MATHEMATISCHES FORSCHUNGSINSTITUT OBERWOLFACH

Tagungsbericht 46/1999

Design and Analysis of Infectious Disease Studies

28.11 - 04.12.1999

This conference organized by N. Becker (Canberra), K. Dietz (Tübingen) and N. Keiding (Copenhagen) gathered 41 participants. While the previous meetings on epidemics in 1989 and 1994 dealt with modelling the spread and the control of infectious diseases, this conference focussed on the associated statistical problems. This topic reflected the great methological progresses made in recent years, for example concerning applications of Monte-Carlo-Markov-Chain methods (MCMC) or new kinds of survival analyses. Main talks took place every morning, additional talks on miscellaneous subjects on the first two afternoons. The morning talks were grouped under the following topics:

1. Data augmentation (Expectation-Maximization algorithm (EM) and MCMC methods)
2. Vaccine efficacy studies
3. Heterogeneities with respect to households, age or time
4. Disease specific applications

The discussion sessions intended to start collaboration on new projects. The resulting ideas for future joint work were addressed by four working groups:

1. Design of studies
2. Networks: Models and Inference
3. Vaccination: Efficacy, Coverage and Policy
4. Space-time: Models and Analysis

Current information about the progresses made within these projects can be obtained from a web page (http://www.ma.hw.ac.uk/epi/index.html) kindly maintained by D. Mollison (Edinburgh). The following abstracts are in alphabetical order.

Hans-Peter Duerr

Bayesian data augmentation for recurrent infections

Kari Auranen, T. Leino, A. K. Takala & E. Arjas

Bayesian data augmentation using Markov chain Monte Carlo (MCMC) methods provides a flexible tool to fit stochastic epidemic models to data with incomplete observations on the underlying processes. In this study, we present a data augmentation model that describes transmission of asymptomatic mucosal infection (carriage) of *Streptococcus pneumoniae* (Pnc) bacteria in families. The panel data comprise measurements of the presence/absence of the three most common Pnc serotypes (6B, 19F, 23F) in all members of 97 families at ten time points over a period of two years. Spread of recurrent Pnc carriage is modelled as a multivariate point process, involving event times of carriage acquisition and clearance in each individual family member. Apart from an SIS type transmission within families, the model includes carriage transmission from the surrounding community. Age-dependent rates of acquiring carriage are modelled non-parametrically. We construct a reversible jump MCMC algorithm to draw samples from the joint posterior distribution of the unobserved events (individual times of acquisition and clearance of Pnc carriage) and the model parameters (acquisition and clearance rates).

Seasonal Transition Matrices for Studying the Population Dynamics of the Lyme Disease Tick

Tamara Awerbuch & S. Sandberg

For the purpose of analysing conditions leading to the emergence of Lyme disease, the various components involved in transmission are linked using a multiple matrix model which was in turn used to follow the population dynamics of the tick. The model was implemented with parameter estimates derived from a long term study in Nantucket Island and Ipswich, Massachusetts. The main inputs of the model are searching efficiency of the tick and abundance of its main hosts (the deer and the white footed mouse). This enabled us to follow tick abundance month by month, and discover the patterns of seasonal fluctuations. A qualitative analysis of annual dynamics indicated that independently of the trend, tick population will osccillate at its initial stages of growth.

Monte-Carlo-within-Metropolis algorithm for household infections data

David Balding

Statistical inference for infectious disease data is problematic, in part because only part of the transmission process is observed. Likelihoods for realistic models then involve integrations over the unobserved values, and are usually intractable. Augmented-data methods can often simplify the likelihood, but at the cost of an increased dimensionality of the parameter space. We propose an alternative approach to this problem, which is applicable in certain settings within a Metropolis-algorithm framework for Bayesian inference. Instead of calculating the likelihood at each step of the algorithm, it is approximated by simulation of a large number of disease transmission events. The basic approximation can be improved via bias corrections using ideas similar to those employed in bootstrap bias estimation. The method is illustrated by application to a classic dataset involving measles transmission within households.

Statistical inference for SIR epidemics among a population of households

Frank Ball

This talk is concerned with a stochastic model for the spread of an SIR epidemic among a population consisting of a large number of small households, with different infection rates for between-household and within-household infections. The threshold behaviour of the model is briefly outlined. Methods for making statistical inferences about the parameters governing such epidemics from final outcome data, when only some of the households in the population are observed, are described and their asymptotic properties, as the number of households becomes large, are determined.

Specifically, a pseudo-likelihood approach is used to obtain estimates of the between- and within-household infection rates, together with associated asymptotic confidence sets, and to derive hypothesis tests for homogeneity of parameters driving different epidemics. Goodness of fit tests are also briefly considered, as are extensions to multitype epidemics. The methodology is illustrated by simulations and by an application to data on influenza epidemics in Tecumseh, Michigan.

The use of data from large-scale health surveillance studies to identify non-random clustering of adverse health outcomes and their determinants. Theoretical investigations and application to the long-term field project in Nouna, Burkina Faso

Heiko Becher

We investigate whether data from demographic surveillance systems (DSS) can be used to identify non-random clustering of adverse health outcomes in a population, for example childhod mortality from malaria. For the project we have data available from some 30000 inhabitants from a study center in Nouna (Burkina Faso) where a DSS has been implemented in the beginning of the 80's. It is expected to have certain restrictions with respect to data quality and completeness, such as incomplete information on number of births / deaths over time, inaccurate migration figures, missing or error-prone information both for mortality and morbidity. We discuss plans to investigate the power to detect disease clusters under different sampling schemes and under different assumptions of variablility of the underlying data. In this analysis assumptions will be made on error stuctures in the data. The methods will be applied to the Burkina Faso data set.

Are Inequalities the Way to Tackle Inequalities?

Niels Becker

Real epidemics occur in heterogeneous settings. Individuals differ and the community has geographic and social structure. What are the consequences of using simplified assumptions when making inferences? How can we best address the estimability problems that arise for multitype epidemic data? We illustrate that inequalities are often an effective way to tackle these questions. Specifically, we illustrate that protective vaccine efficacy can be bounded by parameters that can be estimated, without making the simplifying assumptions necessary for explicit estimation of the vaccine efficacy. Estimation of the reproduction number for multitype epidemics can be dealt with similarly. We also illustrate the use of inequalities to inform us about biases in estimates when we make simplifying assumptions.

"Le Réseau Sentinelles": The French Communicable Diseases network

Pierre-Yves Boëlle, A. Flahault, A.-J. Valleron

The "Réseau Sentinelles" is a network of general practitioners dispersed over the French territory. Since 1984, they have been monitoring 9 conditions, including flu-like syndromes and diarrhea; and childhood diseases (measles, mumps, chickenpox). Over 15 years, approximately 500 000 cases of the different conditions have been described individually. Four different goals are pursued within the Réseau Sentinelles.

1) to monitor the spread of epidemics in France, by the development of early alert methods and of geographical representation tools.

2) to monitor the impact of interventions, like changes in the vaccination coverage.

3) to provide data for research purposes : age distribution of cases, geographical spread of diseases, contact and mixing patterns.

4) to provide a workbench for other selective surveys: case control surveys, incidence and prevalence of diseases monitored in the general population.

The Réseau Sentinelles has been a WHO collaborating centre for the electronic surveillance of diseases since 1997, and in this respect has extended its activity to centralizing communicable diseases data at the scale of the world. As for all the data of the Réseau Sentinelles, those are available without restriction on the internet.

Estimation in epidemics: homogeneity vs. individual heterogeneity

Tom Britton

Suppose an epidemic outbreak resulting in the proportion $\tilde{p}$ infected is observed. Under the assumption of a completely homogeneous community the basic reproduction number R_0 and the critical vaccination coverage v_c have well-known estimators. Are these estimators valid in case there are individual heterogeneities? No! It can be shown that any heterogeneity in susceptibility will have the effect that the estimator underestimates R_0 and v_c. One idea is therefore to separate the community in (hopefully) homogeneous sub-groups and observe $(\tilde{p}_1, \ldots, \tilde{p}_k)$, the observed proportion infected in each subgroup. A negative result is then that R_0 and v_c cannot be estimated consistently. However, one can get consistent estimates on upper bounds (i.e. worst cases) of these quantities.

Design and analysis of transmission experiments

Mart C.M. De Jong

In experiments each animal can only become infected by the experimenters (inoculation) or by other infected animals (contact infection). A count of the number of contact animals that become infected can be used to estimate and compare transmission. Different situations are encountered. In experiments with virus infections it can often only be determined afterwards that the animal has been infected during the experiment. The exact moment of infection is then not known (De Jong & Kimman (1994), Bouma et al. (1995,1996,1997a,b), Muller et al. (1995), De Wit et al. (1998), mostly in Vaccine). In experiments with bacteria in contrast, it is often possible to know during the experiment at which moments the animal is infectious and also the moment at which the animal has become infected. A complication in experiments with bacterial infections is that animals are often infectious beyond the duration of the experiment.
If only final size is known the analysis can be based on the distribution of the number of infected animals at the end of a transmission experiment $f(x \mid R, S_0, I_0, N_0)$. This distribution can be determined for the base line model where the infection event $(S, I) \rightarrow (S-1, I+1)$ occurs with rate $\beta SI/N$ and recovery event $(S, I) \rightarrow (S, I-1)$ occurs with rate $\alpha I (R = \beta/\alpha)$. When we assume that we know S_0, I_0 and N_0 the probabilities only depend on R and thus R can be estimated and compared between treatments (De Jong & Kimman, 1994). As the null hypothesis that the R after vaccination equals the R without vaccination is compound the maximum p-value over all R has to be calculated (Kroese & De Jong, in prep.). Optimal experimental designs have been determined: control group $R \geq 1$ and vaccine group $R \leq 1$, then use larger groups where 50% initially inoculated, both $R \geq 1$, then use pairs.

Prediction of ear infections in incompletely observed immunological processes

Mervi Eerola, A. Andreev & D. Gasbarra

We present a joint dynamic model for the immune system describing the inter-dependence between infection, immuno-response and risk of disease. The three submodels are estimated simultaneously in one large Markov chain Monte Carlo algorithm. As an example, we consider the risk of recurrent ear infections (acute otitis media, AOM) caused by *Streptococcus pneumoniae* (Pnc) when having only partially observed information on the colonisation by Pnc and try to establish the protective role of antibodies induced by certain surface protein antigens of Pnc. Our aim is to combine the different sources of observed data as effectively as possible. In the Bayesian estimation, the three stochastic models are thus the prior information of the system dynamics and the observed data points are used to restrict the

analysis to the realisations of the processes which are supported by the whole data set. Since the complete model involves a large number of unobservables and single parameters are not necessarily well identified, nor even of particular interest, we base our inference on predictive probabilities of ear infection within the infection episodes, given the history of all subprocesses.

Estimation of the time-dependent vaccine efficacy from a measles epidemic

M. Eichner, H. H. Diebner & K. Dietz

We examine a measles epidemic which occurred in 1992/93 in Germany. Data are available on the age of the cases, their date of vaccination if vaccinated and on the overall vaccination coverage of the population. Individual transitions from maternal protection to susceptibility and (given successful vaccination) to protection are described by a set of differential equations. The vaccination coverage was far too low to prevent a measles epidemic and the majority of cases were unvaccinated. According to a maximum likelihood parameter estimation, vaccine efficacy was negligible before 1978; it was about 80% in 1978-82, 97% in 1982-90 and 89% after 1990. Loss of immunity was estimated to be zero.

Estimation of R_0 using serological survey data

Paddy Farrington

In homogeneously mixing populations the basic reproduction number R_0 of an infection in endemic equilibrium can be estimated as $1 + L/A$ where L is the life expectancy and A the average age at infection. However for heterogeneous populations the estimation of R_0 is greatly complicated by the fact that contacts between individuals are usually unobserved. For instance in an age stratified population R_0 cannot be determined without strong assumptions using observations on age-specific incidence.

We seek to control the indeterminacy problem by exploiting information on the route of infection. We postulate that infections with the same route of transmission should have similar contact matrices, up to a constant of proportionality. A specific similarity criterion is proposed based on the likelihood under direct proportionality. This enables us to quantify the relative evidence for different contact structures. Alternatively we can assess model uncertainty using Bayesian methods.

Using data on several infections transmitted by the same route has other advantages. When paired serological survey data on two infections are available we can test for indi-

vidual heterogeneity, and incorporate its effects into the estimates of R_0. These methods are widely applicable since there exist extensive sources of serological data. We illustrate the methods with data on mumps and rubella from the UK.

A time series model for extinction and recurrence of childhood epidemics: Estimation and inference for measles outbreaks

Bärbel F. Finkenstädt & B. T. Grenfell

A key issue in the dynamical modelling of epidemics is the synthesis of complex mathematical models and data by means of time series analysis. We report such an approach, focusing on the particularly well documented case of measles. We propose the use of a discrete-time epidemic model comprising the infected and susceptible class as state variables. The model uses a discrete-time version of the SEIR type epidemic models, which can be fitted to observed disease incidence time series. We describe a method for reconstructing the dynamics of the susceptible class, which is an unobserved state variable of the dynamical system. The model provides a remarkable fit to the data on case reports of measles in England and Wales from 1944 to 1964. Morever, its systematic part explains the well documented predominant biennial cyclic pattern. We study the dynamic behaviour of the time series model and show that episodes of annual cyclicity arise as response to a quicker replenishment of the susceptible class during the baby boom, around 1947.

Using Validation Sets for Exposure to Infection and for Outcomes in Vaccine Field Studies

M. Elizabeth Halloran

Methods for adjusting for bias in estimates due to mismeasured or missing covariates and outcomes have been developed in many types of health studies. These methods can be used for the efficient design and analysis of vaccine studies as well. On the one hand, exposure to infection can influence estimates of vaccine efficacy, but good data on exposure is difficult to obtain. On the other hand, non-specific case definitions can lead to very attenuated efficacy estimates, but confirmation by culture of the infectious agent is also expensive and difficult. Using semiparametric methods, we show how use of small validation sets can correct the bias of the estimate of the large main study while maintaining efficiency. We illustrate the methods for exposure to infection on the example of an HIV vaccine trial (joint with G.T. Golm and I.M. Longini), and for culture-confirmed outcome of true flu on

the example of a flu vaccine trial (joint with I.M. Longini). Use of these efficient designs and methods of analysis for vaccine field studies will improve estimation of vaccine efficacy for both susceptibility and infectiousness.

Aggregation in macroparasitic infections

Valerie Isham

An important feature of macroparasitic infections is that the parasite load is generally aggregated (overdispersed) with a few hosts having most of the parasites. The causes of aggregation include clumped infections where the host is infected simultaneously by a random number of parasites, and host heterogeneity. In the host, immune reactions to parasitic infection may affect the resistance of the host to further parasite establishment, the mortality of the parasites and their fertility. Host heterogeneity means that the strengths of these immune reactions may vary between hosts, or that there may be heterogeneity of exposure because of behavioural differences, or of susceptibility which is reflected in parasite-induced host mortality. The effects of different sources of aggregation will be discussed (with numerical examples) in the context of a general stochastic model of the transmission dynamics, for which many explicit analytic results are possible.

In some parasitic infections, the evolution of resistance against anti-parasitic drugs is an important topical issue. Therefore the model will be used to explore the effects of aggregation on parasite persistence/extinction and, in particular, the persistence of treatment-resistant genotypes.

Further details may be found in Herbert, J. and Isham, V. Stochastic models for host-macroparasite interactions. J. Math. Biol. (2000) to appear; Cornell, S., Isham, V. and Grenfell, B. Drug-resistant parasites and aggregated infection—early season dynamics. (1999) Submitted.

Discrete time branching processes with binomial migrations in epidemiology

Christine Jacob

The binomial Markov chains, especially the Reed-Frost one, are used since a long time to model a disease spread in a closed population. When the population is open or very large, the branching processes offer an interesting stochastic alternative which allow for example the study of the asymptotic behavior of the process itself. We begin to develop such models for studying the propagation of contagious diseases in breedings when the

offsprings are taken into account as well as the controlled migrations of the animals to the outside. These models allow to distinguish between the vertical transmission and the horizontal one. Models are of the size-dependent type with binomial migrations from one state to another one and can include feedback controls. The size-dependence property allows a nonextinction of the infectives with a limit reproduction number equal to one.

Hierarchical modelling of infectious disease incidence

Leonhard Knorr-Held & S. Richardson

We propose a statistical framework based on Bayesian hierarchical models for the analysis of surveillance data on infectious diseases in space and time. Space-time modelling can help to characterize the specific patterns associated with particular diseases and/or predict their development. The proposed model class provides a useful framework for flexible nonparametric modelling of temporal, spatial and spatio-temporal dependences. For example, seasonal patterns, which are inherent features of such data, can be integrated in the formulation in a nonparametric fashion. The models also possibly incorporate the numbers of infected cases at time $t-1$ as additional covariates for the rate of infection at time t. This helps to distinguish between "sporadic" and "epidemic" disease incidence. Throughout, Markov chain Monte Carlo methods will be used for statistical inference. Model selection is done based on the posterior deviance. As an example we analyze the space-time variation of meningitis incidence in France on a monthly basis over the period 1985-1997.

GERMS, A Model For Transmission System Analysis and Study Design

James Koopman, S. Chick & C. Riolo

GERMS is a discrete individual, continuous time framework for stochastic modeling of infection transmission through dynamic contact networks. It facilitates mathematical model analysis and allows for construction of stochastic models with mean equilibrium behavior equal to corresponding deterministic models. Three projects using GERMS are 1) the design of STD surveillance systems, 2) the evaluation of control measures to stop waterborne transmission of pathogens, and 3) the design of statistics that predict network influences on transmission dynamics. GERMS facilitates analyses of the sensitivity of design decisions in these projects to a variety of realistic complexities. In GERMS, individuals are located in geographic and social space and have a set potential to make linkages with other individuals. This total potential is partitioned across geographic and social space.

The duration and nature of the links formed between individuals are assigned randomly from distributions characteristic of the site where linkages are formed. By specifying the geographic and social location of both individuals and the sites where they meet, GERMS can reproduce the network structures of many other models. The theoretical utility of GERMS is demonstrated by mathematical derivations predicting model behavior for situations where the corresponding deterministic model is difficult to analyze and for realistic situations, like high degrees of dynamic concurrency, where corresponding deterministic models are impractical or impossible.

Modern Surveillance of Infectious Diseases: Necessary Structures and Elements of an Effective Surveillance System

Alexander Krämer

The main goal of a modern surveillance system for infectious diseases is to improve the health and quality of life of the population. All attempts for reforms and evaluations of surveillance systems must be tailored towards that aim. By the use of instruments for standardization and innovative information and communication technologies a modern surveillance of infectious diseases can supply policy makers, health professionals and the public with high quality information about the diseases, causing agents and the immunological profile of the population in order to achieve an effective outbreak management and make long-term trend predictions. This process has to take place at all different levels - the national and international level as well as the regional and local level. Modern surveillance systems should be targeted at the users and consumers of the information - patients, health professionals, politicians - by specific communications rather than a mere dissemination of data in order to achieve an active participation of the different groups. This active participation is the key to convert the generated surveillance data into effective public health actions.

Infectious disease studies within general health information systems

Klaus Krickeberg

In some countries like Vietnam there exists a public health network which includes the hospitals and reaches down to commune and village health centers. Such networks can support a "general" or "integrated" health information system which serves at planning, managing, monitoring, and evaluating the normal medical routine activities as well as the

special "vertical" health programmes (tuberculosis, malaria, diarrhoea, acute respiratory infections, AIDS, immunizations, etc.), and includes an epidemic surveillance component. It consists of registers in the hospitals and health centers and of reports, mainly of a statistical nature, to be sent regularly to the relevant higher health authorities which concern both the disease pattern and management and logistics. Such an information system can be used for various other purposes. A few examples are mentioned: epidemiologic studies of risk factors, in particular case-control studies, the analysis of the mechanism of the spread of epidemics, the evaluation of case management strategies (diagnosis and treatment; example: malaria), drug surveillance. Some studies are done by combining knowledge obtained from the records with additional information (example: shigellosis incidence).

Using Stochastic Models to Estimate HIV Vaccine Efficacy

Ira M. Longini, Jr., M. G. Hudgens & M. E. Halloran

We use a Markov model to estimate vaccine efficacy for suceptibility (VE_S) and infectiousness (VE_I) for a simulated HIV vaccine trial. The primary trial is augmented by sexual partnerships in order to estimate the VE_I. We calibrate our simulations to a trial that is currently in the field and we show that both the VE_S and VE_I can be estimated with good precision. We show that this trial is sized to provide good power for rejecting the null hypothesis of no vaccine effect.

Contact Tracing in Stochastic and Deterministic Epidemic Models

Johannes Müller, M. Kretzschmar & K. Dietz

We consider a simple unstructured individual based stochastic epidemic model with contact tracing. Even in the onset of the epidemic, contact tracing implies that infected individuals do not act independently of each other. Nevertheless, it is possible to analyze the embedded non-stationary Galton-Watson process. Based upon this analysis, threshold theorems and also the probability for major outbreaks can be derived. Furthermore, it is possible to obtain a deterministic model that approximates the stochastic process, and in this way, to determine the prevalence of disease in the quasi-stationary state and to investigate the dynamics of the epidemic.

Uses of DNA fingerprinting in epidemiological research, using tuberculosis as an example

Nico JD Nagelkerke

DNA fingerprinting of tuberculosis bacteria is now practised widely. In the Netherlands all culturable TB strains have been fingerprinted since 1993. This allows various analyses to be done. In view of the (assumed) high mutation rate of the commonly used marker RFLP IS 6110 it is reasonable to assume that all individuals sharing the same fingerprint (a cluster) are linked by recent transmission. The problem is that the tree of infection/transmission cannot be inferred from the cluster information. Yet, several important properties can be derived from this cluster information. One is R_0, the other is age specific transmission patterns. Using the incidence of unclustered information as a starting point we can also assign probabilities to which individual in the cluster is its source, and thus information about transmission between (e.g.) nationalities.

Vaccination against Lyme disease: reducing costs for assessing vaccine efficacy using the correlation between titer and onset of disease

Albrecht Neiss

The efficacy E of a vaccine V in preventing a disease D is defined as $E = 1 - P(D \mid V)/P(D \mid nV)$, where nV stands for not vaccinated or vaccinated, but not with the vaccine V, or vaccinated with a different dose of V. In a prospective study information concerning $P(D \mid V)$ and $P(D \mid nV)$ is gathered by observing the $V-$ and the $nV-$ group for a previously defined time interval. If the occurence of D is rare and or the time interval is long the study may be very expensive due to the costs for subjects and for monitoring. So the question is obvious as to whether a surrogate variable exists which can help in lowering the costs of a study. The antibody titer measured after vaccination may be such a variable. For a new vaccine against Lyme disease I discuss the following questions: Is the titer a suitable measure to predict the disease? How can the relationship between titer and risk for the disease be modelled? When can this relationship be used to determine the efficacy? To answer this question data are used from a randomized study which was performed in the USA on about 10,000 subjects in 1995 and 1996.

Markov Chain Monte Carlo methods for analysing infectious disease data

Philip O'Neill

We consider the problem of performing statistical inference given data from infectious disease outbreaks using stochastic epidemic models. In general this is a non-standard problem due to difficulties including dependencies in the data and lack of complete data. Such difficulties can be overcome to an extent by making simplifying assumptions at the expense of model realism. Markov Chain Monte Carlo methods offer, at least in principle, a powerful way of making progress. In particular, such methods allow a large degree of modelling flexibility, are well-suited to missing data problems, and also allow a Bayesian approach to be used. We review a number of examples of models, data sets and MCMC algorithms, highlighting both the advantages and challenges of the approach.

Some problems in using sentinel surveillance data to estimate the global burden of HIV/AIDS: can multiple imputation help?

Andrew W. Roddam

There are many problems in attempting to use sentinel surveillance data from developing countries to estimate the global burden of HIV/AIDS infections. In the majority of countries only prevalence data from antenatal clinics (ANC) is available in the major urban and outside urban areas. To be able to use this data successfully we need to understand (or at least make some assumptions about) the male:female ratio, the fertile:non-fertile female ratio (and how HIV/AIDS might affect this ratio for HIV positive people), the age structure of the population, the ratio of the population in urban and outside urban areas, and the probability of attending an ANC given that a woman is pregnant. In addition most of the time-series from ANC, have many missing observations. Multiple imputation aims to produce a set of possible complete data sets using observed covariates to impute the missing values. As will be described in this talk, it is very difficult to produce such a set of covariates for imputing HIV/AIDS prevalence in these clinics and the use of multiple imputation can have little effect in determining the global burden of disease. Work is currently underway to observe additional covariates and investigate other methodologies for this missing data problem.

Infection Rate Estimation from Repeated Seroprevalence Surveys

Claudio Struchiner, R. Brunet & E. Massad

The methods used to estimate incidence of infection from data on current serological status require steady-state assumptions, so that the only time-like variable is age. When a single cross-sectional data are available, interpretation of the rate estimates thus obtained require the often made stationarity or time homogeneity assumptions which, in the epidemiological jargon, translates into abscence of cohort effects. We show that the time homogeneity assumption will no longer be necessary when repeated cross-sectional observations separated by a time interval are available. By tracking the prevalence of infection along life-lines, one can focus on the instantaneous relationships between prevalence and incidence rates and eliminate any need for knowledge of the history of the system. Thus, birth rates at any time and migration fluxes at previous times are of no concern.

Is it possible to understand mechanisms behind spread of infectious diseases from reported cases?

Åke Svensson

Many aspects of infectious diseases have to be known in order to understand how and why they occur. Most models include features that can not be directely observed. Statistical inference based on reported cases has to take into account that there may be a long way between events that are modelled (infections) and events that are observed (diagnosed and reported diseases). The cases observed have a different stochastic structure than the underlying infections. In general a "thinned" epidemic process is no longer an epidemic process. If this is not taken into account a statistical analysis may give severely biased results. However, some properties are more robust than others. This is true for temporal (durations) patterns and spatial patterns of the spread. Using some examples regarding influenza and campylobacter cases in Sweden I try to illustrate how space/time patterns of reported cases can be used to obtain information of the underlying infections.

A model for markers and latent disease progression

Mei-Ling Ting Lee & V. DeGruttola

In this paper we extend the bivariate Wiener process considered by Whitmore, Crowder and Lawless (1998) and model the joint process of a marker and disease progression. The disease process is assumed to be latent or unobservable. The time to reach the primary endpoint or failure (death, disease onset, etc.) is the time when the latent disease process first crosses a failure threshold level. Inferences for the model are based on two kinds of data: censored survival data and marker measurements. Covariates, such as treatment variables, risk factors and baseline conditions, are related to the model parameters through generalized regression functions. The model offers much richer potential for the study of treatment efficacy than conventional models. Treatment effects can be assessed in terms of their influence on both the failure threshold or the disease process parameters. We derive explicit formulas for the prediction of the onset of disease. Also we discuss model validation. This model does not require the proportional hazards assumption and hence can be widely used. We apply the methods in analyzing data from the protocol 116a of the AIDS Clinical Trials Group, a randomized double blind comparison of two doses of didanosine to zidovudine (ZDV) among patients.

Numerical and algorithmical considerations for the evaluation of epidemiologic parameters of pertussis infection

Stefan Wagenpfeil

We applied a dynamic grouped hazard rate model well-known in survival analysis with the extreme minimal value distribution as response function, in order to estimate the infectiousness function of pertussis index cases and vaccine efficacy out of a whooping cough household contact study simultaneously. Three different algorithmical approaches for jointly estimating the states and hyperparameters are compared: EM (robust with respect to the starting value, but slow convergence, yields only mode estimates), full posterior mode estimation with block-tridiagonal downhill arrowhead Hessian (local quadratic convergence speed for Newton method, can distinguish between local extrema and saddlepoints, yields, however, only mode estimates) and MCMC (full Bayesian analysis, but computationally very demanding, especially for bad starting values initializing the Gibbs sampler or Metropolis-Hastings). The resulting estimates of the states (or, equivalently,

latent variables) can be imputed in the response function to obtain estimates of the infectiousness function and to derive vaccine efficacy.

A two step procedure to estimate the population size of a general epidemic model

Paul S.F. Yip & Q. Chen

A two step procedure based on a conditional likelihood and a Horvitz-Thompson type estimator is proposed to estimate the initial number of susceptibles for a general epidemic model. Alternative estimator is also suggested when the infection process is not completely observed. Asymptotic properties of the estimator are examined. Simulation study is done to illustrate the performance of the proposed estimators.

Participants with e-mail addresses

Auranen, Kari	kari.auranen@rni.helsinki.fi
Awerbuch, Tamara	tamara@hsph.harvard.edu
Balding, David	d.j.balding@reading.ac.uk
Ball, Frank	fgb@maths.nott.ac.uk
Becher, Heiko	heiko.becher@urz.uni-heidelberg.de
Becker, Niels	n.becker@anu.edu.au
Boëlle, Pierre-Yves	boelle@b3e.jussieu.fr
Britton, Tom	tom.britton@math.uu.se
Dietz, Klaus	klaus.dietz@uni-tuebingen.de
Duerr, Hans-Peter	hans-peter.duerr@uni-tuebingen.de
Eerola, Mervi	mervi.eerola@rni.helsinki.fi
Eichner, Martin	martin.eichner@uni-tuebingen.de
Farrington, Paddy	c.p.farrington@open.ac.uk
Finkenstädt , Bärbel	b.f.finkenstadt@warwick.ac.uk
Halloran, M. Elizabeth	mehallo@sph.emory.edu
Ihorst, Gabriele	gabi@imbi.uni-freiburg.de
Isham, Valerie	valerie@stats.ucl.ac.uk
Jacob, Christine	Christine.Jacob@jouy.inra.fr
de Jong, Mart	m.dejong@id.dlo.nl
Keiding, Niels	N.Keiding@biostat.ku.dk
Knorr-Held, Leonhard	leo@stat.uni-muenchen.de
Koopman, Jim	jkoopman@umich.edu
Krämer, Alexander	kraemer@hrz.uni-bielefeld.de
Krickeberg, Klaus	krik@math-info.univ-paris5.fr
Longini, Ira	longini@sph.emory.edu
Mollison, Denis	denis@ma.hw.ac.uk
Müller, Johannes	johannes.mueller@uni-tuebingen.de
Nagelkerke, Nico	Nico.Nagelkerke@rivm.nl
Neiss, Albrecht	albrecht.neiss@imse.med.tu-muenchen.de
O'Neill, Philip	pdo@maths.nott.ac.uk
Richardson, Sylvia	richardson@vjf.inserm.fr
Roddam, Andrew	andrew.roddam@ceid.ox.ac.uk
Scalia-Tomba, Gianpaolo	gianpi@dmmm.uniroma1.it
Smith, Tom	smith@ubaclu.unibas.ch
Stilianakis, Nikolaos	stilianakis@imbe.med.uni-erlangen.de
Struchiner, Claudio	stru@malaria.procc.fiocruz.br
Svensson, Åke	akes@matematik.su.se
Ting-Lee, Mei-Ling	stmei@channing.harvard.edu
Utev, Sergey	sergey.utev@anu.edu.au
Wagenpfeil, Stefan	wagenpfeil@imse.med.tu-muenchen.de
Yip, Paul	sfpyip@hku.hk

Tagungsteilnehmer

Dr. Kari Auranen
Rolf Nevanlinna Institute
P.O. Box 4
Yliopistonkatu 5

SF-00014 Helsinki

Dr. Tamara Awerbuch
DANA-FARBER
Cancer Institute
44 Binney Street

Boston , MA 02115-6084
USA

Prof.Dr. David Balding
Department of Applied Statistics
The University of Reading
PO Box 240, Earley Gate

GB-Reading RG6 &FN

Prof.Dr. Frank G. Ball
Dept. of Mathematics
The University of Nottingham
University Park

GB-Nottingham , NG7 2RD

Prof.Dr. Heiko Becher
Institut für Biostatistik und
Epidemiologie
Universität Heidelberg
Im Neuenheimer Feld 324

69120 Heidelberg

Prof.Dr. Niels Becker
Dept. of Statistics
La Trobe University

Bundoora 3083
AUSTRALIA

Dr. Pierre-Yves Boelle
B3E Faculte de Medicine
Saint-Antoine
27, Rue Cahligny

F-75571 Paris Cedex 12

Dr. Tom Britton
Department of Mathematics
University of Uppsala
P.O. Box 480

S-75106 Uppsala

Dr. Mart C.M. de Jong
Institut for Animal Science and
Health, Dept. of Immunolgy,
Pathobiology and Epidemiology
P.O.Box 65

NL-8200 AB Lelystad

Prof.Dr. Klaus Dietz
Institut für Medizinische Biometrie
Universität Tübingen
Westbahnhofstraße 55

72070 Tübingen

Dr. Hans Peter Dürr
Institut für Medizinische Biometrie
Universität Tübingen
Westbahnhofstraße 55

72070 Tübingen

Prof.Dr. M. Elizabeth Halloran
Department of Biostatistics
Emory University

Atlanta GA 30322
USA

Dr. Mervi Eerola
National Public Health Institute
Rolf Neronlinna Institute
University of Helsinki
P.O.Box 4

SF-00014 Helsinki

Dr. Gabriele Ihorst
Institut für Medizinische Biometrie
und Medizinische Informatik
Klinikum der Universität
Stefan-Meier-Str. 26

79104 Freiburg

Dr. Martin Eichner
Institut für Medizinische Biometrie
Universität Tübingen
Westbahnhofstraße 55

72070 Tübingen

Prof.Dr. Valerie S. Isham
Dept. of Statistical Sciences
University College
Gower Street

GB-London WC1E 6BT

Dr. Conor Patrick Farrington
Department of Statistics
The Open University
Walton Hall

GB-Milton Keynes MK7 6AA

Christine Jacob
Laboratoire de Biometrie
Departement de Mathematiques
I.N.R.A.

F-78352 Jouy-en-Josas Cedex

Dr. Bärbel F. Finkenstädt
Department of Statistics
University of Warwick

GB-Coventry CV4 7AL

Prof.Dr. Niels Keiding
Department of Biostatistics
University of Copenhagen
Blegdamsvej 3

DK-2200 Kobenhavn N

Dr. Leonhard Knorr-Held
Institut für Statistik
Universität München
Ludwigstr. 33

80539 München

Prof.Dr. Ira M. Longini
Dept. of Statistics and Biometry
Emory University
Uppergate House
1462 Clifton Road

Atlanta , GA 30322
USA

Dr. Jim Koopman
Department of Epidemiology SPH-1
University of Michigan
109 Observatory Str.

Ann Arbor MI 48109
USA

Prof.Dr. Denis Mollison
Department of Actuarial Mathematics
and Statistics
Heriot-Watt University

GB-Edinburgh EH14 4AS

Prof.Dr. Alexander Krämer
Fak. für Gesundheitswissenschaften
Unviversität Bielefeld
Postfach 100131

33501 Bielefeld

Johannes Müller
Lehrstuhl für Biomathematik
Universität Tübingen
Auf der Morgenstelle 10

72076 Tübingen

Prof.Dr.Dr.h.c. Klaus Krickeberg
Le Chatelet

F-63270 Manglieu

Dr. Nico J.D. Nagelkerke
R. I. V. M.
Postbus 1

NL-3720 BA Bilthoven

Dr. Mei-Ling Ting Lee
Channing Laboratory
Harvard Medical School
181 Longwood Avenue

Boston , MA 02115
USA

Prof.Dr. Albrecht Neiss
Institut für medizinische Statistik
und Epidemiologie
Technische Universität München
Ismaninger Str. 22

81675 München

Dr. Philip D. O'Neill
Dept. of Mathematical Sciences
University of Liverpool
P. O. Box 147

GB-Liverpool L69 3BX

Dr. Nikolaos Stilianakis
Institut für Medizininformatik,
Biometrie and Epidemiologie
Universität Erlangen
Waldstraße 6

91054 Erlangen

Dr. Sylvia Richardson
Institut de la Sante et de la
Recherche Medicale
INSERM - U170
16, Ave. P. Vaillant-Coututier

F-94807 Villejuif Cedex

Dr. Claudio J. Struchiner
Escola Nacional de Saude Publica
Fundacao Oswaldo Cruz

Rio de Janeiro RJ 21041
BRAZIL

Prof.Dr. Andrew W. Roddam
Wellcome Trust Centre for the
Epidemiology of Infectious Disease
University of Oxford
South Parks Road

GB-Oxford OX1 3PS

Dr. Ake Svensson
Dept. of Mathematics
University of Stockholm
Box 6701

S-10691 Stockholm

Dr. Gianpaolo Scalia-Tomba
Dipartimento di Metodi e Modelli
Matematici per le Scienze Applicate
Universita di Roma "La Sapienza"
Via A. Scarpa 16

I-00161 Roma

Dr. Sergey Utev
National Centre for Epidemiology
and Population Health
The Austr. National University

Canberra , ACT 0200
AUSTRALIA

Dr. Tom Smith
Swiss Tropical Institute
Dept. of Public Health & Epidemiol.
Socinstrasse 57

CH-4002 Basel

Dr. Stefan Wagenpfeil
Institut für Medizinische Statistik
und Epidemiologie
Technische Universität
Ismaninger Str. 22

81675 München

Prof.Dr. Paul Yip
Department of Statistics
The University of Hong Kong
Pokfulam Rd.

Hong Kong
HONG KONG