Mathematisches Forschungsinstitut Oberwolfach

Report No. 48/2009

DOI: 10.4171/OWR/2009/48

Design and Analysis of Infectious Disease Studies

Organised by
Martin Eichner, Tübingen
Elizabeth Halloran, Seattle
Philip O'Neill, Nottingham

November 1st – November 7th, 2009

Abstract. This workshop gathered 45 participants from 16 countries and had a correspondingly multifaceted program covering various infectious diseases, public health applications, and methodological innovations. The discussions and presentations focused on the importance of mathematical models and statistical analyses in understanding the complex transmission systems of infectious diseases and in planning effective intervention strategies. Many different statistical and mathematical approaches were covered. The general unifying theme is that the analyses and models take into account the underlying transmission of the infectious agent among the hosts and/ or vector populations.

Mathematics Subject Classification (2000): 62-07, 62-03 (01A55), 62M05, 62N99, 62P10, 68U20, 91F99, 92N20, 92C60, 92D15, 92D25, 92D30, 92D40N.

Introduction by the Organisers

At the time of the workshop, the novel influenza A (H1N1) pandemic had passed through the southern hemisphere and was in the acute phase of its second wave in the northern hemisphere. Thus several presentations focused on mathematical and statistical methods for assessing the pandemic and planning interventions. Other infectious disease applications included pneumococcus, multiple drug resistant streptococcus, HIV, malaria, and citrus disease, among others.

In recent years network theory and graph theory have provided methodology for understanding the spread of infectious diseases and interventions. A full day of presentations was devoted to network and graph theory, including the role of serial or generation interval in conjunction with network models. Bayesian computation,

in particular Markov chain Monte Carlo (MCMC) methods are quite useful in analyzing infectious disease data where much of the underlying infectious and contact processes are unobserved and generally unobservable. A day was devoted to presentations of developments in these methods and applications.

Viral and bacterial phylodynamics and genomics are integral to infectious disease studies. Several presentations covered the development of statistical methods integrating genetic analysis with dynamic transmission systems. Other topics included statistical methods for analyzing vaccine studies, history of fitting epidemic models to data, and power analyses for hospital based studies of interventions.

A poster session on Monday evening gave participants not giving a talk the opportunity to discuss their current research with others. One evening two discussion groups were formed, one around the topic of the basic reproductive number, particularly within networks, the other around aspects of inference methods. Two early afternoon interactive tutorials were held for junior participants. One was on different aspects of vaccine efficacy and how they are measured. The other was an introduction to MCMC methods.

On Thursday evening, several participants provided a performance of classical music of piano, flute, and viola by Stravinsky, Donizetti, Debussy, Handel, and Mozart, and songs by Schumann and Schubert, followed by more contemporary songs with guitar accompaniment. Readings of poetry and a short story preceded a group song. The evening ended with everyone engaged in a Belgian dance.

Workshop: Design and Analysis of Infectious Disease Studies

Table of Contents

Abstracts

Models of vaccine efficacy for Streptococcus pneumoniae

Kari Auranen

Streptococcus pneumoniae (pneumococcus) is major cause of mortality and morbidity which is at least partly vaccine-preventable. The current pathway to license new pneumococcal conjugate vaccines against severe pneumococcal disease is based on immunological criteria. This approach may be sub-optimal in that use of immunogenicity at the current protective thresholds may not truly reflect the potential of a vaccine to protect. The aim of the PneumoCarr project is to optimize pneumococcal vaccines, their evaluation and their impact. The key to this is nasopharyngeal pneumococcal carriage that precedes disease and is the reservoir for spread of pneumococci between people. One of the particular aims of the project is to introduce statistical tools to estimate vaccine efficacy against pneumococcal carriage in vaccine trials.

There are several obstacles in establishing carriage as and endpoint in vaccine trials. First, there are 91 known pneumococcal serotypes. Between-type competition in the nasopharynx means that the biological efficacy on (e.g.) susceptibility may not be the same as the observed efficacy if such interaction is not taken into account. Second, carriage is common, unlike disease, so that any rare disease assumption cannot be readily made in proposing estimators for vaccine efficacy. Third, carriage is recurrent, even on the serotype level, which means that in principle longitudinal data are required. Finally, carriage is asymptomatic and is therefore only observed in its prevalent state.

For all these reasons, it is not straightforward to define and measure vaccine efficacy against pneumococcal carriage. In this work I review the current knowledge about the vaccine effect on pneumococccal carriage. I then propose an easy-to-use estimator for the (biological) efficacy, applicable in real-life trial settings. The approach is based on cross-sectional measurement of the relative reduction in susceptibility to pneumococcal carriage. It requires employment of a mechanistic, conditional model for the biological efficacy (relative reduction in the rate of acquisition) but is not confined to any particular type of model.

Epidemics on random networks incorporating household structure

Frank Ball
(joint work with David Sirl, Pieter Trapman)

We consider a stochastic SIR (susceptible $\rightarrow$ infective $\rightarrow$ removed) model for the spread of an epidemic amongst a population of individuals, with a random network of social contacts, that is also partitioned into households (Ball *et al.* [1, 2]). The behaviour of the model as the population size tends to infinity in an appropriate fashion is investigated. A threshold parameter which determines whether or not

an epidemic with few initial infectives can become established and lead to a major outbreak is obtained, as are the probability that a major outbreak occurs, the expected proportion of the population that are ultimately infected by a major outbreak and the distribution of the within-household final size in the event of a major outbreak. Monte Carlo simulations demonstrate that these asymptotic quantities accurately reflect the behaviour of finite populations, even for only moderately sized finite populations. The model is compared and contrasted with standard household and standard network models. The effect of the amount of clustering present in the overall population structure on the outcomes of the model is explored. Vaccination is also studied and an example demonstrates that an acquaintance-based vaccination strategy can outperform appreciably one that is household based.

References

[1] F. G. Ball, D. Sirl and P. Trapman, *Threshold behaviour and final outcome of an epidemic on a random network with household structure*, Advances in Applied Probability **41** (2009), 765–796.

[2] F. G. Ball, D. Sirl and P. Trapman, *Analysis of a stochastic SIR epidemic on a random network incorporating household structure*, under revision for Mathematical Biosciences.

Power analyses for cluster-randomized trials for infectious diseases in hospital settings

Martin Bootsma

Clinical studies in small (size=N) hospital units, e.g., intensive care units, with a high turnover of patient are common. Power analyses provide a rational criterion to assess how many patients should be included in study arms to have a pre-specified chance to detect a statistical significant difference between study arms. For infectious disease randomization at the patient level may not be the appropriate way to perform infectious disease studies as the indirect effect that prevention of colonization in one patient lowers the acquisition risk of other patients. Randomization at the unit level is needed to observe the full effect of an intervention. Here we determine how long study periods should be for infectious diseases with randomization at the unit level in case of perfect observation and compare the results with non-infectious diseases. We consider an SI-model with demographic turnover of which a fraction f of the patients is colonized on admission and constant (small) population size, i.e., there are $N + 1$ states distinguished by the number of colonized patients in the unit. We assume that the process is perfectly observed, i.e., we know the colonization status from each patient on admission and we know the exact moment of acquisition if acquisitions occurs. Suppose that in the time interval $(0, T)$, n acquisitions occurred at times $\{t_1, t_2, \ldots t_n\}$ and let t_j^- be the time just before the j^{th} acquisition occurred. The likelihood of these

acquisitions is proportional to:

$$L \propto e^{-\int_0^T \left(\alpha+\beta\frac{C(t)}{N}\right)(N-C(t))dt} \prod_{j=1}^{n} \left(\alpha + \beta\frac{C(t_j^-)}{N}\right)$$

For a non-infectious disease ($\beta = 0$), and the infectious case ($\alpha = 0$), the likelihood can be written as $L \propto \theta^n e^{-\tau\theta}$ with $\theta = \alpha$ and $\tau = \int_0^T (N - C(t))dt$ for non-infectious diseases and $\theta = \beta$ and $\tau = \int_0^T \frac{C(t)}{N}(N - C(t))dt$ for infectious diseases and the MLE for θ equals $\hat{\theta} = \frac{n}{\tau}$. The Fisher-information can be written as:

$$\mathcal{I}(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\log L(\theta)\right) = \mathbb{E}\left(\frac{n}{\theta^2}\right)$$

Asymptotically, the 95% confidence interval for the MLE for the mean prevalence ($p_\alpha(\alpha) = \frac{f+\alpha}{1+\alpha}$ for a non-infectious disease and $p_\beta(\beta) := \sum_{i=1}^{N} \frac{i}{N} p_s^i(\beta)$ with $p_s^i(\beta) = \frac{\left(\frac{f}{1-f}\right)^i \prod_{j=1}^{i}\left(\frac{N-j+1}{j}\right)\left(\frac{\beta(j-1)}{Nf}+1\right)}{1+\sum_{k=1}^{N}\left(\frac{f}{1-f}\right)^k \prod_{j=1}^{k}\left(\frac{N-j+1}{j}\right)\left(\frac{\beta(j-1)}{Nf}+1\right)}$ for an infectious disease) equals $\left(p - \frac{1.96}{\sqrt{NT(p(\theta)-f)}}p'(\theta)\theta, p + \frac{1.96}{\sqrt{NT(p(\theta)-f)}}p'(\theta)\theta\right)$ If we choose β and α for an infectious disease and a non-infectious disease such that $p_\alpha(\alpha) = p_\beta(\beta)$, $\left(\frac{p'_\beta(\beta)\beta}{p'_\alpha(\alpha)\alpha}\right)^2$ determines how much longer study periods for infectious diseases should be to have the same accuracy for the mean prevalence. Infectious diseases always require longer study periods and the ratio for clinically relevant parameters typically exceeds 3. Numerical simulations suggest that the asymptotic expression works fine unless the admission prevalence is low and the unit size is large. A simple method for the ratio with imperfect observations (admission and discharge data are known, but information on the colonization status is imperfect and depends on the scheme at which patients are tested for colonization) is an open problem.

Respondent driven sampling: a survey

Tom Britton

The traditional way of estimating a population mean or fraction is to take a random sample in the community and to take the corresponding sample mean/fraction as the estimate. Some populations are however not possible to take a random sample from, so-called "hidden" or "sensitive" populations (e.g. intraveneous drug-users). An alternative way is then Respondent Driven Sampling, in which a few seeds are selected. They report their response anonymously and also pass on the questionaire to some of their "friends" in the population of interest. The sample is hence obtained by randomly walking around in the population network. In the talk we present this method and how estimates can be obtained. We also illustrate why the method does not work when the network is directed (as is often the case).

Epidemics: the fitting of the first dynamic models to data

Klaus Dietz

Among the models with discrete time I concentrate on the chain-binomial models of Enko (1889), of Reed and Frost (1976) and of Greenwood (1931). The most important deterministic model with continuous time is proposed by Kermack and McKendrick (1927). Finally I consider the stochastic general epidemic model of McKendrick (1926). After a brief introduction of each of these models I provide examples of fitting them to observed epidemics and shall discuss the problems of interpreting the estimated parameters, especially if several models are fitted to the same data.

Individual heterogeneity: effects and estimation from multivariate serological survey data on directly transmitted infectious diseases

Conor Patrick Farrington

The estimation of contact patterns between individuals in a population is a central preoccupation of infectious disease modelling of directly transmitted infections. Much work has been done on estimating the effect of fixed covariates such as age, using data from serological surveys, contact surveys and other means. These reveal a strongly assortative age-related contact structure ([1] - [4]).

However, in contrast to sexually transmitted infections for which relevant individual heterogeneity is perhaps more easily defined and measured, relatively little work has been done on estimating the effects of individual heterogeneity on the transmission of close-contact and airborne infections ([5, 6]).

As shown in ([6]), information on individual heterogeneity in contact rates may be obtained from multivariate serological survey data, using frailty models and information on route of transmission. This basic frailty model can be extended to allow for age-variation in heterogeneity.

A simple framework in which such age-varying heterogeneities may be investigated will be presented and illustrated using multivariate serological survey data. Some ideas for future work in this area will be presented.

References

[1] M. N. Kanaan and C. P. Farrington, *Matrix models for childhood infections: a Bayesian approach with applications to rubella and mumps*, Epidemiology and Infection **133** (2005), 1009-1021.

[2] S. Y. Del Valle, J. M. Hyman, H. W. Hethcote and S. B. Eubank, *Mixing patterns between age groups in social networks*, Social Networks **29** (2007), 539-554.

[3] J. Mossong, N. Hens, M. Jit et al. *Social contacts and mixing patterns relevant to the spread of infectious diseases*, PLoS Medicine **5** (2008), e74 0381-0391.

[4] C. P. Farrington, H. J. Whitaker, J. Wallinga and P. Manfred *Measures of disassortativeness and their application to directly transmitted infections*, Biometrical Journal **51** (2009), 387 407.

[5] F. A. B. Coutinho, E. Massad, L. F. Lopez et al. *Modelling heterogeneities in individual frailties in epidemic models*, Mathematical and Computer Modelling **30** (1999), 97-115.

[6] C. P. Farrington, M. N. Kanaan and N. J. Gay *Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data*, Applied Statistics **50** (2001), 251-283.

Hybrid strategies for model selection and assessment in infectious disease studies

Gavin J. Gibson

The challenge of fitting stochastic dynamical models to imperfect observations is becoming increasingly tractable in a Bayesian frameworks and the recent literature includes numerous examples. Data augmentation, that is the imputation of unobserved information, x, from the course of an epidemic in addition to the observed information, y, is a key, relevant technique. In essence, if the extended data vector (x, y) specifies a likelihood for the model parameter vector θ which is analytically tractable, then by treating the imputed x as additional unknown parameters, it becomes feasible to explore the joint posterior density $\pi(\theta, x, |y)$ using standard Markov Chain Monte Carlo (MCMC) simulation techniques and to recover posterior inferences on θ through marginalisation.

However, the problem of testing the adequacy of models fitted using this approach, or selecting between competing models, is not straightforward. Extending the Bayesian approach, by embedding competing models within an expanded model with model index as a further parameter, can be problematic. Reasons for this include computational complexity, difficulty in designing suitable Markov chains samplers, and the sensitivity of posterior model inferences to the prior densities assigned to the parameters of the respective models. This talk explores alternative approaches to assessing model adequacy that operate by applying classical tests to the extended data vector (y, x) and by investigating the posterior distribution of the associated p-value. Through this hybrid approach, which embeds classical analyses within a Bayesian framework, it is possible, via the posterior distribution of the p-value, to elicit evidence of model inadequacy. Moreover, if the embedded tests are selected to be likelihood ratio tests, then the approach offers the potential to assess one model in direct comparison to a competing formulation.

The approach is illustrated in two scenarios. In the first of these (Streftaris and Gibson, *Proc. Roy. Soc. B* **271**, 1111–1117, 2004) the imputations chains of infection in experimental populations of sheep infected with FMD virus is used to investigate a hypothesis regarding the variation of infectivity with depth in the chain, by investigating the posterior distribution of the p-value arising in an ANOVA test applied to population partitioned by depth in the chain. In the second example (Gibson et al., *Statistics and Computing* **16**, 391-402, 2006) spatio-temporal models for the spread of a fungal pathogen of radish in experimental populations are assessed by imputing exponentially infection thresholds for each individual (with sampling distribution independent of θ). The imputed thresholds are then assessed using a Kolmogorov-Smirnov test for consistency with a random sample from the Exp(1) distribution. For this latter example it is shown that the

analysis shows clear evidence of lack of fit, whereas simpler posterior checks based on the time-course of the number of infections do not reveal lack of fit.

Current work is focussing on applying the techniques to compare spatio-temporal models for arboreal pathogens in larger-scale studies, and on the theoretical understanding of how the extent of imputation affects the sensitivity of the tests.

Influenzanet

Gabriela Gomes

Influenzanet is an internet-based system that monitors influenza-like illness (ILI) in cohorts of self-reporting volunteers. Data collected over six years are presented, analysed and used to parameterize mathematical models of influenza transmission. The system warrants consistency in epidemic monitoring across countries and seasons and is currently tracing the H1N1 pandemic. The recorded epidemic curves were reproduced by models that include time-dependent transmission and host heterogeneities. In addition to previous modelling studies, we detect significant rates of influenza reinfection during the epidemic decay phase.

Reconstructing transmission trees from partially observed epidemic trees: an illustration of H1N1

Niel Hens
(joint work with Jacco Wallinga)

Emerging epidemic outbreaks of infectious diseases are often reported as a transmission tree. In a transmission tree all cases are represented, and when infection is transmitted from one case to another, a link (arrow) is drawn between the two cases to indicate that they form a transmission pair. The advantage of such a tree is that key variables, such as the reproduction number and the generation interval, are easily inferred. However, such a transmission tree ignores that some transmission pairs may be missing, and that some transmission pairs may be misclassified (that is, they actually reflect two cases that did not infect one another). Here, we point out that the missingness and misclassification can result in biased estimates of both the reproduction number and generation interval. We present statistical methods for dealing with missing transmission pairs and for identifying misclassified, unlikely, transmission pairs; these methods are adapted from missing data techniques (the so-called EM algorithm) and from case deletion measures (so-called global influence measures). We apply the methods to a reported outbreak of pandemic influenza A/H1N1v, and we show that the generation interval has to be corrected from 2.43 to 2.20 days.

Structured populations: Epidemics on finite networks

Valerie S. Isham

The General Epidemic (SIR) model is the fundamental model for epidemics in a homogeneously-mixing population. In this talk, we consider the spread of infection or information through a structured population, combining an extension of an SIR epidemic model with a random network to represent population structure. The envisaged application is to the spread of infection/information on social networks. The talk will discuss a) the effect of the population size on thresholds for epidemic/rumour spread; b) the effect of different network structures; c) the adequacy of approximations to the final size distribution.

Survival analysis of epidemic data via infectious contact intervals

Eben Kenah

We argue that the time from infection to infectious contact, which we call the *infectious contact interval*, is a better basis for inference in epidemic data than the generation or serial interval. Since only the first infectious contact with a given susceptible leads to infection, many infectious contact intervals are right-censored and survival analysis is the natural approach to estimation. We derive a likelihood for stochastic SIR models in close-contact groups. For fully-mixed models, we obtain an asymptotic likelihood that requires data only on infected persons, and the estimated infectious contact interval distribution provides a description of the time course of infectiousness and a novel estimator of R_0. This approach avoids some of the restrictive assumptions necessary for an analysis based on generation or serial intervals. It also points to the usefulness of data on the duration of infection and uninfected contacts of cases. An important extension of this work will be the analysis of partially-observed epidemics, for which Bayesian MCMC is a promising tool.

Statistical Analysis of Hospital Infection Data: Models, Inference and Model Choice

Theodore Kypraios

High-profile hospital "superbugs" such as methicillin-resistant Staphylococcus aureus (MRSA) etc have a major impact on healthcare within the UK and elsewhere. Despite enormous research attention, many basic questions concerning the spread of such pathogens remain unanswered. For instance what value do specific control measures such as isolation have? how the spread in the ward is related to "colonisation pressure"? what role do the antibiotics play? how useful it is to have new molecular rapid tests instead of conventional culture-based swab tests? A wide range of biologically-meaningful stochastic transmission models that overcome unrealistic assumptions of methods which have been previously used in the

literature are constructed, in order to address specific scientific hypotheses of interest using detailed data from hospital studies. Efficient Markov Chain Monte Carlo (MCMC) algorithms are developed to draw Bayesian inference for the parameters which govern transmission. The extent to which the data support specific scientific hypotheses is investigated by considering and comparing different models under a Bayesian framework by employing a trans-dimensional MCMC algorithm while a method of matching the within-model prior distributions is discussed how to avoid miscalculation of the Bayes Factors. Finally, the methodology is illustrated by analysing real data which were obtained from a hospital in Boston.

Dynamic random networks

Mathias Lindholm
(joint work with Tom Britton)

A simple Markovian time dynamic random network model is considered where nodes as well as edges are created and removed continously in time. Moreover, when a node is created it is assigned a random social index which affects its possibility of creating edges and which may affect the rate at which edges are attached to it. Some fundamental model properties are analysed and discussed. In particular it is shown that the asymptotic degree distribution is of mixed Poisson type.

Analysis of Influenza Vaccines and Vaccination Strategies: Statistics and Models

Ira M. Longini

In this talk, I give a summary on the large body of work on pandemic influenza A (H1N1) transmission and control. I first describe the pandemic in the US. I then give the statistical estimates of the key parameters that govern transmission and severity, including R0, the generation interval, and the case fatality ratio. After this, I describe estimates of influenza vaccine efficacy in general and then the estimates of pandemic H1N1 vaccine efficacy based on immunogenicity from early phases I and II vaccine trials. I then show modeling results for the mitigation of the pandemic in the US based on the current statistics for the US vaccine supply over time. I show that the later the epidemic peak, the more effective current planned vaccine supply will be on reducing cases. If epidemic peak is late October, the current vaccination strategy will have little effect on attack rates but could prevent many H1N1-related deaths. Vaccinating children early will reduce the attack rate, mortality, and hospitalizations, although it is important to vaccinate the essential workforce and high-risk individuals first. If the epidemic peaks early, non pharmaceutical interventions may be a necessary component of the response.

The dynamics of seasonal influenza

Mick Roberts

The single strain model. Assume that the influenza season starts at time $t = 0$. We measure time in units of the mean infectious period, and neglect demographic changes (births, deaths, immigration, etc.) within a season. Let the proportion of the population infected at time t be $y(t)$, and the proportion that has been infected be $z(t)$. An epidemic is described by:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \mathcal{R}_0 (1 - z) y - y$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = \mathcal{R}_0 y (1 - z)$$

with $0 < y(0) \ll 1$ and $z(0) = z_0$. The proportion of the population infected over the course of the epidemic is $P = z_\infty - z_0$, where

$$z_\infty = \lim_{t\to\infty} z(t) = 1 + \frac{1}{\mathcal{R}_0} \mathcal{W}\left(-\mathcal{R}_0 (1 - z_0) e^{-\mathcal{R}_0 (1 - z_0)}\right)$$

The function $\mathcal{W}$ is the Lambert W-function [4], defined by $y = \mathcal{W}(x)$ if $ye^y = x$ for $x \geqslant -\frac{1}{e}$. In-between seasons a proportion b of the population is replaced with newborn susceptibles, and the protection from infection is reduced due to antigenic drift. Hence the initial condition at the start of the next season is found from $z_0 = pz_\infty = (1 - b) dz_\infty$ where $d < 1$. We define an inter-season map which relates the value of z at the beginning of one season to the value at the beginning of the following season. To do this it is effective to transform the variables to effective reproduction numbers at the beginning of each season, setting $\nu_n = \mathcal{R}_0 (1 - z_0)$. The between-season map is

$$\nu_{n+1} = (1 - p) \mathcal{R}_0 + p\Phi (\nu_n)$$

where $\Phi (\nu_n) = -\mathcal{W} (-\nu_n e^{-\nu_n})$. This map has a unique fixed point, which is stable [1,5]. It is illustrated in Figure 1A, where it is seen that for $\nu_n < 1$ no epidemic takes place, and $\nu_{n+1} > \nu_n$ due to host replacement and antigenic drift; and for $\nu_n > 1$ an epidemic occurs.

The two strain model. Since 1978 two subtypes of influenza A, H1N1 and H3N2, have been represented in seasonal epidemics [2,3]. In our model we assume that exposure to influenza (any subtype) this season may provide non-specific immunity to infection for this season only, exposure to a strain of the same subtype in a previous season provides a degree of protection this season, and previous exposure to the same strain this season provides complete protection. The dynamics of the proportions of the population infected with subtype S, and the proportions immune due to previous exposure to subtype S, are described by the equations:

$$\frac{\mathrm{d}y_S}{\mathrm{d}t} = \mathcal{R}_0^S x_S y_S - y_S \tag{1}$$
$$\frac{\mathrm{d}z_S}{\mathrm{d}t} = \mathcal{R}_0^S y_S (1 - z_S)$$

where $S = 1, 2$. The function $x_S(t)$ is the susceptibility of the population to the predominant strain of subtype S this season.

Let the proportion of the population that has a degree of immunity to subtype S due to exposure to either subtype be u_S, where

$$\frac{\mathrm{d}u_S}{\mathrm{d}t} = \left(\mathcal{R}_0^1 y_1 + \mathcal{R}_0^2 y_2\right)(1 - u_S) \qquad u_S(0) = z_S(0) \tag{2}$$

If the degree of non-specific protection between subtypes is q, with $q = 0$ for no cross-protection and $q = 1$ for complete cross-protection, we have

$$x_S(t) = (1 - q)(1 - z_S) + q(1 - u_S)$$

It follows from equations (1&2) that

$$1 - u_S(t) = \frac{(1 - z_S(t))\left(1 - z_{\hat{S}}(t)\right)}{1 - z_{\hat{S}}(0)}$$

where $\hat{S}$ signifies the alternative subtype. Hence the dynamics are specified by the four-dimensional system (1) with

$$\frac{x_S(t)}{1 - z_S(t)} = 1 - q + q\frac{1 - z_{\hat{S}}(t)}{1 - z_{\hat{S}}(0)}$$

The equations can be solved numerically for a representation of the within-season epidemics of the two strains. It is necessary to specify $z_S(0)$ for $S = 1, 2$, and initial values of $y_S(0)$, with $0 < y_S(0) \ll 1$.

With a slight abuse of notation, let the proportion of the population infected with subtype S over the course of the epidemic be $P_S = z_S(\infty) - z_S(0)$, which may be estimated from equations (1) by noting that

$$\mathcal{R}_0^S \int_{z_S(0)}^{z_S(\infty)} \frac{x_S(t)}{1 - z_S(t)}\,\mathrm{d}z_S(t) + \log\left(\frac{1 - z_S(\infty)}{1 - z_S(0)}\right) = 0$$

We therefore have a final size equation

$$\mathcal{R}_0^S M_S P_S + \log\left(1 - \frac{P_S}{1 - z_S(0)}\right) = 0$$

where

$$M_S = \frac{1}{P_S}\int_{z_S(0)}^{z_S(\infty)} \frac{x_S(t)}{1 - z_S(t)}\,\mathrm{d}z_S(t) \tag{3}$$

The integrand in equation (3) is non-increasing, and we can use the mean value theorem to approximate

$$M_S = \frac{x_S(\bar{t})}{1 - z_S(\bar{t})} = 1 - q + q\frac{1 - z_{\hat{S}}(\bar{t})}{1 - z_{\hat{S}}(0)}$$

for some $\bar{t} > 0$. If there is no hetero-subtype protection, $q = 0$ and $M_S = 1$.

In the same way as for the single-strain model, the initial conditions at the start of the next season are found from $z_S(0) = p_S z_S(\infty) = (1 - b)\,d_S z_S(\infty)$

where $d_S < 1$. Now set $\xi_n = \mathcal{R}_0^1 (1 - z_1(0))$ and $\eta_n = \mathcal{R}_0^2 (1 - z_2(0))$ to obtain a two-dimensional nonlinear between-season map defined by

$$\xi_{n+1} = (1 - p_1)\mathcal{R}_0^1 + \frac{p_1}{M_1}\Phi(M_1\xi_n)$$
$$\eta_{n+1} = (1 - p_2)\mathcal{R}_0^2 + \frac{p_2}{M_2}\Phi(M_2\eta_n) \tag{4}$$

This is illustrated in Figure 1B,where the value of ξ_n is varied from 0 to $\mathcal{R}_0^1$, but η_n is kept fixed. Three distinct types of behaviour can be seen. For ξ_n below the first threshold (left-hand shaded region) there is no epidemic of H1N1, but there is an epidemic of H3N2. For ξ_n above the second threshold (right-hand shaded region) there is an epidemic of H1N1, but not of H3N2. In the intermediate (unshaded) region there is an epidemic of both subtypes. The value of η_n was chosen to correspond to the fixed point, which is seen to be unstable. Numerical results have confirmed periodic solutions, with the period depending on the assumed parameter values [5].

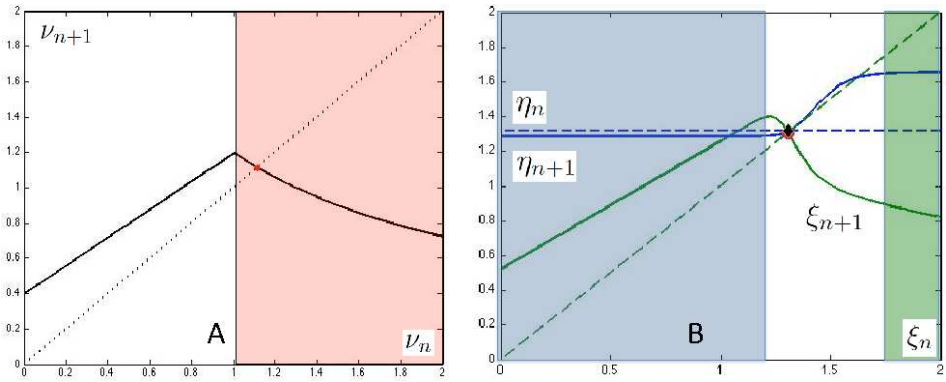

FIGURE 1. A: The map $\nu_n \to \nu_{n+1}$ with the fixed point ν^* shown in red, for $\mathcal{R}_0 = 2.0$ and $p = 0.8$. B: The map $(\xi_n, \eta_n) \to (\xi_{n+1}, \eta_{n+1})$ with $\eta_n = \eta^*$ (green broken line and black diamond); ξ_{n+1} and η_{n+1} are shown as blue and green lines respectively, and the fixed point ξ^* as a red dot. Parameter values are $p_1 = 0.74$, $p_2 = 0.55$ and $\mathcal{R}_0^1 = \mathcal{R}_0^2 = 2.0$.

REFERENCES

[1] Andreasen, V. 2003 Dynamics of annual influenza A epidemics with immuno-selection. *J. Math. Biol.* 46, 504-536.

[2] Finkelman, B. S., Viboud, C., Koelle, K., Ferrari, M. J., Bharti, N. & Grenfell, B. T. 2007 Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: viral coexistence and latitudinal gradients. *PLoS ONE* 2, e1296.

[3] Huang, Q. S., Lopez, L. D., McCallum, L. & Adlam, B. 2008 Influenza surveillance and immunisation in New Zealand, 1997-2006. *Inf. Oth. Resp. Vir.* 2, 139-145.

[4] Ma, J. & Earn, D. J. 2006 Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bull. Math. Biol.* 68, 679-702.
[5] Roberts, M. G. The dynamics of seasonal influenza in New Zealand, *in prep.*

What can the historic record tell us about modern infectious disease epidemics?

Lisa Sattenspiel

The historic record contains a wealth of information on infectious diseases, and it is a resource that is relatively untapped by epidemic modelers. Historical data can provide detailed data sources that can be used in conjunction with models to explore questions about the spread of infectious diseases and, in addition, they can provide more qualitative information about a variety of social and biological factors that influence disease transmission patterns. Specific kinds of information the historic record can provide include aggregate and individual-level data on mortality and morbidity, census and vital statistics data, sources of information on the responses of governments and the public to epidemics (e.g., newspaper articles or government correspondence), and sources of information on other situations that might be relevant to the epidemic experience, such as general nutritional levels, adequacy of health care, or other health conditions that might influence outcomes during an epidemic.

A number of questions about infectious disease transmission and spread have been addressed using epidemic models and historic data. These include major insights that have been derived from the numerous studies of the cycling of measles in England and Wales prior to the availability of vaccination (see [1] and [2] for recent reviews of these studies), the value of historic data for assessing models of the long-term demographic impact of infectious disease epidemics, how historic epidemics such as the 1918-19 flu epidemic can aid in understanding the extent to which epidemic patterns may vary across space, and the potential of historic data to identify and help understand possible interactions among different pathogens, illustrated with data from a 1916-17 measles epidemic and the 1918-19 flu epidemic on the island of Newfoundland.

It is important to keep the limitations of historic data in mind. however. For example, diagnostic criteria may change over time, which happened in the 1980s once AIDS began to be better understood, or records may be incomplete or inaccurate. There may also be incomplete understanding of the relevant biology of microorganisms, as occurred during the 1918-19 flu epidemic when many scientists thought the underlying cause was bacterial rather than a virus. Ideally, for infectious disease models, one would like to know about both mortality and illness (morbidity), but the latter data are very rare. The collection of historic data is also very time consuming. Nonetheless, much has been learned by analyzing epidemics of the past and given the amount of existing data and the variety of other kinds of information, much remains to be learned in the future.

References

[1] Sattenspiel, Lisa (with contributions from Alun Lloyd) (2009). *The Geographic Spread of Infectious Diseases: Models and Applications*. Princeton University Press, Princeton, NJ.

[2] Lloyd, Alun L. and Sattenspiel, Lisa (2009). Spatiotemporal dynamics of measles: Synchrony and persistence in a disease metapopulation. In *Spatial Ecology*, Stephen Cantrell, Chris Cosner, and Shigui Ruan (eds). CRC Press, Boca Raton, FL, pp. 251-272.

Observing generation times, in theory

Gianpaolo Scalia Tomba

(joint work with Ake Svensson, Tommi Asikainen, Johan Giesecke)

The aim of this work is to study the effects of various observation schemes on the generation time distribution during an epidemic outbreak and to propose some statistical methods to achieve unbiased estimation of various aspects of the generation time distribution.

There are two main variants of observation scheme: forward observation, starting from a given infected person and observing his secondary cases, if any, and the corresponding generation times and backward observation, starting from an infected person, his infector is identified and the corresponding generation time observed.

These two approaches have been implemented on the same simulation of a simple Markov SIR epidemic with infectious intensity $\beta = 2$, average infectious period $1/\mu = 1$, i.e. $R_0 = 2$, in a population of 10000 individuals, with 1 initial infective. As can be seen from the two graphs, the behaviour of forward and backward observations is quite different.

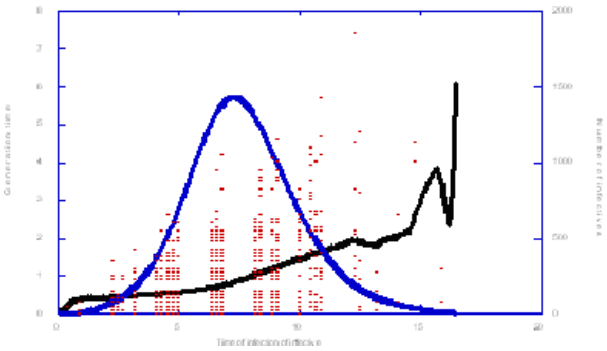

Figure 1. Local average of generation times plotted at time of infectee (backward observation)

In homogeneous models, the distribution seen at a given time, if backward sampling is assumed, is the age of infection of potential infectors at that time and that distribution lenghtens as the epidemic progresses. During the initial exponential phase, the observed GT distribution is not g(a) but $R_0 e^{-ra} g(a)$...In

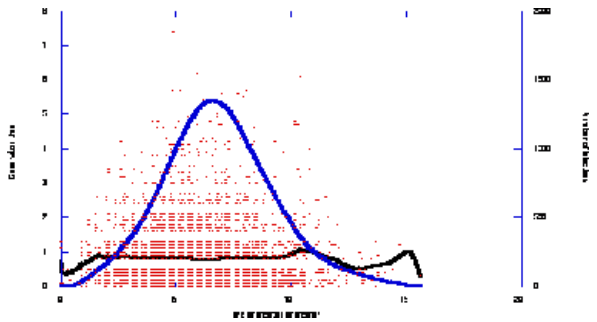

FIGURE 2. Local average of generation times plotted at time of infector (forward observation)

our simulation, the average is 1/2 in the beginning and then grows to about 2. The forward procedure is the one corresponding to the "standard GT distribution", but the values of generation times become shorter the faster susceptibles decline, thus the standard distribution can only be observed at the start and end of an epidemic, in theory, when numbers of susceptibles are relatively stable. Theoretical analysis suggests that the average should be 1, but it is slightly lower near the peak of the epidemic.

If more features, such as latent periods, different infectious period distributions and variable infectivity, are incorporated in the spread model, the phenomena described above always remain to some extent (except in the Reed-frost model...). It is thus complicated to get a "nice" sample of generation times in order to infer about the GT distribution,but one can try to use knowledge about the statistical properties of the system to one's advantage... E.g., in the early exponential phase of an epidemic, with an estimate of r and backward tracing, exp(rT) is approximately unbiased for R_0... Åke Svensson has several ideas for non parametric, martingale based estimators... One may also resort to parametric inference, see e.g. Cauchemez on household data, for estimation of infectious period and infection intensity separately...

Exploration of the Dynamics of Network Topologies during Epidemic Outbreaks

MARKUS SCHWEHM

Looking at the epidemic graphs of the currently ongoing influenza pandemic, a clear pattern is visible: most outbreaks have a very steep ascent while petering out slowly. The graphs are right-skewed. The SEIR-Type models used to model influenza can produce right-skewed graphs, but only for large $R_0 > 4$. The current estimates for R_0 are around 2.0 and most researchers have revised their estimates to much lower values around 1.5. Standard SEIR models with $R_0 < 2$ produce

almost symmetric epidemic graphs and therefore fail to explain the current influenza pandemic. In this contribution I have explored a network-based modelling approach. Networks with heterogeneous degree distribution like the scale-free networks can generate right-skewed epidemic curves even for very low R_0. However, the scale-free network fails to explain another characteristic of the ongoing pandemic. After a first wave of the pandemic, all of the high-degree individuals (the super-spreaders) have become infected and are immune and it is not possible to observe a second and third wave of the epidemic, as currently observed in Mexico. A simple modification of the scale-free network solves this problem. If the long tail of the degree distribution of a scale-free network is cut off, the resulting truncated scale-free network still produces right-skewed epidemic graphs, but the high-degree individuals are not removed in their entirety and a second and third epidemic wave becomes possible. Moreover the outbreak size distribution matches the observed outbreak sizes in Switzerland this summer. While there may be other networks possible that can explain the right-skewed epidemic curves of the current pandemic, at least we have a model that can explain the observed outbreak data. The analysis of the shape of epidemic graphs provides more information than just R_0 and generation time.

Genetic control of vector-borne diseases - Artificial selection and heterogeneity of the immune response

Claudio J. Struchiner
(joint work with Eduardo Massad, J.M.C. Ribeiro)

The malaria model developed in association with the Garki Project in Nigeria by Dietz, Molineaux and Thomas (DMT) explicitly addressed the implications of the human immune response on the transmission dynamics of this mosquito-borne disease. Although this contribution was originally conceived for discussing intervention strategies based on the use, alone or in combination, of house-spraying with propoxur (a residual insecticide) and universal distribution of anti-malarial drugs (sulfalene and pyrimethamine), the presence of this human immune response component also made the model suitable as a starting point for discussing the development of the various malaria vaccines made possible by the new molecular biology paradigm that became widespread in the 80's. Expansions to the Garki's DMT model have shown to be useful to uncover the complex implications of intervention programs against mosquito-borne diseases by focusing on the changes in immune profile of the target human population and the evolution of pathogen virulence and resistance as a result of selection pressures imposed by a vaccine or drug.

The original paradigm described above addresses within-vertebrate host stages of the parasite's life history and was motivated by the possibility of developing vaccines that elicit stage-specific immune responses in humans. Recent advances in molecular genetics and mosquito ecology motivates the expansion of the original paradigm to also encompass those stages of the pathogen that take place in the

vector. These genetic methods for controlling vector transmissions are designed to reduce or eliminate vector populations, to selectively kill only those vectors infected by the pathogen, or to modify (replace) natural vector populations by introgressing genes that eliminate vector competence. However, as became evident from the Garki data since the 70's, the genetic diversity of traits that modulate vector competence posed an important challenge to control programs based on domiciliary spraying of residual insecticides as evidenced by the degree of exophily among the population of individual vectors members of the gambiae complex. More recently, the diversity of the immune response exhibited by vectors, i.e., the means whereby they are able to kill invading pathogens, has been well established by the availability of the genome sequences of vectors, hosts and parasites that enabled genome wide comparative studies. Those advances provide new tools to monitor diversity among the three players, pathogens, vertebrate and invertebrate hosts. In particular, important issues, such as parasite virulence and resistance, are not fixed properties of infection but are affected by the genetic diversity of the players involved, and the environmental conditions under which those players interact.

An expanded paradigm that accounts for the vector-pathogen systems explicitly can contribute to avoid a large underestimation of the pathogen polymorphism as well as polymorphisms of traits that modulate the invertebrate host competence. These two processes together contribute to genetic drift and selection since the higher the pathogen diversity within a host, the greater is the expected genetic change between the original host and the load delivered to the next host. Within-vector competition adds one more level of selective differences affecting the local pathogen diversity since resistance genotypes in vectors could select against some pathogen genotypes more than others.

By using population-genetic models, we explore the evolutionary epidemiologic dimension of genetic strategies currently being proposed to control mosquito-borne diseases. In our three-players system, reciprocal selection pressures determine the distribution of each player's polymorphism. In our work, we focus on the forces that determine the current distribution of traits associated with dimensions of epidemiologic importance such as vertebrate and invertebrate immune response, and pathogen virulence and resistance to drugs. By explicitly describing the forces that determine the current distribution of traits, we hope to add to the discussion of gene drive mechanisms and their relative chances of success in changing the current distribution of traits by the introgression of new genes into the vector population. Finally, we validate our models against the empirical information collected so far from the genome wide studies comparing vectors of medical importance as well as other insects.

Phylodynamics of Infectious Disease Epidemics

Erik M. Volz

Background: Virus phylogenies have been used in many recent studies to infer properties of epidemics. However, these approaches rely on coalescent models that may not be appropriate for infectious diseases. Our aim was to develop a coalescent theory for viral populations whose dynamics fit standard epidemiological models. Our approach allows us to describe the cluster size distribution within a sample or the entire population, as well as predict the patterns of clustering among acutely and chronically infected individuals. Approximations are also derived for epidemics in heterogeneous populations and contact networks, which allows us to determine the effect of network structure on phylogenetic clustering.

Methods: Our coalescent theory uses a standard model from mathematical epidemiology to reproduce epidemic incidence and prevalence. We then calculate the rate of coalescence for a sample taken from the population at any time. Our model was applied to a sample of HIV-1 sequences. The model was fit to the size distribution of clusters by deriving a likelihood function for epidemic parameters, and best-estimates of epidemic parameters were found by Bayesian importance sampling. A 'relaxed clock' approach was used to estimate coalescence times. We checked our results and the efficiency of our fitting algorithm using simulated data, generated by individual-based stochastic models.

Results: We verified by simulation that our model accurately reproduces the cluster size distribution in commonly-used SI, SIR, and SIS epidemic models. When compared to the skyline plot of effective population size from the ACTG data, our best-fit SIR model reproduces consistent growth rates in the exponential phase and a tapering of epidemic prevalence in the early nineties. The estimates of epidemic parameters are also consistent with the known etiology of HIV. When applied to heterogeneous populations, our model predicts that clustering of acute infecteds is largely dependent on host population structure, and does not directly indicate higher transmission probabilities in the acute stage.

Conclusions: The relationship between the effective population size and the absolute number of infected individuals is complex, depending on the details of the epidemiological model. Our mathematical model successfully characterizes aspects of this complexity and explains observed clustering of acute HIV cases. Work is underway to quantify the effect of non-exponential distributions of infectious periods, variability in infectiousness over time, and heterogeneous contact rates.

Modeling Competing Infectious Pathogens from a Bayesian Perspective: with Application to Influenza Studies with Incomplete Laboratory Results

Yang Yang

(joint work with M. Elizabeth Halloran, Michael J. Daniels, Ira M. Longini)

In seasonal influenza epidemics, pathogens such as respiratory syncytial virus (RSV) are often found co-circulating with influenza and cause influenza-like illness (ILI) in human hosts. However, it is often impractical to test for each potential pathogen or to collect specimens for each observed ILI episode, making inference about influenza transmission difficult. In the setting of infectious diseases, missing outcomes impose a particular challenge because of the dependence among individuals. We propose a Bayesian competing-risk model for multiple co-circulating pathogens for inference on transmissibility and intervention efficacies under the assumption that missingness in the biological confirmation of the pathogen is ignorable. Simulation studies indicate a reasonable performance of the proposed model even if the number of potential pathogens is misspecified, and show that a moderate amount of missing laboratory test results has only a small impact on inference about key parameters in the setting of close contact groups. Using the proposed model, we found that a non-pharmaceutical intervention is marginally protective against transmission of influenza A in a recent study conducted in elementary schools.

Reporter: Martin Eichner

Participants

Dr. Kari Auranen
National Institute for Health & Welfare
Mannerheimintie 166
FIN-00300 Helsinki

Prof. Dr. Frank G. Ball
School of Mathematical Sciences
The University of Nottingham
University Park
GB-Nottingham NG7 2RD

Dr. Martin Bootsma
Department of Mathematics
Utrecht University
P.O.Box 80.010
NL-3508 TA Utrecht

Dr. Michiel van Boven
Centre for Infectious Disease Control
National Institute for Public Health and the Environment
PO Box 1
NL-3720 BA Bilthoven

Dr. Tom Britton
Department of Mathematics
Stockholm University
S-10691 Stockholm

Prof. Dr. Klaus Dietz
Institut für Medizinische Biometrie
Universität Tübingen
Westbahnhofstr. 55
72070 Tübingen

Dr. Hans-Peter Dürr
Institut für Medizinische Biometrie
Universität Tübingen
Westbahnhofstr. 55
72070 Tübingen

Prof. Dr. Martin Eichner
Institut für Medizinische Biometrie
Universität Tübingen
Westbahnhofstr. 55
72070 Tübingen

Dr. Conor Patrick Farrington
Department of Mathematics & Statistics
The Open University
Walton Hall
GB-Milton Keynes MK7 6AA

Ashley Ford
Department of Statistics
University of Warwick
GB-Coventry CV4 7AL

Prof. Dr. Gavin Gibson
Department of Actuarial Mathematics and Statistics
Heriot-Watt University
Riccarton
GB-Edinburgh EH14 4AS

Nele Goeyvaerts
Center for Statistics
Hasselt University
Agoralaan Building D
B-3590 Diepenbeek

Dr. Gabriela M. Gomes
Instituto Gulbenkian de Ciencia
Apartado 14
P-Oeiras 2781-901

Prof. Dr. M. Elizabeth Halloran
Department of Biostatics
University of Washington and
Fred Hutchinson Cancer Research Center
1100 Fairview Ave. N, LE-400
Seattle , WA 98109-1024
USA

Dr. Niel Hens
Center for Statistics
Hasselt University
Agoralaan Building D
B-3590 Diepenbeek

Prof. Dr. Valerie S. Isham
Dept. of Statistical Science
University College London
Gower Street
GB-London WC1E 6BT

Dr. Eben E. Kenah
Division of Public Health Science
Fred Hutchinson Cancer Research Center
The University of Washington
1100 Fairview Ave. North, LE-400
Seattle WA 98109-1024
USA

Dr. Jim Koopman
Department of Epidemiology SPH-1
University of Michigan
109 Observatory St.
Ann Arbor MI 48109
USA

Dr. Mirjam Kretzschmar
Centre for Infectious Disease Control
RIVM
Antonie van Leeuwenhoeklaan 9
P.O.Box 1
NL-3720 BA Bilthoven

Dr. Theodore Kypraios
School of Mathematical Sciences
The University of Nottingham
University Park
GB-Nottingham NG7 2RD

An Le Thi Thanh
Interdisziplinäres Zentrum
für Wissenschaftliches Rechnen
Universität Heidelberg
Im Neuenheimer Feld 368
69120 Heidelberg

Dr. Mathias Lindholm
Matematiska Institutionen
Uppsala Universitet
Box 480
S-751 06 Uppsala

Prof. Dr. Ira M. Longini
Division of Public Health Science
Fred Hutchinson Cancer Research Center
The University of Washington
1100 Fairview Ave. North, LE-400
Seattle WA 98109-1024
USA

Prof. Dr. Emma McBryde
Victorian Infectious Disease Service
Royal Melbourne Hospital
The University of Melbourne
Melbourne , VIC 3010
AUSTRALIA

Dr. Alessia Melegaro
Dondena Centre for Research on
Social Dynamics
Universita Bocconi
Via Sarfatti 25
I-20100 Milano

Prof. Dr. Denis Mollison
The Laigh House
Inveresk
GB-Musselburgh EH21 7TD

Dr. Nico J.D. Nagelkerke
Department of Community Medicine
United Arab Emirates University
P.O.Box 17666
AL Ain
United Arab Emirates

Dr. Peter Neal
Department of Mathematics
The University of Manchester
Oxford Road
GB-Manchester M13 9PL

Prof. Dr. Hiroshi Nishiura
Theoretical Epidemiology
University of Utrecht
Yalelaan 7
NL-3584 CL Utrecht

Prof. Dr. Philip D. O'Neill
School of Mathematical Sciences
The University of Nottingham
University Park
GB-Nottingham NG7 2RD

Dr. Lorenzo Pellis
Department of Infectious Disease
Epidemiology, Imperial College
School of Medicine, Norfolk Place
St. Mary's Campus
GB-London W2 1PG

Prof. Dr. Mick Roberts
Institute of Information and
Mathematical Sciences
Massey University, Priv. Bag 102904
North Shore City 0745
Auckland
NEW ZEALAND

Dr. Lisa Sattenspiel
Department of Anthropology
University of Missouri-Columbia
107 Swallow Hall
Columbia MO 65211
USA

Dr. Gianpaolo Scalia-Tomba
Dipartimento di Matematica
Universita di Roma "Tor Vergata"
Via della Ricerca Scientif. 1
I-00133 Roma

Dr. Birgitt Schönfisch
Institut für Medizinische Biometrie
Universität Tübingen
Westbahnhofstr. 55
72070 Tübingen

Dr. Markus Schwehm
ExploSYS GmbH
Otto-Hahn-Weg 6
70771 Leinfelden

Dr. Simon Spencer
Dept. of Mathematics
The University of Nottingham
University Park
GB-Nottingham NG7 2RD

Anette Stauch
Institut für Medizinische Biometrie
Universität Tübingen
Westbahnhofstr. 55
72070 Tübingen

Dr. Claudio J. Struchiner
Escola Nacional de Saude Publica
Fundacao Oswaldo Cruz
Rua Bejamin Batista 22/202
Rio de Janeiro 22461-120
BRAZIL

Dr. Ake Svensson
Department of Mathematics
Stockholm University
S-10691 Stockholm

Dr. Pieter Trapman
Department of Mathematics
Vrije Universiteit Amsterdam
De Boelelaan 1081a
NL-1081 HV Amsterdam

Dr. Simopekka Vänskä
THL
PL30
FIN-00271 Helsinki

Prof. Dr. Erik M. Volz
Department of Epidemiology SPH-1
University of Michigan
109 Observatory St.
Ann Arbor MI 48109
USA

Dr. Yang Yang
Division of Public Health Science
Fred Hutchinson Cancer Research Center
The University of Washington
1100 Fairview Ave. North, LE-400
Seattle WA 98109-1024
USA

Anna Ziach
Institut für Medizinische Biometrie
Universität Tübingen
Westbahnhofstr. 55
72070 Tübingen