Event history analysis

With examples from health surveys

Håkon K. Gjessing

Professor/Principal Investigator

Centre for Fertility and Health
An NRC Centre of excellence, at the Norwegian Institute of Public Health, Oslo

Department of Global Public Health and Primary Care, University of Bergen

NOFE meeting Bergen

Thursday 4 Nov, 2021

CENTRE FOR FERTILITY AND HEALTH

GOOD, OLD DAYS...

Norsk Epidemiologi

Norwegian Journal of Epidemiology

Årgang 15, supplement 1, desember 2005

Utgitt av Norsk forening for epidemiologi

- What did “Norsk Epidemiologi” write about in 2005?

CENTRE FOR FERTILITY AND HEALTH

GOOD, OLD DAYS...

Norsk Epidemiologi

Norwegian Journal of Epidemiology

Årgang 15, supplement 1, desember 2005

Utgitt av Norsk forening for epidemiologi

Redaktør:
Trond Peder Flaten
Institutt for kjemi,
Norges teknisk-naturvitenskapelige universitet, 7491 Trondheim

DEN TRETTENDE NORSKE EPIDEMIOLOGIKONFERANSEN

- Well, aren't ALL NOFE-conferences "trettende"?

CENTRE FOR FERTILITY AND HEALTH

Event History Analysis: What and Why?

New England Journal of Medicine

Editorial, Jan. 6, 2000, p. 42-49

The eleven most important developments in medicine over the past millennium

- Elucidation of human anatomy and physiology
- Discovery of cells and their substructures
- Elucidation of the chemistry of life
- *Application of statistics to medicine*
- Development of anesthesia
- Discovery of the relation of microbes to disease
- Elucidation of inheritance and genetics
- Knowledge of the immune system
- Development of body imaging
- Discovery of antimicrobial agents
- Development of molecular pharmacotherapy

CENTRE FOR FERTILITY AND HEALTH

Event History Analysis: What and Why?

Application of statistics to medicine

"Sir David Cox's 1972 paper on proportional-hazards regression ignited the fields of survival analysis and semiparametric inference (using partial specification of the probability distribution of the outcomes under investigation). Rapid improvements in computer support were essential to the growing role of empirical investigation and statistical inference."

CENTRE FOR FERTILITY AND HEALTH

Time-to-event methods

Common names:

- Survival analysis (Overlevelsesanalyse)
- Event history analysis (Forløpsanalyse)

Some important aspects:

- Better understanding of time scales
- Studies of life histories, recurrent events
- Truncation and censoring at beginning and end of registries
- Time-dependent exposures
- Competing risks, multi-state
- Case-cohort, nested case-control and other study design
- etc.

CENTRE FOR FERTILITY AND HEALTH

Event history analysis: Choice of time scale

Three most common time scales:

1. **Time from inclusion to event (study time)**

   Example: Time from cancer diagnosis to death

   Zero: Date of inclusion (individual)

2. **Calendar time**

   Example: Time from a fixed date (e.g. 01 Jan 2020) to infection with Covid-19

   Zero: Start date (common)

3. **Age**

   Example: Age at death

   Zero: Date of birth (individual)

CENTRE FOR FERTILITY AND HEALTH

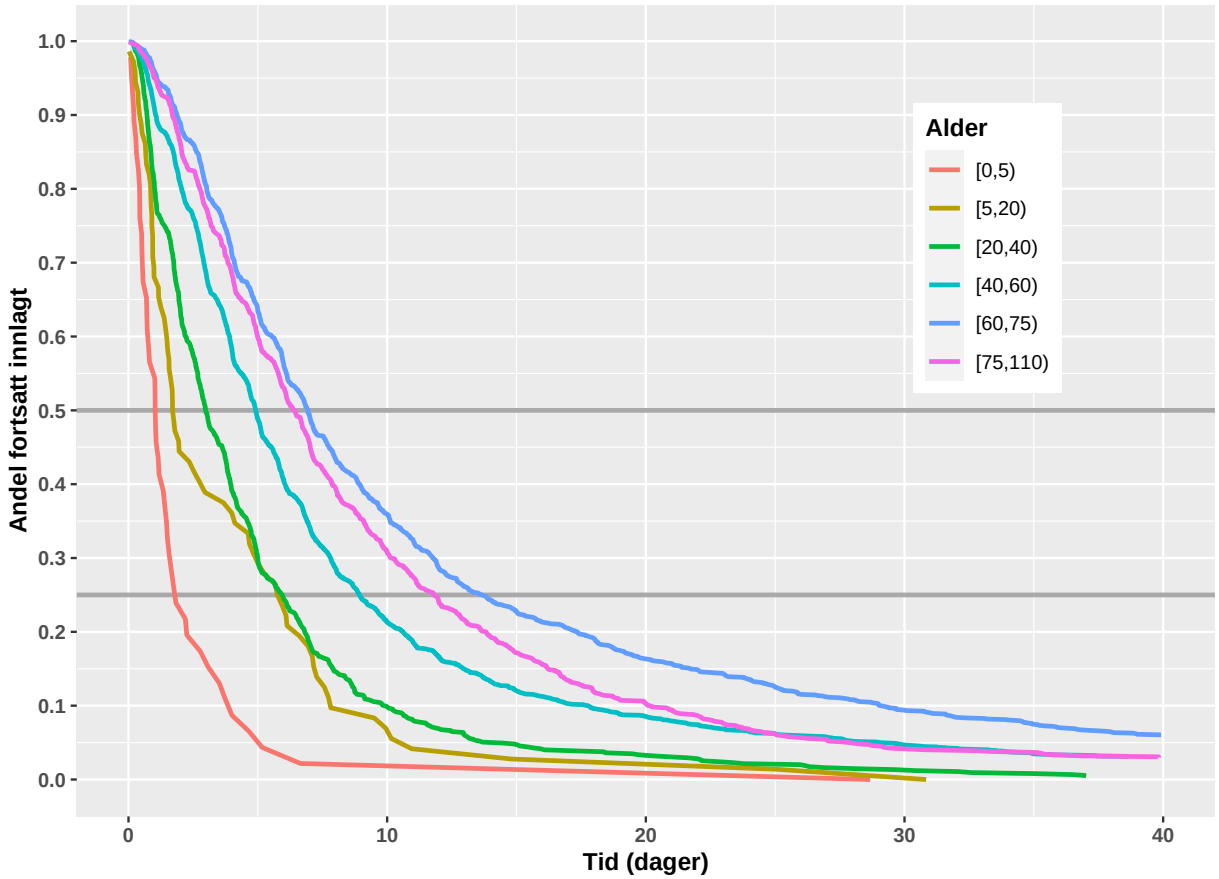
Examples of event history analyses

1. Life insurance, life tables
2. Maternal age at birth of first child
3. Time from first birth to second birth
4. Multi-state: Movement between states employed, retired, sick leave, disability, dead
5. Clinical trials, cancer survival
6. Mortality, cause of death. Competing risk. Loss of life years.
7. Pregnancy: Gestational age as time scale.
   Preterm birth: Birth before week 37. Rest censored at day 259.
8. Time from start of pandemic until an individual becomes infected
   - ... and re-infected
   - May use vaccination status as time-dependent covariate
9. Time from insertion of a new hip prosthesis until failure
10. Etc., etc....

CENTRE FOR FERTILITY AND HEALTH

Example from Covid-19 modeling

Time from hospital admission until discharge:

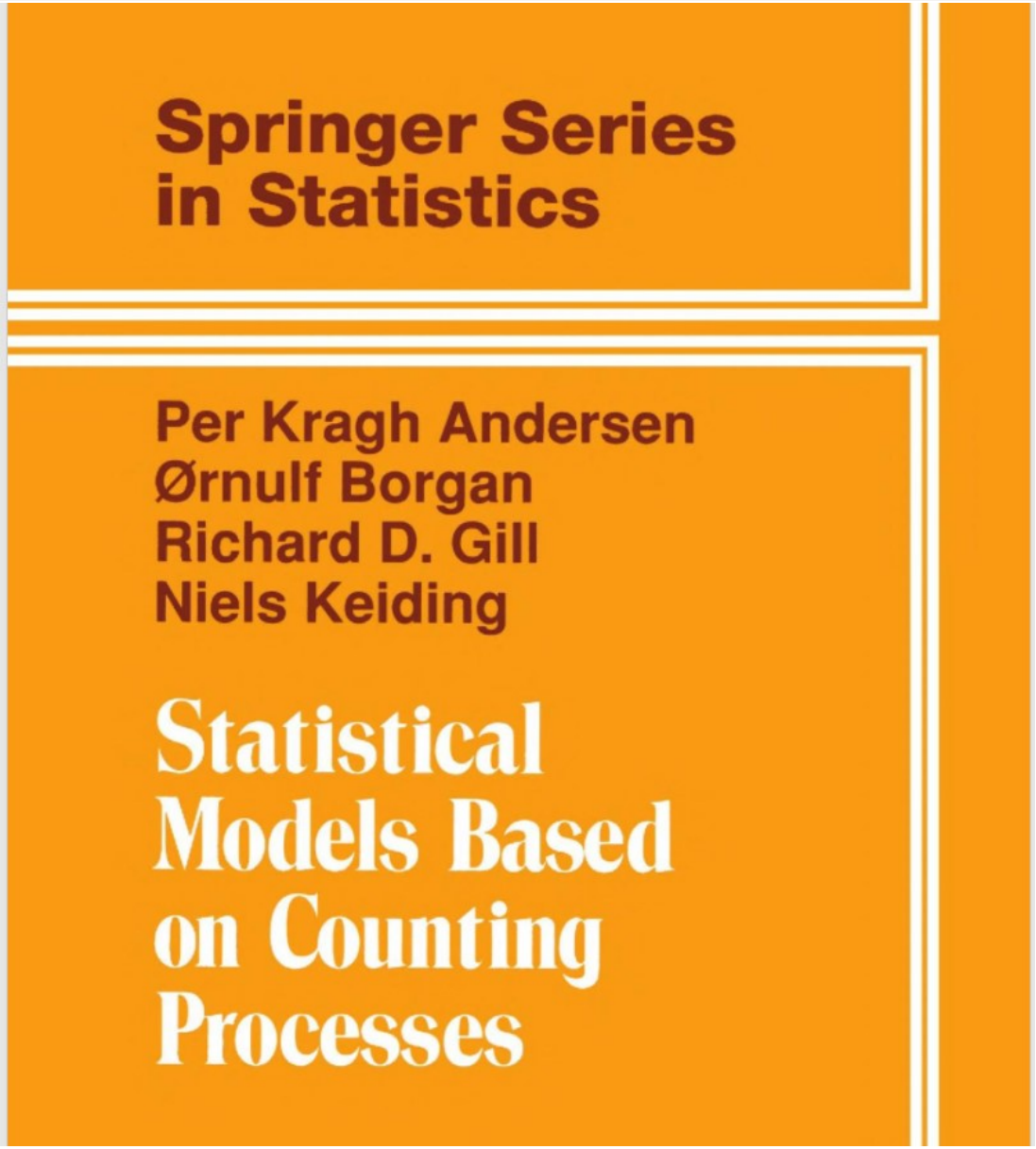


- "Survival" curve: Percentage still in hospital after x days
- Close to current day, hospital stays are "censored" in length

CENTRE FOR FERTILITY AND HEALTH

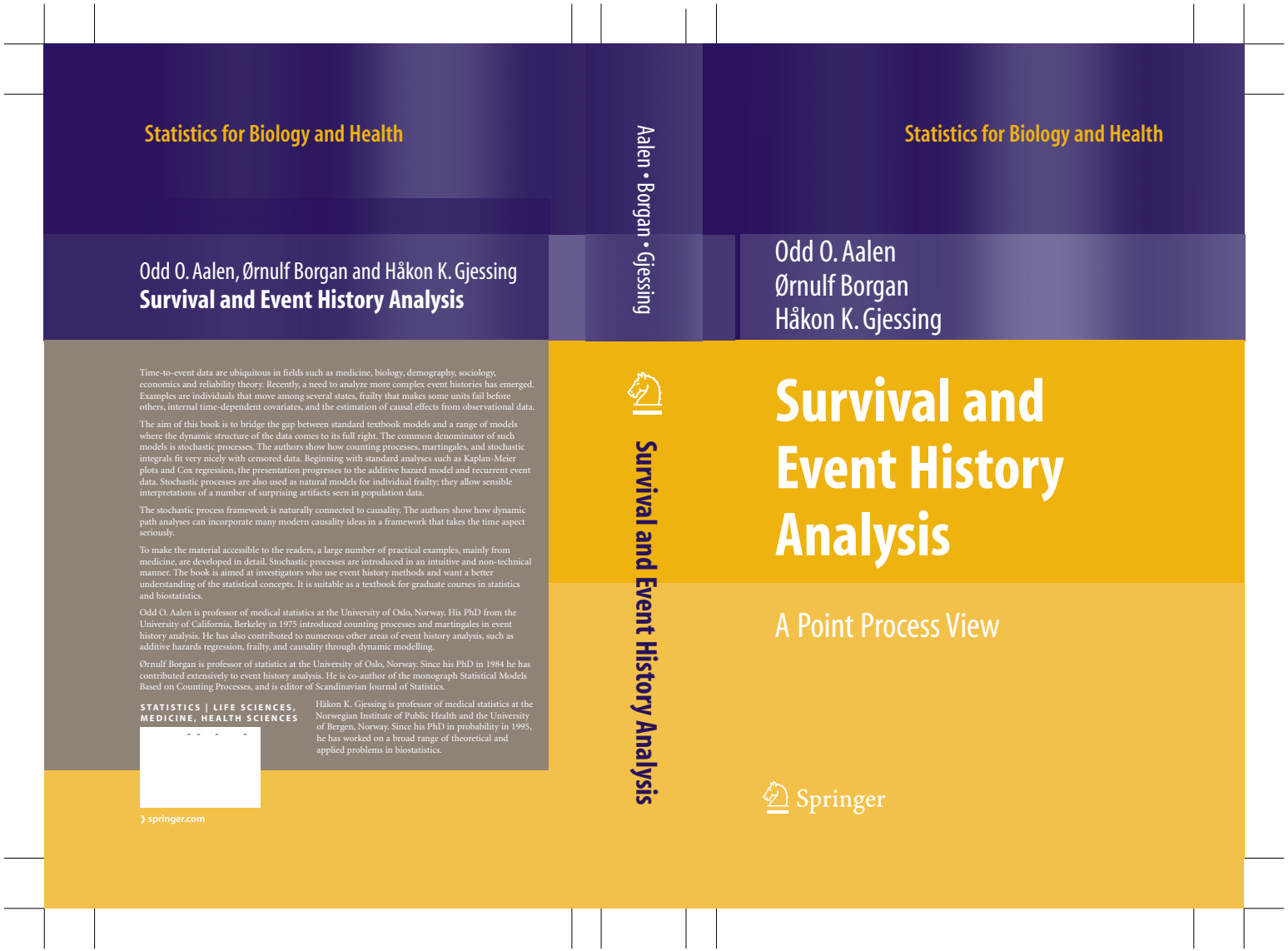
"The Bible"

Norwegian & Nordic tradition in survival analysis



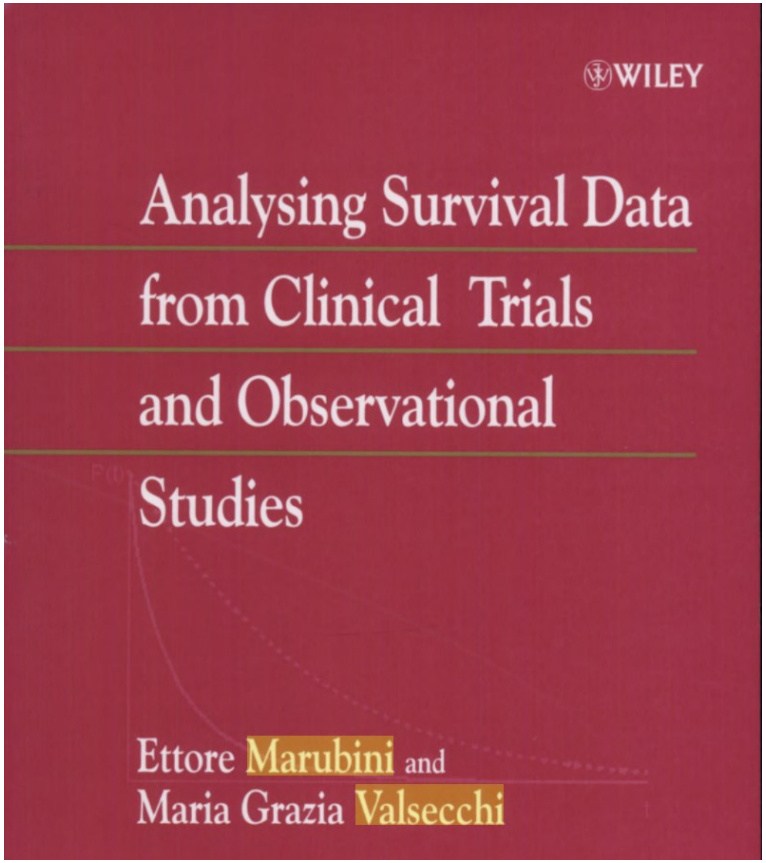
CENTRE FOR
FERTILITY AND HEALTH

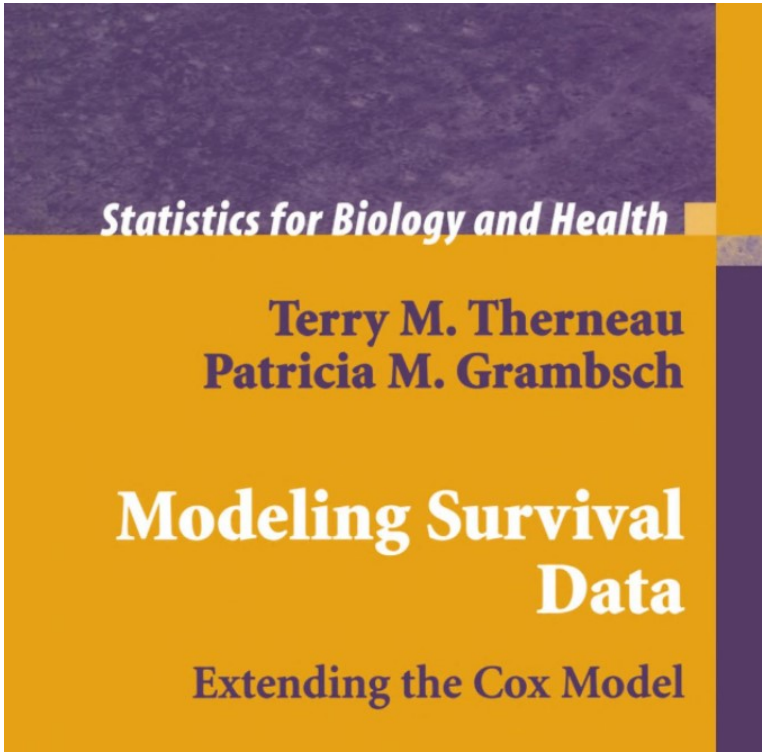
"The New Testament"(?)



- Much emphasis on event histories as *stochastic processes*

But plenty of easier and very useful books out there...

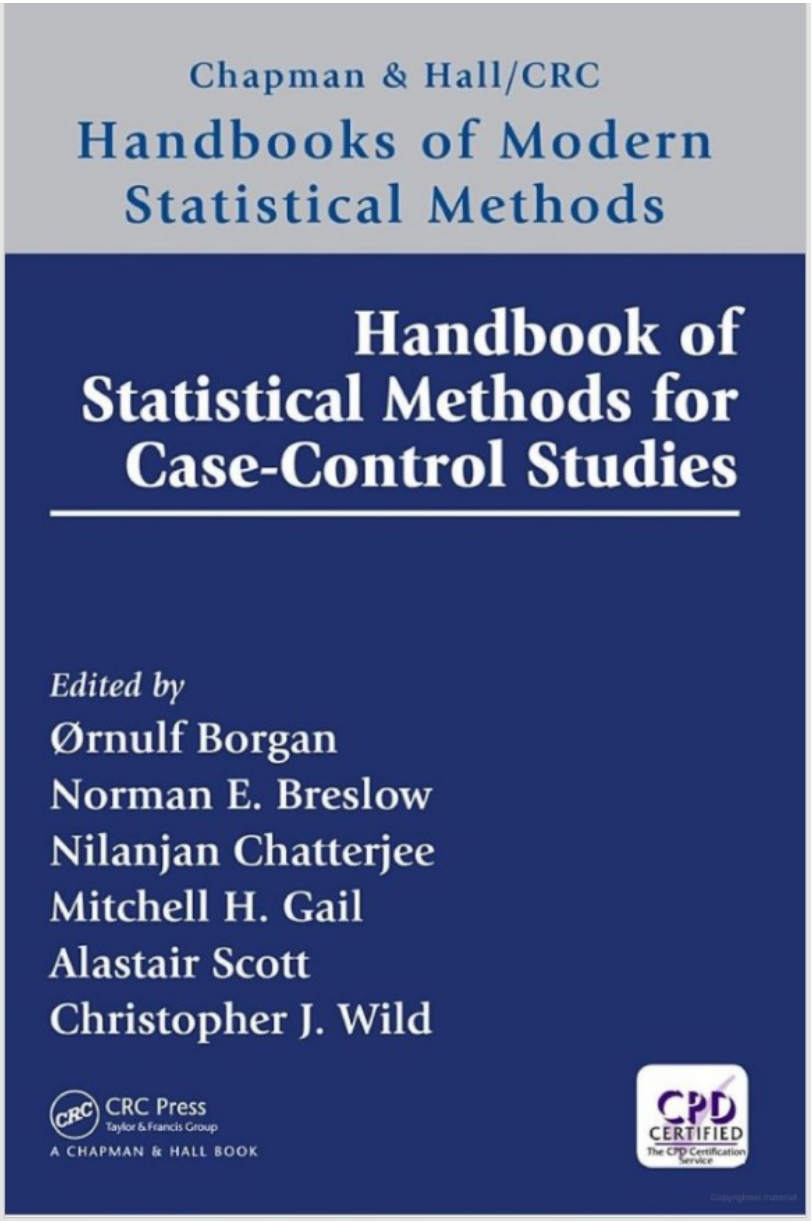


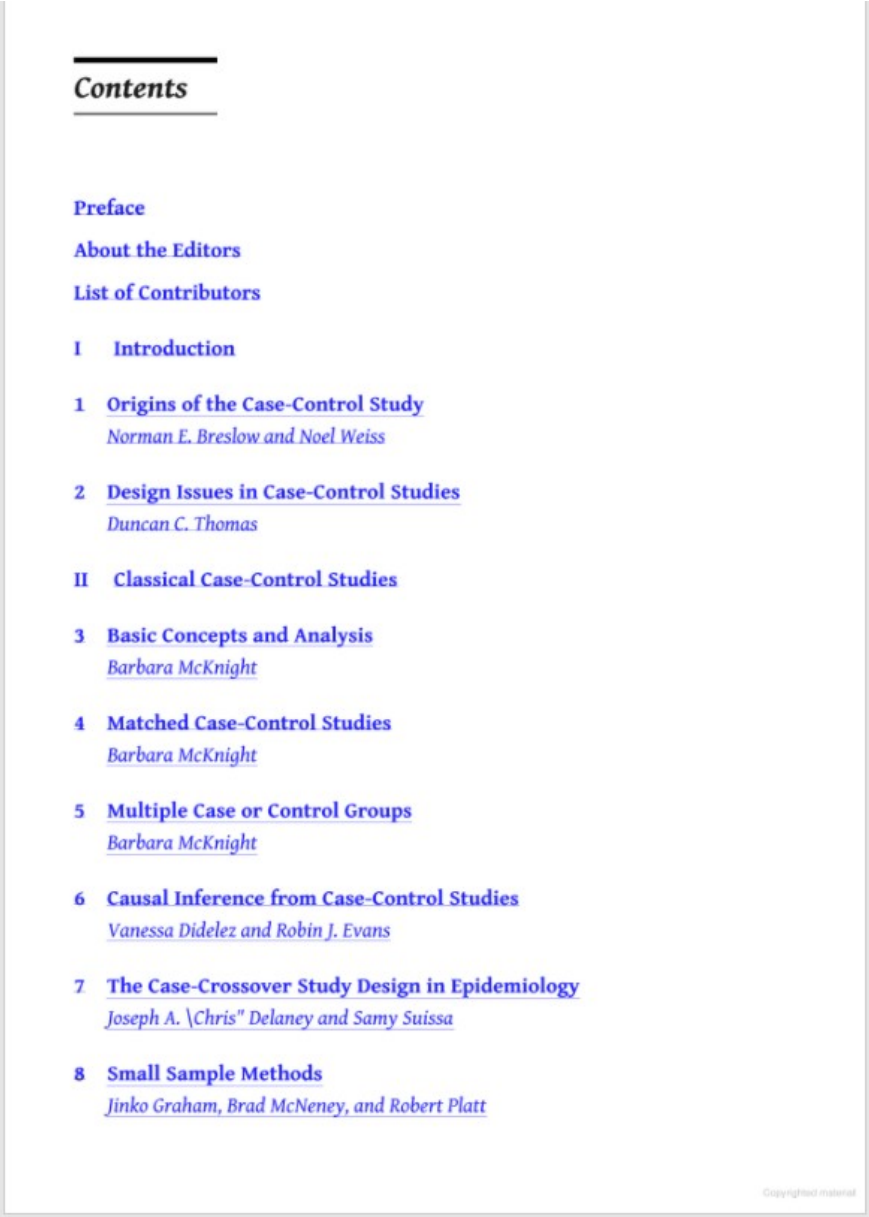


- Two nice examples
- But also many newer ones
- As well as software manuals, e.g. the STATA manuals
- And lots of instructive “tutorial” papers, of course

CENTRE FOR FERTILITY AND HEALTH

Case-cohort, Nested Case-Control, etc.





Contents

Preface

About the Editors

List of Contributors

I Introduction

1 Origins of the Case-Control Study
Norman E. Breslow and Noel Weiss

2 Design Issues in Case-Control Studies
Duncan C. Thomas

II Classical Case-Control Studies

3 Basic Concepts and Analysis
Barbara McKnight

4 Matched Case-Control Studies
Barbara McKnight

5 Multiple Case or Control Groups
Barbara McKnight

6 Causal Inference from Case-Control Studies
Vanessa Didelez and Robin J. Evans

7 The Case-Crossover Study Design in Epidemiology
Joseph A. \Chris" Delaney and Samy Suissa

8 Small Sample Methods
Jinko Graham, Brad McNeney, and Robert Platt

CENTRE FOR FERTILITY AND HEALTH

Case-cohort, Nested Case-Control, etc.

IV Case-Control Studies for Time-to-Event Data 283

16 Cohort Sampling for Time-to-Event Data: An Overview 285
Ørnulf Borgan and Sven Ove Samuelsen

17 Survival Analysis of Case-Control Data: A Sample Survey Approach 303
Norman E. Breslow and Jie Kate Hu

18 Nested Case-Control Studies: A Counting Process Approach 329
Ørnulf Borgan

19 Inverse Probability Weighting in Nested Case-Control Studies 351
Sven Ove Samuelsen and Nathalie Støer

20 Multiple Imputation for Sampled Cohort Data 373
Ruth H. Keogh

21 Maximum Likelihood Estimation for Case-Cohort and Nested Case-Control Studies 391
Donglin Zeng and Dan-Yu Lin

22 The Self-Controlled Case Series Method 405
Paddy Farrington and Heather Whitaker

V Case-Control Studies in Genetic Epidemiology 423

CENTRE FOR FERTILITY AND HEALTH

Ørnulf Borgan

Life event history analysis: Two days seminar in honour of Professor Ørnulf Borgan's 70th anniversary.

We invite you to a two-day seminar celebrating Professor Ørnulf Borgans many and substantial contributions to statistics in general and life event history analysis in particular. The seminar takes place in Auditorium 3 in Helga Engs House, Blindern Campus, University of Oslo, from Thursday 6th 11:00 to Friday 7th, 14:00, 2022.

Time and place: Jan. 6, 2022–Jan. 7, 2022, Auditorium 3 Helga Engs hus
Add to calendar

CENTRE FOR FERTILITY AND HEALTH

NORBIS/UNIVERSITY OF OSLO COURSE, DECEMBER

IMB9335 - Modern methods for analyzing survival and time to event data

IMB9335 - Autumn 2021

Schedule

IMB9335 December 6-10, 2021

Course venue

The course will take place at Runde Auditorium at Domus Medica (see map). Both lectures and practical exercises will take place in the auditorium.

Teachers

-Ørnulf Borgan, Department of Mathematics, University of Oslo.
-Håkon Gjessing, Norwegian Institute of Public Health.
-Morten Valberg, Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital.
-Odd O. Aalen, Oslo Centre for Biostatistics and Epidemiology, Institute of Basic Medical Sciences, University of Oslo.

Standard survival model:

Fig. A.2 *The standard survival model. From its initial state 0 ("alive") the process can move only to state 1 ("dead"). The hazard of moving to state 1 at time t is* $\alpha(t)$.

Illness-death model:

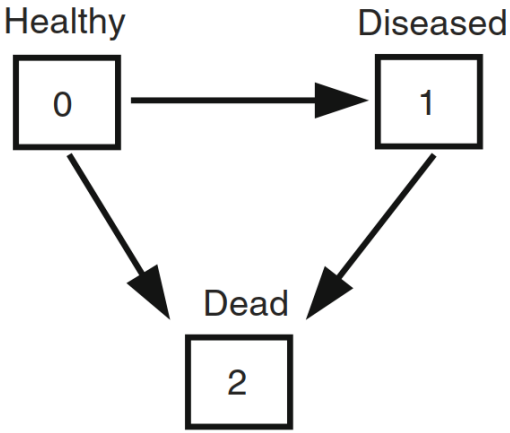


Fig. A.5 *The illness-death model. From its initial ("Healthy") state 0, an individual can either move to state 1 ("Diseased") or die (state 2). In this example, an individual in state 1 can only move to state 2, that is, there is no recovery.*

CENTRE FOR FERTILITY AND HEALTH

Conditional model:

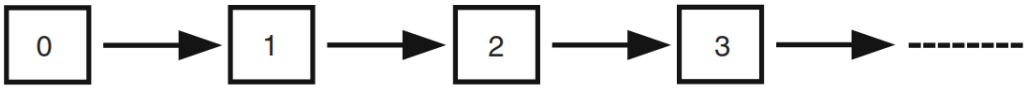


Fig. A.4 *The conditional model. The process moves along a series of stages, with the transition hazards* $\alpha_{01}(t)$, $\alpha_{12}(t)$,... .

Competing risk model:

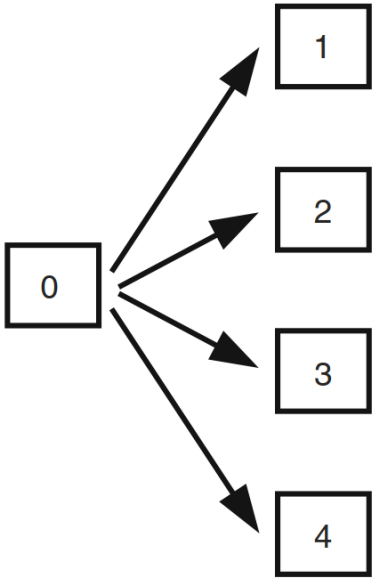
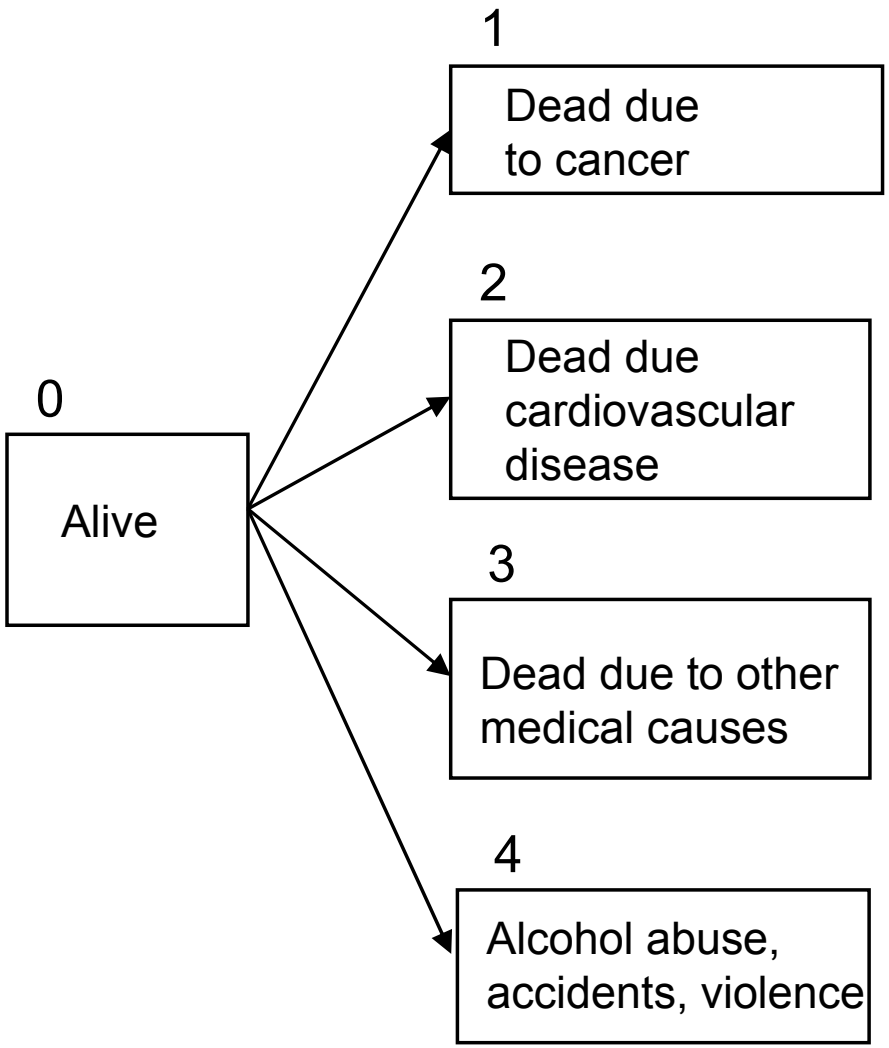


Fig. A.3 *The competing risk model. From its initial state 0 the process can move to any of a number of terminal stages. The hazard of moving to state* j *at time* t *is* $\alpha_{0j}(t)$.

CENTRE FOR FERTILITY AND HEALTH

40-ÅRSUNDERSØKELSEN

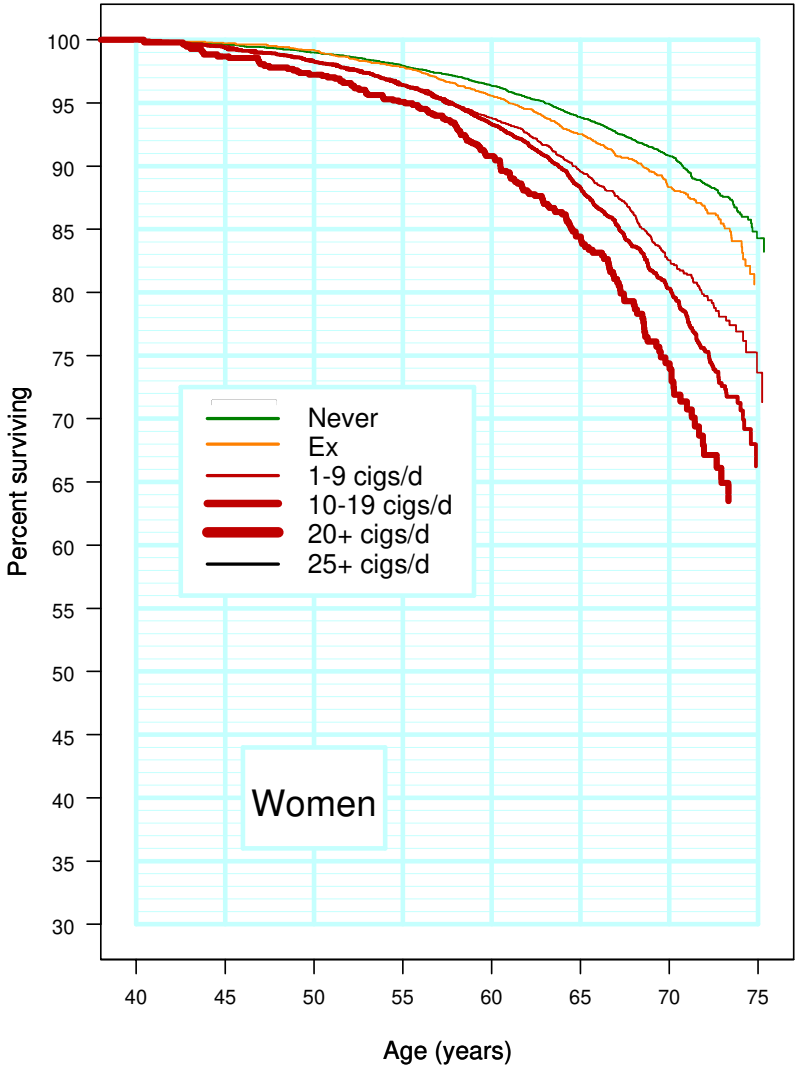
- Health screenings 1974–1978 (Statens Helseundersøkelser)
- Participants roughly 40 years old
- Record linkage Statistics Norway (SSB): Followed until end of 2000

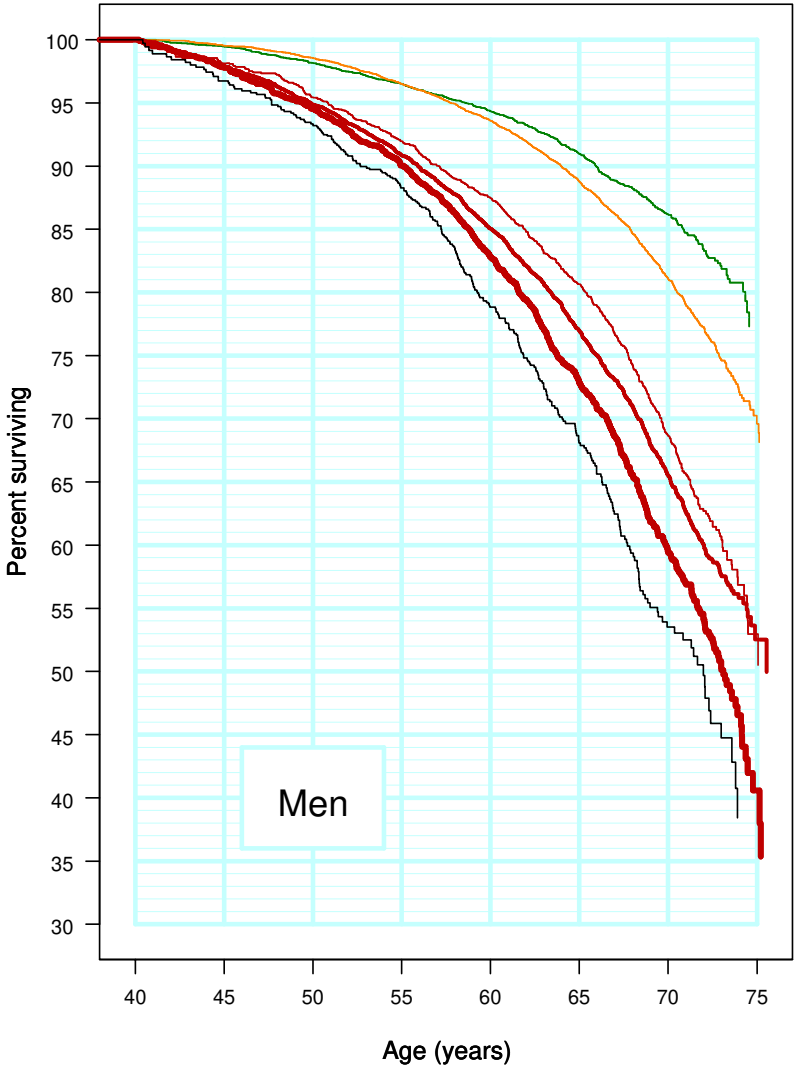


- Vollset, Tverdal, Gjessing (2006) *Annals of Internal Medicine*

CENTRE FOR FERTILITY AND HEALTH

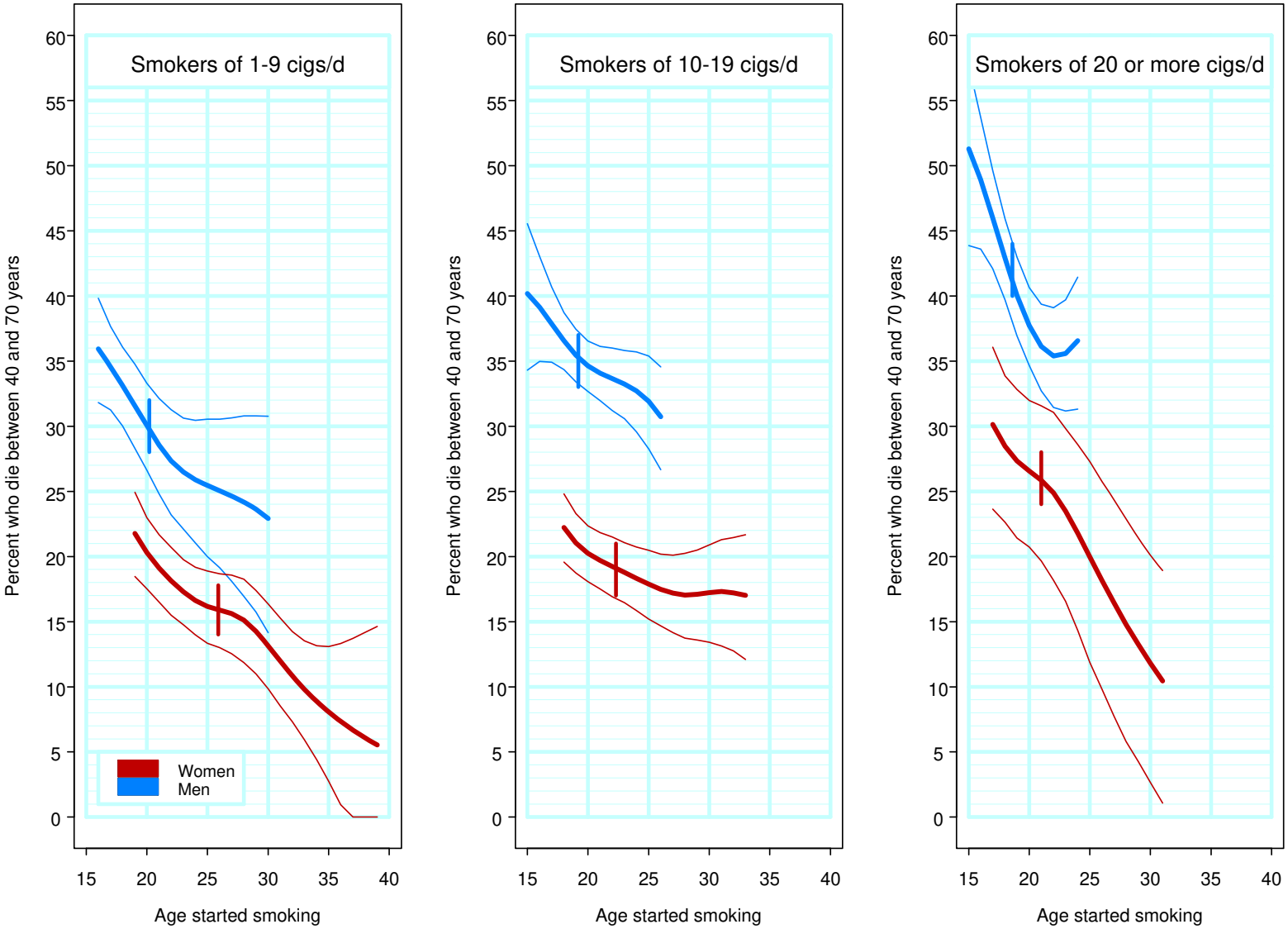
Smoking and deaths 40-70, percent surviving by age





CENTRE FOR FERTILITY AND HEALTH

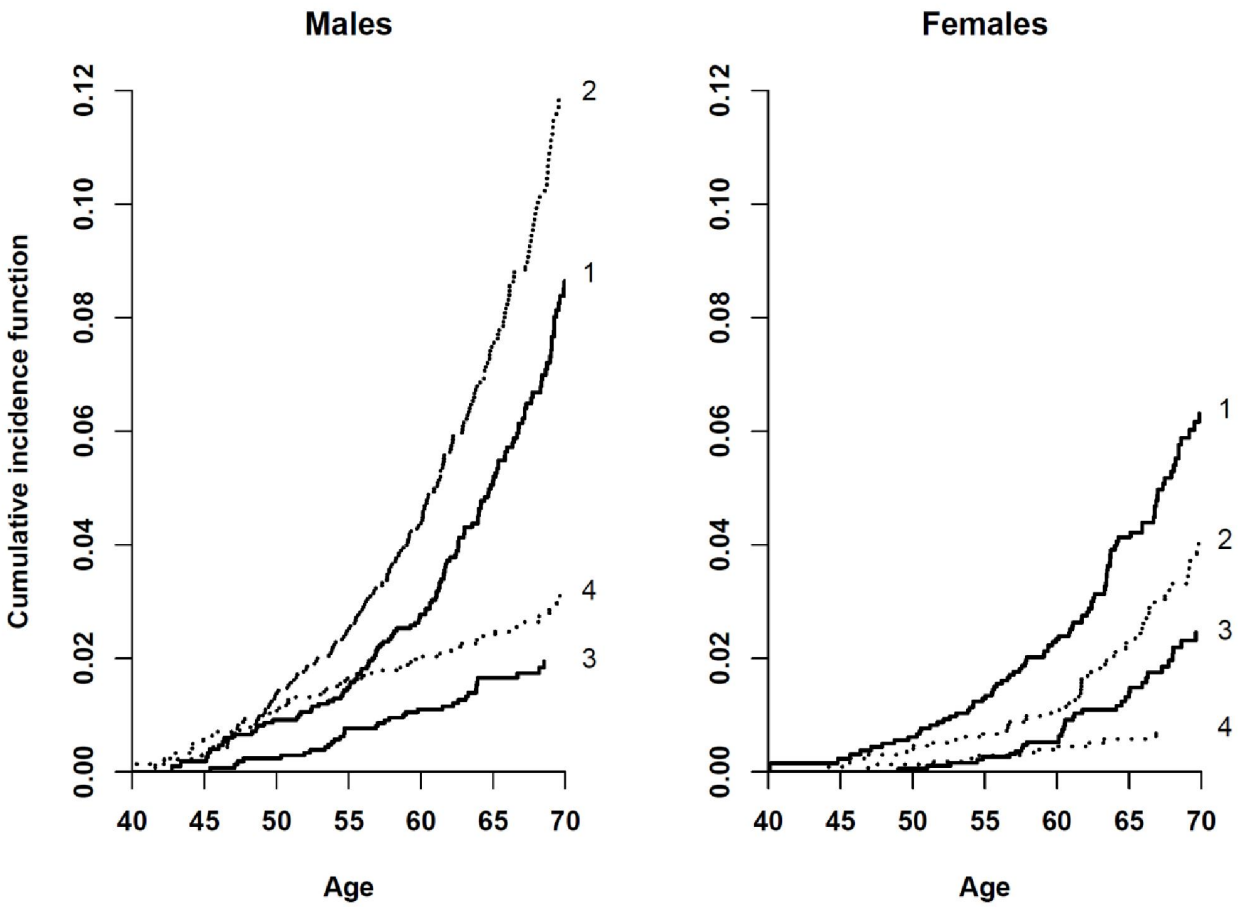
Percent deaths 40-70, effect of age started smoking



- Smooth spline in survival model

Cause-specific cumulative incidence

1. Cancer
2. Cardiovascular disease
3. Other medical
4. Alcohol abuse, violence, accidents



- Estimates of the probability of dying between age 40 and t of cause h
- "Cumulative incidence" is not a good name...

CENTRE FOR FERTILITY AND HEALTH

Smoking and deaths 40-70: Some aspects

- Age as time axis
- Estimate mortality 40-70 even though none of the subjects are followed all the way
  - Left truncation (delayed entry)
  - Right censoring
- “Smooth” effects of age at smoking start
- Handle competing risks of death
  - ... but effects of covariates are always difficult to interpret in a competing risk situation
  - Different causes of death may well be dependent, even conditional on covariates

CENTRE FOR FERTILITY AND HEALTH

H1N1 Pandemic

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Risk of Fetal Death after Pandemic Influenza Virus Infection or Vaccination

Siri E. Håberg, M.D., Ph.D., Lill Trogstad, M.D., Ph.D.,
Nina Gunnes, Ph.D., Allen J. Wilcox, M.D., Ph.D., Håkon K. Gjessing, Ph.D.,
Sven Ove Samuelsen, Ph.D., Anders Skrondal, Ph.D., Inger Cappelen, Ph.D.,
Anders Engeland, Ph.D., Preben Aavitsland, M.D., Steinar Madsen, M.D.,
Ingebjørg Buajordet, Ph.D., Kari Furu, Ph.D., Per Nafstad, M.D., Ph.D.,
Stein Emil Vollset, M.D., Dr.P.H., Berit Feiring, M.Sc.Pharm.,
Hanne Nøkleby, M.D., Per Magnus, M.D., Ph.D.,
and Camilla Stoltenberg, M.D., Ph.D.

ABSTRACT

BACKGROUND

During the 2009 influenza A (H1N1) pandemic, pregnant women were at risk for severe influenza illness. This concern was complicated by questions about vaccine safety in pregnant women that were raised by anecdotal reports of fetal deaths after

From the Norwegian Institute of Health (S.E.H., L.T., N.G., H.K.G., S
A.S., I.C., A.E., P.A., K.F., P.N., S.E.V
H.N., P.M., C.S.), the University of

H1N1 Pandemic

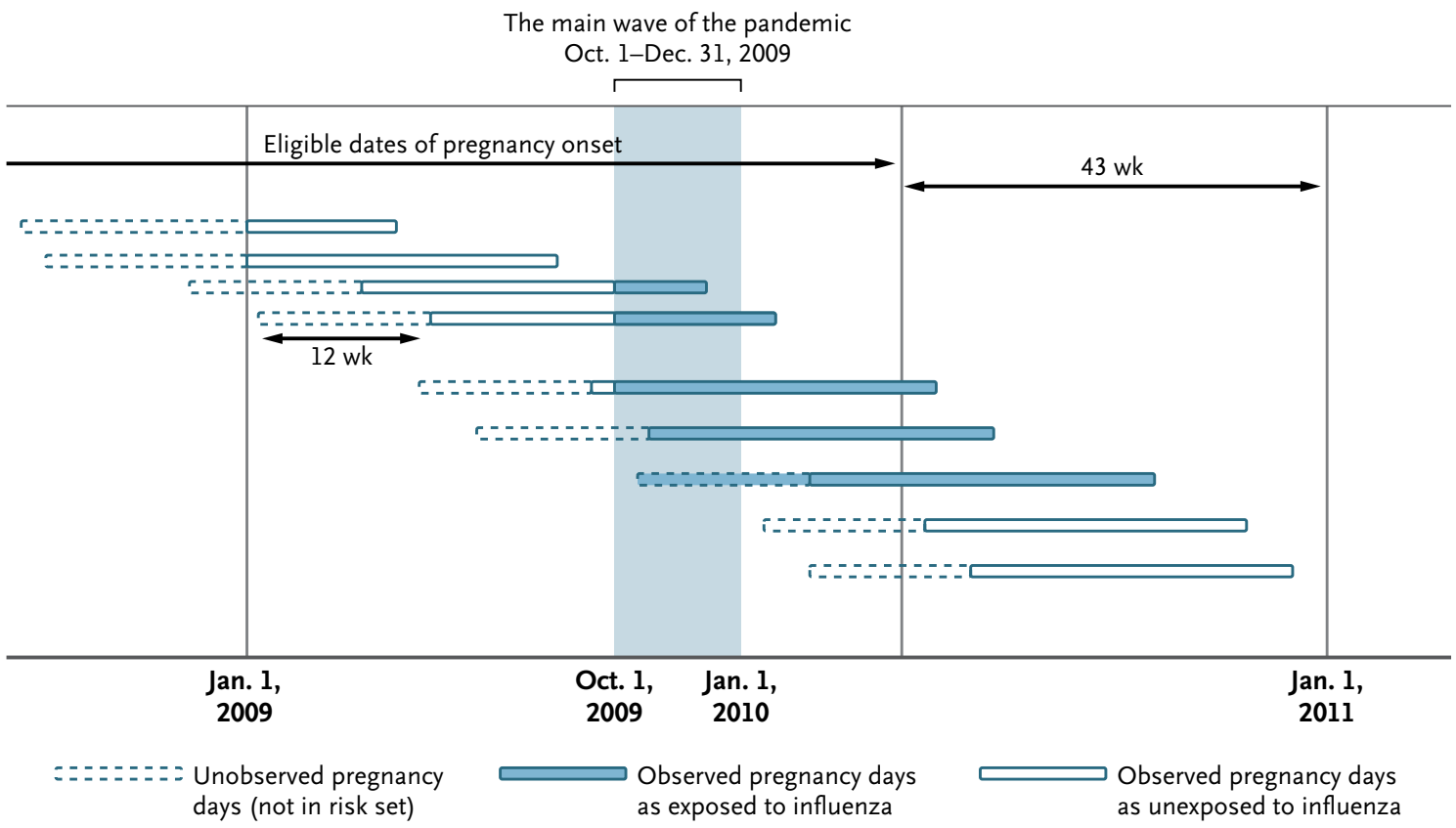


Figure 1. Eligible Pregnancies, Observed Pregnancy Days, and Exposure to the Main Pandemic Wave.
Births in Norway that occurred in 2009 and 2010 were eligible for the study if women had become pregnant at least 43 weeks before December 31, 2010. Eligible pregnancies were classified as involving maternal exposure to the influenza pandemic if any day of pregnancy occurred between October 1, 2009, and December 31, 2009. For a given pregnancy, days at risk were defined as pregnancy days after week 12 that occurred starting on January 1, 2009, and exposure days were defined as all pregnancy days from the first day of exposure until delivery. For simplicity, the figure shows all pregnancies as lasting 9 months. The study included all registered pregnancies lasting at least 12 weeks. The period of the main wave of the influenza pandemic is shaded.

Censoring or truncation?

Right censoring:

- Individuals are followed through the study
- Event (may) happen at an unknown time *later* than the time of censoring

Right truncation:

- Event (may) happen later than the time of truncation
- And for that reason the individual is *not known* to the study

---

Data from the Medical Birth Registry:

- Pregnancies do not enter the registry until birth has taken place
- You risk oversampling short pregnancies close to the end of the study

CENTRE FOR FERTILITY AND HEALTH

Logistic regression, fetal death vs. vaccinated

Total fetal loss

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	SL_VAKSSTATUS	-1.236	.120	106.041	1	.000	.291	.230	.368
	Constant	-3.882	.072	2909.368	1	.000	.021		

a. Variable(s) entered on step 1: SL_VAKSSTATUS.

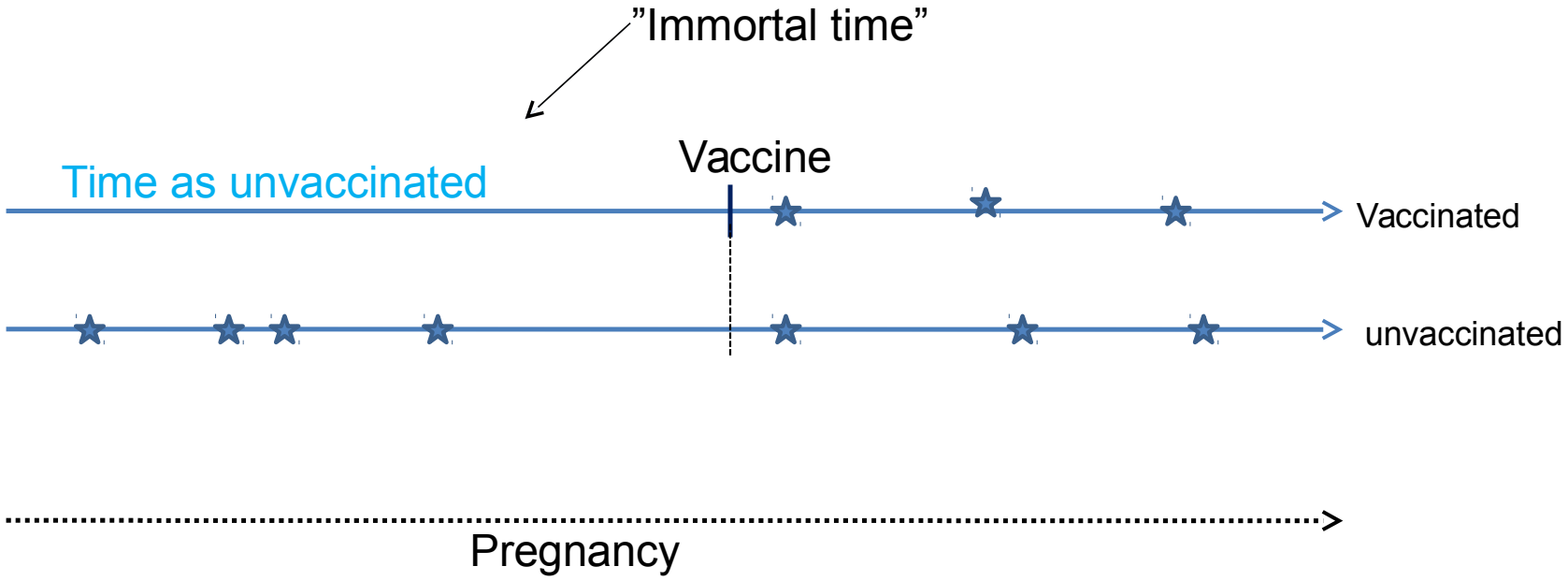
- The vaccine appears to be strongly protective
- ... even when accounting for possible truncation effects

But is this correct?

CENTRE FOR FERTILITY AND HEALTH

Problem: Immortal time bias

Håberg et al.



- Mothers receiving vaccine in the third trimester must by necessity still have an ongoing pregnancy
- Hence the term “immortal time”

RESEARCH METHODS & REPORTING

Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes

Linda E Lévesque,[1 2 3] James A Hanley,[1 4] Abbas Kezouh,[4] Samy Suissa[1 4]

[1]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montréal, Canada
[2]Department of Community Health and Epidemiology, Queen's University, Kingston, Canada
[3]Kingston, Frontenac, Lennox and Addington Public Health, Kingston, Canada
[4]Center for Clinical Epidemiology, Jewish General Hospital, McGill University, Montréal, Canada
Correspondence to: L Lévesque linda.levesque@queensu.ca

Accepted: 25 August 2009

Cite this as: *BMJ* 2010;340:b5087
doi: 10.1136/bmj.b5087

Immortal time in observational studies can bias the results in favour of the treatment group, but it is not difficult to identify and avoid

Well designed observational studies have made important contributions to our understanding of the risks and benefits of drug treatment. Such studies are often the first to identify or confirm important adverse health events associated with drugs, as seen recently with the cardiac effects of ergot derived dopamine agonists[1] and cyclooxygenase 2 inhibitors.[2 3] Observational studies can also assess aspects of drug safety, such as the time varying nature of risk, which cannot be readily appraised using an experimental design.[4]

Cohort studies are often preferred to case-control studies because they are less susceptible to certain biases.[5 6] However, the inappropriate accounting of follow-up time and treatment status in the design and analysis of such studies can introduce immortal time bias.[7]

What is immortal time bias?

Immortal time refers to a period of follow-up during which, by design, death or the study outcome cannot occur.[8] In pharmacoepidemiology studies, immortal time typically arises when the determination of an individual's treatment status involves a delay or wait period during which follow-up time is accrued—for example, waiting for a prescription to be dispensed after discharge from hospital when the discharge date represents the start of follow-up (box 1 see bmj.com).[9-14] This wait period is considered immortal because individuals who end up in the treated or exposed group have to survive (be alive and event free) until the treatment definition is fulfilled. If they have an event before taking up treatment they are in the untreated or unexposed group. Bias is introduced when this period of "immortality" is either misclassified with regards to treatment status or

CENTRE FOR FERTILITY AND HEALTH

Solution: Event history analysis

Cox regression with time-dependent covariates

- Time scale: Gestational age
- Event: Fetal death
- Censoring: Live birth

Time-dependent covariates

- Vaccination status
- Influenza diagnosis
- Pregnancy within influenza period

Constant covariates:

Age, Parity, Marital status, Supplements in pregnancy, Smoking in pregnancy

CENTRE FOR FERTILITY AND HEALTH

Results from Cox

Logistic regression:

Total fetal loss

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	SL_VAKSSTATUS	-1.236	.120	106.041	1	.000	.291	.230	.368
	Constant	-3.882	.072	2909.368	1	.000	.021		

a. Variable(s) entered on step 1: SL_VAKSSTATUS.

Cox:

Adjusted Hazard Ratio (95%CI) for fetal death given vaccination:

$$\mathrm{HR} = 0.88 \quad (0.66 - 1.17)$$

CENTRE FOR FERTILITY AND HEALTH

It all goes back to the Greeks...

Copyright © 1978 by The Johns Hopkins University School of Hygiene and Public Health *Printed in U.S.A.*

A FETAL-INFANT LIFE TABLE BASED ON SINGLE BIRTHS IN NORWAY, 1967–1973

LEIV S. BAKKETEIG,[1] DANIEL G. SEIGEL[2] AND PHYLLIS M. STERNTHAL[2]

Bakketeig, L. S. (U. of Trondheim, Dept. of Community Medicine, Eirik Jarlsgt. 4, 7000 Trondheim, Norway), D. G. Seigel, P. M. Sternthal. A fetal-infant life table based on single births in Norway, 1967–1973. *Am J Epidemiol* 107:216–225, 1978.

The study is based on 440,452 single births occurring in Norway, 1967–1973, with known gestational age. The information was collected through a notification system known as "Medical Registration of Births," covering all births occurring in Norway, and the data are made available through the Medical Birth Registry of Norway, which allows for linkage between births and infant deaths. The life table describes the experience of women still pregnant at a gestational age of 16 completed weeks, and states for each subsequent week the number of pregnancy terminations, the outcome, and the number of women still pregnant. Seven outcomes of pregnancy are considered: fetal death prior to labor, fetal death during labor, death within 24 hours, death 1–6 days, death 7–27 days, death 28 days–1 year, and survival of one year or more. The data in the life table provide information on the probability of pregnancy termination in each week of gestation (after 16 completed weeks), and the probabilities of the various outcomes. The fetal-infant life table is considered as an extension of descriptive perinatal statistics and is of value in monitoring health changes and in comparing

Downloaded from http://aje.oxfordjournals.org/ at

Cox (proportional hazards) regression

Hazard:

$$\alpha(t) = \alpha_0(t) \exp\left(\beta_1 x_1 + \beta_2 x_2 + \cdots\right)$$

- $\alpha_0(t)$ is the *baseline hazard*
- $x_1, x_2, \ldots$ are the covariates
- $\beta_1, \beta_2, \ldots$ are the corresponding parameters

Covariates:

$x_1, x_2, \ldots$ covariates may change over time

Baseline hazard:

$\alpha_0(t)$ is the hazard (at time t) when all $x_1 = x_2 = \cdots = 0$

CENTRE FOR FERTILITY AND HEALTH

Hazard:

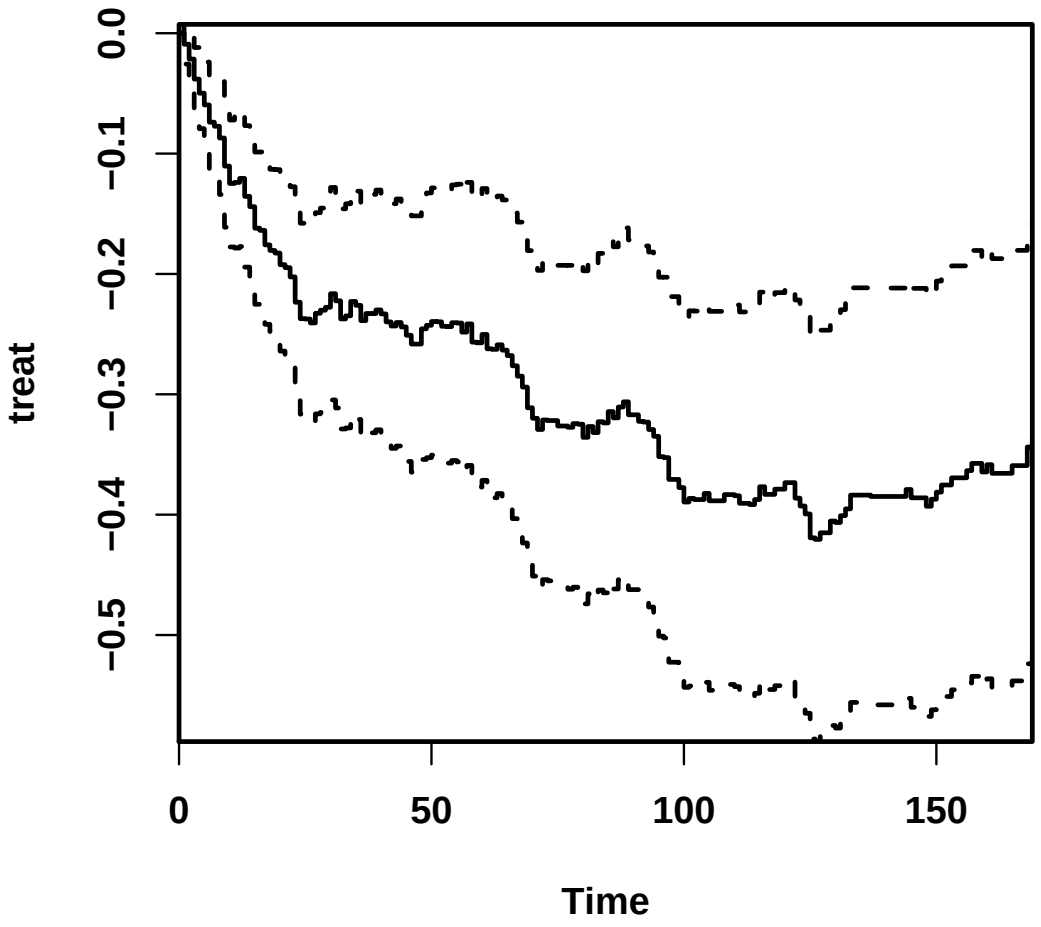
$$\alpha(t) = \beta_0(t) + \beta_1(t)x_1 + \beta_2(t)x_2 + \cdots + \beta_K(t)x_K$$

- $\beta_0(t) = \alpha_0(t)$ is the *baseline hazard*
- $x_1, x_2, \ldots$ are the covariates
- $\beta_1, \beta_2, \ldots$ are the corresponding parameters
- Note the improvement: $\beta_i(t)$ are allowed to depend on time!

Covariates:

$x_1, x_2, \ldots$ covariates may change over time

Aalen regression, example treatment effect (with CI)



- Treatment has a strong effect first 25 days
- Treatment has a moderate effect days 25-100
- Treatment has no effect after 100 days

- Negative slope means reduced hazard
- Flat curve means no effect

CENTRE FOR FERTILITY AND HEALTH

TIME-VARYING...

Time-dependent covariates:

- Exposure (and other covariates) can change over time
- But effect of exposure (and other covariates) assumed the same regardless of time
- Available in both Cox and Aalen

Time-dependent effects:

- The *effect* of exposure (and other covariates) can change over time
- This can be combined with time-dependent covariates
- Available only in Aalen
- (...although various types of Cox-adaptations do exist)

CENTRE FOR FERTILITY AND HEALTH

Cox proportional hazards or Aalen additive hazards?

Cox:

+ Well known

+ Easily summarized
(a single Hazard Ratio for the treatment effect)

- Requires proportional hazards

Aalen:

- Less well known

- Harder to summarize
(cumulative effects curves)

+ Much more flexible than Cox
(allows effect to change over time)

+ No assumption of proportional hazards

CENTRE FOR FERTILITY AND HEALTH

Thank you!

CENTRE FOR
FERTILITY AND HEALTH

MECHANISMS OF DISEASE

Mechanisms of disease

Protection from natural infections with enterotoxigenic *Escherichia coli*: longitudinal study

Hans Steinsland, Palle Valentiner-Branth, Håkon K Gjessing, Peter Aaby, Kåre Mølbak, Halvor Sommerfelt

Summary

Background Enterotoxigenic *Escherichia coli* (ETEC) are an important cause of diarrhoea and diarrhoeal deaths in children living in developing countries and of travellers' diarrhoea. During the past 25 years, vaccine development efforts have been focused on induction of protective immunity against surface colonisation factors (CFs) and the heat-labile enterotoxin. Although vaccines that induce immunity to heat-labile toxin offer protection against diarrhoea from ETEC that produce this toxin, the benefit of including CF antigens remains uncertain. We aimed to estimate the protection that natural ETEC infections induce against new infections.

Methods In Guinea-Bissau, we followed up 200 neonates until up to age 2 years, most of whom were breastfed throughout the study. We collected stool specimens from the children every week irrespective of whether they had diarrhoea. As a measure of protection, we used Cox regression models to estimate the change in infection rates after a primary ETEC infection. We thus estimated the protection attributable to CFs, toxins, and to any other

Introduction

ENTEROTOXIGENIC ESCHERICHIA COLI (ETEC) are an important cause of diarrhoea in children in developing countries—an estimated 380 000 deaths occur annually.[1] ETEC are also the main cause of travellers' diarrhoea.[2,3] Infections with ETEC are often asymptomatic in people living in endemic areas.[4] ETEC that cause disease in human beings produce one or more of three protein enterotoxins—porcine and human heat-stable toxins and heat-labile toxin—and might produce one or more of several different COLONISATION FACTORS (CFs). These factors are protein surface antigens that mediate adherence to the small intestinal mucosa.[4] To date, 21 different CFs have been identified.[5,6]

In developing countries, the incidence of ETEC-associated diarrhoea falls during the first 5 years of life,[7] whereas children and adults from industrialised areas who travel to these countries are susceptible to this type of diarrhoea.[2,3] This difference suggests that children who live in endemic areas develop a natural immunity to ETEC over time, which offers hope that effective vaccines can be developed.

During the past 25 years, the ETEC vaccine development effort has been focused on induction of

CENTRE FOR FERTILITY AND HEALTH

Enterotoxigenic E. coli infections in children

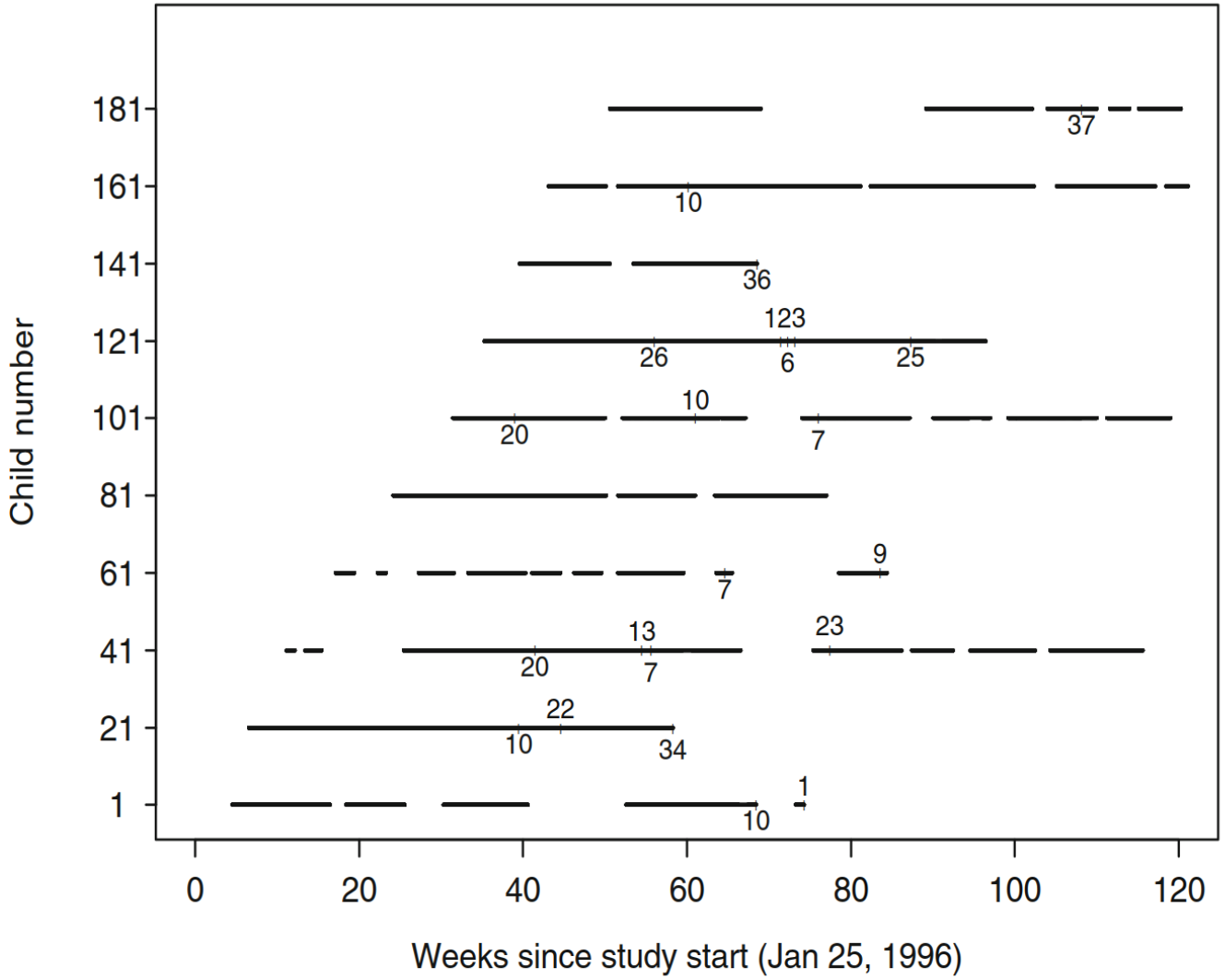


Fig. 8.14 *Follow-up and infection pattern for every 20th child included in the Guinea-Bissau study. Horizontal lines indicate period of follow-up for each child. Numbers above and below lines indicate infecting strain. Horizontal positioning of numbers indicates time at which the infection was observed. Numbers alternate above and below line to enhance readability.*

NTRE FOR
RTILITY AND HEALTH

Is there a protective effect of earlier infections?

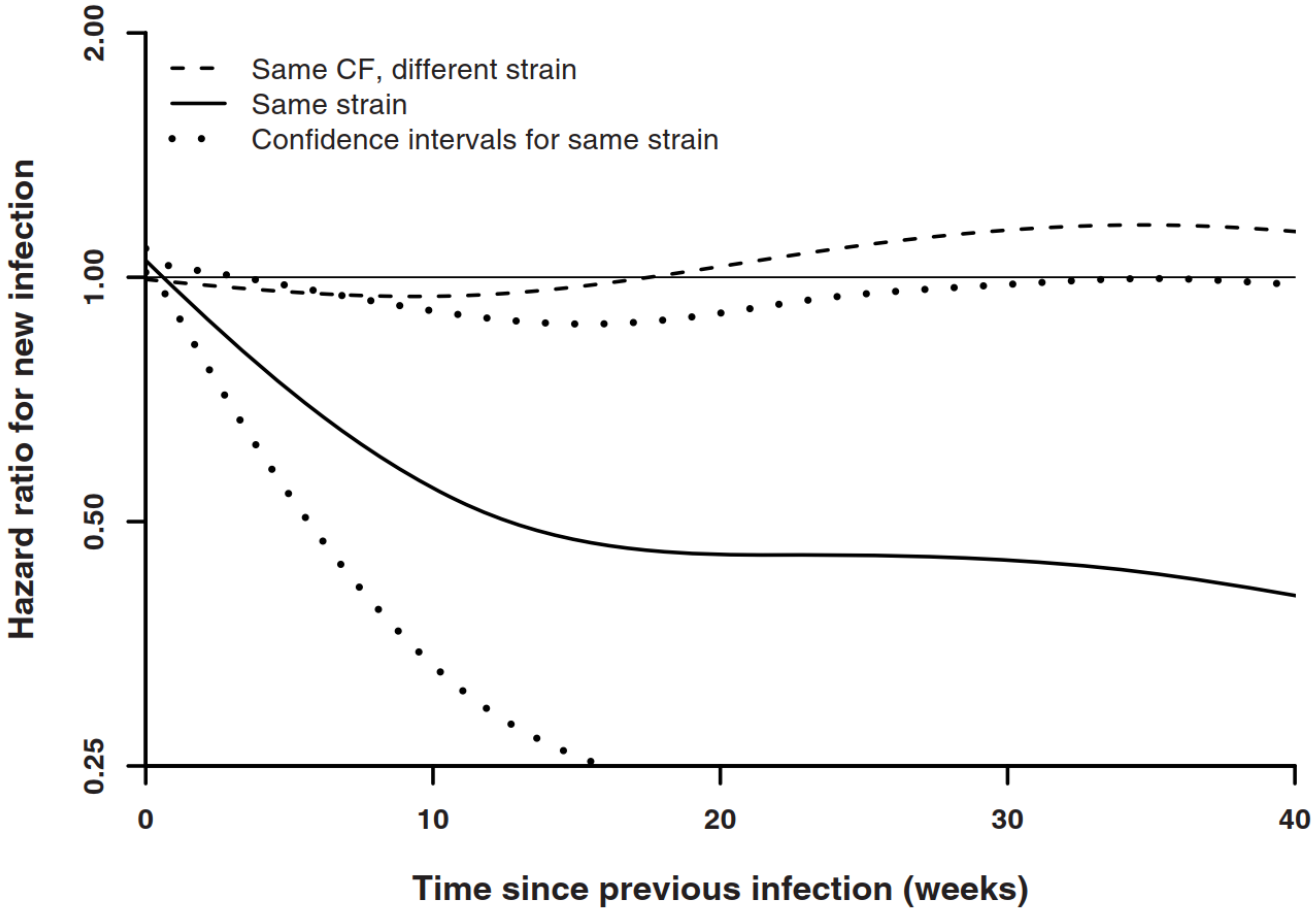


Fig. 8.16 *Protective effect as a function of time since last infection. Solid line shows the effect estimate for same strain infections, with 95% pointwise confidence interval (dotted). Dashed line shows the effect of a previous infection by a different strain with the same colonization factor.*

NTRE FOR
RTILITY AND HEALTH