Demographic register research and multistate models:
rates, probabilities, sojourn time

Bendix Carstensen Steno Diabetes Center Copenhagen
Herlev, Denmark
`http://BendixCarstensen.com`

MEB, KI, Stockholm, Monday 12th May 2025

`http://BendixCarstensen.com/PMM` — Practical Multistate Modeling

From `C:\Bendix\sdc\conf\MEB/DemoReg.tex` Friday 9th May, 2025, 10:40

Topics

- Registers
- Demography
- Scales
- Follow-up representation
- Multistate data
- Multistate likelihood
- Multistate modeling

What's in a register

One record per event (diagnosis):

- person-id
- time of event (a date, usually)
- type of event (T1 / T2 / other)

Some events can occur at most once (diabetes, cancer), other any number of times (CVD, hypoglycemia)

Some registers contain multiple events of a type (NPR, e.g.)

It is **you** who define what an event is

Disease register use: Look-up

- Persons from some study cohort, such as a population survey or a clinical study—what is their:
  - disease **status** (noDM/T1/T2) at a given date
  - disease **date** (T1 / T2)
- by exclusion we also know if a person does **not** have disease (completeness assumption)
- ⇒ data input to existing (cohort) studies where follow-up is already known
  - explanatory variable for known outcome (status)
  - outcome event in an existing cohort (date)

Disease register use: Demography

Demographic **analysis** in the **population**

- incidence and
- mortality rates,
- prevalence
- —and derivatives of basic demographic measures:
  - state probabilities
  - lifetime risk
  - expected lifetime in noDM / T1 / T1
  - lifetime lost

  …these measures need further assumptions
- register events are outcome **events** (d),
  FU-time in population is outcome **risk time** (y)

Disease demography: Scales of inference

-1. Occurrence rates
—the scale of **observed** register data, (d, y) (empirical rate),
measured in time^{-1} (events per person-time)

0. State probabilities (survival function)
—the **integral** of rates w.r.t. time
—requires an origin (such as date of diagnosis)
measured in time^{0} (dimensionless)

1. Sojourn times (time spent in a state)
—the **integral** of state probabilities w.r.t. time
—requires an origin and endpoint
measured in time^{1}

Demographic quantities—functions of time

- occurrence rate:

$$\lambda(t) = \lim_{h \to 0} \mathrm{P}\{\text{event in } (t, t+h) \,|\, \text{alive at } t\}/h$$

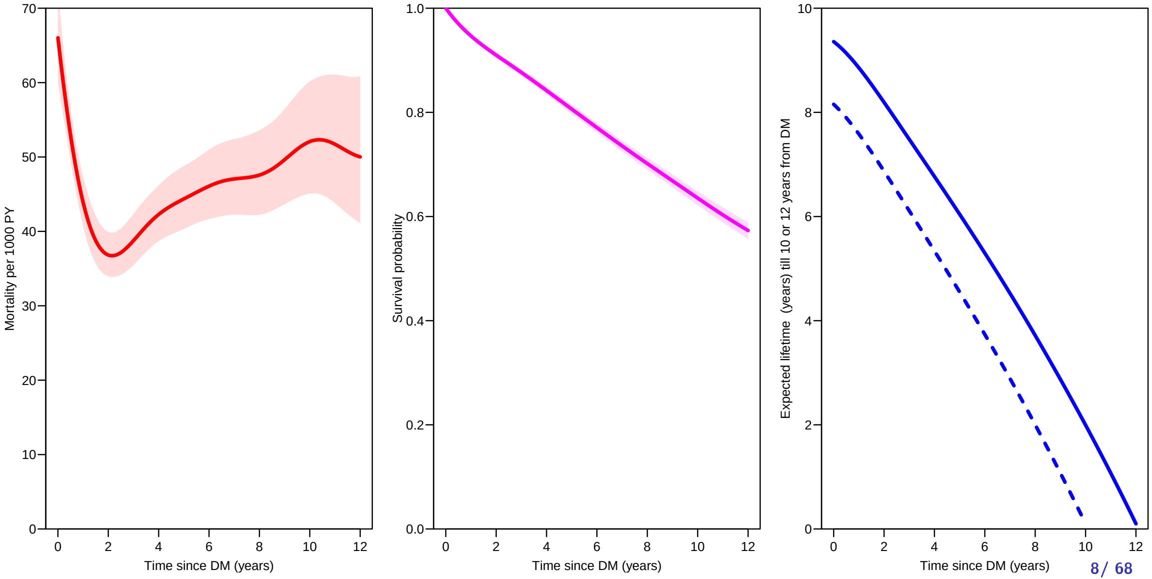
- survival probability (since time a):

$$\mathrm{S}_a(t) = \exp\left(-\int_a^t \lambda(u)\,\mathrm{d}u\right)$$

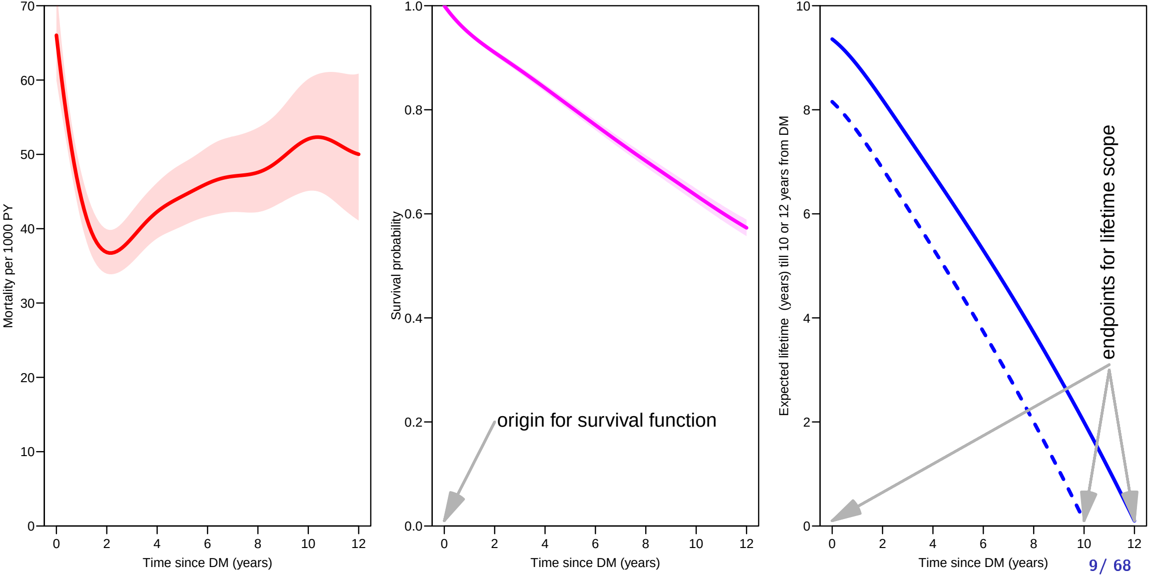
- sojourn time (between a and b)
  (restricted mean survival time till b, RMST):

$$\mathrm{L}(a, b) = \int_a^b \mathrm{S}_a(u)\,\mathrm{d}u$$

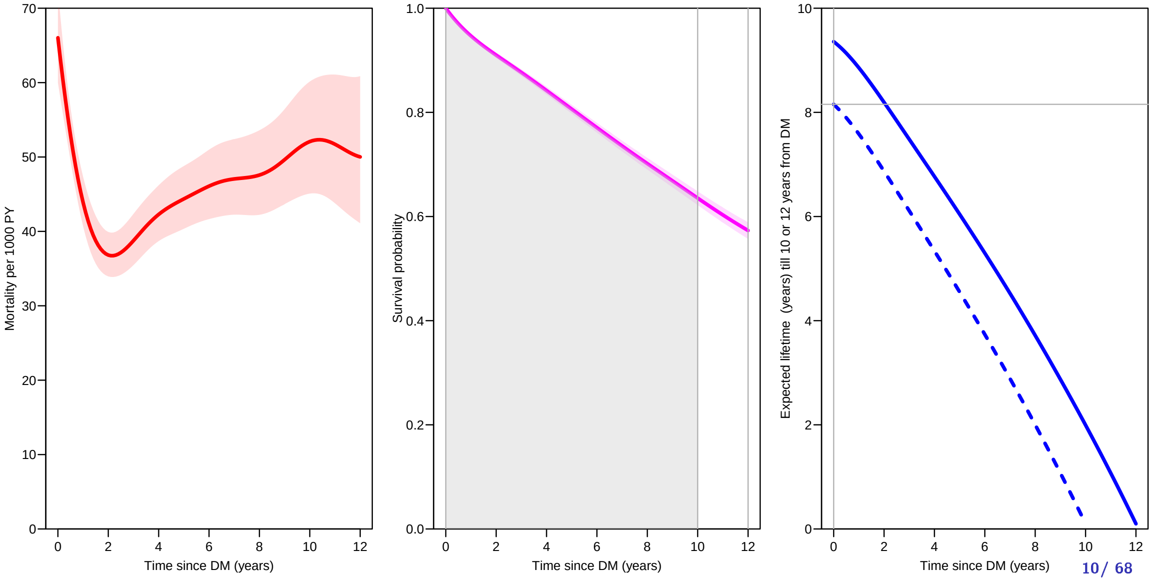
Mortality / survival / life time after DM



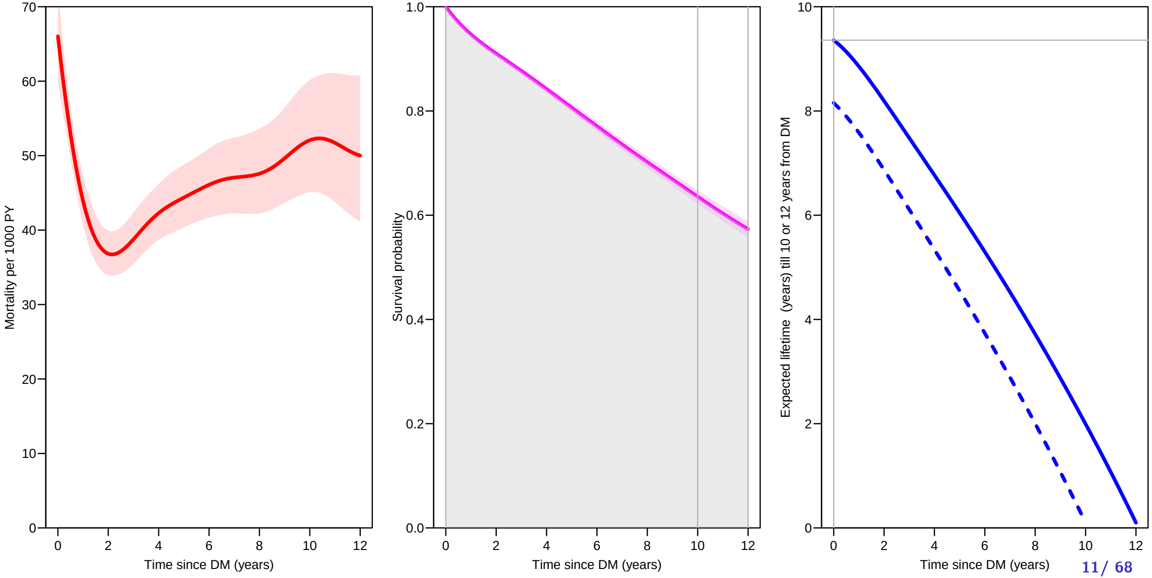
Mortality / survival / life time after DM



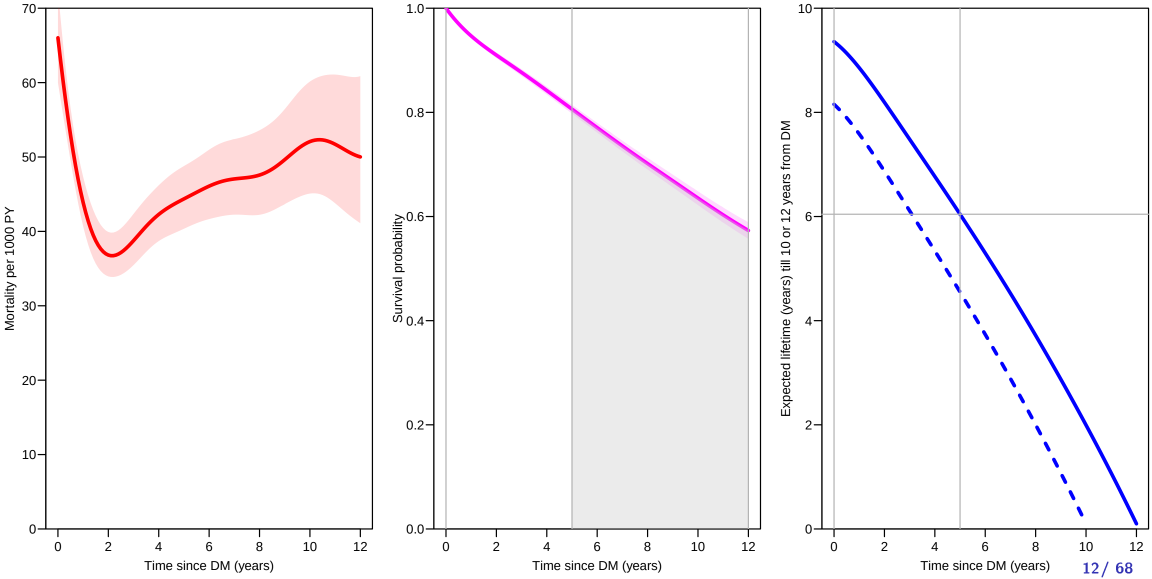
Mortality / survival / life time after DM



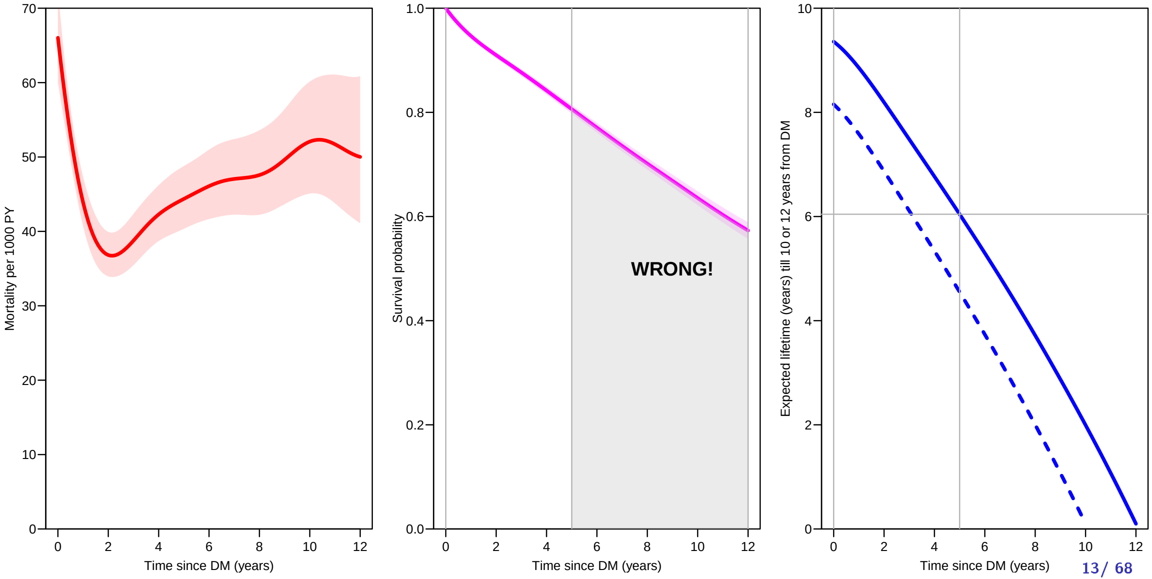
Mortality / survival / life time after DM



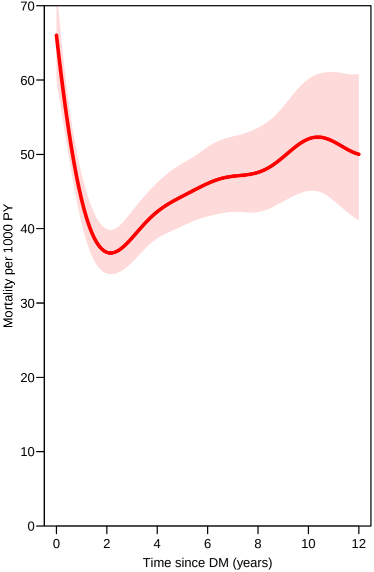
Mortality / survival / life time after DM

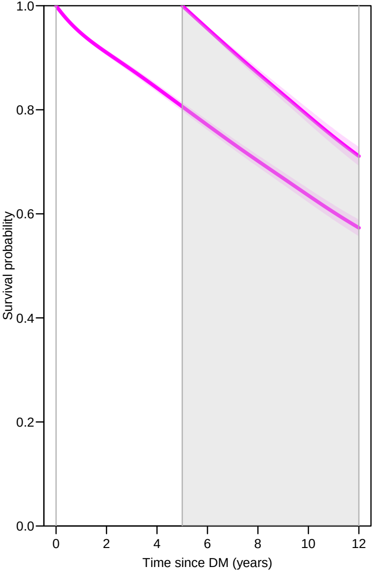


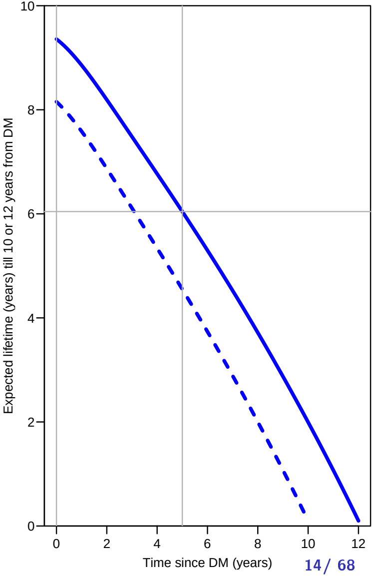
Mortality / survival / life time after DM



Mortality / survival / life time after DM







Disease demography

Demographic analyses of register event rates requires knowledge of **events** as well as **population time** covered by the register:

1. population size (risk time) by sex, age, date and other variables available both in the register and population; **tabular** data, such as that available from Statistikbanken at DST.
   Tabulation of events from the register.
2. **individual level** follow-up for **all** persons in the population — basically knowledge of entry (birth or immigration) and exit (death or emigration).
   In DK available as the LifeLines register at DST:
   individual follow-up of the entire DK population

How does follow-up look in a dataset

- One record per time **interval** (where nothing happens)
- Several records from a person:
  The observational units is (d_{pi}, y_{pi}) from person p, interval i.
- things happen at the **end** of the interval: outcome d
- the FU belongs in a particular **state**, e.g.:
  - noDM / T1 / T2
  - noCKD / CKD
  - no comorb. / 1 comorb. / 2 comorb. / 3 comorb. / ...
  - albuninuria: Norm / Micro / Macro (recurrence possible)
- the **length** of the interval is the outcome y

How does follow-up look in a dataset

- Intervals may further be classified by **time-varying** variables:
    - quantitative **deterministic** variables (time scales): age, date of follow up, disease duration, e.g. age_{pi}
    - quantitative random variables: HbA_{1c}, cholesterol, …
    - categorical random variables: parity, marital status
- States are a special type of time varying covariates: targets of demographic measures (probability, sojourn time)

```
> library(Epi)
> library(tidyverse)
> data(DMlate)
> DMlate[13:19,]
```

```
       sex    dobth     dodm   dodth    dooad    doins      dox
119305   M 1938.107 1997.461 1998.35       NA       NA 1998.350
188248   F 1979.864 1999.684      NA       NA       NA 2009.997
38336    M 1944.420 2002.550      NA       NA 2005.354 2009.997
368534   F 1962.482 2000.355      NA 2001.559       NA 2009.997
139497   F 1956.439 1995.544      NA       NA       NA 2009.997
132331   M 1935.024 1996.746      NA 1997.915 2005.995 2009.997
228434   F 1949.622 2006.783      NA 2006.783       NA 2009.997
```

Each record: relevant dates for a person followed from date of diabetes till death or 2009-12-31 (end of study).

—combination of several registers

—a collection of **dates**

Total follow-up of diabetes ptt.

In terms of follow-up we must define:

- Entry time: `dodm`
- Exit time: `dox`
- Event death: `dodth` = `dox`

states of follow-up by (any) drug-exposure:
no drug / OAD / Insulin

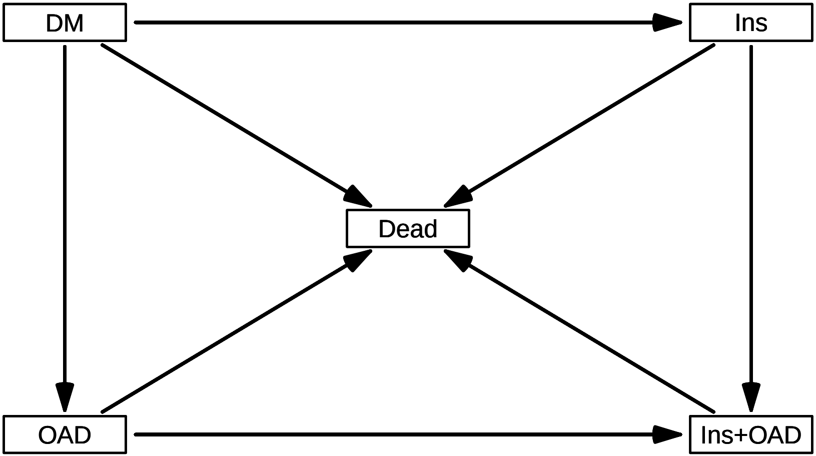
Intermediate register events

Other dates specify occurrence of intermediate events

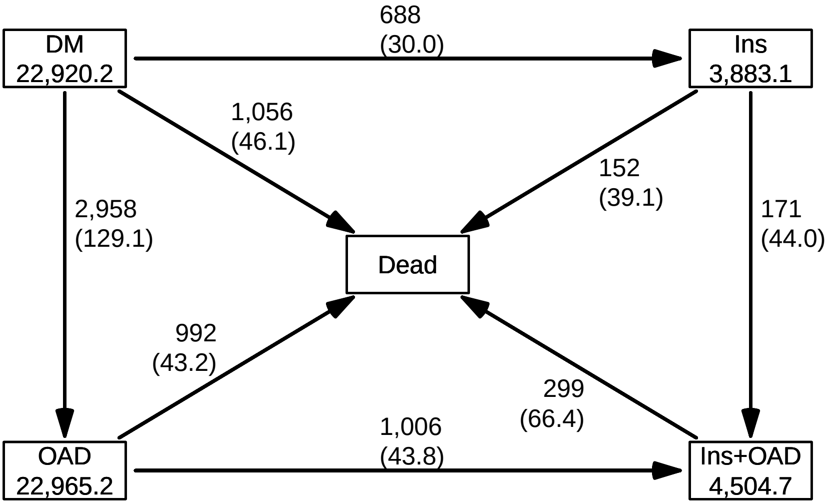
- start of OAD drugs at `doOAD`
- start of insulin at `doIns`
- possible states:
  - `DM`, no drug
  - `OAD` alone
  - `Ins` alone
  - both `OAD` & `Ins`
  - or:
    - `OAD` after `Ins`
    - `Ins` after `OAD`
  - `Dead`

States are derived from data, but **defined** by the investigator

Multi-state model — 5 states, 8 transitions



Multi-state data



DM
22,920.2

Ins
3,883.1

Dead

OAD
22,965.2

Ins+OAD
4,504.7

688
(30.0)

1,056
(46.1)

152
(39.1)

2,958
(129.1)

171
(44.0)

992
(43.2)

299
(66.4)

1,006
(43.8)

Representation of multistate follow-up data

- provide an overview of the follow-up:
  who is where, when, how
- where: state
- when: timescales
- how: other covariates
- provide analytical possibility for **rate** models:
  modeling on the observation scale (observed rates (d, y))

Multi-state data representation with Lexis

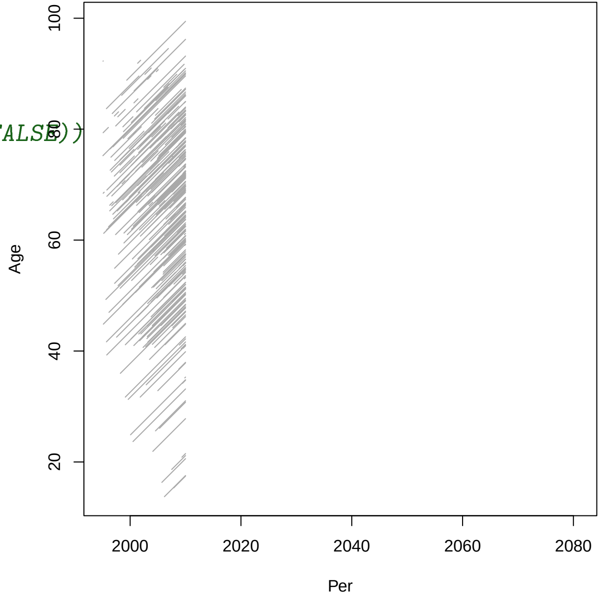
```
> dmL <- Lexis(entry = list(Per = dodm,
+                           Age = dodm - dobth,
+                         DMdur = 0 ),
+               exit = list(Per = dox),
+        exit.status = factor(!is.na(dodth),
+                             labels = c("DM", "Dead")),
+               data = DMlate)
NOTE: entry.status has been set to "DM" for all.
NOTE: Dropping  4  rows with duration of follow up < tol
> summary(dmL)
Transitions:
     To
From   DM Dead  Records:  Events: Risk time:  Persons:
  DM 7497 2499      9996     2499   54273.27      9996
```

Initial set-up for transition `DM -> Dead`, ignoring intermediate events

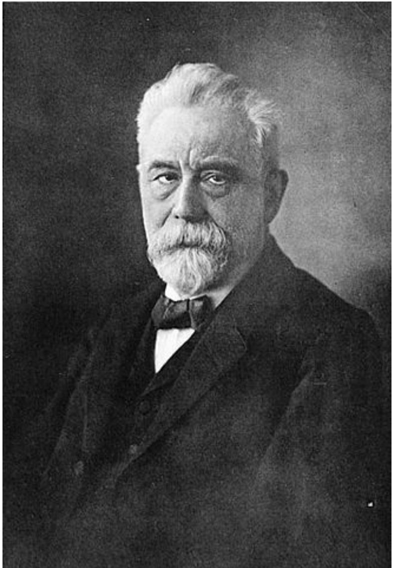
Multiple time scales: `Per`, `Age`, `DMdur`

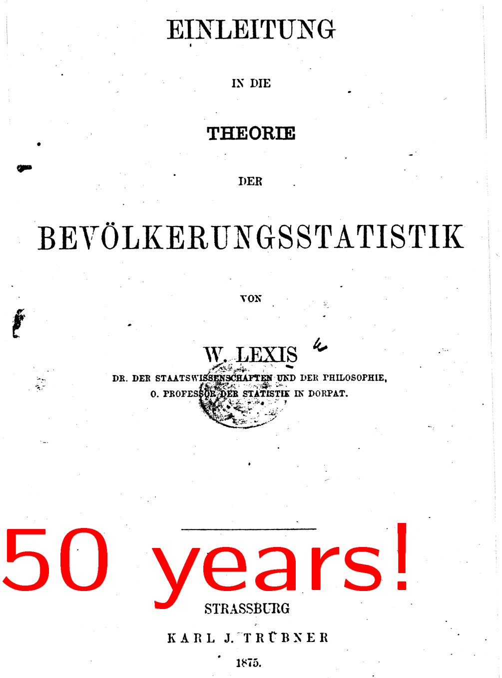
A Lexis diagram

```
> plot(dmL)
> plot(bootLexis(dmL,
+                      300,
+                      replace = FALSE))
```



Wilhelm Lexis





EINLEITUNG

IN DIE

THEORIE

DER

BEVÖLKERUNGSSTATISTIK

VON

W. LEXIS

DR. DER STAATSWISSENSCHAFTEN UND DER PHILOSOPHIE,
O. PROFESSOR DER STATISTIK IN DORPAT.

STRASSBURG

KARL J. TRÜBNER

1875.

150 years!

Multiple states: intermediate events `OAD` and `Ins`

```
> dmIO <- mcutLexis(dmL,
+                      wh = c("dooad", "doins"),
+               timescale = "Per",
+              new.states = c("OAD", "Ins"),
+              seq.states = FALSE,
+            ties.resolve = 1/365.25)
NOTE: Precursor states set to DM
NOTE: 15 records with tied events times resolved (adding 0.002737851 random unifo
      so results are only reproducible if the random number seed was set.
> summary(dmIO)

Transitions:
     To
From          DM Dead  OAD  Ins Ins+OAD  Records:  Events: Risk time:  Persons:
  DM        2830 1056 2958  688       0      7532     4702   22920.25      7532
  OAD          0  992 3327    0    1006      5325     1998   22965.25      5325
  Ins          0  152    0  462     171       785      323    3883.07       785
  Ins+OAD      0  299    0    0     878      1177      299    4504.69      1177
  Sum       2830 2499 6285 1150    2055     14819     7322   54273.27      9996
```

```
lex.id     Per   Age DMdur lex.dur lex.Cst lex.Xst   doins   dooad   dodth
     2 2003.31 64.09     0    6.69      DM      DM      NA 2007.45      NA
    15 2002.55 58.13     0    7.45      DM      DM 2005.35      NA      NA
    18 1996.75 61.72     0   13.25      DM      DM 2005.99 1997.92      NA
   770 1995.22 79.25     0    8.31      DM    Dead 1995.49 1995.64 2003.53
```

```
lex.id     Per   Age DMdur lex.dur lex.Cst lex.Xst doins   dooad dodth
     2 2003.31 64.09  0.00    4.14      DM     OAD    NA 2007.45    NA
     2 2007.45 68.23  4.14    2.55     OAD     OAD    NA 2007.45    NA
lex.id     Per   Age DMdur lex.dur lex.Cst lex.Xst   doins dooad dodth
    15 2002.55 58.13   0.0    2.80      DM     Ins 2005.35    NA    NA
    15 2005.35 60.93   2.8    4.64     Ins     Ins 2005.35    NA    NA
lex.id     Per   Age DMdur lex.dur lex.Cst lex.Xst   doins   dooad dodth
    18 1996.75 61.72  0.00    1.17      DM     OAD 2005.99 1997.92    NA
    18 1997.92 62.89  1.17    8.08     OAD Ins+OAD 2005.99 1997.92    NA
    18 2005.99 70.97  9.25    4.00 Ins+OAD Ins+OAD 2005.99 1997.92    NA
lex.id     Per   Age DMdur lex.dur lex.Cst lex.Xst   doins   dooad   dodth
   770 1995.22 79.25  0.00    0.27      DM     Ins 1995.49 1995.64 2003.53
   770 1995.49 79.52  0.27    0.15     Ins Ins+OAD 1995.49 1995.64 2003.53
   770 1995.64 79.67  0.42    7.89 Ins+OAD    Dead 1995.49 1995.64 2003.53
```

`lex.Cst` is the Current `state`   `lex.Xst` is the eXit `state`

Multistate model data representation

- If all transition times are known (register data):
  - one record per follow-up interval (transient states)
    —representation of follow-up—`Epi` and `survival` package
    "Andersen-Gill" representation
  - one record per likelihood term (transitions)
    stacked data—`mstate` package
- state occupancy known at (some) times (panel data):
  (person `p` is in state `s` at time `t`)
  "prevalence", panel data—`msm` package

We stick to representation of follow-up time (d, y)
—the most natural representation for register-based data

Likelihood for multistate transition rates

- assume all transitions and -times known exactly
- likelihood from one person is a **product** of terms
- constant rates ⇒ log-likelihood a **sum** of terms like:

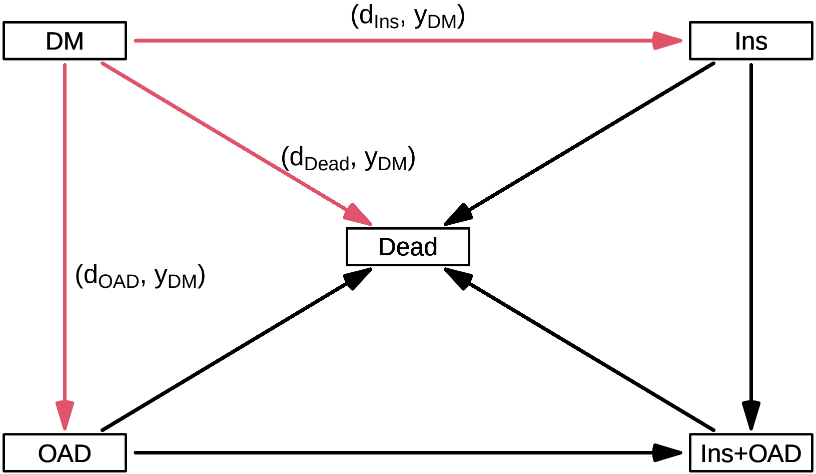
$$d \log(\lambda) - \lambda y$$

- —one term for each **possible** transition between states.
- for state `DM` **one** record but
  three likelihood terms, different ds, same y:

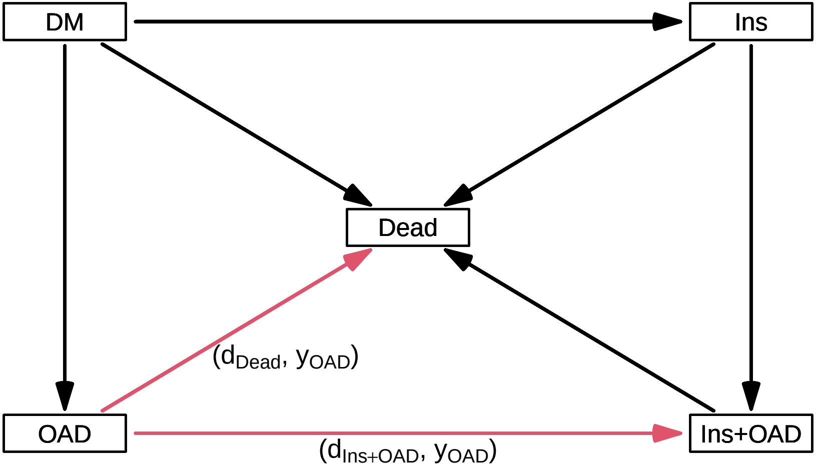
$$d_{\text{OAD}} \log(\lambda_{\text{OAD}}) - \lambda_{\text{OAD}} y_{\text{DM}} + d_{\text{Ins}} \log(\lambda_{\text{Ins}}) - \lambda_{\text{Ins}} y_{\text{DM}} + d_{\text{Dead}} \log(\lambda_{\text{Dead}}) - \lambda_{\text{Dead}} y_{\text{DM}}$$

— looks like independent Poisson variates

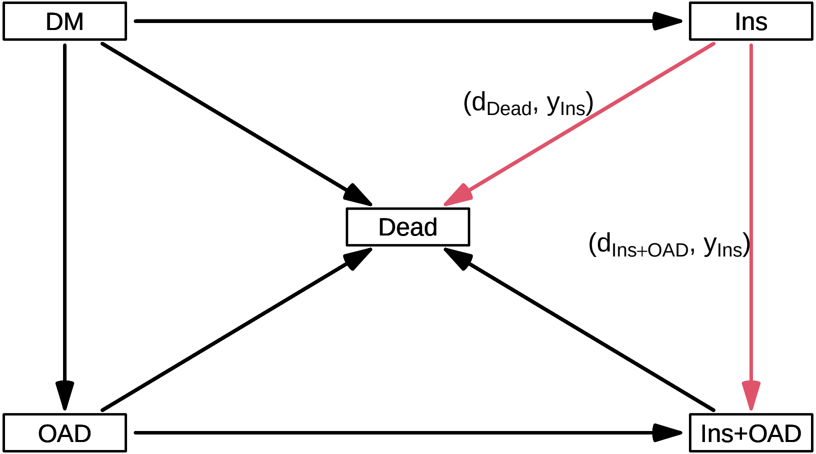
Total multi-state likelihood — 5 states, 8 transitions



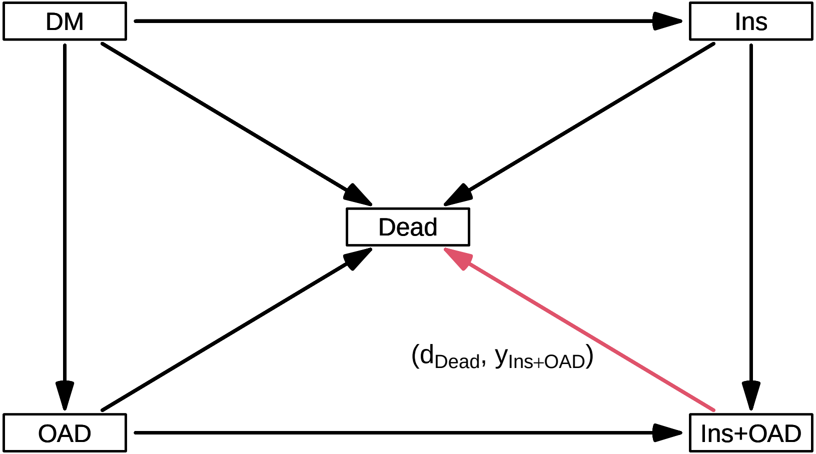
Total multi-state likelihood — 5 states, 8 transitions



Total multi-state likelihood — 5 states, 8 transitions



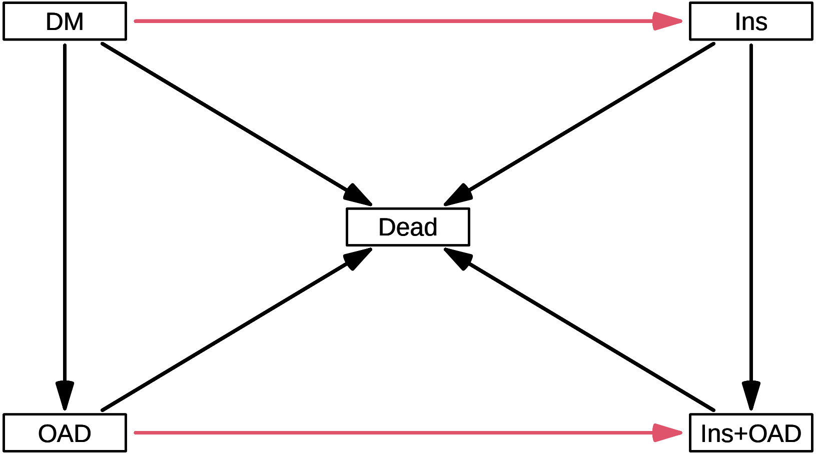
Total multi-state likelihood — 5 states, 8 transitions



Models for transition rates

- For rates in the same model: common parameters possible e.g. same age effect for different rates
- `Lexis` represents FU-time—**not** likelihood terms
- ⇒ analysis of a model for different rates from **different** states can be done based on a `Lexis` object
- Analysis of a model for different rates from **the same** state requires a stacked data frame
- . . . but this is hardly ever relevant, e.g.:
    - do not expect age effect to be the same for rate of `OAD` and `Ins`
    - In practice only rates from **different** origin states are analyzed together, such as `Ins` rates from `DM` resp. `OAD`

Partial multi-state likelihood — rates of `Ins`



DM

Ins

Dead

OAD

Ins+OAD

Modeling rates

- Poisson likelihood is for constant rates:
- $\Rightarrow$ model restricted to constant rate within each FU-record
- remedy: split FU-records in many records with shorter length interval (t_e, t_x), $t_e < t_1 < t_2 < t_x$:

$$\begin{aligned}\mathrm{P}\left\{d \text{ at } t_x \,|\, \text{entry at } t_e\right\} = & \mathrm{P}\left\{\text{survive } (t_e, t_1] \,|\, \text{alive at } t_e\right\} \times \\ & \mathrm{P}\left\{\text{survive } (t_1, t_2] \,|\, \text{alive at } t_1\right\} \times \\ & \mathrm{P}\left\{\text{survive } (t_2, t_x) \,|\, \text{alive at } t_2\right\} \times \\ & \mathrm{P}\left\{d \text{ at } t_x \,|\, \text{alive just before } t_x\right\}\end{aligned}$$

- include the ts as **quantitative** variables
- constant rates only in each small interval
- likelihood is a product of terms

Modeling rates

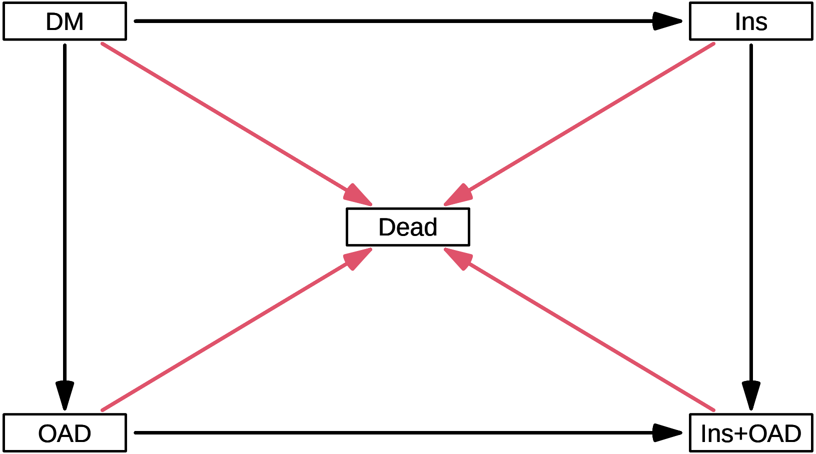
- constant rates only in each small interval
- likelihood is a product of terms
- each term looks like a Poisson likelihood term
- the total likelihood is a product of terms:
- looks as likelihood for independent Poisson variates
- . . . but they are neither independent nor Poisson
- there is not a one-to-one correspondence between models and likelihood—different models can have the same likelihood

```
> summary(dmIO)

Transitions:
     To
From          DM Dead  OAD  Ins Ins+OAD  Records:  Events: Risk time:  Persons:
  DM        2830 1056 2958  688       0      7532     4702   22920.25      7532
  OAD          0  992 3327    0    1006      5325     1998   22965.25      5325
  Ins          0  152    0  462     171       785      323    3883.07       785
  Ins+OAD      0  299    0    0     878      1177      299    4504.69      1177
  Sum       2830 2499 6285 1150    2055     14819     7322   54273.27      9996

> sIO <- splitLexis(dmIO, seq(0,20,0.1), "DMdur")
> summary(sIO)

Transitions:
     To
From            DM Dead    OAD   Ins Ins+OAD  Records:  Events: Risk time:  Persons
  DM        228333 1056   2958   688       0    233035     4702   22920.25      753
  OAD            0  992 231721     0    1006    233719     1998   22965.25      532
  Ins            0  152      0 39203     171     39526      323    3883.07       78
  Ins+OAD        0  299      0     0   45923     46222      299    4504.69      117
  Sum       228333 2499 234679 39891   47100    552502     7322   54273.27      999
```

How cut and split work

```
> subset(dmL , lex.id == 92)[, 1:11]

 lex.id     Per   Age DMdur lex.dur lex.Cst lex.Xst sex   dobth    dodm   dodth
     92 2008.56 55.15     0    0.57      DM    Dead   M 1953.41 2008.56 2009.13

> subset(dmIO, lex.id == 92)[, 1:11]

 lex.id     Per   Age DMdur lex.dur lex.Cst lex.Xst sex   dobth    dodm   dodth
     92 2008.56 55.15  0.00    0.25      DM     OAD   M 1953.41 2008.56 2009.13
     92 2008.81 55.39  0.25    0.33     OAD    Dead   M 1953.41 2008.56 2009.13

> subset( sIO, lex.id == 92)[, 1:11]

 lex.id     Per   Age DMdur lex.dur lex.Cst lex.Xst sex   dobth    dodm   dodth
     92 2008.56 55.15  0.00    0.10      DM      DM   M 1953.41 2008.56 2009.13
     92 2008.66 55.25  0.10    0.10      DM      DM   M 1953.41 2008.56 2009.13
     92 2008.76 55.35  0.20    0.05      DM     OAD   M 1953.41 2008.56 2009.13
     92 2008.81 55.39  0.25    0.05     OAD     OAD   M 1953.41 2008.56 2009.13
     92 2008.86 55.45  0.30    0.10     OAD     OAD   M 1953.41 2008.56 2009.13
     92 2008.96 55.55  0.40    0.10     OAD     OAD   M 1953.41 2008.56 2009.13
     92 2009.06 55.65  0.50    0.07     OAD    Dead   M 1953.41 2008.56 2009.13
```

Multi-state likelihood — mortality rates



DM

Ins

Dead

OAD

Ins+OAD

Mortality rates

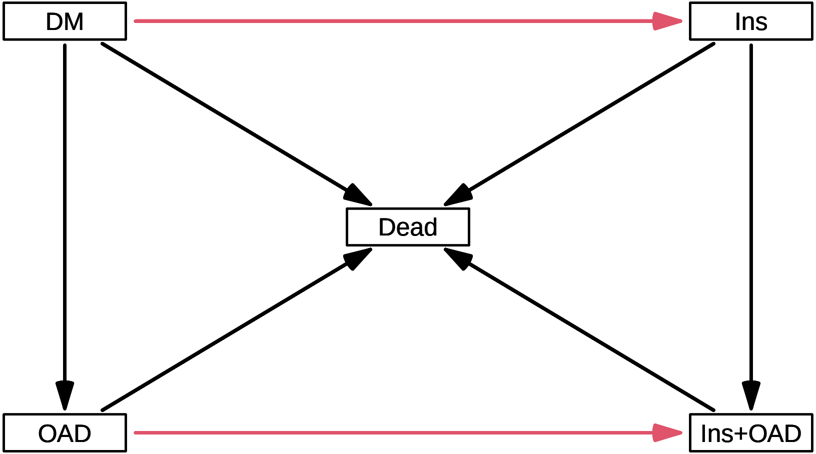
```
> mdth <- glmLexis(sIO, ~ Ns(DMdur, knots=c(0,1,3,6,10)) + lex.Cst,
+                        to = "Dead")
```

```
stats::glm Poisson analysis of Lexis object sIO with log link:
Rates for transitions:
DM->Dead
OAD->Dead
Ins->Dead
Ins+OAD->Dead
```

```
> round(ci.exp(mdth), 3)
```

```
                                       exp(Est.)  2.5% 97.5%
(Intercept)                                0.085 0.075 0.096
Ns(DMdur, knots = c(0, 1, 3, 6, 10))1      0.519 0.433 0.621
Ns(DMdur, knots = c(0, 1, 3, 6, 10))2      0.710 0.605 0.832
Ns(DMdur, knots = c(0, 1, 3, 6, 10))3      0.222 0.159 0.310
Ns(DMdur, knots = c(0, 1, 3, 6, 10))4      0.943 0.836 1.064
lex.CstOAD                                 0.973 0.891 1.063
lex.CstIns                                 0.880 0.742 1.045
lex.CstIns+OAD                             1.508 1.315 1.730
```

Multi-state likelihood — rates of `Ins`



Rates of insulin uptake

```
> mins <- glmLexis(sIO, ~ Ns(DMdur, knots=c(0,1,3,6,10)) + lex.Cst,
+                  from = c("DM", "OAD"),
+                    to = c("Ins", "Ins+OAD"))
```

```
stats::glm Poisson analysis of Lexis object sIO with log link:
Rates for transitions:
DM->Ins
OAD->Ins+OAD
```

```
> round(ci.exp(mins), 3)
```

```
                                         exp(Est.)  2.5% 97.5%
(Intercept)                                  0.215 0.195 0.238
Ns(DMdur, knots = c(0, 1, 3, 6, 10))1       0.137 0.109 0.173
Ns(DMdur, knots = c(0, 1, 3, 6, 10))2       0.358 0.294 0.437
Ns(DMdur, knots = c(0, 1, 3, 6, 10))3       0.002 0.001 0.003
Ns(DMdur, knots = c(0, 1, 3, 6, 10))4       1.608 1.359 1.903
lex.CstOAD                                   1.822 1.650 2.013
```

Rates of OAD uptake

```
> moad <- glmLexis(sIO, ~ Ns(DMdur, knots=c(0,1,3,6,10)) + lex.Cst,
+                    from = c("DM", "Ins"),
+                      to = c("OAD", "Ins+OAD"))
```

```
stats::glm Poisson analysis of Lexis object sIO with log link:
Rates for transitions:
DM->OAD
Ins->Ins+OAD
```

```
> round(ci.exp(moad), 3)
```

```
                                      exp(Est.)  2.5% 97.5%
(Intercept)                               0.732 0.689 0.777
Ns(DMdur, knots = c(0, 1, 3, 6, 10))1     0.213 0.179 0.255
Ns(DMdur, knots = c(0, 1, 3, 6, 10))2     0.155 0.126 0.192
Ns(DMdur, knots = c(0, 1, 3, 6, 10))3     0.004 0.003 0.005
Ns(DMdur, knots = c(0, 1, 3, 6, 10))4     0.366 0.306 0.439
lex.CstIns                                0.469 0.402 0.548
```

What not to do

```
> mDM <- glmLexis(sIO, ~ Ns(DMdur, knots=c(0,1,3,6,10)), from = "DM")
NOTE:
Multiple transitions *from* state ' DM ' - are you sure?
The analysis requested is effectively merging outcome states.
You may want analyses using a *stacked* dataset - see ?stack.Lexis
stats::glm Poisson analysis of Lexis object sIO with log link:
Rates for transitions:
DM->Dead
DM->OAD
DM->Ins
> round(ci.exp(mDM), 3)
                                      exp(Est.)  2.5% 97.5%
(Intercept)                               1.170 1.115 1.229
Ns(DMdur, knots = c(0, 1, 3, 6, 10))1     0.217 0.188 0.250
Ns(DMdur, knots = c(0, 1, 3, 6, 10))2     0.178 0.151 0.211
Ns(DMdur, knots = c(0, 1, 3, 6, 10))3     0.004 0.003 0.005
Ns(DMdur, knots = c(0, 1, 3, 6, 10))4     0.513 0.447 0.588
```

The model is meaningless, not statistically meaningless, but substantially meaningless

—not sensible to have same age effect for different event types

Predictions

Going from rates → probabilities → sojourn times
. . . uses integration, even double integration

- state probabilities not so simple with multiple time scales
- simulation is the way to go
- sojourn times easy from state probabilities or simulated data

Predictions by simulation: `simLexis`

Two things needed for prediction:

- complete **model** for all transitions
  —possibly made up from different models for subsets of transitions: the same model can be used for more than one transition
- prediction **data frame** (baseline):
  one record per person with starting values for **all** covariates

Goal: simulate a cohort starting as the prediction data frame going through time according to the model.

Predictions by `simLexis`: transition rates

Transition models fitted: `mdth`, `mins`, `moad`

```
> Tm <- list(DM = list(Ins = mins,
+                      OAD = moad,
+                     Dead = mdth),
+            Ins = list("Ins+OAD" = moad,
+                          Dead = mdth),
+            OAD = list("Ins+OAD" = mins,
+                          Dead = mdth),
+      "Ins+OAD" = list(Dead = mdth))
> unlist(lapply(Tm, names))
      DM1       DM2       DM3      Ins1      Ins2      OAD1      OAD2   Ins+OAD 
    "Ins"     "OAD"    "Dead" "Ins+OAD"    "Dead" "Ins+OAD"    "Dead"    "Dead" 
```

Predictions by `simLexis`: baseline

Prediction data frame (baseline),
specifies values for all variables in total model (here `DMdur`, `lex.Cst`)
—must be a `Lexis` object (to know timescale variables)

```
> bline <- sIO[1,]
> bline[1, "DMdur"] <- 0
> bline[1, "lex.Cst"] <- "DM"
> bline[, 1:7]
```

```
 lex.id     Per   Age DMdur lex.dur lex.Cst lex.Xst
      1 1998.92 58.66     0     0.1      DM      DM
```

Predictions by `simLexis`: simulated cohort

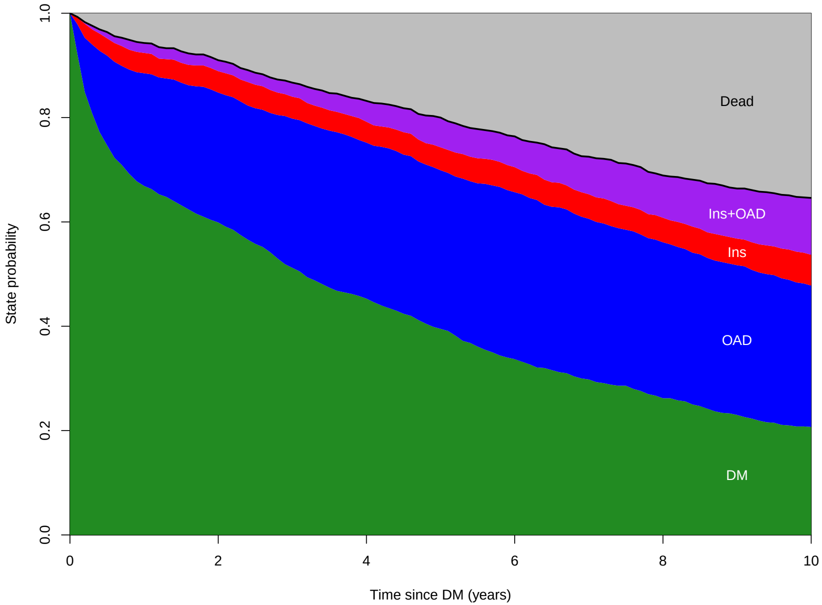
```
> system.time( simL <- simLexis(Tr = Tm, init = bline, N = 1000) )

  bruger    system forløbet 
    1.97      0.23      2.31 

> simL <- Relevel(simL, c("DM", "OAD", "Ins", "Ins+OAD", "Dead"))
> summary(simL)

Transitions:
     To
From      DM OAD Ins Ins+OAD Dead  Records:  Events: Risk time:  Persons:
  DM      50 557 136       0  257      1000      950    5428.24      1000
  OAD      0  90   0     278  189       557      467    4498.08       557
  Ins      0   0  51      32   53       136       85    1015.57       136
  Ins+OAD  0   0   0     170  140       310      140    1986.81       310
  Sum     50 647 187     480  639      2003     1642   12928.70      1000
```

Predictions by `simLexis`: results

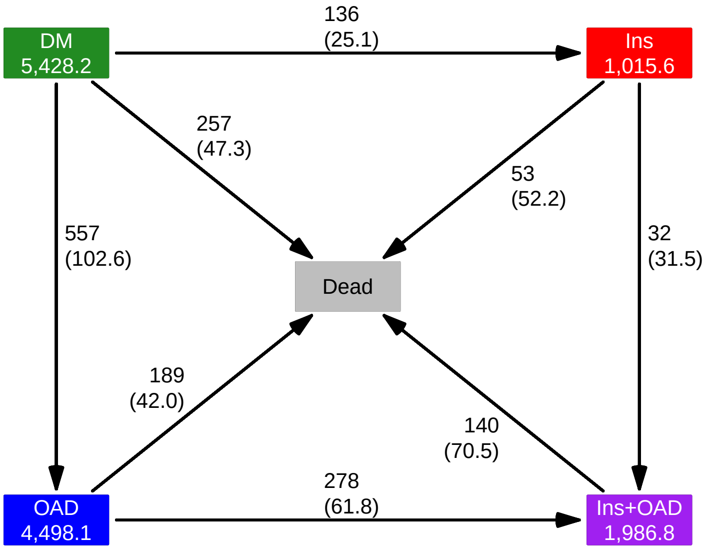
```
> timeScales(simL)

[1] "Per"   "Age"   "DMdur"

> head(nS <- nState(simL, from = 0,
+                         at = seq(0, 10, .1),
+                   time.scale = "DMdur"), 4)

     State
when    DM  OAD  Ins Ins+OAD Dead
  0   1000    0    0       0    0
  0.1  922   57   13       1    7
  0.2  849  104   27       3   17
  0.3  808  132   29       7   24

> head(pS <- pState(nS, perm = 1:5), 4)

     State
when     DM   OAD   Ins Ins+OAD Dead
  0   1.000 1.000 1.000   1.000    1
  0.1 0.922 0.979 0.992   0.993    1
  0.2 0.849 0.953 0.980   0.983    1
  0.3 0.808 0.940 0.969   0.976    1
```

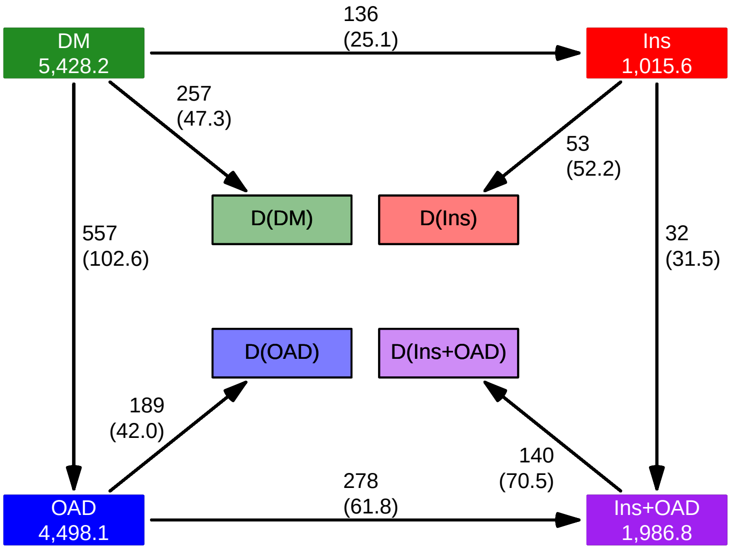
Predictions by `simLexis`: results

```
> clr <- c("forestgreen", "blue", "red", "purple", "gray")
> cld <- c(clr[1:4], adjustcolor(clr[4:1], alpha = 0.3))
> plot(pS, col = clr,
+      xlab = "Time since DM (years)",
+      ylab = "State probability")
> lines(rownames(pS), pS[, 4], lwd = 2)
> #
> mid <- function(x) x[-1] - diff(x) / 2
> text(9, mid(c(0, pS["9",])), colnames(pS), col = c(rep("white", 4), "black"))
```



State probability
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
Time since DM (years)
Dead
Ins+OAD
Ins
OAD
DM


136
(25.1)
DM
5,428.2
Ins
1,015.6
257
(47.3)
53
(52.2)
557
(102.6)
32
(31.5)
Dead
189
(42.0)
140
(70.5)
278
(61.8)
OAD
4,498.1
Ins+OAD
1,986.8



DM
5,428.2

Ins
1,015.6

D(DM)

D(Ins)

D(OAD)

D(Ins+OAD)

OAD
4,498.1

Ins+OAD
1,986.8

136
(25.1)

257
(47.3)

53
(52.2)

557
(102.6)

32
(31.5)

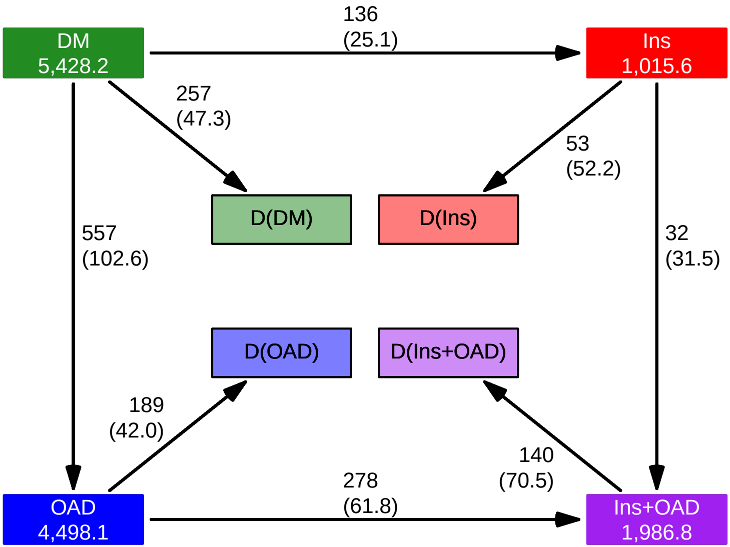
189
(42.0)

140
(70.5)

278
(61.8)

Death subdivided by state at death

```
> simX <- mutate(simL, lex.Xst = as.character(lex.Xst),
+                      lex.Xst = ifelse(lex.Xst == "Dead",
+                                       paste0("D(", lex.Cst, ")"),
+                                       lex.Xst))
> simX <- factorize(simX)
> simX <- Relevel(simX, c(             levels(simX)[1:4],
+                          paste0("D(", levels(simX)[4:1], ")")))
> summary(simX)
Transitions:
     To
From      DM OAD Ins Ins+OAD D(Ins+OAD) D(Ins) D(OAD) D(DM)  Records:  Events: Ri
  DM      50 557 136       0          0      0      0   257      1000      950
  OAD      0  90   0     278          0      0    189     0       557      467
  Ins      0   0  51      32          0     53      0     0       136       85
  Ins+OAD  0   0   0     170        140      0      0     0       310      140
  Sum     50 647 187     480        140     53    189   257      2003     1642

Transitions:
     To
From       Persons:
  DM           1000
```



DM
5,428.2

136
(25.1)

Ins
1,015.6

257
(47.3)

53
(52.2)

D(DM)

D(Ins)

557
(102.6)

32
(31.5)

D(OAD)

D(Ins+OAD)

189
(42.0)

140
(70.5)

278
(61.8)

OAD
4,498.1

Ins+OAD
1,986.8

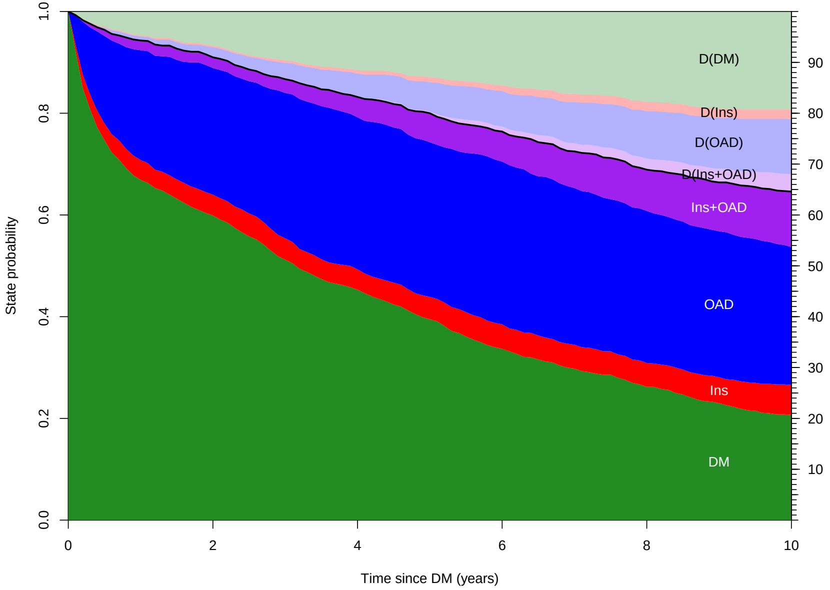
Subdivided death state

```
> head(nS <- nState(simX, from = 0, at = seq(0, 10, .1), time.scale = "DMdur"))
     State
when    DM OAD Ins Ins+OAD D(Ins+OAD) D(Ins) D(OAD) D(DM)
  0   1000   0   0       0          0      0      0     0
  0.1  922  57  13       1          0      0      0     7
  0.2  849 104  27       3          0      0      1    16
  0.3  808 132  29       7          0      1      2    21
  0.4  772 156  33       8          0      2      3    26
  0.5  747 172  33      12          0      2      4    30
```

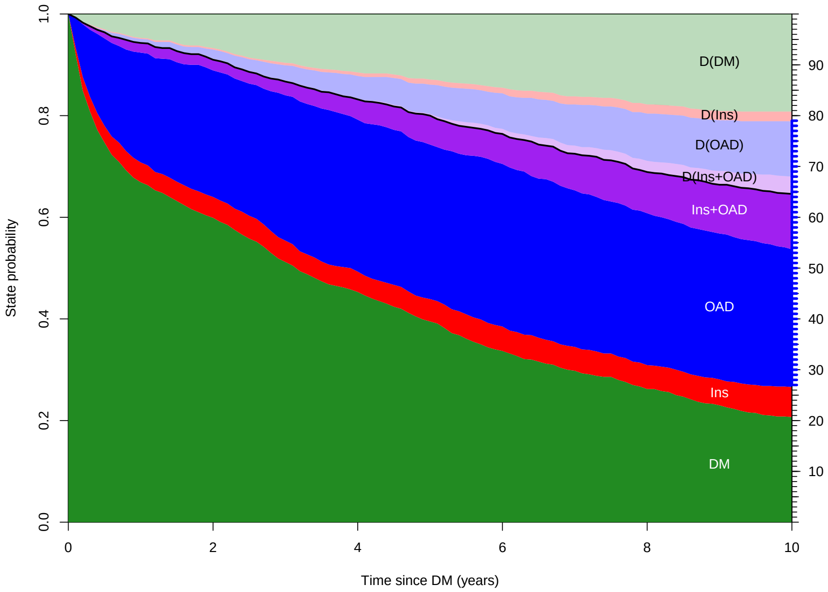
```
> head(pS <- pState(nS))
     State
when     DM   OAD   Ins Ins+OAD D(Ins+OAD) D(Ins) D(OAD) D(DM)
  0   1.000 1.000 1.000   1.000      1.000  1.000  1.000     1
  0.1 0.922 0.979 0.992   0.993      0.993  0.993  0.993     1
  0.2 0.849 0.953 0.980   0.983      0.983  0.983  0.984     1
  0.3 0.808 0.940 0.969   0.976      0.976  0.977  0.979     1
  0.4 0.772 0.928 0.961   0.969      0.969  0.971  0.974     1
  0.5 0.747 0.919 0.952   0.964      0.964  0.966  0.970     1
```

Subdivided death state

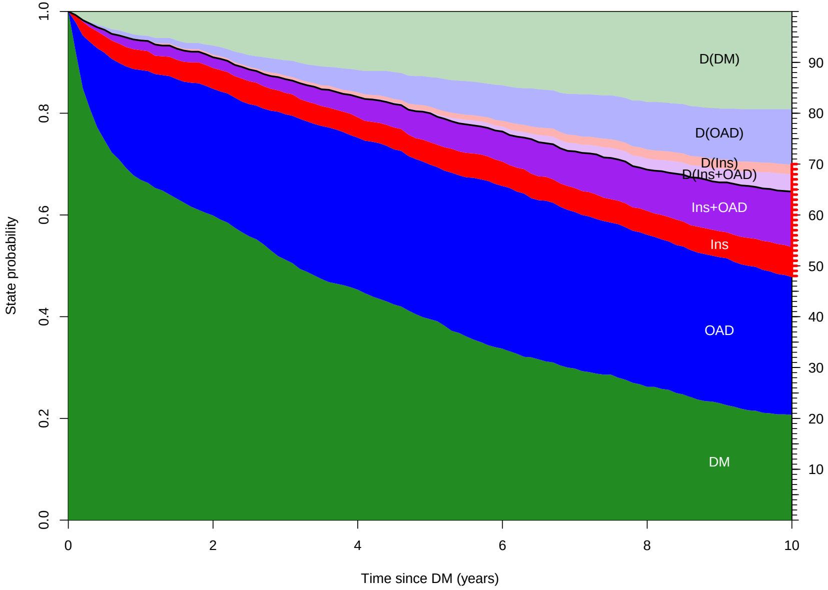
```
> plot(pS, col = clr,
+          xlab = "Time since DM (years)",
+          ylab = "State probability")
> lines(rownames(pS), pS[, 4], lwd = 2)
> mid <- function(x) x[-1] - diff(x) / 2
> text(9, mid(c(0, pS["9",])), colnames(pS), col = rep(c("white", "black"), each
> axis(side = 4, at = tk <- 1:9/10, labels = tk * 100, las = 1)
> axis(side = 4, at = 0:20/20, labels = NA, tcl = -0.4)
> axis(side = 4, at = 1:99/100, labels = NA, tcl = -0.3)
> # hasins <- round(100 * pS["10", c("OAD", "D(Ins)")])
> # axis(side = 4, at = hasins[1]:hasins[2] / 100, labels = NA,
> #       tcl = -0.3, col = "red", lwd =3)
```



State probability
0.0
0.2
0.4
0.6
0.8
1.0
Time since DM (years)
0
2
4
6
8
10
10
20
30
40
50
60
70
80
90
D(DM)
D(Ins)
D(OAD)
D(Ins+OAD)
Ins+OAD
OAD
Ins
DM



P(OAD before 10 y):
$79 - 27 = 52\%$



P(Ins before 10 y):
$70 - 48 = 22\%$

RMST 0–10 years

- What is the expected time spent without medication during the first 5 or 10 years after diagnosis of diabetes?
- This is just the green area in the figure
- We simulated 1000 persons, so `nS` is the state probability in 1/1000s, every 0.1 years until 10 years.

```
> head(nS, 2)

     State
when    DM  OAD  Ins Ins+OAD D(Ins+OAD) D(Ins) D(OAD) D(DM)
  0   1000    0    0       0          0      0      0     0
  0.1  922   57   13       1          0      0      0     7

> round(RMST <- sum(mid(nS[1:51, "DM"]) * 0.1) / 1000, 3)

[1] 2.864

> round(RMST <- sum(mid(nS[1:101, "DM"]) * 0.1) / 1000, 3)

[1] 4.287
```

RMST directly from simulation

`simX` is a `Lexis` object, so we can just take the actual simulated lifetimes (`lex.dur`) before 10 years.

FU must be split at 5 and 10 years:

```
> simS <- splitLexis(simX, c(5,10), "DMdur")
> sum(subset(simS, lex.Cst == "DM" & DMdur <  5)$lex.dur) / 1000

[1] 2.864011

> sum(subset(simS, lex.Cst == "DM" & DMdur < 10)$lex.dur) / 1000

[1] 4.286286
```

Rates, survival, RMST

- **rates** on the observation scale, time^{-1}, depends on time(scales)
- **survival** is one type of state probability
- **state probabilities** are predictions, dimensionless, scale time^{0}, requires:
  - starting time(s)
  - baseline covariates
- **RMST** are integrals of state probabilities, scale time^{1}, requires:
  - starting and ending time (a time interval)
  - baseline covariates
- **Demography** uses expected (residual) lifetime at age a:

$$\mathrm{L}(a) = \int_{a}^{\infty} \mathrm{S}_a(u)\,\mathrm{d}u$$

Summary

- Registers provide **dates** of **events**
- defines **transition times** between defined **states**
- ... or time-dependent variables
- data representation and manipulation in `Lexis` objects
- `cutLexis` to introduce (dates of) intermediate states
- `splitLexis` to make intervals short to allow constant rate assumption
- (parametric) models for transition rates: `glmLexis`, `gamLexis`, `coxphLexis`
- simulation (`simLexis`) using rates used to predict state probabilities (survival) and RMST (expected life time)

Material

- Book: Bendix Carstensen:
  Epidemiology with R, Oxford University Press, 2022
- Book (draft) on line: Practical Multistate Modeling
  `https://bendixcarstensen.com/PMM/`
- Vignettes in the `Epi` package:
  - Analysis of follow-up data using the `Lexis` functions in `Epi`
  - Competing risks with `Lexis`, parametric rates and simulation based confidence intervals
  - Simulation of multistate models with multiple timescales: `simLexis`