EUROPEAN SCIENCE FOUNDATION
SETTING SCIENCE AGENDAS FOR EUROPE

Exploratory Workshop Scheme

Standing Committee for the European Medical Research Councils (EMRC)

ESF Exploratory Workshop on

Mathematical modelling to link contact network analysis and molecular typing of pathogens

Scientific Report

Utrecht (The Netherlands), 5-7 November 2008

Convened by:

Mirjam Kretzschmar ①, Jim van Steenbergen ②, Marc Bonten ① and Marion Koopmans ②

① University Medical Centre Utrecht, NL
② National Institute for Public health and the Environment (RIVM), Bilthoven, NL

ESF EMRC Exploratory Workshop:

Mathematical modelling to link contact network analysis and molecular typing of pathogens

Utrecht (The Netherlands), 5-7 November 2008

Executive summary

Molecular typing of pathogens provides information about the routes that an infectious disease takes through a network of contacts in the population. Therefore, analyzing the clustering of genotypes maps the contact network structure onto a structure that is measurable by modern laboratory methods. However, the contact network together with disease specific parameters such as the duration of the infectious period and the infectivity determine which individuals in a network acquire infection so that contact networks are only partially observed with a bias towards those individuals who have more contacts or who occupy specific network positions. At the same time, the evolution of the pathogen and the specific typing methods determine which groups of isolates are viewed as clusters. Finally, infected individuals are not always observed due to asymptomatic infection, underreporting or other mechanisms. Furthermore, contact networks can be viewed on different scales ranging from local contact patterns between individuals through contacts between risk groups up to contacts between distant geographical locations. Similarly, analysis of molecular sequences occurs on different levels of resolution depending on the typing methods used leading to information on different levels of aggregation. Therefore, the link between contact network structure and molecular clusters is complex and interpretation of clustering of genotypes in terms of contact patterns requires a detailed description of transmission dynamics and all processes linked with pathogen transmission as well as a thorough validation of the genotyping methods used for this purpose. Mathematical modelling provides methodological tools for analyzing how infectious diseases spread through networks. However, these tools have to be developed further to be able to deal with questions concerning the evolutionary dynamics of pathogens and the interpretation of sequence data. Also, assessing possible effects of interventions like contact tracing require the development of new mathematical tools and combining them with insights from molecular typing studies.

To discuss these questions and related research needs, a workshop was organized in November 2008, supported by the European Science Foundation (ESF). The workshop brought together a multidisciplinary group of scientists including virologists, epidemiologists, statisticians, mathematical modelers, and social scientists. The workshop was organized in three sessions. The first session was devoted to introduction and discussion of the research questions concerning contact networks and the spread of pathogens on these networks, the second session gave an overview over molecular typing studies and available data for various types of pathogens (HIV, nosocomial infections, influenza, norovirus, Chikungunya virus, hepatitis A, B, and C virus), the third session attempted to combine mathematical modelling and molecular epidemiology and presented first results in that direction. Finally, the workshop was concluded with a discussion session, in which future research directions and steps to follow up on them were summarized.

Contact networks and their impact on the transmission and spread of pathogens can be viewed on several levels of aggregation. On the micro level, contacts between individuals determine who infects whom, and the heterogeneity in behaviour between individuals therefore influences the reproductive success of specific pathogen strains and genotypes. Individuals with large numbers of contacts (superspreaders) will contribute disproportionally to the pool of genotypes present in a population. On the meso level, risk groups that can also be specified by epidemiological tools determine the genotype distribution. For example in the distribution of the hepatitis B virus, the genotype distribution clearly reflects transmission within well defined risk groups such as homosexual men or injecting drug users. Finally, on the macro scale, geographical distribution of risk populations and their long distance links determine the interrelation of local and global spread of pathogens. Phylogenetic data can help to disentangle these relationships and can in that way sharpen our picture of how pathogens spread among individuals, risk groups and populations.

To accomplish this, large data banks are being established that collect sequence data for many different pathogens. This required standardized laboratory methods for genotyping and knowledge on which regions of a genome should be typed to obtain the most informative phylogenetic tree. This area of research is moving fast and new typing methods are quickly evolving. The amount of available information stored in these data banks is increasing at an enourmous pace, while we are yet lacking the statistical and mathematical power to extract the information that is buried in these vast amounts of data. On the other hand, mathematical modelling has also developed fast within the last decade. Starting from mostly deterministic modelling with constant parameters, the increasing computing power has enabled modellers to develop individual based simulations that in principle are able to describe

infectious disease dynamics across all scales from micro to macro level. Also methodology for estimating parameters for these models has improved, in particular by implementation of Bayesian methods, such that modelling studies have by now become an integral part of infectious disease epidemiology and public health policy making.

While genotyping is performed for many different pathogens, the aims and questions turn out to differ vastly between them. In part the differences derive from the large differences in the clinical course of infection between pathogens. For infections like HIV and HCV with long chronic phases of infection and high variability in viral load and therefore infectivity during the entire infectious period, interpretation of genotype frequencies found in a population have to take different time frames into account than for infections like influenza or norovirus with short generation intervals. In addition, in HIV and HCV the virus evolves within single hosts and displays large within and inter host variability. Other clinical features of a pathogen that determine the observed distributions of genotypes are transmission routes, fraction of asymptomatic infections, and mutation rates of the pathogens.

In mathematical modelling various approaches have been developed for analyzing the influence of contact patterns on the spread of infections, for analyzing phylogenetic trees, and for assessing the impact of various types of intervention on incidence, prevalence and strain composition of pathogen populations. However, there is still a large gap between the level of complexity that these models can deal with and the vast amounts of phylogenetic data that are available. One of the discussion topics of the workshop was how to bridge that gap.

Scientific content of the event

Most of the workshop participants gave a presention of 20-25 minutes followed by a short discussion. All participants engaged actively in the discussions. The range of expertise represented in the workshop was broad and ranged from virologists specializing on laboratory techniques to mathematicians developing theory for describing transmission as stochastic processes on networks. The workshop was opened by a presentation of Prof. Dr. Martin Röllinghof, who introduced the aims and activities of the European Science Foundation. In the following we summarize the scientific presentations and discussions of the different sessions.

Session 1 : Contact networks and disease transmission

Introduction: contact networks and molecular epidemiology

Mirjam Kretzschmar (University Medical Centre, Utrecht, NL)

The session was opened by Mirjam Kretzschmar, who introduced the main question of the workshop, namely the relationship between contact networks and resulting genotype distributions of pathogens. Disease transmission through populations is channelled by underlying contact networks and the resulting genotype distributions found by surveillance or epidemiological studies reflect these contact patterns. However, factors like underdiagnosis and underreporting, uncertainty in genotyping methods, temporal changes in the contact network structure, and many more make a quantitative interpretation of that relationship difficult. In addition, characteristics of specific infections like variable infectivity and long chronic phases of infection lead to problems in interpreting sequence data in terms of pathogen transmission. Some studies have been undertaken to elucidate contact patterns leading to pathogen transmission for sexually transmitted infections as well as for airborne infections. Sampling methods have been developed to probe into contact networks of hidden populations, and these methods are closely related to interventions as contact tracing. On the other hand, contact tracing has been assessed using data from molecular typing studies leading to new insights about the effectiveness of contact tracing and possibilities of enhancing it by including other types of contacts into the intervention. Many questions remain as yet unanswered and a range of methodological questions have to be addressed in order to better understand the relationship between contact networks and molecular epidemiology.

Using environmental transmission theory and data to parameterize contact networks

Jim Koopman (University of Michigan, Ann Arbor, USA)

Most human to human transmitted organisms partially go through the environment during transmission. Current modelling practice is to abstract that away and define contacts and transmission probabilities per contact. When modelling transmission through public venues (and even in households), defining what constitutes a contact and estimating transmission probabilities per contact is difficult because it depends on the extent to which the pathogen travels by air, hand, or fomites. The Center for Advancing Microbial Risk Assessment in the USA is developing an alternative approach that involves modelling human movement, air transmission and fomite transmission. Preliminary analysis reveals that fomites which are touched sequentially by many people differ in their effects from the more common environmental fomite contamination. This approach allows the integration of a large body of data and theory in microbial risk assessment that is currently ignored by most transmission system modelling. It also allows the use of environmental agent identifications as a primary data source for model fitting. Such data is more feasible to collect and far cheaper than risk by venue exposure data. Identification of commensals in environmental samples can inform pathogen transmission model conformations. The potential for both commensal and environmental pathogen sequences from different venues to inform contact structure by venue is being investigated and looks promising.

The fourth dimension: combining epidemiologic and typing data for the transition from data to knowledge for infectious diseases

Dag Harmsen (Universitätsklinik Münster, DE)

Epidemiologic data (time, place, and person) in combination with informatics and molecular laboratory techniques (type, i.e. the fourth dimension) makes fast and accurate early warning outbreak detections for infectious diseases on different geographic levels possible. We have recently shown that rapid MRSA outbreak detection, based on epidemiological and spa typing data, is a suitable alternative of classical approaches (ICP) and can assist in the identification of potential sources of infection in a hospital (A. Mellmann et al. PLoS Med. 2006). We have also shown in an international multicenter study that high inter-laboratory reproducibility of DNA sequence-based typing of bacteria can be achieved due to the unambiguous nature of sequence data. By using dedicated client/server software, Ridom StaphType (Harmsen et al. J. Clin. Microbiol. 2003), a worldwide uniform terminology ("molecular Esperanto") can be ensured. Only high-quality sequence data are automatically accepted by the server (www.SpaServer.ridom.de) and, therefore, no curator is needed for administration of the database. Finally, we are currently busy in a concerted action of European laboratories to build capacities and to harmonize technology for sequence-based typing of microorganisms (www.SeqNet.org). The German National Reference Centre for Meningococci (NRZM, Würzburg) stores information on analyzed meningococcal samples in a central database. The recorded information includes high-resolution typing data, obtained by serogrouping and epitope sequence typing of porA and fetA. We have assembled a server that receives an anonymized subset of the NRZM data. A PostgreSQL database stores the epidemiological information as well as additional static data, such as geographical borders or population figures for counties and federal states. Our custom developed software combines and controls the database and additional open source software components (UNM MapServer and OpenLaszlo) to build an epidemiological geographical information system (GIS). The user accesses the automatically generated maps via the Internet, using a Flash-based application (www.EpiScanGis.org; M. Reinhardt et al. Int J Health Geogr. 2008). The server utilizes SaTScan (developed by M. Kulldorff et al.) to detect significant spatio-temporal clusters, taking the typing-, epidemiologic-data and population-at-risk into account. The SaTScan output is finally visualized within the GIS by depicting significant cluster of cases within the maps. Thus, the application of interactive, Internet-based tools can help achieving better quality control and faster cluster detection and allows for turning surveillance data into knowledge. Community building in an environment of mutual trust and sustainability of such services is crucial for long-term success.

Weighted social networks - modelling and disease control

Ken Eames (University of Cambridge, UK)

We naturally define a contact as any interaction that could facilitate the transmission of infection. However, this definition conceals a great deal of detail. This presentation discussed some of the complexities surrounding networks of contacts, particularly weighted networks and serial monogamy. In weighted networks, not all contacts are equal: some links have higher weight than others, and some social settings (particularly the home) are more likely to display high-weight links, with implications for targeting control strategies. Serial monogamy introduces a new timescale: the delay between partnerships; the evolutionary impact of this is to favour pathogen strains that have an infectious period long enough to allow persistence between partnerships. There is a great deal more to contact networks than meets the eye, and in many cases the question "What is a contact?" has not been adequately answered; combining contact tracing and molecular methods is one way to solve this problem.

Effectiveness of contact tracing in emerging infections

Johannes Müller (Technical University Munich, DE)

Contact tracing is a control method for infectious diseases that is believed to be quite effective. If an infected person is detected (the index case) one tries to find more infected persons via the contact history of the index case. As it is quite simple to keep track especially of the number of detected cases per index case, it is intriguing to ask about the information these data contain w.r.t. rates, reproduction number or contact structure. Up to now, one has basically three approaches: (1) a phenomenological approach that incorporates contact tracing as a linear or nonlinear term in a deterministic framework, where this term is not derived by a submodel on the micro-level. (2) By means of moment closure methods for individual based stochastic models. (3) The third approach formulates the infectious process with contact tracing as a nonlinear branching process. Methods have been developed to analyse this process at the onset of an epidemic. We take up the third approach and focus on the endemic state of an SIS model. As dependencies due to contact tracing as well as dependencies due to the high prevalence of infection are present (I-I contacts cannot be neglected in the endemic state), the analysis is not straight forward. We propose a preliminary method, partially based on heuristic arguments. As we assume a relatively simple model (SIS model), we focus on estimating rates (or better: combinations of rates) from the distribution of detected cases per index case. We furthermore discussed by means of heuristic arguments the effectiveness of contact tracing in comparison with clustering. Core groups and asymptomatic cases may be important aspects that need to be taken into account if one thinks about the effectiveness of contact tracing.

Session 2: Molecular epidemiogical studies of specific pathogens

Molecular typing of HIV, a tool to obtain insight in HIV transmission

Roel Coutinho (National Institute for Public Health and the Environment (RIVM), Bilthoven, NL)

To illustrate the power of molecular typing to obtain insight in the transmission routes of HIV, two topics were covered. The first one was how molecular typing of HIV can be used in forensic transmission investigations. Some years ago we were confronted with a case of deliberate HIV transmission by injection. Molecular typing showed that the viruses from the donor and the victim were very closely related to each other much more than with other HIV isolates from the same area (Amsterdam). Based on that phylogenetic analysis we could conclude that transmission of this virus was highly likely and supported other evidence to incriminate the offender. Also a recent much more complicated case of suspected deliberate transmission through blood or male to male sexual contact from three suspects to 12 victims in Groningen was discussed. Based on that case the caveats of phylogenetic analysis in forensics were discussed: phylogenetic analysis can only support otherwise obtained evidence but can not be used as ultimate proof for transmission itself. The second topic was the recent evidence about the role of primary HIV infection in HIV transmission. Mathematical modelling studies done in the past by several groups including ours indicated that primary HIV infection plays a modest role in HIV transmission. However, recently phylogenetic studies showed clustering of up to 50% of HIV isolates from recent infections indicating transmission between each other. These studies can be done because most new cases of HIV infection are now routinely phylogentically typed for resistance testing. The consequence for public health is that tracing of primary infections – and their contacts – could be an additional tool in controlling the spread of HIV especially among Men who have sex with men.

The use of molecular typing data in model-based analyses of hospital infections

Ben Cooper (Health Protection Agency, London, UK)

In modelling hospital infections and the spread of antibiotic resistant strains in hospitals, precise information is needed on timing of positive tests in patients, colonisation events, and on who might have transmitted to whom. This type of information is usually not complete. Molecular typing data can help to close these information gaps by giving some clues on who has infected whom and indirectly on the timing of transmission events. Use of this data will increase the ability of models to predict the effects of interventions in hospitals.

Phenotype and genotype evolution of influenza A H3N2 virus

Ron Fouchier (Erasmus Medical Center, Rotterdam, NL)

Results were presented from studies of genotype mapping of the influenza A virus. Using this mapping method, a two-dimensional representation can be obtained that describes the genetic drift of influenza A over a number of years. A striking result is that this drift is in essence a one-dimensional path through a plane, therefore enabling prediction of the direction of genetic drift ahead of time. This is important for the development of influenza vaccines and opens up new avenues for the development of pandemic influenza vaccines.

Molecular typing to study transmission of norovirus

Marion Koopmans (National Institute for Public Health and the Environment (RIVM), Bilthoven, NL)

In recent years, norovirus has been recognized as one of the main causative agents of gastro-intestinal illness. Molecular epidemiologic studies of norovirus found in many different outbreaks suggest that the epidemiology of norovirus has gone through a change with new and possibly more virulent strains emerging in recent years. Molecular typing studies of norovirus might help to understand the contribution of different transmission routes to norovirus transmission.

Emerging infections: the example of chikungunya

Giovanni Rezza (Instituto Superiore di Sanita, Roma, IT)

The recent outbreak of chikungunya in Italy was an example of an outbreak of an emerging pathogen in a previously unaffected population after import of one case from an endemic country. The vector had already been endemic in Italy. Molecular epidemiology contributed to elucidating the transmission routes of the outbreak and in corroborating that transmission was indeed ongoing within the country. The import case could be related to genotypes occurring in other regions of the world.

The epidemic history of hepatitis C among injecting drug users in Flanders, Belgium

Catharina Matheï (Catholic University Leuven, BE)

The purpose of this presentation was to demonstrate how the combined use of mathematical methods and molecular methods can be employed to infer the epidemic history of infectious diseases. Therefore a study was presented exploring the epidemic behaviour of hepatitis C subtype 1a and subtype 3a among injecting drug users in Flanders, Belgium, using new gene sequence data sampled among two geographic distinct populations of injecting drug users. In this study the coalescent method was used to estimate the history of changing viral population size using phylogenetic trees that are reconstructed from randomly sampled viral gene sequences. The change in the estimated number of HCV infections over time was then used to infer the growth rate and the basic reproductive number. Evidence was found for different dynamic forces driving both epidemics. Also, the results suggested that the hepatitis C subtype 3a epidemic has reached a steady state, while the hepatitis C 1a epidemic has not, which therefore might become the predominant subtype among injecting drug users. The resulting estimates of the basic reproductive number were rather low and seemed not in accordance with empirical observations. These rather low estimates of the basic reproductive number were shown not to be inconsistent with the epidemiological reality of hepatitis C among IDUs, if one takes into account that a substantial part of the HCV infections among IDUs do resolve spontaneously and that re-infections are possible. In the discussion after the presentation it was suggested that the low estimate of the basic reproductive number might also be the result of an implicit assumption about the distribution of the

generation time (T). In the presented study an exponential distribution was assumed leading to the linear relation R= 1+rT. Assuming a delta distribution of the generation, however, leads to the exponential relation R=exp(rT) and higher estimates of the basic reproductive number. Another topic of discussion concerned the limitations of coalescence theory.

Molecular epidemiology of hepatitis A and B

Jim van Steenbergen (National Institute for Public Health and the Environment (RIVM), Bilthoven, NL)

Hepatitis A occurs in well defined risk groups such as men who have sex with men (MSM), children of migrants with a travel history, persons with HAV cases in their close environment. With molecular techniques, the epidemiology of hepatitis A virus in the Netherlands is unraveled. There are two major separate transmission patterns:

1. Multiple import by people (especially children) returning from endemic countries. Thanks to control efforts (source and contact tracing, immunizing susceptible contacts) by local Public Health Departments, there is only limited spread in families and at school.
2. Ongoing transmission in the gay community after sporadic introduction. Transmission occurs in venues for anonymous sex contact and thus, control efforts fail, as sources and contacts can not be traced.

Molecular epidemiology showed that there is no exchange of viruses between the two groups. Despite detailed knowledge on epidemiology, control officers remain with a practical problem, where further mathematical analysis might assist: each isolated notified case might be the first sign of mounting cluster. However, one case does not justify large scale control efforts (e.g. vaccinating schools). Is it possible to develop an algorithm, incorporating relevant parameters, that assists the control officer in deciding for the best intervention strategy?

Session 3: Linking modeling and analysis of molecular typing data

Molecular epidemiology of HIV in Switzerland

Roger Kouyos (ETH Zürich, Switzerland)

We used a representative set of HIV-1 sequences from 5710 patients enrolled in the Swiss HIV Cohort Study to investigate the structure of the Swiss HIV-1 epidemic. These sequences were pooled with the same number of sequences from foreign epidemics, a phylogenetic tree was inferred, and Swiss transmission chains were identified as clusters that had a minimal size of 10 and contained at least 80% Swiss sequences. More than half (56%) of the Swiss sequences were included within 72 clusters. Most of these were significantly dominated by specific transmission routes, which were either men having sex with men (MSM) [37 clusters, average size 31 patients] or patients who have acquired HIV through heterosexual contacts (HET) or injecting drug use (IDU) [14 clusters, average size 130 patients]. Thus, IDU and HET segregated into fewer, but larger clusters compared with MSMs, suggesting more sub-epidemics and smaller transmission chains for MSM. Mixing between transmission groups was observed frequently for HET patients in MSM clusters (13%), but rarely for MSM patients in IDU-HET clusters (5%) and IDUs in MSM clusters (3%). While 51% of all HET patients infected between 1983-1986 were clustering with IDUs, this percentage was only 9% in later years (2003-2006, p-trend <0.001), suggesting a diminishing role of IDUs on the HET epidemic over time. This decline is also reflected in decreasing absolute numbers of newly infected HETs contained in IDU-HET clusters. Thus, we conclude that prevention measures targeted at IDUs can result in substantial alleviations for other transmission groups as well.

Parameterizing HIV transmission networks using genetic sequences from early infection

Jim Koopman (University of Michigan, Ann Arbor, USA)

HIV genetic sequence data has been mainly used to inform infection transmission system analyses by using coalescent theory approaches. Such approaches make gross simplifying assumptions about the transmission system and must explore tree space first through a phylogenetic analysis approach that is highly insensitive to fine structure in infection transmission system models – especially in the face of the high rates of HIV recombination and immune and therapeutic selective pressures. In a study in Montreal, Canada, we are developing an alternative approach. Montreal has an extremely high coverage of early infection sequences due to the extreme care in preserving privacy of infection status data, a well integrated gay community, and excellent provision of medical care and diagnostic services.

This has allowed the identification of numerous primary HIV infection (PHI) specific genetic clusters. Both the fraction of PHI transmissions and the pattern of PHI clusters have been shown to be highly sensitive to the duration of high risk behavior periods. These relationships are being exploited to estimate HIV transmission system parameters. A more thorough use of genetic information is also being developed that uses transmission system model generated transmission trees to explore genetic tree space. This approach should be less susceptible to degradation by recombination and selective pressures. It facilitates testing hypotheses about transmission system conformation.

Tentative measure of contact dynamics to be accounted in the understanding of outbreak success of AMR
Didier Guillemot (Institut Pasteur, Paris, France)

Plans were presented to conduct a contact study using electronic devices that are able to measure the number of contacts at a certain distance per individual. These data are going to be collected in hospitals to better quantify contact patterns that lead to the spread of antibiotic resistant pathogens.

Experiences of using genotyped MRSA data from an outbreak to detect likely paths of transmission in a regional health care system
Fredrik Liljeros (Stockholm University, SE)

Methicillin resistant *Staphylococcus aureus* (MRSA) constitute a severe threat to modern health care. The bacterium is very often spread between inpatients in different types of health care settings. The fact that the movements of in patients within and between hospitals often are stored in databases opens up new possibilities for contact tracing, especially if information about what type of strain of the bacterium the infected inpatients carries. The data are however usually stored in a way presuppose programming skills for analyzing the date. New more user friendly software is therefore urgently needed.

Pathogen diversity and disease epidemiology: models and realities
Gabriela Gomes (Instituto Gulbenkian de Ciência, Oeiras, PT)

Molecular epidemiology data constitutes an immense, and largely unexplored, resource for the parameterization of infectious disease transmission models. While the workshop focused mainly on the reconstruction of transmission chains using molecular typing of pathogens, in this presentation we discussed the reconstruction of individual histories of sequential infections and the use of this information to challenge competing hypotheses regarding modes of immune protection. Case studies on influenza and tuberculosis revealed the crucial importance of population heterogeneities in the assessment of infection and reinfection rates.

The skyline plot: can model averaging techniques be used in order to account for model uncertainty?
Ziv Shkedy (Hassel University, Diepenbeek, BE)

The estimation of the basic reproductive number from molecular data is based on a demographic model which fitted to the skyline plot. The skyline plot is considered to be a non-parametric estimate for the epidemic curve (or for the effective number of infected individuals) and as such it can be used in order to estimate an appropriate demographic model. Once a demographic model is selected, the basic reproductive number can be derived from the parameter estimate for the growth rate. This procedure introduces a problem of model selection and model uncertainty since the “true” demographic model is unknown and several demographic models can fit the skyline plot equally good. We proposed a method which does not require selecting a demographic model with the best goodness-of-fit but use information from all fitted model. The growth rate is estimated using model averaging techniques which take into account both goodness-of-fit and model uncertainty. In that way, our estimate for the uncertainty which is associated with the growth rate is more realistic. We illustrate the use of the method on simulated skyline plots and pointed out that our goal is to develop the method using real sequence data.

Assessment of the results, contribution to the future direction of the field

There was a general consensus in the group that combining molecular data with mathematical modelling offers huge opportunities but also challenges for future research. It was felt that there was a large gap in communication between different fields (laboratory based research, mathematical modeling, epidemiology) and that we have to work on developing a common language to bridge these gaps. The challenge for mathematical modeling and epidemiology is to generate insight from the enormous amounts of data that are being collected at present. One difficulty herein is that genetic typing methods are continuously in development and standardization between labs is difficult. This hampers the comparability of data from different sources. Furthermore, large data bases are needed to store all the sequence data and data can only be accessed via advanced bioinformatic techniques such as data mining. On the other hand, mathematical modeling has the tendency to build on abstractions of reality that allow for mathematical tractability and might therefore neglect some of the complexities that are inherent to sequence data. Also here the gap between conceptual simplicity and observed complexity has to be bridged in a manner that increases understanding of transmission dynamics and evolution of pathogens.

Concerning specific pathogens, a conclusion from the workshop was that a next step should be to focus on specific questions concerning some groups of pathogens. The transmission dynamics of infections like HIV and HCV with long chronic phases and rapid within host evolution of the pathogen leads to other questions than for pathogens like norovirus and influenza where the infectious period is in the order of days. Other differences are the setting where transmission mainly takes place, for example in small closed populations in the case of hospital infections, or in risk groups with specific behaviours like injecting drug users. It was agreed that meetings of smaller groups should be initiated who discuss and work on the molecular epidemiology and mathematical analysis of specific pathogens. Finally, it was discussed how molecular typing data can be used to understand better the role of re-infection and acquired immunity within individual hosts and their consequences for transmission dynamics on the population level.

If we succeed in converting the accumulated supply of data into an understanding of infection transmission, molecular sequence data can greatly enhance surveillance and control of infectious diseases. Understanding can be gained on several levels. On the micro level we want to understand the dynamics of outbreaks and who infected whom in transmission chains. On an intermediate level we would like to know how risk groups and populations are interconnected and how much transmission is going on between those groups. Finally, on a macro level molecular data can be used to study geographical distribution of infectious diseases and the spread of specific strains across country borders. For generating this type of understanding genotype sequencing has to be accompanied by collecting and analyzing epidemiological data and possibly data about contacts between individuals. Uncovering the routes along which transmission of pathogens takes place in and among populations will give us new options for designing interventions and a quantitative basis for public health decisions. This knowledge might even open up the possibility of anticipating long-term evolutionary effects of interventions and avoiding adverse consequences of pathogen evolution on human health such as deteriorating vaccine effectiveness or development of antimicrobial resistance. However, a lot of work has to be done to reach these ambitious goals. The workshop succeeded in laying the first foundations for future collaboration in that area. There was overall agreement that it is more than worthwhile to follow up on this first initiative with research proposals and efforts to start up collaborative projects. Some of these initiatives will be taken within the newly founded Utrecht Centre for Infection Dynamics (UCID).

ESF EMRC Exploratory Workshop:
Mathematical modelling to link contact network analysis and molecular typing of pathogens
Utrecht (The Netherlands), 5-7 November 2008

FINAL PROGRAMME

Wednesday 5 November 2008

Late afternoon	*Arrival*

Thursday 6 November 2008

08.30-09.00	**Welcome and registration**
09.00-09.15	**Opening: Aim of the workshop**
09.15-09.30	**Presentation of the European Science Foundation (ESF)** **Martin Röllinghoff** (ESF Standing Committee for the European Medical Research Councils)
Session 1:	**Contact networks and disease transmission**
09.30-10.00	**Introduction: contact networks and molecular epidemiology** **Mirjam Kretzschmar** (University Medical Centre, Utrecht, NL)
10.00-10.30	**Using environmental transmission theory and data to parameterize contact networks** **Jim Koopman** (University of Michigan, Ann Arbor, USA)
10.30-11.00	*Coffee break*
11.00-11.30	**The fourth dimension: combining epidemiologic and typing data for the transition from data to knowledge for infectious diseases** **Dag Harmsen** (Universitätsklinik Münster, DE)
11.30-12.00	**Weighted social networks - modelling and disease control** **Ken Eames** (University of Cambridge, UK)
12.00-12.30	**Effectiveness of contact tracing in emerging infections** **Johannes Müller** (Technical University Munich, DE)
12.30-14.00	*Lunch*
Session 2:	**Molecular epidemiogical studies of specific pathogens**
14.00-14.30	**Molecular typing of HIV, a tool to obtain insight in HIV transmission** **Roel Coutinho** (National Institute for Public Health and the Environment (RIVM), Bilthoven, NL)
14.30-15.00	**The use of molecular typing data in model-based analyses of hospital infections** **Ben Cooper** (Health Protection Agency, London, UK)
15.00-15.30	**Phenotype and genotype evolution of influenza A H3N2 virus** **Ron Fouchier** (Erasmus Medical Center, Rotterdam, NL)
15.30-16.00	*Coffee break*
16.00-16.30	**Molecular typing to study transmission of norovirus** **Marion Koopmans** (National Institute for Public Health and the Environment (RIVM), Bilthoven, NL)
16.30-17.00	**Emerging infections: the example of chikungunya** **Giovanni Rezza** (Instituto Superiore di Sanita, Roma, IT)

17.00-17.30 **The epidemic history of hepatitis C among injecting drug users in Flanders, Belgium**
Catharina Matheï (Catholic University Leuven, BE)

17.30-18.00 **Molecular epidemiology of hepatitis A and B**
Jim van Steenbergen (National Institute for Public Health and the Environment (RIVM), Bilthoven, NL)

20.00 *Dinner*

Friday 7 November 2008

Session 3: **Linking modeling and analysis of molecular typing data**

09.00-09.30 **Molecular epidemiology of HIV in Switzerland**
Roger Kouyos (ETH Zürich, Switzerland)

09.30-10.00 **Parameterizing HIV transmission networks using genetic sequences from early infection**
Jim Koopman (University of Michigan, Ann Arbor, USA)

10.00-10.30 **Tentative measure of contact dynamics to be accounted in the understanding of outbreak success of AMR**
Didier Guillemot (Institut Pasteur, Paris, France)

10.30-11.00 *Coffee break*

11.00-11.30 **Experiences of using genotyped MRSA data from an outbreak to detect likely paths of transmission in a regional health care system**
Fredrik Liljeros (Stockholm University, SE)

11.30-12.00 **Pathogen diversity and disease epidemiology: models and realities**
Gabriela Gomes (Instituto Gulbenkian de Ciência, Oeiras, PT)

12.00-12.30 **The skyline plot: can model averaging techniques be used in order to account for model uncertainty?**
Ziv Shkedy (Hassel University, Diepenbeek, BE)

12.30-14.00 *Lunch*

Session 4: **Discussion meeting: future research directions**
Facilitators: **Roel Coutinho and Jacco Wallinga** (National Institute for Public Health and the Environment (RIVM), Bilthoven, NL)

14.00-15.30 **Discussion**

1. to identify relevant research questions for specific pathogens and contact patterns.
2. to define interfaces between different disciplines (what are common interests, where does a common language/approach have to be developed)
3. to inventarize methodology that is available, maybe also from other fields of research
4. to map out which methodology that has to be developed

15.30-16.00 **Wrap up of the meeting**
(Formation of a research network and planning outline for research proposal)

16.00 *Closing of the meeting and departure*

Statistical information on participants (including convenors and rapporteur)

The 23 participants came from 11 different countries (Denmark, United Kingdom, Italy, Portugal, France, Germany, USA, Switzerland, Sweden, Belgium, Netherlands). They were 4 women and 19 men.

Final list of participants

Convenors

Mirjam Kretzschmar
Julius Centre for Health Sciences & Primary Care
University Medical Centre Utrecht
Heidelberglaan 100
3584 CX Utrecht
Phone: +31-(0)88-7568761
m.e.e.kretzschmar@umcutrecht.nl

and

Center for Infectious Disease Control
RIVM
PO Box 1
3720 BA Bilthoven
The Netherlands
Phone: +31-(0)30-2744021
Fax: +31-(0)30-2744409
mirjam.kretzschmar@rivm.nl

Jim van Steenbergen
LCI
Centre for Infectious Disease Control/RIVM
POBox 1 (pb 13)
NL-3720 BA Bilthoven
Phone +31-30-27 47 000
jim.van.steenbergen@rivm.nl

Marc Bonten
Acute Internal Medicine and Infectious Diseases
University Medical Center Utrecht
Heidelberglaan 100
3584 CX Utrecht
the Netherlands
Phone + 31-30-250 6228
Fax + 31-30-252 3741
mbonten@umcutrecht.nl

Marion Koopmans
Diagnostic Laboratory for Infectious Diseases and Perinatal Screening (LIS)
National Institute for Public Health and the Environment (RIVM)
Antoni van Leeuwenhoeklaan 9
3720BA Bilthoven, The Netherlands
Tel: (31) 30.2742889/3945; 31.6.52098601
Fax: (31) 30.2744418
Marion.koopmans@rivm.nl

Rapporteur

Martin Röllinghof
Universität Erlangen-Nürnberg
Institut für Klinische Mikrobiologie der Universität
Wasserturmstrasse 3
91054 Erlangen
Germany
roellinghoff@mikrobio.med.uni-erlangen.de

Participants

Viggo Andreasen
Department of Science
Build. 27.1
Roskilde University
DK-4000 Roskilde
Denmark
office (direct) +45 - 4674 2542
FAX +45 - 4674 3020
dept secretary +45 - 4674 2263
viggo@ruc.dk

Ben Cooper
Statistics, Modelling, and Bioinformatics Department,
Centre for Infections,
Health Protection Agency, London,
United Kingdom
Ben.cooper@hpa.org.uk

Massimo Ciccozzi
Epidemiology Unit
Department of Infectious, Parasite and Immune-mediated diseases
Istituto Superiore di Sanità
Viale Regina Elena, 299
Rome
Italy
Tel. (39) 0649903187
Fax. (39) 0649902755
ciccozzi@iss.it

Roel Coutinho
Center for Infectious Disease Control
RIVM
PO Box 1
NL- 3720 BA Bilthoven
T. +31 30 274 27 67
F. +31 30 274 44 86
Roel.Coutinho@rivm.nl

Ken Eames
London School of Hygiene and Tropical Medicine,
Keppel Street,
London
WC1E 7HT
UK
Ken.Eames@lshtm.ac.uk

Ron Fouchier
Dept. of Virology
Erasmus MC
P.O. Box 2040, 3000CA Rotterdam
Dr. Molewaterplein 50, 3015GE Rotterdam
Room Ee 17.73
The Netherlands
Phone +31 10 7044066
Fax +31 10 7044760
r.fouchier@erasmusmc.nl

Gabriela Gomes
Instituto Gulbenkian de Ciencia
Apartado 14, 2781-901 Oeiras
Portugal
T. +351-214-464-626
F. +351-214-407-970
ggomes@igc.gulbenkian.pt

Didier Guillemot
CeRBEP
Institut Pasteur
Bat Laveran - 2ème étage
25, Rue du Docteur Roux
75724 Paris Cedex 15
Tel: (+33) 1 45 68 82 99
guillemo@pasteur.fr

Dag Harmsen
Poliklinik für Parodontologie
Universitätsklinik Münster
Waldeyerstraße 30
48149 Münster, Germany
Phone: +49 (251) 83-47059
Fax: +49 (251) 83-47134
dharmsen@ridom.de

James Koopman
University of Michigan
Dept. of Epidemiology SPH-1
109 Observatory
Ann Arbor
MI 48109-2029
USA
Phone: +1-734 763-5629
Fax: +1-734 764-3192
jkoopman@umich.edu

Roger Kouyos
Institute for Integrative Biology
Universitaetsstr. 16
ETH Zentrum CHN H 75.1
CH-8092 Zurich
Switzerland
T. +41 44 632 22 17
roger.kouyos@env.ethz.ch

Fredrik Liljeros
Department of Sociology
Stockholm University
S-10691 Stockholm
Sweden
T. +46162000
F. +46 08 612 55 80
liljeros@sociology.su.se

Catharina Matheï
ACHG, KULeuven
Kapucijnenvoer 33, blok J
3000 Leuven
&
Free Clinic
Van Arteveldestraat 64
2060 Antwerpen
Fax: 00 32 (0) 3 216 03 27
Tel: 00 32 (0) 475 571675
catharina.mathei@telenet.be

Johannes Mueller
TU Munich, Tel. +89 / 289 18392
e-maiL. Johannes.mueller@mytum.de
Centre for Mathematical Sciences
Unit M12
Botzmannstr. 3
85747 Garching
&
GSF Munich, Tel. +89 / 3187 4502
Institut für Biomathematik und Biometrie
GSF- Forschungszentrum für Umwelt und Gesundheit, GmbH
Ingolstädter Landstrasse 1
85764 Neuherberg
johannes.mueller@helmholtz-muenchen.de

Giovanni Rezza
Epidemiology Unit
Department of Infectious Diseases
Istituto Superiore di Sanità,
Viale Regina Elena, 299
00161 Roma
Italy
giovanni.rezza@iss.it

Ziv Shkedy
Center for Statistics
Hasselt University
Agoralaan 1
BE - 3590 Diepenbeek
T. +32 11 26 82 62
F. +32 11 26 82 99
ziv.shkedy@uhasselt.be

Minus van Baalen
CNRS-UPMC-ENS-AgroParisTech UMR 7625
Ecologie & Evolution
Université Pierre et Marie Curie
Bat. A, 7eme Etage, CC 237
7 quai St.-Bernard
75252 Paris Cedex 05
France
T. +33 144272545
F. +33 1 44273516
minus.van.baalen@ens.fr

Jacco Wallinga
Center for Infectious Disease Control
RIVM
PO Box 1
NL- 3720 BA Bilthoven
T. +31 30 274
F. +31 30 2744409
Jacco.Wallinga@rivm.nl