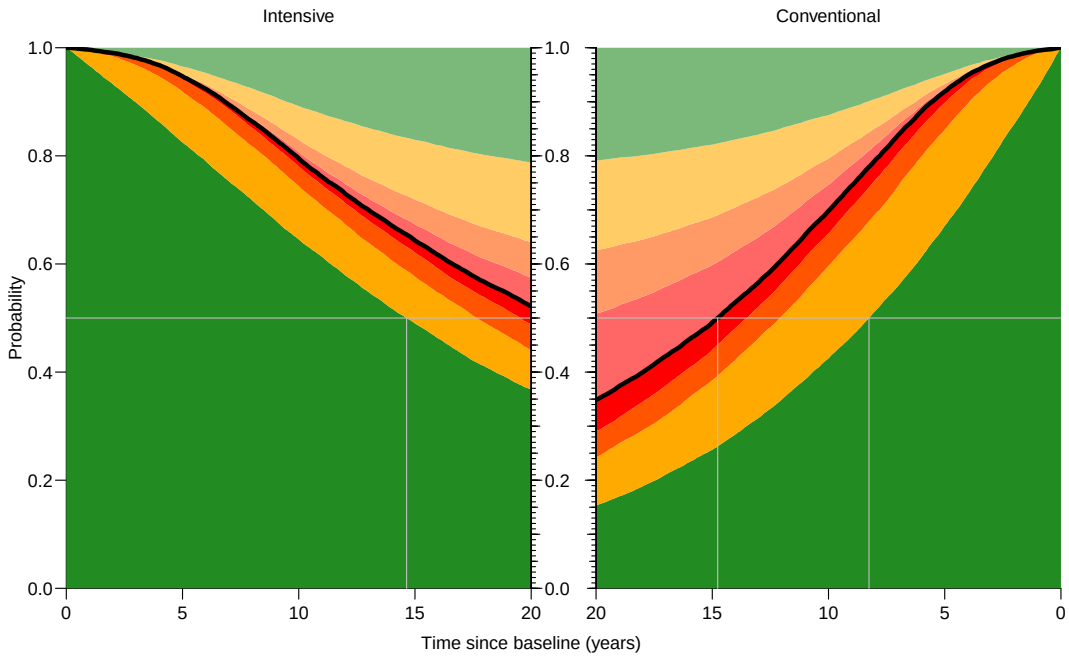
	CVD event	Mortality
HR, Int. vs. Conv.	0.55 (0.39;0.77)	0.83 (0.54; 1.30)
H_0: PH btw. CVD groups	p=0.261	p=0.438
H_0: HR = 1	p=0.001	p=0.425
HR vs. 0 CVD events:		
0 (ref.)	1.00	1.00
1	2.43 (1.67;3.52)	3.08 (1.82; 5.19)
2	3.48 (2.15;5.64)	4.42 (2.36; 8.29)
3+		7.76 (4.11;14.65)

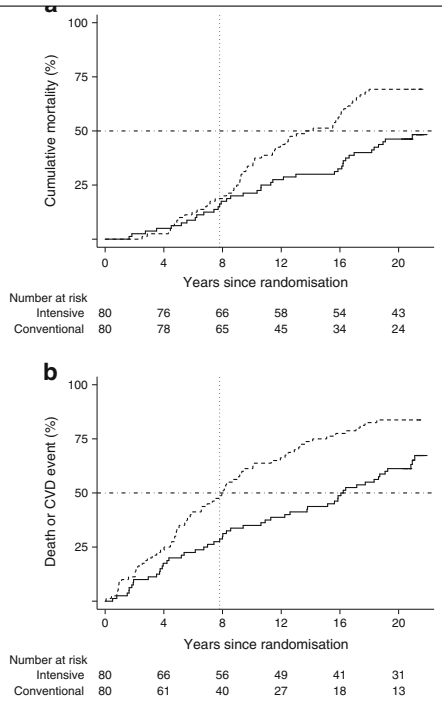
Modeling

- Cut the follow-up time for each person by state
- Split the follow-up time in 1-month intervals
- Poisson model with smooth effect of time since randomization, sex and age:
  - HR estimates
  - Estimates of baseline hazard
  - Hazard for any set of covariates
- Allows calcualtion of expected sojourn time in any state
- — analytically this is totally intractable. . .

Estimating sojourn times

- Use simulation of the state occupancy probabilities:
- `Lexis` machinery in the `Epi` package for multistate representation
- `splitLexis` to subdivide follow-up for analysis
- `simLexis` for simulation to derive probabilities and sojourn times
- — simulates a cohort through the model, so probabilities are just empirical fractions





between groups (HR 0.83 [95% CI 0.54, 1.30], $p=0.43$). Thus, the reduced mortality was primarily due to reduced risk of CVD.

The patients in the intensive group experienced a total of 90 cardiovascular events vs 195 events in the conventional group. Nineteen intensive-group patients (24%) vs 34 conventional-group patients (43%) experienced more than one cardiovascular event. No significant between-group difference in the distribution of specific cardiovascular first-event types was observed (Table 2 and Fig. 4).

Microvascular complications Hazard rates of progression rates in microvascular complications compared with baseline status are shown Fig. 3. Sensitivity analyses showed a negligible effect of the random dates imputation.

Progression of retinopathy was decreased by 33% in the intensive-therapy group (Fig. 5). Blindness in at least one eye was reduced in the intensive-therapy group with an HR of 0.47 (95% CI 0.23, 0.98, $p=0.044$). Autonomic neuropathy was decreased by 41% in the intensive-therapy group (Fig. 5). We observed no difference between groups in the progression of peripheral neuropathy (Fig. 5). Progression to diabetic nephropathy (macroalbuminuria) was reduced by 48% in the intensive-therapy group (Fig. 5). Ten patients in the conventional-therapy groups vs five patients in the intensive-therapy group progressed to end-stage renal disease ($p=0.061$).

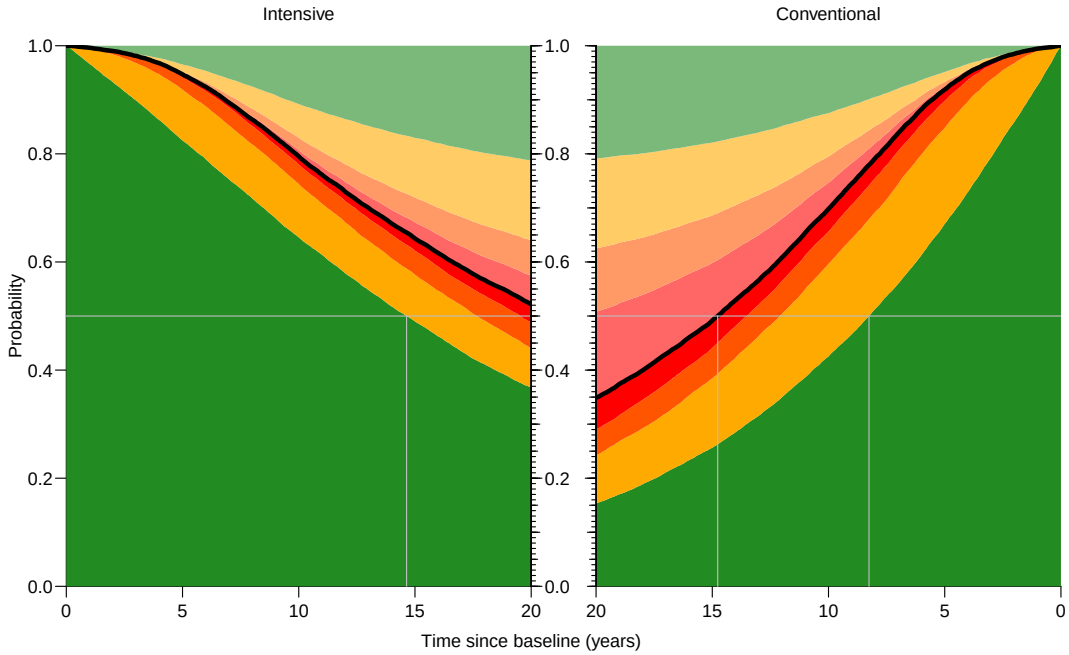
Discussion

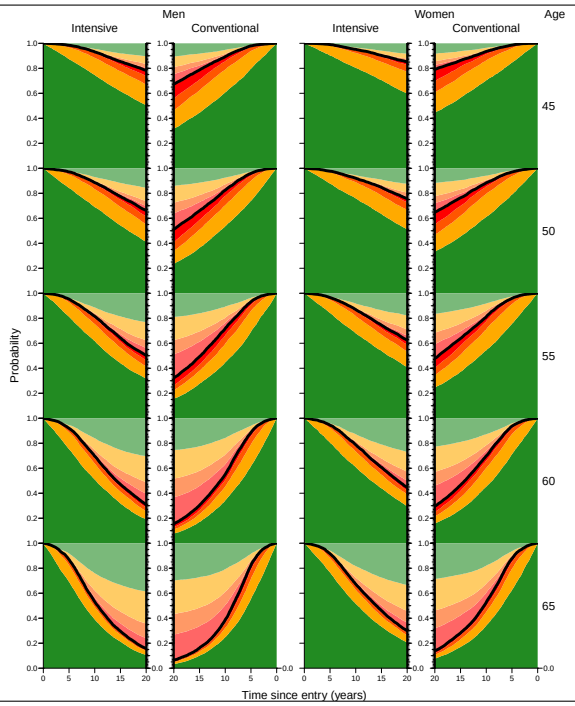
Expected lifetime and YLL (well, gained)

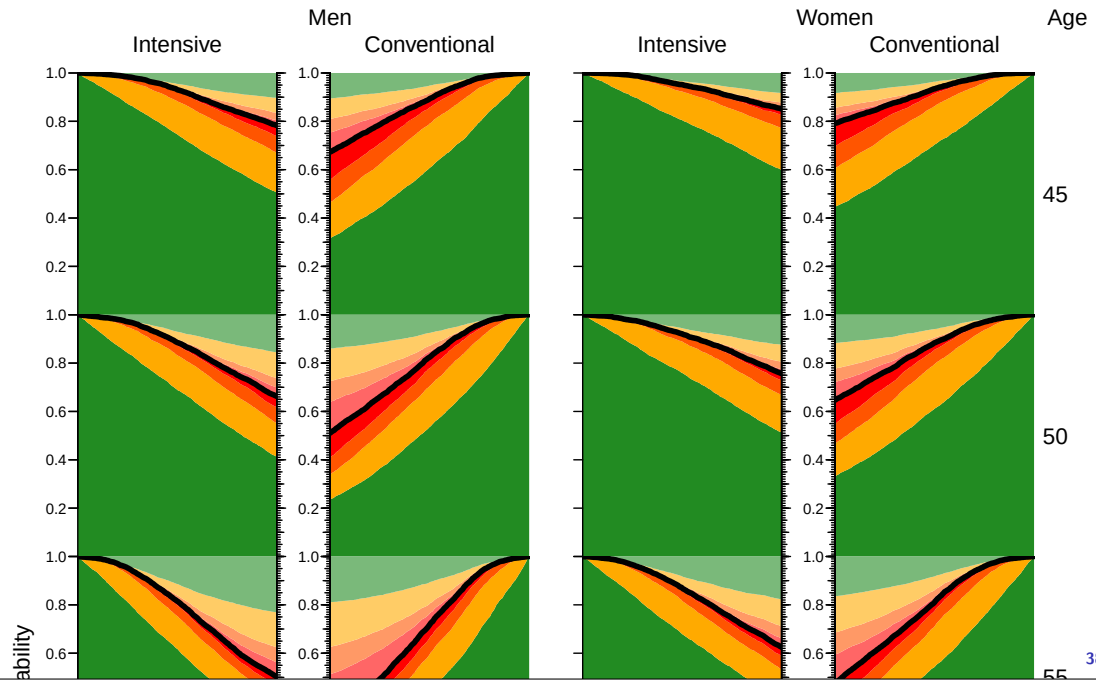
Expected lifetime (years) in the Steno 2 cohort during the first 20 years after baseline by treatment group and CVD status.

State	Intensive	Conventional	Int.−Conv.
Alive	15.6	14.1	1.5
No CVD	12.7	10.0	2.6
Any CVD	3.0	4.1	−1.1

- Simulate a cohort with same covariate dist' as the study
- **Population averaged** years gained alive / CVD-free
- Refer **only** to the Steno 2 trial population
- **Not** generalizable
- . . . but we have a **model**





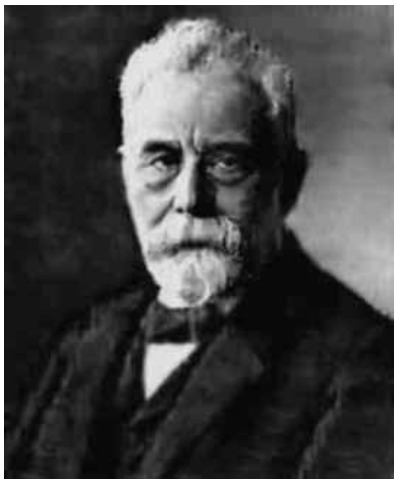


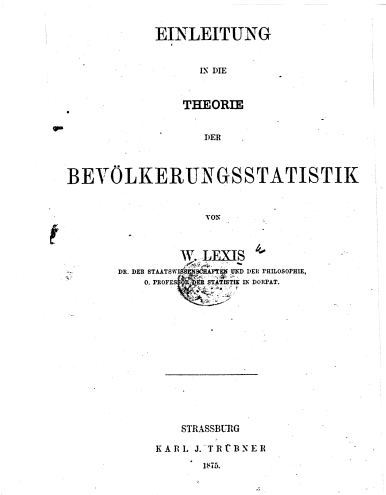
Expected lifetime (years) and $-\mathrm{YLL}$ (YLG) during the first 20 years after baseline by sex, age, treatment group and CVD status.

sex		Men			Women		
state	age	Int.	Conv.	YLG	Int.	Conv.	YLG
Alive	45	18.5	17.5	1.0	19.1	18.4	0.7
	50	17.2	16.1	1.1	18.0	17.2	0.8
	55	15.6	13.8	1.8	17.4	15.9	1.6
	60	13.9	11.6	2.2	15.5	13.7	1.8
	65	11.2	9.5	1.8	13.3	11.4	2.0
No CVD	45	14.9	12.5	2.4	15.8	14.3	1.5
	50	14.0	11.1	2.9	15.1	12.9	2.2
	55	12.2	9.7	2.5	14.3	11.6	2.7
	60	10.9	8.2	2.7	12.4	9.9	2.6
	65	9.0	6.7	2.2	10.7	8.3	2.4

History

- `Epi` package grew out of
  "Statistical Practice in Epidemiology with R"
  annually since 2002 in Tartu Estonia
  `http://BendixCarstensen.com/SPE`
- `Lexis` machinery conceived by Martyn Plummer, IARC
- Naming originally by David Clayton & Michael Hills, `stlexis` in Stata, later renamed `stsplit`
- David Clayton wrote a `lexis` function for the `Epi` package. Obsolete now.




EINLEITUNG

IN DIE

THEORIE

DER

BEVÖLKERUNGSSTATISTIK

VON

W. LEXIS

STRASSBURG

KARL J. TRÜBNER

1875.

Thanks for your attention