

EPJ i Epidemiologisk forskning -kan den bruges?

Mette Nørgaard, overlæge, Ph.d.
Klinisk Epidemiologisk afdeling



Ja - det mener vi

Hvad laver kliniske epidemiologer?

Fastlæggelse af:

- Hyppighed
- Risikofaktorer
- Behandlingseffekt
- Prognose

for klinisk definerede patientgrupper
(fx. cancer, diabetes, lungebetændelse, ...)

Epidemiologiske begreber

Risiko faktorer

Prævalens og incidens

Prognostiske faktorer

Outcome

Sygdom
fx
lungekræft

Death
Disease
Discomfort
Disability
Dis-satisfaction

Rygning
Alder
Mand
Asbest
Radonstråling
Arv?
Social status?

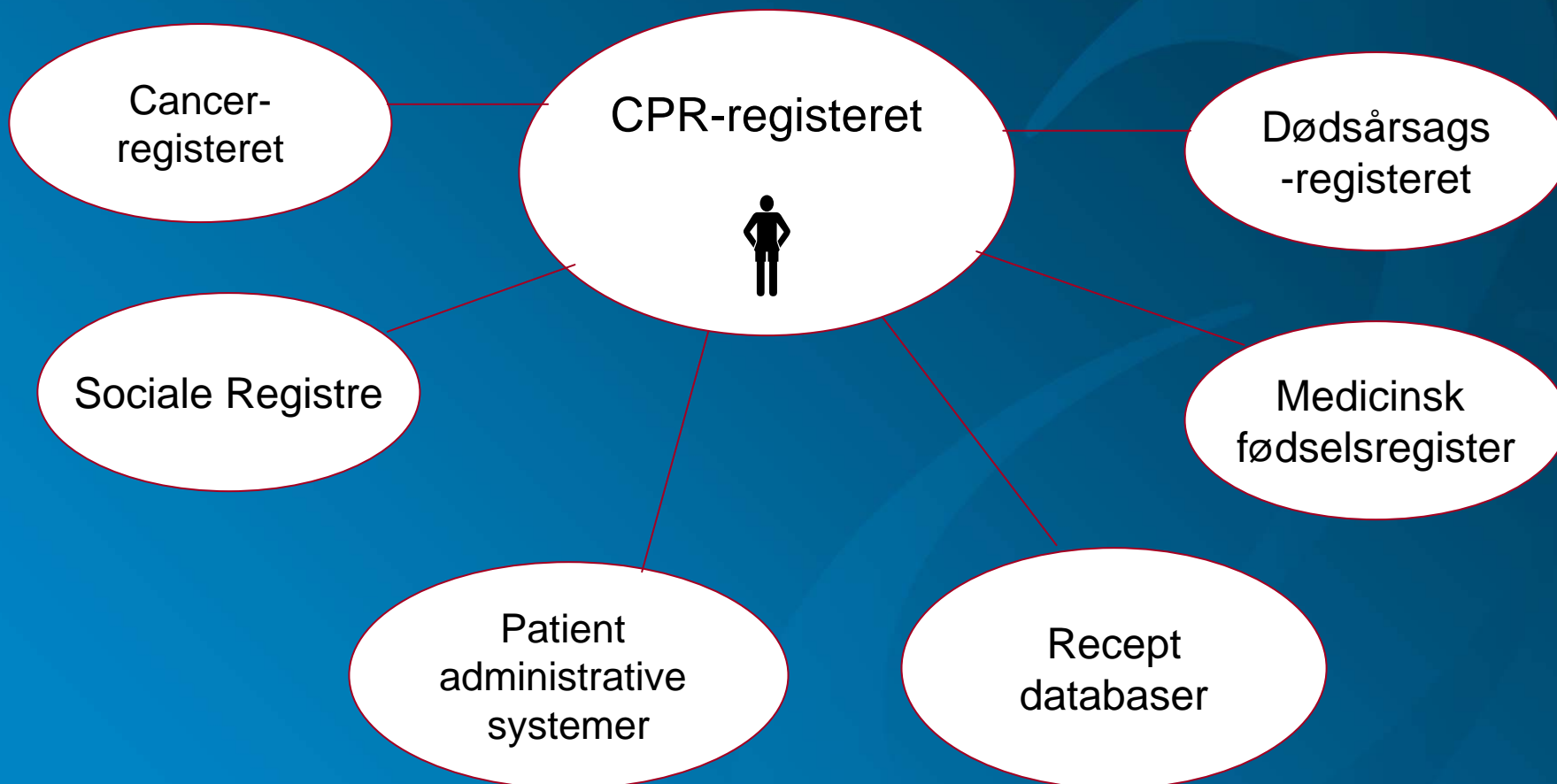
Kræfttype, størrelse, spredning
Alder
Mand
Ko-morbiditet
Diagnosetidspunkt
Behandlingsregimen
Klinikerens "performance"
Patientressourcer, compliance

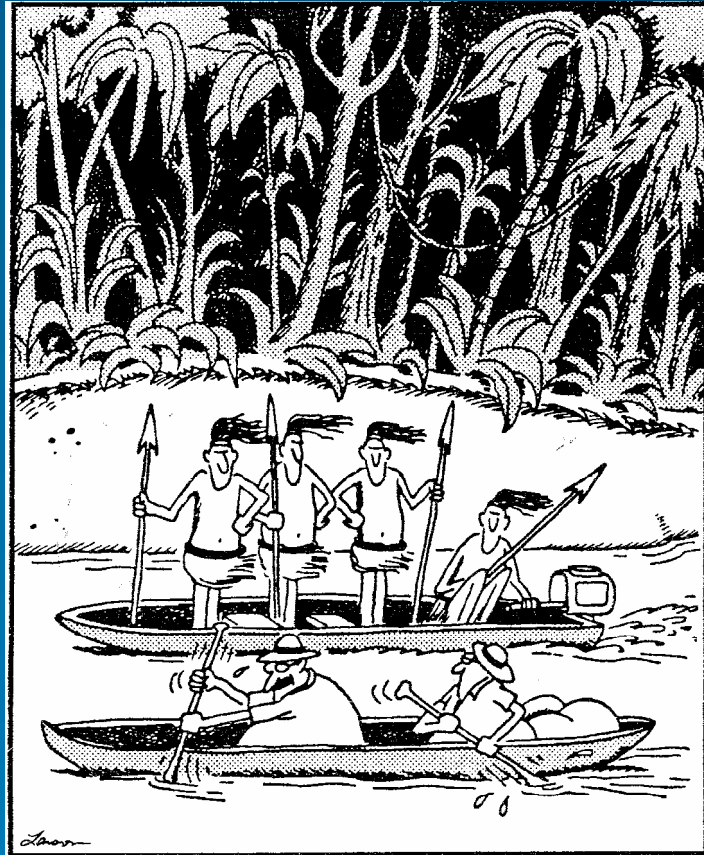
	Disease	Non-disease
Exposed	A	B
Non-exposed	C	D

Kohorte

	Cases	Controls
Exposure, yes	A	B
Exposure, no	C	D

Case kontrol





“Are they gaining, Huxley?”

When an Entire Country Is a Cohort

Denmark has gathered more data on its citizens than any other country. Now scientists are pushing to make this vast array of statistics even more useful

For years, any woman who got an abortion had to accept more than the loss of her fetus: For some unknown reason, she also faced an elevated risk for breast cancer. At least that was what several small case-control studies had suggested before Mads Melbye, an epidemiologist at the Statens Serum Institute in Copenhagen, undertook the largest effort ever to explore the link. He and his colleagues obtained records on 400,000 women in Denmark's national Abortion Register, then checked how many of the same women were listed in the Danish Cancer Register. Their foray into the two databases led to a surprising result: As they reported in *The New England Journal of Medicine* in 1997, there appears to be no connection between abortion and breast cancer.

Their success underscores the value of a trove of data the Danish government has accumulated on its citizenry, which today totals about 5 million people. Other Scandinavian countries have created powerful database systems, but Denmark has earned a preeminent reputation for possessing the most complete and interwoven collection of statistics touching on almost every aspect of life. The Danish government has compiled nearly 200 databases, some begun in the 1930s, on everything from medical records to socioeconomic data on jobs and salaries. What makes the databases a plum research tool is the fact that they can all be linked by a 10-

digit personal identification number, called the CPR, that follows each Dane from cradle to grave. According to Melbye, "our registers allow for instant, large cohort studies that are impossible in most countries."



Beauty in numbers. These Danish twins starred in a variety show at the turn of the 20th century; now it's their medical records, part of a database, that are in demand.

But Melbye and other scientists think they can extract even more from this data gold mine. They argue that not enough money is being spent on maintaining and expanding existing databases, and they say that red tape is hampering studies that require correlation of health and demographic data. The problem is that, while they have unfettered access to more than 80 medical databases maintained

by the Danish Board of Health and public hospitals, their use of 120 demographic databases overseen by the agency Statistics Denmark is tightly restricted. Statistics Denmark won't allow researchers to remove from its premises data coded by CPR, and the procedures for accessing information at all are unwieldy and expensive.

Statistics Denmark officials are reluctant to release data tied to CPRs, citing privacy concerns. "The public should have confidence that information identifying them as individuals does not reside outside of this institution," says the agency's Otto Andersen.

Last month, Danish research minister Birte Weiss formed a committee to break the impasse. Denmark's databases are "a resource which can be used more optimally," she told *Science*. "This should be a scientific flagship."

Working the health databases can yield powerful results. For years the U.S. National Institutes of Health has supported a study following twins, hoping to tease out the relative contributions of genes and lifestyle to aging. Led by University of Southern Denmark gerontologist Kaare Christensen, the project has tapped the Danish Twin Register, which includes 110,000 pairs of twins born since 1870. After following more than 2000 pairs of twins aged 70 or older, Christensen's group has so far tied to genes about a quarter of the variation in human longevity. "The project is made possible by the unmatched age and completeness of the Danish Twin Register," he says.

The health databases have proven invaluable for probing contradictions raised by smaller studies and following disease pro-



Science
2000; 287
2398-9.

CREDIT: PHOTO COURTESY OF THE DANISH TWIN REGISTER, UNIVERSITY OF SOUTHERN DENMARK

Registerdata:

- longitudinelle
- muligt at studere sjældne hændelser
- repræsenterer kliniske dagligdag
- repræsenterer alle patienter uanset alder og komorbiditet
- relativt billige studier

Generelt

1. Indsamlings metoderne er forudbestemt og kan ikke kontrolleres af forskeren og nogle gange er data umulige at validere.
2. Misklassifikation eksisterer i alle data
3. Dårlig data kvalitet udgør en permanent udfordring for register-baseret forskning
4. Den relativt store data mængde kan lede til fisketure og misvisende post-hoc analyser

The data quality of haematological malignancy ICD-10 diagnoses in a population-based Hospital Discharge Registry

M Nørgaard^{1,2}, M V Skriver¹, H Gregersen³, G Pedersen^{3,4},
H C Schönheyder⁵ and H T Sørensen¹

The objectives of this study were to estimate the data quality of haematological malignancy diagnoses in a

in the Hospital Discharge Registry compared with being registered in the Danish Cancer Registry was 0.98 (95% CI

Eur J Cancer Prev.
2005;14:201-206.

Studie I

- at undersøge datakvaliteten af en malign hæmatologisk udskrivningsdiagnose i PAS
- samt vurdere hvilken betydning en eventuel misklassifikation har på den estimerede overlevelse

Studie I

Alle over 15 år med en nydiagnosticeret malign hæmatologisk sygdom

Identificeret i **Det Patient Administrative System (PAS)** i Nordjyllands Amt & i **Cancerregisteret (DCR)**

Vitalstatus og evt. dødsdato via CPR registeret

Studie I

Studieperiode: 1994-1999

1159 patienter var registeret med deres første maligne hæmatologiske sygdom

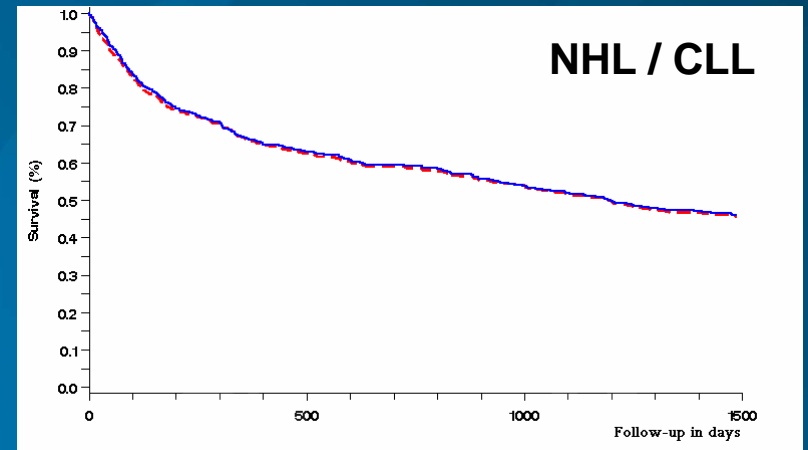
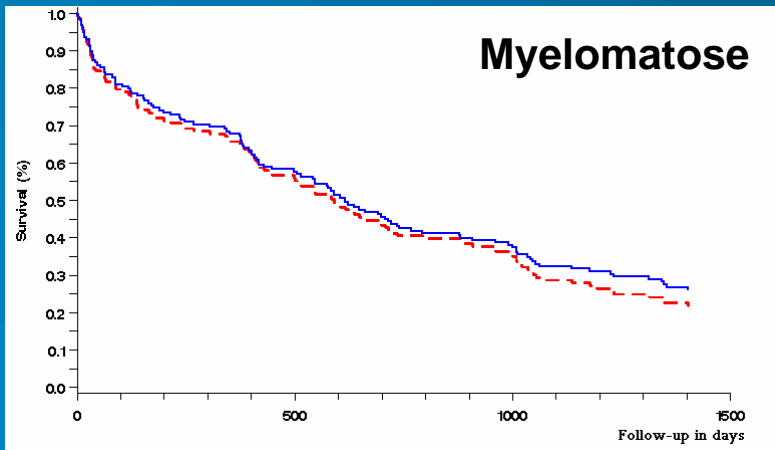
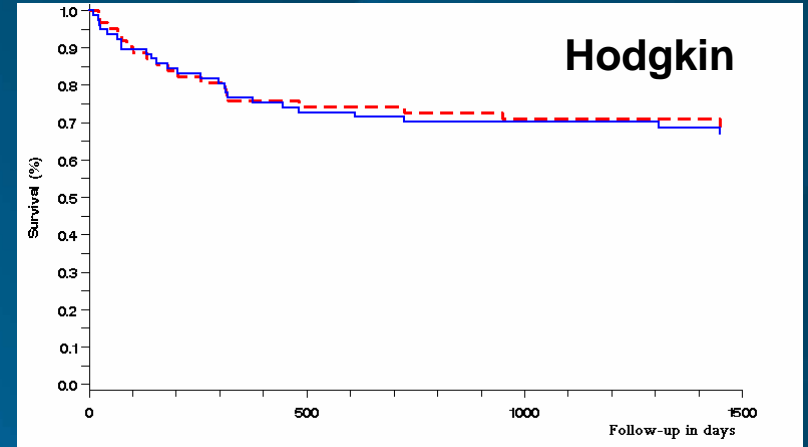
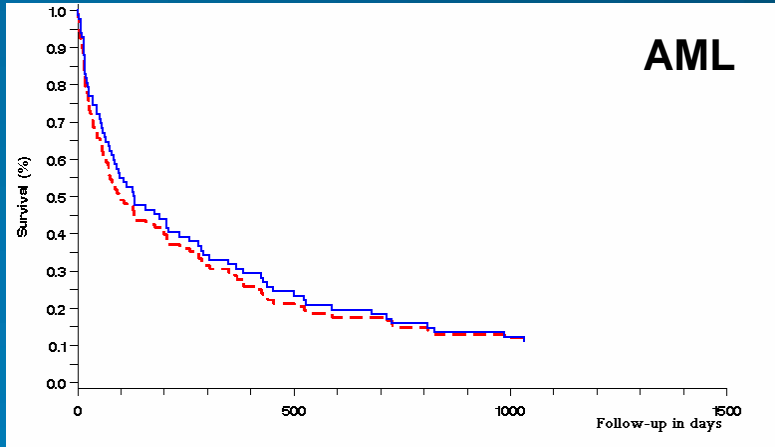
heraf

- 908 (78 %) i begge registre
- 167 (14 %) kun PAS
- 84 (7 %) kun i Cancerregisteret

Studie I

	Antal n	Komplethedegrad %	PPV %
Alle	1,159	91.5	84.5
Akut myeloid leukæmi	117	89.0	67.6
Hodgkin´s lymfom	84	88.7	71.4
Non-Hodgkin lymfom eller kronisk lymfatisk leukæmi	683	88.2	85.3
Myelomatose	171	90.9	82.3

Studie I



--- DCR — PAS

Misklassifikationen gav ingen væsentlig skævhed i overlevelsesberegningerne

Studie I

bjh research paper

Short-term mortality of bacteraemia in elderly patients with haematological malignancies

Mette Nørgaard,¹ Heidi Larsson,¹ Gitte Pedersen,^{2,3} Henrik C. Schönheyder,⁴ Kenneth J. Rothman^{5,6} and Henrik T. Sørensen^{1,5}

Summary

Bacterial infections are important complications in patients with haematological malignancies. We compared the outcome of bacteraemia among elderly and younger patients with haematological malignancies, and

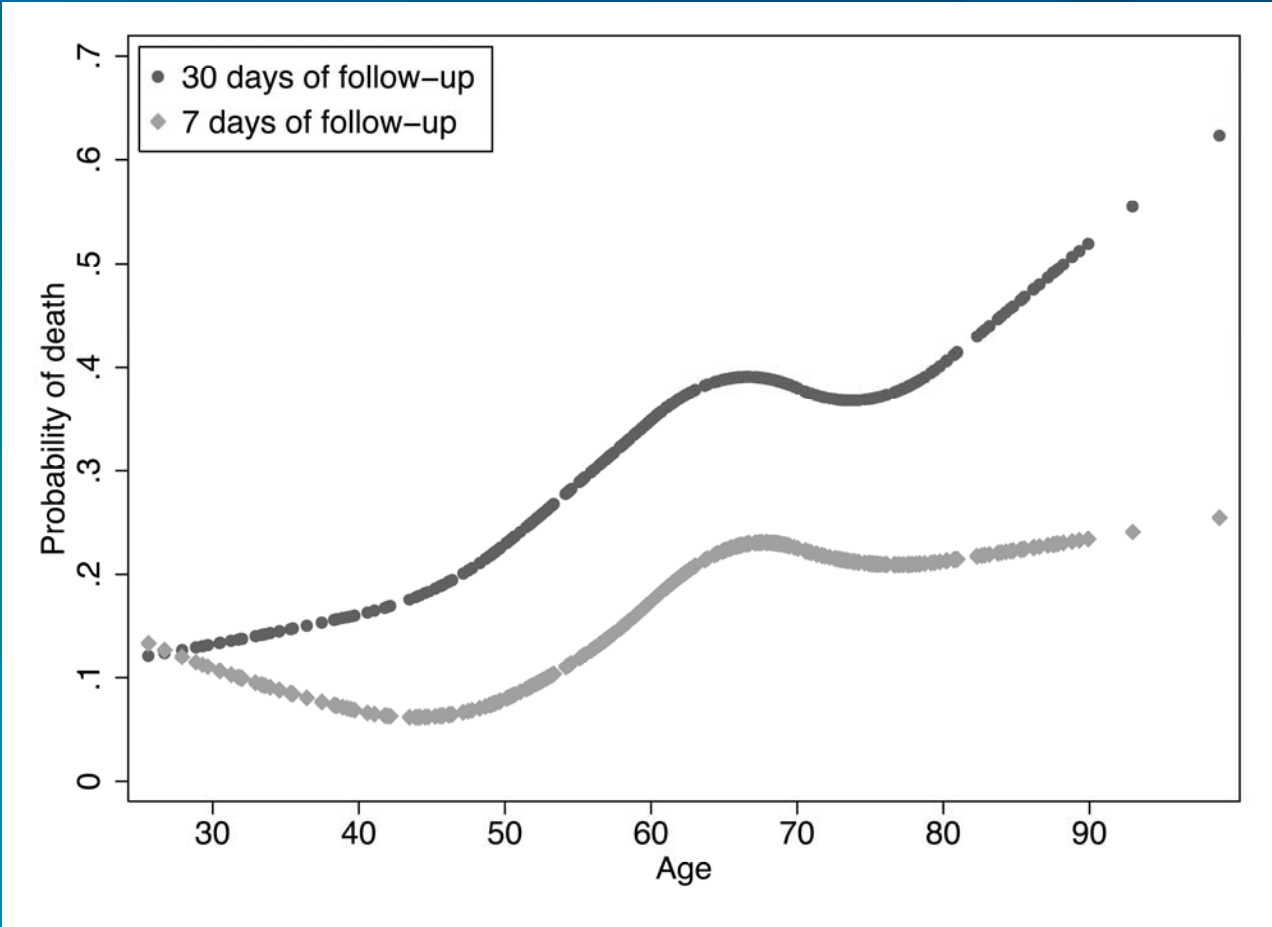
Br J Haematol. 2006;132:25-31.

Studie II

- at belyse sammenhængen mellem alder og overlevelse af bakteriæmi blandt patienter med malign hæmatologisk sygdom

Studie II

Resultater



Stude II

Kan EPJ bidrage med at finde ud af hvorfor:

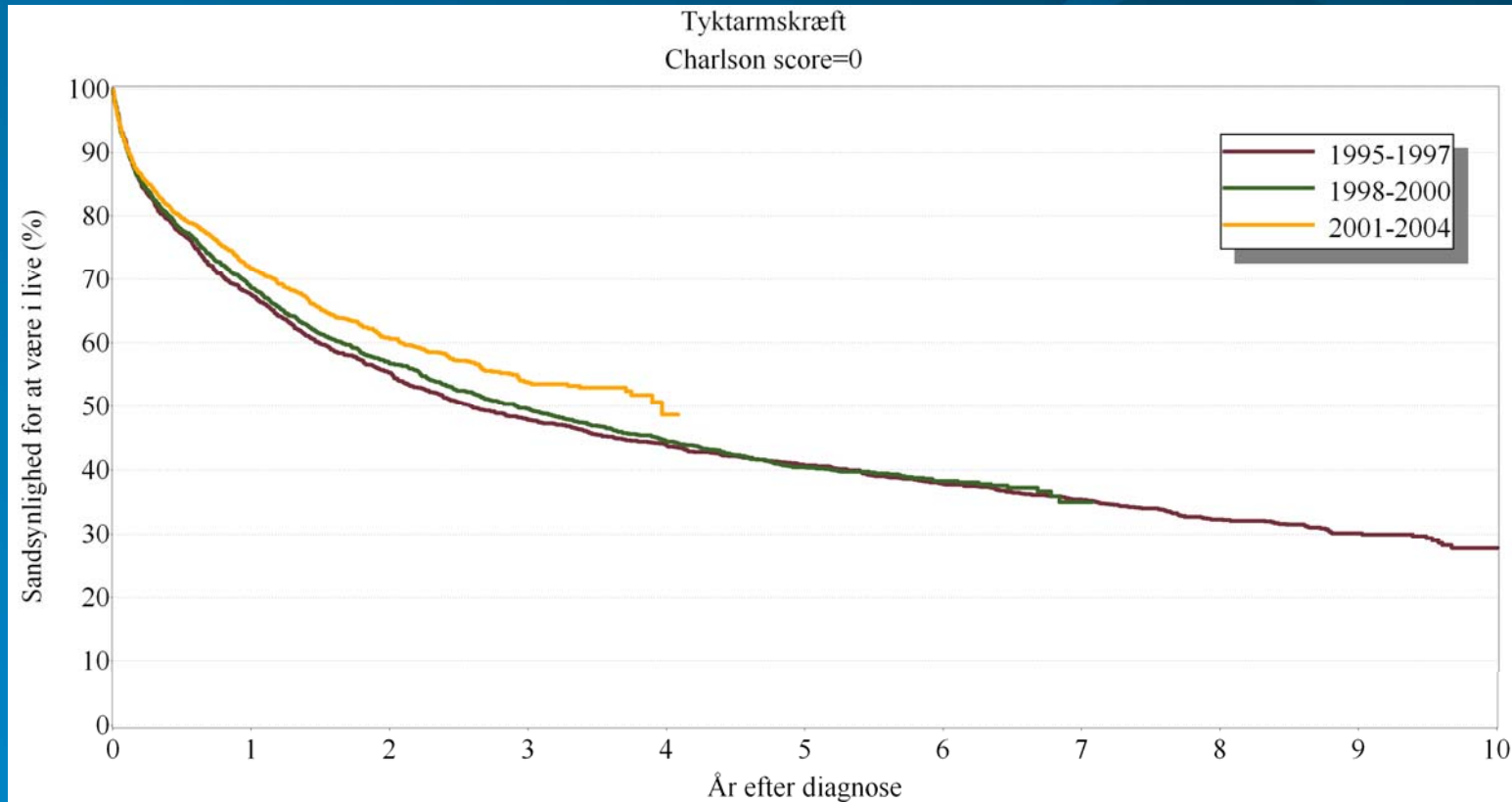
- Får ældre oftere suboptimal behandling?
- Spiller rygning og alkohol en rolle?
- er det udelukkende fordi ældre oftere har andre alvorlige sygdomme?
- er ældre oftere underernærede?

Studie II

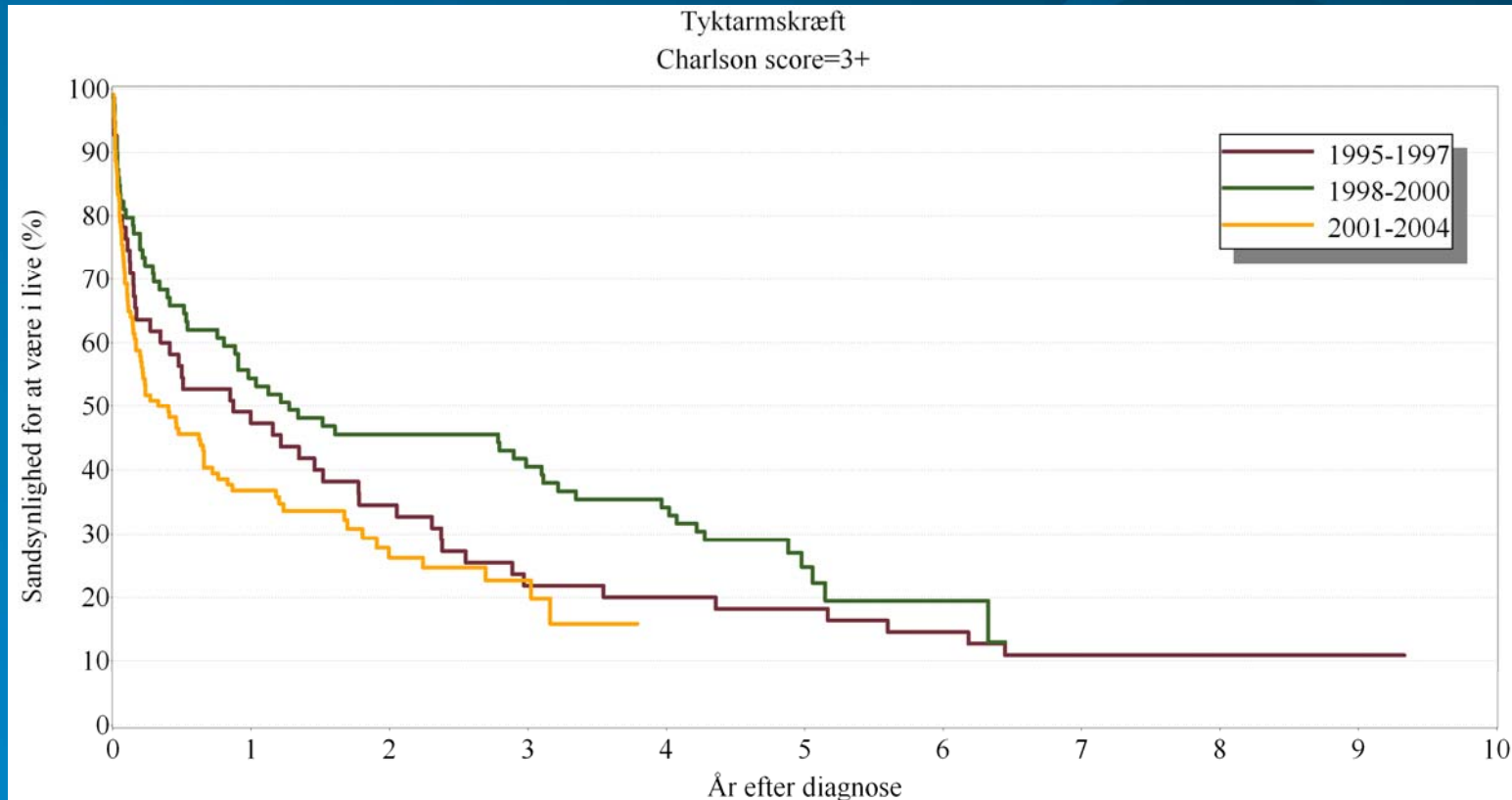
- Er der en sammenhæng mellem graden af andre alvorlige sygdomme (komorbiditet) og overlevelsen efter indlæggelse for kræft?

Studie III

Har overlevelsen i de forskellige komorbiditetsgrupper ændret sig over tid?



Har overlevelsen i de forskellige komorbiditetsgrupper ændret sig?



Hvad bemærker I ved de følgende tre billeder?







Hvordan vil I få en computer
(eller en statistiker) til at genkende
disse ligheder?



missing data:

Vi vil ofte prøve at isolere effekten af en enkelt eksponering:

Eksempel: rygning og risiko for lungekræft

risiko for lungekræft = Basis risiko + a_1 *rygning +
 a_2 *Alder + a_3 *køn + a_4 *erhvervs eksposition

Hvis der mangler oplysninger om en af effekterne kan patienten ikke indgå i analysen!

Can We Teach Compliance?



Tak for opmærksomheden!