Introduction to Clinical Epidemiology

Bendix Carstensen Steno Diabetes Center
Gentofte, Denmark
http://BendixCarstensen.com

NN Tomorrow Forum

Beijing, 14 March 2015

http://BendixCarstensen.com/Epi/NNTF

steno diabetes center

Epidemiology
— what is it all about

Introduction to
Clinical Epidemiology
Beijing, 14 March 2015
NN Tomorrow Forum
http://BendixCarstensen.com/Epi/NNTF

steno
diabetes center

Epidemiology

- **Clinical** epidemiology
- **Analytical** epidemiology
- **Descriptive** epidemiology
- **Medical demography**

. . . **in**creasing **size**
. . . **de**creasing **complexity**

Epidemiology — data

- **Clinical** epidemiology:
  Based on observations in daily care of patients, either in GP, specialised clinics or hospitals
- **Analytical** epidemiology:
  Based on specifically collected (population or patient) data, usually focused on a specific exposure or outcome
- **Descriptive** epidemiology:
  Based on population surveys or registers, hence with limited data on each individual
- **Medical demography**:
  Based on registers or extrapolations from surveys to the entire population

Epidemiology — aims

- **Clinical** epidemiology:
  Describe and monitor patients' course of disease, quality of care
- **Analytical** epidemiology:
  Quantify the effect of specific exposures such as occupation or lifestyle (BMI)
- **Descriptive** epidemiology:
  Population occurrence of diseases, *e.g.* prevalence of diabetes in the population
- **Medical demography**:
  Use population measure to describe how disease will spread in the population, the population burden of disease

Clinical records
e.g. SDC electronic patient records

Clinical epidemiology with a particlar view to the use of:

- **clinical records** available in clinics or hospitals
- **clinical registers** collecting clinical information systematically — compilation of (parts of) clinical records

Clinical records
e.g. SDC electronic patient records

- Complete history of patients:
  - HbA_{1c}
  - lipids
  - blood pressure
  - GFR
  - ...
- Information on:
  - **date** of diagnosis of diabetes (entry?)
  - **dates** of measurement of clinical variables
  - **date** of birth
  - **date** of complications
- **Note:** Intervals between visits depend on patients' status

Clinical records
e.g. SDC electronic patient records

. . . possibilities:

- Prevalence of patients with CKD (chronic kidney disease)
- Prevalence of patients with less than 2 years since last eye examination
- Prevalence of patinets currently prescribed statins

Prevalence means the percentage of patients that meets the criterion . . . more later . . .

Clinical registers
e.g. Danish Adult Diabetes Database

- Data collection (recording) at fixed intervals (once a year, e.g.)
- Clinical data on individuals
- Data collection independent of patients' clinical status w.r.t.
  - HbA_{1c}
  - lipids
  - ...
- Missing data:
  - a patient was not seen for an entire year
  - a patient has moved
  - a patient died (but was not recorded as such)

Clinical registers
e.g. Danish Adult Diabetes Database

. . . possibilities:

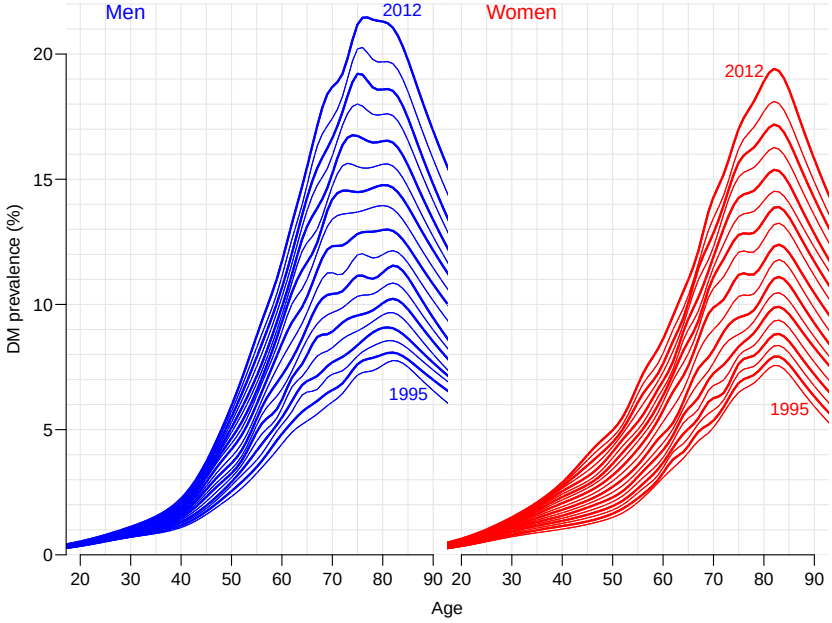
- Population revalence of patients with sysBP < 140
- Median blood pressure in the population

Median means the value where half is higher and half is lower

. . . more about population characteristics later . . .

Population level registers
e.g. Danish National Diabetes Register

- (cl)Aims to cover the entire population
- Limited information on each patient:
  - date of diagnosis of DM
  - date of birth
  - sex
  - date of death
  - ...
- Population levesl monitoring of:
  - DM occurrence (incidence rates)
  - prevalence of DM
  - mortality of DM patients
- Important for long term follow-up of the population



Prevalence of DM in Denmark at: 1.1.1995, 1.1.1996, . . . , 1.1.2012

From the Danish NDR

Study types and data types

Introduction to
Clinical Epidemiology
Beijing, 14 March 2015
NN Tomorrow Forum
http://BendixCarstensen.com/Epi/NNTF

steno
diabetes center

Epidemiological study types

- **Cross-sectional** studies:
  What is disease status at a particular date
- **Follow-up** studies:
  What is the rate of disease occurrence
  - Fixed cohorts, population based surveys
  - Dynamic cohorts
  - An entire population followed through registers
    Medical demography
- **Case-control** studies:
  Compare cases with non-cases.
  - Sampling based on disease status
  - Partial measures of disease occurrence/presence

Epidemiological data types

- **Continuous** (metric) responses can emerge from any observational design.
- **Categorical** response data essentially always **derived** from follow-up data:
  - Tables of counts from a cross-sectional study.
  - Tables of counts and follow-up time.
  - Tables of case-control status and exposure.
- Continuous and categorical explanatory variables occur in any design.

Cross-sectional studies

- What **fraction of the population** has a certain characteristic (such as a diagnosis of diabetes or other disease).
- **Observations:** the entire population (or a sample of it) classified by disease status
- The **likelihood** is a binomial likelihood for

$$p = \mathrm{P}\left\{\text{presence of disease}\right\}$$

- . . . that is, how p depends on explanatory variables such as:
  - sex
  - age
  - . . .

Cross-sectional studies

- What is the **population distribution** of a certain characteristic (such as HbA1c or other clinical measurement).
- **Observations:** The measured characteristics in the (patient)population (or a sample of it)
- The **likelihood** is a normal likelihood for the measurement

$$y_i = \mathrm{P}\left\{\text{measured value in individual}\right\} i$$

- . . . that is, how y_i depends on explanatory variables such as:
  - sex
  - age
  - duration of DM
  - . . .

Follow-up studies

- **Medical demography** — describing the entire population w.r.t. disease status over time
  - An entire population is followed for a particular event of interest (CVD, death, . . .)
- **Epidemiological** (observational) study
  Part of the population (a cohort) is followed for a limited period of time
  - May not necessarily be generalizable.
  - — but can elucidate the size of exposure effects on disease occurrence.
  - Neither exposures nor outcomes need be representative — only their relationship.

Follow-up studies: clinical measurements

- Measurements at each clinic visit
- Many measurements per person: (measurement,date): y_{it}
- Changes in measurements over time described by:
  - population mean
  - variation between individuals in: level & slope

Follow-up studies: clinical cohorts

- Entry to the cohort is **diagnosis** of disease
- Events are occurrence of *e.g.* **complications**
- Target is **rates** of complications
- . . . and how they depend on explanatory variables such as:
  - sex
  - age
  - duration of DM
  - . . .
- Measures derived from rates:
  - survival (without complications)
  - lifetime risk
  - . . . these are probabilities.

Follow-up studies

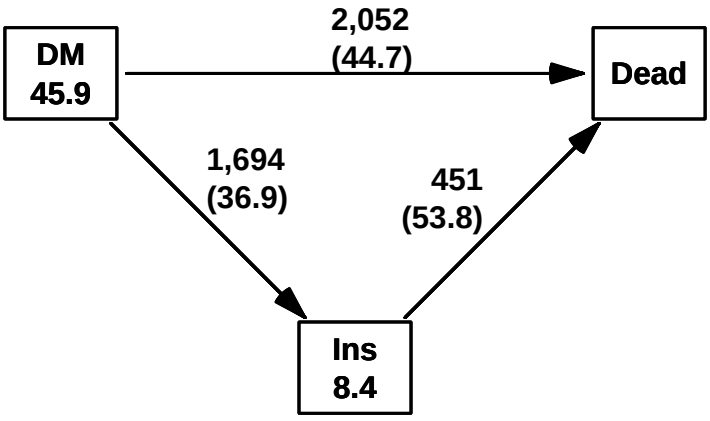
- **Observations** are (empirical) rates:
  (d, y): d events during y follow-up time
  (risk time, exposure time, person-years)
- **Models** for occurrence rates:

$$\lambda(t) = \mathrm{P}\left\{\text{event in } (t, t+h) | \text{ no event till } t\right\}/h$$

- The **likelihood** for this is proportional to a Poisson likelihood (if λ is constant):

$$\text{log-lik} = \ell(\lambda | d, y) = d\log(\lambda) + \lambda y$$

How a follow-up study looks



Follow-up studies — modelling

- Each transition can be considered separately
- Rates modelled separately (or jointly)
- Probabilities can be derived from estimated rates
- Simplest probability is:

$$S(t) = \mathrm{P}\left\{\text{survive till time } t\right\}$$

- Other probabilities of interest, *e.g.*:

$$P_c(t) = \mathrm{P}\left\{\text{die from cause } c \text{ before } t\right\}$$

— depends on more than one rate.

Case-control studies

- Events (cases) are sampled.
- But risk time is not...
  — it is replaced by a carefully chosen sample of the non-event persons.
- The likelihood is a binomial likelihood for

$$p = \mathrm{P}\left\{\text{case} \mid \text{included in the study}\right\}$$

  which contains the parameters of interest (and some not of any interest) *e.g.* rate-ratios.

Prevalence calculation: examples

Introduction to
Clinical Epidemiology
Beijing, 14 March 2015
NN Tomorrow Forum
http://BendixCarstensen.com/Epi/NNTF

steno
diabetes center

Computing a prevalence from clinical records

The prevalence of CKD (Chronic Kidney Disease) among SDC patients. . . . at 1 January 2008, say:

- Find all (n) patients in SDC records alive at this date:
  - date of diagnosis of DM $<$ 1 January 2012
  - date of death $>$ 1 January 2008
  - date of ESRD $>$ 1 January 2008
  - data of leaving SDC $>$ 1 January 2008
- Among these, find those (x) with CKD at this date:
  - date of diagnosis of CKD $<$1 January 2008

Note: it is all about using dates.

Computing a prevalence from clinical records

- $n = 1,421$ patients at the SDC clininc at 1.1.2012
- $x = 558$ of these had CKD
- Prevalence of CKD: $p = x/n = 0.393$
- Formula for a confidence interval:

$$\begin{aligned} \text{erf} &= \exp\left(1.96/\sqrt{np(1-p)}\right) \\ p_{\text{lower}} &= p/\left(p + (1-p) \times \text{erf}\right) \\ p_{\text{upper}} &= p/\left(p + (1-p) \ / \ \text{erf}\right) \end{aligned}$$

which gives: $(0.418, 0.368)$

Computing a prevalence from a population register

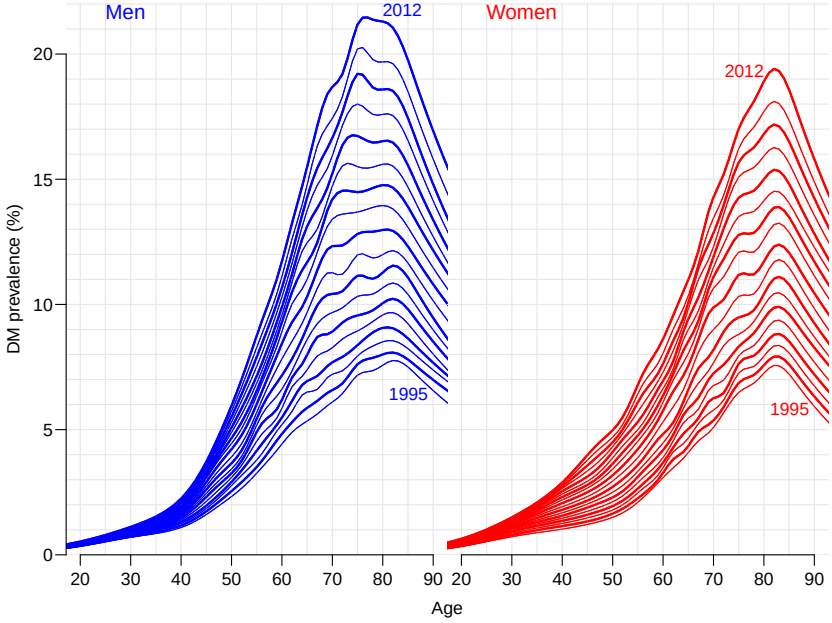
. . . at the 1 January 2008, say.

- Find all persons (x) in the register with:
  - date of diagnosis of DM $<$1.1.2008
  - date of death $>$1.1.2008 (or no date of death)
- Use Statistics Denmark to obtain the number of persons in the entire population as of 1.1.2008 (n)
- Calculation of the prevalence is a before: $p = x/n$
- Confidece intervals too:
  —they will be tiny beacuse of the large numbers.

Computing a prevalence from a population data base

In practice, the calculation is done as for the (x, n) except that:

- The numbers are large, so the confidence intervals are narrow — not worth considering
- The calculation is done for each (1-year) age-class
- The resulting age-curve is smoothed
- The exercise is repeated for the dates 1.1.2011, 1.1.2010, . . . , 1.1.1995
- Everything is done separately for men and women



Prevalence of DM in Denmark at: 1.1.1995, 1.1.1996, . . . , 1.1.2012

From the Danish NDR

Nephropathy complications example

Introduction to
Clinical Epidemiology
Beijing, 14 March 2015
NN Tomorrow Forum
http://BendixCarstensen.com/Epi/NNTF

steno
diabetes center

Renal disease and CVD in SDC T1 patients

- Patients with DN (diabetic nephropathy)
- Occurrence of ESRD (end stage renal disease: dialysis or transplant)
- Death
- How do rates of death and ESRD depend on clinical parameters?
- How is long-term outcome dependent on clinical status?

Improved Survival and Renal Prognosis of Patients With Type 2 Diabetes and Nephropathy With Improved Control of Risk Factors

Gudbjörg Andrésdóttir,[1] Majken L. Jensen,[1] Bendix Carstensen,[1] Hans-Henrik Parving,[2,3,4] Kasper Rossing,[5] Tine W. Hansen,[1] and Peter Rossing[1,3,4]

DOI: 10.2337/dc13-2036

OBJECTIVE

RESEARCH DESIGN AND METHODS

All patients with type 2 diabetes and DN (*n* = 543) at Steno Diabetes Center were followed during 2000–2010. GFR was measured yearly with ^{51}Cr-EDTA plasma clearance. Annual decline in GFR was determined in patients with at least three measurements over a minimum of 3 years (ΔGFR cohort, *n* = 286). Results were compared with historical data, obtained using identical criteria at our hospital, before implementation of current treatment guidelines.

RESULTS

Baseline mean (SD) GFR was 74 (32) mL/min/1.73 m^2. More than 93% received RAS inhibition. During median 7.8 (interquartile range 5.7–9.8) years, mean (SE) annual GFR decline was 4.4 (0.24) compared with previously 5.2 (0.27) mL/min/1.73 m^2/year (*P* = 0.04). Doubling of plasma creatinine or end-stage renal disease (ESRD) developed in 19%, and 37% died during 5.7 (3.3–8.8) years. Mortality from onset of DN in the ΔGFR cohort was compared with that of our prior ΔGFR cohort from 1983 to 2003 (*n* = 227). Crude mortality risk was reduced by 42% and after age adjustment by 50% (*P* < 0.001 for both). In a multistate model accounting for competing risks of ESRD and death, prior cardiovascular disease and lower GFR were predictors of mortality, whereas albuminuria, HbA$_{1c}$, and low GFR predicted ESRD.

CONCLUSIONS

Overall prognosis has improved considerably with current multifactorial treatment of DN in type 2 diabetes, including long-term RAS inhibition.

Diabetic nephropathy (DN) is a major complication of type 2 diabetes, characterized by elevated urinary albumin excretion rate (UAER), increase in blood pressure (BP), and decline in renal function leading to end-stage renal disease (ESRD). In addition, these patients have a high risk of cardiovascular disease (CVD) (1), which further

[1]*Steno Diabetes Center, Gentofte, Denmark*
[2]*Department of Medical Endocrinology, Rigshospitalet, Copenhagen, Denmark*
[3]*Faculty of Health Science, University of Copenhagen, Copenhagen, Denmark*
[4]*HEALTH, University of Aarhus, Aarhus, Denmark*
[5]*Department of Cardiology, Rigshospitalet, Copenhagen, Denmark*

Corresponding author: Gudbjörg Andrésdóttir, gbad@steno.dk or gugga77@gmail.com.

Received 29 August 2013 and accepted 27 January 2014.

This article contains Supplementary Data online at http://care.diabetesjournals.org/lookup/suppl/doi:10.2337/dc13-2036/-/DC1.

Improved prognosis of diabetic nephropathy in type 1 diabetes

Gudbjörg Andrésdóttir[1], Majken L. Jensen[1], Bendix Carstensen[1], Hans-Henrik Parving[2,3,4], Peter Hovind[1], Tine W. Hansen[1] and Peter Rossing[1,3,4]

[1]Steno Diabetes Center, Gentofte, Denmark; [2]Department of Medical Endocrinology, Rigshospitalet, Copenhagen, Denmark; [3]Faculty of Health Science, University of Copenhagen, Copenhagen, Denmark and [4]HEALTH, University of Aarhus, Aarhus, Denmark

The natural history of diabetic nephropathy offered an average survival of only 5–7 years. During the past decades, multiple changes in therapy and lifestyle have occurred. The prognosis of diabetic nephropathy after implementing stricter control of blood pressure (including increased use of long-term renin–angiotensin system inhibition), lipids, and glycemia, along with less smoking and other lifestyle and treatment advancements, is inadequately analyzed. To clarify this, we studied 497 patients with type 1 diabetes and

Diabetic nephropathy (DN) is a major complication of diabetes, characterized by elevated urinary albumin excretion rate (UAER), increase in blood pressure (BP), and a relentless decline in renal function. During the natural course of DN, mean survival after the onset of persistent proteinuria was 5–7 years.[1,2] Despite better treatment, including antihypertensive agents and dialysis, DN is still the leading cause of end-stage renal disease (ESRD) in industrialized countries. Furthermore, the increased mortality observed in diabetes

criteria at our hospital. The glomerular filtration rate, measured yearly by ^{51}Cr-EDTA plasma clearance, was a mean of 71 ml/min per 1.73 m^2 at baseline. The mean glomerular filtration rate decline was significantly reduced by 19% (95% confidence interval 5–34) from previously 4.0 to 3.3 ml/min per 1.73 m^2/year. During a median follow-up of 9.1 years, 29% of participants doubled their plasma creatinine or developed end-stage renal disease. Mortality risk was similar to our prior study (hazard ratio 1.05 (0.76–1.43). However, after age adjustment, as both diabetes and nephropathy onset occurred later in life, mortality was reduced by 30%. Risk factors for decline in glomerular filtration rate, death, and other renal end points were generally in agreement with prior studies. Thus, with current treatment of nephropathy in type 1 diabetes, the prognosis and loss of renal function has improved along with better control of modifiable risk factors.

Kidney International advance online publication, 11 June 2014; doi:10.1038/ki.2014.206

KEYWORDS: diabetic nephropathy; diabetes mellitus; end-stage renal disease; glomerular filtration rate; mortality; progression of chronic renal failure

Correspondence: *Gudbjörg Andrésdóttir, Steno Diabetes Center, Niels*

ESRD.

Antihypertensive treatment[5,6] and particularly inhibition of the renin–angiotensin system (RAS) has become a cornerstone in the treatment of patients with diabetes and albuminuria. This is based on randomized studies showing RAS inhibition to delay renal end points and death.[7–10] However, the prognosis with longer-term clinical use is inadequately analyzed.

For patients with type 1 diabetes (T1DM) and DN, RAS inhibition became a fully implemented part of standard therapy after reinforcement of local guidelines in 2000. These guidelines also stressed the importance of control of BP, lipids, glycemia, and smoking.

This current study evaluates the loss of renal function and prognosis of patients with T1DM and DN from 2000 to 2010 by assessing change in ^{51}Cr-EDTA plasma clearance (glomerular filtration rate (GFR)), progression to ESRD, and mortality rate. A multistate model is used to account for CVD, the competing risks of ESRD, and death. Results are compared with patients identified and followed up with the same criteria and methods, at the same hospital,[11,12] before these guidelines.

RESULTS

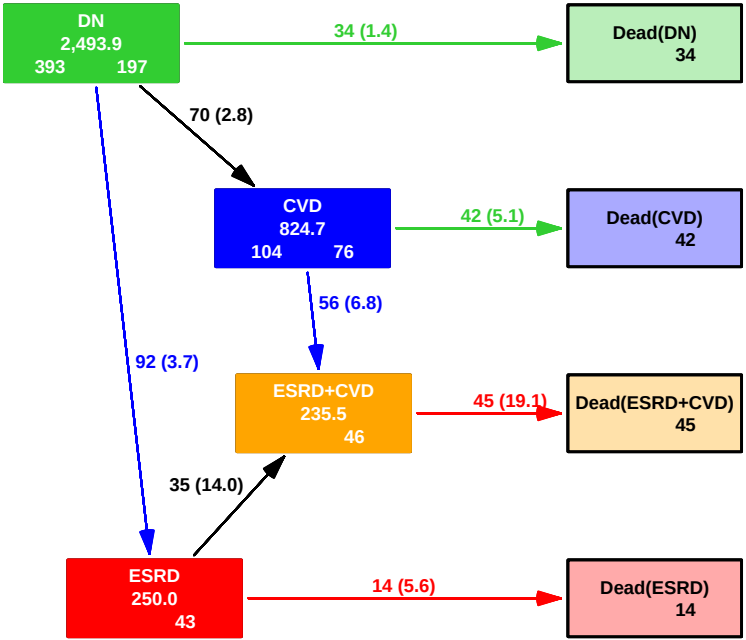
Study participants

We identified 497 eligible patients with T1DM and DN. The mean (s.d.) baseline GFR was 71 (32) ml/min per 1.73 m^2, hemoglobin A_{1c} (HbA_{1c}) was 9.1 (1.4) %, and median (interquartile range) albumin excretion rate (UAER) was 483

- Well defined patient population:
  - events well defined
  - when do DN, CVD, ESRD occur
- Well defined research question:
  - effect of clinical variables on rates
  - on long-term outcome

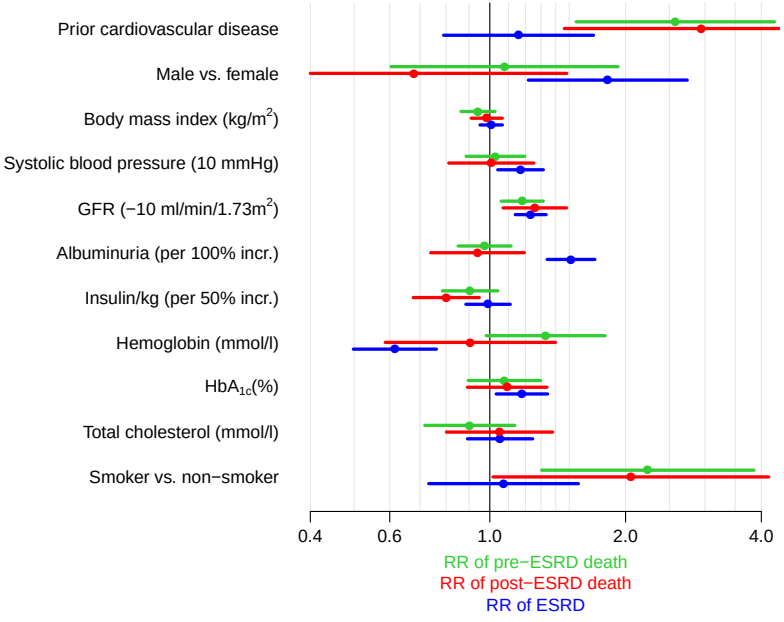
Logistics of the research work

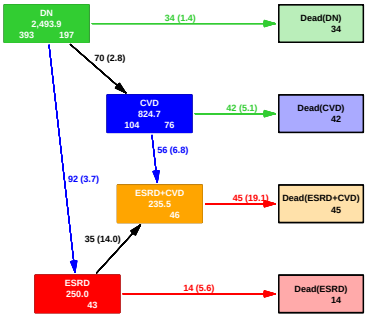
- Only possible through close collaboration between
- Clinical researchers: what is relevant, what is available, what is reliable
- Statistician: what is possible, what is relevant, what data is needed
- The project took many hours of joint discussion to get the boxes right, and the hypotheses properly hammered out.

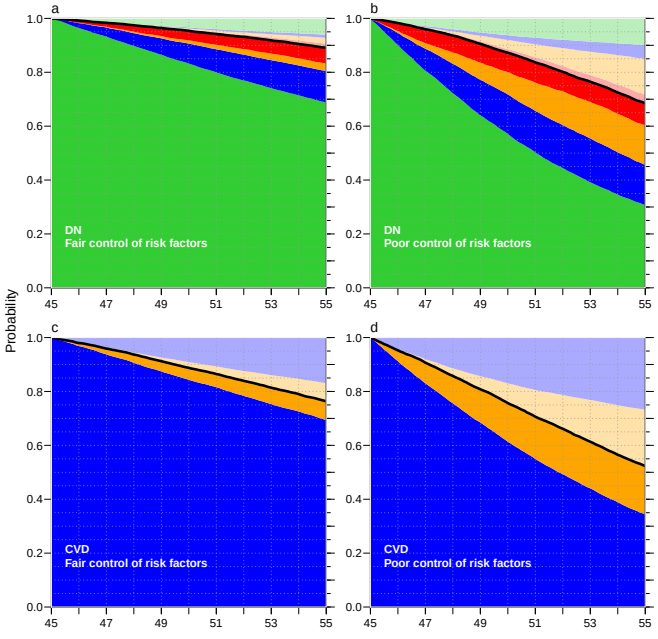

DN
2,493.9
393 197
34 (1.4)
Dead(DN)
34
70 (2.8)
CVD
824.7
104 76
42 (5.1)
Dead(CVD)
42
56 (6.8)
92 (3.7)
ESRD+CVD
235.5
46
45 (19.1)
Dead(ESRD+CVD)
45
35 (14.0)
ESRD
250.0
43
14 (5.6)
Dead(ESRD)
14

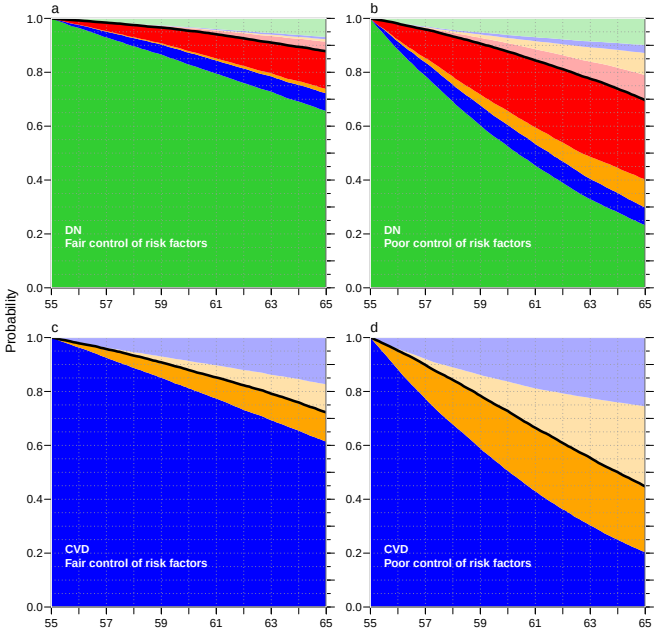
Data requirements

- Clinical status for patients at baseline:
  BMI, HbA1c, bloodpressure, GFR, insulindose, lipids, ...
- Dates of:
  - Birth
  - DM
  - CKD
  - CVD
  - ESRD
  - Death


Prior cardiovascular disease
Male vs. female
Body mass index (kg/m^2)
Systolic blood pressure (10 mmHg)
GFR (−10 ml/min/1.73m^2)
Albuminuria (per 100% incr.)
Insulin/kg (per 50% incr.)
Hemoglobin (mmol/l)
HbA$_{1c}$(%)
Total cholesterol (mmol/l)
Smoker vs. non−smoker
0.4
0.6
1.0
2.0
4.0
RR of pre−ESRD death
RR of post−ESRD death
RR of ESRD


DN
2,493.9
393 197
34 (1.4)
Dead(DN)
34
70 (2.8)
CVD
824.7
104 76
42 (5.1)
Dead(CVD)
42
56 (6.8)
92 (3.7)
ESRD+CVD
235.5
46
45 (19.1)
Dead(ESRD+CVD)
45
35 (14.0)
ESRD
250.0
43
14 (5.6)
Dead(ESRD)
14


a
b
c
d
DN
Fair control of risk factors
DN
Poor control of risk factors
CVD
Fair control of risk factors
CVD
Poor control of risk factors
Probability
Age at follow−up
0.0
0.2
0.4
0.6
0.8
1.0
45
47
49
51
53
55


a
b
c
d
DN
Fair control of risk factors
DN
Poor control of risk factors
CVD
Fair control of risk factors
CVD
Poor control of risk factors
Probability
Age at follow−up
0.0
0.2
0.4
0.6
0.8
1.0
55
57
59
61
63
65

Use of clinical database

- Are patients reasonably representative to show relevant results?
- Is the data sufficiently reliable in terms of:
  - completeness
  - accuracy

  of measurements and dates

Using clinical records in research

Introduction to
Clinical Epidemiology
Beijing, 14 March 2015
NN Tomorrow Forum
http://BendixCarstensen.com/Epi/NNTF

summ

steno
diabetes center

Use of clinical records: Statistical methods

Continuous outcomes:

- HbA_{1c}
- lipids
- GFR
- ...

require repeated measures models (aka. mixed models, random effects models)

Use of clinical records: Statistical methods

Event type outcome:

- death
- ESRD
- retinopathy

require survival-type analysis:

- death: survival analysis
- all other: competing risks or multistate models

Clinical records: data

- Describe data:
  - Who
  - What
  - When
  - Why
- Describe hypothesis or research question:
  - **What** quantity
  - depends on **what**
  - and in particular **how much**
- Always specify research question in **quantitative** terms,
- never "is there an effect of...".
- There is always one, but it may be so small that we do not bother.

Clinical records: Summary

- Hypothesis
- Data access
- Data limitations
- Never throw data away:
  - never dichotomize variables (BMI>25)
  - never throw away patients from analysis
- Statististical methods to handle data as they are
- . . . you might want to consult a statistician