Mathematisches Forschungsinstitut Oberwolfach

Report No. 49/2004

Design and Analysis of Infectious Disease Studies

Organised by
Niels Becker (Canberra)
Klaus Dietz (Tübingen)
Niels Keiding (København)

October 17th – October 23rd, 2004

Mathematics Subject Classification (2000): 62-07, 62P10, 68U20, 91F99, 92B20, 92C60, 92D15, 92D25, 92D30, 92D40.

Introduction by the Organisers

This workshop gathered 44 participants from 19 countries and represented a correspondingly multifaceted program concerning various diseases, public health issues and methodological innovations.

The presentations and discussions highlighted again the crucial role that mathematical models and statistical analyses play in understanding the transmission of infectious diseases and in the development of strategies for their control. Mathematical transmission models and analyses are needed to assess potential control strategies (including determination of optimal strategies), to develop statistical analyses that allow for the dependence in data generated by transmission of infection and to keep track of aspects of the infection dynamics that one cannot observe in practice.

Modeling approaches, data analyses and parameter inferences were applied to diseases like malaria, yellow fever, pneumococcal and meningococcal infections, hepatitis C and smallpox. Due to their current relevance, emerging infectious diseases and SARS attracted special attention and generated much discussion. These were complemented by talks on the public health issues of bioterrorism and interventions like vaccination and contact tracing and general aspects of control and eradicability of infections.

While the research focus is motivated by real world applications, the need to accommodate complex population structures makes this area one requiring diverse

and innovative mathematics and statistics. For example, the models draw on differential calculus, graph theory and multi-type stochastic processes with novel specifications of 'type'. The statistical analyses based on such models often require modern computer intensive methods with novel features, such as the use of random graphs for the transmission chains as latent variables. Furthermore, the methodological spectrum in this workshop comprised Bayesian approaches, computer simulations and the modelling of spatial structures and households.

An evening discussion session on reproduction numbers focussed on the problems of its applicability, in particular with respect to infections in which density-dependent processes operate. The variety of topics made it necessary to organize a session for short talks and an evening with presentations where speakers presented current work on their individual laptops. This provided a very stimulating platform for collaborations and discussions because lecturers could interactively present the work on his/her specific computer environment to a specifically interested audience. This mode of presentation, which turned out to be a novelty at the Oberwolfach institute, could represent an appropriate alternative to poster presentations in future workshops.

The workshop closed with a discussion session about future activities and open problems.

On Thursday evening after a first class concert with piano recitals of Bach, Mozart and Chopin and Lieder of Schubert and Schumann an old Oberwolfach tradition was revived when participants from nearly all represented countries contributed to a most enjoyable variety program of poems and songs.

Niels Becker
Klaus Dietz
Niels Keiding

Workshop: Design and Analysis of Infectious Disease Studies

Table of Contents

Abstracts

Optimal vaccination strategies for epidemics with two types of infection

Frank Ball
(joint work with Niels Becker)

This talk is concerned with vaccination strategies for the control of an epidemic among a community of households (cf. [1] and [3]), in which the severity of disease depends on the dose ingested at exposure. There are two types of infective, mild and severe cases, which, in the absence of vaccination, arise from low- and high-dose exposures, respectively. Individuals mix homogeneously within households and, at a much lower rate, within the community at large. Two models for vaccine action are considered. In the first model, the vaccine response of an individual is described by the realisation of a random vector (A, B), where A describes relative susceptibility compared to an unvaccinated individual and B describes relative infectivity should the vaccinee become infected. In the second model, a vaccinated individual avoids infection when exposed to a low dose and becomes a mild case when exposed to a high dose.

A reproduction number for the epidemic model is derived. Optimal vaccination strategies, which reduce the reproduction number to its threshold value of one with minimum vaccination coverage, are considered for the case when the disease is highly infectious within households (cf. [2]). For the first model of vaccine action, this optimisation problem is solved completely. For the second model of vaccine action, the problem is solved when community mixing is proportionate (i.e. when the proportion of community contacts that are low-dose exposures is the same for contacts made by both mild and severe infectives), and also when all the households in the population have the same size.

Suppose that every household in the population has n members. Then under the first model for vaccine action, with $E(A) < 1$ and $E(B) < 1$, the optimal strategy has the equalising form, which, for given vaccination coverage v, minimises the variance of the number of people vaccinated in a randomly chosen household. Thus the numbers of people vaccinated in the different households are made as equal as possible. Under the second model for vaccine action, depending on the parameter values, the form of the optimal strategy is either equalising, or to vaccinate whole households (which maximises the above variance). Moreover, for some choices of parameter values, the form of the optimal strategy can also depend on the household size n.

References

[1] F. G. Ball and O. D. Lyne, *Optimal vaccination policies for stochastic epidemics among a population of households*, Mathematical Biosciences **177-178** (2002), 333–354.

[2] N. G. Becker and K. Dietz *The effect of household distribution on transmission and control of highly infectious diseases*, Mathematical Biosciences **127** (1995), 207–219.

[3] N. G. Becker and D. N. Starczak *Optimal vaccination strategies for a community of households*, Mathematical Biosciences **139** (1997), 117–132.

On the analysis of randomised malaria trials

Heiko Becher

In this presentation, some aspects in the analysis of randomized malaria trials are outlined. Data available for analysis vary between trials. Usually, participants are visited daily or every two days. In case of fever, blood samples are taken for malaria diagnosis. In addition, cross-sectional surveys are performed at which blood samples are taken from all individuals. Individual data (anthropometric measures and clinical parameters etc.) Observation time varies between less than one year to several years.
Among the difficulties in the analysis are the following:

- How to define the outcome variable.
- how to take the observed parasite count into account
- how to adjust for mortality (malaria as main cause, concomitant cause, other cause of death)
- how to adjust for standard treatment effect
- how to use cross-sectional and longitudinal data simultaneously?

Possible standard analysis include Poisson regression with number of malaria episodes as the outcome variable and the logarithm of the risk period as an offset. In defining the risk period the time under treatment for a given malaria period must be taken into consideration.

The statistical methods used and their inherent limitations will be outlined using two examples:
i) an ongoing 2-arm trial on insecticide-treated bednets (Group 1: bednet from birth; group 2: bednet after 6 months)
ii) a published randomized double-blind placebo-controlled trial on zinc supplementation (Müller et al., 2001)

References

[1] O. Müller, H. Becher, A. Baltussen, Y. Ye, A. Diallo, A. T. Konate, I. Nebie, A. Gbangou, B. Kouyaté, M. Garenne, *Effect of zinc supplementation on malaria and other causes of morbidity in Westafrican children: A randomized double blind placebo controlled trial*, British Medical Journal **322** (2001), 1–6.

Predicting severe sequelae of injection-related Hepatitis C

Sheila M. Bird
(joint work with Sharon J. Hutchinson, David J. Goldberg)

The work presented, which expands on the data and ideas in [1, 2] is from the doctoral thesis of Sharon J. Hutchinson, which was co-supervised by David J. Goldberg and Sheila M. Bird.

Hepatitis C (HCV) is distinguished from Hepatitis B by the absence of vaccine, low transmissibility sexually but 80% chronic carriage; from HIV by 10 times higher transmission risk from needle-stick and by only 5 - 10% 20-year progression rate to severe disease, and 50 - 60% sustained viral clearance by treatment (with pegylated interferon plus Ribavirin). Severe HCV sequelae are cirrhosis, decompensated cirrhosis (DC) and hepatocellular carcinoma (HCC); incubation period is counted in decades; and prognostic factors are age, gender, heavy alcohol consumption and co-infection (such as with HIV or hepatitis A or B).

In Scotland (population: 5 millions), HCV prevalence is 0.3 - 0.4% in pregnant women, but in the mid 1990s was 76% (95% CI: 70 - 81%) for injection drug users (IDUs), only 30% of whom had been HCV-tested. To the end of December 2001, Scotland had registered 13,500 HCV diagnoses, 90% of them in individuals with a history of injection drug use. Diagnoses increase by about 2000 per annum in Scotland where a national capture-recapture study estimated about 25,000 current IDUs (of known demography) in 2000, and HCV incidence in mid to late 1990s was estimated at 20 - 34 per 100 susceptible injector-years.

Against this background, our aim was to project severe sequelae of injection-related HCV, primarily for Glasgow IDUs about whom there is a wealth of behavioural data from HIV/HCV surveys in the 1990s. The work was in three parts: estimating incidence and cessation of injecting (Modified delphi technique); estimating the incidence of HCV among injectors (HCV transmission model); and estimating progression from HCV infection to severe disease among ever-IDUs (HCV progression model, which took account of covariate influences (age at infection, gender, heavy alcohol use, and co-infection with HIV) - as determined by meta-analysis of HCV progression studies). Also critical was a Database Linkage Outcomes Study because, unlike for HIV/AIDS, there is no national register of diagnoses of severe liver sequelae of injection-related HCV. Thus, these outcomes had to be determined by matching of the HCV diagnosis register to hospitalisations, cancer registrations, HIV test database, and deaths (by cause) register. Briefly, key results were the following:

a) experts' opinion about (B) incidence and (C) cessation of injection drug use in 1960-2000 in Glasgow were not coherent with their beliefs about (A) IDU prevalence. By a form of rejection-sampling, coherence distributions for (B) and (C) were derived which were consistent both with (A) and with capture-recapture estimates for Glasgow's prevalent IDUs at the end of the 1980s and in 2000.

b) stochastic modelling was then used to follow HCV infection status - on an individual basis - of current injectors from IDU onset to death/cessation. The

above coherence disrtibutions (B) and (C) were used together with behavioural data (on percent of IDUs who had shared needle-syringe (N/S) in the past year; conditional geometric distribution for number of N/S partners; and percentage of injecting episodes (assumed 3 per day for 48 weeks per year) that were with used N/S - decreasing with decreased number of sharing partners), and viral factors (transmissibility and persistence of HCV infection). By incorporating a 10-fold increase in infectiousness in an initial high viraemia phase of 6 - 8 weeks was HCV prevalence in 1980 - 2000 adequately modelled - otherwise it was systematically under-estimated compared to survey data whereas HIV prevalence was properly accounted for by use of same behavioural assumptions.

c) stochastic modelling was used to follow individuals - in annual cycles - from IDU onset through HCV progression states. Lacking behavioural data on the percentage of IDUs with heavy alcohol consumption [Relative Risk for progression: 2.3 (1.7 to 3.3)], three scenarios (0%, 20% and 40%) were investigated, noly the highest of which gave adequate fit to Glasgow's estimated IDU-related 179 (169 - 193) cases of DC in 1996-2001 - provided also that progression rate to cirrhosis at 20 years was revised up from 6.5% (3.5 - 9.5%, based on Freeman et al.[3]) to 7.5% (5 - 10%).

Public health implications were as follows: projected doubling of Glasgow IDUs HCV-related cases of DC between 2000 and 2020; in 2005, Glasgow's former IDUs will include about 13,000 aged 30-39 years, one sixth of whom have moderate to severe (treatable) HCV disease, and 5,100 aged 40-49 years, one third of whom have moderate or severe HCV disease; modelled disease toll by 2000 (and by 2010 - when around 40 years of age) among 1800 new Glasgow IDUs in 1985 (peak year) was: HCV related death 1% (3%), former IDU death 3% (8%), current IDU death 10% (11%) - mainly overdose deaths, moderate HCV disease or worse 18% (32%) with only 29% (29%) having been HCV uninfected/recovered.

Model criticisms include: need to question constant drugs-related death rate irrespective of gender, age and calendar time; and need for better information about cause-specific death rate of former IDUs and about (current and former) IDUs' heavy alcohol consumption.

References

[1] S. M. Bird, D. J. Goldberg, S. J. Hutchinson, *Projecting severe sequelae of injection-related hepatitis C virus epidemic in the UK. Part 1: Critical hepatitis C and injector data*, Journal of Epidemiology and Biostatistics **6** (2001), 243–265.

[2] S. M. Bird, D. J. Goldberg, S. J. Hutchinson, *Projecting severe sequelae of injection-related hepatitis C virus epidemic in the UK. Part 2: Preliminary UK estimates of prevalent injection-related hepatitis C carriers, and derivation of progression rates to liver cirrhosis by gender and age at hepatitis C virus infection*, Journal of Epidemiology and Biostatistics **6** (2001), 243–265.

Estimating transmission parameters for infectious diseases in small hospital units

Martin C. J. Bootsma
(joint work with Marc J. M. Bonten, Odo Diekmann)

Resistant pathogens in hospitals form an emerging health care problem and effective strategies to prevent their spread are required. However, the efficacy of control measures depends on the nature of the spreading mechanisms. Therefore the identification of the most important route is important. Recently, two articles appeared [1, 2] to determine the relative importance of outgrowth of pathogens, that were already present at an undetectable level, due to antibiotic selection and cross transmission. Both methods have some disadvantages. Both assume constant discharge rates and a constant bed occupancy and cannot distinguish between admission of colonized patients and the outgrowth of already present pathogens. Moreover, the model in [1] assumes that acquisition took place at the moment of detection, while [2] only looks at infection data. However infections only represent the tip of the iceberg as most patients carry pathogens asymptomatically.

We use a Markov chain approach based on likelihood methods, to make optimal use of the available data (date of admission, date of discharge, dates of the microbiological culturing and the results of these cultures) to determine the importance of different routes. This approach does not require genotyping. We use a discrete time framework where the up-dating (on a day by day basis) consists of 4 parts: 1) Evolution according to the mechanistic model, 2) Use of the results of the culturing, 3) Removal of the patients from the state space for which the colonization status is certain, 4) Incorporate new patients in the state space for which the colonization status becomes uncertain. Within the framework, additional patient characteristics can be incorporated as well. In this framework, we glue together state spaces of different size according to the need as exposed by the data.

To determine optimal culture strategies, we simulated a unit of 10 beds. Of course, the more frequent patients are cultured, the faster the most important infection route can be determined statistically significantly. However, for typical values of the parameters, a regime of culturing patients on admission and afterwards twice per week gives the most information per performed culture.

References

[1] I. Pelupessy, M. J. M. Bonten and O. Diekmann, *How to assess the relative importance of different colonization routes of pathogens within hospital settings*, Proc. Natl. Acad. Sci. **99** (2002), 5601-5605.

[2] B. Cooper, M. Lipsitch, *The analysis of hospital infection data using hidden Markov models*, Biostatistics **5** (2004), 223-237.

Epidemic models with different severities

Tom Britton
(joint work with Frank Ball)

Inspired by discussions at an epidemic meeting in Mariefred, Sweden, 2003, this talk is concerned with stochastic epidemic models where there are two types of severity, mild and severe, of infectious individuals. In the first model the type of severity depends on the amount of infectious exposure an individual receives, and infectives are always initially mild but may become severe if additionally exposed. In the second model an infective can only be one type, but the type typically depends on the type you were infected by. Large population properties of the models are derived and compared. The first model has the property that the basic reproduction number only depends on parameters of the mild infectious state whereas the final size in case of a major outbreak also depends on parameters of the severe state, and the limiting final size proportions need not even be continuous in the parameters. The second model resembles a so-called competing epidemic model having a more complex final outcome behaviour. In the talk we will discuss pros and cons of the models and look forward to comments (criticism) from the audience.

References

[1] F. G. Ball and T. Britton, *Epidemic models with varying severity*, (2004), (In preparation).

Survival analysis of Taiwan SARS data by a semiparametric mixture model

I-Shou Chang
(joint work with Tsung-Hsi Wang, Yung-Hsiang Huang, Donald Dah-Shyong Jiang, Jhy-Yuan Yang, Ting-Hsiang Lin, Che-Chi Yang, Yuh-Jenn Wu, Chi-Chung Wen, Ih-Jen Su, Chao A. Hsiung)

A semiparametric mixture model for competing risks problem is used for the analysis of SARS(severe acute respiratory syndrome) patient data, where a patient admitted to a hospital experiences either a death or a discharge from the hospital. This model includes an explicit modelling on covariate-specific fatality rate, and a semiparametric modelling on the time to death and the time to discharge. Identifiability of the parameters in the model and the existence and consistency of the standard nonparametric maximum likelihood estimate (NPMLE) are established. Self consistency equations derived from the score functions are used to get an iterative algorithm for the computation of NPMLE; bootstrape method is used to provide confidence intervals of the estimates [3].

This model was first proposed by Fine [6], which generalizes the parametric mixture model of Larson and Dinse [9]. Estimation in this model was also discussed by Fine and Gray [8], Betensky and Schoenfeld [2], Fine [7], Andersen [1], Donnelly

et al. [5]. Readers are referred to Lingappa et al. [10] for a general introduction of SARS epidemic.

Using this method, we report the covariate-specific fatality rate, covariate-specific time of onset-to-death, and covariate-specific time of onset-to-discharge of SARS patients, based on Taiwan data in the year of 2003 (cf. [4]). All the statistical significance results are based on 95% confidence intervals. We find that age is a significant covariate, especially in the group with positive laboratory test result. In fact, the estimated age-specific fatality rate is a strictly increasing function of age, and the hypothesis that fatality rate doesn't change with age can be rejected at significance level less than 0.05. For the curable ones, whom we do not know in advance, both the elderly and the laboratory confirmed cases spend longer time in hospital before discharge. For the incurable ones, whom we do not know in advance either, the laboratory confirmed cases spend longer time in hospital before death. Neither gender nor time from onset to admission play any statistically significant role in the fatality rate, in the time to death, or in the time to discharge.

References

[1] P. K. Andersen, *Competing risks as a multi-state model*, Statistical Methods in Medical Research **11** (2002), 203-215.

[2] R. A. Betensky and D. A. Schoenfeld, *Nonparametric estimation in a cure model with random cure times*, Biometrics **57,1** (2001), 282-286.

[3] I. S. Chang, C. A. Hsiung, C. C. Wen, Y. J. Wu, and C. C. Yang, *Nonparametric maximum likelihood estimate in a semiparametric model for competing risks data*, In preparation (2004).

[4] I. S. Chang, T. H. Wang, Y. H. Huang, D. D. S. Jiang, J. Y. Yang, T. H. Lin, C. C. Yang, Y. J. Wu, C. C. Wen, I. J. Su, and C. A. Hsiung, *Survival analysis of SARS probable cases in Taiwan 2003*, In preparation (2004).

[5] C. A. Donnelly, A. C. Ghanl, G. M. Leung, A. J. Hedley, C. Fraser, S. Rifey, L. J. Abu-Radded, L. M. Ho, T. Q. Thach, P. Chau, K. P. Chan, T. H. Lam, L. Y. Tse, T. Tsang, S. H. Llu, J. H. B. Kong, E. M. C. Lau, N. M. Ferguson, and R. M. Anderson, *Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Knog*, Lancet **361** (2003), 1761-1766.

[6] J. P. Fine, *Analysing competing risks data with transformation models*, J. R. Statist. Soc. B **61,4** (1999), 817-830.

[7] J. P. Fine, *Regression modelling of competing crude failure probabilities*, Biostatistics **2,1** (2001), 85-97.

[8] J. P. Fine and R. Gray, *A proportional hazards model for the subdistribution of a competing risk*, Journal of the American Statistical Association **94** (1999), 496-509.

[9] M. G. Larson and G. E. Dinse, *A mixture model for the regression analysis of competing risks data*, Appl. Statist. **34,3** (1985), 201-211.

[10] J. R. Lingappa, L. C. McDonald, P. Simone , U. D. Parashar, *Wresting SARS from uncertainty*, Emerg Infect Dis.**10** (2004), 167–70.

The impact of density-dependent processes on the eradicability of parasitic diseases

Hans Peter Duerr
(joint work with Martin Eichner & Klaus Dietz)

The control of parasitic diseases has been subject to past and current WHO activities. Although some examples of (local) parasite elimination exist (schistosomiasis, lymphatic filariasis), the global control campaigns have either not achieved eradication (e.g. onchocerciasis) or are in progress (e.g. lymphatic filariasis). It remains unclear if these infections can be (globally) eradicated so that our predictions of intervention success rely on modeling approaches. The prerequisites of eradicability are transmission thresholds [1] (*a vector density* below which the infection cannot persist) and breakpoints [2] (*a parasite density* below which the infection cannot persist). In contrast to transmission thresholds, which are known since long, the effects of breakpoint-inducing processes are widely unexplored. Breakpoints can result from particular density-dependence, e.g. facilitated infection [3] which has been suggested to occur in filarial diseases as a consequence of parasite-induced immunosuppression.

We investigate numerically how density-dependent processes (i.e. facilitation and limitation processes) impact on the eradicability of parasitic diseases. It seems to be a rule that facilitation processes increase transmission thresholds and induce breakpoints into the transmission of an infection whereas limitation processes lower transmission thresholds or potentially existing breakpoints to a value where these practically can disappear. This can be summarized with respect to intervention campaigns, because facilitation processes will 'facilitate' the eradicability of a parasite, whereas limitation processes will 'limit' the prospects of such a success. With respect to mathematical models, this means, that predictions of the success of intervention programs will be over-optimistic if the degree of facilitation is overestimated or if the degree of limitation is underestimated. Vice versa, model predictions will be over-pessimistic if the degree of facilitation is underestimated or if the degree of limitation is overestimated. Both types of density-dependence may coexist in the transmission cycle and even in the same host, and their sum decides whether the net effect on eradicability will be positive or negative.

In this context it would be helpful to have measures of density-dependence available by which the net result of coexisting processes could be quantified. The design of future control activities and the predictability of intervention success highly depend on our knowledge about parameters by which the infection dynamics are driven. Our sensitivity analysis identifies relevant parameters and provides information where future research should focus on [4].

References

[1] R. Ross, *The prevention of malaria*, John Murray (1911).
[2] G. Macdonald, *The dynamics of helminth infections, with special reference to schistosomes*, Trans. R. Soc. Trop. Med. Hyg. **59** (1965), 489–506.

[3] R.R. Regoes, D. Ebert, S. Bonhoeffer, *Dose-dependent infection rates of parasites produce the Allee effect in epidemiology*, Proc. R. Soc. London Ser. B **269** (2002), 271–279.
[4] H. P. Duerr, K. Dietz, M. Eichner *Determinants of the eradicability of filarial diseases*, Trends Parasitology (to appear).

Can residual protection conferred by previous smallpox vaccination be harmful?

Martin Eichner
(joint work with Markus Schwehm)

Recent discussion about the use of smallpox by bioterrorists have rekindled the interest in the one and only infectious disease defeated by a huge concerted vaccination action. About half of the population of most industrial countries have experienced smallpox vaccination some decades ago. Although previously vaccinated individuals retain some degree of protection against disease, most of them have lost their protection against infection. We assume that a previous vaccination reduces susceptibility to smallpox by a factor f_S. Furthermore previous vaccination of an infected individual may reduce infectivity to contacts by a factor f_I. But previous vaccination can also have a dark side which may actually lead to a more devastating outbreak. Previously vaccinated individuals might show less symptoms and thus be recognized (by a factor f_D) later as carriers of the infection than unvaccinated cases with clearly visible symptoms. Finally, previously vaccinated cases developing only a mild infection may move around more freely, contacting (by a factor f_M) more people per unit of time, whereas severe cases concentrate most of their contacts to few close contacts and caregivers.

We use a deterministic mathematical model and an individual-based stochastic computer simulation to explore the effect of previous vaccinations on the outcome of an epidemic for a wide range of parameter constellations. Both models consider case detection, isolation and contact tracing as interventions. The average outbreak size of a basic stochastic model corresponds well with the outbreak size predicted by the deterministic model over a wide range of parameter constellations. However, the range of outbreak sizes computed by the stochastic model are too large to be ignored. This observation becomes amplified when a more realistic and detailed stochastic model (including scale-free contact network topology, close and remote contacts, gamma-sampled state transitions and limited intervention capacities) is used. Finally the computed results are compared with historical reports of smallpox outbreaks.

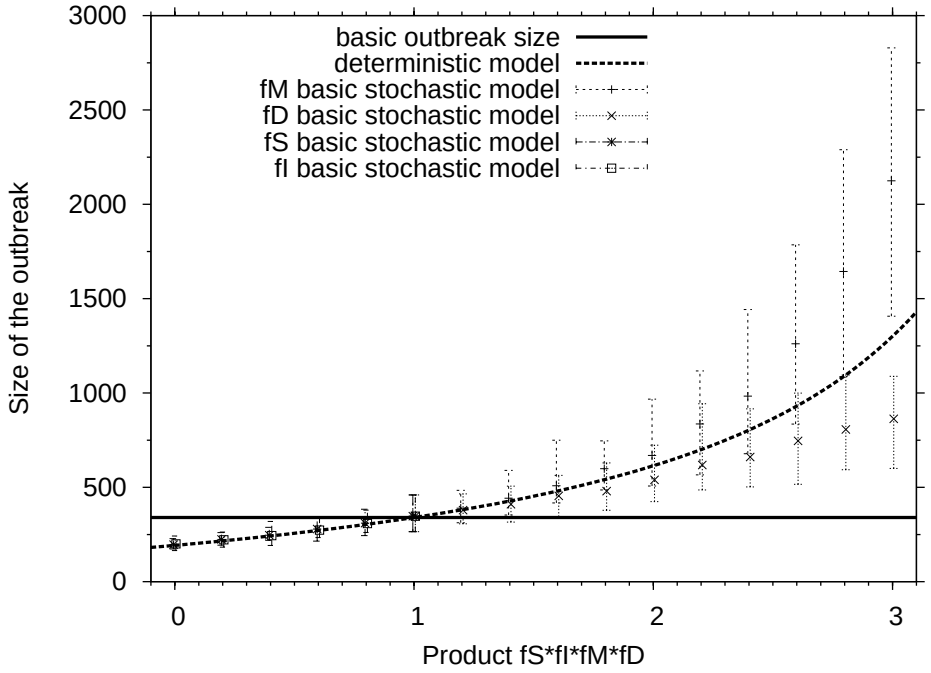

FIGURE 1. Comparing the "basic outbreak size" (fS*fI*fM*fD=1; full line) with the expected outbreak size for different products fS*fI*fM*fD (dashed line) and with outbreak scenarios in an individual-based stochastic model with exponentially distributed sojourn times.

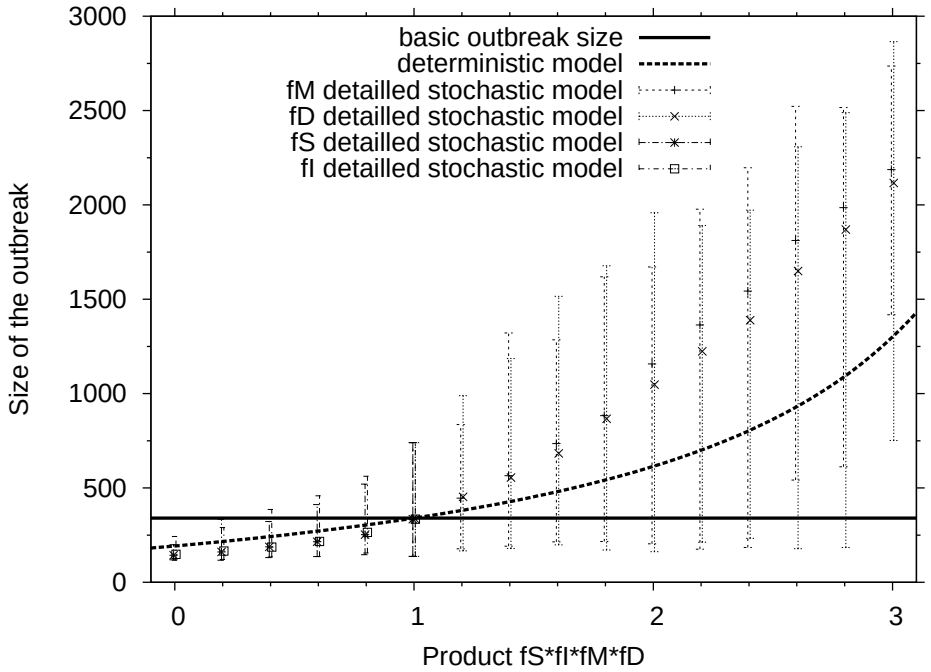

FIGURE 2. Like Fig. 1, but under more realistic assumptions (e.g. gamma-distributed sojourn times).

SARS Incubation and Quarantine Times

Vernon T. Farewell
(joint work with A.M. Herzberg, K.W. James, L.M. Ho, G.M. Leung)

Quarantine was one of the key aspects of infection control introduced during the recent SARS (severe acute respiratory syndrome) epidemic. An important paper on epidemiological aspects of SARS was that of Donnelly *et al* [1]. The work reported here arose from a question related to the confidence a community should have that an individual who has passed through the SARS quarantine period is disease-free. The concept of a maximum incubation time could be relevant to these considerations.

Consider a gamma distribution for incubation times. Thus if T is the random variable representing an incubation time, with an observed value of $T = t$, then a gamma distribution for T is specified by the probability density function

$$g(t) = \frac{1}{s^a \Gamma(a)} t^{(a-1)} e^{-(t/s)},$$

where $t > 0, a > 0$ and $s > 0$.

If this distribution is truncated at some time M, so that $0 < T < M$, then the density function for T becomes $f(t) = g(t)/G(M)$, where $G(M) = \int_0^M g(t)dt$.

Estimation of the parameters a, s and M can be based on the likelihood function denoted $L(a, s, M)$. The profile likelihood for the parameter M is defined by $L_P(M) = L(\tilde{a}(M), \tilde{s}(M), M)$, where $\tilde{a}(M)$ and $\tilde{s}(M)$ are the MLEs of a and s for a fixed value of M. This function can be standardized as $L_P^*(M) = \frac{L_P(M)}{L_P(\hat{M})}$, where $\hat{M}$ is the MLE of M.

The MLE of M is at the largest t_i. However, standard asymptotic distributional results for MLEs will not be applicable for the parameter M. Here, it is sufficient to consider the comparative shapes of likelihood functions, regarded simply as representing the information *available from the data* for inference concerning the unknown parameters.

Alternatives to the truncated gamma can be considered and the truncated log-normal is specifically examined. A more general alternative is the so-called log-gamma distribution of Farewell and Prentice [5] which represents a reparameterization and extension of a generalized gamma distribution.

With $\alpha, q \in R$ and $\sigma > 0$, the log-gamma model can be written as the location scale model $y = log(t) = \alpha + \sigma w$, where the density $f(w; q)$ for w is

$$| q | (q^{-2})^{q^{-2}} \exp[q^{-2}\{qw - \exp(qw)\}/\Gamma(q^{-2})$$

if $q \neq 0$ and, when q=0, is the standard normal distribution. The log-gamma distribution includes the Weibull ($q = 1$) and exponential ($q = \sigma = 1$) distributions as special cases as well as the gamma ($q = 1/\sigma$) and log-normal (q=0).

We consider data from 67 SARS cases in Hong Kong whose interval of possible exposure times is less than 5 days. The data consist of a maximum and minimum possible incubation period for each case. For the averaged times, Figure 1 presents

the profile likelihoods, $L_P^*(M)$, based on the gamma, log-normal and log-gamma models.

For the truncated gamma model, the profile likelihood never drops below 60% suggesting that any value of M greater than the maximum time observed, 14, is plausible. However, the situation is different for the truncated log-normal model. For this model, any value for M greater than 19.5 days makes the data more than 10 times less plausible than does the MLE of 14 days. For the log-gamma model that includes both other models as special cases, the profile likelihood for M is more informative than that based on a gamma model, but it never falls below a value of 20%. This is true even though the maximum likelhood estimate of q is -0.13, a value close to the value $q = 0$ corresponding to the log-normal model.

We also consider likelihood estimation based explicitly on the maximum and minimum incubation times. The general pattern of the likelihoods is similar to that in Figure 1, but, with interval-censoring, not even the log-normal likelihood drops to less than the 10% level.

When there is insufficient evidence to be very confident about a maximum incubation time, an alternative approach is to set a quarantine time on the basis of percentile estimation. The estimated 95th percentiles for the untruncated log-gamma and log-normal distributions are 10.66 and 12.09. Corresponding 95% confidence intervals are (9.24, 13.68) and (9.95, 15.34). Note that if interval-censored data are used to fit the log-gamma model, then the estimated 95th percentile is 10.2 with a confidence interval of (8.64, 13.68), an interval 14% longer than that for the averaged data.

It appears a quarantine time of 10 days for SARS might release one infectious patient in twenty. The larger the quarantined population, the larger the number of released infectious individuals. Thus the length of a quarantine period might well be set in light of the expected number of quarantined indivduals.

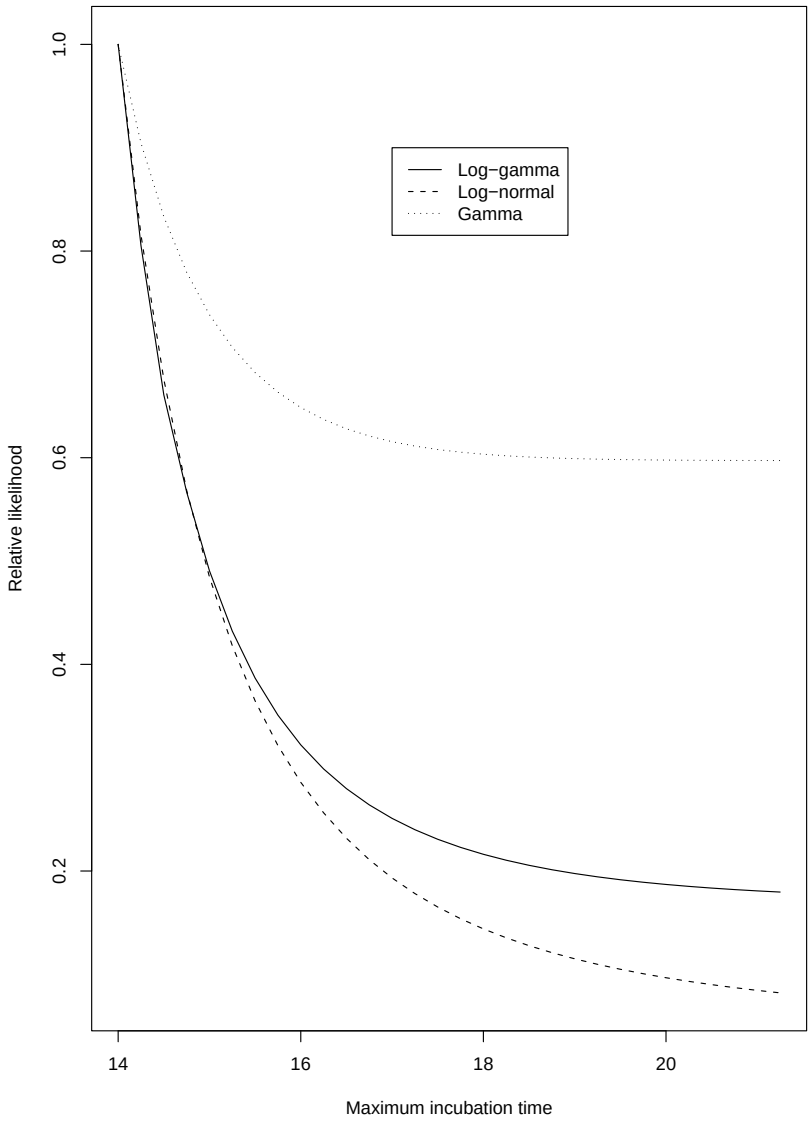

FIGURE 1. Profile likelihoods based on 67 SARS cases

REFERENCES

[1] C. A. Donnelly, A. C. Ghani, G. M. Leung, A. J. Hedley, C. Fraser, S. Riley, L. J. Abu-Raddad, L. M. Ho,T-Q. Thach, P. Chau, K-P. Chan, T-H. Lam, L-Y. Tse, T. Tsang, S-H. Liu, J. H. B. Kong, E. M. C. Lau, N. M. Ferguson, R. M. Anderson, *Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong*, Lancet **361** (2003), 1761–1766.

[2] V. T. Farewell, R. L.Prentice, *A study of distributional shape in life testing*, Technometrics **19** (1977), 69–75.

Evolution and Epidemiology of Bacterial Pathogens

Christophe Fraser
(joint work with William P. Hanage and Brian G. Spratt)

Transmissible bacteria are a fascinating evolutionary system. They are complex enough to engage in sophisticated evolutionary strategies, including several mechanisms of gene exchange (i.e. bacterial sex). Bacterial populations often exhibit complex strain structure. However their fate is also intimately tied in with their host. The basic act of bacterial reproduction is transmission from one host to another, which means that for human-carried bacteria, human sociology can have a major impact on bacterial evolution. Furthermore, colonising bacteria must not only compete against each other, but also against various host immune responses.

Evolutionary Questions

The study of bacterial evolution is often motivated by practical problems. How will bacterial populations respond to imperfect, strain specific vaccines? For example for the new pneumococcal conjugate vaccines that targets seven of the most common serotypes, replacement by non-vaccine strains is of concern. How do bacterial populations respond to specific selection pressures such as exerted by antibiotics? Methycillin resistance has become widespread in hospital populations of *Staphylococcus aureus* (MRSA) but not yet in the community. Will this change? However, there are also more basic questions that should be asked, such as identifying evolutionary models which best explain observed population structures.

Epidemiology Questions

Genetic typing of bacteria is widely used to identify clusters of transmission. However doing this without an underlying evolutionary model is problematic, since it is difficult to identify whether two strains are identical because of recent transmission or because of common evolutionary descent. If this problem can be resolved, genetic studies may be invaluable for inferring both local and global patterns of transmission. Their study may contribute to our understanding of disease transmission networks, which underlie predictions of for example the spread and possible control of future influenza pandemics.

The studies and methods

Our approach is data-driven. We focus on four studies of three pathogens that together cause substantial global morbidity and mortality ([2, 3, 4, 5]): *Streptococcus pneumoniae*, *Neisseria meningitidis* and *Staphylococcus aureus*. These three species share several features. First, they are accidental pathogens. Infection only very rarely leads to disease. Healthy carriage is the norm. The reason the net burden of disease is large is that infection rates are extremely high. Our four studies thus focus on the natural populations, i.e. samples of carried bacteria. Second, they recombine, in that small segments of their genome are transposed (directly, or via several indirect routes). Third, they exhibit complicated strain structure.

One sample was collected from each healthy volunteer in each study. Each sample were characterised by multi-locus sequence typing, a method where seven housekeeping genes (circa 450 base pairs long) are directly sequenced (see www.mlst.net for more details). This method has recently become the gold standard for bacterial strain typing.

A basic model

A basic model of evolution was formulated, analysed and compared to the data. The model extends the classical Wright-Fisher population genetic model to several loci (=genes) and to varying levels of recombination. In each discrete time-step in the model, each infection seeds a number of new infections, dependent on the fitness of the strain. The total number of infections is kept constant. In each new infection, any of the genes can mutate with probability m, in which case it is assumed that a new previously unseen allele is generated. There is also a probability r that a gene is transformed by recombination, in which case a donor for the gene is chosen at random from the parent population. Eventually, this model reaches equilibrium (as measured by any metric), and the model is then compared to the data.

The metric that was chosen for model comparisons was the proportion of pairs of samples that differed at none, one, two, etc. of the seven sequenced genes. It was found that the null (neutral) model of evolution, in which each strain is considered equally fit, described the model quite well, but statistically significant deviations were observed. We thus searched for the most plausible and parsimonious explanation for the deviation.

Local transmission and the excess of identical strains

All of studies were based on individuals chosen from cities, with an above average chance of knowing each other. We thus surmised that there was a substantial chance that the samples collected could have been taken from a single transmission cluster, all sharing an identical genotype. We thus formulated a model whereby samples are collected from clusters, and clusters are seeded from a globally neutral population, as described in the model above. We found that this model provided a very good fit to all four studies.

Testing the model

To test the model, we used three other metrics:
i) For each sample, how many identical genes it shares with the most similar but not identical other sample in the population (of the seven studied).
ii) How many 'clonal complexes' and 'singletons' were identified by the phylogenetic program eBURST (see www.mlst.net).
iii) The inferred recombination rate as compared to that estimated by two other commonly used (but much less efficient) methods [6, 7].
In each case the model provided a good fit. Furthermore, a number of other alternative parsimonious models involving selection were analysed, and did not provide

a good fit to the data. Frailty was tested by simulating from a multi-factorial model which included positive selection at a hidden gene, and transmission clusters. The neutral model was unable to fit to this simulated data.

Conclusions

We describe the evolutionary picture that emerges from our analysis as neutral micro-epidemic evolution. At the local level, bacteria are transmitted epidemically, with single genotypes sweeping through families, daycare centres, workplaces. However at the global level, the effect of these epidemics is smoothed out. We found no evidence for strong differences in fitness between circulating strains. The bacterial population can effectively be considered selectively neutral. The analysis allowed us to estimate roughly the size and number of transmission clusters in our study, and we are currently investigating what we may infer for the communities at large, and how this relates to socially 'realistic' models of pathogen transmission.

The method allowed us to estimate both mutation and recombination rates without the substantial computational burden associated with current methods. We are also gathering more samples, which may allow us to elucidate both local and global transmission rates. The method will be extended to more bacterial species, to more genes, and to samples where selection is more likely to be seen in action.

Our model provides a calibrated and validated framework within which questions of specific selective pressures associated with vaccine or antibiotic usage can be addressed. The analysis highlights that for transmitted bacteria, evolutionary and epidemiological questions are intimately linked and cannot be answered independently.

References

[1] C. Fraser, W. Hanage, B. G. Spratt, *Neutral micro-epidemic evolution of bacterial pathogens*, Proc. Natl. Acad. Sci. U.S.A. **32**, in press.

[2] K. A. Jolley et al., *Carried meningococci in the Czech Republic: a diverse recombining population*, J Clin Microbiol **38** (2000), 4492–4498.

[3] W. P. Hanage et al., *Ability of pneumococcal serotypes and clones to cause acute otitis media: implications for the prevention of otitis media by conjugate vaccines*, Infect Immun **72** (2004), 76–81.

[4] N. P. Day et al., *A link between virulence and ecological abundance in natural populations of Staphylococcus aureus*, Science **292** (2001), 114–116.

[5] E. Meats et al., *Stability of serotypes during nasopharyngeal carriage of Streptococcus pneumoniae*, J Clin Microbiol **41** (2003), 386–392.

[6] G. McVean et al., *A coalescent-based method for detecting and estimating recombination from gene sequences*, Genetics **160** (2002), 1231–1241 .

[7] E. J. Feil et al., *The relative contributions of recombination and mutation to the divergence of clones of Neisseria meningitidis*, Mol Biol Evol **16** (1999), 1496–1502.

Inferring the presence of undiagnosed cases during infectious disease outbreaks

Kathryn Glass
(joint work with Niels G. Becker, Mark S. Clements)

During an outbreak of an infectious disease, data can only be collected on individuals that are diagnosed. Individuals that experience a mild or asymptomatic form of the disease are unlikely to seek medical attention, and so will not be diagnosed and recorded. If the numbers of undiagnosed individuals are sufficiently large, it becomes difficult to infer the effect that control measures are having on disease transmission, and in particular, it is difficult to be sure when the outbreak is over. This is particularly problematic when there is an outbreak of a newly-emerged infection (such as SARS) for which limited information is available on which to estimate the rates of undiagnosed infections.

We use a two-type branching process to model disease transmission in generations [**?**]. Individuals in generation t are classified as either diagnosed (D_t), or undiagnosed (U_t). We assume that a fixed proportion, π of new cases are undiagnosed, so that the number of cases in generation $t+1$ is given by:

$$\begin{aligned} D_{t+1} &\sim \text{P}oisson(\pi(\mu_t D_t + \mu U_t)) \\ U_{t+1} &\sim \text{P}oisson((1-\pi)(\mu_t D_t + \mu U_t)). \end{aligned}$$

We assume that we have data from N generations of cases, and that control measures were applied from generation M onwards. These control measures are assumed to affect the diagnosed cases only, so that the decline in cases is driven by the impact of control on diagnosed cases. We model this by setting $\mu_t = \mu$ for $t \leq M$, and $\mu_t = \mu_c$ for $t > M$. Thus, the model includes three parameters: the reproduction number before control (μ), the reproduction number of diagnosed cases after control (μ_c), and the proportion of cases that are undiagnosed (π).

Here, inferences in the Bayesian framework are proposed. In order to estimate the posterior distributions of the parameters, we write down the likelihood function.

$$\begin{aligned} L(\mu, \mu_c, \pi) &= \prod_{t=1}^{M} \frac{[\mu(D_{t-1}+U_{t-1})]^{D_t+U_t} \pi^{D_t}(1-\pi)^{U_t} e^{-\mu(D_{t-1}+U_{t-1})}}{D_t! U_t!} \\ &\quad \times \prod_{t=M+1}^{N} \frac{[\mu_c D_{t-1} + \mu U_{t-1}]^{D_t+U_t} \pi^{D_t}(1-\pi)^{U_t} e^{-(\mu_c D_{t-1}+\mu U_{t-1})}}{D_t! U_t!} \end{aligned}$$

Note that although data will exist for the observed number of diagnosed cases in each generation (D_t), the number of undiagnosed cases in each generation (U_t) will remain unknown and must be treated as a latent variable. We generate estimates for the parameters and latent variables in the model using a Metropolis-Hastings algorithm [2]. The distributions of the parameters provide information on the type-specific reproduction numbers of the infection, and on the fraction of cases

that are undiagnosed. The distribution of number of undiagnosed cases in the last generation can be used to quantify the probability that the epidemic is over.

We first test the algorithm on artificial data generated by the model, both with and without undiagnosed cases. The algorithm provides a good estimate of μ in both cases, and distinguishes between the cases with and without undiagnosed cases in the distribution of π.

Data from the SARS outbreaks is available by date of onset of symptoms [3]. We use data from Hong Kong, Taiwan, and the first wave of infection in Canada, and group these into generations, assuming that a generation corresponds to 10 days. Applying the algorithm to this data, we estimate a basic reproduction number of between 1.5 and 3.0, and that around 80% of cases are diagnosed.

In order to keep the model relatively simple, a number of assumptions have been made about the patterns of transmission. We explore the effect of relaxing some of these assumptions by generating some artificial data sets and looking at the resulting parameter estimates. We find that the algorithm may overestimate the basic reproduction number (μ) if control measures were in place earlier than the model assumes. The model also has trouble distinguishing between 'undiagnosed' cases and 'uncontrolled' cases that are diagnosed, but do not experience any reduction in the reproduction number after control. However, even allowing for this, the number of undiagnosed (or uncontrolled) cases remains a valuable indicator of whether the outbreak has been successfully controlled.

In future work, we intend to perform a systematic appraisal of the algorithm to determine the type and quantity of data needed for this form of analysis. The current algorithm requires outbreak data to be grouped into generations before analysis is performed. We intend to explore the use of the continuous-time Bellman-Harris branching model [1] to avoid the need to identify generations.

References

[1] P. Jagers, *Branching processes with biological applications*, John Wiley and Sons: London (1975).

[2] P. D. O'Neill, *A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods*, Math. Biosciences **180** (2002), 103–114.

[3] World Health Organisation, *Probable cases of sars by date of onset*, available at WHO site: `http://www.who.int/csr/sars/epicurve/en/epicurves2003_06_17.pdf`.

Timely Identification of Control Strategies for Emerging Infectious Diseases: Severe Acute Respiratory Syndrome in Singapore

John Glasser

(joint work with Zhilan Feng)

Background

Within weeks of a traveler from Guangdong Province, China, with Severe Acute Respiratory Syndrome (SARS) infecting others in their Hong Kong hotel, several

case-series were published and the responsible pathogen was identified (Gerberding 2003). Public health responses to the ensuing outbreaks in Hong Kong and elsewhere were equally swift, but not equally effective.

Objectives

To facilitate intervening effectively in outbreaks of new diseases, we modeled a generic emerging infectious disease apparently transmitted by close contact, but about which little else is known. Initial objectives were 1) to elucidate social phenomena affecting disease transmission, which well-crafted health communiqués could influence, and 2) to determine if quarantine accelerated control of SARS, or if timely control of this new disease could have been achieved by influencing those social phenomena. But results motivated us 3) to derive analytical expressions for the impact of all possible interventions that could be evaluated quickly in future.

Methods

1) Our model is a system of non-autonomous differential equations (DE) in which proportions seeking medical care during the prodrome and being diagnosed and effectively isolated may evolve, and their contacts be quarantined soon after exposure, whereas contacts of those misdiagnosed or who present with acute respiratory symptoms may be identified too late. 2) We estimated stage-specific infection rates, conditional on early clinical observations and these social phenomena, by minimizing disparities between predicted and observed hospitalizations following the 2003 importation of SARS to Singapore. 3) We derived an expression for the average number of infections per infectious person, R ,which must be < 1 for control, from the autonomous DE underlying our new disease model, and described its relationship to R_0, the reproductive number in a wholly susceptible population absent intervention. 4) To evaluate various social phenomena that Singapore's Ministry of Health orchestrated, as well as its quarantine of possible contacts, we took partial derivatives with respect to each intervention.

Results

During the outbreak in Singapore, people with compatible symptoms sought medical care earlier (proportion hospitalized within 4 days of onset increased from 0.3 to 0.9) and clinicians became proficient at diagnosing them (proportion isolated on admission also increased from 0.3 to 0.9). We do not yet know which patients had been quarantined, but use the proportion isolated within one day of symptom onset, which increased from 0.05 to 0.6, as a surrogate. Conditional on other relevant social phenomena, we estimate that quarantine had only modest impact (five cases and one death averted). Because 7,863 people were quarantined, only 11 of whom became ill (Tan 2004), the societal cost of this intervention was enormous. We show that, for biologically reasonable pathogen, and fixed but otherwise reasonable host response parameters, 7% more cases seeking care during their prodrome versus acute illnesses is equivalent to quarantining 87% of contacts (cf. only 5%

were quarantined in Singapore and Taiwan).

Conclusions
1) SARS was controllable solely by ensuring that people with compatible symptoms sought medical care during the prodrome, especially ones who might have been exposed, and effectively isolating those diagnosed (Fraser et al. 2004). 2) Given biological parameters estimable from early case-series or experience with related pathogens, our analytical expressions permit identification of the most promising responses to emerging infectious diseases. 3) By refining initial parameter estimates as information accumulated and monitoring intervention effects, modelers could ensure timelier and otherwise more advantageous allocation of resources in future outbreaks of new diseases.

References

[1] C. Fraser, S. Riley, R. M. Anderson, N. M. Ferguson, *Factors that make infectious disease outbreaks controllable*, PNAS **101** (2004), 6146–6151.
[2] J. L. Gerberding, *Faster..., but fast enough? Responding to the epidemic of Severe Acute Respiratory Syndrome*, N Engl J Med **348** (2003), 2030–2031.
[3] C. C Tan, *Public health response: a view from Singapore*, in *SARS: The First New Plague of the 21st Century*, Blackwell Publishing, Oxford (2004).

Causal vaccine effects on binary post-infection outcomes

M. Elizabeth Halloran
(joint work with Michael G. Hudgens)

Evaluation of many effects of prophylactic vaccines on outcomes such as severe disease, death, or transmission to others, condition on being infected. Préziosi and Halloran have estimated the beneficial effects of pertussis vaccination on reducing transmission to others [5] and severe disease [4] in breakthrough cases. On the other hand, concern has been raised that vaccines might have harmful post-infection effects, increasing interest in testing for such effects ([2] and [3]). Conditioning on an event that occurs posttreatment, in our case infection subsequent to assignment to vaccine or control, could result in selection bias, because the people who become infected in the vaccinated group might not be comparable to those who become infected in the unvaccinated group.

In this talk, we consider identifiability and estimation of causal effects of vaccination on binary post-infection outcomes such as transmission to others, severe disease, and death. We use the Frangakis and Rubin [1] approach to define causal effects within principal strata of individuals who have the same joint potential infection values under vaccine and control. We develop a likelihood model to define and to estimate the causal estimands. In general, the causal estimand for post-infection outcomes is not identifiable without unverifiable assumptions. Under the assumption of no selection bias, our causal estimand equals the usual net post-infection efficacy measure. We derive closed forms for the maximum likelihood

estimates of the post-infection causal estimates under the lower and upper bound extreme selection bias models. The bounds depend on the configuration of the data. We show the conditions under which the upper bound can be negative and the lower bound can be positive. We present three methods for sensitivity analysis based on varying assumptions of the selection bias. We analyze data from field studies of a rotavirus vaccine candidate and a pertussis vaccine.

References

[1] C. E. Frangakis and D. B. Rubin, *Principal stratification in causal inference*, Biometrics **58** (2002), 21–29.

[2] P. B. Gilbert, R. J. Bosch and M. G. Hudgens, *Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials*, Biometrics **59** (2003), 531–541.

[3] M. G. Hudgens, A. Hoering and S. G. Self, *On the analysis of viral load endpoints in HIV vaccine trials*, Statistics in Medicine **22** (2003), 2281–2298.

[4] M-P. Préziosi and M. E. Halloran, *Effects of pertussis vaccination on severity: vaccine efficacy for clinical severity*, Clinical Infectious Diseases *37* (2003), 772–779.

[5] M-P. Préziosi and M. E. Halloran, *Effects of pertussis vaccination on transmission: vaccine efficacy for infectiousness*, Vaccine *21* (2003), 1853–1861.

Adverse effects of smallpox vaccination: new analysis of old data

Mirjam Kretzschmar

(joint work with Jacco Wallinga, Peter Teunis, Shuqin Xing, Rafael Mikolajczyk)

With concerns rising that the smallpox virus could be used in a bioterror attack, many countries have been planning for large scale vaccination programs in case an outbreak of smallpox occurs. As of yet, there are no new vaccines available and one has to rely on vaccines that were used during the last century, and especially those that were used during the WHO eradication campaign. Considering the large fraction of unvaccinated people in present societies, the question arises how many people would be expected to suffer from adverse effects and death after smallpox vaccination. There is a lot of information about adverse effects of vaccination scattered in the literature up to the 1970's, the most systematic investigations being those published by Lane *et al.* ([2], [3]) and Neff *et al.* ([4], [5]). In the recent literature an attempt has been made to compare outcomes of those studies with techniques of meta-analysis [1], but this paper was only based on US studies and did not employ sophisticated statistical methods. In our study, we attempt to give a more complete picture of adverse events after smallpox vaccination by including also data from various European countries as published in the (non-English) literature. We use Bayesian methods to analyze the frequency of adverse effects depending on vaccine strain and on age of the vaccinees. Here we report on some preliminary results concerning the occurrence of post-vaccinal encephalitis (PVE) and death after primary vaccination.

Data sources

We conducted a systematic search of the literature using Medline and Social Medicine and Public Health Database (SOMED). For data specifically about Germany we searched the "Bundesgesundheitsblatt" from 1959 to 1985. We extracted data about numbers of primo- and revaccinations, age groups, strains, and all adverse effects, where available. Here we report on our analysis of the data about the occurrence of post-vaccinal encephalitis (PVE) and mortality. Besides the very complete and systematic studies conducted in the USA by Neff and Lane and their coworkers, we were able to extract a considerable amount of information about vaccinations in Germany, although the studies reported there were less systematic and the data sometimes hard to interpret.

In all studies in the USA the vaccine strain New York City Board of Health (NYCBH) was used. In Europe the situation was less clear and many different strains were used in different countries, regions and time periods. In Germany in the 1950's and 1960's the strain Bern was mostly used, also in Austria. In the USSR the strain EM-63, which was derived from the NYCBH strain was used up to 1971. In the UK the strain Lister/Elstree was developed and used. In the Netherlands, the strain Copenhagen was used up to 1962, after that Lister/Elstree was used. In 1968 in an effort to standardize vaccines worldwide and ensure vaccine quality, the WHO recommended that either NYCBH or Lister be used in the worldwide eradication campaign.

Methods

We denote by y the number adverse events given N vaccinations with strain s in the age interval $[la, ua]$. Assume that we have n observations $Y_i = (y_i, N_i, s_i, ua_i, la_i)$, $i = 1, ...n$. We want to estimate parameters $\Theta = (\theta_1, ...\theta_n)$ that describe the probability of occurrence of an adverse event given vaccination at a certain age with a certain strain. We assume that the observed number of adverse events follows a Poisson distribution with mean $\theta_i N_i$ for all i. The likelihood function for Y is then given by

$$L(Y|\Theta) = \prod_{i=1}^{n} f(y_i|\theta_i N_i) \tag{1}$$

where f is the probability density of the Poisson distribution with mean $\theta_i N_i$. We assume that the probability θ consists of a base probability ζ_s that may depend on the vaccine strain and a function $\phi_s(a)$ that describes per strain the effect of age on the probability of an adverse event. As a prior for ζ_s we take a gamma distribution (restricted to the interval $[0, 1]$) with strain dependent parameters:

$$\zeta_s \sim Gamma(r(s), l(s)) \tag{2}$$

The priors for the parameters r and l are chosen such that $\log r$ and the log mean of the gamma distribution are normally distributed. To describe the age effect we define the function ϕ_s as

$$\phi_s(a) = \exp(-\beta_s a) + \frac{\sigma_s a}{a + \rho_s} \tag{3}$$

with strain dependent parameters β_s, σ_s, and ρ_s that have gamma distributed priors. In the estimation procedure, this function is evaluated at the midpoints of the observed age intervals. We analyzed the model using the Winbugs software developed by the MRC and Imperial College, UK [6].

The analysis was conducted in several steps. First, an analysis was conducted to investigate the effect of different vaccine strains on the frequency of PVE or death. Then the analysis was extended to also include the effects of age. Finally, we investigated the question, whether information obtained during a first wave of ring vaccination could be used to judge whether the vaccine strain has increased its virulence in comparison with historical data.

Frequency of PVE and death

For primovaccinations, we find that there are large differences between vaccine strains. As we did not know for all data sets which strains were used for the vaccinations, we considered two different options. One was that the USSR had used EM-63 in the reported vaccinations, and Sweden had used Copenhagen. In an alternative estimation, we assumed that the USSR and Sweden had used Lister/Elstree. In both cases we assumed that EM-63 had been used in GDR, the former East Germany. Independent of those assumptions we find that Bern is by far the most virulent vaccine strain, while NYCBH is the most benign. Lister/Elstree is intermediate. The results for the remaining two strains depend very much on the assumptions.

Also concerning the age effects, the uncertainty for the strains Copenhagen and EM-63 is large, reflecting the scarce data underlying the estimates. Interestingly, we find different age dependent patterns of virulence for NYCBH and Bern. For NYCBH the frequency of PVE is highest in the youngest age group of $0-1$ year olds, then drops to a minimum and rises again for ages > 3. For Bern we see an monotonous increase with age in the frequency of PVE. Those results are already reflected in historical or present vaccination schemes. In Germany and Austria primovaccinations of children were not allowed above the age of 3, while in the present United States vaccination guidelines primovaccination of children under the age of 1 is contraindicated.

We show that with the probability of adverse events being as low as they are, the data collected in a first wave of ring vaccination is not sufficient to revise our historical knowledge about the virulence of a vaccine strain. For NYCBH which would be used for ring vaccination in the US, the expectation is that with vaccination of say 5000 persons in different age groups, no case of PVE or death will be observed. Observation of one or more cases would fall outside the 95% credible interval of the estimates for 5000 vaccinations. For NYCBH we extended the analysis to estimate the mortality after revaccination, so that a total estimate of the expected number of vaccination related deaths in case of a mass vaccination campaign can be obtained.

References

[1] T. J. Aragon, S. Ulrich, S. Fernyak, G. W. Rutherford, *Risks of serious complications and death from smallpox vaccination: a systematic review of the United States experience, 1963 - 1968*, BMC Public Health **3** (2003), 26–37.

[2] J. M. Lane, F. L. Ruben, J. M. Neff, J. D. Millar, *Complications of smallpox vaccination, 1968. National Surveillance in the United States*, NEJM **281** (1969), 1201–1208.

[3] J. M. Lane, F. L. Ruben, J. M. Neff, J. D. Millar, *Complications of smallpox vaccination, 1968: Results of ten statewide surveys*, J Infect Dis **122** (1970), 303–309.

[4] J. M. Neff, J. M. Lane, J. H. Pert, R. Moore, J. D. Millar, D. A. Henderson, *Complications of smallpox vaccination. I. National Surveillance in the United States, 1963*, NEJM **276** (1967), 125–132.

[5] J. M. Neff, R. H. Levine, J. M. Lane, E. A. Ager, H. Moore, B. J. Rosenstein, J. D. Millar, D. A. Henderson, *Complications of smallpox vaccination United States 1963. II. Results obtained by four statewide surveys*, Pediatrics **39** (1967), 916–923.

[6] http://www.mrc-bsu.cam.ac.uk/bugs

Issues in Online Surveillance of Infectious Diseases

Andrew B. Lawson

Counter terrorism surveillance of health events is now a major concern of health surveillance community in the US following 9/11. An example of this is the now annual conference on Syndromic Surveillance sponsored by the New York Academy of Medicine. Online surveillance systems have to tackle a large scale data mining problem: surveying large amounts of health data (in the form of linked time series and disease maps) to attempt to detect aberrations in the record that may suggest evidence of an attack. This problem is particularly important for infectious disease as potential for the use of infectious agents in bioterrorism is clear. In this talk I outline some basic issues in this area, and also hope to highlight the need for research into the statistical issues relating to data mining of complex disease data bases. In particular, I will briefly discuss: data capture and ascertainment, syndromic detection, online process control, spatial imprints (trend, clustering, discontinuity), change detection (before and during an epidemic) and Bayesian optimal surveillance models. For further coverage see [1]

References

[1] A. B. Lawson and K. Kleinman (eds) *Spatial and Syndromic Surveillance for Public Health* Wiley, New York, (2005).

Stochastic Models for Bioterrorist and Naturally Emerging Infectious Disease Threats

Ira M. Longini
(joint work with M. Elizabeth Halloran, Azhar Nizam, Yang Yang, Shufu Xu)

How would we contain a large bioterrorist smallpox attack or the introduction of pandemic influenza into a U.S. metropolitan area? In this talk, we present stochastic microsimulation models for development of control strategies for just such threats. These models are built for population structures based on U.S. census data. We use a graph theoretic approach to construct populations that have connectivity that is statistically similar to a typical U.S. population. The models are calibrated to data from past outbreaks. This includes statistical estimation of key parameters such as household secondary attack rates as well as intervention efficacy such as vaccine and antimicrobial agent efficacy. We will give the particular example of containing the first wave of pandemic influenza. In this case, there would be little or no influenza vaccine available. We will use the model to determine the best use of antiviral agents and compare the effectiveness of such strategies to those involving vaccine. We will present an outline of what we could do to prepare for and contain the next influenza pandemic. We will end the presentation with a discussion of how these models can be integrated in the national effort to control infectious disease threats.

Inference issues for stochastic multitype SIR epidemics among a population of households

Owen D. Lyne
(joint work with Frank G. Ball)

This talk is concerned with a stochastic SIR (susceptible $\rightarrow$ infective $\rightarrow$ removed) epidemic among a closed, finite population partitioned into households that contain several classes of individual. I will discuss inference via pseudo-likelihood for this model and possible approaches to dealing with an identifiability problem which arises. The talk will be illustrated with an application to real data.

The households model discussed here was first analysed in [5]. In [5] the exact distribution of the final outcome of the epidemic is outlined, a threshold theorem is proven and a central limit theorem for the final outcome of epidemics which take off is derived.

To conduct inference for this model from final size data the pseudo-likelihood method uses the independent households model of [1] to generate a likelihood as if the households were independent. Inference then needs to be corrected for the true dependence between households in this model.

Since final size data contains a limited amount of information about between-household infection rate parameters, an identifiability problem arises when there is more than one class of individual. This can be overcome if multiple data sets exist

or if assumptions are made about the form of parameters. This latter approach can be used to analyse the variola minor data of [2].

Vaccination schemes for this model are considered in [3] and [4].

References

[1] C. L. Addy, I. M. Longini and M. Haber, *A generalized stochastic model for the analysis of infectious disease final size data*, Biometrics **47** (1991), 961–74.

[2] J. J. Angulo, *Variola minor in Bragança Paulista County, 1956:Overall description of the epidemic and of its study*, InternationalJournal of Epidemiology **5** (1976), 359–366.

[3] F. G. Ball, T. Britton and O. D. Lyne, *Stochastic multitype epidemics in a community of households: Estimation of threshold parameter R_* and secure vaccination coverage*, Biometrika **91** (2004), 345–362.

[4] F. G. Ball, T. Britton and O. D. Lyne, *Stochastic multitype epidemics in a community of households: estimation and form of optimal vaccination schemes*, Mathematical Biosciences, **191** (2004), 19–40.

[5] F. G. Ball and O. D. Lyne, *Stochastic multitype SIR epidemics among a population partitioned into households*, Advances in Applied Probability **33** (2001), 99–123.

A Game-Theoretical Model in Yellow Fever Vaccination

Eduardo Massad

(joint work with h F.A.B.Coutinho, M.N. Burattini, L.F.Lopez and C.J.Struchiner)

Introduction

Yellow fever (YF) can be prevented by a live attenuated vaccine prepared from the 17D strain of YF virus, that induces seroconversion in more than 95% of recipients and provides immunity for 30 years or longer. YF vaccine has been incorporated into routine vaccination programmes in South America, but in Africa, coverages rates are low. Unfortunately, the vaccine has shown several adverse effects, including mortality in an average rate of 2.5 cases per million doses.

Vaccination policies has ranged from preemptive mass vaccination to post-outbreak ring vaccination. Basing on self-interest, individuals may decide whether to vaccinate during the preemptive mass vaccination campaigns (called vaccinators hereafter), or wait until he/she feel threatened by an outbreak (called delayers hereafter). Unfortunately, when the proportion of delayers is too high, the level of herd immunity achieved may differ from what would be best for the population as a whole.

The Game

Our vaccination game is a population game or nonatomic game, meaning that the payoff to an individual choosing a particular strategy depends on the average behavior of the population. The two basic strategies are "vaccinator" (obtain preemptive vaccination) and "delayer" (decline preemptive vaccination but seek vaccination in the event of an attack). For any strategy, the payoff to an individual

is measured in terms of a cost function for the risks of death due to vaccination and/or a yellow fever outbreak.

To solve the game, we seek a Nash equilibrium strategy. In a population where all individuals play such a strategy, it is impossible for a few individuals to increase their payoffs by switching to a different strategy. Vaccinator cannot be a Nash equilibrium for the reason indicated in the introduction (an individual who chooses the delayer strategy when population coverage is at 100% reaps the benefits of high population immunity without suffering the risk of vaccine complications).

By comparison, delayer can be a Nash equilibrium under certain conditions, such as when the attack risk is sufficiently low or the risk of death due to vaccination is sufficiently high. In other situations, it might be best for some individuals to be vaccinated preemptively and for others to delay. To allow for this we consider mixed strategies whereby individuals choose the vaccinator strategy with probability P ($0 < P < 1$) and the delayer strategy otherwise. If all individuals play the mixed strategy P, then a proportion p = P of the population is preemptively vaccinated.

The Model

$$
\begin{aligned}
\tfrac{dM_s}{dt} &= -caM_sH_i/N_h + (\alpha_M + \mu_M)M_s \\
\tfrac{dM_i}{dt} &= caM_s(t-\tau)H_i(t-\tau)/N_h(t-\tau) - (\alpha_M + \mu_M)M_s \\
\tfrac{dH'_s}{dt} &= -baM_iH'_s/N_h - (\nu_h + \mu_h)H'_s \\
\tfrac{dH^{"}_s}{dt} &= -(\mu_\nu + \mu_h)H^{"}_s \\
\tfrac{dH_i}{dt} &= baM_iH'_s/N_h - (\gamma_h + \mu_h + \alpha_h)H_i \\
\tfrac{dH_v}{dt} &= \nu_hH'_s - (\mu_h + \mu_\nu)H_v \\
\tfrac{dH_r}{dt} &= \gamma_hH_i - \mu_hH_r
\end{aligned}
$$

whose threshold for persistence is given by

$$The = \frac{N_m a^2 bce^{-\mu_m \tau}}{N_h(\gamma + \mu_h + \alpha_h)(\mu_m + \alpha_m)}$$

We simulated the dynamical system above with the following initial conditions

$$\begin{aligned} H_s'(0) &= (1-p)N_h \\ H_s^{"}(0) &= pN_h \\ H_i(0) &= 10 \\ H_v(0) &= 0 \\ H_r(0) &= 0 \\ M_s(0) &= N_m \\ M_i(0) &= 0 \end{aligned}$$

Let us define

$$I = \int_0^\infty baM_iH_s'/N_h dt \rightarrow \phi_s(p) = \frac{I}{(1-p)N_h}$$

$$D = \int_0^\infty (\mu_h + \alpha_h)H_i dt \rightarrow d_s = \frac{D}{(1-p)N_h}$$

$$V = \int_0^\infty \nu_h H_s' dt \rightarrow \phi_v(p) = \frac{V}{(1-p)N_h}$$

$$D_v = \int_0^\infty \mu_\nu H_v dt \rightarrow d_v = \frac{D_v}{(1-p)N_h}$$

The payoff to an individual choosing the vaccinator strategy is

$$E_{vac} = -d_v$$

The payoff to an individual choosing the delayer strategy is

$$E_{del}(p) = -r\left[\phi_s(p)d_s + \phi_v(p)d_v\right]$$

If

$$E_{vac} > E_{del}(0),$$

there is a unique Nash equilibrium p_{ind} $(0 < p_{ind} < 1)$ that can be found by solving for p_{ind} in the equation

$$E_{vac} = E_{del}(p_{ind})$$

If

$$E_{vac} \leq E_{del}(0),$$

then the pure delayer strategy $(p_{ind} = 0)$ is the unique Nash equilibrium.

Now, if p is the proportion of the population that is preemptively vaccinated in campaigns before outbreaks, we can express the expected cost $C(p)$ due to vaccination and potential yellow fever outbreaks as

$$C(p) = pd_v + r(1-p)\left[(d_s - d_v)\,\phi_s(p) + d_v\right]$$

We then minimize $C(p)$ on the unit interval $(0 \leq p \leq 1)$ to determine the group optimum p_{gr}, which is the coverage level that would have to be imposed to minimize the total expected number of deaths.

Preliminary results show that whenver the risk of outbreak is too low and the fatality rate due to the vaccine is too high, the threshold proportion to vaccinate

tends to zero if the threshold condition for the infection establish itself in the population is lower than 3.0.

References

[1] E. Massad, F. A. B Coutinho, M. N. Burattini and L. F. Lopez, *The risk of yellow fever in a dengue-infested area*, Trans. Roy. Soc. Trop. Med. Hyg. **95** (2001), 370–374.

[2] C. T. Bauch, A. P. Galvani and D. J. Earn, *Group interest versus self-interest in smallpox vaccination policy*, Proc. Natl. Acad. Sci. U.S.A. **100** (2003), 10564–10547.

Pair approximations for spatial structures?

Denis Mollison

This work explores the success of pair approximations in capturing local correlations and the spatial structure of population contact networks, especially in respect of the rate of spread of epidemics.

Networks of interest range from the local extreme where interactions are only between nearest neighbours in some low dimensional space, and the infinite-dimensional 'mean-field' extreme where all interact equally with all [8, 4, 5, 9]. Intermediate cases of practical interest include 'small-world' and meta-population models [2, 13, 11].

One of the obvious distinctions between homogeneous mixing and spatial population structures lies in their local correlations: if 'AB' means 'A is a neighbour of B', then $P(AC|AB, BC) \gg P(AC)$ for the spatial case.

Pair approximation differential equations (PAs), that add second order variables such as [SI], the mean number of (S,I) pairs of neighbours to a standard SIR differential equation model [10, 7], have recently been widely used to approximate spatial ecological and epidemic processes [6, 12]. How well do they do this?

There are theoretical reasons why PAs should be better at approximating mean-field than spatial networks. Figure 1 shows how PAs provide excellent approximations to mean-field SIRs for a wide range of the correlation parameter ϕ. In particular, the duration of the epidemic is of order $\log(N)$, where N is the population size.

Now for a spatial SIR with local contacts – Figure 2 shows a nearest-neighbour SIR on a sphere – the duration is of order $\sqrt{N}$, so the PA cannot be expected to approximate this well, as is confirmed by the time plot for the spatial SIR (curve 'S' in Figure 1), which is very different from the PA with the same value of the correlation parameter ($\phi = 0.4$).

There may seem to be a paradox here, in that the spatial network is an element of the set of random graphs $G(N, \phi)$ that have the same number of sites and the same value of the correlation parameter, although members of that set can generally be assumed to be mean-field in character. The resolution of the paradox is that, within $G(N, \phi)$, such spatial or near-spatial networks are of almost infinite improbability, what we might call 'Adams-improbable' [1].

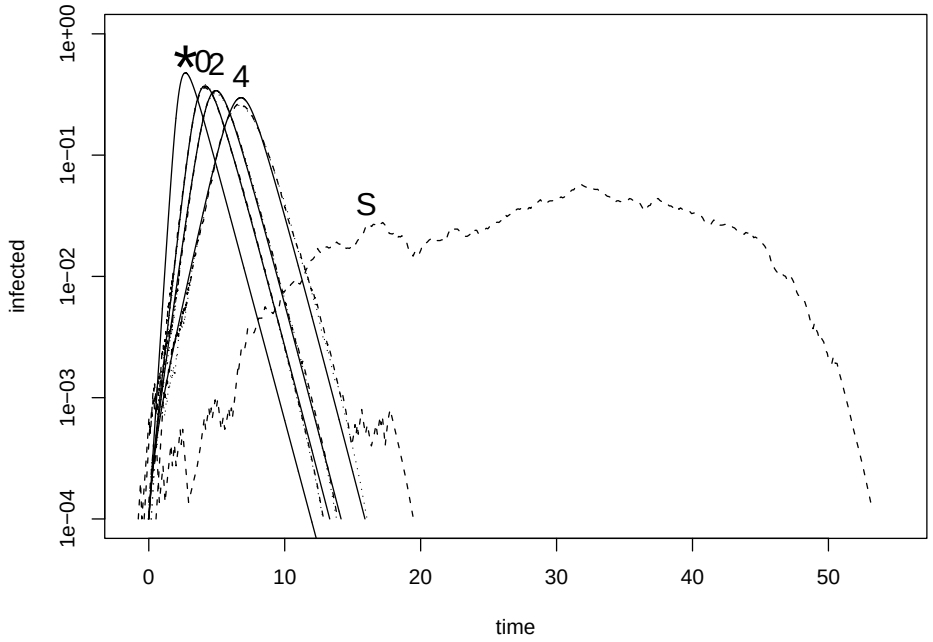

FIGURE 1. Comparison between stochastic SIRs on simple random graphs, constrained to have varying correlation parameter ϕ ($= 0, 0.2, 0.4$; 2 simulations each, dashed curves), and PAs of the same ϕ (solid curves). Also shown are the standard SIR DE ('$\star$') and a simulation of a spatial stochastic SIR ('S') – see Figure 2.

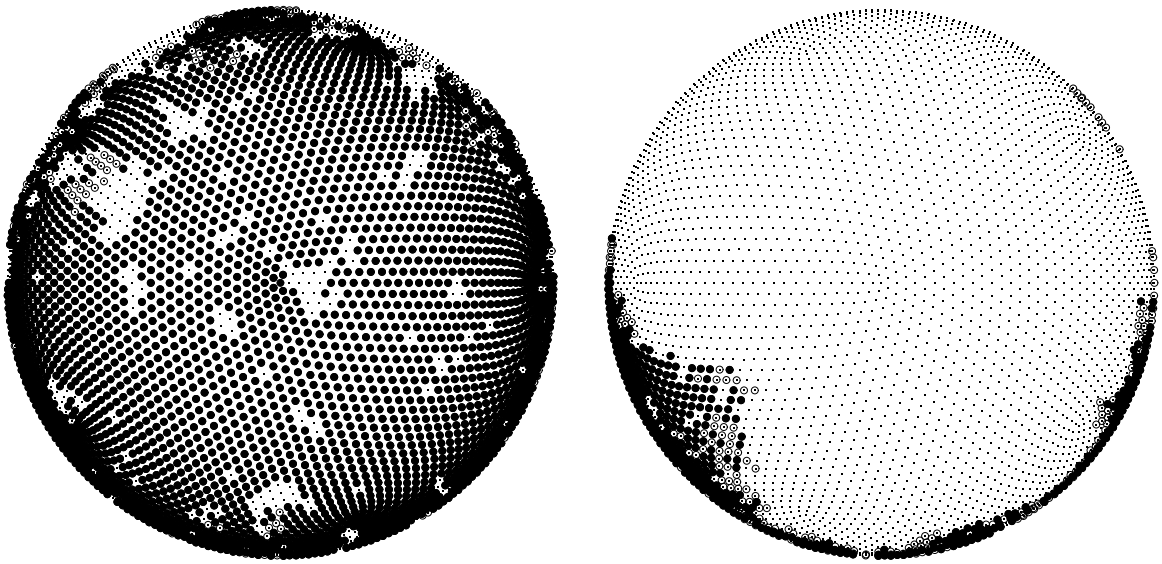

FIGURE 2. Simulation of a nearest-neighbour SIR on a hexagonal$^{(\star)}$ lattice on a sphere: $\cdot$ susceptible, $\odot$ infectious, $\bullet$ removed. This outbreak started at the north pole (left), and has just reached the southern hemisphere (right).
[($\star$) Note: An exact hexagonal lattice on a sphere is not possible; here there are 12 sites that each have only 5 neighbours.]

More broadly, this work in progress tends to support the generalisation that spatial processes need explicit spatial modelling [3, 9].

References

[1] D. Adams, *The Hitchhiker's Guide to the Galaxy* (1979), Pan Books.

[2] F. G. Ball, D. Mollison and G-P. Scalia-Tomba, *Epidemics in populations with two levels of mixing*, Ann. Appl. Prob. **7** (1997), 46-89.

[3] B. Bolker, *Spatial dynamics group report*, in *Ecology of Infectious Diseases in Natural Populations* (eds. B. T. Grenfell and A. P. Dobson), Cambridge UP (1995), 399-420.

[4] A. Barbour and D. Mollison, *Epidemics and random graphs*, in *Stochastic Processes in Epidemic Theory* (eds. J-P. Gabriel, C. Lefèvre and P. Picard), Springer-Verlag (1990), 86-89.

[5] R. Durrett and S. A. Levin, *The importance of being discrete (and spatial)*, Theor. Pop. Biol. **46** (1994), 363-394.

[6] N. Ferguson, C. Donnelly and R. Anderson, *The Foot-and-Mouth epidemic in Great Britain: pattern of spread and impact of interventions*, Science **292** (2001), 1155-1160.

[7] M. J. Keeling, *The effects of local spatial structure on epidemiological invasions*, Proc. R. Soc. Lond. **B 266** (1999), 859-867.

[8] D. Mollison, *Spatial contact models for ecological and epidemic spread*, J. Roy. Statist. Soc. **B 39** (1977), 283-326.

[9] D. Mollison and S. A. Levin, *Spatial dynamics of parasitism*, in *Ecology of Infectious Diseases in Natural Populations* (eds. B. T. Grenfell and A. P. Dobson), Cambridge UP (1995), 384-398.

[10] A. J. Morris, *Representing spatial interactions in simple epidemiological models*, PhD Thesis, Warwick University (1997).

[11] M. E. J. Newman, *Models of the Small World: A Review*, (2000), http://www.arxiv.org/abs/cond-mat/0001118.

[12] D. A. Rand, *Correlation equations and pair approximations for spatial ecologies*, in *Advanced Ecological Theory* (ed. J. McGlade) (1999), 100-142.

[13] D. J. Watts and S. H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature **393** (1998), 440-442.

Contact tracing and stochastic graphs

Johannes Müller

(joint work with Martin Möhle, Mirjam Kretzschmar, Klaus Dietz)

Contact tracing is still not completely understood. Especially how to obtain estimates for parameters from data (of the type that is easily to collect) is not clear at all. There are several attempts made to develop treatable models for contact tracing [2, 5, 6, 3, 4, 1]. Most of them are deterministic or use pair approximation. Here, we follow the stochastic model [6] and develop the analysis further.

We first consider a birth-death process (or stochastic branching process) of independent particles. Particles give birth to children with constant rate β, and die according to some age dependent rate $\mu(a)$. This stochastic process generates a graph: the nodes are the (living) particles while a directed edge goes from mother to daughter. If there is no death, the graph is a tree. Death destroys this connectivity such we are left with a forest. We are interested in the statistics of the

connected components (size and structure). In order to investigate these aspects, we condition on the age of the root of the connected component, and consider the probability measure induced on the set of finite (directed) trees.

Definition 1: *The connected components of the branching process, which have a root with age a, induce a random measure P_b on the set of finite trees.*

We introduce another stochastic process, which we call the uniformly growing tree, that also generates a random measure on the set of finite (directed) trees.

Definition 2: *Let $q \in [0,1)$. A uniformly growing tree is generated by the following algorithm:*

Initialize: Start with the tree consisting of one node.

Step 1: Stop the process with probability q.

Step 2: If the process is not stopped, add a node to the tree in drawing a directed edge from a randomly chosen node that is already there (all nodes have the same probability to be chosen) to the new node. Proceed with step 1 until the process stops.

Let P_{ut} be the probability measure generated by this process.

We now are able to prove two theorems. These theorems are direct generalizations of the case that μ does not depend on age, given in [7]. Also the proof parallels that given there.

Theorem 1: *Consider a connected component generated by the birth-death process and assume that the age of the root is known to be a.*

The size Y_a of the tree is one plus a geometrically distributed random variable,

$$Y_a - 1 \sim Geom(\tilde{q}), \qquad \tilde{q} = \exp\left(-\beta \int_0^a \kappa(\tau)\, d\tau\right),$$

with $\kappa(a) = \exp(-\int_0^a \mu(\tau)\, d\tau)$.

Proof: (Idea) Let Y_a be the random variable that gives the size of a tree with root of age a, X_a denote the number of living children of a particle of age a and $\kappa_a(.)$ denotes the age distribution of living children of a particle of age a. We immediately find the compound equation

$$Y_a = 1 + \sum_{i=1}^{X_a} \int_0^a \kappa_a(b) Y_b\, db.$$

Using this equation, it is possible to work out the proof of the theorem.

□

Theorem 2: *Consider the same situation as in Theorem 1. The birth-death process and uniformly growing trees imply the same random measures on the set of finite trees, $P_b = P_{ut}$ a.s., where $q = \tilde{q}$.*

Proof: (Idea) If we go backward in time and ask for the youngest node, we find in any case (branching process and uniformly growing tree) that any leaf is

the youngest node with the same probability. This leads to a recursive argument that shows the statement.

□

We may now introduce contact tracing. Therefore we assume the death rate to be age independent. In addition to "natural" age we introduce a screening rate σ. With this rate, individuals are detected (they e.g. feel sick and visit the doctor) and become an index case. The index case is removed and every (former) adjoint node has a probability p also to be detected. These secondary detected cases become new index cases, such that we consider a complete recursive process (another variant of the model would e.g. trace at most paths of length n). I.e., we consider a branching process on a branching process. We have been able to prove also in this case, that the probability measure induced by the connected components with a root of given age a coincides with the random measure created by uniformly growing trees (where we need to choose q in an appropriate way).

A still open question is the statistics of size and structure of connected components if we do not condition on the age of the root. In this case, we need to know the Malthusian parameter of the branching process; due to the dependencies of the particles, this parameter is difficult to determine. However, since the process is split into small connected components, there is hope to find asymptotic results for the desired statistics.

References

[1] C. Fraser, S. Riley, R. M. Anderson, and N. M. Ferguson, *Factors that make an infectious disease outbreak controllable* PNAS **101** (2004), 6146–6151.

[2] H. W. Hethcote and J. A. Yorke, *Gonorrhea transmission dynamics and control*, Lecture Notes in Biomathematics **56** (1984).

[3] R. Huerta and L.S. Tsimring, *Contact tracing and epidemics control in social networks*, Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **66** (2002), 056115.

[4] J. Hyman, J. Li and E.A. Stanley, *Modeling the impact of random screening and contact tracing in reducing the spread of HIV*, Math. Biosc. **181** (2003), 17–54.

[5] R. Lounes and H. de Arazoza, *A two-type model for the cuban national programme on HIV/AIDS*, IMA J. Appl. Med. Biol. **16** (1999), 143–154.

[6] J. Müller, M. Kretzschmar and K. Dietz, *Contact tracing in deterministic and stochastic models*, Math. Biosc. **164** (2000), 39–64.

[7] J. Müller and M. Möhle, *Family tries for time-continuous birth-death processes*, J. Appl. Prob. **40** (2003) 980–994.

Interactions between infections

Nico J.D. Nagelkerke

(joint work with Sake J. de Vlas)

Traditionally much of mathematical modelling has been devoted to infectious diseases caused by single pathogens with well-understood aetiology. Problems that have traditionally received most attention are the implications of complex contact patterns and the impact of interventions such as vaccination. In addition, the

statistical methodology of inferring disease parameters (e.g. R_0) and contact patterns from observed epidemiological patterns is developing rapidly. However, the potential of mathematical modelling and statistics in clarifying the aetiology of (infectious) diseases is not yet fully exploited. For example, modelling appears to have played no role in identifying the infectious aetiology of cervix carcinoma, kaposi sarcoma, or peptic ulcer. Perhaps because, as yet, there exist no properly developed study designs to recognize the aetiology of a disease from its epidemiology. An outstanding, rare, example of how mathematics can help unravel the infectious aetiology of disease is given by Griffith, who in 1967 predicted the pathogenesis of Scrapie (and thus also other prion diseases such as Kuru, CJD, and BSE) as an infectious protein. The epidemiology of many other diseases might similarly suggest an infectious aetiology if approached properly. Notably complex are diseases in which several (unknown) infectious agents are involved, or if transmission depends on unrecognised mechanisms, for then disease may seem to be non-infectious or sporadic. In order to make modelling a more useful tool for this purpose, what would seem needed is a comprehensive classification of infectious mechanisms and the development of study designs and methods to recognize such mechanisms from observations.

One of the least explored areas, both by epidemiologists and more notably by mathematicians, is that of interactions between infections. Yet, there are both practical examples of such interactions, and sound biological and immunological arguments to suspect that such interactions are important. As most humans are invariably host to a large variety of infectious organisms, co-infections are the rule rather than the exception. In addition, the immunological memory of one infection may influence the course of subsequent exposure to another pathogen so that interaction is not even restricted to simultaneous infection.

Practical examples of interaction are: 1) the cross immunity that may exist between different competing pathogens; 2) disease enhancement, as in dengue and RSV infections, after prior infection with a different strain; 3) the well-recognized role of viral respiratory infections in facilitating subsequent bacterial disease, and 4) the role of conventional sexually transmitted infections (STIs) in spreading HIV infection, and the role of HIV in facilitating spread of STIs. Many more may exist but may long remain unrecognised. Recently, we proposed the existence of an as yet unidentified STI in the progression from *M.tuberculosis* infection to disease. Even when interaction between two well-recognized infections is suspected, identifying the exact mechanism from observational data is complicated. For example, in a recent study on the effect of bacterial STIs on the risk of HIV infection, a clear correlation between the two was seen, yet prevention of bacterial STIs by antibiotic prophylaxis did not have any impact on HIV incidence.

The usefulness of modelling interacting infections is that it may yield more realistic predictions. Also, apparently "random" events, such as disease caused by a pathogen after infection, can be explicitly modelled in terms of co-infection with another pathogen. While formulating equations for specific combinations of infections poses few problems, the generic properties and conceptual framework of

interacting infections is not yet well developed. For example, R_0 is a more elusive concept. It depends on whether infection A is introduced into a population where infection B is already established, or whether the simultaneous introduction in a susceptible population is considered. Thus, specifying conditions for a pair of infections of becoming endemic is more complex than for a single infection.

References

[1] J. S. Griffith, *Self-replication and scrapie*, Nature **215** (1967), 1043–1044.
[2] N. J. D. Nagelkerke, S. J. De Vlas, K. S. MacDonald, H. L. Rieder, *Tuberculosis and sexually transmitted infections*, Emerg Infect Dis. **10** (2004), 2005–2006.
[3] R. Kaul, J. Kimani, N. Nagelkerke, *Monthly antibiotic chemoprophylaxis and incidence of sexually transmitted infections and HIV-1 infection in Kenyan sex workers: a randomized controlled trial*, JAMA **291(21)** (2004), 2555–2562.

Bayesian inference for final outcome data using Random Graphs

Philip D. O'Neill

We describe methodology for conducting Bayesian inference given final outcome data in a structured multitype population model setting. For illustration we focus on models with two levels of mixing as defined in Ball et al [1]. In such models a population consisting of individuals is partitioned into groups, such as households, and mixing is allowed to occur both within groups and between groups.

Given data on the population structure and the final outcome of an epidemic, the objective is to perform inference for the two mixing rate parameters. However, the problem is complicated by the fact that the likelihood of the final outcome of an epidemic is numerically and analytically intractable for all but very small population sizes. Our approach involves imputing missing information that leads to a tractable (augmented) likelihood. The choice of imputed variable is not immediately obvious in this setting, and we choose a certain random graph that essentially describes the contacts in the underlying population. Such graphs, and in particular their link with the final outcome of stochastic epidemic models, are well-known (e.g. [2]), but they have not previously been exploited for inference purposes.

The approach employs Markov chain Monte Carlo methods, which are natural for this kind of missing-data problem. The methods have several attractive features, including (i) they do not rely on either asymptotics or assumptions about being above threshold (ii) they yield detailed information about the spread of the epidemic and the relative contribution of different types of mixing (iii) they are widely applicable across different models (iv) they naturally allow model choice questions to be addressed. The approach is illustrated with a number of data sets.

References

[1] F. G. Ball, D. Mollison, G. Scalia-Tomba, *Epidemics with two levels of mixing*, Ann. Appl. Prob. **7** (1997), 46–89.

[2] H. Andersson and T. Britton, *Stochastic Epidemic models and their Statistical Analysis*, Springer Lecture Notes in Statistics, Springer: New York **151** (2000).

Invasion of exotic infections

Mick Roberts

The global epidemic of severe acute respiratory syndrome (SARS) in 2003, and concerns that terrorists may use an infectious agent such as smallpox as a weapon, demonstrate the need to determine control strategies for exotic infections. The prior determination of such strategies, and the use of mathematical models to assist this, are hampered by the obvious lack of data. An integral equation model of Kermack-McKendrick type has been used to compare strategies based on the isolation of infectious individuals and targeted vaccination. The model structures the incidence of infection according to the location of exposure, and requires some knowledge of the infectivity kernel and the initial rate of exponential increase of cases. The model has been used to design strategies to minimize the risks of SARS and smallpox in a previously unexposed community [1,2].

The equation for the incidence of infection is (see [3])

$$i_k(t) = \alpha_k(t) + w_k S(t) \sum_{\ell=1}^{4} \int_0^\infty p(\tau) C_{k\ell}(\tau) i_\ell(t-\tau)\, d\tau \tag{1}$$

with subscripts $k = 1, 2, 3, 4$ relating to infections taking place in the home, the workplace (including school), the wider community, and a health-care facility respectively [1]. The function $p(\tau)$ is the probability of transmission given contact at time τ after infection, and $\mathbf{C}(\tau)$ is the matrix of contact rates per host. The index cases are represented by the incidence term $\alpha_k(t)$, and the weights w_k are used to relate the total number susceptible in the population,

$$S(t) = S(0) - \left| \int_0^t i(\tau)\, d\tau \right| \tag{2}$$

to the effective number susceptible at each location.

For an invading infection we can assume that the entire population is initially susceptible, and that the prevalence of infection is small. Hence, approximating $S(t) \equiv N$ (the population size) linearises equation (1) which may then be solved in the Laplace transform domain to obtain

$$\overline{i}(s) = \left(\mathbf{I} - \overline{\mathbf{K}}(s)\right)^{-1} \overline{\boldsymbol{\alpha}} \tag{3}$$

where

$$\left(\overline{\mathbf{K}}(s)\right)_{k\ell} = w_k N \int_0^\infty p(t) C_{k\ell}(t) e^{-st}\, dt \tag{4}$$

The basic reproduction number for the model defined by equation (1) is $R_0 = \rho\left(\overline{\mathbf{K}}(0)\right)$ (the largest eigenvalue of $\overline{\mathbf{K}}(0)$). If $R_0 < 1$ the solution is exponentially bounded, corresponding to a small epidemic. If $R_0 > 1$ a large epidemic occurs:

the incidence of infection increases approximately exponentially with $|i(t)| \sim e^{rt}$ where r is the solution of $\mathbf{I} - \overline{\mathbf{K}}(r) = 0$ with largest positive real part.

A convenient approximation to $p(\tau)$ is the trapezium function

$$p(\tau) = \begin{cases} p_0 \dfrac{\tau - \tau_a}{\tau_b - \tau_a} & : \quad \tau \in (\tau_a, \tau_b) \\ p_0 & : \quad \tau \in (\tau_b, \tau_c) \\ p_0 \dfrac{\tau - \tau_c}{\tau_d - \tau_c} & : \quad \tau \in (\tau_c, \tau_d) \\ 0 & : \quad \text{otherwise} \end{cases}$$

which can be specified using a minimum amount of data, and which with the assumption of constant contact rates leads to an explicit Laplace transform solution. For SARS we used $(\tau_a, \tau_b, \tau_c, \tau_d) = (4, 7, 11, 14)$ days [1], and for smallpox $(\tau_a, \tau_b, \tau_c, \tau_d) = (14, 16, 25, 27)$ days [2]. The relative sizes of the weights w_k were fixed using expert opinion, and their overall magnitudes by assuming $R_0 = 3.3$ for SARS [4] and $R_0 = 3.2$ for smallpox [5].

Upon the introduction of an invading infection $R_0 > 1$, and we obtain the large epidemic solution which may be determined numerically. The only control measure available to contain an emerging infection is the isolation of cases. Suppose that when a time t_q has elapsed after the infection of the index case, a policy is introduced that effectively prevents a fraction $q_{k\ell}(\tau)$ of infecteds of type ℓ from contacting susceptibles at location k at time τ after they were exposed. Then, for times $t > t_q$ equation (1) would become

$$i_k(t) = w_k S(t) \sum_{\ell=1}^{4} \int_0^\infty p(\tau) C_{k\ell}(\tau) \left\{1 - u(t - \tau - t_q) q_{k\ell}(\tau)\right\} i_\ell(t - \tau)\, d\tau \tag{5}$$

where $u(\tau) = 1$ when $\tau > 0$ and zero otherwise. The basic reproduction number in the presence of the isolation policy is $R_q = \rho\left(\mathbf{K}_q\right)$ where

$$\left(\mathbf{K}_q\right)_{k\ell} = w_k N \int_0^\infty p(\tau) C_{k\ell}(\tau) \left\{1 - q_{k\ell}(\tau)\right\}\, d\tau \tag{6}$$

For smallpox a vaccine is also available. If a proportion $v_k(t)$ of transmission were prevented by the vaccine, equation (1) would become

$$i_k(t) = \alpha_k(t) + \left(1 - v_k(t)\right) w_k S(t) \sum_{\ell=1}^{4} \int_0^\infty p(\tau) C_{k\ell}(\tau) i_\ell(t - \tau)\, d\tau \tag{7}$$

If the v_k are constant with time, the basic reproduction number in the presence of vaccination is $R_v = \rho\left(\mathbf{K}_v\right)$ where

$$\left(\mathbf{K}_v\right)_{k\ell} = \left(1 - v_k\right) w_k N \int_0^\infty p(\tau) C_{k\ell}(\tau)\, d\tau \tag{8}$$

The style of model presented here may be less familiar than, for example, a stochastic simulation model used for SARS [6] or compartmental models used for smallpox [7], despite their historical pedigree. Their advantage is that they can be parameterised with a minimal set of data, and can be readily analysed to compare

proposed control strategies. This makes them ideally suited to studies of invading exotic infections.

References

[1] M. G. Roberts, *Modelling strategies for minimizing the impact of an imported exotic infection*, Proc. Roy. Soc., Ser. B (2004), in press.

[2] G. K. Aldis, M. G. Roberts, *A model for the control of an outbreak of smallpox*, Math. Biosci., submitted.

[3] O. Diekmann, J. A. P. Heesterbeek, *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, Chichester: Wiley (2000).

[4] J. Wallinga, P. Teunis, *Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures*, Am. J. Epidemiol. **160** (2004), 509–516.

[5] M. E. Halloran, I. M. Longini, A. Nizam, Y. Yang, *Containing bioterrorist smallpox*, Science **298** (2002), 1428-1432.

[6] S. Riley, C. Fraser, C. A. Donnelly , A. C. Ghani, L. J. Abu-Raddad, A. J. Hedley, G. M. Leung, L-M. Ho, T-H. Lam, T. Q. Thach, P. Chau, K-P. Chan, S-V. Lo, P-Y. Leung, T. Tsang, H. Ho, K-H. Lee, E. M .C. Lau, N. M. Ferguson, R. M. Anderson, *Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions*, Science **300** (2003), 1961-1966.

[7] N. M. Ferguson, M. J. Keeling, W. J. Edmunds, R. Gani, B. T. Grenfell, R. M. Anderson, S. Leach, *Planning for smallpox outbreaks*, Nature **425** (2003), 681-685.

Identification of reservoirs of infection

Mick Roberts

(joint work with Hans Heesterbeek)

We have formulated a new threshold quantity for the analysis of the epidemiology of infectious diseases, which we refer to as the *type reproduction number* (T) [1,2]. The quantity is similar in concept to the familiar basic reproduction number (R_0), but it singles out particular host types instead of providing a criterion that averages over all host types. The calculation of T enables us to identify the long-term effects of control strategies for particular sub-groups of the population, and to estimate the level of control necessary when targeting a subset of host types. This identifies those host types that constitute a reservoir for infection, and hence must be targeted for successful eradication of the infection.

There are many situations where control effort is targeted at particular host types. Consider for example vector-transmitted infections such as malaria or dengue. If each infected human infects K_{21} mosquitoes, and each infected mosquito infects K_{12} humans, then the next generation matrix [3] is

$$\mathbf{K} = \begin{pmatrix} 0 & K_{12} \\ K_{21} & 0 \end{pmatrix} \tag{1}$$

This has spectral radius $R_0 = \sqrt{T} = \sqrt{K_{12}K_{21}}$. If a vaccine were available, the proportion of humans that would need to be protected should exceed $1 - 1/R_0^2$ which is equal to $1 - 1/T$. Both R_0 and T are valid threshold quantities, and in the literature we find both $R_0 = \rho(\mathbf{K})$ and $R_0 = \rho(\mathbf{K}^2)$.

The mosquito *Aedes albopictus* can transmit dengue transovarially. This modifies the next generation matrix to

$$\mathbf{K} = \begin{pmatrix} 0 & K_{12} \\ K_{21} & K_{22} \end{pmatrix} \tag{2}$$

with $K_{22} < 1$. The value of T is found by summing over all possible paths from host type 1 (human) to type 1, hence

$$T = K_{21}K_{12} + K_{21}K_{22}K_{12} + K_{21}K_{22}^2K_{12} + \ldots \tag{3}$$

$$= \frac{K_{21}K_{12}}{1 - K_{22}} \tag{4}$$

whereas

$$R_0 = \rho(\mathbf{K}) = \frac{1}{2}\left(K_{22} + \sqrt{K_{22}^2 + 4K_{21}K_{12}}\right) \tag{5}$$

If a proportion v of humans were effectively vaccinated, then K_{12} would be multiplied by $1 - v$ and $\rho(\mathbf{K}) < 1$ iff $v > 1 - 1/T$. If a proportion w of humans were prevented from transmitting infection to mosquitoes, then K_{21} would be multiplied by $1 - w$ and $\rho(\mathbf{K}) < 1$ iff $w > 1 - 1/T$. Hence there is a simple relationship between the control effort required to achieve eradication and T, but no similar relationship exists with R_0.

The concept described above may be generalized to an arbitrary number of host types with a focus on any subset of them. We start with an $n \times n$ next generation matrix $\mathbf{K}$, and focus our attention on ℓ host types. Defining an $n \times n$ projection matrix $(\mathbf{P}_\ell)_{ii} = 1$ for $i = 1 \ldots \ell$, $(\mathbf{P}_\ell)_{ij} = 0$ otherwise, and an $n \times \ell$ projection matrix $(\mathbf{E}_\ell)_{ii} = 1$ for $i = 1 \ldots \ell$, $(\mathbf{E}_\ell)_{ij} = 0$ otherwise, we can collapse the transmission of infection onto the selected ℓ host types by

$$\mathbf{M}_\ell = \mathbf{E}_\ell' \mathbf{K} \left(\mathbf{I} - (\mathbf{I} - \mathbf{P}_\ell)\mathbf{K}\right)^{-1} \mathbf{E}_\ell \tag{6}$$

and define $T = \rho(\mathbf{M}_\ell)$. We have proved [1] that $R_0 = \rho(\mathbf{K}) < 1$ iff $T = \rho(\mathbf{M}_\ell) < 1$.

Equation (6) only makes sense if $\rho((\mathbf{I} - \mathbf{P}_\ell)\mathbf{K}) < 1$. The matrix $\mathbf{M}_\ell$ is the next generation matrix for the ℓ host types, with infection pathways that pass through the other $n - \ell$ host types condensed and summed. The matrix $(\mathbf{I} - \mathbf{P}_\ell)\mathbf{K}$ is the next generation matrix for the other $n - \ell$ host types, but with infection pathways that pass through the first ℓ host types deleted. Hence if $\rho((\mathbf{I} - \mathbf{P}_\ell)\mathbf{K}) > 1$ these $n - \ell$ host types are capable of maintaining the infection without the presence of the ℓ targeted host types, and hence they constitute a reservoir of infection. In biological terms, if one of the first ℓ host types is infected it will transmit infection to this group of $n - \ell$ host types, where the infection will persist and continue to be transmitted back into the first ℓ types indefinitely.

In conclusion, the type reproduction number T has the same threshold property as R_0, but can be used directly to calculate the critical control effort required for eradication. This result generalises to ℓ dimensions, and provides threshold criteria for control measures. As a by-product, our methodology identifies particular groups of host types that constitute a reservoir of infection.

References

[1] M. G. Roberts, J. A. P. Heesterbeek, *A new method for estimating the effort required to control infectious diseases*, Proc. Roy. Soc., Ser. B **270** (2003), 1359–1364.

[2] J. A. P. Heesterbeek, M. G. Roberts, *The type-reproduction number T in models for infectious disease control*, Math. Biosci. (2005) in press.

[3] O. Diekmann, J. A. P. Heesterbeek, *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, Chichester: Wiley (2000).

Using an individual-based model to study the spread of infectious diseases among Canadian fur-trapping populations

Lisa Sattenspiel

(joint work with Connie Carpenter)

Models that incorporate realistic contact structures are being used increasingly to study the impact of social interactions on the spread of infectious diseases within and among populations. The majority of such models have used a differential equations framework, but because of the small size of many real social groups in humans and other animals, individual-based models are being used more frequently to address this question. I compare here results obtained using both modeling approaches to study the same epidemic scenario. My models focus on the impact of the 1918-19 influenza epidemic among native Canadian fur trappers in central Manitoba.

The differential equations model, described in [1], divides the population into three discrete groups corresponding to three HudsonÕs Bay Company (HBC) trading posts, and allows individuals to move among these communities in the course of their daily activities. Ethnohistoric analyses of parish and HBC post records have been used to estimate the mobility parameters of the model and to provide information on the social context within which disease transmission occurred. Results from simulations of this model clearly point out the importance of social interactions within the communities in determining the overall impact of an epidemic (see, for example, [2]).

Ethnographic data, however, indicate that at the time of the epidemic during the winter of 1918, the real populations were not clustered into these three communities; rather, the bulk of the population was dispersed across the landscape into much smaller camps that averaged around 15 individuals. During the summer, the family groups were aggregated at the HBC posts, but the largest of the study communities numbered at most only about 750 individuals. This has made it essential to move towards the development of an individual-based stochastic model in order to progress further in understanding how diseases spread in the study communities. Consequently, recent research has centered on the development of such a model, which incorporates a more realistic wintertime structure with satellite camps linked to a larger cluster representing one of the HBC posts and a summertime structure with all family groups aggregated at the posts.

Because the individual-based model considered only one of the three HBC posts modeled in the ODE system, direct comparisons of the results of simulations of the individual-based model are relatively limited. One very marked difference relates to the impact of seasonal changes in social structure and mobility patterns. Results from the ODE model suggested that seasonal variation in the rates and patterns of travel had little or no effect on the size of epidemics, even though it could significantly affect the timing of epidemic spread. Results from the individual-based model were quite different. In these simulations, winter and summer epidemics differed markedly from each other, with summer epidemics earlier, shorter, and much more intense than winter epidemics. This result is a consequence of the summer aggregation and winter dispersal of the study population, features of the real population that could not be adequately captured by the ODE model.

Further analysis of variation in the parameters of the individual-based model indicated that simulated winter epidemics were sensitive to variation in nearly all parameters. On the other hand, simulated summer epidemics were not sensitive to variation in probability of traveling to the post on any given day, distance between the camps and the fort, distribution of the population between the fort and the camps, and the age-sex composition of the population. The first three factors had no effect because the summer model assumed that the entire population was aggregated at the post; hence, these parameters were not relevant for the summer population structure. The fourth parameter, age-sex composition, had a significant impact on winter epidemics because only men traveled from the camps to the post, but again because of the summer population structure, the impact of differences in the age-sex composition of the population would be minimal at that time.

Results of this work clearly emphasize the basic insight derived from the ODE model – that patterns of social interaction at a local scale can have major impacts on the spread of infectious diseases within and among populations. Results also point out, however, that previous conclusions about how population mobility affects the spread of infectious disease epidemics must be modified to reflect new insights derived from more realistic models.

References

[1] L. Sattenspiel and K. Dietz, *A structured epidemic model incorporating geographic mobility among regions*, Mathematical Biosciences **128** (1995), 71–91.

[2] L. Sattenspiel and D. A. Herring, *Structured epidemic models and the spread of influenza in the central Canadian subarctic*, Human Biology **70** (1998), 91–115.

The Data Analysis Plan for the New Zealand Meningitis B Vaccination Programme

David J. Scott

(joint work with Shanthi Ameratunga, Alex Macmillan, Diana Lennon, Joanna Stewart, Sue Crengle, Kim Mulholland, Mick Roberts)

In this talk I gave an outline of the data analysis plan for the New Zealand Meningitis B Vaccination Programme [1] prepared by the Vaccine Effectiveness Evaluation Team, of which I was a member. An account of this plan was recently presented at the Fourth World Conference on Vaccines and Immunisation in Tokyo, Japan by the lead author, Shanti Ameratunga. The paper is to appear in a special issue of the journal *Vaccine*, see [2]. The following is the abstract from that paper.

A nationwide strategy to control a group B meningococcal disease epidemic in New Zealand using an epidemic strain-specific vaccine (MeNZB^{TM}) commenced in 2004. In the absence of randomised controlled trials investigating the efficacy of this particular vaccine, a complement of observational methods are planned to evaluate the post-licensure effectiveness of this vaccine strategy. The two main approaches involve a Poisson regression model investigating the overall impact of the MeNZB^{TM} programme on disease rates over time capitalising on detailed population-based disease surveillance data and the staged roll-out of the vaccine campaign, and a case-control study that aims to estimate vaccine effectiveness in pre-school children. The studies are designed to minimise the potential biases inherent in all observational methods and provide critical data on the effectiveness of a major public health intervention.

References

[1] S. Ameratunga et al., *Post-licensure Effectiveness Evaluation of the Meningococcal B Vaccine in New Zealand (MeNZB^{TM}): Data Analysis Plan*, Report prepared for the Meningococcal Management Team by the Department of Paediatrics, University of Auckland, (2003), 65pp.

[2] S. Ameratunga, et al., *Evaluating the post-licensure effectiveness of a Group B meningococcal vaccine in New Zealand: a multi-faceted strategy*, Vaccine, to appear.

Estimating the duration of common persistent infections

Tom Smith

(joint work with Wilson Sama)

We consider the general problem of estimating the average duration of parasitic infection when (i) infection is frequent (ii) both infection and clearance processes may depend on acquired immunity (iii) the means of detection has an unknown sensitivity less than 100%, which may also depend on acquired immunity. (iv) Repeated determinations of parasitological status are available for a panel of exposed individuals. This problem arises in studies of bacterial carriage in the upper respiratory tract, and in *Plasmodium falciparum* malaria. In both these cases panel

datasets are available where infection status was determined at regular intervals for a sample of exposed hosts.

We have examined the following approaches:

(i) If the pathogen population is treated as homogeneous, estimates of duration of infection can be made by fitting hidden Markov models (HMM) where the hidden underlying dynamics are described by the 'catalytic' model (corresponding to Ross' original malaria model):

$$\frac{dp}{dt} = \lambda(1-p) - \mu p$$

where p is the probability of being infected at time (or age) t, λ is the force of infection, μ is the clearance rate, and infections are detected with some probability s.

Extensions using alternative survival models, and/or including effects host heterogeneity and of age or immunity are all possible but the parameter estimates are very sensitive to details of the model and are often not in a plausible range. The same model for the underlying process can give very different results if it is fitted to data summarised cross-sectionally or longitudinally.

(ii) When serological or molecular typing data are available, an extension is to explicitly model heterogeneity in the infection process between parasite types, (the multiple-strain SIS model) by adding random effect terms to the HMM. This achieves identifiability but leads to highly parameterised models requiring Markov chain Monte Carlo approaches for fitting [2].

(iii) An alternative is the infinite-strain SI(R/S) model, in which each infection event is assumed to correspond to a different strain and is hence distinguishable. The number of concurrent infections in a host (n) can then be described by:

$$\frac{dn}{dt} = \lambda - \mu n$$

corresponding to the malaria model intended by Macdonald [1]. We assume an analogous observation process to that in (i). Extensions of this model are less highly parameterized than those of the multiple-strain SIS model. We consider whether this makes it possible to identify the effect of acquired immunity on infection duration, when immunity is also known to affect the detection process.

References

[1] G. Macdonald, *The Epidemiology and Control of Malaria*, Oxford University Press, London (1957).

[2] T. Smith and P. Vounatsou, *Estimation of infection and recovery rates for highly polymorphic parasites when detectability is imperfect, using hidden Markov models*, Stat. Med. **22** (2003), 1709–1724.

Population Dynamics of Transposable Elements: Copy Number Regulation and Species Invasion Requirements

Claudio J. Struchiner
(joint work with Margareth G. Kidwell, José M.C. Ribeiro)

Transposable elements (TE) are sequences of DNA that invade animal and plant species genomes over a period of several generations. Class I TE's are retroviral-like elements that transpose by means of a reverse transcriptase, while class II TE's transpose by means of a transposase enzyme. In general, class I TE's have long (> 100 base pairs terminal repeats), while class II TE's have relatively short (ca. 30 bp) inverted repeat sequences. Both element types can insert in different places within a host genome (transposition), changing location and/or increasing copy number during successive generations. After several generations, copy number may stabilize to a few (< 10) or to as much as several thousand copies, depending on the element type and host factors. Stabilization of copy number is associated with lack of transposition of the TE at the individual level, and with total invasion of the host species by the TE at the population level (i.e., every member of the species carries the element). Eventually, mutation of most, or all, element copies leads to degeneration, and only remnants of previous invasions are detected in the form of degenerate sequences of reverse transcriptases, transposases or inverted repeats [2]. Perpetuation of the TE family requires invasion of another host species before total loss of the potential to transpose occurs within the original host species [4].

While transposing within their host genome, TE's may disrupt essential gene functions (e.g. by TE insertion within the coding region of a gene) and thus decrease host fitness. Although increase in fitness may follow transposition events, such instances are considered rare. Transposition of P elements in Drosophila may cause from 1% to 30% lethality per transposition event. Mean homozygous fitness, viability and fertility were reduced by, respectively, 55, 28 and 40% in chromosome lines invaded by P elements. Despite decreasing their hosts' fitness, such elements fix in sexual populations because they increase their representation in their hosts' gametes to a greater extent than they decrease their hosts' fitness due to transposition. For example, the F_1 generation of a cross between an individual containing TE's and one not containing TE's would be expected to have 50% of the gametes containing TE's, but due to transposition of the element, 70% of the gametes could contain one or more copies of the element, thus having a 20% (70-50%) advantage. In this example, if a total reduction in fitness of 19% occurred, the element would still have a 1% advantage and would fix in the host population [4]. Successful elements thus "cheat" the meiotic segregation laws more than they harm their hosts.

The phenomenon of transposition thus creates conflicting results by generally decreasing host fitness and by increasing the frequency of TE-bearing gametes. If transposition were to remain unregulated, eventually a majority of the host genome would consist of TE's, and indeed some organisms have more than 50% of their genome constituted by TE's. This high frequency is not found, however, with

most organisms, indicating the existence of regulatory mechanisms that limit TE copy number. Several models have attempted to describe the population dynamics of TE's and copy number regulation [1], but such models usually ascribe a fixed and limited maximum number of sites that could be either vacant or occupied by TE's, thus limiting the generality of the model. We here present a general model of the population dynamics of TE's based on the transposition efficiency of the TE, the burden it causes on host fitness per transposition event, and on the modification of the transposition efficiency as a function of copy number. Initial conditions for successful species invasion are also simulated. The present model is deterministic, producing an "average" interpretation of the phenomena.

A population dynamics model of the spread of transposable elements in sexually reproducing populations is presented. The population is modeled by a 3 parameter equation describing host reproductive capacity, population size and the strength of the density dependence, while transposable element dynamics were modeled based also on 3 parameters, the maximum ability of the element to copy itself in the absence of regulation (T_0), the regulatory effect of copy number decreasing transposition ($C_{0.5}$), and the deleterious effect of each new transposition on host fitness (d). The model is general because no assumptions about the mechanism of transposition control are made, except that this is some function of copy number. Results indicate that non-regulated elements cannot fix in host populations, and that prediction of stable copy number following successful invasion is mainly a function of the combination of T_0 and $C_{0.5}$ values. Fitness reduction does not affect the final copy number after successful invasion of the element. Fitness reduction, however, will affect the surface of the $\{T_0 \times C_{0.5}\}$ parameter space leading to successful invasion of the transposable element. Invasion of host populations by eight or more individuals containing elements with appropriate parameters will lead to successful element fixation at any size of the host population. A small area of the $\{T_0 \times C_{0.5}\}$ parameter space indicates the possibility of host population extinction due to the invasion of transposable elements. The results are robust as indicated by their insensitivity to host population dynamics parameters, or the shape of the functions defining regulation of transposition.

References

[1] B. Charlesworth and C. H. Langley, *The population genetics of Drosophila transposable elements*, Annu. Rev. Gen. **23** (1989), 251–287.

[2] J. B. Clark and M. G. Kidwell, *A phylogenetic perspective on P transposable element evolution in Drosophila*, Proc. Natl. Acad. Sci. USA. **94** (1997), 11428–11433.

[3] D. A. O'Brochta, *Transgenic mosquitoes: the state of the art*, In Ecological aspects for application of genetically modified mosquitoes, edited by W. Takken, and T.W. Scott. Kluwer, Dordrecht, The Netherlands, (2003).

[4] J. M. C. Ribeiro and M. G. KIDWELL, *Transposable elements as population drive mechanisms: Specification of critical parameter values*, J. Med. Entomol. **31** (1994), 10–16.

Dynamic population models and invasive pneumococcal diseases

Stefan Wagenpfeil
(joint work with B. Hellriegel)

In general, children under the age of five and elderly people with a weak immune system are especially susceptible to pneumococcal infections. For prevention different vaccines exist for children and adults. Since the year 2000 there is a general vaccination recommendation against streptococcus pneumoniae in the USA for young children aged 2 years or less. In Germany there is no general recommendation in favour of vaccinating young children against streptococcus pneumoniae in contrast to elderly aged 60 years or more. To explore possible consequences for the most severe invasive pneumococcal disease, pneumococcal meningitis, we analysed different virtual vaccination strategies using a stochastic and a deterministic age-structured population model. Special attention was paid to possible effects of herd immunity. A compartment model is used to describe the course of infection in a target population. We defined eight age classes (0-2, 3-4, 5-12, 13-20, 21-39, 40-64, >65): young children to be vaccinated or not, the age group of older brothers and sisters, school-aged children, teenagers, parents with small children or adults, parents with adult children and grandparents. Directed transitions model changes in the numbers of a specific compartment like newborns with and without maternal antibodies, of susceptible, infected and vaccinated individuals. Failure of immunization by the vaccine is also accounted for. Each person could undergo at most one transition within a prespecified time period of half a year. The stochastic model is individual-based and derived from a previous deterministic version [1]. Deterministic population-based approaches are common tools for analysing infectious diseases like varicella [2], smallpox [3], etc. Halloran et al. [4] present an alternative modelling approach to smallpox using a stochastic simulator. Whitaker and Farrington [5] extend classical deterministic approaches allowing for varying contact rates in the example of varicella.

Initial values for the comparments and transition rates in our model came from expert knowledge and the literature. Demographic variables came from Statistical Yearbooks for the respective population. There are 380 cases of invasive pneumococcal diseases in Germany each year, in which 140 occur under the age of two years, for example.

Data for the situtation in the USA are much better available [6]. After validating the model with data of invasive pneumococcal diseases like sepsis, bacteraemia or meningitis in the USA, we used it to simulate the dynamics of pneumococcal meningitis cases in Germany using German demographic parameter values.

Results indicate that vaccinating children under 2 years of age with a coverage rate of about 80% could lead to a 75% reduction of disease cases within 30 years. The accuracy of these results is indicated by 95% confidence intervals from stochastic simulations not to be obtained from deterministic modelling approaches. Furthermore there seem to be some effects of herd immunity.

The stochastic model leads to a faster initial decrease in the number of infected people but levelled off much earlier and at higher numbers than the comparable deterministic model. Further time series data are needed in order to refine the models. An interesting and stimulating discussion on principle differences between stochastic and deterministic population models is contained in [7].

References

[1] S. Wagenpfeil, *Dynamische, altersstrukturierte Populationsmodelle zur Simulation von Erkrankungszahlen und Impfstrategien am Beispiel von Pneumokokken-Erkrankungen*, Abstract 2003; 49th Biometrical Colloquium, Wuppertal.

[2] M. E. Halloran, L. Watelet, C. J. Struchiner, *Epidemiologic effects of vaccines with complex direct effects in an age-structured population*, Mathematical Biosciences **121** (1994), 193–225.

[3] E. H. Kaplan, D. L. Craft, L. M. Wein, *Emergency response to a smallpox attack: The case for mass vaccination*, Proc.Natl.Acad.Sci.U.S.A. **99** (2002), 10935–10940.

[4] M. E. Halloran, Jr. I. M Longini, A. Nizam, Y. Yang, *Containing Bioterrorist Smallpox*, Science **298** (2002), 1428–1432.

[5] H. J. Whitaker, C. P. Farrington, *Infections with varying contact rates: Application to varicella*, Biometrics **60(3)** (2004), 615–623.

[6] Whitney et al., ISPPD (3rd International Symposium on Pneumococci and Pneumococcal Diseases), May 5-8 (2002), Anchorage, Alaska.

[7] M. E. Halloran, Jr I. M. Longini, *Containing Bioterrorist Smallpox: Response*, Science **300** (2003), 1503–1504.

Real-time tracking of reproduction numbers during epidemic outbreaks

Jacco Wallinga

(joint work with Peter Teunis)

The reproduction number R measures the transmission potential of an infectious disease. We are interested in inferring the temporal pattern of the effective reproduction numbers R_t from the observed time series of symptom onset dates of reported cases. In an ideal world with perfect information we could assess R_t by simply counting the number of cases that have been infected by cases who developed symptoms on day t. Here we show how this simple trick of counting secondary infections also can be applied when we only have information on the generation interval.

The generation interval is the duration between symptom onset of a secondary case and symptom onset of its primary case. We denote the distribution of the generation interval by $g(x|\theta)$, where x refers to the generation interval and θ to the parameters of the distribution. We denote by $y(t)$ the number of cases with symptom onset on day t. The expected reproduction number for cases with

symptom onset on day t is

$$E(R_t) = \sum_{u=0}^{\infty} \frac{y(u)g(u-t|\theta)}{\sum_{v=0}^{\infty} y(v)g(u-v|\theta)}. \tag{1}$$

In this equation we use several variables that refer to time of onset of first symptoms: t refers to the symptom onset date of the cases we are interested in, u to the symptom onset dates of their possible secondary cases, v to the symptom onset dates of all plausible parents of these secondary cases. The equation is obtained by integrating over the likelihood of all possible transmission trees, given the observed dates of symptom onset [1].

We have applied this likelihood-based estimation equation to available data for SARS outbreaks in Hong Kong, Vietnam, Singapore and Canada in 2003. The effective reproduction numbers reveal that epidemics in the various affected regions are characterized by a markedly similar transmission potential of the disease and similar effectiveness of control measures. For controlling SARS outbreaks, the timely alert has been essential: delaying the institution of control measures by one week would have nearly tripled the epidemic size and would have increased the expected epidemic duration by four weeks [1].

The estimating equation for the reproduction number generalizes other estimating equations for the reproduction number that are based on the exponential growth rate r of an epidemic. To show this, We consider the special case where the number of cases increases exponentially over time with exponential growth rate r

$$y(t) = y(0)e^{rt}.$$

We substitute this exponential growth of new cases into the above equation for the reproduction number, and obtain, after rearranging,

$$R_t = \frac{1}{\sum_{\tau=0}^{\infty} e^{-r\tau} g(\tau|\theta)} = \frac{1}{M_{g(\tau|\theta)}(-r)}.$$

where M indicates the moment generating function of the distribution of generation intervals. This estimating equation coincides, as it should, with the relation between reproduction number R and exponential growth rate r as known from the standard renewal equation for incidence of infection $i(t)$ with a serial interval $g(\tau)$ and reproduction number R [2, 3]:

$$i(t) = \int_{\tau=0}^{\tau=\infty} i(t-\tau)g(\tau)R$$

and substituting exponential growth $i(t) = ce^{rt}$ yields

$$1 = \int_{\tau=0}^{\tau=\infty} e^{-r\tau} g(\tau) R$$
$$1 = R \int_{\tau=0}^{\tau=\infty} e^{-r\tau} g(\tau)$$
$$1 = R M_\tau(-r)$$
$$R = \frac{1}{M_{g(\tau)}(-r)}$$

if $M(-r)$ exists.

When we have a SEIR model where λ_1 is the rate of becoming infectious and λ_2 is the recovery rate, the distritubion of the generation interval is

$$g(x|\theta) = \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1}(e^{-\lambda_1 x} - e^{-\lambda_2 x})$$

with mean $\frac{1}{\lambda_1} + \frac{1}{\lambda_2}$, and the moment generating function is

$$M(z) = \frac{\lambda_1}{\lambda_1 - z} \frac{\lambda_2}{\lambda_2 - z}$$

provided that $r > -\lambda_1, \lambda_2$, and the resulting relation between reproduction number and exponential growth rate is

$$R = (1 + \frac{r}{\lambda_1})(1 + \frac{r}{\lambda_2}),$$

which is exactly the same equation as used by Lipsitch et al [4]. The estimating equation for $E(R_t)$ mentioned above generalizes this equation by removing the restrictive assumption of an exponential growing number of cases.

Our methodological contribution toward real-time analysis of incoming notifications is to remove the need for restrictive assumptions such as exponential growth, and to provide a few simple computational steps for transforming the time series of cases into a time series of estimated values for the instantaneous reproduction number on each day. The transformation is from calendar time to generation time and from number to relative increase in number. Perhaps surprisingly, such a transformation is in many cases computationally trivial. This opens up perspectives for real-time estimation of reproduction numbers from incoming case notifications of any emerging infectious disease [5, 6].

References

[1] J. Wallinga, P. Teunis, *Different epidemic curves for Severe Acute Respiratory Syndrome reveal similar impacts of control measures*, American Journal of Epidemiology **160** (2004), 509–516.

[2] L. I. Dublin, A. J. Lotka, *On the true rate of natural increase, as exemplified by the population of the United Satates, 1920*, Journal of the American Statistical Association **150** (1925), 305–339.

[3] W. Feller, *On the integral equation of renewal theory*, Annals of Mathematical Statistics **12** (1941), 243–267.

[4] M. Lipsitch et al., *Transmission dynamics and control of severe acute respiratory syndrome*, Science **300** (2003), 1966–1970.
[5] M. Lipsitch, C. T. Bergstrom, *Invited Commentary: Real-time tracking of control measures for emerging infections*, American Journal of Epidemiology **160** (2004), 517–519.
[6] J. Wallinga, P. Teunis, *respond to "Real-time tracking of infection control measures"* , American Journal of Epidemiology **160** (2004), 520.

A study of real time fatality rate for an emerging disease: a case for severe acute respiratory syndrome (SARS) in Hong Kong

Paul S.F. Yip
(joint work with KF Lam, Eric Lau, Pui-Hing Chau, Kenneth W Tsang and Anne Chao)

Severe acute respiratory syndrome (SARS) is a highly contagious and severe atypical pneumonia caused by a new coronavirus, and is predominately transmitted by droplets [3]. SARS has rapidly spread worldwide and there were totally 8098 reported cases with 774 fatalities as at July 31, 2003 (WHO, 2003). An ongoing controversial topic is on the estimation of the fatality rate of SARS. The usual definition of the fatality rate adopted by the World Health Organization (WHO, 2003) is the ratio of the cumulative number of deaths to the cumulative number of infected persons in the course of the epidemic. When the outbreak was not over and there were patients still in hospitals over the course of the epidemic, the WHO estimate in the midst of the epidemic assume implicitly that all current SARS inpatients would eventually recover, it therefore tends to under- or over-estimate the final case fatality rate which really depend very much on the stage and the progress of the outbreak and the outcome of the inpatients.

In contrast to the constant case-fatality rate [2], a new fatality rate, termed the 'real-time' fatality rate using a competing risk model with counting process, is proposed in this paper for monitoring the new emerging epidemic on a population level. Furthermore, it can provide information on the efficacy of a certain treatment and management for the disease. A competing risk model via counting process is used to estimate the real time fatality rate in SARS epidemic [1, 5, 6]. It can capture and reflect the time varying nature of the fatality rate over the course of the outbreak in a more timely and accurate manner. The method has been applied to the SARS data in Hong Kong (see Figures 1). The magnitudes and patterns of the estimated fatalities are more or less the same except in Beijing which has a lower rate. It is speculated the effect is linked to the different treatment protocols used. The standard estimate used by the World Health Organization has been shown to be unable to provide useful information to monitor the time varying fatalities caused by the epidemic.

References

[1] P. K. Andersen, Borgan, R. D. Gill and N. Keiding, *Statistical Models Based on Counting Processes*, Springer-Verlag, New York; 1993.

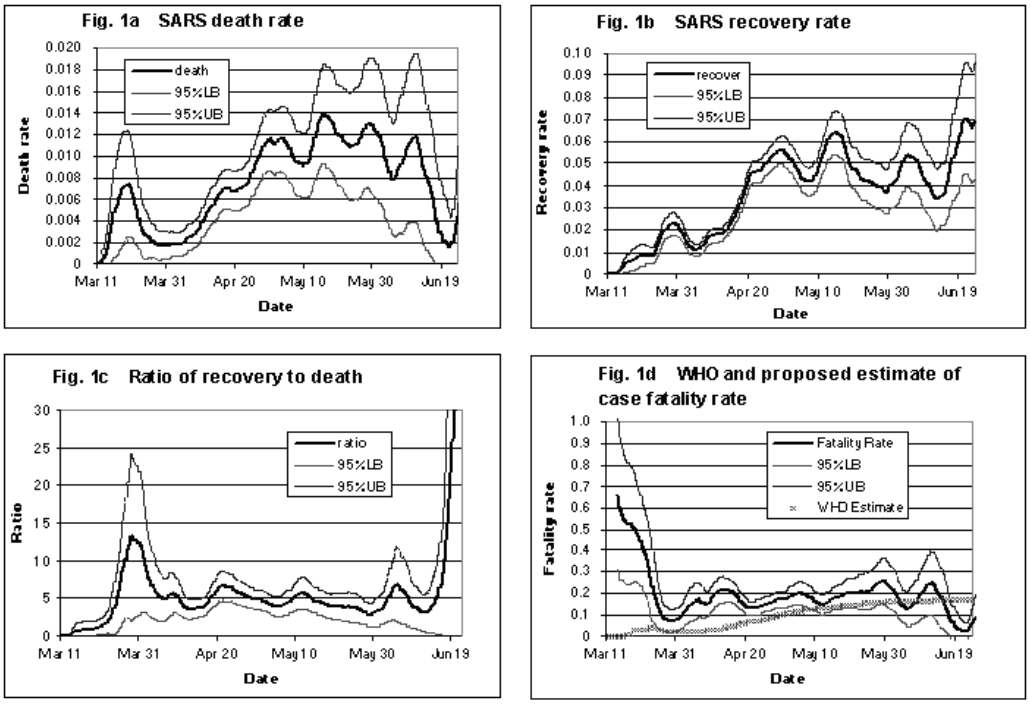

FIGURE 1. Kernel estimates of the SARS outbreak. (a) Instantaneous death rate $\gamma_1(t)$. (b) Instantaneous recovery rate $\gamma_2(t)$. (c) Ratio of instantaneous recovery/death rates $\theta(t)$. (d) Time-varying fatality rate $\pi(t)$ and WHO estimate.

[2] C. A. Donnelly, A C. Ghani, G. M. Leung, A. J. Hedley, C. Fraser, S. Riley, etal., *Epidemiological determinants of spread of causal agent of SARS in Hong Kong*, Lancet **361** (2003), 1761–1766.

[3] K. M. Tsang, P. L. Ho, G. C. Ooi, W. K. Yee, T. Wang, M. Chan-Yeung, et al., *A cluster of cases of severe acute respiratory syndrome in Hong Kong*, N Engl J Med **348** (2003), 1977–1985.

[4] P. S. F. Yip and K. F. Lam, *A multivariate non-parametric tests for the equality of failure rates in a competing risks model*, Communication in Statistics **11** (1993), 3199–3222.

[5] P. S. F. Yip and K. F. Lam, *A non-parametric test for the equality failure rates in competing risks model*, Communication in Statistics **21** (1992), 2541–2556.

[6] P. S. F. Yip, K. F. Lam, E. H. T. Lau, P. H. Chau, A. Chao and K. W. T. Tsang, *A comparison study of real time fatality rate: SARS in Hong Kong, Singapore, Taiwan, Toronto and Beijing, China*, to appear in Journal of Royal Statistical Society, Series A (2005).

Reporter: Hans Peter Duerr

Participants

Prof. Dr. Elja Arjas

Department of Mathematics
and Statistics
University of Helsinki
FIN-00014 University of Helsinki

Dr. Kari Auranen

Department of Mathematics
and Statistics
University of Helsinki
FIN-00014 University of Helsinki

Prof. Dr. Frank G. Ball

School of Mathematical Sciences
University of Nottingham
University Park
GB-Nottingham NG7 2RD

Prof. Dr. Heiko Becher

Hygiene-Institut
Universität Heidelberg
Im Neuenheimer Feld 324
69129 Heidelberg

Prof. Dr. Niels Becker

National Centre for Epidemiology
and Population Health
The Austr. National University
Canberra, ACT 0200
AUSTRALIA

Prof. Dr. Sheila M. Bird

Medical Research Council Centre
Biostatistics Unit
Robinson Way
GB-Cambridge CB2 2SR

Dr. Pierre-Yves Boelle

B3E Faculte de Medicine
Saint-Antoine
27, Rue Cahligny
F-75571 Paris Cedex 12

Dr. Martin Bootsma

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

Dr. Michiel van Boven

Animal Sciences Group
Wageningen University and Research
Center
P.O.Box 65
NL-8200 AB Lelystad

Dr. Tom Britton

Dept. of Mathematics
University of Stockholm
S-10691 Stockholm

Prof. Dr. I-Shou Chang

National Health Research Institutes
128 Yen-Chiu-Yuan Road, Sec. 2
Taipei 115
Taiwan

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

PD Dr. Martin Eichner

Institut für Medizinische Biometrie
Universität Tübingen
Westbahnhofstraße 55
72070 Tübingen

Prof. Dr. Vernon T. Farewell

MRC Biostatistics Unit
Institute of Public Health
Robinson Way
GB-Cambridge CB2 2SR

Dr. Christophe Fraser

Department of Epidemiology and
Public Health, Imperial College
School of Medicine, Norfolk Place
St Mary's Campus
GB-London W2 1PG

Dr. Kathryn Glass

National Centre for Epidemiology
and Population Health
The Austr. National University
Canberra, ACT 0200
AUSTRALIA

Dr. John W. Glasser

Centers for Disease Control
and Prevention
Mail Stop E-61
1600 Clifton Road NE
Atlanta GA 30333
USA

Prof. Dr. M. Elizabeth Halloran

Department of Biostatistics
Rollins School of Public Health
Emory University
1518 Clifton Road, N.E.
Atlanta GA 30322
USA

Prof. Dr. Niels Keiding

Department of Biostatistics
University of Copenhagen
Blegdamsvej 3
DK-2200 Kobenhavn N

Dr. Jim Koopman

Department of Epidemiology SPH-1
University of Michigan
611 Church Str.
Ann Arbor MI 48104
USA

Dr. Mirjam Kretzschmar

R. I. V. M.
Dept. of Infectious Dis. Epidem.
Antonie van Leeuwenhoeklaan 9
P.O.Box 1
NL-3720 BA Bilthoven

Dr. Andrew B. Lawson

Department of Epidemiology and
Biostatistics
Arnold School of Public Health
University of South Carolina
Columbia SC 29209
USA

Prof. Dr. Ira M. Longini

Department of Biostatistics
Rollins School of Public Health
Emory University
1518 Clifton Road, N.E.
Atlanta GA 30322
USA

Dr. Owen D. Lyne

Institute of Mathematics,
Statistics & Actuarial Science
University of Kent
GB-Canterbury Kent CT2 7NF

Prof. Dr. Eduardo Massad

School of Medicine
The University of Sao Paulo
Av. Dr. Arnaldo 455
Sao Paulo SP 01246-903
Brazil

Prof. Dr. Denis Mollison

Department of Actuarial Mathematics
and Statistics
Heriot-Watt University
Riccarton
GB-Edinburgh EH14 4AS

Prof. Dr. Johannes Müller

Zentrum Mathematik
TU München
Boltzmannstr. 3
85748 Garching bei München

Dr. Nico J.D. Nagelkerke

R. I. V. M.
Dept. of Infectious Dis. Epidem.
Antonie van Leeuwenhoeklaan 9
P.O.Box 1
NL-3720 BA Bilthoven

Dr. Philip D. O'Neill

School of Mathematical Sciences
University of Nottingham
University Park
GB-Nottingham NG7 2RD

Prof. Dr. Mick Roberts

Insitute of Information and
Mathematical Sciences
Massey University, Priv. Bag 102904
North Shore Mail Centre
Auckland
New Zealand

Muntaser Safan

Institut für Medizinische Biometrie
Universität Tübingen
Westbahnhofstraße 55
72070 Tübingen

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"
V.della Ricerca Scientifica, 1
I-00133 Roma

Prof. Dr. David Scott

Department of Statistics
Tamaki Campus
The University of Auckland
Private Bag 92019
Auckland
NEW ZEALAND

Prof. Dr. Tom Smith

Swiss Tropical Institute
Dept. of Public Health & Epidemiol.
Socinstrasse 57
CH-4002 Basel

Dr. Nikolaos Stilianakis

Joint Research Centre
European Commission
T.P. 441
Via E. Fermi 1
I-21020 Ispra (Varese)

Dr. Claudio J. Struchiner

Escola Nacional de Saude Publica
Fundacao Oswaldo Cruz
Rio de Janeiro RJ 21041
BRAZIL

Dr. Ake Svensson

Dept. of Mathematics
University of Stockholm
S-10691 Stockholm

Dr. Sake J. de Vlas

Department of Public Health
Erasmus MC
University Medical Centre Rotterdam
P.O.Box 1738
NL-3000 DR Rotterdam

PD Dr. Stefan Wagenpfeil

Institut für Medizinische Statistik
und Epidemiologie
Technische Universität München
Ismaninger Str. 22
81675 München

Dr. Jacco Wallinga

R. I. V. M.
Dept. of Infectious Dis. Epidem.
Antonie van Leeuwenhoeklaan 9
P.O.Box 1
NL-3720 BA Bilthoven

Prof. Dr. Paul Yip

Department of Statistics
The University of Hong Kong
Pokfulam Rd.
Hong Kong
HONG KONG

Yidi Zhang

Institut für Medizinische Biometrie
Universität Tübingen
Westbahnhofstraße 55
72070 Tübingen