DISS. ETH NO. 18971

Models of epidemics:
How contact characteristics shape the spread of infectious diseases

A dissertation submitted to the
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
TIMO SMIESZEK
Master of Sciences ETH
born January 31st, 1980

citizen of the Federal Republic of Germany

accepted on the recommendation of

Prof. Dr. Roland W. Scholz, examiner
Prof. Dr. Kay W. Axhausen, co-examiner
Prof. Dr. Fredrik Liljeros, co-examiner
PD Dr. Jakob Zinsstag-Klopfenstein, co-examiner

2010

Summary

This thesis examines *how the configuration and the quality of contacts between hosts shape the course of epidemics*. Being able to evaluate the relevance of such contact characteristics is highly relevant for constructing adequate mathematical or computer models of disease spread, which are nowadays one of the main techniques for assessing the effectiveness of interventions against epidemics of infectious diseases. Nonetheless, many modelers still apply the random mixing assumption which assumes contacts to be completely random and transient without scrutinizing the adequacy of this assumption for the respective infectious disease.

For treating the above mentioned research question, this thesis takes an interdisciplinary perspective that is rooted in system theoretic thinking: We understand and treat the spread of an infectious disease as the result of the dynamics of a coupled transmission system, comprising the host species, the pathogen(s) and environmental factors and their various interactions. This thinking in transmission systems forms the basis of all four contributions which build the main body of this thesis:

The first contribution gives answers to the question *when contact repetition and clustering should be included in models of epidemics*. This assessment is done by comparing the outcomes of models including repetitive and clustered contacts between hosts with models assuming randomly mixed and transient contacts. The differences between both model types are systematically tested for a multidimensional parameter space of social and biological influence factors. One of our findings is that the relevance of contact structure for the model outcome depends highly on the pathogen biology: Highly infectious and easily transmissible pathogens can be adequately modeled under the random mixing assumption; models of diseases that need close physical interaction for transmission should include more details about the actual contact structure.

The second contribution highlights the role of contact quality by investigating how *the duration and the intensity of contacts* affect the pathogen transmission probabilities. A model is presented that allows for calculating the probability that a specific infected host infects a specific susceptible host based on the duration and the intensity of their interaction. With the help of empirical contact data and this model, it is also shown that a large number of different contact partners is not sufficient for qualifying as a super-spreader inducing disproportionate amounts of secondary cases. Super-spreading events can only be explained when also the pathogen shedding rate of the infector is disproportionately high.

The third contribution is about *contacts between Swiss poultry farms*, which presumably play a role in avian influenza spread. One of the main findings of this empirical work is that non-

commercial farms are more important parts of the network of poultry farms than assumed by previous work which focused on contacts of commercial farms. For instance, non-commercial farms also have long-distance contacts that are crucial for disease control. Further, they are functionally connected with the commercial farms. As a result, non-commercial poultry farms must be included in models of avian influenza spread.

Finally, the fourth contribution presents a *reconstruction of the 2003/2004 H3N2 influenza epidemic in Switzerland* with an individual-based model. Successfully reconstructing past influenza outbreaks is presented as a strategy for validating epidemic models which shall be used for investigating hypothetical, future pandemics. We were able to reproduce spatial, temporal and age patterns of a past seasonal influenza epidemic with a detailed individual-based model integrating social and biological factors.

The four contributions give new insights of how contact characteristics shape the spread of infectious diseases. Moreover, this thesis shows that an interdisciplinary, systemic thinking in infection transmission systems is needed for understanding the spread of infectious diseases. We conclude that future research in this field should not only generate new knowledge about the relevance of the different factors that govern disease spread. Rather, knowledge integration and systematization is also needed for guiding field researchers when to investigate which specific elements of a certain transmission system.

Zusammenfassung

Diese Doktorarbeit untersucht, *wie die Anordnung und die Eigenschaften von Kontakten zwischen Wirtsorganismen den Verlauf von Epidemien prägen*. Die mathematische oder computergestützte Modellierung von Epidemien ist heutzutage die Standardmethode, um die Effektivität von Interventionsmassnahmen gegen Epidemien zu bewerten. Um hinreichend gute Krankheitsausbreitungsmodelle entwerfen zu können, muss man in der Lage sein die Relevanz der oben genannten Kontaktcharakteristiken (Anordnung und Eigenschaften) einschätzen zu können. Gleichwohl nehmen viele Modellierer an, dass sich eine Wirtspopulation homogen mischt und dass Kontakte stets kurzlebig sind – ohne zu testen, ob diese Annahme für die Modellierung der betreffenden Infektionskrankheit angemessen ist.

Für die Bearbeitung der genannten Forschungsfrage wählen wir eine interdisziplinäre Perspektive, welche auf systemtheoretischem Denken gründet: Wir verstehen und behandeln die Ausbreitung von Infektionskrankheiten als das Ergebnis eines Wechselspiels innerhalb eines Infektionsübertragungssystems. Ein solches System besteht dabei aus der oder den Wirtsspezies, dem Krankheitserreger, Umweltfaktoren sowie deren vielfältigen Wechselwirkungen. Dieses Denken in Infektionsübertragungssystemen ist die Basis aller vier Beiträge, die den Hauptteil dieser Doktorarbeit bilden.

Der erste Beitrag gibt Antworten auf die Frage, *wann das wiederholte Treffen bestimmter Kontaktpersonen sowie die Ausbildung von Gruppen („clustering") in Krankheitsausbreitungsmodellen berücksichtigt werden sollten*. Dazu vergleichen wir Simulationsergebnisse von Modellen, welche diese Kontaktcharakteristiken aufweisen, mit solchen, die homogen gemischte, kurzlebige Kontakte annehmen. Die Unterschiede beider Modelltypen werden systematisch für einen mehrdimensionalen Parameterraum bestehend aus sozialen und biologischen Einflussfaktoren getestet. Ein Ergebnis ist dabei, dass die Relevanz der Kontaktstruktur für die Simulationsergebnisse zu einem hohen Mass von der Erregerbiologie abhängt: Hochinfektiöse Erreger können gut mit der Annahme homogener Durchmischung modelliert werden; Ausbreitungsmodelle für Erreger, die durch engen, physischen Kontakt übertragen werden, sollten hingegen die tatsächliche Kontaktstruktur der Wirtspopulation berücksichtigen.

Der zweite Beitrag beleuchtet die *Rolle der Kontakteigenschaften*. Es wird darin untersucht, wie die Dauer und die Intensität eines Kontakts die individuelle Übertragungswahrscheinlichkeit beeinflusst. Es wird ein Modell eingeführt, mit dem sich die Übertragungswahrscheinlichkeit für ein bestimmtes Wirtspaar anhand der Dauer und der Intensität ihrer Interaktion berechnen lässt. Mit Hilfe empirischer Kontaktdaten und dem Modell können wir zeigen, dass eine grosse Zahl unterschiedlicher Kontaktpartner nicht ausreicht, um als sogenannter „Superspreader"

eine Vielzahl von Sekundärfällen zu initiieren. Vielmehr können solche „super-spreading" Ereignisse nur mit einer gleichzeitig überdurchschnittlich starken Ausscheidung von Erregern erklärt werden.

Der dritte Beitrag beleuchtet die *Kontakte zwischen Schweizerischen Geflügelhaltungen*: Von diesen wird angenommen, dass sie im Falle einer aviären Influenzaepidemie eine wichtige Rolle für deren Ausbreitung spielen. Eines der Hauptergebnisse dieser empirischen Arbeit ist, dass nicht-kommerzielle Haltungen ein wichtigerer Bestandteil des Netzwerks Schweizerischer Geflügelhaltungen ist, als in vorangegangenen Arbeiten angenommen wurde. So verfügen beispielsweise auch nicht-kommerzielle Haltungen über weit entfernte Kontaktpartner – ein Faktor, der für die Eindämmung einer Infektionskrankheit eine bedeutende Rolle spielt. Darüber hinaus sind kommerzielle und nicht-kommerzielle Haltungen über verschiedene Kontaktpfade miteinander verbunden. Demzufolge sollten auch Freizeithaltungen in Ausbreitungsmodellen für die aviäre Influenza berücksichtigt werden.

Der vierte Beitrag *rekonstruiert die H3N2 Influenzaepidemie der Saison 2003/2004 in der Schweiz mit Hilfe eines Individuen-basierten Modells*. Wir begreifen die erfolgreiche Rekonstruktion vergangener Influenzaepidemien als eine Strategie zur Validierung von Ausbreitungsmodellen, die auch für hypothetische Szenarien künftiger Influenzapandemien angewendet werden sollen. Wir konnten empirisch gemessene Muster bezüglich Raum, Zeit und Alter der Infizierten anhand eines detaillierten Individuen-basierten Modells reproduzieren. Dieses Modell integriert dabei sowohl soziale als auch biologische Einflussfaktoren.

Alle vier Beiträge bieten neue Erkenntnisse darüber, wie Kontaktcharakteristiken die Ausbreitung von Infektionskrankheiten beeinflussen. Darüber hinaus zeigt diese Doktorarbeit, dass eine interdisziplinäre, systemische Denkweise notwendig ist, um die Ausbreitung von Infektionskrankheiten verstehen zu können. Wir schlussfolgern, dass künftige Forschung nicht nur neues Wissen über die Relevanz verschiedener Einflussfaktoren auf den Verlauf von Epidemien schaffen sollte. Vielmehr sollte auch bestehendes Wissen geordnet und verflochten werden, um empirisch arbeitenden Forschen eine Orientierung bieten zu können, welche spezifischen Elemente eines bestimmten Infektionsübertragungssystems erhoben werden sollten.

Some remarks

This thesis is a cumulative thesis. Most of the papers that build the main body of thesis were written by several authors and use, thus, the first-person plural "we" as personal pronoun. For achieving a consistent style throughout the thesis, I will use consistently the pluralis auctoris also in those parts, which were authored solely by me.

For the same reason, we adapt slightly the style and some other minor details (e.g. the numbering of the figures, equations and tables) of the four papers. However, the content is identical with the published versions. A complete list of all references is given at the end of the thesis (Chapter 8), but not after each contribution individually. Accordingly, the numbering of the references is not identical with the numbering in the published articles.

In Chapter 1, we use several figures, which are taken from publications of other authors or from a publication from the author of this thesis that is not part of this thesis. The copyright status of these Figures is described subsequently:

Figure 1.2 has been taken from Calvin Schwabe's book "Veterinary medicine and human health". This book was originally published by Kluwer, but Kluwer sold the copyright to Mosby, which is now part of Elsevier Health. Meanwhile, the title has gone out-of-print and Elsevier has reverted the rights to the author. Calvin Schwabe died in 2006 and we do not know who the heirs are. Consequently, the copyright status of Figure 1.2 remains unclear.

Figure 1.4 has been taken from the paper "Appropriate models for the management of infectious diseases" by Wearing, Rohani and Keeling. Permission to do so is granted by the CCAL license, which explicitly allows to copy, to modify and to distribute papers and parts of papers under this license.

Figure 1.5 has been taken from the conference paper "Social patterns and communicable diseases: What epidemiology (still) can learn from SNA" by Fiebig, Smieszek and Zinsstag. The figure has been created by the author of this thesis and the authors of this conference paper are the copyright holders.

Table of contents

1 Introduction

Infectious diseases were a major scourge of mankind from time immemorial [154] and continue to be a main reason for death, malaise and economic losses [173, 225, 289, 366]. Probably more than ten million deaths per year can be directly attributed to infections with pathogens and four out of the ten most important causes of death are infectious diseases (Table 1.1). Thus, understanding the spread of infectious diseases (and how it can be influenced) is not only a fascinating scientific exercise, but also a challenge of enormous practical relevance.

In light of that, our knowledge about many infectious diseases is rather fragmented: We have extensive and detailed knowledge of many aspects of diseases, but only unreliable or even missing information on many other, crucial aspects. For example, thousands of human influenza genomes with millions of nucleotides have been completely sequenced allowing insights into the evolutionary dynamics of influenza [130, 273], but the relative importance of different transmission pathways – crucial knowledge for effective public health measures – is still highly controversial for influenza [15, 47, 200, 330].

Important factors that shape the spread of infectious diseases are the configuration and the quality of potentially contagious contacts between potential hosts. For a long time, the role of contact structure was almost ignored by epidemiological research and by modelers in particular. Instead, the attention was mainly on the biological factors of disease spread [231]. This situation has changed vastly during the last thirty years and a growing body of literature has emerged, which deals with the impact of contact structure on the patterns of disease spread [232].

Nevertheless, often models of epidemics are published that do not adequately reflect the appropriateness of their underlying model assumptions with respect to contact structure [122, 248, 277, 308]. And still, host contact structure continues to be inconsiderately approximated with the random mixing assumption, which deems contact between hosts to be completely random and transient. Random mixing models are well-proven and provided valuable insights into disease dynamics [22], but there are also many cases where there are clear indicators that random mixing models might lead to vastly biased model outcomes (cf., e.g., the review of Morris [231]).

This thesis investigates and discusses how several contact characteristics can be included in computer models of epidemics, in which way this affects the model outcomes, and what this means for health policy. In particular, we provide answers to the two following *general* questions:

1) Under which conditions is the inclusion of the concrete configuration of contacts (i.e., how contacts are arranged between hosts) into models of epidemics necessary and when might the simplifying random mixing assumption be adequate?

2) What role does the quality of contacts (i.e. how long, intense, and frequent contact events are between hosts) play for model outcomes?

Furthermore, we will examine two *concrete* cases: avian influenza preparedness for Swiss poultry farms and human seasonal influenza spread in Switzerland. Thereby, we will give answers to the following questions:

3) How might the concrete configuration of contacts between farms affect the risk of avian influenza transmission on the farm and national levels?

4) Is it possible to reconstruct measured characteristics of seasonal influenza spread with a detailed computer model incorporating details on human travel, contact characteristics, and the biology of hosts and pathogens?

Cause of death	Deaths (10^6)	1/100 of deaths	Cause of death	Deaths (10^6)	1/100 of deaths
World			***Low-income countries***		
Ischaemic heart disease	7.2	12.2	Lower respiratory infections	2.9	11.2
Cerebrovascular disease	5.7	9.7	Ischaemic heart disease	2.5	9.4
Lower respiratory infections	4.2	7.1	Diarrhoeal diseases	1.8	6.9
COPD	3.0	5.1	HIV/AIDS	1.5	5.7
Diarrhoeal diseases	2.2	3.7	Cerebrovascular disease	1.5	5.6
HIV/AIDS	2.0	3.5	COPD	0.9	3.6
Tuberculosis	1.5	2.5	Tuberculosis	0.9	3.5
Trachea, bronchus, lung cancers	1.3	2.3	Neonatal infections	0.9	3.4
Road traffic accidents	1.3	2.2	Malaria	0.9	3.3
Prematurity / low birth weight	1.2	2.0	Prematurity / low birth weight	0.8	3.2
Middle-income countries			***High-income countries***		
Cerebrovascular disease	3.5	14.2	Ischaemic heart disease	1.3	16.3
Ischaemic heart disease	3.4	13.9	Cerebrovascular disease	0.8	9.3
COPD	1.8	7.4	Trachea, bronchus, lung cancers	0.5	5.9
Lower respiratory infections	0.9	3.8	Lower respiratory infections	0.3	3.8
Trachea, bronchus, lung cancers	0.7	2.8	COPD	0.3	3.5
Road traffic accidents	0.7	2.8	Alzheimer and other dementias	0.3	3.4
Hypertensive heart disease	0.6	2.5	Colon and rectum cancers	0.3	3.3
Stomach cancer	0.5	2.2	Diabetes mellitus	0.2	2.8
Tuberculosis	0.5	2.2	Breast cancer	0.2	2.0
Diabetes mellitus	0.5	2.1	Stomach cancer	0.1	1.8

Table 1.1: Leading causes of death by income group, 2004. Adapted from [366], p. 12. Low income was defined as $825 or less; high income as $10066 or more.

We believe that neither the concentration on the biological factors of infection risk nor the one-sided focus on the configuration of contacts between hosts can lead to an adequate understanding of the spread of infectious diseases. Even though this thesis highlights the role of contact characteristics, we understand that this should not and cannot be done without considering the context composed of the biology of both host and pathogen as well as of parameters of the physical environment. An example supporting our believe is that one has to know the possible and the likely pathways of transmission to know what kind of contacts are relevant and have to be surveyed in epidemiological contact studies.

The indicated need for integrating knowledge belonging to different classical disciplines like biology, mathematics or sociology demonstrates that interdisciplinary work is necessary to tackle problems in the field of disease spread. In fact, we believe that the borders of classical disciplines like medicine, biology, or sociology are often too dysfunctional to adequately answer scientific questions on infectious disease spread. Instead, we suggest conceptualizing the spread of infectious diseases in an interdisciplinary, system theoretical framework, which integrates relevant mechanisms and elements belonging to different disciplinary realms.

To prepare the paradigmatic ground for the research work, we first review in Section 1.1 different concepts of scientific inquiry and describe schools of thought that all go beyond classical disciplines and that all incorporate system theoretical considerations. We then synthesize this review work and define the position underlying this thesis.

The integration of the various factors governing disease spread and the investigation of the dynamics of transmission systems are usually done with the help of mathematical or computer models. There are different views in philosophy of sciences about (i) what models are; (ii) what we can learn from models; and (iii) what makes a model a good model. To embed our own modeling work, these three questions will be discussed briefly and the viewpoint taken in this thesis will be clarified in Section 1.2.

Sections 1.3 and 1.4 contain brief reviews about transmission models on the level of individuals and models of epidemic spread on the level of populations. As we will see, the history of mixing concepts in epidemiology is tightly bound to the development of certain modeling approaches and their specific limitations. For assessing the role of contact characteristics in disease spread, it is helpful to understand where often-used concepts and metaphors of mixing come from. Hence, Section 1.4 provides some background knowledge on the co-evolution of mixing concepts and modeling approaches and helps to put the guiding questions of this thesis in context. Section 1.3 lays the broader foundation for the integration of biological processes and host behavior needed to answer the second question outlined above.

Then, the final section of this chapter, Section 1.5, introduces the four contributions that build the research core of this thesis. These contributions correspond (in order) with the four guiding questions and are titled:

1) Models of epidemics: When contact repetition and clustering should be included. (Chapter 2)

2) A mechanistic model of infection: Why duration and intensity of contacts should be included in models of disease spread. (Chapter 3)

3) Contacts between poultry farms, their spatial dimension and their relevance for avian influenza preparedness. (Chapter 4)

4) Reconstructing the 2003/2004 H3N2 influenza epidemic in Switzerland with a spatially explicit, individual-based model. (Chapter 5)

Finally, Chapter 6 will provide an overall conclusion about the findings elaborated in the preceding chapters of this thesis. Furthermore, further research needs which are tied in with out work and findings will be presented.

1.1 Interdisciplinary and system theoretical thinking in science and medicine

We already stated that there is a need for systemic approaches going beyond the borders of classical disciplines for understanding the spread of infectious diseases. This section reviews interdisciplinary thinking in general and with respect to disease spread. In the first subsection, we will provide a short view on the interdisciplinarity debate in the sciences in general. The second subsection reviews four related concepts and schools of thought that contribute to a better understanding of disease spread. Namely, we will discuss the one-medicine concept, conservation medicine, the human-environment systems framework, and transmission systems analysis. In the final section, we will describe the specific framework underlying the research work of this thesis.

1.1.1 Going beyond classical disciplines in science

Higher education in the ancient and medieval world was founded on the study of the *artes liberales*[1] [186] in order to grant a comprehensive, encyclopedic knowledge to the "freeman". During the Late Middle Ages the term discipline appeared in the university system to distinguish three branches of (professional) training: theology and the arts in Paris, law in Bologna, and

[1] Grammar, rhetoric, logic, arithmetic, geometry, music, astronomy

medicine in Salerno [276]. These early professional specializations mirrored a societal need, i.e., were external to educational institutions, whereas later subdivisions were often stimulated by intra-scientific processes [324]. By the 19th century, research and education were entirely on a reductionist track: Reality became decomposed into more and more granular parts.

Reductionism is a mode of scientific inquiry that aims at understanding "the whole of something by examining its parts" (p. 379) [172]. Reductionism in general has been overly successful and has led to the wealth of knowledge and multitude of comfortable technologies modern societies enjoy today. However, reductionism also has faced several inherent limits, which led to claims for a re-unification of science or – at least – a synergistic co-operation between the sciences to overcome those limitations. The origins of the concept of interdisciplinarity are highly disputed [186], but we think that one can distinguish two different observations that have led to calls for research beyond the classical disciplines:

1) The observation that there are striking *analogies* in the concepts, mechanisms, and structures postulated in different disciplines.

2) The observation that disciplinary boundaries are often *dysfunctional* for understanding and solving real-world problems.

Observation 1; analogy of concepts, mechanisms and structures in various disciplines: We can observe that there are disciplines and professions that seem to be "useful" for problem solving regardless of which disciplinary realm the problem belongs to. This holds particularly true for applied mathematics and statistics. Leslie Kish, in those days the president of the American Statistical Association, emphasized in his 1978 presidential address:

> "Statistics is a peculiar kind of enterprise of contradictory character because it is at the same time so special and so general. Statistics exists only at the interfaces of chance and empirical data. But it exists at every such interface, which I propose to be both necessary and sufficient for an activity to be properly called statistics." (p. 1) [184]

While other disciplinary researchers focus necessarily on a more or less well-defined part of the world, statisticians apply their knowledge to all kinds of problems and data. The same is true for applied mathematics: According to Hersh, *universality*, i.e. that the "mathematic we know is the only mathematics there can be" (p. 37) [159], is one of the generally accepted myths of mathematics. At latest with Newton's *Philosophiae Naturalis Principia Mathematica* the mathematical approach to science was well-established and until today it proved to be a powerful way of

deriving insightful scientific conclusions without direct reference to the material reality [360] – be it in physics, biology, chemistry, or the social sciences.

Since the 1950s several concepts, schools, sciences, and paradigms emerged, which aim at identifying general principles overarching the established scientific disciplines and which can be subsumed under the label *theories of systems*. One of the most prominent is von Bertalanffy's *General System Theory*, which he defines as a "new discipline" with "the formulation and derivation of those principles which are valid for 'systems' [cf. Figure 1.1, TS] in general." (p. 76) [340] as subject matter. He starts with the observation that some "similar general viewpoints and conceptions" (p. 75) like organization, wholeness, and dynamic interaction have emerged in otherwise independent disciplines [340]. As a core of General System Theory von Bertalanffy postulates (i) that there *are* structural similarities or even isomorphies (cf. Subsection 1.2.1) between different realms of the world; (ii) that (self-)organization, (self-)regulation, and directiveness are characteristics of systems of interest; and (iii) that those structural similarities go beyond the character of weak or narrative analogies.

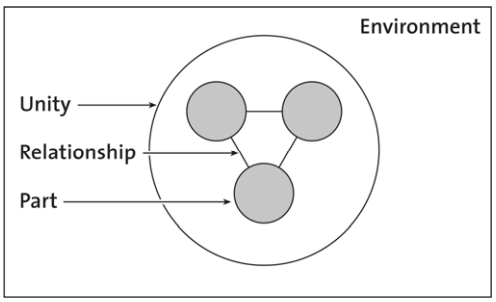


Figure 1.1: General scheme of a system: A system is composed of its parts or elements and the relationships between these parts. A system is an indivisible unit (any change would lead to a different functioning). Any part, which does not belong to it, is part of its environment.

Further approaches, to which General System Theory is closely knit, are: (i) *cybernetics* [14, 357], i.e., the study of the structure of regulatory systems. Cybernetics deals with circular relationships (feedback loops), which allow systems to regulate their internal dynamics and to adapt to external disturbances and forces; (ii) the *theory of information* [309], which relates informational content to the notion of entropy and has proven useful for understanding diverse communication processes ranging from human conversation to the transmission of heritable information via DNA or RNA [186]; (iii) *structuralism*, which perceives reality as a complex system of interrelated parts and seeks "the underlying formal structures, the deep structures" [222]; (iv) recent concepts including *complex adaptive systems*, *complex systems science*, *and network theory* [23, 99, 292, 360]. They form a broad field of diverse but similar concepts that try to explain complex macro-behaviors by micro-interaction independent of the disciplinary realm the case example might belong to.

Observation 2; dysfunctionality of disciplines for solving real-world problems: The second observation is that many problems humankind faces today cannot sufficiently be tackled by the classical disciplines through which research is predominantly organized [48, 172, 186, 207]. The ongoing fragmentation of knowledge leads to a situation where no single "expert" can tackle realistic problems adequately [48] and where an integral assessment can only be done with great difficulty as the "parts can no longer be put together easily" (p. 379) [172]. Integrating the different parts fails often (i) because there is no interface between the different realms of knowledge; (ii) because there are different cultures and frames of reference between different disciplines; (iii) because they use different languages; and (iv) because various disciplines utilize different methods of scientific inquiry [48].

Environmental problems are classical examples of problems that cannot be tackled with purely disciplinary inquiry [186, 324]. There is a vast body of examples where disciplinary knowledge offered "solutions" for problems that – after an initial improvement – engendered secondary problems often more severe than the original ones. To mention three of them: The pesticide dichlorodiphenyltrichloroethane (DDT) was widely used as a means against the vectors of Typhus [356] and Malaria [310]. However, it not only lost effectiveness in vector control, but was also found to be harmful to man and wildlife [310]. Other pesticides, if applied on acid soils, are proven to contribute to carcinogenic nitrosamines, which, when ingested via comestibles, were found to cause harm to man [211]. As a final example, chlorofluorocarbons (CFCs) were investigated and promoted in the late 1920s as a refrigerant to replace the common but toxic ammonia, chloromethane, and sulfur dioxide [61]. From an engineering point of view, CFCs were thought to be safe and environmental-friendly substances as they are generally non-reactive (consequently also non-flammable) and nontoxic [105]. It took until the 1970s to finally recognize that CFCs are inert enough to reach the stratosphere, but that an ozone-depleting chain reaction would be initiated by photolysis of the CFCs [224] resulting in the holes in the ozone layer we experience today

How observations 1 and 2 are interlinked: Both observations disclose a need for scientific inquiry beyond the classical disciplines, but seen from a different angle: The first observation highlights communalities of the sciences and demonstrates that an idea like the unity of science is still a partly realistic utopia. The second observation makes it clear that there are complex real-world problems that cannot be tackled successfully by scientific disciplines that were defined following other logics than those of the problems to be solved. Although both observations look at the disciplinarity-interdisciplinarity question form the opposite angle, they are complementary when the shortcomings described should be overcome: Karlqvist states that one "seek[s] a kind of metaknowledge" (p. 379) [172], when science wants to contribute to the solution of real-

world problems with scientific approaches. According to Karlqvist, credible interdisciplinary research requires an adequate understanding of the disciplines, but one also has to know how to connect this knowledge. To achieve knowledge integration, the communalities of the different realms of life and the gap-bridging tools of mathematics, statistics, and computer-based modeling can and must be employed.

1.1.2 Four problem-oriented concepts for understanding disease spread

Subsequently, four selected concepts, all of which can be applied for conceptualizing disease spread, will be introduced. All of them are rooted in the observations 1 and 2 described in the previous Subsection 1.1.1. This subsection starts with the one medicine concept, which originates from the observation that the split between human and veterinary medicine is dysfunctional. Then, we lead over to the more general concept of conservation medicine. Finally, two concepts – the human-environment systems framework and transmission system analysis – will be presented which both come from a system theoretic background.

One medicine concept

The *one medicine concept* is based on the belief that the split of medicine into human and animal subdisciplines is unnatural, arbitrary, and not useful. Integrative thinking in health professions is quite old as the origins of healing in all parts of the world made no systematic differentiation between humans and animals [301]. Today, traditional, integrated healing is still practiced, e.g., by African pastoralists [328]. Part of the explanation for the divide between human and veterinary medicine goes back to a religious motivation as many Christian schools of thought believe(d) in the uniqueness of man, separating him from fauna [301]. But also earlier sources distinguish human and veterinary medicine. Although science and education were and are dominated by distinct human and animal health disciplines, integrative thinking was constantly present in the discourse: it was a component of the foundation of universities in Europe [288] and medical pioneers like Rudolf Virchow held that "between animal and human medicine there is no dividing line—nor should there be" (p. 2) [301]. Finally, it was Calvin Schwabe, who coined the term *one medicine* denoting a concept of integrated human and animal health research [301].

The links between human and veterinary medicine founding the one medicine concept are manifold and depicted in Figure 1.2. Amongst one medicine's various components, comparative medicine, i.e., the study of human disease by comparison with corresponding diseases of animals, is the most obvious. For instance, ethical concerns and regulations often do not allow studying the course of human disease in an unaltered way, as patients have to be treated if a treatment is available for a disease. Comparative medicine offers the opportunity to study

diseases under natural conditions and to thereby gain insights that a researcher of ethically bound human medicine could never offer [301].

Schwabe further emphasizes that veterinary medicine has a focus on population or herd medicine [7, 301]. In contrast to that, he sees human medicine often focused on the "reductionistic methods" (p. 9) [301] of diagnosing and treating individuals isolated from their context. In his opinion, veterinary practice could contribute to human public health because veterinary medicine was more familiar with epidemiological thinking than human medicine [301].

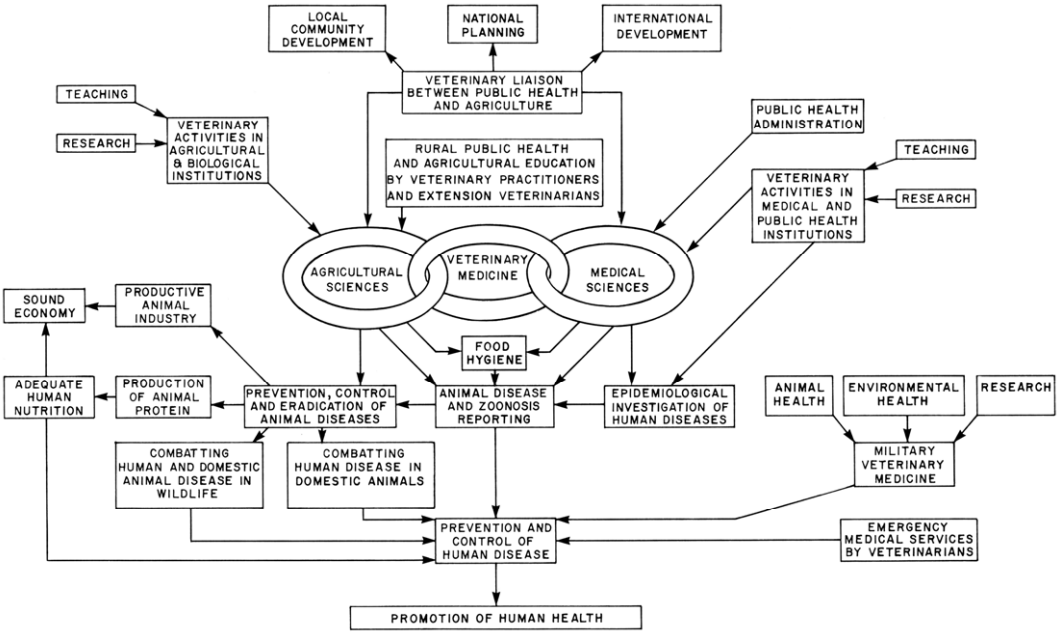


Figure 1.2: Links between veterinary medicine, (human) medical sciences, and agricultural sciences. Figure has been taken from [301]. Copyright status is discussed in the remarks.

Animals can be used as sentinels for natural or xenobiotic toxic substances [57], as well as for infectious agents [270]. Advantages of using animals as sentinels for infectious agents can be: (i) certain animals could be more sensitive than humans to a particular infectious agent; (ii) animals could have shorter incubation periods; or (iii) animals could be more exposed [271]. A conjoint, systemic view on human and animal cases can help to successfully investigate and mitigate outbreaks. For example, an Anthrax outbreak in the USSR could only be attributed to a the release of spores from a military facility because not only human cases, but also patterns in sheep were investigated. Furthermore, the West Nile Virus in the New York City area could have been detected earlier if medical doctors and veterinarians would have worked together more closely [271].

The importance of a more systemic view of infectious disease transmission becomes evident when looking at the quantitative and qualitative importance of zoonotic diseases: Schwabe estimates that four-fifths of all infectious diseases occurring in humans also occur in other vertebrates [301]. According to Molyneux et al., approximately three-fourths (132 of 175) of all emerging infectious diseases in humans have first been observed in other organisms [226]. Both understanding and fighting old and emerging human infectious diseases with animal hosts thus require a systemic view that maps the intra-human or intra-animal flows and patterns of disease spread, as well as the cross-species transmission.

Conservation medicine

As with one medicine, there are various similar concepts that are used almost synonymous for conservation medicine (e.g. ecosystem health [275]). The boundaries between these similar concepts are fuzzy – even Schwabe's [301] work on one medicine already includes some ideas belonging to conservation medicine. In the following, we concentrate on conservation medicine, which we understand as a health discipline including the relationship between human health, environmental health, and environmental conditions [352], while being focused on the well-being of humans or animals.

One part of this field is to identify and utilize analogies regarding phenomena and methods in medicine and ecosystem research [275] (cf. observation 1, Subsection 1.1.1). Rapport found, for instance, that both (i) try to measure the vital signs of their respective system; (ii) try to understand their systems capacity to deal with stress; and (iii) try to identify factors of disease and malfunction [274]. However, for understanding the spread of infectious diseases, we are mainly interested in those parts of conservation medicine that deal with host-parasite ecology and, thus, link both fields[2].

Following Daszak et al. [78, cf. also 300] the key factor for the formation of most emerging infectious diseases (EID) is a change in host-parasite ecology. Their (not exhaustive) listing of causes for EID encompasses, amongst other things, changes (i) in the international travel and commerce; (ii) in population demographics and host behavior [84, 124]; (iii) in agriculture and food processing; (iv) in the global climate system; (v) in technology; or (vi) in the interaction patterns with wildlife [78]. Accordingly, there are environmental processes and elements beyond the individual host, the host population and the pathogen, which influence the spread of infectious diseases.

[2] As described in the paragraph "How observation 1 and 2 are interlinked", Subsection 1.1.1

One environmental factor that may lead to new infectious diseases in humans or change the patterns of the spread of existing diseases might be the occurrence of toxic substances in the environment. We know that there are xenobiotic substances like some pesticides (e.g., γ-hexachlorocyclohexane) or flame retardants (e.g., hexabromobenzene or decabromodiphenyl ether) that bio-accumulate along the food chain [239]. As a consequence, predators are most prone to becoming poisoned by toxic anthropogenic substances. The selective poisoning of predators may lead to an increase in parasite abundance as, for instance, small invertebrates feed on trematode cercariae (a larval form of these parasitic flatworms) [299] and other predators eat snails that are an intermediate host of trematodes [28]. Removal of predators may, thus, increase the risk of infections because an important regulatory feedback loop has been disturbed.

Another impact factor possibly altering the risk of disease transmission is changes in the global nutrient cycles. In their review on this issue McKenzie and Townsend [218] describe two mechanisms by which nutrient pollution can increase the risk of the transmission of infectious diseases: (i) there can be an increase in the resources for intermediate hosts or vectors (e.g., in case of malaria or dengue fever) or (ii) the additional nutrients are a directly exploitable resource for pathogens (e.g., cholera). McKenzie and Townsend reviewed studies on the relationship between parasites or pathogens and nutrients and found for 55 case examples only four cases in which the burden of disease was likely to decrease as a response to additional nutrient supply [218].

Changes in the built environment and land use – like new roads or changes in watercourses – can also alter infection transmission systems (cf. transmission system analysis, pp. 14). Changes in the water system, such as new irrigation schemes or dams, have been a source for increased infection risk worldwide. A positive association between the malarial incidence and the irrigated area was found for the two Indian states of Nagaland and Punjab [179]. The construction of a dam at Diama – 40 km upriver from the estuary of the Senegal River – to prevent the intrusion of salty water into the river lead to the appearance of intestinal schistosomiasis: The altered ecology ameliorated the conditions for *Biomphalaria pfeifferi*, which is an intermediate host for the parasite *Schistosoma mansoni*. Before the dam was built, the predominant snail species was *Bulinus globosus*, a snail which copes with brackish water during the dry season and is not a host for *Schistosoma mansoni* [318].

A final example of the benefits of conservation medicine approaches to disease spread is the impact of global climate change on disease transmission. Several infectious agents and hosts are known to be sensitive to climatic conditions [262, 280]. Seasonal influenza, for instance, appears only in epidemic extent during the cold season [85, 374] (cf. Chapter 5). Salmonella and cholera bacteria replicate more rapidly at higher temperatures [147]. Several recent studies

showed that there is a relationship between short-term changes in the mesoclimate and the occurrence of infectious diseases [129]: In Asia and South America a correlation between the prevalence of malaria prevalence and the El Niño-Southern Oscillation cycles could be observed [44-46]. El Niño / La Niña events further seem to have triggered dengue fever outbreaks in Asia [148, 149, 167]. Finally, variations in the climatic conditions appear to have affected Ross River virus outbreaks in Australia [333, 361].

Human-environment systems framework

As previously became apparent, the spread of infectious diseases through (human) populations can be conceptualized as a coupled human-environment system. Scholz [297] introduced seven postulates within his human-environment systems framework to "conceptualize the structure and dynamics of human-environment interactions" for "coping with the complexity of most relationships within and between human and environmental systems" (p. 549) (cf. Table 1.2). Like the one medicine concept, conservation medicine or ecosystem health, Scholz' human-environment systems framework is a paradigmatic[3] approach as all of them deal to a varying extent with the following questions: (i) What is to be observed and which questions have to be asked? (ii) How should research be structured to answer these questions? (iii) How should the results of such scientific enquiry be interpreted?

In the following, we briefly discuss the postulates applied to infectious diseases.

Label	Contents
Complementarity	Human and environmental systems are complementary.
Hierarchy	Human systems can be investigated according to hierarchies within and among them.
Interference	Different levels of human system, from the micro to the macro level, interact.
Feedback	There are different types of feedback loops within and between human and environment systems.
Decision	Human systems can be conceived of as decision makers that have goals and strategies to strive for.
Awareness	Human systems have different types of environmental awareness
Environment first	After problem definition, the analysis of a human-environment system should be based on an adequate analysis of the environment.

Table 1.2: Postulates of the human-environment system framework. Table adapted from Scholz [297].

Complementarity, the first postulate, distinguishes two kinds of complementary subsystems: a human and an environmental one, both of which have different rationales, but both of which are coupled. Investigating them in isolation is often dysfunctional for understanding human-environmental problems. This distinction is purely definitional: One could consider both parts as

[3] in the etymological meaning of paradigm as pattern, example, model

unity [297], but splitting the universe in two coupled subsystems is a pragmatic decision that makes conceptualizing human-environment problems easier.

The definition of what is part of the human sub-system and what is part of the complementary environment depends on the hierarchical level (cf. *hierarchy postulate*) from which the entire system is looked at: If one takes the angle of an individual human, the relevant material environment consists of other individuals that can transmit pathogens either directly (by interaction between two individuals) or indirectly (e.g., by contaminating the inanimate environment with causative agents). The social environment of an individual includes, inter alia, the societal rules defined by the institutions of the respective society (including curfews, school closures, face mask obligation, sexual norms).

Relevant hierarchical levels of the human sub-system with respect to infectious diseases are, for example: (i) the *individuals' tissues and cells*, which are targets for pathogens and parasites as sites of and instruments for replication; (ii) the *individuals* themselves as potential hosts of a pathogen or parasite; (iii) *organizations*, e.g., pharmaceutical companies that develop and provide vaccines, antibiotic or antiviral drugs, or non-governmental organizations (NGOs) such as patient movements; (iv) *national or sub-national institutions*, which are commissioned to mitigate epidemics and pandemics, such as the health care systems, universities, and research institutions, or the national centers for disease control; (v) *international entities* – the World Health Organization (WHO) being one among others – that organize coordinated action and decision-making beyond the nation states.

All of these different hierarchical levels have their own rationales, which might be in accordance with each other or in opposition. That means the various hierarchical levels can interfere with one another (*interference postulate*). One example of such interference is that national health care systems often prioritize certain groups for vaccination and treatment such that the highest possible total number of lives saved or DALYs[4] averted can be achieved [265]. In contrast to that, many concerned individuals want to receive the best possible treatment regardless of what the "optimal" solution for the entire population might be.

An important and integral part of understanding complex human-environment systems and, thus, of infectious disease spread as a particular case example of a human-environment system is the idea of feedback loops (*feedback postulate*). The human-environmental systems framework [297] distinguishes between primary and secondary feedback loops: Primary feedback in a

[4] Disability-adjusted life years

human-environment system can be understood as simple and instantaneous cause-effect relationships observed in a system. One example of primary feedback in infectious disease spread is the immediate reaction of the system to a change in behavior – for instance, that influenza infection rates decrease during school vacation [65]. Secondary feedback refers to changes in the system's composition or dynamics itself, which can mean changes in the environmental dynamics or in the human system by reflected learning processes [297]. One example of such a secondary feedback loop in infectious disease spread is that changing mobility patterns with increasing frequency of remote journeys [295] increase the likelihood that highly virulent infectious agents evolve. It has been shown theoretically [41] and empirically [40] that a loss of locality in contact patterns leads to the evolution of more virulent strains of infectious agents.

Finally, the human-environment systems framework is rooted in a decision theoretic tradition besides its recourse to system theory: The *decision* and the *awareness postulate* (i) state that human systems are intentional systems whose rationales have to be understood for an accurate understanding of the functioning of the system and (ii) emphasize the role of human perception and cognition (which build up *environmental awareness*) for human behavior. The importance of knowing the problem awareness and goals of key actors for anticipating the behavior of a transmission system is apparent for the evaluation of intervention measures. When models of epidemics are used to test the effect of social distancing measures like school closures, this is usually done under the *ceteris paribus* assumption, but we do not know whether children really confine themselves at home when schools are closed. Furthermore, we have learned from various studies that even health care workers can have alarmingly little knowledge about influenza transmission and countermeasures [233, 306, 322], what might lead to counterproductive behavior in case of a pandemic. There are various historical and literary records of how people adapting their behavior when faced with a perceived risk of infection [35, 58], but an adequate understanding of human behavior in the face of potentially lethal diseases is still lacking [109].

Transmission system analysis

The term "transmission system analysis" was coined by Koopman and co-authors in several publications [189-191, 201]. Koopman defines infection transmission systems as systems that "circulate infection through complex contact patterns related to both contact patterns and patterns of factors that affect the risk of transmission given contact" (p. S3; all subsequent quotes in this paragraph refer also to p. S3) [191]. The transmission systems thinking as presented by Koopman distinguishes between "transmission system components" and "transmission system processes". As system components he mentions the "infectious agents, hosts, the [physical, TS] space in which hosts move, and environments affecting how infectious agents

reproduce or survive outside of hosts". As system processes he mentions "those that generate the natural history of infection and contagiousness, mechanisms by which agents leave hosts, all the modes of transmission that enable agents to survive as they transmit form one host to another, the processes that lead to contact between hosts, and the evolutionary mechanisms of the agent and hosts".

As with all the other approaches presented before, transmission system analysis is a problem-oriented concept. The aim of research efforts in this field is to ameliorate the living conditions of humankind through better (and better informed) infectious disease control [191]. In that sense, transmission system analysis aims to answer questions like the following:

> "Should control be sought with interventions directed to the entire population or will contact tracing and quarantine be more productive? What symptoms and circumstances should lead to isolation of sick individuals? Which types of contact with which kinds of individuals should lead to quarantine of healthy individuals? Which populations or places deserve concentrated intensive surveillance or control efforts like quarantine, chemoprophylaxis, symptomatic treatment, vaccination, or decontamination? When should whole groups of people like those living in an apartment complex be quarantined? When should we close contact venues like schools, or cancel sporting events and concerts? Which kind of hygiene or barrier precautions should be instituted? Which studies should we undertake to support those entrusted to make difficult decisions, such as restricting the freedom of individuals?" (pp. 303) [190]

The school of thought represented by Koopman is rooted in a clear *system theoretic* tradition, as he clearly defines what is part of the system and what lies beyond the system's boundaries and as he distinguishes between system elements or components and system processes (i.e., relations between the components; cf. Figure 1.1). As in the human-environment system framework presented before, clearly the *human parts* (No. 1 and 4 in Figure 1.3) and *environmental parts* (No. 2 and 3 in Figure 1.3) of the entire system are distinguished. The host-internal dynamics (No. 5) are partly driven by the human component and partly by the infectious agent.

Further, the idea of *hierarchy* plays a role in Koopman's concept of transmission system analysis (e.g. reflected in the links to multilevel analysis [83, 120]); seen from a granular, sub-individual perspective, transmission systems are driven by the interaction between intra-host replication of pathogens and the counteracting immune forces of the host (No. 5 in Figure 1.3). If an individual is immunosuppressed (be it due to HIV [194], immunosuppressive drugs [115], or simply stress [269]), this individual might become ill due to an otherwise harmless micro-

organism that would be eliminated effectively by the immune system under normal conditions. On the other side, people acquire or "learn" specific immunity whenever they are challenged with a specific infectious agent and are usually capable of suppressing or – at least – allaying secondary infections afterwards [116, 293]. On an individual level, infection is triggered by the actual condition of and interaction(s) with the social (No. 4 in Figure 1.3) and inanimate environments (No. 3 of Figure 1.3) of an individual. If no pathogen-shedding individual is around or if the physical environment is not contaminated with pathogens, there is no risk of infection. Finally, on a supra-individual level, the population-wide arrangements of contacts as well as the biological, chemical and physical properties of different venues determine the actual patterns and dynamics of disease spread in a system.

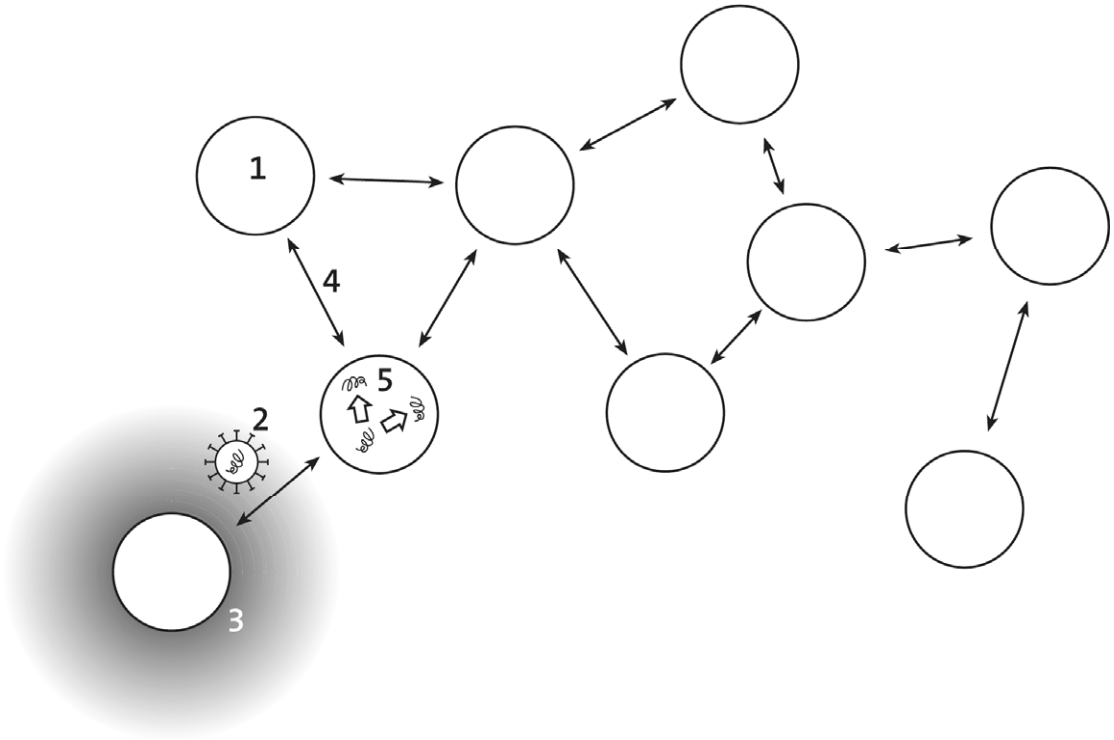


Figure 1.3: Scheme of an infection transmission system. The system includes individual hosts (1), specific pathogens (2), and the inanimate environment of hosts and pathogens (3). There can be epidemiologically relevant interaction (contact) between two hosts (4). Another relevant component of infection transmission systems is the internal dynamics in an infected host (5).

From a disciplinary perspective, Koopman sees transmission system analysis at the brink of becoming a new "science", i.e., a new discipline within the existing canon of scientific disciplines. For forming this new "transmission science" he calls for inter- and transdisciplinary knowledge integration (e.g., on p. S7 of [191]). We see three clusters of classical disciplines that the transmission system analysis concept makes use of: (1) natural science disciplines like medicine, biology, physics, and (partly) chemistry provide the foundation for understanding

intra-host dynamics, interference between different strains [189] and the environmental conditions that modulate the risk of transmission given exposure [201]; (2) social science disciplines provide insights into the contact patterns of humans and the determinants of risky behavior [189-191]; and (3) overarching disciplines like mathematics, statistics, or computer science allow the integration of processes and mechanisms uncovered by the previous two clusters [189-191] (cf. Subsection 1.1.1).

It is explicitly not the aim of transmission system analysis to include every detail of a specific transmission process in a model. Also this school of thought emphasizes the necessity and advantages of simplification. However, Koopman claims that transmission system approaches allow better identification which details have to be included and which details are omissible than classical disciplinary approaches [191]. They allow determination of which role different properties of a transmission system (e.g., heterogeneity in contact structure or different modes of transmission) play for the spread of infectious disease and under which conditions they have to be included in an appropriate analysis to understand transmission [190]. One way of doing so is robustness analysis, i.e., testing whether (and if so, under which conditions) the conclusions of a transmission system analysis changes if certain details are taken out of the analysis [189].

1.1.3 Synthesis and perspective taken in this thesis

All concepts and schools of thought presented in the previous subsection have in common that they are problem-oriented instead of discipline-oriented. They all recognize that understanding infectious disease in humans or animals can neither be understood by concentrating on the single individual, nor by focusing on the pathogen, nor by solely observing populations. Instead, all concepts take to varying extent a system theoretical stance. For a brief synoptic synthesis of the four approaches, see Table 1.3.

Here, we also take a problem-oriented perspective. The aim of the research presented is to contribute to the understanding of infectious disease spread and of the various consequences of different human disease management options. Therefore, we utilize theories, methods and concepts coming from various disciplines. Further, this thesis also stands in a system theoretical tradition. Instead of defining the boundaries of the analyses along disciplines, we try to identify the relevant elements and processes that drive the dynamics of the phenomena we are interested in.

Concept	Focus
One medicine	The rationale behind the one medicine concept is to show the analogies of human and veterinary medicine. Hence, one major focus is on the dysfunctionality of their specific disciplinary boundaries. According to representatives of this concept, both veterinary and human medicine could learn from each other as both have peculiar strengths. Systemic thinking particularly occurs with respect to common human and animal diseases: Both understanding zoonoses und using animals as sentinels for human outbreaks are conceptualized as coupled human-animal systems.
Conservation medicine	Conservation medicine can be understood as an enhancement of the one medicine concept. While one medicine emphasizes the benefits from a merger of human and veterinary medicine, conservation medicine wants to integrate the medicines and ecosystem research. Conservation medicine acknowledges that the coupling between human hosts and animal hosts (as already mentioned in the one medicine concept) is not static. Instead, one has to understand changes in an ecosystem's composition and dynamics as well as human-ecosystem interactions to explain, e.g., emerging infectious diseases.
Human-environment systems	The human-environment systems framework is a general framework for organizing research at the interface of human and environment systems. It is not bound to a specific domain like infectious disease spread, but can be applied to it. While the one medicine and the conservation medicine concept rather elaborate on the benefit of specific interdisciplinary cooperation (but stay within the idea of disciplines), the human-environment systems framework takes a rigid system theoretical perspective beyond the disciplinary view. It offers seven postulates to structure research on human-environment systems for a better understanding of their dynamics. Amongst other things, it proposes structuring such systems along the human system / environment system divide and along hierarchical levels. Feedback loops play a central role, be it short-term reactions of the system or long-term changes in the entire system's composition. The system theoretic view is complemented by decision theoretical parts, which reflect the role of human decision making for the dynamics of human-environment systems.
Transmission system analysis	The concepts of transmission system analysis and human-environment systems have strong similarities. Transmission system analysis is also grounded in a strong system theoretical tradition and it also has been emancipated from the tight boundaries of the classical disciplines. However, while "human-environment systems" is a broad framework applicable to all such systems, transmission system analysis focuses purely on infectious diseases. Another difference between the two concepts is that transmission system analysis does not explicitly demand research on human decision making: Human behavior explicitly plays a role (e.g. in form of the configuration of contacts), but questions like compliance with intervention measures and what triggers certain behavior are not explicitly addressed.

Table 1.3: Synoptic comparison of the foci of the four concepts presented in Subsection 1.1.2.

We want to understand how contact characteristics shape the spread of infectious disease. As we concentrate on single epidemics, the research presented within this thesis has a short-term perspective. Consequently, we do not explicitly model the co-evolution of pathogens and hosts (however, this plays a role as a constraint in Chapter 5). We further analyze both human and animal systems (Chapters 5 and 4). However, for the spatial setting (Switzerland) and time scale (one influenza season) of our analyses, the disease transmission between the human and the

animal systems appears to be rather irrelevant. Therefore, we concentrate on one host species per analysis. Existing and known links between the human subsystem and other host species are not included in the various models presented in this thesis. Finally, also the concrete decision making processes of human systems (cf. the decision and awareness postulates in Table 1.2) are ignored in this thesis, although we acknowledge their relevance during actual epidemic outbreaks of diseases.

We explicitly investigate processes on various hierarchical levels and discuss how these levels are linked: Chapters 2, 4, and 5 focus mainly on population level dynamics driven by the behavior of individuals (cf. Section 1.4); Chapter 3 presents an individual level model of disease transmission (cf. Section 1.3), which reflects the processes that lead to transmission. This is done with a higher resolution than is usually applied in population level models. To do so, we also utilize knowledge coming from various disciplinary contexts: Measures of contact patterns are closest to the realm of sociology (used in Chapters 2, 3, 4, and 5); models of individual infection risks originate from microbiology (used in Chapter 3); population-wide patterns of immunity have been researched in epidemiology (used in Chapter 5). Knowledge integration is done to a large extent with modeling (cf. Subsection 1.1.1; used in Chapters 3 and 5).

1.2 The role of models in science

The terms *model* and *modeling* play a central role in all parts of this thesis. Therefore, it is utile for the understanding of this thesis to clarify our stance towards three fundamental questions in the philosophy of science regarding models:

1) What are models?
2) What can we learn from models?
3) What are good models?

The following subsections shall elucidate these questions in a pragmatic way. We do not claim completeness regarding the portrayal of the past and ongoing discourses in philosophy. Rather, the subsequent answers shall guide us in setting the presented research in context.

1.2.1 What are models?

Thinking in models is as old as humankind [241]. Defining what the term *model* means belongs to the competence of semantics, i.e., the study of meaning; to the competence of ontology, i.e., the study of the nature of being; and to epistemology, i.e., the study of knowledge. A very broad definition of the term *model* is given by Haag [144], who defines it as a

> "material or ideal (re-) production of an object by means of *analogies* realized by a cognitive subject" (p. 4).

This definition includes *material* representations like miniatures of buildings and constructions in architecture [257] or hydrodynamically similar models in engineering [281], model organisms in biological and medical research [114], or the calotte models of molecules in chemistry [73], as well as *abstract* representations like the Lotka-Volterra model of a predator-prey system [209, 339], threshold models of riots [138], and Ising-like models of behavioral pattern formation [303] in social sciences or the model of black-body radiation in physics [266]. When we use the term model in this thesis, we always refer to abstract, not to material models.

Models can further be *static* or *dynamic*. Architectural models are typical examples of static models, whereas abstract, mathematical models applied in biology, epidemiology, physics, or social sciences are typically dynamical models. Some authors distinguish between the terms *model* and *simulation* according to the dynamic aspect of the respective representation of reality: they name a *model simulation*, if and only if the model is a dynamic model with a specific and concrete purpose for application [128, 235]. Our understanding of a model of epidemics is always a dynamic one. The dynamics of disease transmission systems are modeled (represented) by equations or rules. We use the term *simulation* as concrete realizations of an abstract model.

Haag's definition of *model* requires a model to be an *analogous* reproduction of an object. More or less strict requirements of similarity are the common ground of most model definitions [131, 329]. There is the restrictive claim that reality and model must be isomorphic [335] or – at least – partly isomorphic [76] to each other. Ashby defines *isomorphism* as two systems[5] (here model and 'reality') differing just in the type of their elements, but not in their number or relations [14]. In its stricter sense, the concept of isomorphism would allow no simplification or idealization in models: Neither Aristotelian (i.e., leaving out "irrelevant" properties of the system [119]) nor Galilean (deliberate distortions that make a system more tractable [119]) idealizations would be accepted as models of reality. In its less strict sense, Aristotelian idealizations would be

[5] Here, a *system* shall be defined by its elements (inclusive definition from a set theoretic definition), the relations between these elements and its boundaries (exclusive definition; the complement in set theory). See also Figure 1.1.

accepted, but not Galilean[6]. Thus, many statistical and phenomenological models would not be accepted as models under this concept.

Constructivism and (neo-)pragmatism stand in opposition to this restrictive claim [241, 319]. Already physicist Ernst Mach reasoned that models were not images of reality but constructs. He called theories and models 'auxiliary notions' ("Hülfsvorstellungen") for the representation of matters of fact [241]. Constructing a model requires making decisions about how to represent reality, what to include in the model, and how to parameterize it [351, 355]. It further implies that models come into existence by the synthesis of formerly unconnected, observed elements and that their relation to reality emerges by using them [235, 241].

With respect to the requirement of similarity or analogy, constructivism also implies that it must be specified which aspects of model and 'reality' must be similar and to what extent [131]. A radical constructivist would argue that trying to compare 'reality' with models, which are products of human thought, is absurd, as the idea of reality itself is a human construct. A moderate constructivist would accept such an attempt, as human senses at least perceive signals of the external world. From a moderate constructivist stance, such a decision has no underlying objective truth and cannot be made purely on the basis of philosophy of science [329] – it rather "depends on the problem at hand and the larger scientific context", as Frigg and Hartmann [119] point out.

Pragmatic model definitions further take a functional perspective: Relevance and interest are major drivers of model construction and, at the same time, are an explanation for the fact that the same subject matter can be represented in diverse ways [119]. Taking the neo-pragmatic stance, models are constructed by and for specific "cognitive subjects" in order to meet the specific modeling goals of these subjects [319]. This focus on function also means – in contrast to the claim for isomorphism – that both *a priori* models (i.e., mechanistic models aiming for a reproduction of the 'real' processes behind a phenomenon) and *a posteriori* models (i.e., data-driven, statistical models) are acceptable models as long as they fulfill the tasks for which they were defined (cf. Subsection 1.2.2).

In this thesis, we take a moderate constructivist and pragmatic stance: We believe that reality exists independently from the observer and that we can approach reality through scientific inquiry. One way of integrating the various signals we receive from the world to achieve a

[6] This less strict definition of *isomorphism* (allowing for Aristotelian, but not Galilean idealizations) equals Ashby's definition of *homomorphism* [14].

representation of reality is modeling (cf. the discussion on functional validity in Subsection 1.2.3). Furthermore, we see modeling as a purposeful operation that aims to explain or predict future states of specific phenomena of interest. We see modeling as a partly objective and partly artistic task with a certain inherent arbitrariness: partly objective, because certain quality criteria can be defined that measure how well a model fulfills its function – partly artistic (cf. [230]) as both the exact definition of the model function as well as certain structural decisions are acts of volition, which are not tractable in terms of right or wrong.

1.2.2 What can we learn from models?

The question of what we can learn from models belongs to the realm of epistemology, i.e., the study of "the nature, sources and limits of knowledge" [213]. In the previous subsection we introduced the distinction between *a priori* and *a posteriori* models. Here we describe how they can be distinguished from an epistemological perspective. In philosophy, these two terms are usually employed to disclose the origin of propositions. Baehr [17] distinguishes them as follows:

> "A given proposition is knowable a priori if it can be known independent of any experience other than the experience of learning the language in which the proposition is expressed, whereas a proposition that is knowable a posteriori is known on the basis of experience."

The division between *a priori* and *a posteriori* knowledge goes back to the controversy between rationalism (→ Descartes, Spinoza, Leibniz, Wolff) and empiricism (→ Bacon, Hume, Locke) in epistemology. A radical rationalism postulates concepts and ideas to be eternal and existing since the dawn of time (and new ideas only to come into existence by combining previous ideas or deducing them from previous ideas, respectively [213]), while a strict empiricism solely accepts experience as a source of insight and as the criterion for the verification or falsification of ideas [156].

Immanuel Kant "reinvented" epistemology and, by cutting his own path, reconciled both branches, the rationalism and the empiricism: Kant "advocated a 'transcendental' form of justification" [17], which believes in rational insight, but accepts that every concept is directly or indirectly connected to empirical experience [17, 156, 171].

Without going deeply into the details of the philosophical discourse on the origins of knowledge, we take a position in this thesis that stands in a Kantian tradition. We accept that (i) concepts and ideas are entities of their own and that (ii) new concepts can be derived from existing concepts (for instance, but not solely following the system of rules of mathematics). However, we also take up the positions that (i) all knowledge is related directly or indirectly to

experience and that (ii) concepts that do not correlate with observations may exist, but are dysfunctional.

In this thesis we use the term *a priori model* as a synonym for *mechanistic model*. An a priori model is a model that starts from theory. A priori models use an (pre-existing) understanding of the mechanisms driving the phenomena we want to investigate, describe, or predict. For example, with the mathematical representation of the three laws forming classical mechanics, all movements of the macroscopic world can be described based on this *a priori* knowledge. In contrast to that, we use *a posteriori models* for data-driven models (typically statistical models). Such models often have an explorative character or are grounded in rather general theories; i.e., theories that postulate an interrelation between two entities without describing the underlying mechanisms exactly.

Both *a priori* and *a posteriori* models have a certain epistemological value. However, they differ in what we can learn from them: Oftentimes, *a posteriori* models outdo *a priori* models with respect to predictive power. However, they can become invalid when their underlying determinants ("hidden variables") change (see, e.g., Section 3.2 or [313]) or when they are used beyond the interval for which they were calculated[7]. *A priori* or *mechanistic* models aim at incorporating the mechanisms that drive a certain system's dynamics. Therefore, their intended space of application is larger; the number of case scenarios and the data interval for which they can be used is larger. On the other side, such *mechanistic* models – particularly when social processes are involved! – often show poor predictive power compared to well-fitted *a posteriori* models.

This epistemic distinction is closely tied to a functional classification of models: We distinguish *models for understanding* and *models for prediction*. With the former, researchers attempt to gain insight into a system's dynamics. Non-linear phenomena (e.g., the occurrence of thresholds) of complex systems can be investigated by means of such models. They can be used for theory building and falsification. In contrast, models for prediction can be completely wrong in terms of isomorphism / homomorphism, but are still good models when they fulfill their function by correctly predicting system states. Ideally, a model can fulfill both functions: helping

[7] There are several examples for phenomena that are non-linear, but can be approximated with a linear model for certain intervals. In Subsection 3.3.1, for instance, we introduce a model to describe the probability that contact between an infector and a susceptible host leads to infection. The assumed interrelation between the duration of such contact and the probability of infection is non-linear. For only slightly infectious diseases and/or low exposure, a linear approximation results in similar predictions.

to understand the dynamics of a system and providing good predictions. Nonetheless, very often a trade-off between these two functions has to be made, which is then a pragmatic decision of the researcher (cf. Subsection 1.2.1).

Modeling can be seen as a third approach of scientific inquiry apart from formal theory building and empirical studies [302]. Modeling can have epistemological components of each of the other approaches: As in theory building, the modeler has to translate observed or assumed processes from the sphere of 'reality' into relations between abstract entities (here mostly variables; cf. Figure 1.1). This can be understood as an encoding process [144]. On the other side, deductions derived from the formal model system are decoded [144] to 'reality' again for gaining understanding or for predicting its behavior. In that respect, model computations are like (physical) experiments: given "appropriate numerical tools" or simulation capacities, "a researcher can explore the behavior of physical [or other, TS] systems, as predicted by a set of governing equations, and look for interesting new effects" [359] (cf. Chapter 2).

1.2.3 What are good models?

In scientific discussion the term *validity* (lat. validus = strong, robust, or powerful) and the underlying concepts are frequently used for determining the quality of a model. In general, validity stands for the argumentative power of a proposition, theory, or investigation. The meaning of this word is closely related to the discussions about the nature and origins of knowledge introduced above.

Validity is an ambiguous term that is used for different, but similar concepts by different disciplines and schools in science [e.g., 4, 62, 298, 320]. Even within subfields of disciplines and for specific approaches, there is often no clear consensus on what *validity* means exactly or how it can be measured. For the field of modeling and simulation Kleindorfer et al. [187] exemplify this plurality as follows:

> "If one culls out the sections on validation from any sample of simulation papers, one is immediately struck by the wide variation to be found. There will be descriptions about model behavior, success in application, reservations and restrictions, personal experiences, descriptions of success in the field or lab—in short anything that the experimenter deems relevant to the experience of formulating and applying the model. This diversity, it seems to us, is an indication that at a fundamental level there is still confusion about what "validation" involves or in some cases if it is even feasible to talk about it." (p. 1088)

In agreement with our prior positioning, we query that it is possible to identify something like a "satisfactory model" that is "absolutely true" by means of any validation procedure [187]. Instead we agree with Oreskes et al. [255] that in "practice, few (if any) models are entirely confirmed by observational data, and *few* are entirely refuted" (p. 643). Following Barlas and Carpenter [25], validation "(confirmation) is inevitably relative. It is a matter of social conversation rather than objective confrontation" (p. 163). However, we explicitly do not take a radical relativist position that any model is as good as any other model and that there are no (partly) objective criteria with which model quality could be assessed [108, 187]. There is an intended space of applications for every model and a model has to perform relatively well with respect to the function it is intended to fulfill in order to be a good model.

From a functionalist perspective, it first has to be defined what the exact purpose of a specific model shall be. In the case of a *model for prediction*, for instance, it is important that the model is able to predict (usually for a more or less clearly defined case and a clearly defined parameter range) the outcome more or less correctly. Indicators have been defined in statistics to distinguish between good and bad models regarding prediction accuracy, e.g., by analyzing whether a model up with significantly better results than random guesses or by calculating the relative proportion of the variance explained by the model. In the case of a *model for understanding*, we usually judge qualitatively whether the model can reproduce empirically observable behavior or patterns. The actual purpose of a model has to be communicated by research and appropriate criteria on how to evaluate the validity of the model with respect to the purpose have to be discussed in the community. Thereby, such criteria can be algorithmic, but it will not always be possible to entirely define algorithmic criteria.

We see an analogy between (i) functionalism in the perception of the environment by a single human being and (ii) model / theory validation by a single researcher or within the scientific community. Our understanding of *validity* comprises both the deliberate choice of the model's function and the accuracy with which the model fulfills this function. This kind of validity is called "*functional validity*" (p. 346) by Scholz et al. [298] and is related to the understanding of human perception in psychology. As in Brunswik's lens model of perception [50], every complex model is based upon and further receives empirical cues about the reality it tries to reflect. The information upon which a complex model is based is often not consistent; it lacks "univocality" (p. 37) [298]. In practice, nobody will reject a model when it has failed once, but at the same time, nobody will trust a model that only delivers accurate outcomes from time to time. Therefore, we believe validation to be an evolutionary, probabilistic approach (similar to the idea of vicarious mediation by Brunswik [50, 51]) in which, in the long run, ongoing scientific dispute

and new supporting as well as contradicting empirical information can be used to figure out whether a model is "valid" or not (this idea is underlying the work presented in Chapter 5).

A further often-required quality criterion is *simplicity* or *parsimony* (often labeled as "Ockham's Razor")[8]. In the subsequent paragraphs, we will develop a position towards *parsimony*. Where not stated differently, we refer to Baker [18] throughout these paragraphs, as he provides a comprehensive review on the definition(s), the history, the relevance and various justifications of *simplicity* in science. Particularly for a deeper understanding of the justifications, we refer to Baker's work.

Although a vast majority of modelers and philosophers agree that models should be simple, the actual meaning of *simplicity* is as vague as that of *validity*, and the term has been used in various ways throughout the history of science. The most common formulation of Ockham's Razor is:

> "Entities are not the be multiplied beyond necessity." [18]

A common reinterpretation in contemporary philosophy of science is the following formulation:

> "Other things being equal, if T1 is more [...] parsimonious than T2 then it is rational to prefer T1 to T2." [18]

Parsimony can, thereby, be interpreted as an epistemic (the simpler the theory, the more likely it is true) or as a methodological (it is pragmatic to choose the simpler or two successful theories) principle. Neither formulation demands to always choose the simpler of two hypotheses, models, or theories. Instead, they restrict the principle with the words "necessity" and "other things being equal". That means they ask implicitly for a trade-off between *simplicity* and *validity*. Therefore, both formulations are of limited practical value, because "cases where competing hypotheses explain a phenomenon equally well are comparatively rare" (p. 145) [166].

Under different research paradigms various ways have been defined to weigh simplicity against validity: Baccini and Bader [16] suggested a quality measure for models, which is based on the reciprocal value of the sum of the number of model parameters and the deviation between model and measured data. For multiple regression analysis, several tests have been developed to distinguish predictor variables with a significant contribution to the prediction from variables

[8] However, one should be aware that there is also a minority position within the philosophy of science and various scientific communities, who actually take the opposite position and ask for principles of plentitude!

with an insignificant contribution (e.g., test for significance of beta-weights, cf. p. 436 in Bortz [43]). For curve-fitting there is also a rich variety of approaches weighing parsimony (as a measure against overfitting) against goodness-of-fit (seen as an indicator for validity) [126].

Kuhn [195] emphasized that the decision on the relative importance of particular scientific virtues, like simplicity or validity, was only a matter of personal or community preference. As with other decisions in modeling, there was no underlying objective truth with regard to the relative weight of simplicity versus accuracy. In line with that, we take the position that – analogous to the problem of validation – the question of parsimony has to be discussed and agreed upon individually for each specific model.

1.3 Individual level models of disease transmission

As stated earlier, one must have a sufficient understanding of the interaction between pathogen biology, host biology, environmental conditions, and the hosts' behavior, to be able to formulate adequate models of disease spread (cf. Subsection 1.1.3). In order to know, what *kind of host contacts* have the potential to transmit disease, a sound analysis of the host and the pathogen biology as well as the conditions of the physical environment have to come first (cf. the environment first postulate 7 in Table 1.2).

Host-pathogen interactions are multifaceted, reflecting the enormous variety of pathogens (for details on pathogen classification and characterization see Haas et al. [146], pp. 19-29). Modeling the risk of infection on an individual level therefore requires the integration of (i) a specific exposure model, which describes the amount of pathogens the host is exposed to, and (ii) a specific dose-response model, which describes the host's reaction to a certain exposure [146]. For the former, one has to consider the possible pathways of transmission, i.e., how the pathogen leaves the infector, how it distributes in the environment and how it enters the host (see Table 1.4). For the latter, an understanding of the host-pathogen interactions is needed. In the following, we will introduce some basic considerations for exposure and dose-response modeling with the help of the example of airborne diseases.

Exposure analysis always begins with the shedding of pathogens. In the case of airborne transmission, an infector sheds infectious material by talking, sneezing, coughing, or even breathing [261]. In an indoor setting, such an infector contaminates the indoor air and all susceptible persons who are in the same room are exposed to pathogens. If we assume – what is usually done [282, 287] – that one or more infectors generate aerosolized pathogens at a constant rate $q \ \left[s^{-1}\right]$ and if we know the supply of outdoor air $Q \ \left[m^3 \cdot s^{-1}\right]$, we can calculate the steady-state

concentration of pathogens in the well-mixed indoor air as $C = q/Q$ [9]. If the infection situation is not overly complicated, non-steady-state concentrations can also be derived analytically[10]. As soon as other – more complex – transmission pathways than the airborne route are involved, exposure models can become arbitrarily complicated [cf., e.g., 245]. Heterogeneities in shedding [64], environmental factors (mainly temperature or humidity) that inactivate pathogens [327], or settling [89] can further complicate aerosol transmission.

The prevailing and most parsimonious (cf. Subsection 1.2.3) theory of infection is that pathogens act independently (cf. Subsection 3.3.1). Thus, to become infected, a host must ingest at least one pathogen. Then, such an ingested pathogen must remain infectious and may not be removed before it reaches its target site in the host's body [145]. Thereby, the expected ingested dose $\overline{d}$ is a function of the exposure (in our example, the pathogen concentration in the indoor air C) and the consumption per exposure (in our example, the breathing rate p). If exposure and consumption are independent, the expected ingested dose $\overline{d}$ can be computed as the product of the means (cf. p. 162 of [146]):

$$\overline{d} = \overline{Cp} \tag{1.1}$$

The actually ingested dose is usually assumed to follow a Poisson distribution with parameter $\overline{d}$ [98].

Dose-response relations are usually modeled as stochastic processes (exemptions are deterministic threshold models as, e.g., used by Eubank [102]). As mentioned, there is evidence that one single pathogen is in principle sufficient to cause infection [220, 286, 354], but invading pathogens are continuously removed by the host's immune system or inactivated by environmental conditions. Riley et al. [282] were the first to formulate a probabilistic model of the airborne transmission of measles based on earlier observations of Wells [354]. They found that the

---

[9] $V\frac{dC}{dt} = q - Q \cdot C$; V $\left[m^3\right]$ is the volume of the room; the steady-state condition is $\frac{dC}{dt} = 0$.

[10] After the infector enters the room, the pathogen concentration is $C_{\uparrow}(t) = \frac{q}{Q}\left[1 - \exp\left(-\frac{Q}{V}t\right)\right]$; t $[s]$ is the time since the infector entered the room. When she/he leaves the room again, the pathogen concentration follows $C_{\downarrow}(t) = \left\{\frac{q}{Q}\left[1 - \exp\left(-\frac{Q}{V}t^*\right)\right]\right\} \cdot \exp\left[-\frac{Q}{V}\left(t - t^*\right)\right]$; t^* $[s]$ is the point in time when the infector left the room (relative to the time, when the infector entered the room).

exposure-time-dependent probability that a fully susceptible individual gets infected in a steady-state indoor air setting can be described as

$$P = 1 - \exp\left(-\frac{Ipqt}{Q}\right) \tag{1.2}$$

with I being the number of infectors in the indoor space; p being the volume inhaled per time unit; q being the quantum[11] generation rate; t stands for the exposure time and Q is the (already mentioned) fresh air supply.

Transmission pathway	Description
Droplet	In this thesis, droplet transmission means transmission only via large droplets. As such droplets fall out quickly, transmission can only take place up to a distance of max. two meters from the infector. Droplets are expelled by talking, sneezing, coughing, etc. Uptake is usually via the respiratory path. Deposition and subsequent contact transmission can play a role. [47, 89, 146, 330]
Airborne / aerosol	Airborne or aerosol transmission differs from what we call droplet transmission by the size of droplets and, thus, by the physical behavior of these droplets. Aerosol particles are small enough to remain suspended in the air for a long time due to their low settling velocities. Such small particles are known to be generated by the same mechanisms as larger droplets. Furthermore, activities like breathing, but also vomiting or flushing the toilet can lead to aerosolized pathogens. Uptake is via the respiratory path. [47, 89, 146, 261, 330, 354]
Direct / indirect contact	Surfaces like hands, doorknobs or items of clothing can be contaminated with pathogenic organisms (see also droplet transmission) or patients can have purulent lesions / affected mucous membranes. Transmission takes place when contaminated surfaces and afterwards locations of entry are touched. Depending on the pathogen, relevant locations of entry can be specific mucous membranes or the skin. [146]
Sexual intercourse	Transmission takes place via the surfaces in contact (typically bacteria) or from secretions that carry infectious agents (typically viruses). Transmission can be penile-vaginal, penile-oral, penile-anal or vaginal-oral.
Fecal-oral	In the fecal-oral route, pathogen-laden feces contaminate drinking water or food that will be ingested or surfaces that might be touched. Uptake is via the mouth. The airborne route can also play a role in transmission of such pathogens (e.g., when flushing the toilet). [146]
Iatrogenic	Iatrogenic transmission includes all infections caused by medical treatment. Typically pathogens enter the host during surgeries (hygiene) or with transplants or contaminated infusions. [146]
Vector-borne	Transmission via intermediate hosts of other species, often insects. The typical route of entry is via bites or stings.

Table 1.4: Selected transmission pathways and their definitions.

[11] A quantum is a dummy measure that was introduced by Wells. It is defined as the amount of infectious material "infecting 63.2 per cent of homogeneously exposed hosts" (p. 123) [354]. The quantum generation rate is the amount of quanta generated per time unit.

Of course, host responses are much more complex than represented in the described Wells-Riley model of airborne transmission. Exposure can result in more different states than captured by the binary distinction between the infected and non-infected. Epidemiologists often distinguish between (i) exposure without colonization, (ii) infection without clinical symptoms, and (iii) clinical cases. The last group is often differentiated into mild, moderate, severe and fatal cases (cf. p. 32 of [146]; see also Chapter 5). The susceptibility of a host is affected by various factors including age, alcoholism, chronic disease, double infection, pre-exposure serological status, nutritional status [146]. Furthermore, it is a naïve assumption that the risk of infection would depend solely on the *total* dose ingested. In fact, the risk also depends on the temporal patterns with which the pathogens are ingested (see Chapter 3).

Nonetheless, more or less complex models of transmission risk can complement population level analysis of disease spread. The kind of analysis presented in this section can help to identify *relevant* kinds of contact between hosts, which then can be incorporated in population level models. Additionally, models of transmission probabilities allow weighing different kinds of potentially contagious contacts, which otherwise are usually treated all the same in models of epidemics. In the following section, we will review population level models of disease spread and indicate how the individual-based insights can be integrated on a population level.

1.4 Population level models of disease spread

Mathematical modeling of infectious disease spread is quite old and goes – at least – back to the 18th century. In the 1760's Daniel Bernoulli, for instance, used an actuarial approach to investigate the excess morbidity and mortality caused by smallpox. He further described the emergence of (partial) immunity after infection and calculated the benefit, if people would be variolated (i.e., purposefully infected with pox) as a means of immunization [30]. Since then, manifold approaches to describe disease propagation on a population level have been developed, be it simple mathematical descriptions as in Hamer [150], early compartmental approaches like that of Kermack and McKendrick [180] or modern, sophisticated, individual-based models [102].

In the following subsections we will introduce three commonly used model types used for modeling disease spread and describe their advantages and shortcomings from a transmission system perspective: deterministic compartmental models, network models and individual-based models.

1.4.1 Deterministic compartmental models

Compartmental models are composed of a number of variables ($\in \mathbb{R}$) that represent compartments or partitions of a system (cf. Figure 1.1). Dynamics are governed by flows between the compartments and flows from or to the environment, which are described by means of differential equations [134]. In general, compartmental models with n compartments are mathematically expressed as follows [cf. 291]:

$$\begin{aligned} &\frac{dx_i}{dt} = f_{io} + \sum_{j=1}^{n} f_{ij} - \sum_{j=1}^{n} f_{ji} - f_{oi} \\ &t \geq 0 \\ &x_i(0) = x_{oi} \\ &i = 1,2,\ldots,n \end{aligned} \tag{1.3}$$

Thereby x_i is the value (e.g., number of individuals, concentration of a substance, amount of material, money in an account, etc.) attributed to compartment i; x_{oi} is the initial value of compartment i; f_{io} denotes the flow from the environment o to compartment i; f_{ij} stands for the flow from compartment j to compartment i; f_{ji} gives the flow from compartment i to compartment j; finally, f_{oi} is the flow from compartment i to the environment.

As Godfrey [134] points out, the "fact that many [...] differential equations are compartmental without necessarily being described as such makes tracing the exact origins of compartmental models rather difficult" (p. 6). Compartmental models are widely used in various thematic fields such as the global carbon cycle or phosphorus concentrations in lakes [168], the flow of radioactive tracers in the environment [376], the distribution of xenobiotic chemicals [240] and nanomaterials [135] in the environment, material flows in the anthroposphere [16], species relations in ecology [253], or in pharmacokinetics [331]. In the context of epidemiological questions, the use of compartmental models can be traced back approximately one century. The history of compartmental models for describing disease spread is closely linked to the idea of mass action coming from chemical kinetics. Therefore, we will subsequently discuss the mass action principle and its implications for modeling disease spread, before going back to the advantages and shortcomings of deterministic compartmental models for disease spread in general.

The mass action principle and compartmental models

The mass action principle applied in chemistry describes the reaction rate v of a chemical reaction. If two kinds of reactants are involved in the reaction, the corresponding mathematical equation reads

$$v = k \cdot [A] \cdot [B] \tag{1.4}$$

under the condition that the reactants A and B are well-stirred. Then k is the rate constant and $[A]$ and $[B]$ are the concentrations of the respective reactants.

According to Heesterbeek [155] the analogy between collisions of chemical substances (leading to a chemical reaction) and the meeting of individuals (leading to infection) was first formulated simultaneously by Ross and McKendrick. The mass action metaphor underlying most compartmental models assumes that (i) individuals mix homogeneously ("well-stirred"), (ii) that there are only two kinds of individuals which drive the dynamics of on outbreak – susceptible ones and infectors – which don't differ in their other characteristics, and (iii) that contacts are transient and that there is no path dependence or memory (i.e., there are no stable relations between individuals). Of McKendrick it is known that he looked intensively into chemical kinetics (cf. p. 93 of [155]) and that he explicitly made metaphorical reference to the idea of two gas molecules colliding when he introduced his formalism to describe infectious disease:

> "The rate at which this epidemic will spread depends obviously in the number of infected animals, and also on the number of animals which remain to be infected—in other words the occurrence of a new infection depends on a *collision* [italicized by TS] between an infected and an uninfected animal." (p. 54) [217]

Analogue to reaction kinetics, the rate of new infections is proportional either to the number, the density, or the proportion (this depends on the definition of mass action, cf. [175]) of infectious and susceptible individuals.

Most deterministic compartmental models of disease spread rely on the mass action assumption to describe the mixing between the various groups of individuals. Applied to the common SIR[12]-model of disease spread (with demography) the resulting system of equations reads [8, 178]

$$\begin{aligned} \frac{dS}{dt} &= \mu - \beta SI - \mu S \\ \frac{dI}{dt} &= \beta SI - \gamma I - \mu I \\ \frac{dR}{dt} &= \gamma I - \mu R \\ S + I + R &= 1 \end{aligned} \tag{1.5}$$

with S, I, R being the proportions of susceptible, infectious, and recovered individuals. β is the product of contact rate and per-contact transmission probability. μ is the natural mortality (note that the total population in Equations 1.5 is constant as $\mu - \mu S - \mu I - \mu R = 0$). γ stands for the recovery rate. As can be seen, the SIR-model assumes lifelong immunity after recovery, because there is no flow from the recovered compartment back to the susceptible compartment nor to the infectious compartment. Analogous definitions exist for host-pathogen constellations in which infection confers no immunity (SIS-models) or in which an intermediate "exposed" or "latency" state is assumed between the susceptible and infectious compartments (SEIR-models).

Advantages and disadvantages of compartmental models

The advantages of these classical compartmental models are manifold: (i) The analogy between molecules colliding in a gas and reacting to build another molecule and individuals meeting in space and infecting each other makes intuitively sense as a very rough approximation to reality. At least we have with the mass action principle a more or less plausible, *mechanistic* model of infection that can be captured with simple mathematical expressions. (ii) Due to their mathematical simplicity, these compartmental models can be studied analytically, whicht makes them, on the one hand, easily manageable and, on the other hand, interesting for scientific theorizing. Some of the most important indices in epidemiology (e.g. the basic reproduction number, for Equations 1.5 $R_0 = \beta/(\gamma + \mu)$, which is the number of secondary cases induced by one infector introduced in a fully susceptible population) were defined and derived on the basis of these simple models (for further examples refer to [178]). (iii) From a parsimony perspective (cf. Subsection 1.2.3), these classical, simple compartmental models are also "good" models, because they incorporate no more details than needed to reproduce and explain observed behavior (in fact, they reflect too few mechanisms – we will discuss this subsequently).

[12] S=susceptible, I=infectious, R=recovered

However, from a transmission system perspective (cf. Subsection 1.1.3), compartmental models of disease spread also show vast disadvantages. As early as 1929 it was Soper [316], who stated that it is, "perhaps, a false analogy between infection in disease and the mechanism understood under the name of chemical mass action" (p. 54). He points to the fact that – in contrast to the optimal and controlled conditions in a chemical reactor – societies exhibit "imperfect mixture". In fact, *all* mixing assumptions underlying mass action do not hold in societal reality: Everyday experience and a plentitude of scientific studies [94, 95, 203, 204, 221, 237, 278, 313] show us (i) that contacts between humans are not made randomly; (ii) that there are transient contacts, but also very stable long-term relations (such as stable sexual partnerships or family relations); and (iii) that there is intra- and inter-individual variability in the number of contact partners people have.

Furthermore, not only is the *social* component of mixing represented with questionable assumptions in classical compartmental models –*biological* qualities are also represented in a simplified manner. One of the implicit assumptions of the classical SIR-, SIS-, or SEIR-models is that there are only two kinds of compartments essentially driving disease spread dynamics: the compartment representing the group of infectors and the compartment representing the susceptible group. In contrast to that, reality is often far more complex: For instance, in the case of influenza we observe complex patterns of pre-existing immunity (making the "group" of susceptible individuals heterogeneous) and different levels of viral shedding (consequently, infectors differ in their infectiousness) – for details see Subsections 5.3.5–5.3.7. β, the product of contact rate and per-contact transmission probability, is the parameter in Equations 1.5 determining how severe an outbreak will be. In Chapter 3 we will see that not only contact patterns differ inter-individually – also the per-contact transmission probability is highly variable. Another unrealistic biological feature of the classical compartmental models is the exponentially distributed infectious period ($1/\gamma$), which conflicts with the knowledge that the infectious period has a strong central tendency for most diseases and gamma distributed infectious periods would be much more appropriate (cf. Subsection 2.4.4 and [349]).

Two further shortcomings of deterministic compartmental models lie in their mathematical structure: As they are mathematically nothing else than a set of ordinary differential equations, such models are inherently continuous and non-stochastic. The non-discrete nature of compartmental models leads to some peculiarities, which are hard to interpret (e.g., non-integer values for the number of infectors). Along with this is the fact that a disease never dies out completely in a model as described with Equations 1.5. The compartment I can become arbitrarily small, but never approaches zero. The fact that classical compartmental models are deterministic does not play a negative role when the proportion of infectors I is not close to zero and the

infectiousness and the contact rate are sufficiently high (i.e., $\beta >> \gamma + \mu$). If, however, a disease is only slightly contagious or if an epidemic is still in its very beginning, stochastic effects dominate the further fate of the outbreak (cf. to Chapters 2 and 5).

Most of the disadvantages described above can be addressed to a certain degree *within* the framework of deterministic compartmental models – however, not without cost and not perfectly. Two shortcomings, the lacking heterogeneity in social contact patterns as well as the unrealistic infectious period, can be tackled with more fine-grained compartmentalization:

1) *Contact heterogeneity* can be introduced into compartmental models by dividing all existing compartments (i.e., the susceptible, the infectious and the recovered compartment in the case of an SIR-model) into sub-compartments. Typical sub-divisions are according to age, risk behavior, or spatial entities. Age-dependent differences regarding contact rate and transmission probability can be treated by dividing the population into n age classes. The model then can be parameterized by means of an $n \times n$ transmission rate matrix replacing β, which is usually called the WAIFW[13] matrix (cf. pp. 175 & 675 of [8]). A typical example for a sub-compartmentalization along risk groups is the model of gonorrhea transmission by Hethcote and Yorke [160]. They differentiate in their model two groups of individuals of which the "core group" is ten times as sexually active as the other group. Spatial heterogeneity is typically implemented by so-called metapopulation models [e.g., 141, 142, 206]. Metapopulation models make sense whenever mixing is sufficiently patchy for sub-compartmentalizing the population in a reasonable manner [175, 178]. Usually, human populations are divided along municipal, state, or national borders.

2) The assumption of *exponentially distributed infectious periods* can be easily relaxed by dividing the infectious class into n subclasses, which are passed in sequential order [349]. For $n = 1$ we have the classical SIR model with exponentially distributed infectious periods as given in Equations 1.5. For $n \to \infty$ we approach a fixed infectious period. All n in between those two extremes lead to

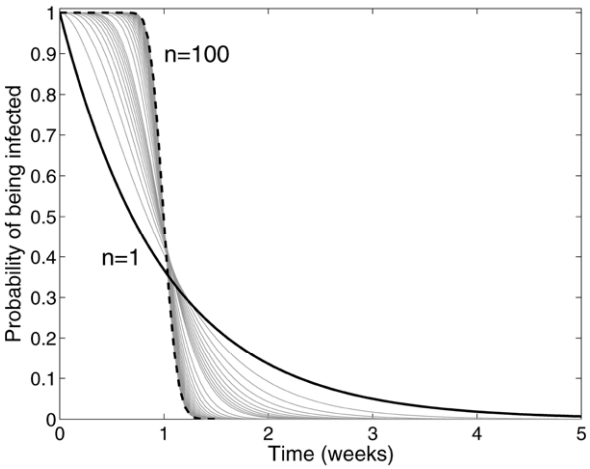


Figure 1.4: Distribution of the infectious period versus the number of subcompartments *n*. Figure has been taken from [349] under the CCAL license.

[13] WAIFW=**W**ho **A**cquires **I**nfection **F**rom **W**hom

gamma-distributed infectious periods of various shapes (as can be seen in Figure 1.4), which are deemed to be more realistic than the two extremes (cf. Subsection 2.4.4, as well as [205, 349]).

Although sub-compartmentalization offers the possibility to relax some of the most unrealistic assumptions of deterministic compartmental models, several drawbacks persist and new ones emerge in further compartmentalization. As Koopman points out, even highly partitioned compartmental models "retain an essential core of mass-action" (p. 147) [192], as at the most granular level, mixing still follows the mathematics of mass action. That means, contacts are still random and transient and subpopulations still behave homogeneously. Furthermore, metapopulation models partitioning populations into various spatial patches may lead to asynchrony between the patches, but for $t \to \infty$ they are homogeneous [175]. That means spatial steady-state heterogeneity (which can be observed in other model classes) still does not occur. Finally, the most important drawback of highly partitioned compartmental models is that their major advantage of simplicity vanishes (cf. Subsection 1.2.3): Every additional property, along which the population is further partitioned, vastly increases the total number of compartments and flows to describe. The InfluSim pandemic influenza model developed by Eichner et al. [96], which is based on far more than 1000 differential equations without being spatially explicit, might serve as an example.

1.4.2 Network models

As pointed out in the previous section, compartmental models fail to adequately represent certain properties of social structure relevant for disease spread. However granular the compartmentalization might be – on the most detailed level, random mixing persists. Network models offer a way to elegantly include many kinds of contact heterogeneity in models of disease spread, which cannot be represented in compartmental models. Network representations proved to be helpful for describing and understanding particular properties of the world wide web, the Internet (hardware), collaborations of scientists and movie actors, human sexual contacts, interactions between cells in organisms, predator-prey relations, language, power grids, the brain of nematodes, and protein folding [5]. As with other exchangeable entities (products, ideas, norms, information, money, etc.), social relations serve as "channels" for infectious diseases by which pathogens "move" through society [231].

Network models go back to graph theory [38]. In their simplest form, they consist of *nodes* and *links*, which connect nodes. Attributes of the nodes are called *composition variables*, whereas attributes of the links are *structural variables*. Depending on the scientific community in which network models are applied, the terminology can differ. In the field of infectious disease spread, nodes usually stand for hosts and links are contacts between hosts, which are, in principle,

sufficient to transmit a certain infectious disease. In the case of droplet-transmitted diseases, there is usually only one host type included in the model. The corresponding network representation is called a *one-mode network* (cf. p. 29 [347]). There are also examples with a bipartite population: In a heterosexual population, for instance, nodes can only be linked to nodes of the other sex [244]. As we have two types of nodes, we call such a network representation a *two-mode network* (also p. 29 [347]). Finally, there is the case in which infectors do not transmit infection directly to susceptible hosts, but contaminate parts of the inanimate environment. Such indirect transmission is best represented in *affiliation networks* (p. 30 [347]) that relate individuals to places. Networks are often represented in the form of adjacency matrices $\mathbf{X}$, in which the (usually binary) value of the matrix element x_{ij} signals whether node i is linked to node j.

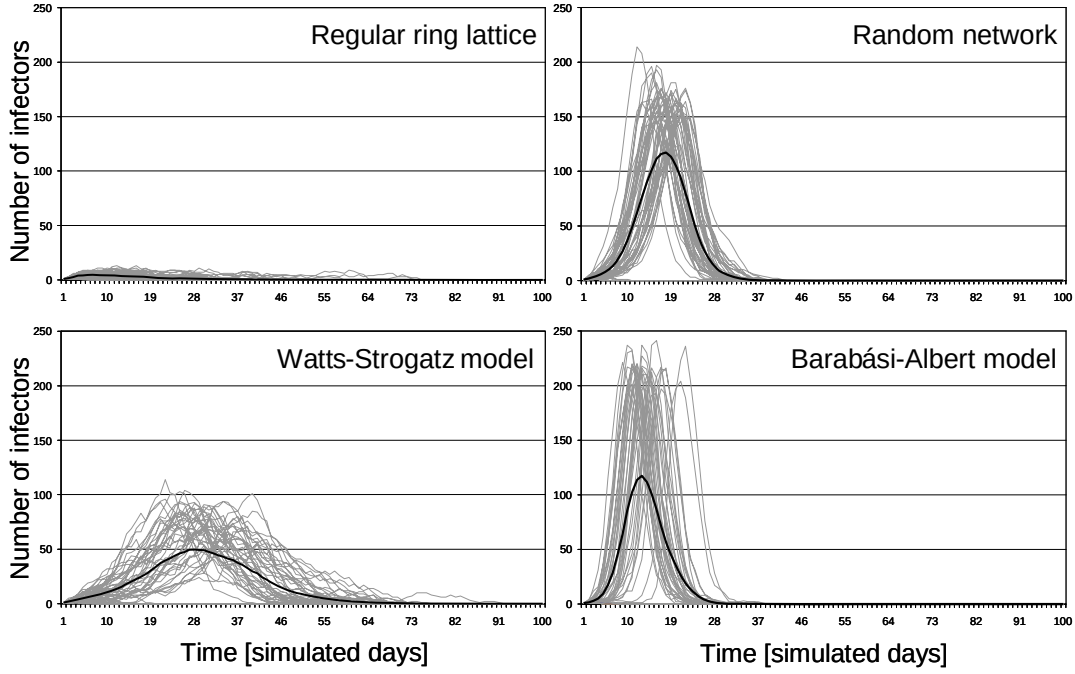


Figure 1.5: Number of infectors versus time for four different network archetypes. Besides the contact structure, all simulation parameters were equal. In each subfigure, every grey line shows one out of 50 simulation runs. The black line gives the arithmetic mean of all 50 runs. The figure has been taken from [113]; the copyright holder is the author of this thesis.

The difference contact heterogeneity makes can be seen in Figure 1.5. This figure shows the results of SIR-type simulations that were computed for four different archetypical network structures. All four examples were parameterized identically: the population size was set to 500; every node had in average six contact partners (i.e., 1500 links were distributed over the entire network); the infectious period was fixed to three days; the probability that a contact between an infector and a susceptible host leads to transmission during one day was set to 0.11. The only

difference between all four examples was the actual configuration of the contacts. In Figure 1.5 it is evident that the infection dynamics differ greatly depending on the social structure. Subsequently, we will introduce all four network archetypes and discuss the differences and peculiarities observed in Figure 1.5.

Infectious disease spread in four network archetypes

Regular lattices and random networks (see below) are the simplest network structures. The basic characteristic of regular lattices is that the nodes are connected only to their nearest neighborhood in a regular manner. Often-used schemes to connect nodes for generating regular lattices are the von Neumann neighborhood, the Moore neighborhood, and hexagonal neighborhoods (for details refer to [3]). Hence, the structure underlying cellular automata is nothing more than a regular lattice network representations [175, 231].

In Figure 1.5, we observe for the regular ring lattice (for details see the paragraphs about small-world networks and [348]) under the given parameterization that the epidemic is rather mild compared to those of the other network structures. This can be explained with the locality and the one-dimensional character of this archetypical contact structure: In a regular lattice, disease spreads in wave-like patterns with a clear front line. In contrast to a two-dimensional lattice, in a one-dimensional lattice the outbreak can only move along the line of nodes in two directions.

A two-dimensional regular lattice can be seen as an exceedingly idealized representation of the contact structure of pre-industrialized human populations, because these populations only had very slow means of transportation and, consequently, very small daily cruising radii. The spread patterns generated with two-dimensional regular lattice models resemble those described for the bubonic plague in medieval Europe [69, 154, 375]. Disease spread in wild animals has some similarities with simulated spread in regular lattices, too: In the mid-twentieth century, rabies spread in wave-like patterns through Europe [258], a phenomenon that was also captured with two-dimensional cellular automatons [307].

In *random networks*, nodes are linked in a random manner: Each pair of nodes has an equal chance to be connected. Although contacts are not transient as under the mass-action assumption, random networks are the network representation conceptually closest to compartmental models and they also behave similarly [24]. Random networks show interesting properties such as a phase transition when the average number of contacts per nodes is increased[14] [100] and are – like mass action models – accessible to analytical treatment. The

[14] A giant component of interconnected nodes emerges at an average of one contact per node.

shape of the curve shown in Figure 1.5 resembles that of a compartmental SIR-model with a fixed infectious period.

Within the model class of *small-world networks*, the Watts-Strogatz model [348] is the particular type most often referred to. The Watts-Strogatz model can be defined as a blend of a regular ring lattice model and a random network. To generate such models, one starts with a regular (in this case ring) lattice and re-wires the existing links with a defined probability P_{rw}. The two extremes of the Watts-Strogatz model are a pure regular ring lattice ($P_{rw} = 0$) and a pure random network ($P_{rw} = 1$). Consequently, if small-world networks are used in simulations of disease spread, they also show a mixture of the characteristics of both underlying network extremes. In two-dimensional space, we observe diffusive spread with clear front lines (coming from the lattice component), which are ruptured by new sources of infection popping up at remote places of the social space. In time, we observe small-world models to be retarded compared to random mixing models and to be accelerated compared to cellular automata. With an increasing number of remote contacts, there is a transition from independent outbreaks in different parts of the network to one synchronized outbreak [197].

Similarities between observed patterns of disease spread and characteristics of a two-dimensional small-world epidemic model are, for example, given in the case of the 1918/19 influenza pandemic in the USA. The pandemic seems to have started at the two coasts of the USA and moved towards the interior in a wave-like manner. However, before the front line reached the central parts of the country, several local epidemics broke out and established new centers of a regional wave-like spread of the disease (cf. spread map on p. 65 in [75]). Also, in some animal societies like whales, small-world-like contact patterns have been described that caused concern over how vulnerable such societies might be against severe infectious disease epidemics [143].

Finally, *scale-free networks*, here represented by the Barabási-Albert model [5], exhibit epidemiological dynamics that are mainly driven by the dispersion of the links per node. Scale-free networks assume a power-law distribution of the nodal degree (= number of links), which results in many nodes having only a few contacts and few nodes having many contacts. Those highly connected individuals are often referred to as “hubs”.

In our concrete simulation runs, the epidemic either did not take off at all or accelerated greatly. This can be explained by the interplay of initial conditions and the structural properties of the Barabási-Albert model: If the initially infected cases are rather isolated individuals, there is a considerable chance that the disease will die out before a substantial outbreak occurs; if the chain of infections, however, reaches such a hub, this hub then acts as a super-spreader and has the chance to generate plenty of secondary cases.

In a network structure, hubs play a central role in accelerating the velocity with which a disease spreads and they stabilize outbreaks to that effects that it becomes less likely that a disease will die out just by chance. In fact, Pastor-Satorras and Vespignani were able to prove mathematically, that in the case of an infinite scale-free network, a disease can persist "at whatever spreading rate the epidemic agents possess" (p. 3200) [263]. Liljeros et al. were able to show for Sweden that the distribution of the number of sexual partners follows a power-law [203]. In light of such findings, scale-free networks are mostly discussed in the context of sexually-transmitted diseases.

The archetypical, idealized network models described above do not capture all the epidemiologically relevant properties of contact structure observed in the real world. If the factors that classify an interaction between two hosts as potentially contagious contact are known, one can, in principle, measure contact networks empirically. A book edited by Morris [232], for instance, gives plenty of guidance for survey designs and data collection in the field of network epidemiology. In various fields of network research, helpful indicators for characterizing such complex networks have been developed. Several of these indicators can be re-interpreted in a meaningful way for epidemiological applications. The following paragraphs will offer some insights as to what kind of indices and measures network analysis can offer to better understand disease transmission systems.

Epidemiologically relevant network indices

One important factor that affects the individual risk of becoming infected is the *centrality* of the individual. Several centrality indices have been defined, all of which measure how central the position of a node is in the entire network. Here, we discuss *degree centrality*, *closeness centrality* and *betweenness centrality*.

The *degree centrality* is nothing more than the degree of a node[15]. In social network analysis the term degree refers to the number of links a node has (i.e., for epidemiological applications, the number of potentially infectious contacts a host has). It is obvious that a high degree centrality corresponds to an increased risk of becoming infected (more contact partners, all of which are potential infectors) as well as to more secondary cases on average. We know from several

[15] $C_D(n_i) = d(n_i) = \sum_j x_{ij}$ with x_{ij} being the matrix elements of the adjacency matrix $\mathbf{X}$, n_i denoting a specific node i, and i and j being indices pointing at specific nodes. The degree centrality can be standardized by dividing $C_D(n_i)$ by the amount of other nodes in the network. For details see pp. 178-179 in [347]

studies that a high dispersion in degree centrality decreases the epidemic threshold on the population level [22]. That means that only slightly infectious pathogens can be sustained in such a host population and that such pathogens are even capable of causing major outbreaks (see the paragraphs on scale-free networks).

The rationale underlying *closeness centrality* is similar to that of degree centrality: Both measure how well a node can be reached by other nodes. The difference lies in the shift from the individual (egocentric) to the network perspective. While degree centrality focuses on the individual and its direct alters, closeness centrality measures how close a node is to *all* other nodes in the network. The most common version of closeness centrality has been provided by Sabidussi [290], who defined it as the inverse of the sum of the lengths of all shortest distances between a certain node and all other nodes[16]. If a node is rather distant to all other nodes, the value of the closeness centrality will be small. If all other parts can be reached with only few intermediate nodes, the closeness centrality index will have a high value. In disease spread network models, individuals with a high closeness centrality are likely to be infected early, as they can be reached from all parts of the network within a short time. Further, it will be difficult to encapsulate an outbreak locally if central individuals are involved.

Betweenness centrality measures a completely different aspect of centrality than degree and closeness centrality. To our knowledge, Bavelas [27] and Shimbel [311] were the first to express what is meant by betweenness centrality. Shimbel wrote:

> "Suppose that in order for site *i* [i.e., node, TS] to contact site *j*, site *k must* be used as an intermediate station. Site *k* in such a network has a certain "responsibility" to sites *i* and *j*.
>
> If we count *all* of the *minimum* paths which pass through site *k*, then we have a measure of the "stress" which site *k* must undergo during the activity of the network." (p. 507) [311]

[16] $C_C(n_i)=\left[\sum_j d(n_i,n_j)\right]^{-1}$ with $d(n_i,n_j)$ being the shortest distance between node i and node j (p. 184-186 in [347] and [290]). The shortest distance between two nodes is defined as the minimal number of links that have to be passed to come from one node to another.

Freeman [118] quantified betweenness centrality by defining it as the sum of the proportion of all shortest paths between all pairs of nodes that go through the node of interest[17]. Consequently, all nodes with a high betweenness centrality act as "bridges" between different "islands" of the network. For epidemiological applications, betweenness centrality can help to identify individuals, which are effective targets for interventions: When, for example, an individual with a high betweenness centrality gets vaccinated, the distances between many other individuals will increase greatly.

All different indices of centrality have been proven to be strong predictors for the risk of acquiring infectious diseases, particularly in studies of sexually transmitted infections [70, 129].

Another index relevant for the dynamics of infectious disease spread is *clustering*. Clustering basically describes the degree to which "friends of my friends are also my friends". One common mathematical definition of the clustering coefficient CC is the ratio between the number of triangular loops and the total number of possible triplets [175, 348]. In disease transmission systems, clusters fulfill two functions:

1) If one or a few individuals belonging to a certain cluster are infectors, it is rather likely that all or – at least – a major part of the other individuals belonging to this cluster will also be infected. Clustering generates redundancy: In a dense cluster of mutually interconnected people, there will be many short paths between any given pair of individuals belonging to this cluster. Consequently, it makes sense, e.g., in case of sexually transmitted diseases, to identify and target entire clusters instead of single individuals and their alters.

2) On a population level, clustering acts as a "break" [314]. As mentioned, clustering increases the probability that individuals within a highly clustered environment become infected if one individual within the cluster in infected. The same mechanism, however, leads to an increase in locality of resources, i.e., susceptible individuals available for infection. Under the random mixing assumption, the probability that an infector "collides" with a susceptible individual is proportional to the density of susceptible individuals in the whole population. In a network model, clustering can lead to a situation where in the local neighborhood of an infector almost no susceptible individual is left although, on a population level, this group still constitutes a large proportion of the entire population.

---

[17] $C_B(n_i) = \sum_{j<k} \frac{g_{jk}(n_i)}{g_{jk}}$ with g_{jk} being the number of all shortest paths between nodes j and k and $g_{jk}(n_i)$ being all shortest paths between nodes j and k going through node i (p. 190-191 in [347] and [118]).

Advantages and disadvantages of network models

Despite all the advantages of network models compared to compartmental models, there are several properties of network models that can turn out to be problematic for models of disease spread. Two of them will be discussed here: 1) the neglected quality of contacts and 2) the lack of manageable dynamic network approaches.

1) *Quality of contacts*: Usually, adjacency matrices representing (epidemic) contact networks are binary. Thus, such network models ignore – like most compartmental models – that contacts between hosts can differ in type (e.g., direct vs. indirect), duration, intensity, and other relevant manners. However, transmission probabilities and, thus, the epidemiological relevance of certain contacts can depend on such characteristics. The dependency between duration, intensity, and transmission probability in the case of droplet transmitted diseases will be shown in depth in Chapter 3. In principle, this drawback could be overcome by weighing the links connecting the nodes directly by their individual transmission probabilities. Nonetheless, contacts are not equally intensive and of the same duration every day, so a dynamic representation of contact quality allowing for varying daily transmission probabilities would be needed (see below). Furthermore, Pujol et al. [268] showed that even the transmission probabilities within one day are inherently governed by the concrete *dynamics* of the infection situation. A prolonged but non-intensive contact is not the same as a short, but very intensive contact – even if the same amount of infectious material is exchanged in both situations. Dealing with these time-dependent properties of contacts adequately within the network paradigm appears to be impossible.

2) *Dynamic networks*: The second disadvantage of the network paradigm with respect to models of epidemics is the static character of the common network models. In that sense, network models are the complementary extreme to mass action models, as the latter assume constantly changing contacts while the former assume an entirely stable structure. Both extremes are unrealistic representations of reality as humans have both stable relations which result in high frequencies of potentially contagious events as well as transient encounters, which never will never be repeated. In Chapter 2 we will see that both extremes can be dysfunctional for adequately understanding or predicting the spread of infectious diseases.

There are several approaches – within the field of disease spread and in general – to make the network paradigm dynamic [278, 315, 358]. Nonetheless, most of the indicators described for network analysis are only defined for static networks. For instance, the concept of clustering (p. 42) needs a static structure to be calculated (cf. to Chapter 2). Also the centrality measures described above and other valuable concepts like reachability matrices [347] are not defined for constantly changing contact configurations.

1.4.3 Individual-based models

The last category of models applicable in the field of infectious disease spread that we want to discuss here are individual-based models – nowadays widely used to simulate the spread of infectious diseases [e.g., 22, 41, 63, 65, 72, 77, 102, 110, 111]. Individual-based models are the most flexible type of models, overcoming all the disadvantages mentioned for the other two model types.

As indicated by their name, the unit of operation in an individual-based model is the individual. All individual-based models are built upon a community of simulated individuals that are situated in an environment. Thereby, every individual has its role and interacts with its environment [132, 353]. This focus on the individual makes them predestined for epidemic models: As in models of evolution, where "individuals are the unit of selection" (p. 11) [170] individuals are the entities, that take up pathogens with their behavior and are the carriers of pathogens and transmit disease by setting pathogens free. For this reason, individual-based models are also said to be better and more intuitively understandable for empiricists than matrix algebra (for social network analysis) or differential equations (in compartmental models) [170]. Also, for the communication with decision-makers, stakeholders in general or the public, this intuitive comprehensibility of individual-based models is advantageous.

Compared to compartmental models or network models, it is not so easy to define the boundaries of the model class of individual-based models. There is also a fuzziness in terminology. In principle, network models can be seen as a special case of individual-based models. This becomes clear when looking at the regular lattice model / cellular automaton relation: Cellular automata belong – without any doubt – to the class of individual-based models as their unit of operation are the cells of the lattice representing individual entities. At the same time "cellular automata models are a narrow class of networks in which only nearest-neighbour individuals are connected" (p. 120) [175].

There are several terms often used for similar kinds of modeling approaches that have individuals as units of operation, e.g., individual-based model, individual-based simulation, microsimulation, multi-agent model, multi-agent system, etc. Differences between the various concepts can be found in the literature, but mutual discriminations remain fuzzy as there are no universally accepted definitions of the various terms (cf., e.g., to Wooldridge [362]). Besides the focus on individuals as units of operation, we see two further characteristics that are distinctive for individual-based models (cf. Holland [164]):

1) *Parallelism*: All dynamics of individual-based models are generated by the interaction of their individuals acting in parallel. All individuals act on their respective simulated environment (which usually includes parts or sometimes all of the other individuals) simultaneously, and all individuals are affected by the state of their simulated environment (i.e., the actions of the other individuals) at the same time.

2) *Conditional action*: Changes in the model behavior of an individual in an individual-based model usually depends on the signals it receives from its environment. The reaction to the specific environment of an individual can be entirely deterministic or can include stochastic elements. Typically, such reactions are implemented in a rule-based manner (IF-THEN operations). For models of disease spread, such a rule can be IF the individual is susceptible and if there was a contact with an infector, THEN switch the own status from susceptible to infectious with a certain probability p (cf. Figure 2.1 and Section 5.3).

Another characteristic often mentioned in this context (e.g., by Holland [164] as a characterization of complex adaptive systems) is adaptation, learning or evolution. We use this model feature to distinguish individual-based models in general from agent-based models, which we see as a subset of individual-based models. Agent-based models in our understanding are tightly bound to the concept and meaning of agency (Merriam-Webster defines agency as "the capacity, condition, or state of acting or of exerting power" [2]), which is often associated with intelligence, learning [169], or flexibility [362].

In most cases, individual-based models of epidemics are not agent-based according to this distinction, as they involve no learning processes. Usually, individual-based models of epidemics incorporate rules, how individuals make contact or react to illness (e.g., by home confinement), but do not allow for adaptive changes of these rules. However, in a broader transmission system context (cf. Section 1.1.2), including evolutionary or learning process can be insightful. Boots and Sasaki were, for instance, able to simulate the interdependence of host contact patterns and pathogen evolution by including evolutionary processes in their individual-bases model [41]. Furthermore, optimal rules regarding quarantine can be identified by building adaptive transmission models applying genetic algorithms [370].

Advantages and disadvantages of individual-based models

The advantages of individual-based models are obvious: They are an intuitively comprehensible and flexible tool that fits – due to its flexibility – almost any modeling problem in the broader field of infection transmission systems. They offer a distinct explanatory power over previous approaches, "in that observed phenomena emerge from interactions, rather than being imposed by the modeling framework" [343]. Thereby, they also avoid the fallacy of aggregation, i.e.,

the erroneous belief that occurs "when we aggregate purely individual relationships and assume that collectives will behave accordingly" (p. 8) [26]. Another advantage of individual-based models is that they allow for including naturalistic mechanisms for every desired aspect of the model reality (cf. the discussion on *a priori* and *a posteriori* model on p. 23). They can mirror individual considerations of activity and location choice (see Chapter 5), model explicitly the mechanisms that lead to potentially contagious contacts between individuals [192], or include complex biological mechanisms (cf. p. 269 [178], see also Chapter 3).

However, all these advantages of individual-based models come at the cost of a reduced manageability exhibited by many individual-based models. Rather often researchers face the situation where they feel confident about the relevant mechanisms to be included in an epidemic model, but the empirical data to parameterize the model are lacking (cf. [178], see also Chapter 5). Furthermore, complicated and detailed models are always difficult to test [170]. As individual-based models of epidemics often include more mechanisms and more detailed information than the other two types of models presented before, sensitivity analyses are more laborious than, e.g., in case of a compartmental model. But it is not only the level of details, which makes individual-based models often hard to test. Due to their often stochastic nature and the fact that they cannot be treated analytically, sensitivity analysis builds upon a multitude of simulation runs for which high numbers of parameter combinations have to be simulated repeatedly to achieve stable insights into the model's sensitivity. This is costly in terms of computing time and storage space – compared to the easily done sensitivity analyses for compartmental models.

Finally, there is the problem of adequate communication of the model structure and the underlying assumptions. Models have to be communicated to the scientific community to make them available for criticism. Usually this is done by means of papers in scientific journals. Compartmental models are usually easily presented by a figure and or a small set of differential equations (however, with some detailed compartmental models, one also reaches the limits to do so; cf., e.g., to [96]). For individual-based models, the computer code can be quite long [170]. This makes it difficult to effectively communicate the model even when a journal offers the option of online supplementary material.

1.5 Structure of the research and contributions to science

This doctoral thesis addresses the role of contact characteristics in shaping the spread of infectious diseases. Particularly, we want to elaborate how the *configuration of contacts*, i.e., the arrangement of links between individuals in the social and physical space, and the *quality of contacts*, i.e., their duration, frequency, and intensity, affect the actual *course of an epidemic* and corresponding *mitigation options* in the context of *host and pathogen biology*.

The main part of this thesis consists of three peer-reviewed papers and one manuscript to be submitted to a peer-review journal.

The first two papers are contributions which aim at uncovering rather *general* mechanisms and laws in the field of disease spread. They are neither pathogen nor case-specific. The aim of these two papers is to elaborate (i) under what conditions the random mixing assumption in epidemiological models might fail, and (ii) how the differing quality of contacts could be included in such models and in what respect this differs from the common assumption of equally infectious contacts.

The other two contributions focus, in contrast, on two different case examples and answer *specific* questions. The third contribution investigates the contact structure of the poultry sector in Switzerland, the fourth aims at reconstructing a seasonal influenza epidemic that occurred in Switzerland.

Contribution 1:
Models of epidemics: When contact repetition and clustering should be included

Focus and design of the study

The aim of the first paper is to better understand when (i) repeated contact with the same partners and (ii) the clustering of contact partners (i.e., a static, structured network, cf. Subsection 1.4.2) lead to significantly different model outcomes when compared to random mixing. We tested the influence of these structural parameters under the condition of varying context parameters, i.e., we systematically varied the levels of (i) the per-contact transmission probability, (ii) the infectious period, (iii) the number of different contact partners per day, (iv) contact clustering, and (v) the ratio of repeating and transient contacts. The main outcome variable we used for our comparisons is the total outbreak size (i.e., the sum of all new cases over the entire simulation time); however, in the supplementary material, peak size and the time to peak are also provided. The focus is on droplet and contact transmitted diseases of the SIR-type. Never-

theless, some of the inferences made can also cautiously be generalized to SIS- and SEIR-settings and to airborne diseases.

Contributions to the field

The effects of clustering, of contact repetition and of the number of contact partners on the spread of infectious disease all have been investigated previously, but in isolation. The paper offers a systematic view on the effect of one parameter on disease spread in the context of the other parameters. This allows grasping not only the effects of the parameters in isolation, but also of their interaction. In particular, we show that the relative importance of the *social parameters* clustering and repetition depends on the specific *pathogen biology* (cf. Section 1.1).

Furthermore, for various concrete infectious diseases we identify corresponding parameter combinations and make statements about whether the simplifying random mixing assumption might be applicable for them. This allows field researchers to estimate to what extent the contact structure has to be surveyed and included in modeling work in order to understand and adequately reproduce specific disease spread patterns.

Contribution 2:
A mechanistic model of infection: Why duration and intensity of contacts should be included in models of disease spread

Focus and design of the study

While the first paper focuses on the *configuration of contacts*, the second paper highlights how the variability in the *quality of contacts* affects disease spread. To do so, we introduce a mechanistic model of infection for diseases transmitted by droplets or close contact (cf. Section 1.3). This model incorporates the *duration* and the *intensity* of contacts as model parameters. We calculated individual-based transmission probabilities for an empirical data set of conversational and physical contacts. These results are compared with calculations under the common assumption of constant per-contact transmission probabilities (cf. Subsections 1.4.1 and 1.4.2). This allows us to estimate to what extent the simplifying assumption of equal transmission probabilities biases model outcomes.

Contributions to the field

For the first time, we propose a mechanistic approach to quantify the transmission risks of droplet and contact transmitted diseases. Furthermore, we are able to show with empirical data that transmission risks of contacts are systematically linked with the number of contact partners of an individual. Hence, concentrating on the configuration of contacts and ignoring their quality might result in misleading inferences. Particularly, we find that super-nodes, i.e.,

highly connected individuals, are not necessarily super-spreaders as suggested by previous research (cf. Subsection 1.4.2 and [185, 263]).

Contribution 3:
Contacts between poultry farms, their spatial dimension and their relevance for avian influenza preparedness

Focus and design of the study

For this paper, we measured contacts between poultry farms in Switzerland since contacts between farms are deemed to be relevant for avian influenza spread in poultry (particularly in light of Contribution 1). Accordingly, we shift the focus from contacts between individual creatures to groups of creatures defined by the organizational structure of this sector. Furthermore, we compare the difference between an individual-based and a farm-based perspective regarding spatial density: We assume bird density to be an indicator for infection pressure in case of an outbreak. Farm density, on the other side, gives an indication of how likely accidental, contagious contacts within the immediate vicinity of a farm might be.

We particularly focus on the structural importance and interconnectedness of small, non-commercial poultry farms, as their role in the poultry farm network was controversial in previous publications. We elaborate answers to the questions of whether non-commercial farms truly act rather locally and whether they are really completely separated from the commercial production, as suggested in other work.

Contributions to the field

First, we are able to show that there are areas with a low bird density, which nevertheless have a high farm density. This has implications for disease surveillance and control as every single farm has to be instructed and controlled. Further, we find that – against the prevailing assumption – small, non-commercial farms are highly involved in long-range poultry movement and that the assumed functional division between commercial and non-commercial poultry farming does not entirely hold true for Switzerland. Consequently, we conclude that non-commercial farming should be included in surveillance systems and in models of avian influenza spread. Furthermore, we advocate one of the various competing, published modeling strategies for avian influenza spread based on the empirical findings of our study.

Contribution 4:
Reconstructing the 2003/2004 H3N2 influenza epidemic in Switzerland with a spatially explicit, individual-based model

Focus and design of the study

With the fourth contribution, we aim at reconstructing the 2003/2004 H3N2 influenza epidemic in Switzerland using an individual-based modeling approach. The core of this contribution is a validation strategy for models of epidemics, which helps to gain confidence that a certain modeling approach is reasonable and meaningful. As described in Subsection 1.2.3, we see validation not as a purely algorithmic task. Instead, validation is a matter of argumentation. In this sense, we propose a multi-criteria-based validation strategy and argue that if model and measured data are in good agreement for multiple criteria and if the mechanisms behind the model are based on plausible theories on how influenza spreads in reality, we can be relatively sure that the model is valid (cf. also to p. 25). Based on a detailed and spatially explicit *model of host-host contacts* and detailed information on *host and pathogen biology* (cf. Subsections 1.1.2 – 1.1.3), we aimed at reconstructing (i) the shape of the epidemic curve, overall infection rate, and reproduction number; (ii) age-dependent infection rates and time of infection; (iii) spatial patterns.

Contributions to the field

We present a simulation model that is able to reproduce main characteristics of the 2003/2004 H3N2 epidemic in Switzerland and seasonal influenza in general. Our chosen combination of biological and social mechanisms allows us to reproduce typical age dependencies and spatial patterns of seasonal influenza in Switzerland shown with the example of the 2003/2004 H3N2 epidemic. This gives us hints that the presented model would also produce realistic results if applied to future pandemic scenarios.

Nonetheless, we disclose that for an accurate reproduction of epidemic outbreaks, additional empirical data is needed. As a complement to the sentinel data based on reports of general practitioners, representative pre- and post-season serological data would be highly beneficial to get a better grip on the actual infection rates.

2 Models of epidemics: When contact repetition and clustering should be included

Smieszek T, Fiebig L, Scholz RW. Theor Biol Med Model 2009, 6:11

2.1 Abstract

Background

The spread of infectious disease is determined by biological factors, e.g. the duration of the infectious period, and social factors, e.g. the arrangement of potentially contagious contacts. Repetitiveness and clustering of contacts are known to be relevant factors influencing the transmission of droplet or contact transmitted diseases. However, we do not yet completely know under what conditions repetitiveness and clustering should be included for realistically modelling disease spread.

Methods

We compare two different types of individual-based models: One assumes random mixing without repetition of contacts, whereas the other assumes that the same contacts repeat day-by-day. The latter exists in two variants, with and without clustering. We systematically test and compare how the total size of an outbreak differs between these model types depending on the key parameters transmission probability, number of contacts per day, duration of the infectious period, different levels of clustering and varying proportions of repetitive contacts.

Results

The simulation runs under different parameter constellations provide the following results: The difference between both model types is highest for low numbers of contacts per day and low transmission probabilities. The number of contacts and the transmission probability have a higher influence on this difference than the duration of the infectious period. Even when only minor parts of the daily contacts are repetitive and clustered can there be relevant differences compared to a purely random mixing model.

Conclusions

We show that random mixing models provide acceptable estimates of the total outbreak size if the number of contacts per day is high or if the per-contact transmission probability is high, as seen in typical childhood diseases such as measles. In the case of very short infectious periods, for instance, as in Norovirus, models assuming repeating contacts will also behave similarly as random mixing models. If the number of daily contacts or the transmission probability is low, as assumed for MRSA or Ebola, particular consideration should be given to the actual structure of potentially contagious contacts when designing the model.

2.2 Background

The spread of infectious disease is determined by an interplay of biological and social factors [191]. Biological factors are, among others, the virulence of an infectious agent, pre-existing immunity and the pathways of transmission. A major social factor influencing disease spread is the arrangement of potentially contagious contacts between hosts. For instance, the distribution of contacts among the members of a population (degree distribution) strongly impacts population spread patterns: Highly connected individuals become infected very early in the course of an epidemic, while those that are nearly isolated become infected very late, if at all [8, 160]. For a high dispersion of the degree distribution, the transmission probability above which diseases spread is lower than for a low dispersion [8, 87, 160]. If the degree distribution follows a power law, the transmission probability necessary to sustain a disease even tends to zero [57-59].

Another important structural property influencing the spread of diseases is the clustering of contacts. Clustering deals with how many of an individual's contacts also have contact among each other. High clustering of contacts means more local spread (within cliques) and thus a rapid local depletion of susceptible individuals. In extreme cases, infections get trapped within highly cohesive clusters. Random mixing is known to overestimate the size of an outbreak [373], whereas the local depletion caused by clustering remarkably lowers the rates of disease spread [61, 62]: Clustering results in polynomial instead of exponential growth, which can be expected for unclustered contact structures [325].

For most of the diseases transmitted by droplet particles or through close physical contact, the number of contacts that can be realistically made within the infectious period has a clear upper limit. The mean value of potentially contagious contacts can be interpreted in a meaningful way, since the distribution of daily contacts is unimodal with a clear "typical" number of contacts [31, 95, 221, 237]. Potentially dominant properties of the underlying contact structure are

the clustering of such contacts and their repetitiveness, i.e. whether contacts repeat within the infectious period or not.

A recent study combining a survey and modelling showed that the repetition of contacts plays a relevant role in the spread of diseases transmitted via close physical contact. Contrarily, the impact of repetitiveness seems to be negligible in case of conversational contacts [278]. However, the generality of these findings is limited, as they are based on a small, unrepresentative sample and as the specific patterns of such contacts vary depending on the national and cultural context [237]. A more theoretical work showed that the dampening effect of contact repetition is further increased by contact clustering and is more pronounced if the number of contacts per day is low [90].

The aim of this paper is to better understand the conditions under which the inclusion of contact repetition and clustering is relevant in models of disease spread compared to a reference case assuming random mixing. This is pertinent, as many researchers still use the random mixing assumption without thoroughly discussing its adequacy for the respective case study [122, 243, 248, 277, 308]. In particular, we test and discuss the influence of transmission probability, number of contacts per day, duration of the infectious period, clustering and proportion of repetitive contacts on the total outbreak size of a disease. This helps modellers and epidemiologists make informed decisions on whether the simplifying random mixing assumption provides adequate results for a particular public health problem.

2.3 Methods

2.3.1 Stochastic SIR models

We assess the influence of repetitive contacts and clustering on the total outbreak size I_{tot} (number of new infections over simulation time) for a simple SIR structure [8, 180] under which every individual is either fully susceptible or infectious or recovered (= immune) (cf. Figure 2.1a).We construct two different types of individual-based models: one assuming random mixing (i.e. contacts are unique and not clustered), the other assuming complete contact repetitiveness (i.e. the set of contacts of a specific individual is identical for every simulation day) and allowing for clustering (cf. Figure 2.1b and additional file 2.1). Both model types can be blended in varying proportions. In our models, every infectious individual infects susceptible contacts at a daily probability β, which is equal for all infectious-susceptible pairs. Individuals remain infectious for an infectious period τ, which is exactly defined and not stochastic in its duration. Infectious individuals turn into the recovered state as soon as the infectious period passed by. We assume that infection confers full immunity for the time scale of the simulation. Hence, recovered individuals cannot be reinfected by further contacts with infectious persons.

There are no birth or death processes: Hence, the population size is constant. All possible state transitions are delineated in Figure 2.1a.

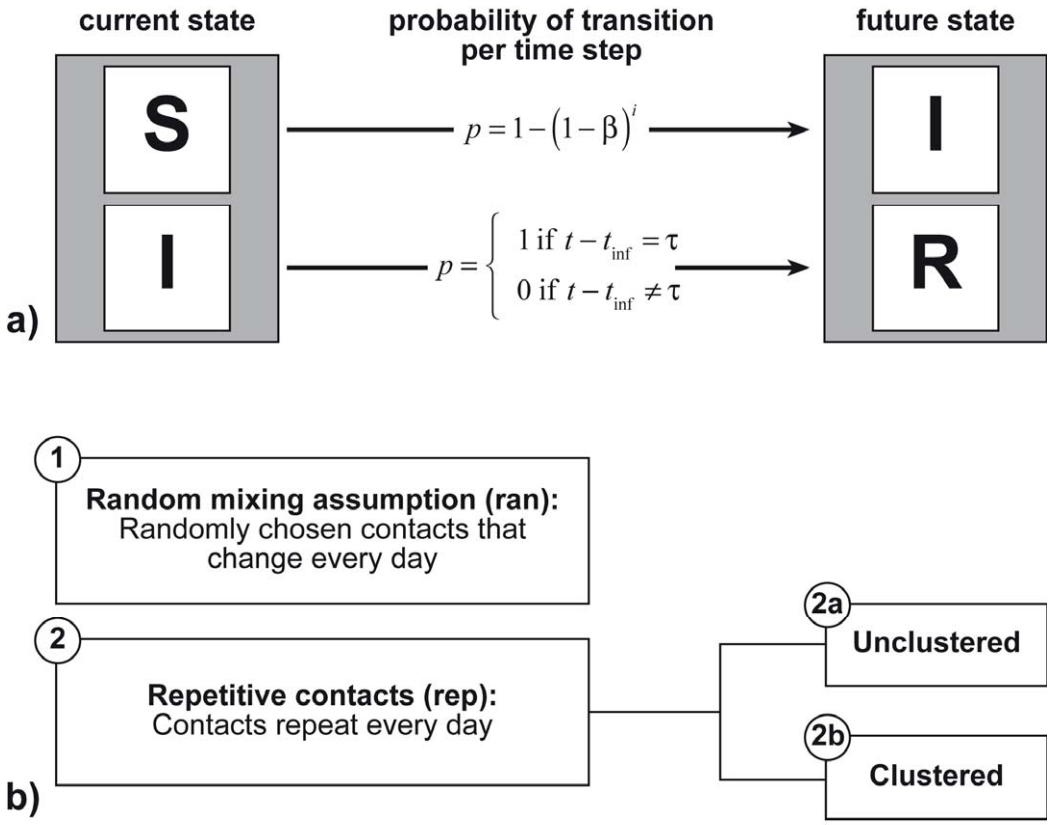


Figure 2.1: State transitions and contact structures. Subfigure a: Two transitions are allowed between three different states an individual can take: (S)usceptible to (I)nfectious and (I)nfectious to (R)ecovered. β denotes the transmission probability of one susceptible-infectious pair per time step. i stands for the number of infectious contacts that a specific susceptible individual has at the current time step. t gives the current simulation time, whereas t_{inf} gives the time step at which the individual was infected. τ is the infectious period. Subfigure b: We compare two model types: the contacts in the first type change daily while those in the second type are constant over time. The second model type assuming repetitive contacts exists in the two variants 2a and 2b.

Under the random mixing assumption (in mathematical terms denoted by index *ran*), n contacts are randomly chosen out of the whole population (including susceptible, infectious and recovered individuals) for every individual and every day. There is neither contact repetition nor clustering, as our algorithm ensures, that no contact partner is picked twice by the same individual.

In fact, clustering is neither properly defined nor is it a reasonable concept under the random mixing assumption for theoretical and practical reasons: In this paper we refer to the common definition that the clustering coefficient CC is the ratio of closed triplets to possible triplets [348], where a closed triplet is defined as three individuals with mutual contact. This definition is based on static networks. As in random mixing models contacts change daily, different clustering coefficients could be calculated for every single simulation time step. However, no epidemiologically relevant effect of such clusters could be observed, because any new infection comes into effect only in the following time step when contacts are already rearranged. As a consequence, there is no local depletion of susceptible individuals observable under this definition, even for high clustering coefficients. If clustering would be defined for an extended time interval (e.g., the infectious period), an enormous amount of closed triplets would be necessary to attain only slight clustering coefficients as the total number of contacts over such a long time is very high. For such huge cliques, there is no meaningful interpretation and no analogy in the real world.

Repetitive contacts (in mathematical terms denoted by index rep) are implemented by generating a static network with n links for every individual. The links of this network represent stable, mutual, daily contacts between individuals. As mentioned, the model type assuming repetitive contacts exists in two variants. For the variant without clustering, individuals are linked completely at random. Nonetheless, for repetitive contacts, clustering is a meaningful concept as contacts are static and as clusters correspond to observable entities in the real world: Family or work contacts, for instance, are usually clustered and tend to be highly repetitive. In this paper, predefined average clustering coefficients are achieved by alternately generating random links and triplet closures, as suggested by Eames [90], until the clustering aim is achieved in average for the whole population. When the target value of closed triplets is reached, the network is filled up with random contacts until all individuals have n contacts.

This paper compares most parameter settings for a model assuming either full random mixing or perfect repetitiveness of contacts. This comparison allows for estimating the maximal possible difference between both antipodal simplifications of reality. However, real world dynamics of networks are far more complicated; therein some contacts are repeated daily, others on certain days of the week and others only once in a while. In order to investigate the effect of different proportions of repetitive contacts, we vary the fractions of repetitive contacts.

2.3.2 Parameter space to be tested

In the following section, we describe some important factors in the spread of infectious diseases that will be systematically tested for their influence on the difference between the random mixing model and the model assuming repetitiveness (with and without clustering).

Important biological factors influencing the spread of infectious diseases are the duration of the infectious period τ and the per-contact transmission probability β.

The *infectious period* τ stands for the number of days (simulation time steps) a newly infected individual will remain infectious. The effect of repetitive contacts is tested for diseases with τ values between 2 and 14 days (see τ values given for various diseases in Table 2.1).

The *transmission probability* β is defined as the probability that an infectious-susceptible pair results in disease transmission within one single time step of the simulation. β is equal for every infectious-susceptible pair. The effect of β on the impact of repetitive contacts compared to the reference case (without repetitive contacts) is analyzed via systematic variation.

In the results section, we show all results for $\beta \cdot n \cdot \tau$ values instead of pure β values to assure comparability of the outcomes: $\beta \cdot n \cdot \tau$ equals the basic reproduction number R_0 for the random mixing model and thus models with the same $\beta \cdot n \cdot \tau$ result in a similar total outbreak size. Referring to $\beta \cdot n \cdot \tau$ values assures that model comparisons are always made for a relevant range of β. The effect of repetitive contacts is tested for $\beta \cdot n \cdot \tau$ values between 1.2 and 4.0 in increments of 0.2. The epidemic threshold of random mixing models is $\beta \cdot n \cdot \tau = 1.0$. As we are only interested in diseases that can cause an epidemic, we set the lower boundary to 1.2. The upper boundary is chosen arbitrarily.

Social factors considered in this paper are the number of contacts per day n, the proportion of repetitive contacts and the clustering coefficient.

For every single simulation run, the *number of contacts per day* n is constant and equal for all individuals. n counts every contact an individual has within one simulation step, regardless of the alter's infection status (susceptible, infectious or recovered) and regardless of whether the contact is repetitive. The effect of repetitive contacts on the simulation outcome is tested for n values between 4 and 20 with a step width of 2 (mean values for conversational contacts lie in this range [237]).

In order to investigate the effect of varying *fractions of repetitive contacts*, we simulate the total outbreak size for 0%, 25%, 50%, 75% and 100% repetitive contacts. Thereby, 25% repetitive contacts means that one fourth of all contacts on a given day repeat daily but that three fourth of the contacts on a given day are unique.

In the case of repetitive contacts, *clustering coefficients* between $CC = 0.0$ and 0.6 with a step width of 0.2 are accounted for. This span covers a wide range of existing transmission systems from highly infectious diseases with a high number of contacts per day and with clustering coefficients close to zero to highly structured settings with a considerable proportion of clustered contacts like in hospitals [204].

Disease [d]	R_0	τ [d]	Transmission pathways [162]
Chickenpox (Varicella)	7-12 [8]	10-11 [8]	Direct contact, airborne, droplet, contact with infectious material
Ebola	1.34 [67][a] 1.79 [112] 1.83 [67][b] 2.13 [112][c,a] 3.07 [112][c,b]	14 [112]	Direct contact, contact with infectious material, monkey-to-person
Influenza	1.3;1.8;3.1 [308][d] 1.39 [121] 1.58;2.52;3.41[247][e] 1.7-2.0 [110] 2-3 [223][f] 3.77 [349]	2-3 [8] 2.27 [349] 3-7 [79]	Direct contact, airborne, droplet [47]
Measles	5-18 [8] 7.17-45.41 [344][g,h] 7.7 [238] 15-17 [162] 16.32 [344][g]	6-7 [8]	Direct contact, airborne, droplet, contact with infectious secretions
MRSA [i]	1.2 [42][j]	As long as purulent lesions continue to drain [267]	Direct contact, contact with infectious material [267] •
Mumps	7-14 [8] 4.4 [93][h] 10-12 [162]	4-8 [8]	Direct contact, airborne, droplet, contact with infectious secretions
Norovirus	3.74 [337][j]	1.8 [337]	Direct contact, droplet (vomiting), contaminated food [88, 103][k]
SARS [k]	1.43 [112][l] 1.5 [112][m] 1.6 [219] 2.2-3.7 [283] >2.37 [345]	4 [345] 5 [112]	Close direct contact
Whooping cough (Pertussis)	10-18 [8] 15-17 [162]	7-10 [8]	Direct contact, airborne, droplet, contact with infectious secretion

Table 2.1: Key transmission parameters of selected diseases. Abbreviations, data sources and methods for the calculation of R_0, as far as known: [a] outbreak Uganda 2000 [259]; [b] outbreak Congo 1995 [182]; [c] regression estimates; [d] 1918 pandemic data from an institutional setting in New Zealand [308]; [e] 1918 pandemic data from Prussia; assuming serial intervals of 1, 3 and 5 days [247]; [f] 1918 pandemic data from 45 cities of the United States [223]; [g] data from six Western European countries [344]; [h] age structured homogenous mixing model; [i] MRSA, Methicillin-Resistant Staphylococcus Aureus; [j] hospital outbreaks; [k] SARS, Severe Acute Respiratory Syndrome; [l] outbreak Singapore 2003 [364]; [m] outbreak Hong Kong 2003 [364].

For all runs of the simulation model, the total population N was fixed to 20000 individuals. As initial seed 15 randomly chosen individuals are set to infectious every simulation run. For each combination of model parameters 350 runs were performed to achieve stable mean values of the outcome variables. A simulation run was terminated when no infectious individual was left.

2.3.3 Overview on performed analyses

We test the influence of the abovementioned parameters on the difference between the model typed in three distinct analyses. First, we show how strongly the total outbreak sizes $I_{tot,ran}$ and $I_{tot,rep}$ differ depending on τ, n and β. In the second analysis we vary n and β and the clustering coefficient CC for the case of repetitive contacts. Thirdly, we show how the total outbreak size changes under various n, β and CC, when repetitive and random contacts are mixed in varying proportions. Details for the three analyses are given in Table 2.2.

In addition to the total outbreak size, we present further epidemiologically relevant indicators in the additional files. Epidemic curves can be found in additional file 2.2, findings on the model differences regarding the average peak size of the outbreaks and the average time to peak are given in additional file 2.3.

	n	τ [d]	$\beta \cdot n \cdot \tau$	CC	Proportion repetitive contacts
Analysis 1					
a	4-20; 2	2-14; 1	1.6	0.0	0.0 vs. 1.0
b	4-20; 2	14	1.2-4.0; 0.2	0.0	0.0 vs. 1.0
c	4	2-14; 1	1.2-4.0; 0.2	0.0	0.0 vs. 1.0
Analysis 2	4-20; 2	14	1.2-4.0; 0.2	0.0-0.6; 0.2	0.0 vs. 1.0
Analysis 3	8-20; 4	14	1.2-3.0; 0.6	0.0-0.6; 0.2	0.0-1.0; 0.25

Table 2.2: Parameter settings of the analyses. Parameter ranges are given before the semi-colon; the increment is given after the semicolon. Single numbers stand for fixed values.

2.4 Results and Discussion

2.4.1 Analysis 1: The effect of contact repetition depending on τ, n and β

As described in the methods section, τ, n and $\beta \cdot n \cdot \tau$ have been varied systematically to investigate the difference between the mean values of the outbreak sizes $\bar{I}_{tot,rep}$ and $\bar{I}_{tot,ran}$ under different parameter constellations. Figures 2.2a-c show three contour plots in which the difference between both model types $\left(\bar{I}_{tot,ran} - \bar{I}_{tot,rep}\right)/N$ is given for various τ, n and β values. Figure 2.2a gives $\left(\bar{I}_{tot,ran} - \bar{I}_{tot,rep}\right)/N$ depending on $4 \le n \le 20$ and $2 \le \tau \le 14$ with a fixed $\beta \cdot n \cdot \tau = 1.6$. The total outbreak size depends strongly on the number of contacts per day n but only slightly on the infectious period τ. In case of an infectious period between two and four days, there is a considerable change of $\left(\bar{I}_{tot,ran} - \bar{I}_{tot,rep}\right)/N$ with $\Delta\tau$; for $4 < \tau \le 8$, slight changes

are observable; in case of infectious periods over eight days, the difference between both models depends mainly on n. Figure 2.2b gives $(\bar{I}_{tot,ran} - \bar{I}_{tot,rep})/N$ depending on $4 \le n \le 20$ and $1.2 \le \beta \cdot n \cdot \tau \le 4.0$ with a fixed $\tau = 14$. It shows that the difference between both models depends strongly on both parameters, the number of daily contacts n and the transmission probability β. Differences are large for a small n or small β but negligible for a large n when β is large at the same time. Figure 2.2c, showing $(\bar{I}_{tot,ran} - \bar{I}_{tot,rep})/N$ for $1.2 \le \beta \cdot n \cdot \tau \le 4.0$, $2 \le \tau \le 14$ and $n = 4$, is consistent with the observations made for the other two figures.

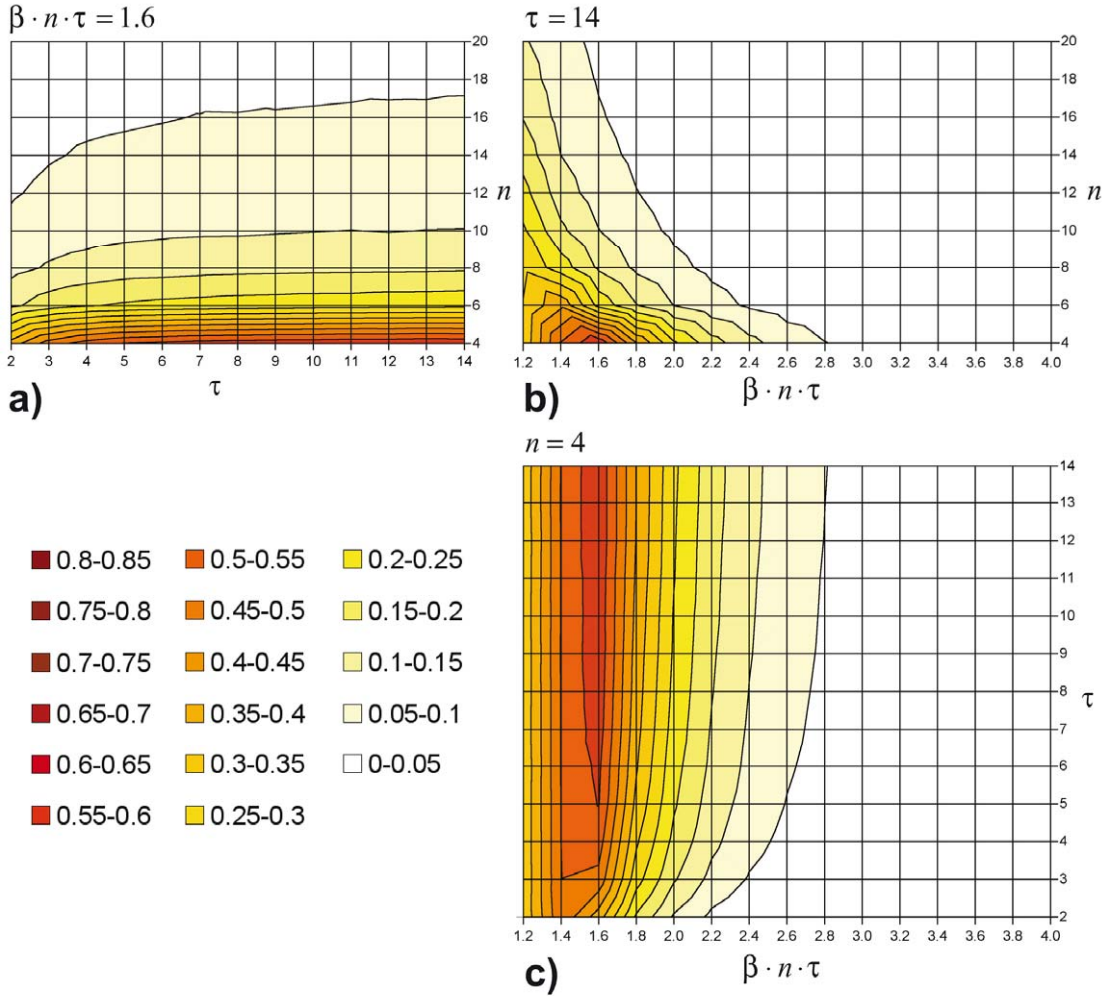


Figure 2.2: Model differences depending on τ, n and β. Subfigures a-c show the difference in the total outbreak size between a pure random mixing model and a model assuming complete repetitiveness (without clustering) relative to the population size N. Contour plots are interpolated from a grid of measurement points using Microsoft® Office Excel 2003. (a) infectious period: $2 \le \tau \le 14$, step width (sw): sw = 1; daily number of contacts: $4 \le n \le 20$, sw = 2; per-contact transmission probability: $\beta \cdot n \cdot \tau = 1.6$. (b) $1.2 \le \beta \cdot n \cdot \tau \le 4.0$, sw = .2; $4 \le n \le 20$, sw = 2; $\tau = 14$. (c) $1.2 \le \beta \cdot n \cdot \tau \le 4.0$, sw = .2; $2 \le \tau \le 14$, sw = 1; $n = 4$.

Effect of contact number: The increasing difference between $\bar{I}_{tot,rep}$ and $\bar{I}_{tot,ran}$ with decreasing n can be explained by two lines of reasoning.

First, in the case of contact repetition, there is always at least one out of the n contacts per day that is already infected (and thus not available for new infection): As contacts are stable over time, the infector of a susceptible individual is included in the subsequent contact list of that individual even when said individual has changed to the infectious state. Thus, at the least, the contact that originally transmitted the infection is not susceptible. In contrast, contacts change in every time step under the random mixing assumption: Hence, the infector is not more likely to appear in the contact set than any other individual. This difference between $\bar{I}_{tot,rep}$ and $\bar{I}_{tot,ran}$ is more pronounced for small n because one non-susceptible individual out of a small set of contacts means a relatively higher decrease in local resources than does one out of a large set of contacts.

Secondly, any new infection means that the infector will have one susceptible contact less for all subsequent time steps. This local depletion of resources is more pronounced for small n for the same reason as in the first argument. Further, stochasticity acts stronger in small local environments than in large ones [177].

Both effects can also be seen in the Equation 2.1, which gives $R_{0,rep}$ as a function of $R_{0,ran}$, n and τ (see also Figure 2.3a; details for Equation 2.1 are given in additional file 2.4):

$$R_{0,rep} \cong (n-1)\cdot\left[1-\left(1-\frac{R_{0,ran}}{n\cdot\tau}\right)^{\tau}\right] \qquad (2.1)$$

In this equation the number of susceptible individuals in the local environment is reduced by 1 compared to the random mixing case, as we assume that every contact except the one that originally transmitted the infection is susceptible. This number of susceptible individuals $(n-1)$ is multiplied by the probability that such an individual becomes infected during the infectious period τ. As $(n-1)$ is smaller than n and $\left[1-(1-\beta)^{\tau}\right]$ is smaller (or equal for $\tau=1$) than $\beta\cdot\tau$, the expected number of secondary cases caused by an infectious individual in a population with a huge number of susceptible and few infected ones is always smaller in the repetitive case.

Effect of the per-contact transmission probability: The difference between $\bar{I}_{tot,rep}$ and $\bar{I}_{tot,ran}$ decreases rapidly with increasing β. The reason is that practically every individual will be reached and infected in case of large transmission probabilities, regardless of the underlying contact structure. Differences between both models may appear in the shape of the outbreak curve (cf. to additional files 2.2 and 2.3), but in terms of I_{tot} both models are equivalent. In case of small transmission probabilities, differences in the effective number of secondary cases

generated by an infectious individual can become visible, as only a fraction of the whole population will be infected under both assumptions.

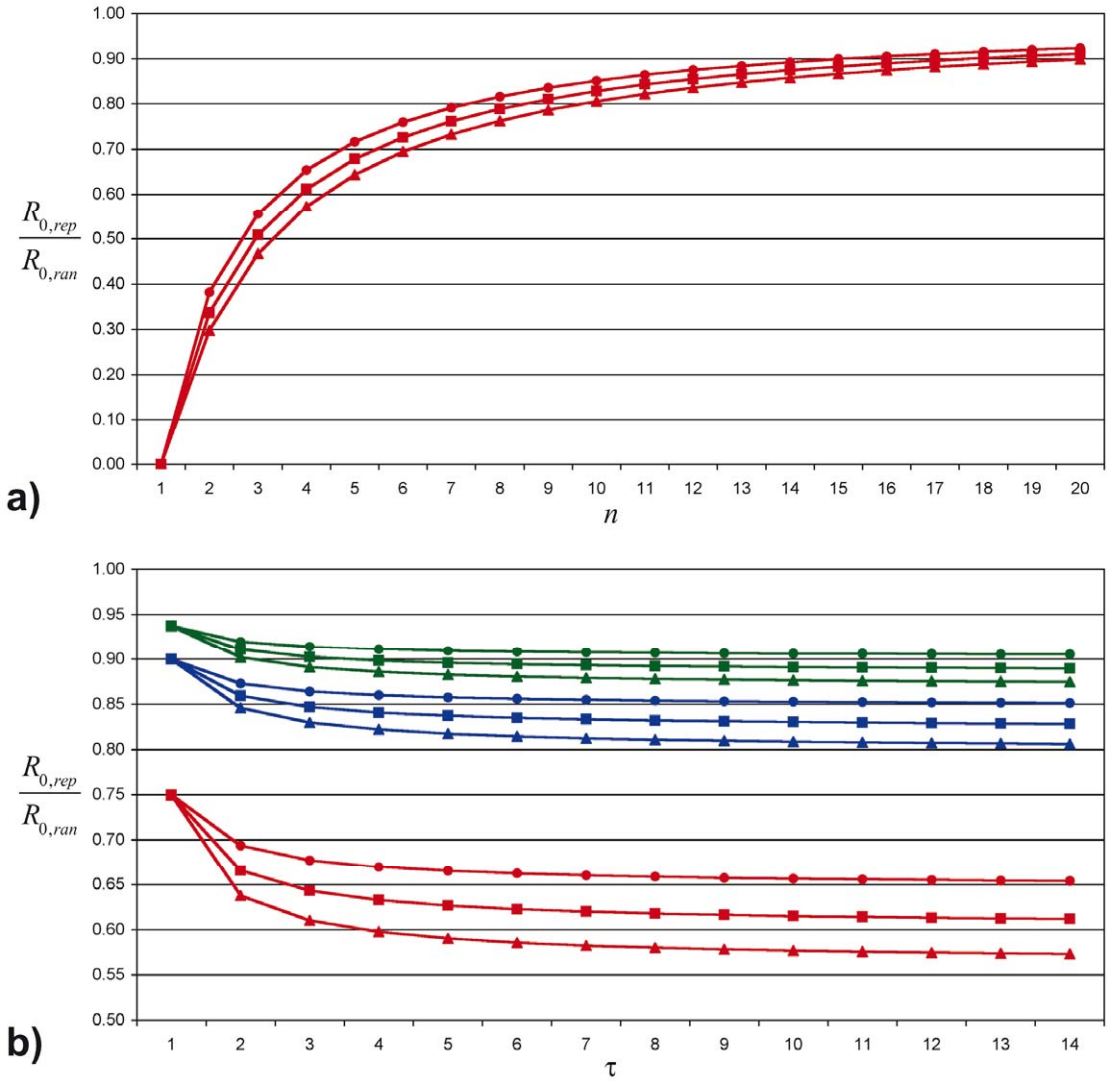


Figure 2.3 : Ratio of the basic reproduction numbers. Subfigure a shows the ratio $R_{0,rep}/R_{0,ran}$ (as defined in equation 1) for $1 \le n \le 20$ (number of daily contacts) and $\tau = 14$ (infectious period). Triangles stand for $\beta \cdot n \cdot \tau = R_{0,ran} = 2.4$, squares for $R_{0,ran} = 1.8$ and circles for $R_{0,ran} = 1.2$. Subfigure b gives $R_{0,rep}/R_{0,ran}$ depending on the infectious period τ. Red lines and symbols are for $n = 4$, and blue lines stand for $n = 10$, whereas green lines represent $n = 16$. The meaning of the symbols is identical as in subfigure a.

Effect of the infectious period: As expected, the difference between $\bar{I}_{tot,rep}$ and $\bar{I}_{tot,ran}$ increases with increasing τ. However, the change in difference is largest for $\Delta\tau$ in a range of low τ values, but is almost irrelevant for high values of τ. This observation is explained by the τ-dependence of $R_{0,rep}$ (Equation 2.1, see also Figure 2.3b): The longer the infectious period, the smaller the chances for a specific contact to remain uninfected. However, this increase in individual infection probability is partly compensated by a lower per-day transmission probability, which is needed to achieve constant $R_{0,ran}$. The interaction of these antagonistic effects results in a stabilization of $R_{0,rep}/R_{0,ran}$ for a large τ.

2.4.2 Analysis 2: The effect of contact repetition combined with clustering depending on *n* and β

The results presented previously show that $\left(\bar{I}_{tot,ran}-\bar{I}_{tot,rep}\right)\!/N$ depends mainly on n and β. In a second step, we investigate how the difference between model type 1 and 2 changes, if clustering is introduced in the latter. Figures 2.4a-d show the difference between both model types for clustering coefficients CC between 0.0 and 0.6 when τ is fixed to 14 days and when n and $\beta\cdot n\cdot\tau$ vary in the ranges mentioned above. As expected, clustering results in an increased difference between both model assumptions. This increase is most pronounced for small numbers of contacts per day. The peak of $\left(\bar{I}_{tot,ran}-\bar{I}_{tot,rep}\right)\!/N$ is constantly at $n=4$ but shows a right shift on the $\beta\cdot n\cdot\tau$ axis for increasing CC.

The further dampening of disease spread by clustering can be explained by increased locality of resources: While repetition limits the number of available susceptible individuals by keeping previously infected ones in the set of contacts, clustering reduces the number of susceptible contacts because there is a higher likelihood that contacts of an infector have already become infected by others during the infectious period, as infections spread rapidly within cliques. The reason why this effect is more pronounced for small n rather than for large n is the same as in the case of unclustered, pure contact repetition: Any reduction of susceptible individuals in the set of contacts weights relatively stronger in the case of few contacts than in the case of many. The right shift of the peak of $\left(\bar{I}_{tot,ran}-\bar{I}_{tot,rep}\right)$ can be explained by the increased transmission probability β needed to pass the epidemic threshold under increased clustering compared to the constantly low levels of β necessary under the random mixing assumption [9].

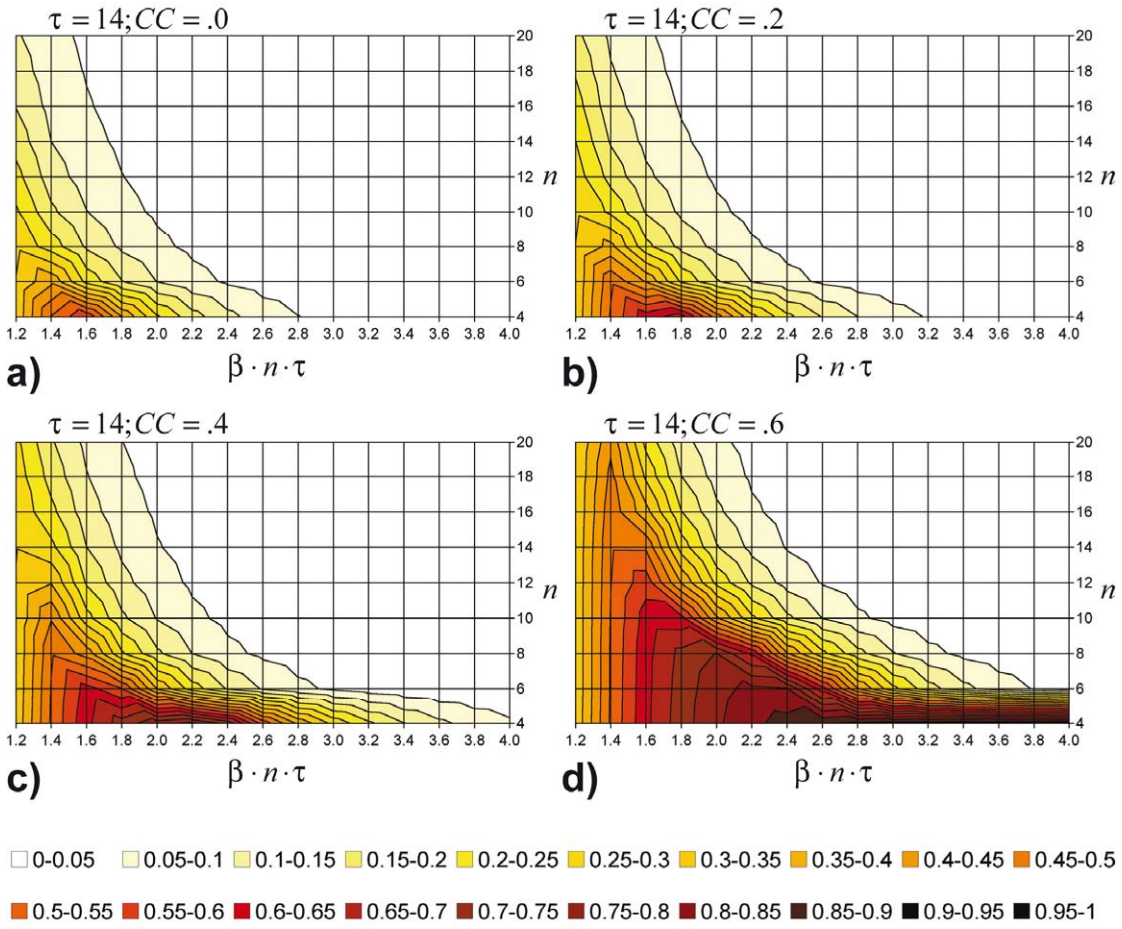


Figure 2.4: Dampening effect of clustering. Subfigures a-d show the difference in the total outbreak size between a pure random mixing model and a model assuming complete repetitiveness (with different levels of clustering) relative to the population size N for $4 \le n \le 20$, $1.2 \le \beta \cdot n \cdot \tau \le 4.0$ and $\tau = 14$. Subfigure 4a is identical with subfigure 2b. The clustering coefficient CC is increased picture-wise in steps of .2.

2.4.3 Analysis 3: Varying proportions of contact repetition, clustering and β

We simulated the difference between both model assumptions for all possible combinations of $n = 8$, 12, 16 and 20, $\beta \cdot n \cdot \tau = 1.2$, 1.8, 2.4 and 3.0, $\tau = 14$ and $CC = 0.0$, 0.2, 0.4 and 0.6. The simulation results are shown in Figures 2.5a-p. The relation between the proportion of repetitive contacts per day and the average difference between this mixed model and a model assuming purely random mixing is approximately linear in the absence of clustering (for all tested cases, linear regressions between the proportion of repetitive contacts per day and the deviation of $\bar{I}_{tot}$ from the purely random mixing model achieve $R^2 > .98$). However, the deviation from the random mixing model increases disproportionately with the fraction of repetitive contacts when clustering is introduced (cf. to Figures 2.5b-d, 2.5f-h, 2.5j-l and 2.5n-p).

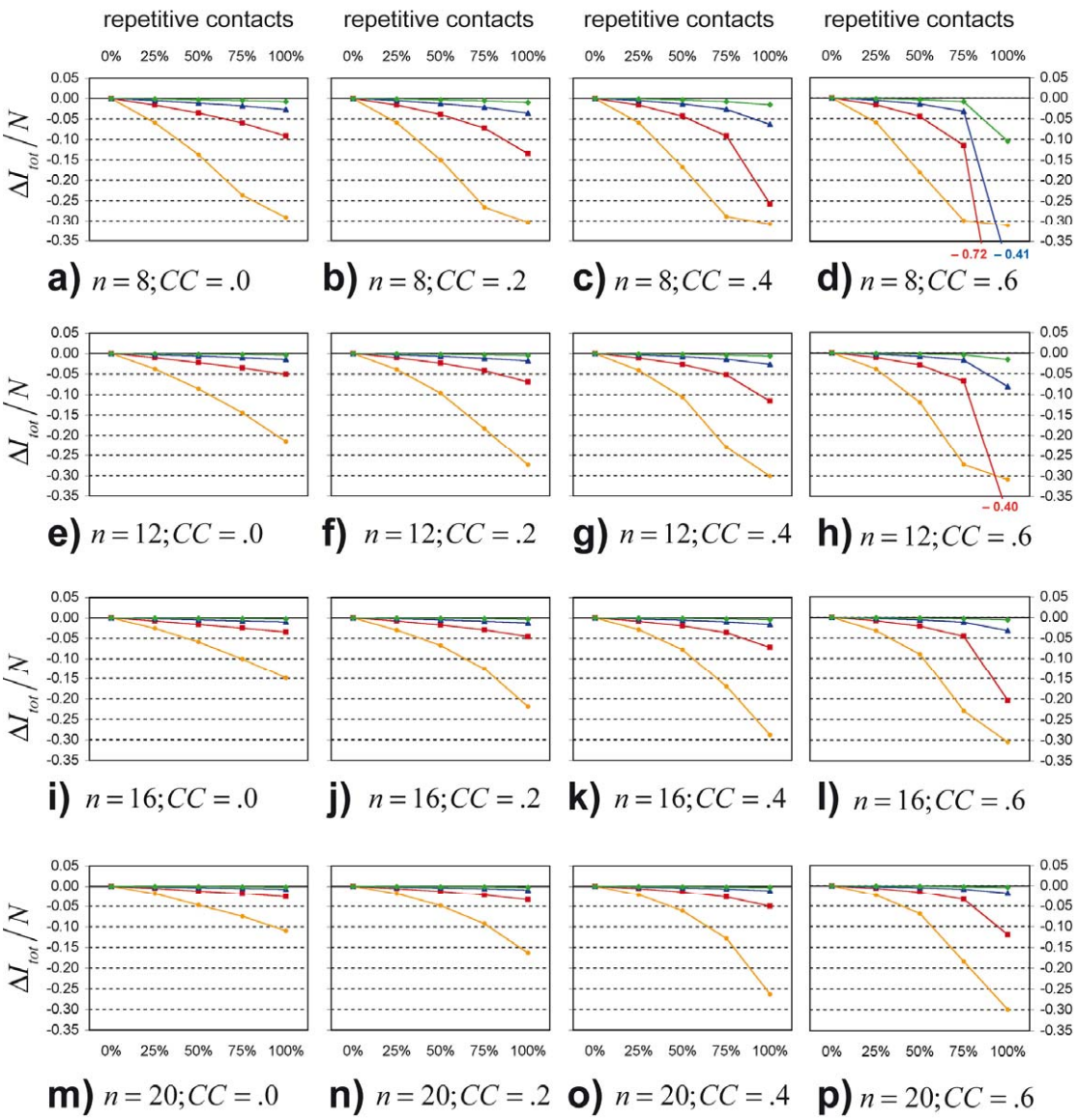


Figure 2.5: Mixed models. Subfigures a-p show the decrease of the total outbreak size relative to the size of the total population when the fraction of repetitive and clustered contacts is increased. 25% rep means that one fourth of all contacts on a given day repeat every day but that three fourths of the contacts on a given day are unique. Clustering coefficients CC are only defined and calculated for the repetitive fraction of the contacts. All simulations were calculated for an infectious period of 14 days. Orange circles stand for $\beta \cdot n \cdot \tau = 1.2$, red squares for $\beta \cdot n \cdot \tau = 1.8$, blue triangles for $\beta \cdot n \cdot \tau = 2.4$ and green rhombi for $\beta \cdot n \cdot \tau = 3.0$. The number of daily contacts n increases in steps of 4 per line of the subfigures, beginning with $n = 8$ in the first line. The first column of the subfigures shows $CC = .0$, the second column $CC = .2$, the third column $CC = .4$ and the fourth column $CC = .6$.

One mechanism driving this non-linear relation when clustering is present is the local depletion of resources. Repetitive contacts of an infector have a much higher chance of becoming infected than do non-repetitive contacts. Moreover, if these repetitive contacts are also highly clustered, it is likely that the disease will become trapped in those cohesive social subgroups. However, if

only a few non-repetitive, non-clustered contacts are added per day, the chances of spreading the disease between otherwise unrelated regions of the social network greatly increase.

2.4.4 Limitations

This paper systematically investigates a variety of epidemiologically relevant parameters needed to describe real-world transmission systems of diseases spread by droplet particles or direct physical contact. However, real-world social and biological processes involved in the transmission of infectious diseases are far more complex than captured by the archetypical model structures presented. Conceptual decisions and simplifications which could have potentially influenced the results are critically discussed in the following:

Model structure: We designed our two model types as SIR models, assuming that every individual is either susceptible, infectious or immune with respect to a certain disease. Transitions are only allowed from susceptible to infectious or from infectious to immune. The SIR structure is a fairly good representation for many diseases which lead to full immunity after recovery (e.g., measles). However, many diseases require other representations, as relevant intermediate states need to be covered, e.g., as with a long latency period in SEIR (Susceptible-Exposed-Infectious-Recovered) models. Another common deviation from the SIR structure arises, when recovery confers only partial or no immunity. In such cases, SIS (Susceptible-Infectious-Susceptible) representations are often chosen. In SIR or SEIR models, a total outbreak size can be defined (because the disease fades out at the end of an epidemic), whereas SIS models typically achieve an equilibrium $I(t)$ in the long run, but the disease does not die out. Despite all the differences in model behaviour, we expect the rough picture to be the same for SIR, SEIR and SIS models, as the mechanisms behind the observed differences for SIR models that we discussed also apply to SIS and SEIR models. Thus, the general conclusions derived in this paper should also hold true for these model types.

Degree distribution: The number of daily contacts n is fixed and equal for the entire population in both modelling approaches presented. This is a reasonable simplification for the purpose of this paper, as it keeps the investigated number of interactions manageable. However, in real world systems, the number of daily contacts appears to follow a negative binomial distribution [221, 237] with some people having a relatively high number of contacts and others being almost isolated. It is known that the variance of the degree distribution impacts the spread of infectious disease, for instance, by decreasing the transmission probability needed to cause an epidemic [22]. Particularly relevant for the difference between random mixing models and models accounting for contact repetition and clustering are the correlations between the number of contacts per day and contact repetition and clustering, respectively. It is plausible to assume

that individuals with many contacts tend to also have many unrepeated contacts, whereas individuals with few contacts tend to have disproportionately high levels of repetitive contacts. If the proportion of repetitive contacts and clustering is correlated with the number of contacts, individuals with few contacts are likely to be dead-end streets for infectious diseases. In contrast, highly connected individuals could be structurally more important than expected, as they bridge distinct cliques.

Occasional contact repetition: In our simulations, contacts repeat either daily or never. Intermediate states between both extremes of complete random mixing and complete contact repetition have been investigated by combining both models in defined proportions. However, in reality, specific persons can be met at any frequency between never and daily. It is plausible to assume that intermediate frequencies reduce the effect of repetitiveness depending on the duration of the infectious period τ: For short infectious periods, those with low contact frequencies might appear as unrepeated contacts whereas they unfold their full dampening potential for long infectious periods.

Contact intensity and duration: In our models all contacts between an infector and a susceptible individual are equally likely to result in the transmission of the infectious disease. This simplification is not a good representation of the real world: The transmission probability depends on the amount of infectious material ingested by a susceptible person [146, 354]. The uptake correlates with contact duration and intensity. Contact duration is long for highly repetitive contacts, while unrepeated contacts tend to have short duration (unpublished data). Accordingly, it can be expected that the interaction of clustering, contact repetitiveness and contact duration leads to a rapid infection of all closely tied clusters (primarily families, then workgroups and cliques at school and childcare institutions), leaving behind the people connected via mainly short, unclustered, occasional contacts.

Distribution of infectious period: The infectious period τ is fixed in our model, which contrasts to the design of classical mean-field models assuming exponentially distributed infectious periods [8, 180]. Keeling and Grenfell argue that R_0 is smaller for exponential period models than for fixed period models under otherwise identical conditions, because individuals with a long τ rapidly exhaust the susceptible in their local neighbourhood and, therefore, cannot compensate for the large majority of individuals with extremely short infectious periods [176, 177]. However, the often assumed exponential distribution is highly unrealistic, as observed infectious periods tend to be closely centred around a mean period and are thus less dispersed [205]. Thus, assuming a fixed infectious period is a reasonable simplification of the reality that is not likely to have a major influence on $\bar{I}_{tot}$ as only very few individuals will use up their local susceptible resources during the infectious period in most cases. Moreover, if the infection probability is

high enough to exploit almost the entire local environment (such that deviations of τ could affect the individual reproduction ratio), $\bar{I}_{tot}$ will reach the order of magnitude of the population size in either the fixed or the exponential case.

2.4.5 Implications for some exemplar diseases

Information on the per-contact transmission rate β and the number of potentially contagious contacts n is often not easily accessible or available and has to be measured (or fitted) if included in models of disease spread. However, rough estimates of both variables can be obtained when R_0 estimates are available and when the possible pathways of transmission are known, because β and n are linked to the basic reproduction number by $R_{0,ran} = \beta \cdot n \cdot \tau$ and the possible pathways reveal information on the possible number and structure of contacts at risk: At one extreme there is transmission via close physical contacts, which correlate mostly with intense social relations and are typically rare, repetitive and highly clustered. The other extreme is airborne transmission via tiny droplet nuclei that remain suspended indoors for a long time. In this case, vast numbers of persons can potentially be exposed, and such casual contacts are neither highly repetitive nor strongly clustered.

Table 2.1 provides information about the infectious period τ, R_0 estimates and the possible pathways of transmission for a variety of infectious diseases. The implications of clustering and contact repetition for models of the diseases listed in this table are discussed below.

Typical childhood diseases like mumps, measles, pertussis (whopping cough) or chickenpox have comparatively high R_0 estimates [53, 83-86], which means that one infector generates many secondary cases if a sufficient number of susceptible contact partners are available. These diseases are highly communicable – in fact, measles is one of the most highly communicable diseases in the world [236] – and thus, very short and non-intense contacts have the potential to confer infection. Accordingly, both the number of contacts per day n and the per-contact transmission probability β are very high. We further assume that a high proportion of the contacts are casual contacts, because the threshold for a contact to be potentially contagious is very low with respect to duration and intensity. Consequently, the levels of repetitiveness and clustering are low, which means that the contact patterns for such childhood diseases are structurally similar to random mixing. Considering that high numbers of daily contacts n make both types of models that we discussed behave similarly and considering that under high transmission probabilities β almost every individual will be reached, random mixing models achieve almost the same results as more elaborate models including a certain amount of contact repetition and clustering. Also in case of Norovirus, the difference $\left(\bar{I}_{tot,ran} - \bar{I}_{tot,rep}\right)$ is probably small, as the infectious period of this infectious agent is very short [337] and as at the

same time the basic reproduction number is comparatively high [337] (because the disease is easily communicable [88, 103]).

On the other side, there are diseases with comparatively low R_0 estimates and typically low numbers of contacts that still qualify for potential transmission. Methicillin-resistant *Staphylococcus aureus* (MRSA), for instance, is an infectious agent mostly transmitted in health care and nursing institutions. It needs close physical contact for transmission [267] and R_0 estimates given in the literature are close to the epidemic threshold [42]. Accordingly, both β and n are low. At the same time, health care settings tend to be highly structured regarding who cares for whom and who shares a room with whom. Hence, high levels of contact repetitiveness and clustering can be assumed [204]. Modelling MRSA under the random mixing assumption is likely to overestimate the total number of cases for given n, β and τ. If, in contrast, a random mixing model is fitted to measured data from an outbreak, either the infectivity or the number of potentially infectious contacts will be underestimated to meet the measured outbreak size. A similar argumentation applies to Ebola, which is transmitted via direct contact with infected blood, secretions, organs or semen (thus, n is rather low) and seems to be only moderately infectious [67, 112, 182, 259]. As a consequence, random mixing models of Ebola [199] are of limited validity.

Finally, there are some diseases not easily attributable to one or the other class. Severe Acute Respiratory Syndrome (SARS) and Influenza, for instance, have a range of R_0 estimates between 1.43 and 3.7 [112, 219, 283, 345, 364] and between 1.3 and 3.77 [79, 110, 121, 223, 247, 308, 349], respectively. No definite consensus has been reached on whether Influenza is transmitted predominantly by large droplets and close contact or by very small droplets that disseminate quickly and stay suspended in indoor air for a long time [47]. In the latter case, a large amount of people would be at risk of infection, so random mixing would be a reasonable approximation of the real contact patterns. In the case of transmission by close contact and large droplets (that fall out quickly), the mean number of potentially contagious contacts per day lies between 8 and 18, depending on the national and cultural context [237]. Considering that not all contacts are equally likely to transmit influenza, but that long and intense contacts (such as household contacts [111]) are more prone to do so and that such contacts also tend to be more repetitive and clustered, it is likely that random mixing models also overestimate the outbreak size for given n, β and τ. However, problems will definitely arise when the impact of social distancing measures (decrease of n) or of antiviral treatment (decrease of β) are estimated under the random mixing assumption: Both interventions will be much more effective in a more elaborate model than in a random mixing model when n, β and τ are the same for both model types. This argumentation is consistent with recent findings on the impact of other network

properties on influenza spread: Heterogeneity in degree distribution does not influence the outbreak size in case of highly contagious influenza strains, but does so for moderately contagious strains; however, it does influence the total outbreak size when interventions are simulated – even in case of highly contagious strains [87].

2.5 Conclusions

Real-world contact patterns are complex. They typically show all kinds of intermediate states ranging from contacts repeating on a daily basis to and never again. There are various clearly defined, cohesive groups with typically high intra-group clustering coefficients (e.g. households, workgroups, peer groups at school) and, at the same time, random contacts, e.g., in a leisure setting. Moreover, contacts differ in intensity and duration, which further complicates the dynamics of disease spread in such settings. This paper simplifies these complex patterns to a manageable model and parameter space that can be investigated systematically. Our research applies to diseases transmitted via conversational or direct contact, for which a typical number of contacts per day can be defined. For such diseases, our findings can help modellers judge whether a specific transmission system consisting of a specific infectious agent and a specific human system at risk can be represented by a simple random mixing model or if more elaborate models are necessary.

Random mixing models result in acceptable estimates of the total outbreak size $\bar{I}_{tot}$ even if the real world contacts are highly repetitive and clustered

- if the number of potentially infectious contacts per day is high and
- if the transmission probability for a single infectious-susceptible pair is high and
- particularly, if the infectious period is just one to three days.

If the number of contacts per day or the transmission probability is low, particular consideration should be given to the actual structure of potentially contagious contacts in designing the model.

2.6 Competing interests

The authors declare that they have no competing interests.

2.7 Authors' contributions

TS carried out the majority of the model design, implemented the model, computed the analyses and prepared the manuscript as the lead writer. LF participated in the model design, contributed to the epidemiological interpretation of the model results, reviewed the literature

on model parameters for specific diseases and helped to draft the manuscript. RWS participated in the model design and helped to draft the manuscript. All authors read and approved the final manuscript.

2.8 Additional material

Additional file 2.1

Algorithms. Provides a description of the key algorithms used for this paper following the ISO 5807-1985 standard.

http://www.biomedcentral.com/content/supplementary/1742-4682-6-11-S1.pdf

Additional file 2.2

Epidemic curves. This document provides exemplary epidemic curves for selected parameter settings.

http://www.biomedcentral.com/content/supplementary/1742-4682-6-11-S2.pdf

Additional file 2.3

Contour plots & tables. Additional contour plots for the differences in peak size and the differences in the simulation time till the peak is reached are given. In addition, data tables of means and standard deviations are provided for many analyses presented here.

http://www.biomedcentral.com/content/supplementary/1742-4682-6-11-S3.pdf

Additional file 2.4

Reproduction numbers. This document shows how Equation 2.1 can be derived.

http://www.biomedcentral.com/content/supplementary/1742-4682-6-11-S4.pdf

3 A mechanistic model of infection: Why duration and intensity of contacts should be included in models of disease spread

Smieszek T. Theor Biol Med Model 2009, 6:25

3.1 Abstract

Background

Mathematical models and simulations of disease spread often assume a constant per-contact transmission probability. This assumption ignores the heterogeneity in transmission probabilities, e.g. due to the varying intensity and duration of potentially contagious contacts. Ignoring such heterogeneities might lead to erroneous conclusions from simulation results. In this paper, we show how a mechanistic model of disease transmission differs from this commonly used assumption of a constant per-contact transmission probability.

Methods

We present an exposure-based, mechanistic model of disease transmission that reflects heterogeneities in contact duration and intensity. Based on empirical contact data, we calculate the expected number of secondary cases induced by an infector (i) for the mechanistic model and (ii) under the classical assumption of a constant per-contact transmission probability. The results of both approaches are compared for different basic reproduction numbers R_0.

Results

The outcomes of the mechanistic model differ significantly from those of the assumption of a constant per-contact transmission probability. In particular, cases with many different contacts have much lower expected numbers of secondary cases when using the mechanistic model instead of the common assumption. This is due to the fact that the proportion of long, intensive contacts decreases in the contact dataset with an increasing total number of contacts.

Conclusions

The importance of highly connected individuals, so-called super-spreaders, for disease spread seems to be overestimated when a constant per-contact transmission probability is assumed. This holds particularly for diseases with low basic reproduction numbers. Simulations of disease spread should weight contacts by duration and intensity.

3.2 Background

Research has shown that the arrangement of potentially contagious contacts among the individuals of a society is a determining factor of disease spread: Both the repetition and the clustering of contacts diminish the size of an outbreak compared to a random mixing model [90, 174, 314]. Further, the epidemic threshold is low if the degree distribution shows a high dispersion [8, 160]. In contrast to the vast body of literature that exists on the importance of network structure, only little emphasis has been put on the *quality* of such potentially contagious contacts, i.e. how long they last and how intensive they are. In fact, mathematical models and computer simulations of disease propagation often assume a constant per-contact transmission probability [cf. 8, 178, 180, 242, e.g.: 308, 377]. This approach ignores that, for instance, a short random encounter of two persons on a public bus is less likely to transmit a certain communicable disease than a rendezvous that lasted several hours.

Treating all contacts equally may lead to an overestimation of the individual transmission probability in cases of short, non-intense contacts and an underestimation in cases of intense, prolonged contacts. Allowing for heterogeneous transmission probabilities may then affect the model behaviour in various ways (e.g., altering the shape of the epidemic curves or changing the predictions of the effectiveness of intervention measures). In particular, the valuation of certain "risk groups," such as so-called super-spreaders defined as highly connected individuals [178], may change.

Several authors have already introduced heterogeneous transmission probabilities in their models . To do so, field data was typically analysed statistically to extract differences due to age, the susceptible individuals' immune responses, the levels of infectiousness of the infectors, and different contact situations [116-119]. For instance, in their model for Ebola epidemics, Legrand et al. differentiated the infection potential of hospital, funeral, and community settings [199], while Ferguson et al. distinguished household and non-household contacts in their model for an influenza pandemic [110]. The disadvantage of such *a posteriori* statistical models is that they become invalid when their underlying determinants (e.g., how individuals interact with other individuals) change.

Only few epidemic simulations model infection processes mechanistically (i.e., based on an *a priori* model instead of purely statistical analysis) to determine the transmission probability of differing contact situations: Alexandersen et al. [6] and Sørensen et al. [317], for example, show that basing large scale simulation models on quantities, such as intensity and duration of an exposure to infectious material, is possible and expedient. Existing mechanistic transmission models applied in simulations of disease propagation focus almost exclusively on aerosol transmission, but do not cover transmission by droplets and physical contact ("close contact"). Hence, simple mechanistic models of close contact contagion that can be used in simulations of disease spread are needed.

This paper is intended to highlight why mechanistic models of disease transmission are needed, to provide an example of how they can be built, and to show how they differ from the often-used transmission model that assumes a constant per-contact transmission probability. The proposed mechanistic approach for including the heterogeneity of transmission probabilities into disease spread simulations concentrates exclusively on diseases that are transmitted via close contact between an infector and a susceptible individual. We build on the fundamental knowledge that the risk of disease transmission is not only a function of the infectivity of the infectious agent and the quality of the immune response but also of the host's exposure to a specific infectious agent [146, 354]. Particularly, we present evidence suggesting that the common assumption that highly connected individuals act as super-spreaders [56, 57, 112] might be misleading.

3.3 Methods and Material

In this section, we first describe a formula that models transmission probabilities based on mechanistic considerations. Then, we introduce and describe an empirical data set of self-reported contacts qualified to transmit infectious disease. This data set was used to test the impact of the proposed transmission model. Finally, we introduce the scheme that describes how the outcome of both transmission models, i.e., the proposed mechanistic model assuming exposure dependency and the classical model assuming equally weighted contacts, were compared. Subsequently, we will refer to the first transmission model as the "mechanistic model" and the second model as the "classical model."

3.3.1 A mechanistic transmission model

The probability of contracting a disease is closely linked to exposure to infectious organisms. A susceptible individual can only become infected if she/he is exposed to infectious organisms. Thereby, the transmission probability increases with an increase in the number of infectious organisms to which a susceptible individual is exposed. Subsequently, we refer to exposure as

the cumulative, average amount of infectious medium ingested by a susceptible individual within a time period of interest due to close contact with an infectious person.

We base our proposed transmission model on the exponential relationship between the ingested dose and the infection risk as derived in Haas et al. [146] and used in several other publications [110]. Details describing how the following assumptions 1, 3, and 5 translate into an exponential dose-response model can be found in Haas et al [146]. As an extension to this general formulation of an exponential relation between exposure and the risk of infection, we extrapolate the actual exposure from information about the duration and intensity of a contact between an infector and a susceptible individual. The proposed mechanistic model is based on the following underlying assumptions:

1) In principle, one infectious organism is sufficient to cause infectious disease. This hypothesis has been repeatedly supported by various studies against the alternative hypothesis assuming a threshold dose of infectious organisms must be passed to cause infection [220, 286, 354].

2) Every ingested infectious organism has a certain probability to survive until it reaches its target tissue and can initiate infection [12, 146].

3) We assume that this survival probability is a constant, i.e., factors like the susceptible hosts' immune responses are assumed to be equally effective for all individuals. This assumption is a simplification of reality since susceptibility is known to differ between individual hosts [13]. However, for the purpose of this paper, such a simplification that keeps the model and the interpretation of its results manageable is justified.

4) The average dose of infectious material that is ingested by an individual is a linear function of the duration and intensity of the contact with an infectious individual. Research has shown that these measures are good predictors for individual attack rates of SARS [132]. In theory, we recognize that contact can be any kind of interaction between two individuals that is sufficient to exchange body fluids that can carry infectious particles. However, for reasons of manageability and measurability, we concentrate on conversational and physical contacts.

5) The actual amount of infectious organisms ingested by an individual follows a Poisson probability distribution with the average dose (defined in assumption 4) as parameter [146, 354]. Thereby, we model the total (i.e., cumulative) average dose ingested during an entire simulation time step. This can lead to biased results in extreme cases [268], but given the fact that this assumption has proven to work well in the past and considering other uncertainties, utilizing this simplification is justified.

Based on this, the probability P_{n,t_x} that individual n becomes infected during simulation time step t_x can be derived as

$$P_{n,t_x} = 1 - \exp\left(-\Theta \sum_{m=1}^{I} q_{m,t_x} \cdot i_{nm,t_x} \cdot t_{nm,t_x} \right) \tag{3.1}$$

where I is the total number of infectors; q_{m,t_x} $\left[s^{-1}\right]$ is the shedding rate (~ microbial load) of infector m at simulation time step t_x; i_{nm,t_x} [1] is the contact intensity between the infector and the susceptible individual, which corresponds with the proportion of infectious material spread by infector m that is actually ingested by n; and t_{nm,t_x} [s] is the time individuals n and m actually interact during time step t_x. Finally, Θ is a calibration parameter that accounts for all relevant factors that are not explicitly represented, such as survival probability of the infectious agent. Simulation models can be fitted to measured epidemiological data, such as epidemic curves, or to targeted reproduction rates by means of Θ. We used Θ to achieve predefined reproduction rates for the contact structure introduced in the following section.

3.3.2 Empirical contact structure

In the subsequently described test setting, empirical contact data is needed to compare the mechanistic transmission model with the classical one. We rely on contact data reported in a contact diary study that was conducted in Switzerland. A convenience sample of 54 participants was asked to report their potentially contagious contacts (as defined below) for 14 different days. Although a convenience sample is not representative for the whole population, the sample used here represents a very diverse cross-section of the population as can be seen in Table 3.1.

The design of the diary is similar to that used by Mikolajczyk et al. [221]. A potentially contagious contact is defined as (1) a mutual conversation of more than 10 words within a short distance (<2m), (2) physical contact in general, or (3) contact involving kisses. The participants were asked to categorize their contacts according to these three categories and to estimate how long they interacted with each reported contact person during an entire day based on six provided categories. However, for the analysis, we need concrete values instead of categories to calculate transmission probabilities as defined in Equation 3.1. Therefore, we assume a concrete duration, the arithmetic mean of the upper and lower bounds, for each category as given in Table 3.2.

One diary had to be revoked due to deficient data quality; three of the remaining 53 participants provided only information for 5, 7, or 8 days, resulting in a total of 720 different person days with 7145 reported contact partners. In 36 of 7145 records, the information about contact duration was missing. These missing values were imputed based on probability distributions observed for the complete records. The processed data is provided in Additional File 1.

Gender of participants	
Female	27 (50.9%)
Male	26 (49.1%)
Age distribution of participants	
Mean and standard deviation	37.48 (SD = 16.71)
Min	20
25% percentile	24
50% percentile	29
75% percentile	52.25
Max	76
Occupational status of participants [a]	
Student	21 (39.6%)
Employed	35 (66.0%)
Neither student nor employed	11 (20.8%)
Distribution of contact partners per day	
Mean and standard deviation	9.92 (SD = 7.64)
Min	0
25% percentile	4
50% percentile	8
75% percentile	13
Max	51

Table 3.1: Basic information about the sample and the contact structure. [a] This does not sum up to 100% because multiple answers were possible.

Category in diary	Time value used for calculations
Less than 5 min	2.5 min
5-15 min	10.0 min
15-60 min	37.5 min
1-2 h	90.0 min
2-4 h	180.0 min
More than 4 h	360.0 min

Table 3.2: Time categories and translation into concrete values.

3.3.3 Test setting for transmission models

In the results section, we compare how the proposed mechanistic transmission model differs from the classical model assuming an equal transmission probability for all contacts. Thereby, both the contact structure and the basic reproduction number R_0 are fixed for both transmission models. We use the classical definition of R_0 as the average number of secondary cases generated by an infected individual being introduced into a fully susceptible population [8].

We first analyse the effect of the observed patterns of contact duration and assume the intensities i_{nm,t_x} to be equal and constant for all contacts. Then, we analyse the impact of contact intensity in a qualitative way. Information on shedding rates and inter-individual differences is available for many diseases (e.g., influenza cf. [110]). However, as we are more interested in exposure differences due to contact structure than in the impact of shedding rate differences, we also assume q_{m,t_x} to be equal for all infectors m. We further concentrate on hypothetical diseases with an infectious period of one day and basic reproduction numbers $R_0 = 1.5$, 3.0, 4.5, and 6.0. With these assumptions, the contact intensity and the shedding rate can be included in a new calibration parameter $\Theta' = \Theta \cdot \bar{q} \cdot \bar{i}$, and Equation 3.1 can be simplified to

$$P_n = 1 - \exp\left(-\Theta' \sum_{m=1}^{I} t_{nm} \right) \tag{3.2}$$

The expected number of secondary cases SC generated by a specific infector m if introduced into a completely susceptible population can then be calculated as follows:

$$SC_m = \sum_{n=1}^{S} P_n \tag{3.3}$$

with S represents the total number of susceptible individuals infector m has contact with during the day m is infectious. Finally, the equation

$$R_0 = \frac{1}{X} \sum_{m=1}^{X} SC_m \tag{3.4}$$

reveals the basic reproduction number as defined previously when X includes the total population of interest.

The following two analyses are used to contrast the effect of the mechanistic model (Equation 3.2) against the classical model:

1) We illustrate the relationship between the expected number of secondary cases SC and the number of contacts S by calculating SC for the 720 person days as separate units of observation. SC is calculated according to Equations 3.2 and 3.3 and based on the contact durations measured with the contact diaries. We group the SC -values by S and show the so grouped SC -values in box plots. We do this for different values of Θ'; Θ' is determined such that $R_0 = 1.5$, 3.0, 4.5, and 6.0 for the given test population according to Equation 3.4. The contact intensity i and the shedding rate q are assumed to be constants. We then compare the number of secondary cases SC of the simplified mechanistic model with the analogue SC value when the assumption of a constant per-contact transmission probability is used (also grouped by S).

2) In the second analysis, we calculate how the contact intensity is related to the duration of a contact in the empirical data set of potentially contagious contacts. This allows a qualitative discussion related to how the inclusion of variable contact intensities instead of a constant might affect the results found in analysis 1.

3.4 Results

Figures 3.1a-d show how the expected number of secondary cases of an infector introduced into a fully susceptible neighbourhood is related to the number of contacts (following Equation 3.3). Each subfigure represents another level of infectivity of the hypothetical infectious agent. Despite all of the random fluctuations, the following trends are quite clear:

1) Unsurprisingly, the expected number of secondary cases SC tends to be higher for highly connected individuals than for those with only few contacts.

2) For low contact numbers, the median expected number of secondary cases $\widetilde{SC}$ and the number of susceptible contact partners S appear to be linearly related. For high contact numbers, the gradient $\widetilde{SC}/S$ is less steep than for low contact numbers.

3) As a disease becomes more infectious, the relationship between $\widetilde{SC}$ and S seems to come closer to linearity. In Figure 3.1a ($R_0 = 1.5$), $\widetilde{SC}$ seems to reach a more or less stable plateau for $S > 10$, while in Figure 3.1d ($R_0 = 6.0$), $\widetilde{SC}$ appears to be an almost linear function of S. This impression is supported by regression analysis: If S is used as independent variable in a linear regression model to explain SC, the variance explained by this linear model equals 0.249 for $R_0 = 1.5$, 0.339 for $R_0 = 3.0$, 0.493 for $R_0 = 4.5$ and 0.696 for $R_0 = 6.0$ (all four linear regression analyses refuse the null hypothesis $R^2 = 0$ on a significance level of $p < 0.01$ using a F-test).

Figure 3.2 shows how the proposed mechanistic model deviates from the classical transmission model if both are fitted to the same basic reproduction number and have the same underlying contact structure. Average deviations are shown for the whole range of S and R_0. The average deviations were normalized by the basic reproduction number R_0. Figure 3.2 reveals that individuals with less than 11 contacts have a slightly higher number of expected secondary cases when the transmission model depends on contact duration as compared to the case of a constant per-contact transmission probability. At the same time the classical model exceeds the mechanistic one in reference to highly connected individuals. The slight differences in case of individuals with less than 11 contacts can compensate for the rather pronounced differences of highly connected individuals as the majority of the person days reported eight or less contacts while highly connected individuals are rather seldom.

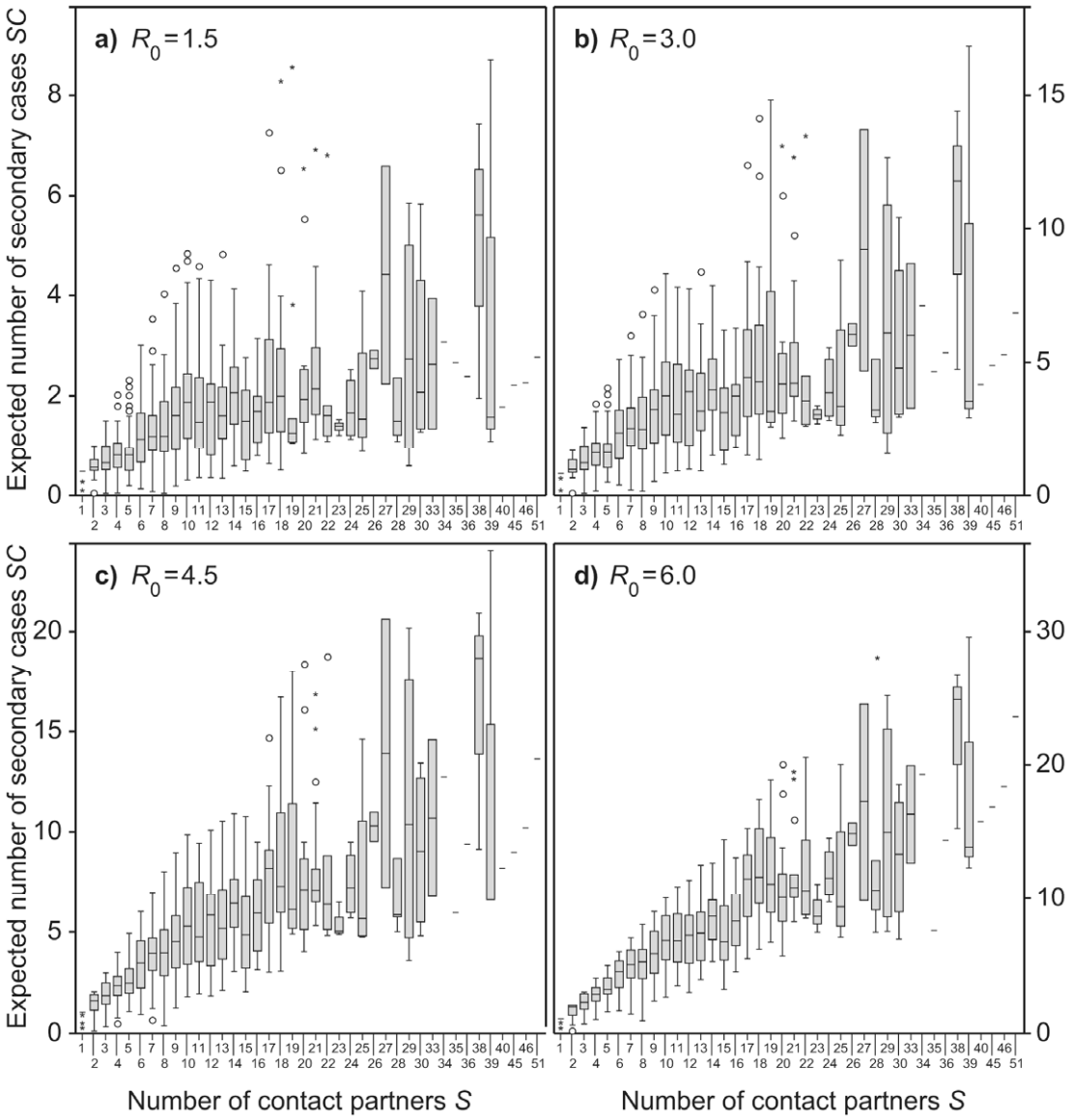


Figure 3.1: Expected number of secondary cases versus number of contacts. The boxplots show the distribution of the expected number of secondary cases that are induced by one infector that is introduced into a fully susceptible population. The values are grouped by the number of contact partners of the infectors. The boxes represent the interquartile range (IQR) with the median values marked as horizontal line. The whiskers are defined as max. $\pm 1.5 \cdot IQR$. Circles are outliers and asterisks are extreme outliers. The subfigures represent the following basic reproduction numbers: a) $R_0 = 1.5$; b) $R_0 = 3.0$; c) $R_0 = 4.5$; d) $R_0 = 6.0$. Subfigures a-c are cropped such that one outlier lies outside the displayed range. The corresponding person day had 28 reported contacts and amounts $SC = 14.15$ for subfigure a, $SC = 24.00$ for b, and $SC = 27.76$ for c. The rationale for this outlier is presumably a reporting bias from the participant; i.e., the participant stated that she or he had close contact lasting for hours with a large number of persons at a festivity. Interacting closely with a large number of persons at a festivity over long time is almost impossible when the rigid contact definition in the diary is used.

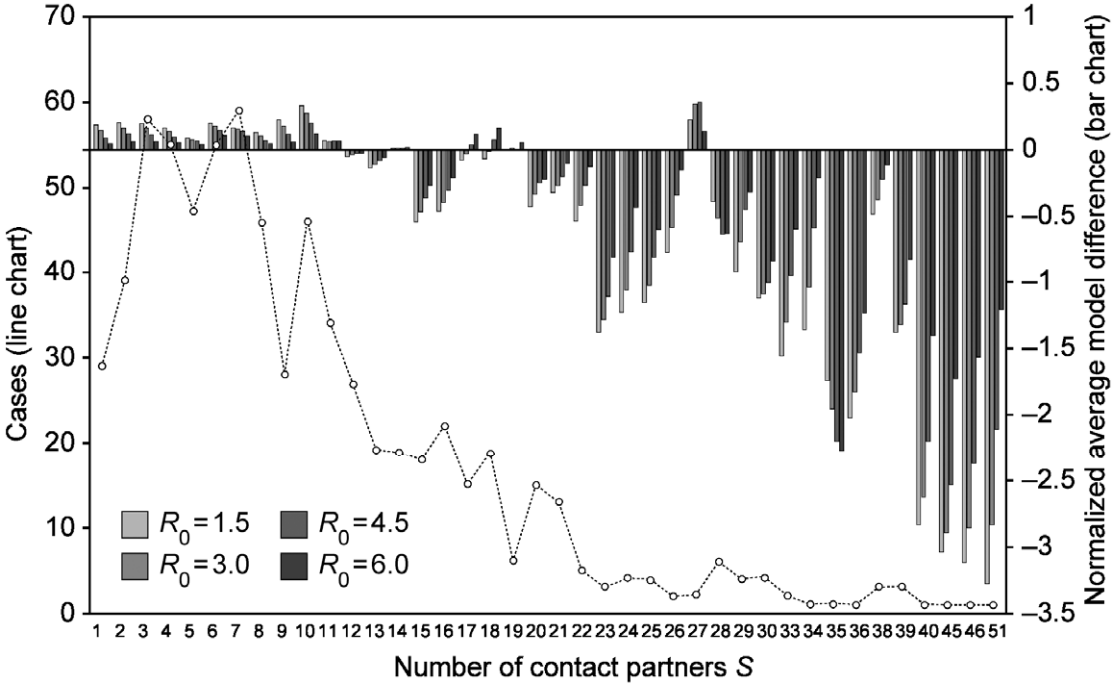


Figure 3.2: Mechanistic versus classical model. The bars show the average difference between the expected number of secondary cases of the mechanistic model SC_{mech} and that of the classical model SC_{clas} when normalized by the basic reproduction number R_0 and grouped by the number of contact partners S. The sequence of reproduction numbers R_0 within each category S goes from $R_0 = 1.5$ on the left (light grey bars) to $R_0 = 6.0$ on the right (black bars) in steps of 1.5. The line shows how many person days with exactly S contact partners exist in the sample.

Figures 3.1a-d and Figure 3.2 are based on Equation 3.2, which accounts for contact duration but ignores the influence of contact intensity. Figure 3.3 reveals how contact duration and contact intensity are interrelated, thereby allowing an interpretation of how the consideration of the contact intensity might alter the findings presented in Figures 3.1a-d and 3.2. Figure 3.3 shows separately for the three different levels of contact intensity how the reported numbers for the six duration categories deviate from the expected numbers (i.e., assuming no relation): Far more contacts of less then 5 minutes were observed than expected within the purely conversational contacts, while contacts of more than 4 hours are overrepresented in the most intensive contact category. This finding is also reflected in a positive correlation coefficient between these two ordinal variables: The non-parametric Kendall rank correlation results in $\tau = 0.388$, which is significantly different from zero at the 0.01 level.

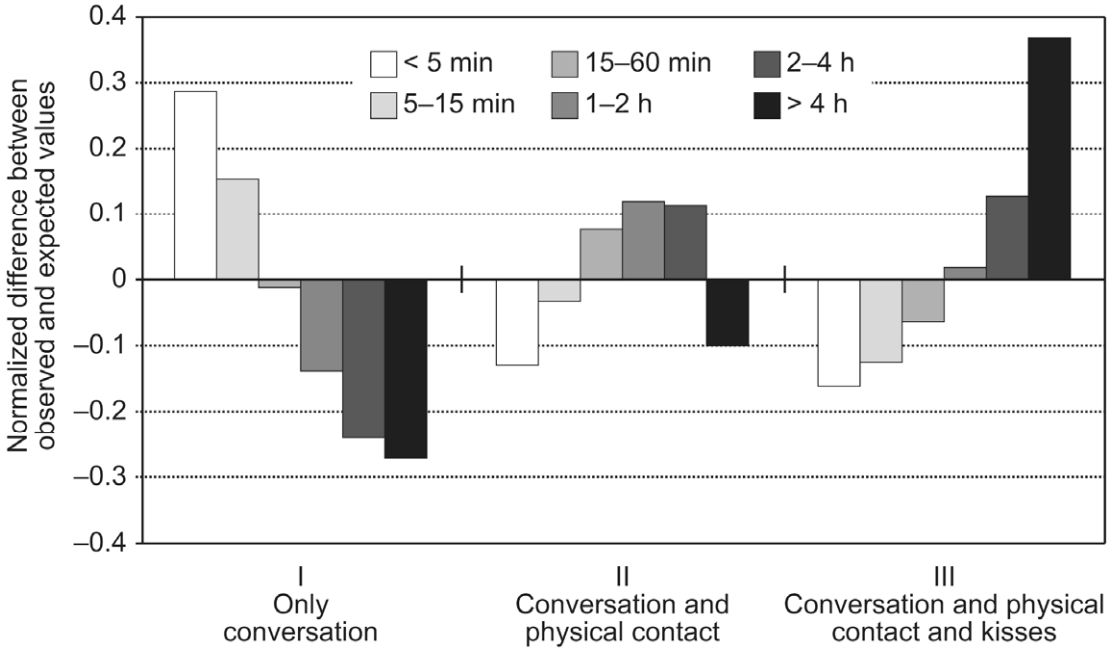


Figure 3.3: Relationship between contact duration and intensity. This figure shows the relationship between contact intensity (I, II, III) and duration (six categories). The bar charts result from subtracting the expected values (assuming no relation between intensity and duration of contacts) from the numbers of observations for each possible combination of duration and intensity. Hence, a positive value means that there were more observations of a certain duration-intensity-combination than would be expected if duration and intensity are independent. A negative value means that there were less observations than expected. The value zero means that the expected and observed numbers are the same. Every bar is normalized by the total number of observations for every time category.

3.5 Discussion

3.5.1 Implications of the results presented

The presented results elucidate the implications of accounting for contact duration and intensity in simulations of disease spread. Figure 3.2 suggests that the importance of highly connected individuals, so-called super-spreaders, is strongly overestimated when all contacts are assumed to be equally likely to transmit infectious disease. This finding is particularly important because some well cited publications have concluded that highly connected individuals are major drivers of disease spread without accounting for the heterogeneity of the inter-individual transmission probabilities [185, 263]. The results suggest that in the case of a disease with a low reproduction number, the expected secondary cases induced by individuals with many contacts are in the same range as those induced by individuals with medium numbers of contacts. Only when R_0 is close to the mean number of contacts $\bar{S}$ is the expected number of secondary cases

approximately linearly related to the number of contacts (see Figure 3.1 and the linear regressions shown in the results section).

This finding can be easily explained by the fact that the marginal total contact time decreases with every further contact person. In other words, most people have only a small set of persons (usually at home or at work) with whom they meet and interact for long periods during a day. Those individuals who meet with far more people than the average spend on average less time with every single contact person than those persons who have some or only a few contact partners per day. This can be illustrated with the example of train conductors, flight attendants, or supermarket cashiers; indeed, all of them have contact with hundreds of people a day, but they interact with each single contact only for a very short time.

As a consequence, highly connected individuals have more potentially contagious contacts than others, but these contacts are simultaneously on average less likely to transmit disease. Highly connected persons can reach their full "super-spreading" potential only if a disease is so contagious that almost every contact with a susceptible person leads to infection. Similar findings have been reported for sexually transmitted diseases: Research has shown that individuals with many different sexual partners per year are less important for disease propagation than often assumed because they have less sex acts per partner and in total than individuals with a smaller number of partners [49, 251].

The conclusion that highly connected individuals are overestimated in their importance if a constant per-contact transmission probability is assumed is further supported by the analysis of the contact intensity as reported in the contact study. Theoretically, a short interaction between a susceptible and an infectious person could lead to a comparable amount of ingested infectious material as that of a long interaction assuming that the short interaction is more intensive than the long one. However, prolonged contacts tend to be more intensive than short contacts because they more often involve closer interaction, such as physical contact and kisses. This finding is plausible because those persons with whom the individuals spend much time are in most cases their loved ones, thereby indicating the higher likelihood of more intimate interactions than with casual acquaintances. As a result, the conclusions from the analysis of the pure contact durations are even further pronounced by taking contact intensities into account.

Therefore, our results suggest that sole concentration on the connectedness of individuals to explain super-spreading events is not valid. The explanation for super-spreading events might lie in a combination of many contacts and high shedding rates (cf. the notion of "super-shedder" [178]). Extreme numbers of secondary cases can only be achieved when the shedding rate q_{m,t_x}

of an infector m is much higher than the average shedding rate and only if this infector m has many susceptible contacts S.

3.5.2 Limitations

We see three limitations in our study. First, the empirical data set used to test the proposed exposure-based transmission model is rather small and not representative of the population. Furthermore, the person days used to calculate secondary cases are not statistically independent from each other as every person participating in the study contributed diary entries for several days. Finally, contact patterns are dependent on the cultural background and may look differently in Italy, Germany, Thailand, or Sudan [237]. Thus, the generalizability of our results may be questioned. Although this limitation exists, it is not likely to bias the presented results in a relevant manner. The observed contact patterns are plausible and theoretically grounded. An increasing number of contact partners per time unit naturally results in a decrease in the time spent with each single contact partner. Additionally, most people plausibly have only a very limited set of persons with whom they interact very closely. Additionally, the attributes of our contact structure are in complete agreement with other empirical studies on potentially contagious contacts that have also addressed similar attributes [237].

Secondly, the six time categories of the diary study offer rather imprecise information on the actual time that two persons interacted, and the three intensity categories are too vague to be translatable into concrete i_{nm,t_x} values for use in Equation 3.1. Hence, the results presented have to be qualified in a quantitative rather than qualitative sense. For every time category, we defined a precise value (the arithmetic mean of the upper and lower boundaries) that was used for all calculations. However, a sensitivity analysis that alters the actual duration defined for every category within the given boundaries does not lead to qualitatively different results (see Additional File 2). Although the measured intensity indicator is not sufficiently precise to allow inclusion in a mathematical sense, the analysis clearly indicated that inclusion of contact intensity would amplify the observed phenomena rather than falsify our conclusions.

Finally, this paper makes statements about the expected number of secondary cases of infected individuals in a fully susceptible population. In a simulation model of disease spread, the importance of an individual also depends on her/his position within the contact network structure; i.e., the network position of every individual determines the likelihood of becoming infected as well as the susceptibility status of the surrounding individuals. Due to the complex nature of simulation models of disease spread, complete simulation models must be designed and tested for sensitivity to the changed transmission model proposed in order to allow precise statements on the impact of exposure-based transmission models on simulation outcomes.

3.6 Conclusions

The goal of this paper was to provide evidence for the need of exposure-dependent transmission models and to suggest a mechanistic transmission model that can be used in simulations of disease spread. One remarkable result is that individuals with many contact partners seem to be less important for the transmission of diseases that are transmitted by droplets or physical contact than suggested by the classical assumption that all contacts are equally infectious. Particularly with only slightly infectious diseases, contacts should be differentiated by their potential to transmit infection when simulating disease spread.

This paper proposes an approach that enables the replacement of the problematic assumption of equally weighted contacts or purely statistical approaches to differentiate potentially contagious contacts with a mechanistic model. The proposed transmission model is based on well-established dose-response models that were developed in microbiology and builds upon assumptions that are closer to reality and better justifiable than the assumption that all contacts have the same transmission potential.

The spread of infectious disease is governed by a complex interplay of social and biological factors and to fully grasp its dynamics, processes on both the individual and the population level have to be understood [191, 338]. Therefore we suggest including *a priory*, mechanistic models in simulations of disease spread and combining them with an *a posteriori*, statistical approach: Often data is available that allows fitting a simulation model that includes such mechanistic elements to empirical data, thereby making use of the advantages of both approaches.

3.7 Competing interests

The author declares that he has no competing interest.

3.8 Authors' contributions

TS is the sole author of this paper.

3.9 Additional material

Additional file 3.1

Contact Data. This additional file contains the processed empirical contact data that was used for calculating the results presented in this paper.

http://www.biomedcentral.com/content/supplementary/1742-4682-6-25-S1.pdf

Additional file 3.2

Sensitivity Analysis. We assumed the arithmetic mean of upper and lower bound as precise representations of the duration categories. In this additional file we provide an analysis on how sensitive the results react on changes to this assumption.

http://www.biomedcentral.com/content/supplementary/1742-4682-6-25-S2.pdf

4 Contacts between poultry farms, their spatial dimension and their relevance for avian influenza preparedness

Fiebig L., Smieszek T., Saurina J., Hattendorf J., Zinsstag J. Geospat Health 2009, 4: 79-95

4.1 Abstract

Ongoing economic losses by and exposure of humans to highly pathogenic avian influenza (HPAI) in poultry flocks across Asia and parts of Africa and Europe motivate also outbreak-free countries such as Switzerland to invest in preparedness planning. Country specific population data on between-farm contacts are required to anticipate probable patterns of pathogen spread. Information is scarce; in particular on how strongly small, non-commercial poultry farms are involved in between-farm contacts. We aimed to identify between-farm contacts of interest for HPAI spread at both commercial and non-commercial farms in a non-outbreak situation: whether or not commercial and non-commercial farms were involved in poultry and person movements and shared resources by company integration. Focus was on poultry movements for the purpose of purchase, sale and poultry show visits, their spatial dimension, their frequencies and the farm types they connected.

This was to inform the discussion on whether at all, and under what circumstances poultry farms, and non-commercial farms in particular, play a role in the sector's connectedness and how they should be considered in the HPAI surveillance system and in pertinent transmission models.

Of the total 49437 recorded poultry farms in Switzerland, 95% had less than 500 birds. The farm number resulted in densities of up to 8 poultry farms per square kilometer and a median number of 47 neighbour farms within a 3 km radius around the farms. Person movements and shared resources were identified in 78% of the surveyed farms (93% among commercials, 67% among non-commercials). Poultry trading movements over extensive spatial ranges were stated at 65% (79% among commercials, 55% among non-commercials). Movement frequencies depended on farm specialization and were higher for commercial than for non-commercial farms except for poultry show visits. Estimates however for the entire population revealed 3.5 times higher chances of a poultry purchase, and 14.6 times higher chances of exhibiting birds at

poultry shows occurring in a given time by a farm smaller than 500 birds (non-commercial farm) than by a larger (commercial) farm.

These findings indicate that both commercial and non-commercial farms are involved in neighbourhood and remote between-farm contacts relevant to HPAI spread. It is necessary to include all poultry farms, irrespective of their size and purpose in both livestock registration and disease surveillance systems, as well as in transmission models for poultry and zoonotic diseases.

4.2 Introduction

Highly pathogenic avian influenza (HPAI), has been noted for decades as an animal disease with high economic impact. Although well documented and reported, HPAI received little public attention until 1997 when, for the first time, human infections due to the H5N1 HPAI virus strain were confirmed [80] and caused 262 confirmed fatal human cases to date [367]. Since December 2003, HPAI viruses, mainly H5N1, have reached poultry populations across Asia, parts Africa and Europe causing high economic losses [81, 106, 193, 350]. Switzerland has been free from HPAI in domestic poultry since the 1930's but in early 2006 34 cases of H5N1 HPAI infected dead water fowl were identified [163]. Both wild birds [183] and the import of poultry and poultry products represent a certain risk of HPAI virus introduction into the Swiss poultry sector [153].

HPAI virus transmission to susceptible birds occurs by direct contact with excretions and secretions from infected birds and indirectly via contaminated water, feed and equipment used on a farm. Between-farm transmission can occur through direct bird-to-bird contact when subclinically infected poultry is traded or exhibited at poultry shows. Other animals such as wild birds, martens, or domestic cats are known to potentially act as vectors [188, 252, 368]. People can contribute to virus spread by introducing contaminated fomites into a susceptible flock. Such between-farm contacts are also depending on the organization of the local structure of poultry industry [60].

It is known from post-outbreak investigations that such potentially contagious contacts, in particular livestock movements between farms, strongly influence the course of epidemics [312]. The distribution of number of contacts (degree distribution) among the members of a population (here poultry farms) was shown to be relevant for identifying members with high probabilities of being infected early in a course of epidemic because of having many incoming contacts. Members having many outgoing contacts were causing high numbers of secondary cases [22, 29, 363]. Furthermore, it was shown that high dispersions of degree distributions lowered the epidemic threshold, and thus were an important factor to consider when predicting epidemic dynamics [8, 87, 160, 263]. Clustering, describing how many of a member's contact

partners have contact amongst one another, and other structural properties such as the stability of contacts further influence the spread of disease. To assume that all members have equal numbers of contacts and that they randomly chose contact partners, changing them continuously as is often done in transmission models, is known to overestimate the size of an outbreak for many infectious diseases [210, 314, 373].

Only rarely detailed contact information in its spatial context has been systematically integrated in models for HPAI transmission and used for the planning of preparedness and control strategies. Boender et al. [36] performed a spatial analysis of the HPAI outbreak occurred in 2003 in the Netherlands. They modeled HPAI transmission from infected to uninfected farms as a function of inter-farm distance and farm density. Resulting risk maps help to define areas where preemptive culling is advisable. Truscott et al. [334] showed that transmission models taking both density-dependent spatial transmission and periodic network contacts into account were particularly suitable to reflect HPAI spread within the Great Britain poultry flock. Other countries, especially those not yet experiencing HPAI outbreaks can draw on these findings in their own preparedness planning.

Country specific information on the spatial distribution, structural composition, and the connectedness of the poultry sector is required to develop transmission models properly. In particular it has to be clarified to what extent non-commercial poultry farms should be considered. Their role in between-farm transmission is controversial. Often non-commercial farms were defined by small flock sizes and were assumed to have a small poultry movement distances. However, Garber et al. [123] investigated destination locations for "birds sold or given away" by non-commercial farms in the USA and found movements beyond the State and beyond the USA borders. Capua et al. [59] suggested defining non-commercial backyard poultry farms not only by small flock size but primarily by the absence of functional connection to commercial poultry production systems. Such definition would imply that specific information on the interconnectedness of the poultry sector is available. Boender et al. [36] considered only commercial flocks in their model. In Great Britain, only farms with 50 or more birds kept have to be registered, and are thus included in models. Distant contacts were only taken into account for farms keeping 500 or more birds [334] or 1000 and more birds [81]. This makes it difficult to judge the actual role of non-commercial poultry husbandries in between-farm transmission scenarios.

This study was aimed to identify between-farm contacts of interest for HPAI spread at both commercial and non-commercial farms in a non-outbreak situation. We took advantage of available data in Switzerland where registration of poultry farms irrespective of size and purpose has been introduced in 2005 on a communal and cantonal level [304]. We geo-

referenced the locations of poultry farms to understand where occasional between-farm contacts within a neighbourhood were most probable. We then identified in a cross-sectional study whether commercial and non-commercial farms were involved in person movements, such as employees shared by two farms, and shared resources by company integration (affiliation to poultry marketing organizations). Of particular interest were poultry movements for the purpose of purchase, sale and poultry show visits, their spatial dimensions, their frequencies and the farm types they connected.

This was to inform the discussion on whether at all, and under what circumstances poultry farms, and non-commercial farms in particular, play a role in the sector's connectedness and how they should be considered in the HPAI surveillance system and in pertinent transmission models.

4.3 Material and methods

4.3.1 Study population and density of poultry farms

The population investigated in this study are the poultry farms of Switzerland. By "poultry farm" we understand all sites where one or more domestic chicken (*Gallus gallus domesticus*), turkey (*Meleagris gallopavo*), duck (*Anas platyrhynchos domesticus* or *Cairina moschata*), goose (*Anser anser*), quail (*Coturnix coturnix*), guinea fowl (*Numida meleagris*), peafowl (*Pavo cristatus*), ostrich (*Struthio camelus*), and/or pigeon (*Columba livia*) are kept.

We established a single list of all recorded poultry keepers and farms (data from 2005 to 2007) in Switzerland out of 23 registers maintained by the 26 Swiss cantons (some cantons cooperate), and the federal livestock register database "Agrar information system (AGIS)" from 2005 [55]. The AGIS contains only farms receiving direct government subsidy. The cantons recorded either all their poultry farms or only those not included in AGIS. Therefore data from all sources had to be merged and duplicates to be eliminated electronically privileging the more recent cantonal records. This lead to a single list subsequently called "census" containing a total of 49437 countrywide identified poultry keepers. Captured attributes included farm address and total number of birds kept. Further farm details were provided in the original registers, however not in a standardized way. Manual checks revealed similar entries of farms under different names. Thus, the census might still contain some duplicates.

The address data from the census were geo-referenced and read into a base map from Swisstopo 2008®. An accuracy of exact localization was reached for 78% of the farms. For 6% and for 15% only precision on the street level and on the postal code level could be achieved,

respectively. The census was used to investigate the density distributions of poultry farms and birds kept for the entire country and to depict them in density maps.

4.3.2 Survey design

The investigation of the between-farm contacts and their determinants followed a mixed methods research design. First a quantitative cross-sectional study among poultry keepers was conducted. In addition five experts from companies integrating commercial poultry farms (poultry and egg marketing organizations) were interviewed (qualitative part).

Cross-sectional study among poultry keepers

The census was used as sampling frame of which a random sample of 3978 poultry keepers was drawn. The poultry keepers' probability of being selected for the cross-sectional study was proportional to the square root of the number of birds kept on their farm (farm size), to ensure a sufficient number of the less numerous larger poultry farms.

A mail-out/mail-back survey among the 3978 selected poultry keepers was conducted between August and December 2007. As survey instrument a structured questionnaire was developed in the German language and translated into French and Italian; national languages of Switzerland. Topics covered between-farm contacts, a self-assessment of the farm type by the respondent, a section on disease awareness, and one on wild bird observations in the poultry free-range area if existing. The two latter topics are presented in other manuscripts [294] and unpublished data.

Defining relevant contacts

Between-farm contacts potentially relevant for HPAI transmission were identified based on available literature [82, 137, 332, 365, 368] and based on consultation with poultry experts. The investigated contact relations included farm neighbourhood and neighbourhood-related contacts. Farm neighbourhoods are commonly considered to allow for casual contacts between the poultry keepers and overlapping movement ranges of potential vectors such as sparrows and freely moving domestic animals such as cats being potential vectors for HPAI viruses [196, 279]. This is reflected in the implementation of control and surveillance zones with 3 km and 10 km radii as a HPAI control measurement regulated in the Animal Health Act [56] and 1 km bands for risk zones in other appraisals [152]. Therefore, the number of the participants' neighbour farms within all 1, 3, and 10 km radii was based on the addresses given in the poultry farm census. Contacts surpassing a 10 km radius were defined as remote contacts.

Investigated contact relations beyond neighborhoods included human movements, shared resources and poultry movements (Table 4.1). Poultry movements for the purpose of "purchase" and "sale" had one direction, those for "exhibiting birds at poultry shows" were bidirectional.

The questionnaire allowed specifying of up to six different contact partners for each purchase, sale and show visits. Date (month/year), site (postal code), and types of contacts (hatchery, other farm, or abattoir/butcher) or name of poultry show were inquired. The frequency of poultry trade and show visits was captured in "x times per year" and "less than once a year" which was coded as 0.5 times per year in the analyses. The term "poultry" included here live birds of the species described above, one-day chicks and also hatching eggs.

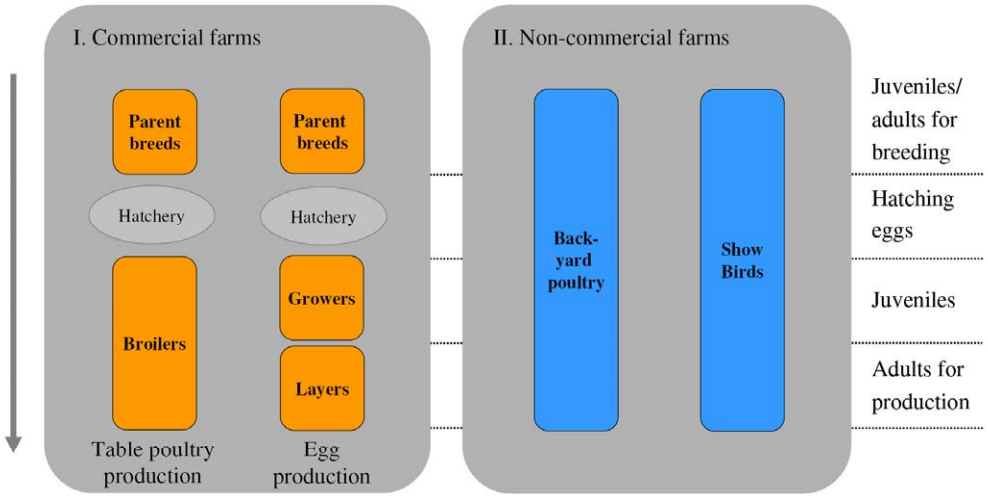
Contact relation	Vector	Connection through	Source of information
Neighborhood			
Neighborhood to other poultry farms within 1, 3, and 10 km	Human and animal vectors	Proximity	Poultry farm census
Person movements and shared resources			
Poultry show (visiting only)	Person	Co-attending show	Questionnaire
Co-working	Person, equipment	Staff and equipment	Questionnaire/interviews
Dead stock collection	Person, equipment	Co-accessing communal dead stock collection point	Questionnaire/interviews
Company integration	Person, equipment	Staff and shared resources	Questionnaire/interviews
Poultry movements			
Poultry purchase	Live birds/hatching eggs	Transport (unidirectional)	Questionnaire/interviews
Poultry sale	Live birds/hatching eggs	Transport (unidirectional)	Questionnaire/interviews
Poultry show (exhibiting birds)	Live birds	Co-attending show	Questionnaire

Table 4.1: Overview on contact relations under study.

4.3.3 Data processing and analysis

Data of the returned and completed questionnaires were double-entered into a database, compared and cleaned. Presented analyses rely on data of 1317 (33%) questionnaires that contained valid contact information. Spatial data were collected for all poultry movements, "show visits", and "co-working" in the form of the postal code of the contact partner or event. Postal codes were geo-referenced. Maximum air-line distances in km between respondents and contacts were calculated for each contact relation if the postal code was given. Map presentations were completed using the maptools and spatstat libraries in R and base maps from Swisstopo 2008®.

Two participant groups were formed based on the respondents' self-description in the questionnaire: "commercial" and "non-commercial" poultry farms. Further information on these groups is provided in Box 4.1.



Poultry is kept in different livestock production systems for different purposes and on different professional levels that is in different farm types. Legal definitions for these farms types differ between countries or are missing. It is commonly understood that commercial poultry farms (I) add essentially to earning a living and are operating with more than 500 and up to tens of thousands of birds. These farms, represented by orange boxes, typically cover only one step in either the table poultry production line or the egg production line. The arrow indicates the general direction of production. In Switzerland, parent breeds (F1) are imported from few global companies. Imports of breeding eggs by hatcheries and of one-day chicks by producers supplement the domestic production. Hatcheries don not keep live birds and are not considered as farms. Non-commercial poultry farms (II), represented by blue boxes, comprise all farms with smaller flock-sizes. Birds are kept for sideline production, subsistence farming and/or leisure. We distinguish between backyard poultry, when the emphasis is on the production of table poultry and/or eggs for human consumption, and show birds, when fancy birds or rare species are kept for breeding and preservation.

Box 1

Multinomial models with poultry movement distances as an outcome were used to investigate the following explanatory variables: number of birds kept (farm size), the respondent's farm type, and flock composition. Estimates and confidence intervals for the poultry movement frequency of the entire poultry sector were constructed using Bootstrap resampling with 2000 replications.

4.3.4 Interviews with experts from poultry industry

For the purpose of data triangulation and complementary information on between-farm contacts, interviews with experts from the poultry industry were conducted in addition to the survey. Five companies integrating commercial poultry farms in Switzerland were selected for interviews. The selection was based on whether the companies were frequently named by the survey participants and in order to include different areas of the poultry industry, including broiler and egg production. Company I and II, integrating about 400 farms each, covered the entire broiler production line from the hatchery and to the abattoir. Companies III to V were involved in egg production; company III contracted about 100 farms with laying hens, company IV regrouped 110 organic farms on different levels. Company V covered around 60 farms levels

plus one hatchery. All together the experts represented about one-half of the some 2000 commercial poultry farms in Switzerland.

Main topics of the interview were between-farm contacts among the company's integrated farms, contacts to outsiders, and shared resources. The experts were asked to describe production cycles, numbers, and specifics of their integrated farms. An interview guideline was used to systematically probe on issues not mentioned spontaneously by the experts. Information on poultry trade and shared resources was depicted by expert and interviewer together on paper (mapping tool). Here, different colors were used to draw the studied contact relations (Table 4.1) amongst the company's farms, and to outsider farms. The interview protocols including notes from experts and the interviewer were transcribed and underwent qualitative content analysis according to Mayring [215].

4.4 Results

4.4.1 Poultry farm density and neighborhood

The identified number of poultry farms in Switzerland was 49437 until May 2007. The largest poultry flock comprised of 47300 birds and the smallest had 1 bird. 95% farms had less than 500 birds, and 90% had less than 50 birds. The poultry farm density differed amongst regions. High density areas with more than 8 farms per square km were presented in purple, areas with moderate farm density in yellow and with very low farm density and no farms in white. Light areas were congruent with high altitudes in the Alps in Southern Switzerland (Figure 4.1). The distribution of the number of birds kept per square kilometer resembled roughly the farm density distribution with low densities in the Alps. Maxima with more than 2500 birds per square kilometer were, however, more in the west of the country between Berne and Lausanne reflecting the location of several large commercial farms (Figure 4.2). South of Bellinzona farm density was at a maximum, but low numbers of birds were kept per square kilometer reflecting the sparsity of large commercial farms in that area.

In the sample of 1317 poultry farms, 543 were self-described as commercial farms and 783 as non-commercial farms. Similar group sizes were due to the weighted sampling privileging the less frequent large farms. The median total number of birds kept was 4500 for commercial farms and 15 for non-commercial farms (Table 4.2). The threshold between both farm groups was roughly around 500 birds. 97% of farms had other farms within 1 km of the farm. Equal median numbers of neighbor farms representing potential contacts were found for both commercial and non-commercial farms with a median of 11 poultry farms within 1 km, 47 within 3 km, and 283 within 10 km (Table 4.2).

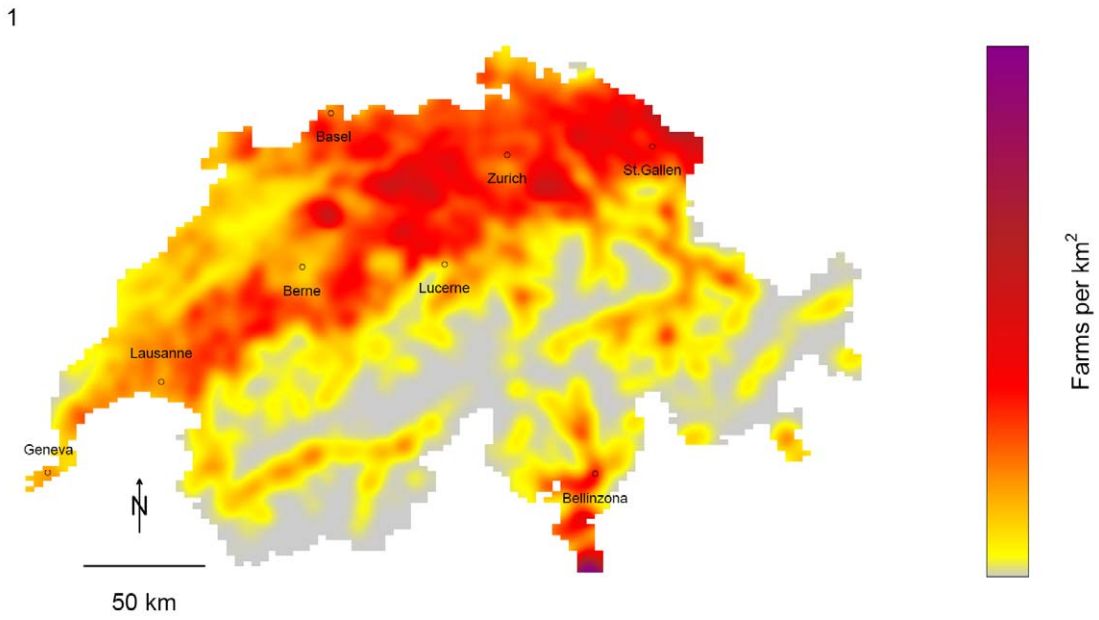


Figure 4.1: Density distribution of poultry farms in Switzerland (in farms per km²). Locations of important cities of Switzerland are given for orientation.

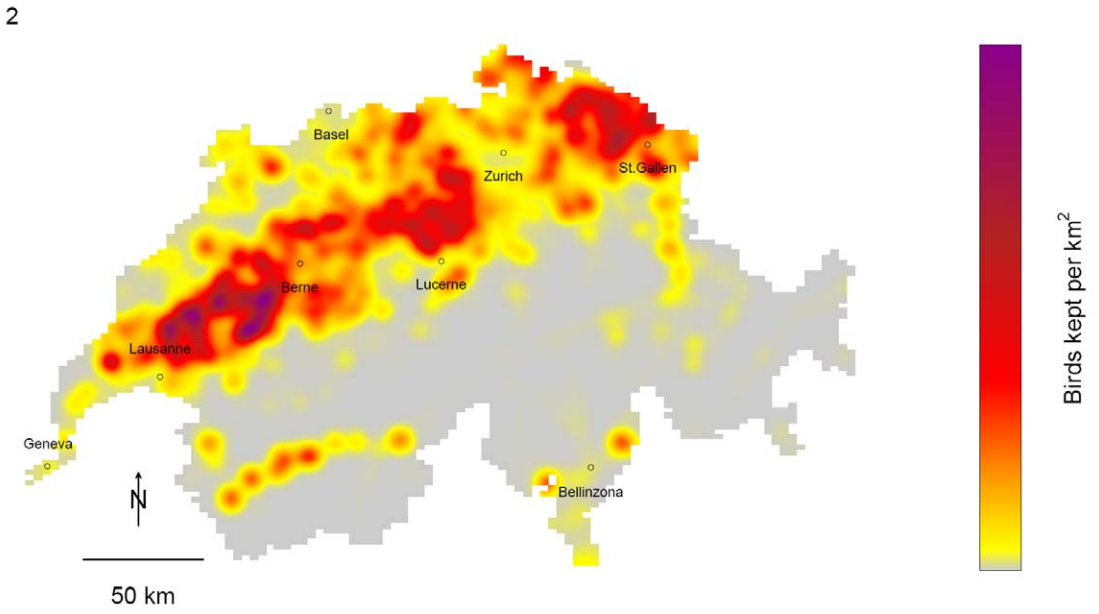


Figure 4.2: Density distribution of birds kept in Switzerland (in farms per km²). Locations of important cities of Switzerland are given for orientation.

Potential human and animal vectors (cats) were found on commercial and non-commercial farms. In both groups a median of 3 people were, on average, present on the farm during a normal working day. These persons were mostly described as "staff" at commercial farms and as "residents" and "guests" on non-commercial farms. One or more cats were kept on 65% of the farms without significant difference between commercial and non-commercial farms (unpublished data).

	Commercial	Non-commercial	All	Extrapolation to CH [c] poultry sector
No. of birds kept per farm	n=534	n=783	n=1317	n=1317
(m [a] [IQR [b]])	4500 [2000-8610]	15 [7-30]	37 [12-3807]	11 [6-23]
Fraction of farms having neighbor farms in radii of	n = 532	n = 780	n = 1312	n = 1312
1 km	98%	96%	97%	97%
3 km	100%	99.7%	99.8%	99.4%
10 km	100%	100%	100%	100%
No. of neighbor farms in radii of				
1 km (m [IQR])	11 [7-18]	11 [6-19]	11 [7-19]	11 [6-19]
3 km (m [IQR])	47 [29-75.5]	47.5 [25-74.5]	47 [28-75]	46 [25-73]
10 km (m [IQR])	289 [162.5-402]	279 [142-381]	283 [152-393]	277 [144-386]

Table 4.2: Farm specifics and neighborhood of the commercial and non-commercial farm group and data extrapolation to the entire Swiss poultry sector. [a] median; [b] inter-quartile range; [c] Switzerland.

4.4.2 Person movements and shared resources

At least one incident of human movement and shared resources was present at 78% of the participating farms (93% for commercial and 67% for non-commercial farms). "Use of dead stock collection points" was the most frequent response with 75%, "company integration" was stated by 30%, "poultry shows (visiting only)" by 7%, and "co-working" on other farms by 4% of the respondents. "Use of dead stock collection points", "company integration", and "co-working" on another poultry farm were more common among commercial farms. Non-commercial farms were virtually non-integrated into companies and visited more often visited poultry shows (Table 4.3). Median distances were available for "poultry shows (visiting only)" and "co-working". Visited poultry shows were in a median distance of 12 km from the farm, with 27 km for the commercial and 8 km for non-commercial farms. This difference was explained by the commercial farm group mostly indicating visits to national agricultural expositions, and the non-commercial group mostly indicating visits to local shows and markets. "Co-working" on

other farm was mainly indicated by the commercial farm group (Table 4.3). Between farms sharing employees a median distance of 2 km was identified. Thus sharing employees happened within a neighborhood and should not be classified as a remote contact.

	Commercial	Non-commercial	All	Extrapolation to CH [c] poultry sector
Poultry show (visiting only)	n = 518 7%	n = 754 9%	n = 1272 9%	n = 1272 8%
Co-working	n = 534 10%	n = 782 1%	n = 1316 4%	n = 1316 1%
Dead stock collection points	n = 533 92%	n = 782 63%	n = 1315 75%	n = 1315 62%
Company integration	n = 534 73%	n = 783 0.3%	n = 1317 30%	n = 1317 3%
Fraction of farms having one or more of above incidents	n = 517 93%	n = 752 67%	n = 1269 78%	n = 1269 65%
Distances				
Poultry show (visiting only) km (m [a] [IQR [b]])	n = 22 27 [9-37]	n = 51 8 [5-27]	n = 73 12 [6-34]	n = 73 8 [6-34]
Co-working km (m [IQR])	n = 44 2 [1-4]	n = 5 3 [2-3]	n = 49 2 [1-4]	n = 49 2 [2-4]

Table 4.3: Prevalence of contact relations under study among the commercial and non-commercial farm group and data extrapolation to the entire Swiss poultry sector. [a] median; [b] inter-quartile range; [c] Switzerland.

4.4.3 Poultry movements

Poultry movements were identified for 65% of the participating farms, with 79% among commercial and 55% among non-commercial farms. Purchase of poultry occurred more often (61%) than sale (25%) and exhibiting birds at poultry shows (3%), with a higher contribution of commercial farms except for poultry shows (Table 4.4).

Geo-mapping of the air-line distances showed a geographical overlap of all poultry movements by commercial and non-commercial farms in farm dense areas. Itemizing poultry movements by type of origin and destination contact revealed characteristic patterns. Purchase from hatcheries (Figure 4.3a) and sale to abattoirs/butchers (Figure 4.5a) by commercial farms was focused. The foci were the same for farms integrated into the same company, confirmed by the interviewed experts. Commercial farms were not always affiliated to the company whose hatchery and abattoir were closest to the farm. Each of the companies had contract farms in up to 19 of the 26 Swiss cantons. That implies same suppliers, consulters and veterinarians serve contract farms over large parts of the country. Commercial farms' purchases from other farms were mainly identified as laying farms buying laying hens from growers. Non-commercial farms

had essentially other farms as contact partners, clear centers in the overall pattern were not identified (Figures 4.3a-4.6a).

The air-line distances of poultry purchase increased significantly with increasing farm size. For purchases from hatcheries, the increase was estimated as 0.75 km per farm size increase by 1000 birds (P = 0.026) (Figure 4.3b), for purchases from other farms the increase was 1.80 km (P <0.0001)) (Figure 4.4b). Sales to abattoirs/butchers (P = 0.378), to other farms (P = 0.718), and distances to poultry shows where a farm's own birds were exhibited (P = 0.582) did not depend on the farm size (Figures 4.5b-4.7b).

Comparison of median distances between participant groups revealed poultry purchase (25.1 km median distance) being more than twice as distant for commercial farms (40 km) than for non-commercials (16 km). Median poultry sale distances (20 km) were 25 km for commercial farms and 10 km for non-commercials, explained by the commercials' longer journeys to abattoirs (31 km). In contrast to distances for "poultry shows (visiting only)", distances to poultry shows where owned birds were exhibited were about equal for commercial (median distance of 28 km) and non-commercial farms (27 km) (Table 4.4). Within the non-commercial group show participation was mainly attributed to farms self-described as "show bird breeders" (OR = 8.0; 95% CI = 4.9-13.2, n = 783). Among the commercial farms, 6 out of 9 responses were attributable to self-described "layer farms".

	Commercial	Non-commercial	All	Extrapolation to CH [c] poultry sector
Purchase (total)	n = 534 75%	n = 783 52%	n = 1317 61%	n = 1317 50%
Sale (total)	n = 534 50%	n = 783 8%	n = 1317 25%	n = 1317 8%
Poultry show (exhibiting birds)	n = 518 2%	n = 754 4%	n = 1272 3%	n = 1272 3%
Fraction of farms having one or more of above incidents	n = 518 79%	n = 754 55%	n = 1272 65%	n = 1272 52%
Distances				
Purchase of poultry (total) km (m [a] [IQR [b]])	n = 337 40 [23-74]	n = 337 16 [8-29]	n = 674 25 [12-51]	n = 674 16 [8-32]
Purchase from hatchery	n = 223 37 [23-74]	n = 46 23 [13-37]	n = 269 36 [22-68]	n = 269 26 [16-51]
Purchase from other farm	n = 134 37 [18-67]	n = 311 15 [7-28]	n = 445 18 [8-40]	n = 445 15 [7-28]
Sale of poultry (total) km (m [a] [IQR [b]])	n = 148 25 [13-60]	n = 40 10 [2-17]	n = 188 20 [9-51]	n = 188 10 [3-21]
Sale to hatchery	n = 6 18 [6-25]	n = 2 9 [7-12]	n = 8 12 [6-24]	n = 8 12 [7-12]
Sale to abattoir/butcher	n = 119 31 [15-72]	n = 2 18 [15-21]	n = 121 30 [15-71]	n = 121 29 [15-64]
Sale to other farm	n = 27 9 [3-19]	n = 38 10 [2-17]	n = 65 10 [2-17]	n = 65 9 [2-17]
Poultry show (exhibiting birds) km (m [IQR])	n = 9 28 [16-44]	n = 25 27 [9-56]	n = 34 28 [12-48]	n = 34 18 [9-45]

Table 4.4: Contact relations and median maximum distances to contact partners in km by the commercial and non-commercial farm group and data extrapolation to the entire Swiss poultry sector. [a] median; [b] inter-quartile range; [c] Switzerland.

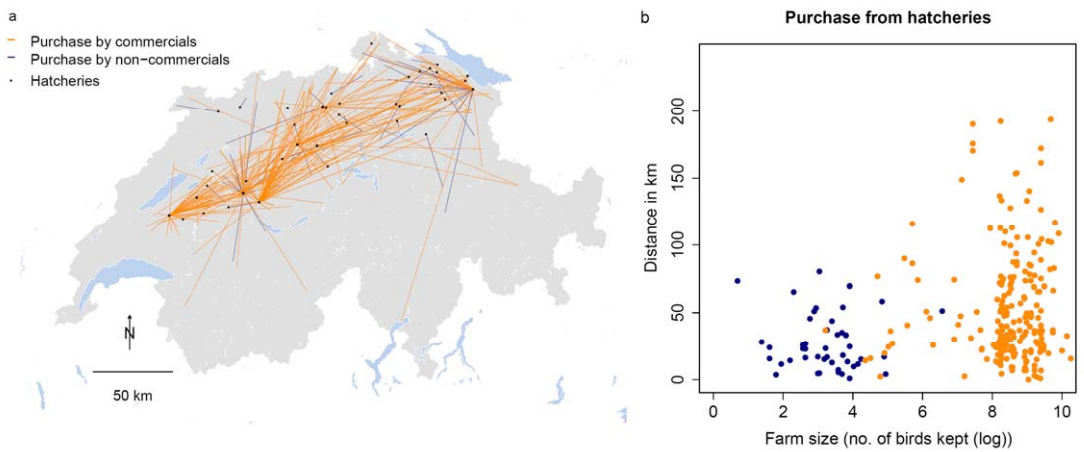


Figure 4.3: Poultry movements. Map 4.3a indicates airline distances for purchase from hatcheries (black dots) by commercial (orange lines) and non-commercial farms (blue lines). In the scatter plot 4.3b correlation between farm size (log) and airline distances is shown. Non-commercial farms are represented by blue dots, commercial farms by orange dots.

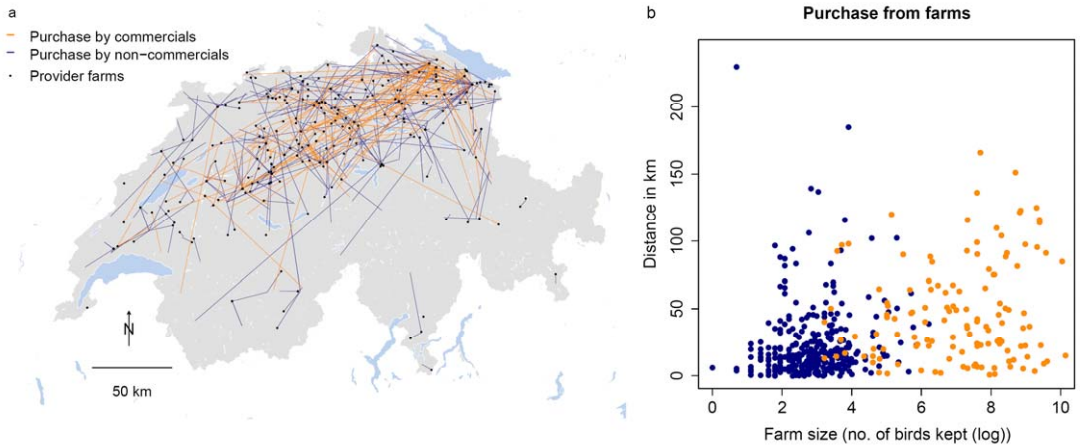


Figure 4.4: Poultry movements. Map 4.4a indicates airline distances for purchase from other farms (black dots) by commercial (orange lines) and non-commercial farms (blue lines). In the scatter plot 4.4b correlation between farm size (log) and airline distances is shown. Non-commercial farms are represented by blue dots, commercial farms by orange dots.

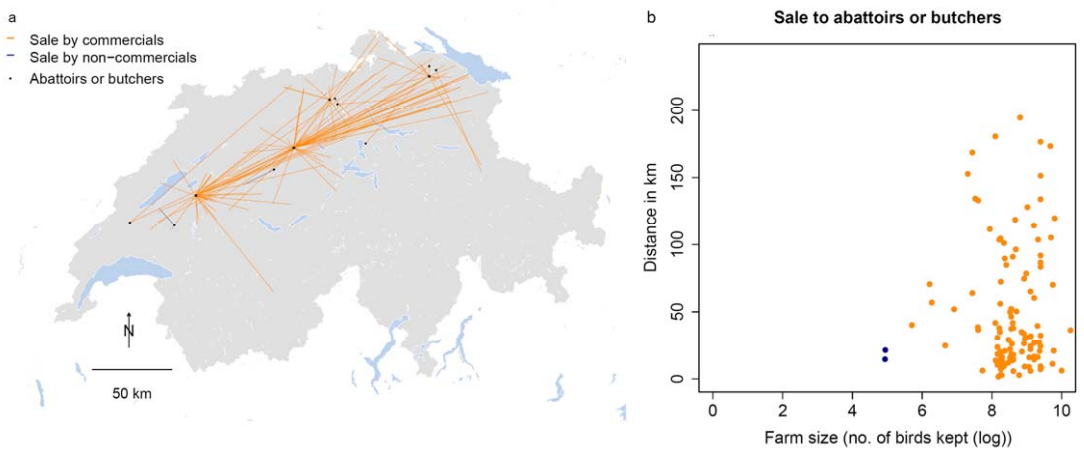


Figure 4.5: Poultry movements. Map 4.5a indicates airline distances for sales to abattoirs or butchers (black dots) by commercial (orange lines) and non-commercial farms (blue lines). No significant correlation between farm size (log) and airline distances was found (scatter plot 4.5b). Non-commercial farms are represented by blue dots, commercial farms by orange dots.

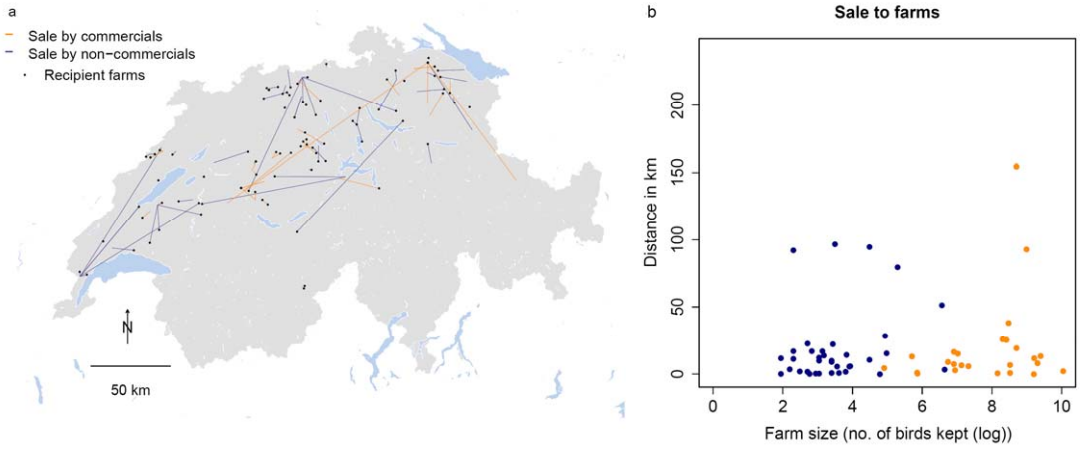


Figure 4.6: Poultry movements. Map 4.6a indicates airline distances for sales to other farms (black dots) by commercial (orange lines) and non-commercial farms (blue lines). No significant correlation between farm size (log) and airline distances was found (scatter plot 4.6b). Non-commercial farms are represented by blue dots, commercial farms by orange dots.

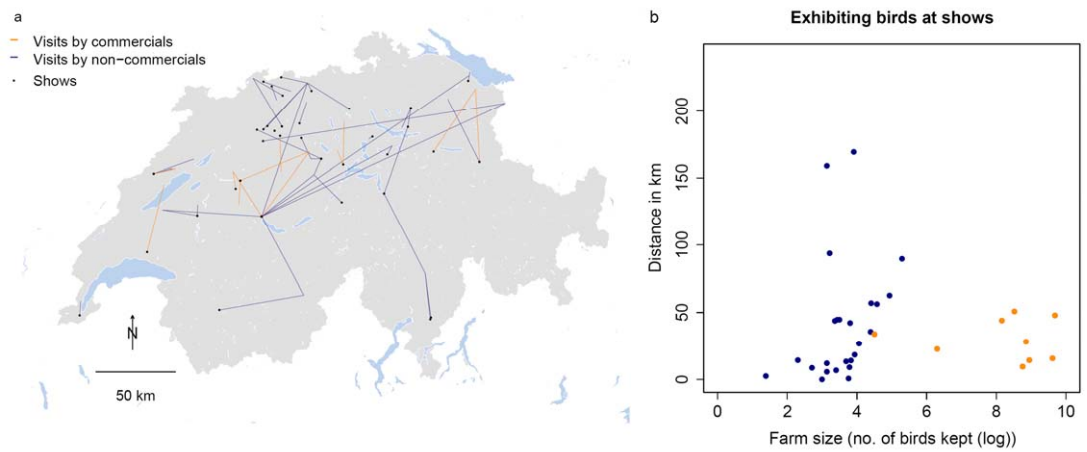


Figure 4.7: Poultry movements. Map 4.7a indicates airline distances for poultry show visits where own birds were exhibited (black dots) by commercial (orange lines) and non-commercial farms (blue lines). No significant correlation between farm size (log) and airline distances was found (scatter plot 4.7b). Non-commercial farms are represented by blue dots, commercial farms by orange dots.

4.4.4 Poultry movements across the farm groups

Commercial and non-commercial farms were directly connected by between-farm poultry movements. Out of a total of 767 specified purchases and sales between farms, 212 (28%) contacts were within the commercial farm group only, and 198 (26%) within the non-commercial farm group only. Across group contacts were mainly from commercial to non-commercial (347 (45%)) and 10 times (1%) from non-commercial to commercial farm types. Commercial to non-commercial contacts were mainly identified to be from grower and layer farms to backyard poultry farms. The experts from Companies III, IV, and V confirmed that some grower farms produced an excess of laying hens knowing the market opportunity to supply non-commercial farmers. Several layer farms were known to sell their hens, sorted out after one year of production, at low price to non-commercial farmers rather than disposing of them or supplying them to soup-hen production. Non-commercial to commercial farm contacts were attributed to several commercial farms keeping small flocks in a hen house separate from the commercial production although this was not recommended by the companies.

Further connections were found through the access to the same hatcheries in 4 cases (Figure 4.3a) and the same poultry shows in 2 cases (Figure 4.7a) by both commercial and the non-commercial farms. The “use of dead stock collection point”, the officially recommended practice for the disposal of dead livestock and pets, created a further link (although not through live poultry movement) as commercial and non-commercial farms share the same facilities.

4.4.5 Number of different contact partners

Only one contact partner per each origin (hatchery and other farm) and destination (hatchery, abattoir/butcher, and other farm) contact relation was found in most cases. Exceptions were observed in the few specialized farms. Grower farms supplied up to hundreds of commercial layer farms with laying hens. The experts confirmed that this distribution of the number of contact partners (degree distribution) was highly skewed and that the contacts were mostly stable over time.

4.4.6 Frequency of poultry movements

Movement frequencies were higher at commercial farms compared to non-commercial farms. Higher figures for commercial farms were explained by 6 to 8 transactions a year at broiler farms for purchase and sale, one purchase and sale by layer farms, and up to 80 purchases per year and daily sales by the few specialized farms (parents or grower farms cf. Box 4.1, or farms having more than one production level). Non-commercial farms had purchases and sales one time or less per year. If owned birds were exhibited at poultry shows, this was commonly done twice a year for both commercial and non-commercial farms. Both groups had outliers with 20 to 30 show attendances per year.

4.4.7 Data extrapolation to the entire poultry sector in Switzerland

Contact data were collected on a sample where the poultry keepers' probability of being selected was proportional to farm size, to ensure a sufficient number of the less numerous larger poultry farms. To provide contact estimates for the entire poultry sector, contact data were extrapolated on the entire poultry sector taking the sampling weight into account. Except for contact relations uncommon among non-commercial farms (such as sales to abattoirs/butchers), the extrapolated values were in the same range as in the non-commercial farm group (right column of Tables 4.2-4.5).

Estimates of the median number of poultry movements per month in Switzerland were calculated, ignoring seasonal variations of layer farms. Accordingly, 488 (95% CI = 443-538) purchases per month would be performed by farms with 500 or more birds kept (basically commercial farms), and 1686 (95% CI = 1665 – 1707), 3.5 (95% CI = 3.1 – 3.9) times more, by farms smaller than 500 birds (basically non-commercial) farms. Poultry sales would be in the same range with 1092 (95% CI = 880 – 1327) for large and 1018 (95% CI = 925 – 1116) for small farms. Poultry movements to poultry shows would be 45 (95% CI = 31 – 63) by large, and 655 (95% CI = 624 – 687), 14.6 (95% CI = 9.9 – 22.2) times more transactions, by small farms.

	Commercial	Non-commercial	All	Extrapolation to CH [c] poultry sector
Purchase (total)	n = 395	n = 405	n = 800	n = 800
times per year (m [a] [IQR [b]])	5 [1-7]	1 [0.5-1]	1 [0.75-5]	1 [0.5-1]
Sale (total)	n = 262	n = 62	n = 324	n = 324
times per year (m [a] [IQR [b]])	6 [2-7]	1 [0.5-2]	6 [2-7]	2 [0.5-3]
Poultry show (exhibiting birds)	n = 9	n = 27	n = 36	n = 36
times per year (m [a] [IQR [b]])	2 [1-10]	2 [1-3]	2 [1-3]	2 [1-3]

Table 4.5: Frequency of poultry movements in times per year by the commercial and non-commercial farm group and data extrapolation to the entire Swiss poultry sector. [a] median; [b] inter-quartile range; [c] Switzerland.

4.5 Discussion

We aimed at identifying between-farm contacts potentially allowing for highly pathogenic avian influenza (HPAI) to be spread between and amongst poultry farms in Switzerland.

At the completion of this study, countrywide density maps for both poultry farms and birds kept were produced for the first time for Switzerland. Both density maps provided complementary information. Bird density is an import factor to assess infection pressure. Farm density is relevant to HPAI control measurements such as the implementation of control and surveillance zones around farms. When only commercial poultry farms are included in farm density maps it might be concluded that areas such as south of Bellinzona in Switzerland have a very low farm density and thus are of minor importance for HPAI surveillance. In fact, the area south of Bellinzona is the most dense for poultry farms in Switzerland with more than 8 poultry farms per square kilometer when non-commercial farms are included in the dataset.

Our findings support the concept of 'farm neighborhood' as a potential contact in poultry farm population models. The two participant groups, poultry keepers with commercial (large) and non-commercial (small) farms were found to have equal neighborhood characteristics: (i) the number of other poultry farms in the neighborhood and (ii) the potential human and animal vectors such as cats and small birds (unpublished data) present on the farms. Free-range systems, facilitating vectors' access to domestic poultry and thus the risk of HPAI virus dissemination, were more common among non-commercial farms (92%) compared to commercial farms (61%) (unpublished data). Sharing employees within a neighborhood was, in contrast, more common among commercial (10%) compared to non-commercial farms (1%). This could increase the risk of HPAI virus dissemination amongst commercial farms, in the case where hygiene measurements are deficient.

The majority of farms were involved in human movements and shared resources (78%) and/or poultry movements (65%). The fraction was higher among commercial farms and distances were larger compared to non-commercial farms, except for those that exhibited birds at poultry shows. The number of different contact partners and poultry movement frequencies had skewed distributions. Few specialized grower and parent farms had high rates, mainly of outgoing contacts. The majority had low rates or no contacts at all. Highly connected farms are critical for a rapid spread of an epidemic [29]. These farms must therefore be well surveyed by veterinary authorities. The operating companies and producer must be particular vigilant at maintaining good farm hygiene management practices.

Poultry movement frequencies were higher at commercial farms compared to non-commercial farms. Estimates however for the entire population of poultry farms revealed 3.5 times higher chances of a poultry purchase, and 14.6 times higher chances of exhibiting birds at poultry shows occurring in a given time by a farm smaller 500 than birds (non-commercial farm) than by a larger (commercial) farm. This is because 95% of poultry farms in Switzerland keep less than 500 birds.

The common assumption of a closed circuit of the commercial poultry production without connections to non-commercial farms does not entirely hold true. Commercial and non-commercial farms were functionally connected through direct purchase and sale interactions (mainly from commercial to non-commercial), access to the same dead stock collection points and hatcheries, and visits of the same poultry shows.

The pattern of contacts between poultry farms has been investigated in terms of whether or not contact incidents were present. This was ignoring the strength of contacts (e.g. number of birds moved per transaction) and hygiene precautions taken by the poultry keepers. Contact partners were identified on a postal code level for data protection and the respondent's convenience resulting in only approximate air-line distances. Knowledge on effective transport routes may identify potential critical control points for remote contacts. We assume a slight under-reporting of contacts in the questionnaire: in follow-up interviews with 28 of the non-commercial respondents, it was sporadically explained that respondents had received birds as a gift that they had not declared in the postal questionnaire [181]. Interviews with experts from poultry industry indicated that commercial broiler producers do not always own the flock but raise birds on contract. This may explain why only 76.8% of the broiler subgroup indicated 'purchase of poultry/hatching eggs'.

There is a need to better understand why and under what conditions non-commercial keepers trade over long distances even though they have many other poultry farms in their direct neighborhood. Further the identified structural properties of the poultry sector must be complemented with data of biological factors sound predictions of outbreak dynamics. For instance, HPAI susceptibilities could be flock specific depending on virus strain and species kept, as described for the H7N7 outbreak in the Netherlands in 2003 [321].

Our findings have both local and global implications; for instance on zoning (geographical division) and compartmentalization (functional division by biosecurity measures). These are strategies introduced by the World Organization of Animal Health (OIE) to allow unaffected parts or segments of larger countries to continue trading during an epidemic [52]. Geographical and functional connections between commercial and non-commercial poultry farm subpopulations, as found in Switzerland, might also exist in larger countries. Geographical separations might be especially difficult to establish and maintain when poultry farm density is high over larger areas. Further, the present study helps to strengthen awareness for the importance of comprehensive and well organized epidemiological baseline data on the poultry population. The legislative basis for a mandatory notification of all poultry on a national level has been created in Switzerland [305]. The future national poultry register would, ideally, be entirely geo-referenced, maintained in a relational database format, and linked up with data on poultry movements and data on presence of wild birds and waterfowl as main reservoirs. Regarding other livestock species, movement databases for cattle have shown to capture spatio-temporal data in nearly real-time [284]. Such data support authorities in the timely prevention, surveillance and control of HPAI and any other poultry epidemic or zoonotic disease. Maps are a well-proven utility for combined presentations of data on agricultural, wildlife and ecosystem factors in preventive [91, 92] and post-outbreak investigations concerning HPAI [346].

As for models for HPAI transmission, the study results indicate that contact patterns are far from random given close neighborhood, farm type specific long distance contacts, and strong influence of the farms' affiliation to companies. To reflect the population's contact characteristics the combination of diffusion models (to reflect neighborhood contacts) and network models (to reflect long distance poultry movement contacts) as suggested by Truscott et al. [334] should be considered. However, this should not only be done for commercial farms but also for non-commercial farms.

Our findings indicate that both commercial and non-commercial farms are involved in neighborhood and remote between-farm contacts relevant to HPAI spread. It is necessary to include all poultry farms, irrespective of their size and purpose in both livestock registration and

disease surveillance systems, as well as in transmission models for poultry and zoonotic diseases.

4.6 Competing interests

The authors declare that they have no competing interests.

4.7 Authors' contributions

LF planned, conducted, and analyzed the quantitative and qualitative parts of the study, and drafted the manuscript as the lead writer. TS participated in the survey design, the analytical framework, and helped to draft the manuscript. JS planned and conducted the quantitative study part, and helped to draft the manuscript. JH provided support for data management, for the quantitative analyses, and for preparing the maps. JZ supervised the project, participated in the study design, and helped to draft the manuscript. All authors have read and approved the final manuscript.

5 Reconstructing the 2003/2004 H3N2 influenza epidemic in Switzerland with a spatially explicit, individual-based model

Smieszek T, Balmer M, Hattendorf J, Axhausen KW, Zinsstag J, Scholz RW. In prep.

5.1 Abstract

Background

Simulation models of influenza spread play an important role for pandemic preparedness. However, as the world has not faced a severe pandemic for decades, except for the rather mild H1N1 in 2009, pandemic influenza models are inherently hypothetical and validation is, thus, difficult. We aim at reconstructing a recent seasonal influenza epidemic that occurred in Switzerland and deem this to be a promising validation strategy for influenza spread models.

Methods

We present a spatially explicit, individual-based simulation model of influenza spread. The simulation model is based on (i) simulated human travel data, (ii) data on human contact patterns and (iii) empirical knowledge on the etiology of influenza. For model validation we compare the simulation outcomes with empirical knowledge regarding (i) the shape of the epidemic curve, overall infection rate and reproduction number, (ii) age-dependent infection rates and time of infection, and (iii) spatial patterns.

Results

The simulation model is capable of reproducing the shape of the 2003/2004 H3N2 epidemic curve of Switzerland and generates an overall infection rate (14.9%) and reproduction numbers (between 1.2 and 1.3), which are realistic for seasonal influenza epidemics. Age and spatial patterns observed in empirical data are also reflected by the model: The highest infection rates are in children between 5 and 14 and the disease spreads along the main transport axes from west to east.

Conclusions

We show that finding evidence for the validity of simulation models of influenza spread by challenging them with seasonal influenza outbreak data is possible and promising. Simulation models for pandemic spread gain more credibility if they are able to reproduce seasonal influenza outbreaks. For more robust modeling of seasonal influenza, serological data complementing sentinel information would be beneficial.

5.2 Background

Mathematical models and computer simulations of influenza spread have become increasingly important for pandemic preparedness within the last few years and have influenced the decisions of public health authorities [39, 104]. A non-systematic search in the common publication databases identified plenty of studies modeling the spread of (mostly pandemic) influenza outbreaks [11, 63, 72, 77, 86, 97, 110, 111, 250, 254]. However, models of pandemic spread are in most cases completely hypothetical because they focus on future pandemics [e.g. 72, 77, 86, 250, 254] and, thus, are not validated with empirical data. In contrast, some models of historical case examples explicitly address the model validation issue [e.g. 63]. As there is no alternative to prospective, hypothetical modeling for addressing scientifically potential future problems, new strategies for model validation are needed.

By validation we understand confirming that a certain model provides a good reproduction of the real-world behavior we are trying to simulate [132]. In agreement with the prevailing view in philosophy of science, we see model validation as a rather non-algorithmic, but argumentative process [25, 187]: The more real world data sets can be reproduced with a certain model and the more known characteristics can be reproduced with a model, the more reason we have to believe that the model is valid. Inherently, models forecasting future events cannot be checked against empirical data, but applying such a model successfully to past events provides some certainty that the model is valid *per se* and can be used for the comparative assessment of different scenarios including interventions.

This validation strategy has successfully been applied in various fields, such as climate research [216] or disease spread, where Carpenter and Sattenspiel [63] reconstructed meticulously the characteristics of the 1918 influenza outbreak in an indigenous Canadian community by means of an individual-based model. The problem when influenza pandemic models are challenged with data about past pandemics is that the last pandemics date back so far that they are only of limited value as mobility patterns and contact structures changed vastly and detailed data are scarce.

An alternative approach is to challenge simulation models with data from seasonal influenza outbreaks. Seasonal influenza outbreaks feature some particularities, which increase the system complexity and make them quite often resistant against attempts to reproduce them successfully in simulation models: Most pandemic models assume that there is no pre-existing host immunity due to the novelty of a pandemic strain [e.g. 110, 111, 254], but in case of seasonal influenza this cannot be ignored. Further, seasonal influenza epidemics are often characterized by several co-existent strains, which have to be treated as distinct diseases but which can interact in a complex manner at the same time. Nevertheless, challenging models with seasonal influenza data is a promising validation strategy because for seasonal influenza we have topical data.

In this paper, we present and describe a spatially explicit, individual-based model of influenza spread in Switzerland. This model makes use of disaggregated human travel data of all of Switzerland generated by the open source transport simulation software MATSim [1]. We further select data from the 2003/2004 H3N2 influenza epidemic in Switzerland and delineate why it is a good example for reconstruction and model validation. Our aim is to show that the simulation outcome is consistent with measured data and empirically based knowledge about the following aspects of seasonal influenza:

1) Epidemic curve, overall infection rate and reproduction number

2) Age-dependent infection rates and time of infection

3) Spatial patterns of influenza spread.

5.3 Methods and material

In this section, we first substantiate why we chose to model the 2003/2004 H3N2 influenza epidemic for model validation. We then describe the available data on this epidemic in Switzerland. Finally, we present all social and biological processes and assumptions which constitute our simulation model of influenza spread.

5.3.1 Choosing the influenza season to be reconstructed

Seasonal influenza epidemics are very complex in their dynamics and their effective drivers. On top, there are enormous uncertainties regarding the involved transmission processes, pre-existing host immunity, the proportions of asymptomatic, mild, moderate and severe courses of infection as well as the “true” infection rates. For a successful reconstruction of a preceding epidemic, certain preconditions should be fulfilled by this exemplary epidemic to achieve a meaningful and manageable model that can be validated.

1) *Only one dominant influenza strain in the period of interest*. Quite often we see two or more strains circulating simultaneously in a population. Having more than one strain at a time leads to an increase in system complexity that cannot be tackled in a simulation model: Two different strains must be modeled as two distinct diseases, but usually existing sentinel data only monitors on the basis of ILI symptoms and does not differentiate by strain. Therefore, there is no data on spatial patterns and age dependency of different strains. In Switzerland, from the eleven influenza seasons between 1995 and 2006, just three fulfilled the criterion of having a dominant strain which is responsible for almost all analyzed cases: 1997/1998; 1999/2000 and 2003/2004 (cf. Table 5.1).

2) *Clear empirical spatial patterns can serve as one indicator amongst others for the validity of a simulation model*. If there are observable spatial differences in the course of an influenza epidemic, these differences can be tried to be reproduced. If a simulation model reproduces observed spatial-temporal patterns (amongst other characteristics), then this is a further cue that the structural decisions and model assumptions underlying the simulation are appropriate. On the European continent, during several seasons a clear west-east-direction of influenza spread was observable [260]. Such a west-east trend can also be observed over several years within Switzerland. However, of all the years with a single dominant strain, this west-east trend was most pronounced in the 2003/2004 season.

3) *Inflows of infected persons from outside the system boundaries should not affect the internal spread dynamics in a relevant manner*. Every simulation model has system boundaries – typically national borders. In the case of Switzerland there are between 167000 (2003) and 213000 (2008) cross-border commuters [34], who travel daily from Germany, France, Italy or Austria to Switzerland and who are potentially an important source of imported influenza infections. The condition that the dynamics should be primarily internal is perfectly fulfilled by the 2003/2004 H3N2 epidemic. The epidemic took off in the westernmost part of Switzerland around Geneva during the 47th and 48th calendar weeks of 2003 (cf. Figure 5.1) – preceded by the westernmost countries in Europe (England, Scotland, Portugal and Spain had peaked during the 45th and 46th calendar weeks) – where it probably was introduced from France [10, 260]. Within Switzerland the epidemic moved toward the east within the next few weeks with a nationwide peak in the second week of 2004 (cf. Figures 5.1 and 5.2). Spatial dynamics seen within Switzerland might be mainly governed by internal processes because the other two neighboring countries of Switzerland with a long common border, Germany and Italy, experienced only an extremely mild and late influenza season. Germany had much fewer excess physician consultations than in the four preceding seasons and Baden-Württemberg (the German federal state neighbouring

Switzerland) peaked in the 9th week of 2004 when the epidemic in Switzerland was almost over [10]; Italy as a whole peaked in the 6th week [260].

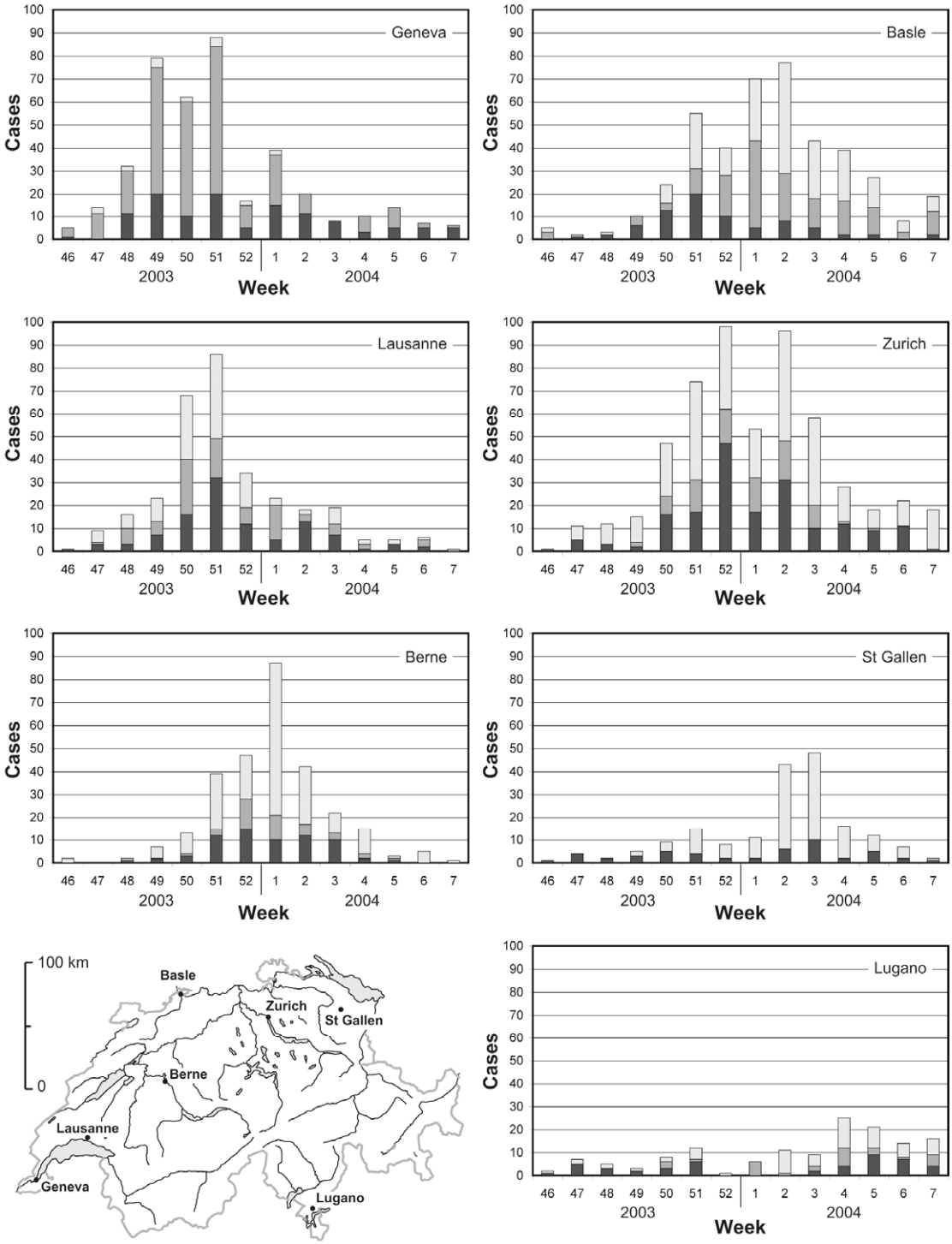


Figure 5.1: Reported cases of seven Swiss cities and their hinterland. The figure shows the reported cases of (i) seven Swiss cities (black bars); (ii) municipalities, whose centre is in a range of 7.5 km from the centre of the respective city (dark grey bars); (iii) municipalities with a centre in the range of 15 km (light grey bars).

Season	A H1N1		A H1N2		A H3N2		B		Total
	cases	*proportion*	*cases*	*proportion*	*cases*	*proportion*	*cases*	*proportion*	
1995/96	146	51%	0	0%	109	38%	30	11%	285
1996/97	2	1%	0	0%	234	68%	109	32%	345
1997/98	5	2%	0	0%	321	98%	0	0%	326
1998/99	0	0%	0	0%	83	37%	143	63%	226
1999/00	0	0%	0	0%	115	100%	0	0%	115
2000/01	110	89%	0	0%	1	1%	13	10%	124
2001/02	0	0%	1	0%	103	44%	130	56%	234
2002/03	1	1%	5	3%	125	68%	52	28%	183
2003/04	1	0%	0	0%	225	99%	2	1%	228
2004/05	35	12%	0	0%	225	75%	41	14%	301
2005/06	9	4%	0	0%	13	6%	183	89%	205

Table 5.1: Frequency of different influenza strains in analyzed Swiss samples. The data has kindly been provided by the Swiss Federal Office of Public Health.

5.3.2 Swiss influenza sentinel data

Approximately 3% (some 200; cf. Figure 5.2, orange line) of the general practitioners participate voluntarily in the Swiss sentinel system. They report the number of ILI diagnoses and the total number of consultations in their surgery on a weekly basis [54]. The nationwide incidence of influenza cases leading to a consultation is calculated by the Swiss Office of Public Health based on this sentinel data and the nationwide number of consultations known from the health insurance companies. The extrapolated number of consultations due to influenza for the 2003/2004 H3N2 epidemic in Switzerland is shown in Figure 5.2 (black bars).

The Swiss influenza sentinel data further allow coarse spatial analyses as every reported case is linked to a concrete practitioner and as for every practitioner it is known to which municipality she or he belongs.

5.3.3 Transmission pathways of influenza and contact definition

Known transmission pathways for human influenza viruses include direct physical contact, indirect physical contacts, and direct transmission via large droplets or indirectly by aerosols. No consensus has been achieved on the relative importance of these pathways under natural conditions. Some deem aerosols an important path [15, 140, 330] while others emphasize that close interaction is usually needed for transmission [47, 200]. For example, a past influenza outbreak on an airplane was best explained with aerosol transmission [234], whereas a recent H1N1 outbreak appeared to be caused solely by close, direct conversation [151]. It is also possible that the relative importance of the pathways depends on climatic conditions and virus strain specific characteristics.

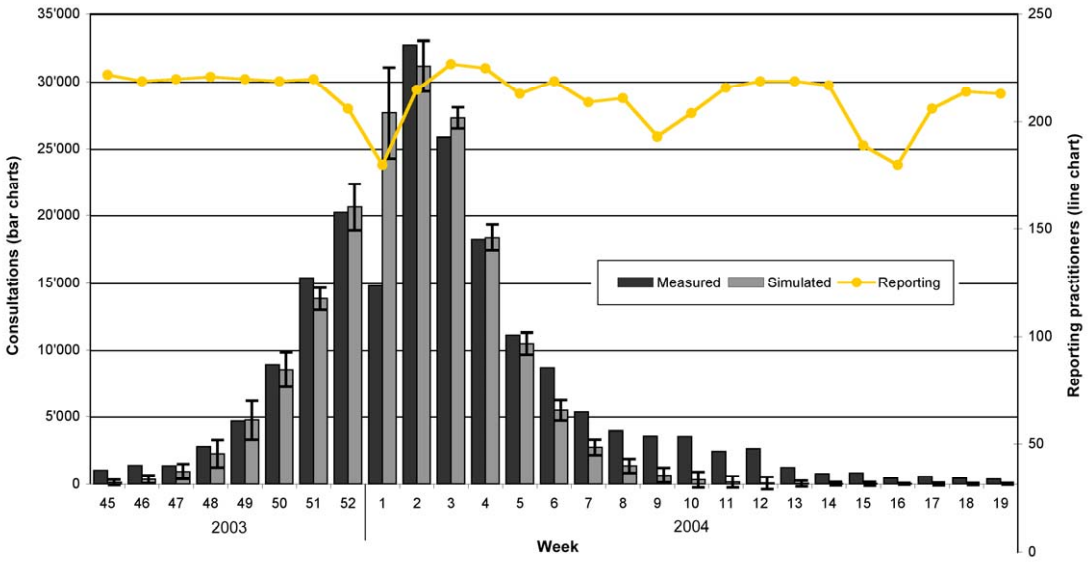


Figure 5.2: Epidemic curves and reporting practitioners. The black bars show the extrapolated reported cases coming from the Swiss sentinel system. The grey bars show the average simulated number of cases. The whiskers represent the standard deviation. The orange line stands for the number of reporting practitioners during the course of time.

The decision on which pathway of transmission a model should be based has far-reaching implications for the modeling strategy: aerosol transmission would be best represented by dynamic affiliation networks (individuals contaminate places; cf. [347]) and the transmission risk depends vastly on ventilation conditions and less on social interaction [202]. Transmission by large droplets and physical contact are best described by dynamic one-mode networks (individuals infect individuals directly, cf. [347]); the related risk of infection can be estimated based on intensity and duration of the respective contact [313].

Here, we subscribe to the assumption that transmission via large droplets and close contact are the dominant pathways of transmission and thus direct and close interaction between individuals is a precondition for infection in our model. Information on contact duration and intensity is not included in the current version of the model; instead, all contacts are weighed equally.

5.3.4 Spatio-social contact structure

Identifying sets of potential contact partners

Understanding and reproducing the spatial patterns of human movements is the common interest of spatially explicit epidemic models and transport simulation models. Hence, it is reasonable to link transport modeling and epidemiological modeling in a synergistic way as has previously been done e.g. for smallpox spread in the Portland region [101, 102] or for respiratory

diseases [272]. To identify individuals that possibly have contact with each other, we utilize a validated [20, 71] open source toolbox called MATSim [1], which was developed for implementing large-scale agent-based transport simulations. Details on MATSim are given elsewhere [1, 19, 21].

The wish or the need to perform certain activities is the main driver for human movement [21]. Individuals choose activities in different locations. That generates traffic and potentially infectious contacts [19]. Sequences of activities ("schedules") are generated and equilibrated in MATSim based on information about the characteristics of every individual of the (synthetic) population, i.e. their needs and preferences, and about the social and built environment, e.g., opening hours (cf. Table 5.2) [19, 21]. The MATSim simulation process has three steps: (i) scenario creation (establishing a synthetic population [21], the transport network [341] and the set of locations with their corresponding capacities and open times), (ii) initial demand modeling (creating an initial schedule for every individual based on given data [107]) and (iii) demand equilibration (allowing individuals to iteratively re-decide on their schedules for achieving more realistic results) [21]. Compared to state-of-practice transport modeling processes [256], MATSim comes up with an integrated assignment and demand equilibration completely based on time-dynamic schedules for every member of the synthetic population. These individual schedules are saved in an XML-file (an excerpt of such a file is shown in Figure 5.3) that can be used for further analysis or as input data for other simulations – like in the epidemic model presented in this paper.

Source	Description and information used
Volkszählung 2000	The Swiss census is surveyed once a decade and includes the entire legal resident popu-ation (~7.2 million in 2000). It links inhabitants to concrete buildings and houseolds and provides information about demographic characteristics, place(s) of residence, place of work, etc.
Mikrozensus Verkehr 2005	The "Mikrozensus Verkehr" is a travel behavior survey carried out by the Swiss Federal Statistical Office every five years. In 2005 a sample of 33390 inhabitants participated and answered questions about their mobility behavior. In particular, they filled in a travel and activity diary.
Betriebszählung 2000	The enterprise census ("Betriebszählung") records data on all public or private enterprises in Switzerland. There are ~398000 workplaces in Switzerland and for all of them the exact location, the sector they belong to and the number of employees is known.
Nationales Netzmodell	The national transportation planning network consists of ~24000 nodes and 60000 links of the Swiss transportation grid.

Table 5.2: Sources of information used for the Swiss MATSim scenario.

```
<person id="1020708" sex="m" age="57" license="yes" car_avail="always" employed="yes">
    ...
    <plan score="164.3011341159086" selected="yes">
        <act type="home" link="119428" facility="133427" x="692210" y="286910"
         start_time="00:00:00" dur="07:19:58" end_time="07:19:58" />
        <leg mode="bike" dep_time="07:19:58" trav_time="00:15:13" arr_time="07:35:11">
        </leg>
        <act type="work_sector3" link="110071" facility="10278517" x="689048" y="284557"
         start_time="07:35:11" dur="02:32:59" end_time="10:08:10" />
        <leg mode="bike" dep_time="10:08:10" trav_time="00:15:13" arr_time="10:23:23">
        </leg>
        <act type="home" link="119428" facility="133427" x="692210" y="286910"
         start_time="10:23:23" dur="00:51:05" end_time="11:14:28" />
        <leg mode="bike" dep_time="11:14:28" trav_time="00:15:13" arr_time="11:29:41">
        </leg>
        <act type="work_sector3" link="110071" facility="10278517" x="689048" y="284557"
         start_time="11:29:41" dur="07:05:25" end_time="18:35:06" />
        <leg mode="bike" dep_time="18:35:06" trav_time="00:13:58" arr_time="18:49:04">
        </leg>
        <act type="shop" link="110070" facility="10286070" x="692047" y="286350"
         start_time="18:49:04" dur="00:54:59" end_time="19:44:03" />
        <leg mode="bike" dep_time="19:44:03" trav_time="00:00:07" arr_time="19:44:10">
        </leg>
        <act type="leisure" link="110070" facility="10286078" x="692022" y="286369"
         start_time="19:44:10" dur="04:06:55" end_time="23:51:05" />
        <leg mode="bike" dep_time="23:51:05" trav_time="00:01:46" arr_time="23:52:51">
        </leg>
        <act type="home" link="119428" facility="133427" x="692210" y="286910"
         start_time="23:52:51" />
    </plan>
</person>
```

Figure 5.3: MATSim output – Sequence of activities. This figure shows an excerpt of the XML data structure generated by MATSim. The highlighted information is used in the here presented influenza model: id refers to an existing unique identifier in the Swiss census; age gives the age in years of the respective agent; every activity has a type – we distinguish home, work, education, shop and leisure; x and y are coordinates and refer to the Swiss military grid.

Based on this information, we can identify individuals who share the same location and thus have a chance to infect each other. In the current version of our model we map all individuals on a geographic grid with a side length of 500 by 500 meters. In principle, the set of potential contact partners for each individual consists of those people who perform the same activity within the same grid cell. However, for some activity categories, we further narrowed down these sets as follows:

Home contacts: We assume that every individual has daily contact with every other individual who lives in the same household. Information on the exact household compositions is taken from the Swiss census (cf. Table 5.2). Contacts at home beyond the regular cohabitants are not included in the simulation model.

Work contacts: All individuals who work within a certain grid cell are allocated to work groups. The targeted work group sizes are assumed to follow a normal distribution with a mean of ten work group members and a standard deviation of two (values rounded to the nearest integer value). Every individual can only belong to one work group per grid cell. If there are not enough individuals in a grid cell to form a workgroup of the targeted size, they form a workgroup of the

maximally possible size. All individuals within the same work group as well as the whole set of working individuals within the same grid cell are eligible as contact partners according to the selection rules explained subsequently.

Educational contacts: All individuals who perform any kind of education activity within a certain grid cell are allocated to classes. Based on the data published in the education statistics of the canton Zurich [32], we assume that 20 students form a school class. Up to the age of 18, only students of the same age can form a class. Educational activities of adults include everything from university to evening classes. For these individuals, we also form classes of 20 – but these classes are not stratified by age. All individuals within the class as well as the whole set of individuals within the same grid cell performing education activities are eligible as contact partners according to the selection rules explained next.

Contacts during shopping or leisure time: For both the shopping and leisure categories, we assume no further structure: All individuals who perform either shopping or leisure activities within a certain spatial cell are eligible as contact partners for another individual with the respective activity within this cell.

Realizing contacts

Relying on the data from the Swiss MATSim scenario, we were able to identify sets of potential contact partners for every individual. However, people do not have contact every day with all persons they could possibly have contact with: Students do not have contact with the same classmates every day and employees have meetings with different co-workers each day. Thus, only a subset of the potential contact partners will be chosen as realized contacts during one simulation time.

To do so, we rely on age-dependent distributions of potentially contagious contacts measured in ten European countries [237]. We calculate the mean number of household contacts (all other household members) for each age group and subtract it from the empirical means [237], which results in the remaining average number of contacts to be achieved during the daily non-home activities. By calculating how many distinct activities are in average performed by the members of a certain age group, we can figure out how many contacts have to be made in average per activity in order to reach the contact number targets. We assume that the contact distribution per activity and age group follows a negative binomial distribution [237]. Details are given in Table 5.3.

Age group	Empirical mean [237]	Mean household contacts in CH	Mean activities per person	Mean contacts per activity [SD] [a]	Median contacts per activity [IQR] [a]
0-4	10.21	2.95	1.29	5.62 [4.0]	5 [3-8]
5-9	14.81	3.23	1.52	7.6 [5.2]	7 [4-10]
10-14	18.22	3.23	1.54	9.75 [6.4]	9 [5-13]
15-19	17.58	2.98	2.29	6.38 [4.5]	6 [3-9]
20-29	13.57	2.12	2.06	5.56 [4.0]	5 [3-8]
30-39	14.14	2.40	1.91	6.14 [4.3]	5 [3-8]
40-49	13.83	2.52	1.90	5.95 [4.2]	5 [3-8]
50-59	12.30	1.63	1.87	5.70 [4.1]	5 [3-8]
60-69	9.21	1.32	1.72	4.58 [3.4]	4 [2-6]
70+	6.89	1.39	1.82	3.02 [2.5]	3 [1-4]

Table 5.3: Contact data. [a] We assume that the number of contacts per activity follows a negative binomial distribution.

Based on the so-defined probability distributions, we allocate a target number of contact partners to each distinct combination of a person performing a certain activity at a certain location. The allocation is a random process taking place at every simulation time step and is completely independent from former allocations. To note, individuals, who perform many activities meet on average more other people than individuals with only a few activities. With the Swiss MATSim scenario we achieve in total an average number of contacts per day of 12.4 (SD=8.9; mode=7), which is plausible [237].

The target numbers of contact partners allocated as described are like stubs which have to be connected in order to become contacts. We connect stubs on demand, i.e., we look only for contact partners for those individuals who are infectious. If all of a certain individual's stubs are connected to other individuals' stubs, this individual is no longer available as a contact partner.

In principle, contact partners are chosen from the set of other individuals performing the same activity within the same grid cell. In addition, for two out of four non-home activities, fixed subgroups have been defined: Employees are allocated to work groups and students are allocated to classes. We adopt the assumption of Ferguson et al. and assume that with a probability of 0.75 the counterpart for an unconnected stub comes from the work group or the class, respectively, whereas the others are chosen from the total set on the grid cell level (cf. p. 8 of the supplementary material of [110]). As we know that more contacts are made with people of the same age than with people who are considerably younger or older [94, 237], we introduced a slight age-dependent bias for the partner selection out of the total set: The relative probability weight of being picked is $P = \exp(c \cdot \Delta_{age})$, where Δ_{age} is the age difference (in years) of two potential contact partners and c is a shape parameter modulating how sensitive the selection process is for age differences. We arbitrarily assume $c = -0.02$.

Approximately 45% of all contacts are made within the defined subgroups (household, classes and work groups) – which appears to us to be realistic [314]. If there are no stubs left within the group or if all individuals within the group have already established contacts with the individual to be connected, the contact partner will be automatically chosen from the total set of potential partners. If there are not enough partners left in the total set, we allow connecting to individuals from adjacent grid cells (von Neumann neighborhood). This exemption occurs in less than 2 per mil of all individuals to be connected.

5.3.5 Pre-existing immunity, acquired immunity and vaccination

Modeling seasonal influenza, in contrast to pandemic influenza, must account for patterns of pre-existing partial or full immunity. For Switzerland no such measures like the pre-episode hemagglutination-inhibiting (HI) antibody titers (cf. Table 5 of [116]) are available. Therefore, we assume here that individuals older than 20 years have a halved transmission risk [323], i.e., we multiply the transmission probability for every contact between an infector and a susceptible individual older than 20 years by 0.5.

A part of the Swiss population gets vaccinated against the dominant seasonal influenza strains every year. Several sources report vaccination efficacy values between 70 and 90% [198, 246, 323]. In this paper, we assume vaccination to reduce the per contact transmission probability by 80% for all age groups. We had no data about the age distribution of vaccinated people in Switzerland for 2003/2004, but as the overall vaccination rates (and vaccination strategies) are similar in the neighboring countries [326], we took the age distribution of Germany [165] and adjusted it proportionally to achieve approximately 1.31 million doses – a realistic number for Switzerland as a whole [53].

After recovery, we assume individuals to be fully immune for the rest of the simulation run, i.e., they cannot be re-infected.

5.3.6 Severity of infection and shedding

Only a minor proportion of all seasonal influenza infections leads to such severe illness that a practitioner's advice – or even hospitalization – is needed. The yearly infection rates reported in the literature range between 5 and 30% (most of them between 10 and 20%) [68, 198, 214, 228, 293], but only approximately 2-3% of the Swiss population seeks out a general practitioner due to an ILI during the influenza season every year.

Influenza occurs with varying severity and can be even completely asymptomatic [371]. The exact proportion of asymptomatic infection and of the different levels of severity is unknown. Modeling studies used proportions of asymptomatic infection between 30 and 50% [72, 110, 111,

127, 208, 223]. Wright et al. reported an asymptomatic infection rate of 37% in children [369]. Carrat et al. re-analyzed volunteer challenge studies and found that the proportion of infected individuals showing any kind of symptoms was 66.9% (CI: 58.3, 74.5) – however, only 34.9% (CI: 26.7, 44.2) reported fever [64]. Fox et al. [116] showed that the proportions of different levels of severity depend on age and pre-episode hemagglutination-inhibiting (HI) antibody titer.

In addition, the relative infectiousness of asymptomatic, mild, moderate and severe cases is not known. The infectiousness can be represented as a function of the viral shedding [313], quite often simplified as a linear relation [64, 110]. There is only little known about the shedding behavior of asymptomatic cases (and, thus, about their infectiousness) [64, 371]. However, in two studies a positive correlation was found between the quantity of viruses per positive specimen and the reported severity of illness [33, 74]. For the case of the 1918/19 pandemic influenza in Geneva, Chowell et al. estimated the relative infectiousness of asymptomatic cases to be negligibly small [66]. Ferguson et al. assume severe cases to be twice as infectious as mild cases based on household data [110, 111]. Eichner et al. assume asymptomatic cases to be half as infectious as moderate or severe cases [96]. The empirical work of Couch et al. showed that the mean quantity of viruses detected in specimens of nasal wash was only approximately half in cases of asymptomatic infection compared to symptomatic infection (cf. Figure 1 of [74]).

For this paper we decided to distinguish two groups of illness severity: We assume that 50% of the infections end up in moderate to severe illness (subsequently referred to as symptomatic cases) and that 50% of the infections are asymptomatic or very mild cases (referred to as asymptomatic cases). We further assume that the latter group is half as infectious as the first group. Although in reality there is an interaction between age, pre-existing immunity and illness severity, we simplify here and assume illness severity to be a random outcome completely independent of the other two factors.

5.3.7 Incubation period and infectiousness over time

Anderson and May specify the incubation period of influenza as between one and three days [8]. Lange and Vogel deem the onset of symptoms to be 24 hours after inoculation (cf. Figure 3.2 in [198]). Most empirical studies report the onset of symptoms to be within the first one or two days after inoculation [64, 110]. In our simulation model the step width is one day, i.e., we can implement the course of a disease with such precision. We assume here – simplifying reality – that virus shedding always starts the day after inoculation.

Various numbers for the infectious period can be found in literature: Anderson and May assume the infectious period to be two to three days [8]. Wearing et al. use for their model of an influenza outbreak in an English boarding school infectious periods of 2.1 to 2.2 days (depending

on the make of the model) [349]. Eichner et al. implement a "fully contagious period" of 4.1 days for asymptomatic and moderately sick adults and of 7.0 days for all other groups [96]. However, most simulation models assume the infectiousness to be constant during the infectious period. We describe the infectiousness over time by a function estimated by Ferguson et al. [111] based on shedding data: They assume the infectiousness over time to follow a lognormal function $\kappa(T,\delta,\gamma)$ with $\delta[\log(d)]=-0.72$ and $\gamma[\log(d)]=-1.8$ (for details cf. to pp. 10 and Table SI1 of the supplementary material of [111]). We truncated this function after seven days for both the symptomatic and the asymptomatic group.

The total infectiousness (over the seven days of infection) can be varied by multiplying the values for every single day by a constant factor.

5.3.8 Initial seed

Defining the initial seed is a particularly critical modeling decision and complicated task. Regarding the seasonality of influenza, it is not entirely clear to what degree influenza viruses outlive the summer between two influenza seasons and to what degree they are reintroduced from other world regions [198]. Usually, seasonal influenza is also prevalent in several regions within Europe at the same time – and therefore multiple ports of entry are possible. To complicate matters further, we have to face the fact that the geospatial information coming from sentinel systems is the worst at the very beginning of an epidemic: In contrast to the epidemic phase during which the positive predictive value of practitioners' diagnoses is around four-fifths [37, 227, 372], the positive predictive value in absence of a known epidemic is around one-third in various studies and different age groups [136, 198, 264]. At the same time, random variation is known to be particularly decisive for the further spread during the early phase of an outbreak [249].

In this framework, we decided to take the sentinel report of the 49th week as the starting point for the initialization. For 293 cases, an agent with the same age and residential municipality was randomly identified in the synthetic population and symptomatically infected. The actual status in the course of infectiousness was attributed randomly using a uniform probability distribution.

The chosen approach combines various requirements: (i) The chosen week is the first sentinel week with clear signs of an outbreak in the westernmost parts of Switzerland while the rest of Switzerland is still pre-epidemic. Such a "headstart" of the Geneva region is necessary to be able to reproduce the spatial patterns observed. At the same time early cases in the rest of Switzerland are realistically captured by this initialization. (ii) Although the age distribution of the cases reported within the sentinel system is biased (the help seeking behavior varies across

age groups), we can assume that the reported cases reflect the real age distribution better than, e.g., random choice. (iii) 293 initial cases are negligible compared to the total population size and, thus, still a reasonable starting point for model initialization.

5.3.9 Model calibration

What remains to be calibrated is the overall infectiousness of the disease and the proportion of symptomatic cases which is diagnosed with influenza by a practitioner. We assume the latter proportion to be constant over the entire epidemic season. We further assume perfect diagnosis accuracy, i.e., 100% sensitivity and specificity, and that the extrapolated sentinel data can directly be compared with the respective proportion of the symptomatic cases coming from the simulation runs. We further compare only weeks 47 to 52 (2003) and 2 to 15 (2004). The first week of 2004 is excluded from the analysis and calibration procedure as we do not know to what degree the observed drop in cases reflects a real decline and to what degree this is a simple measurement error (cf. to the discussion below). In order to compare the weekly sentinel data with the daily simulation data, the latter also had to be converted into weekly data. We did this with floating week boundaries, i.e., the definition of which simulation days should be the beginning of a report week was flexible and chosen in a way to achieve maximum fit with the measured data. The peak of the epidemic was assumed to be in the second week of 2004.

We started our calibration procedure with an arbitrarily chosen mean (i.e., averaged over the seven days an individual stays infectious) per-contact infection probability for symptomatic cases of $1.129 \cdot 10^{-2}$. We compared the fit (least squares) of this parameter setting with three higher and three lower values for the mean per-contact infection probability. The increment was $5.65 \cdot 10^{-4}$ (i.e., 5% of the start value). The two parameter values that fit best were the start value and the next higher value (which also fit far better than the start value). We then analyzed the parameter range between the two best values starting from the higher value and approaching the lower with an increment of $1.6 \cdot 10^{-4}$. The best accordance with the empirical epidemic curve was achieved for a mean per-contact infection probability of $1.1693 \cdot 10^{-2}$. The proportion of symptomatic cases that has to be diagnosed by a practitioner to reproduce the extrapolated empirical curve best (least squares) is approximately one-third.

5.4 Results and discussion

5.4.1 Epidemic curve, overall infection rate and reproduction number

Epidemic curve: The extrapolated epidemic curve of the 2003/2004 H3N2 epidemic and the corresponding average values and standard deviations of 30 simulation runs are shown in Figure 5.2. The weeks of the year 2003 and weeks two through five of 2004 are reproduced quite well by the model. The model overshoots the extrapolated value of the first week of 2004 (began on 29 December 2003) a lot and it undershoots the extrapolated sentinel data of weeks six to twelve systematically.

There are several explanations for the difference in the first week of 2004, including:

(i) This week is part of the Swiss Christmas vacation and children do not go to school. Children are seen as a major pacemaker for the spread of seasonal influenza [157] and school is one of the most important places to make contact with other children [133, 158]. A study conducted by Cauchemez et al. showed a relation between school vacation and a decrease in infection rates in France [65]; Heymann et al. were able to show similar results for the 1999/2000 influenza epidemic in Israel when a teachers' strike broke out during the influenza season [161]. However, another study conducted by Rodriguez et al. failed to show such an association in a survey about absenteeism in King County (Washington) public schools [285]. Hence, there might be a real drop in cases during the Swiss Christmas vacations due to school closure, but such an effect is neither proven nor quantified.

(ii) It is not only plausible to expect fewer infections during school vacations, but also to expect fewer consultations: Many patients with influenza-like symptoms normally do not seek out the practitioner primarily because they want treatment, but for a sick note for school or the employer. As many offices and all schools were closed in the first week of 2004, many sick people also did not need such a medical certificate. Consequently, most likely a smaller proportion of the symptomatic cases sought out a practitioner than usual.

It is further not clear what causality is behind the undershooting during the tapering phase of the epidemic curve. Plausible explanations are:

(i) The epidemic peaked quite late in the German and Italian neighborhood of Switzerland. A possible explanation for the comparatively slow decay in the Swiss sentinel data might be that new sources of infection were constantly introduced from the neighborhood and prevented the disease from dying out earlier. The spatial distribution of the reported cases at the end of the 2003/2004 epidemic appears to support this interpretation as the focal points of infection seem

to be at the German and Italian borders. However, such transborder processes are not reflected in our model.

(ii) We have seen that the diagnosis accuracy is good during the peak phase, but comparatively poor when influenza is not abundant. Hence, it is likely that sentinel extrapolations overestimate the number of new cases during the early and the late phase of an epidemic. Particularly at the end of an outbreak after a phase when influenza diagnoses were quite common, practitioners might be prone to diagnose influenza too inattentively.

(iii) We assume the infectiousness of the infective agent to be constant over time. However, influenza is a quickly mutating virus and we have evidence that the risk of infection is partly modulated by, e.g., climatic conditions [85, 374]. If – for whatever reason – there was a rise in infectiousness of the 2003/2004 H3N2 strain, this could also explain why the extrapolated measurements decayed more slowly than the simulated values.

Overall infection rate: The simulated overall infection rate is 14.9%. There is no serological data for the 2003/2004 H3N2 epidemic in Switzerland we could compare this simulated value to. However, various sources state that between 5 and 20% [68] or between 10 and 20% [198] of the population in a Western context become sick with influenza every year, which is the range of our results.

Reproduction number: In case of seasonal influenza it makes more sense to refer to the effective reproduction number than to the basic reproduction number, because the infectious agent never hits a fully susceptible population and, thus, the basic reproduction number is only of theoretical value. We show the time-development of the reproduction number of our simulation runs in Figure 5.4. When the disease hits the unaffected population during the early phase of the epidemic, reproduction numbers range between 1.2 and 1.3. This is in accordance with the results reported in studies on influenza spread: Chowell et al. analyzed influenza seasons in the USA, France and Australia from 1972 to 1997 and found for all three countries mean reproduction numbers of 1.3 (CI 1.2-1.4) [68]. Nishiura et al. report maximum likelihood estimates for the reproduction number of the 2009 H1N1 epidemic in Japan between 1.21 and 1.35 [249].

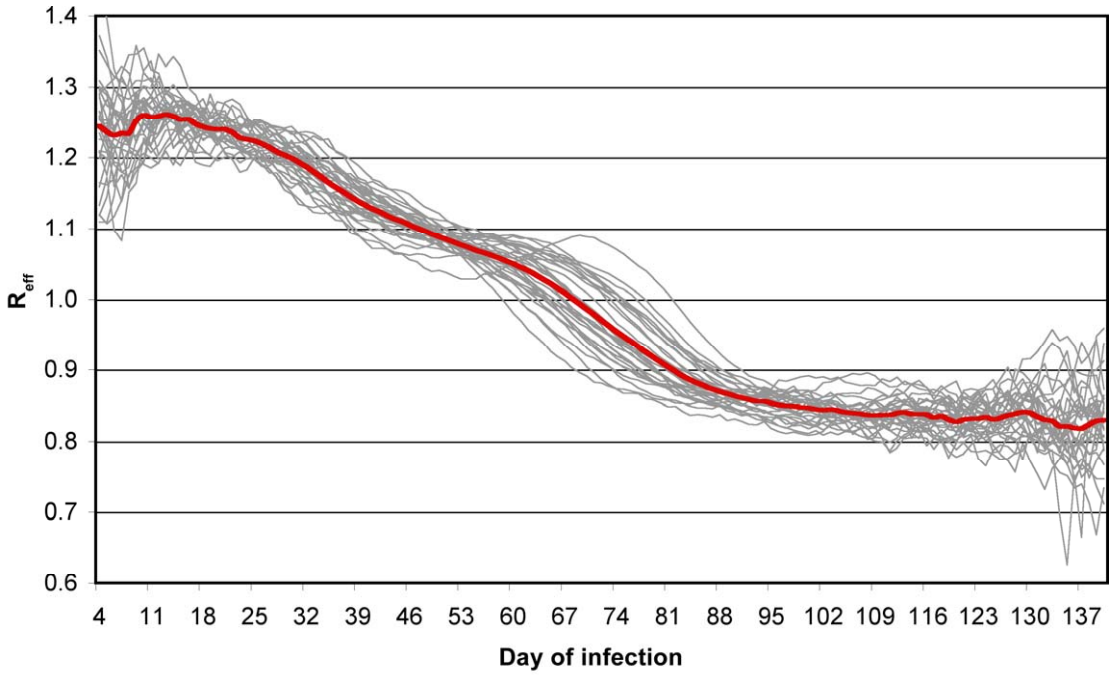


Figure 5.4: Reproduction number versus time of infection. The grey lines show the development of the reproduction number (defined as the average number of secondary cases) of each individual simulation run. The red line is the average of all runs. The abscissa represents the time-point (in days) when an infector started shedding.

5.4.2 Age-dependent infection rates and time of infection

Infection rates: Figure 5.5 shows the simulated mean age-dependent infection rates. Children (all persons younger than 20) clearly have the highest simulated infection rates with 33.0%, whereas adults only show attack rates of 9.5%. These simulated results are in accordance with general empirical data on seasonal influenza outbreaks: Several sources state that children are considerably more affected than adults [198, 214]. Nicholson et al. estimate that approximately 20% of children and 5% of adults worldwide develop symptomatic influenza every year [246]. Sauerbrei et al. see the annual infection rates of children in a range between 20 and 30% [293].

Figures 5.5a and 5.5b further show serologic age-dependent infection rates measured for five influenza seasons in Tecumseh, Michigan, [229] and Seattle, Washington [117]. Although there is a pronounced variability in infection rates, clear patterns are visible in the serologic data, which are also reproduced by the model in a qualitative sense: (i) In most cases, the infection rates of children are twice to three times as high as those of adults. (ii) Very young children and adolescents (almost adult individuals) show lower infection rates than children between five and 14 years of age.

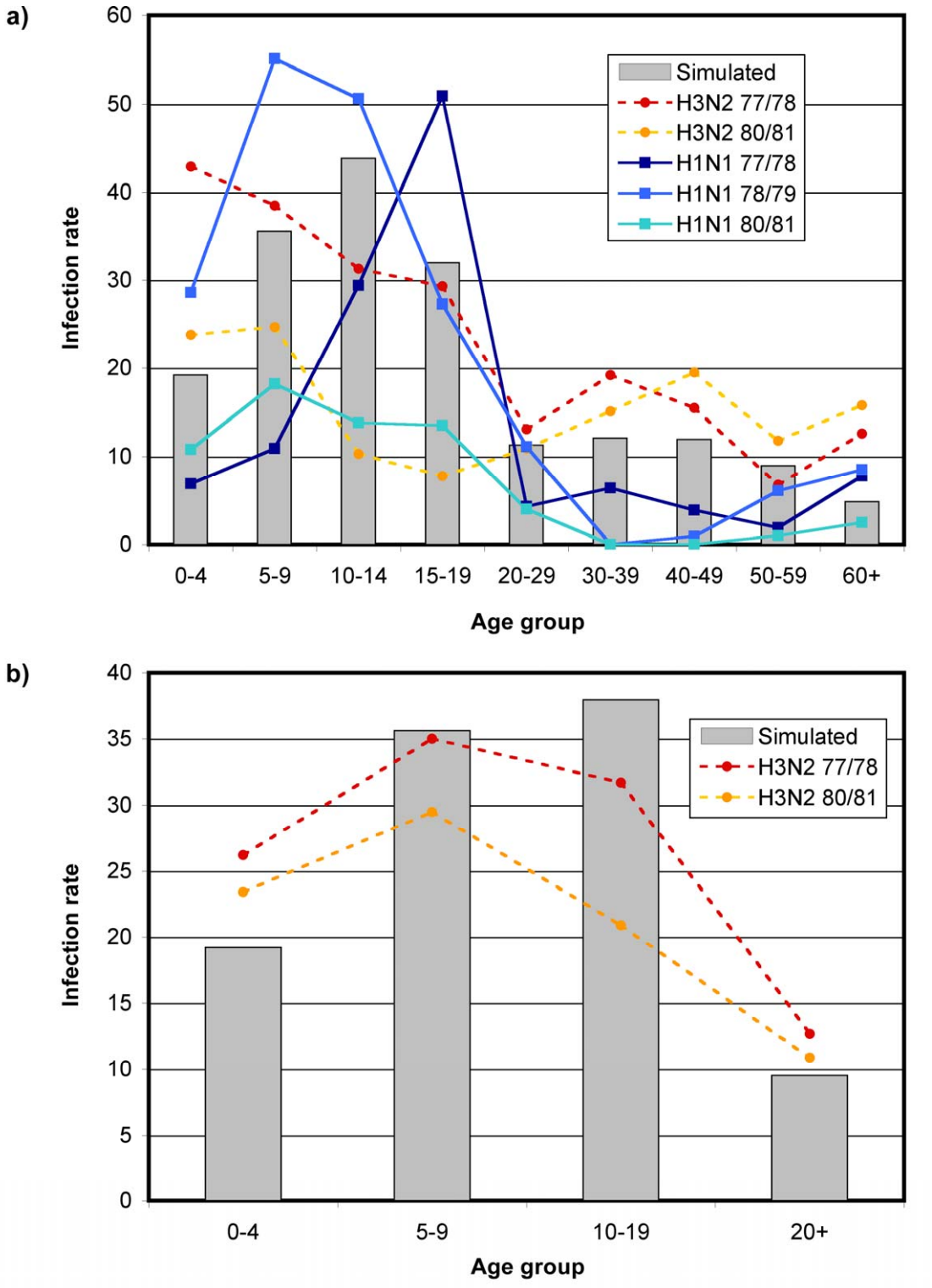


Figure 5.5: Age-dependent infection rates. Subfigure a: The lines show the serological age-dependent infection rates measured for five influenza seasons in Tecumseh, Michigan [229]; the bars show the corresponding average infection rates of our simulation model. Subfigure b: The lines show the serological age-dependent infection rates measures for two influenza seasons in Seattle, Washington [117]; the bars show the corresponding simulated infection rates.

The patterns observed in the serologic data as well as in the simulation results can be explained by the following mechanisms: (i) Practically every child over twelve years of age has been challenged with influenza during their lifetime [293]. Every exposure to influenza confers specific immunity against the respective strain and partial immunity against kindred strains. We have implemented this mechanism by assuming that adulthood is 50% protective. (ii) Differences in the mean number of contact partners can also partly explain the observed differences. People with many contacts have more chances of infecting others or of becoming infected than people with few contacts. Particularly, the differences within the age group of children can be partly attributed to differences in the number of contact partners (cf. Table 5.3 and Figures 5.5a and 5.5b). (iii) The actual arrangement of contacts also determines infection rates. It is known, for instance, that children tend to have most of their contacts with other children of the same age. If a certain age group has on average more contacts than another, such age assortativeness means that individuals with many contacts preferentially meet other individuals with many contacts, which results in a further elevated risk of infection.

Time of infection: If we relate the average age of infected individuals to their time of infection, we see a clear increase in average age in both sentinel and simulated data. We calculated linear regression models with point in time as the independent variable and average age of reported cases as the dependent variable.

For the sentinel data, we observe an increase in the average age of reported cases of approx. 0.32 years every day, if we look only at the swelling phase of the epidemic between week 49 in 2003 and week two in 2004. After the peak in week two has been passed, the average age of the reported cases decreases again. However, for the whole epidemic phase (from week 48 in 2003 to week five in 2004), there is still an increase in the average age of 0.10 years each day. In case of the simulated data, we calculate a median slope of 0.081 years of age per day (IQR 0.076 – 0.092) for the simulation time range between days 30 and 100.

These increases in the average age of the cases can be explained by the hypothesized pacemaker function of children. Theoretical modelling studies showed that highly connected persons get infected earlier and more often and play a more important role as infectors than rather isolated persons [22, 29]. We know from literature [237] and have implemented in our model that children have on average more contact partners per day than adults. The increase in average age by time we see analogously in the sentinel and simulated data is most likely caused by this mechanism: First, the susceptible and highly connected children of a region get infected and then, the partly-immune and less connected adults follow.

5.4.3 Spatial spread of the influenza epidemic

The spatial dynamics of the 2003/2004 H3N2 epidemic are visible in Figure 5.1, which shows the epidemic curves of seven Swiss cities and their hinterland based on the sentinel data. It is clearly observable that the epidemic broke out in the Geneva region, which is the westernmost part of Switzerland. It moved from there over Lausanne and Berne toward the north-eastern parts of Switzerland with the cities of Zurich and St Gallen. At last, it reached the canton of Ticino in the south-east of Switzerland, here represented by the city of Lugano.

Table 5.4 provides the median duration from the peak in one city to the peak in another city as well as the inter-quartile range as shown in the 30 simulation runs. The median durations of the 30 simulation runs are consistent with the time lags that can be observed in the measured data. Most of the simulation runs also showed a clear movement from the west to the east with Ticino having the latest of all peaks. One exemplary run is shown in Figure 5.6.

	Geneva	Lausanne	Berne	Basle	Zurich	Lugano
St Gallen	4.0 [3.1-4.3]	2.1 [1.3-2.7]	0.4 [-0.4-1.5]	-0.5 [-1.3-0.1]	0.0 [-0.6-0.6]	-2.0 [-3.5- -0.9]
Lugano	5.8 [4.8-7.1]	4.5 [3.6-5.6]	2.9 [1.7-3.9]	1.8 [0.2-3.0]	2.4 [1.1-3.0]	
Zurich	3.9 [3.0-4.6]	2.2 [1.2-3.0]	0.5 [-0.1-1.4]	-0.7 [-1.3-0.3]		
Basle	4.6 [3.0-5.4]	2.7 [1.6-3.7]	1.1 [0.1-2.1]			
Berne	3.1 [2.4-4.0]	1.7 [0.8-2.4]				
Lausanne	1.4 [1.1-2.0]					

Table 5.4: Median duration from peak to peak and inter-quartile range. The underlying epidemic curves were smoothed with a moving average of five days for identifying stable peak days.

These observations regarding the spatial patterns of influenza spread can be interpreted in many ways. First, it must be stated that the spatial spread of influenza given certain initial conditions is a stochastic process. Nishiura et al. [249] showed clearly that during the early phase of an outbreak parameters like the reproduction rate can be highly fluctuating – and in a spatially explicit model we observe these random fluctuations at every place where the disease is newly introduced: Before a certain point of no return has been passed *locally*, it is a random process whether the local outbreak dies off again or can be sustained.

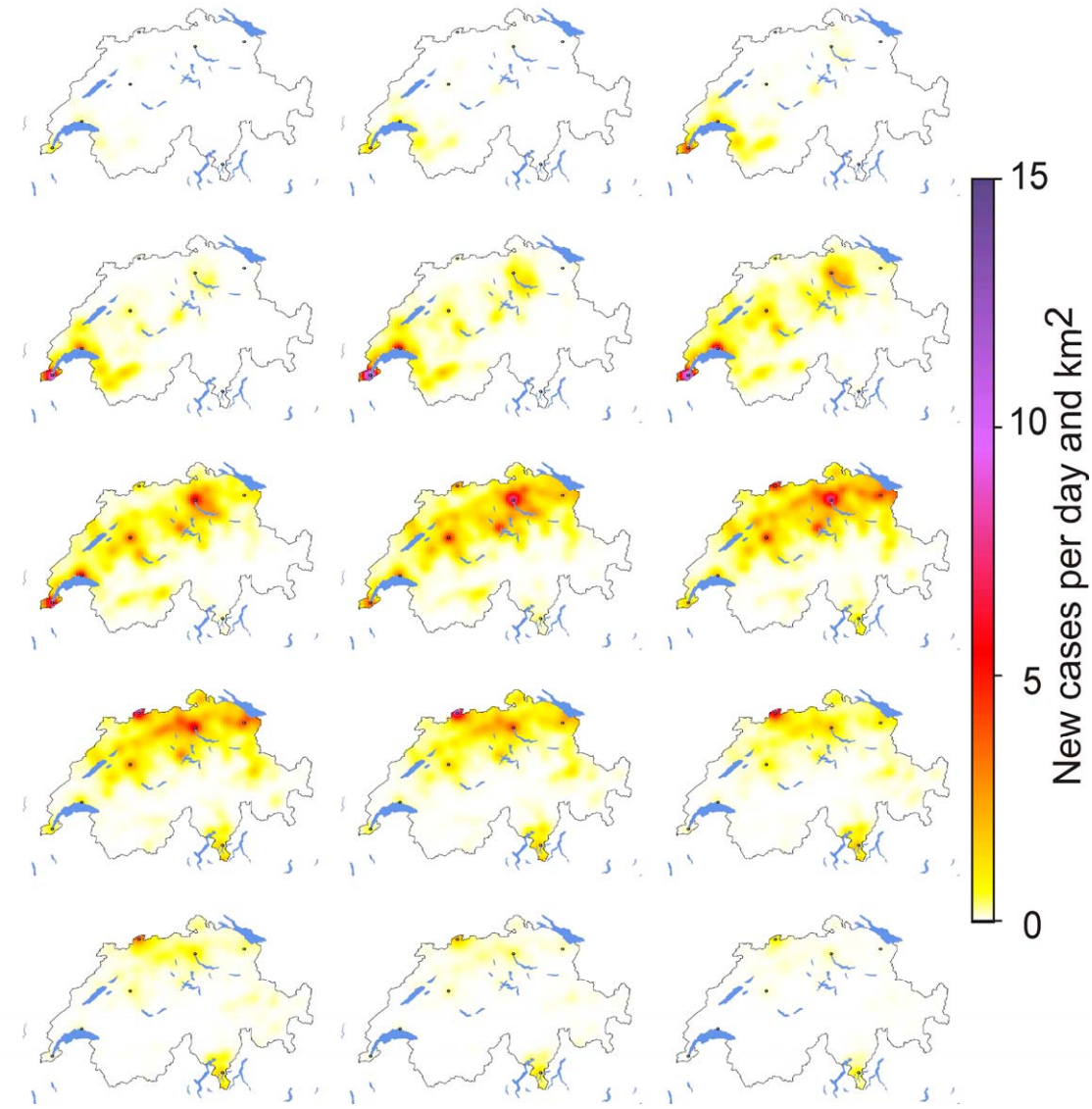


Figure 5.6: Simulated spatial outbreak patterns. This figure shows the prevalence of influenza in the course of simulation time for one simulation run.

It is also a matter of chance whether there is, e.g., an infector travelling very early from Geneva to Lugano and causing an outbreak there or whether Lugano is reached rather late. However, we claim that there are typical time lags between the peaks in different regions of Switzerland: There are hubs – like Zurich – which are origins, stopovers or final destinations for many activities. These hubs are highly connected with the rest of the country by large traffic flows and are, thus, likely to become affected by an epidemic quite early. On the other hand, there are remote regions – like the lower Engadin in the south-western part of the Alps – which have a good chance of being spared for quite a long time. The canton of Ticino with the city of Lugano is geographically isolated from the rest of Switzerland by the Alps and is strongly orientated toward Italy, which might explain why it was affected last by the 2003/2004 epidemic.

Consequently, we believe that analyzing spatial patterns and cautiously comparing simulated patterns with observed patterns is meaningful, although random variation in both reality and simulation can lead to "atypical" outcomes.

5.4.4 Limitations of the model and the study design

Limitations of the overall study design: We chose a particular influenza season in order to reconstruct it and justified this case selection. We took the stance that every single aspect of an influenza outbreak which can be reproduced is a cue for the validity of a modeling approach. However, reproducing one particular outbreak can be just a starting point for a validation strategy. Further suitable data sets should be identified for challenging the proposed influenza model for gaining more certainty that the presented model captures influenza dynamics reasonably well.

Limitations of the empirical data: The reconstruction of a seasonal influenza outbreak in Switzerland rests upon general knowledge about seasonal influenza reported in the literature and upon sentinel information about Switzerland and the surrounding countries.

On the one hand, seasonal influenza is extremely variable in its actual characteristics (cf., e.g., the enormous variability in the age-dependent attack rates shown in Figure 5.5a): Hence, findings reported in older studies can only serve as heuristics and not as precise parameter estimates. Further, one has to note that important factors such as social contact structure vary in different cultural contexts (cf. differences between ten European countries reported in [237]) and, thus, the generalizability of findings to a different social context is limited.

On the other hand, sentinel data can give only a biased image of reality and have to be interpreted with caution: As mentioned, there is a tendency to overestimate the number of new cases in the early and late phase of an outbreak when true cases are not abundant. Sentinel systems can only record correctly diagnosed, symptomatic cases which seek out a practitioner. Accordingly, all asymptomatic cases and all sick individuals that do not see a practitioner and all false diagnoses remain undetected. Unfortunately, asymptomatic infection is age-dependent and the consultation rates also differ most likely between age groups and between the different parts of Switzerland. Finally, it is not clear whether and – if so – to what degree vacations lead to reporting biases.

For future studies, it would be optimal to have seasonal epidemics for which both sentinel data and serologic information about the overall and the age-dependent infection rate is available. Relying solely on sentinel data is accompanied by a considerable uncertainty regarding the

infection rate estimate and age-dependent infection rates as one can only reproduce general knowledge instead of outbreak-specific data.

Limitations of the model structure: There are several limitations in the model itself. For instance, it is known that traffic and activity patterns differ vastly between weekdays and weekends [296] – but we assumed weekday patterns throughout the simulation. We further have not included voluntary home confinement nor changed patterns due to vacations or other events in our model. We think the former is an acceptable assumption given that we do not know real-world rates of home confinement and given that the severity of symptoms seems to lag the shedding rate slightly [198] and, thus, infectors are likely to expose the community to influenza during their most infectious phase. The latter might be part of the explanation for the difference between the model outcome and the extrapolated sentinel data in the first week of 2004. We also do not distinguish the probability of becoming a symptomatic or an asymptomatic case depending on the immunity status of the respective host. Finally, we assume all contacts between an infector and a susceptible person to lead to infection with equal probability irrespective of the duration and the intensity of the contact, which is also known to be a simplification of reality [313].

5.5 Conclusions

This paper had two purposes: (i) A spatially explicit, individual-based model for influenza spread should be introduced, and (ii) it should be validated with empirical data from the 2003/2004 H3N2 influenza epidemic. We succeeded in compiling a simulation model, which is capable of reproducing main characteristics of the 2003/2004 H3N2 epidemic in Switzerland and seasonal influenza in general. Namely, we were able to reproduce the following characteristics:

1) The shape of the epidemic curve was similar in model and extrapolated Swiss sentinel data; overall infection rates and reproductive numbers were in the same range as reported for several outbreaks in various countries.

2) As we had no empirical data on the age-dependent infection rates for the 2003/2004 H3N2 epidemic in Switzerland, simulated patterns had to be compared with observed patterns from other contexts. Nonetheless, the simulated infection rates and the patterns observed during seven outbreaks in two US municipalities are in good qualitative accordance.

3) The empirical data show a clear spread along the main transport axes: The epidemic started at the westernmost tip of Switzerland and moved along the main axes to the north-east and from there to the rather isolated canton of Ticino in the south-east. This pattern could be

reproduced with our model and median time lags between different cities were in good accordance with the observed data.

With this study, we showed that finding evidence for the validity of influenza models by challenging them with seasonal influenza outbreak data is possible. Validating the structure of influenza models with seasonal influenza data seems us to be a good strategy for making – inherently hypothetical – pandemic models more credible: If a model succeeded in reproducing several aspects of seasonal influenza, it is also likely to generate adequate results for pandemic scenarios.

Further case examples should be reconstructed to gain more certainty regarding the model's validity. The highest uncertainties in our model lie in the patterns of pre-existing immunity, the proportion and infectiousness of asymptomatic cases and age-dependent attack rates. All these uncertainties could be approached with a longitudinal serological study measuring antibody titers before and after the seasonal influenza period designed as a complement to the existing sentinel systems. For better pandemic planning and more informed decisions in the field of seasonal influenza, public health authorities should consider investigating this *terra incognita* in more depth.

5.6 Competing interests

The authors declare that they have no competing interests.

5.7 Authors' contributions

TS designed the influenza model, carried out the epidemiological parts of the simulation work, contributed to the quality control of the Swiss MATSim scenario, reviewed the literature and wrote this paper as the lead author. MB is one of the head developers of the MATSim project. He carried out the transportation part of the simulation work and helped to draft the manuscript. JH parameterized the infectiousness over time and contributed to the geographical analysis of the simulated data and helped to draft the manuscript. KWA and RWS contributed to the contact network parts of the study design whereas JZ contributed to the disease biology part of the study design. All of them contributed to the manuscript. All authors read and approved the final manuscript.

6 Conclusions

The leitmotif of this thesis was to examine how contact characteristics shape the spread of infectious diseases. In particular, we investigated how the configuration and the quality of host-to-host contacts can inhibit or accelerate the spread of infectious diseases.

In this chapter, we present the main conclusions of this thesis and identify further research needs. Our conclusion will be structured along the guiding questions which we posed in the introduction. Finally, there is an overall conclusion sketching a general position on future requirements for research about and with models of epidemics.

Configuration of contacts

We subsumed various measures and phenomena under the label *configuration of contacts*, all of which have to do with the arrangement of potentially contagious contacts between different hosts. Characteristics, which we explicitly considered in our four contributions, are the average number of contact partners per individual (called average *degree* in network analysis) and its dispersion as well as mechanisms that increase the locality of contacts, such as clustering or interaction within defined groups (e.g., school classes, households etc.). Particularly the first two contributions showed that one should not investigate the relevance of such contact configuration isolated from biological factors and contact quality. We were able to differentiate transmission systems, for which random mixing models (ignoring all kinds of social structure) appear to result in good approximations of the observed epidemic behavior. However, there are other cases in which random mixing models lead to notably biased results.

The actual configuration of contacts seems to play a subordinate role in case of typical childhood diseases. The characteristic of diseases such as measles or mumps is that they are highly contagious (R_0 is in the range of 10^1), and they spread effectively on the airborne route. Consequently, the corresponding transmission systems are characterized by large numbers of contact partners (as the disease spreads through aerosols and non-intense contacts also qualify for transmission) and high per-contact transmission probabilities. The large number of non-intense contacts includes also many incidental interactions between otherwise unrelated people, which are neither clustered nor highly structured in any other way. This also means that the actual arrangement of contacts between individuals is closer to random-mixing than in other cases. As we indicated in Chapter 2, these are good premises that random mixing models may lead to approximately correct reproductions of measured epidemic curves. In fact, spread is mostly governed by biological factors, such as the infectious period, infectivity or patterns of immunity in the host population.

In contrast, other infectious diseases need intense contact for transmission. The spread of methicillin-resistant Staphylococcus aureus or Ebola epidemics are characterized by the fact that direct contact with infectious secretions is necessary for transmission. One consequence of this is the typically low basic reproduction numbers estimated for such diseases. Due to low numbers of contact partners which are typically highly clustered and due to rather low infectivity, such diseases are more governed by contact structure and stochasticity and less by biological factors. Applying the random mixing assumption, e.g., in models of Ebola spread [e.g., 199] might lead to strongly biased model outcomes or erroneous conclusions on the effectiveness of simulated interventions.

Particularly the findings presented in Chapter 2 can serve as a heuristic for the decision on the type and detailedness of a model of disease spread. However, in addition to the clear extreme cases, there are many other incidences where the applicability of the random mixing assumption is a matter of argumentation. Furthermore, in Chapter 2 we examined small set of characteristics of transmission systems, which were all operationalized in a rather simplified way. Thus, there are other factors, for instance, the spatial location of individuals and their contact partners (cf. Chapters 4 and 5), which further constrain the applicability of simplifying mixing assumptions.

Ultimately, the kind and details of structural properties of contacts that must be included in the model are defined by the actual function a model is ought to fulfill (cf. Section 1.2). Even in the case of the highly infectious measles, we find phenomena, which cannot be explained without a closer look at the configuration of contacts. For example, in 1999 few initial cases of measles in the Netherlands resulted in a total outbreak of 2961 reported cases, even though between 95% and 96% of the Dutch children were reported to have been vaccinated [336]. In a random mixing model, a vaccination coverage of 93.3% – 94.4% (with $q_c = 1 - 1/R_0$, where q_c is the critical vaccination coverage and R_0 is assumed to be between 15 and 18; cf. Table 2.1) is sufficient to contain outbreaks. This specific outbreak can be explained by a cluster of non-vaccinated individuals who were members of a specific religious community that refused vaccination. Accordingly, no heuristic for model selection can replace a thorough reasoning of the biological and social mechanisms which are at work.

Needs for further research

The set of parameters that was systematically analyzed was rather limited. It would be helpful to know how additional factors shape models of epidemics given a certain context, such as (i) the assortativeness of contacts regarding age and physical space, (ii) the dispersion of the

number of contact partners or (iii) higher-order clustering (the common definition of clustering refers to triplets, but in principle quadruples or any other n-tuple might be of relevance).

Further, we must acknowledge that in practice models for prediction are fitted to field data. For instance, the parameters of a deterministic compartmental model are varied until a sufficiently good approximation (usually in terms of least-squares) is achieved between model outcome and measured data. Thus, the baseline scenario (i.e., without any intervention) of such a model is *per se* a good reconstruction of the measured reality in terms of model fit. Hence, we need systematic research on how strongly simplifying mixing assumptions affects predictions of intervention effectiveness *if* the baseline scenario is fitted to empirical outbreak data in the absence of interventions.

Quality of contacts

Research with models of epidemics often records contacts between hosts as a binary measure; a potentially contagious contact between one host and another host is either present or absent. Two reasons for this are apparent. First, many network indices have originally been defined for unvalued graphs and are often not or not easily transferable to valued graphs. Also deriving inferences with means of mathematical operations is often easier with unvalued graphs. Secondly, for many concrete transmission systems, it is difficult to actually value contacts between hosts. Often, we do not know exactly how a certain behavior translates into risk of infection and, moreover, many kinds of interaction are hard to measure in a host's everyday life.

In Chapter 3, we proposed a mechanistic approach (cf. also p. 23) of how to include duration and intensity of contacts as two examples of contact quality in models of disease spread. The mechanistic model of transmission described in Chapter 3 relates the probability of infection to the duration and the intensity of a contact. Thereby, we concentrate on diseases, which are predominantly transmitted via large droplets or close physical contact. An approach for truly airborne diseases is described in Section 1.3; work on sexually transmitted diseases has been conducted by other researchers [49, 251]. Our proposed mechanistic model of transmission was used with empirical contact data.

The results presented in Chapter 3 suggest that there is a clear correlation between the configuration and the quality of contacts. We found that individuals, who reported a high number of different contact partners on a certain day, spent on average less time with each of these contact partners than other persons. Moreover, we found a correlation between the duration of a contact and its intensity. Short encounters are likely to remain purely conversational. Long-lasting interaction between humans often involves physical contact including kisses.

The resulting differences in transmission probabilities and, particularly, the correlation between configuration and quality have implications for the understanding of the population dynamics of infectious diseases. We have learned from previous work that the dispersion of the number of potentially contagious contacts modulates the epidemic threshold (cf. Subsection 1.4.2). It was even suggested that under a scale-free network topology the actual infectivity of causative agents is completely irrelevant for sustaining disease spread [263]. All of these findings are based on the assumption that all contacts between hosts weigh equally. In contrast, our results show that such statements should be moderated. In the case of most infectious agents, it is not only the high number of different contact partners which makes an individual a super-spreader. To be a super-spreader, higher-than-average shedding rates also seem to be necessary.

Needs for further research

The work presented in Chapter 3 allows for a general insight into the interdependency of individual transmission probabilities and the configuration of contacts. We further proposed an approach, which makes it *in principle* possible to include the quality of contacts into population-level models of epidemics. However, contact duration is just a proxy measure for the amount of infectious material exchanged, and quantifying the effect of various contact intensities seems currently impossible.

This problem has two origins. First, it is technically difficult to dissect how much more infectious material is transferred from an infector to another host when they are kissing or touching each other instead of simply having conversation. Secondly, it is almost impossible to measure more detailed information on contact intensity with a classical diary approach than we have done here. The rather rough way of measuring contacts presented in Chapter 3 has already overstrained several study participants.

What should be done to tackle the first problem is to find or to measure appropriate epidemiological field data to calibrate the transmission model. Having large enough and precise enough connected data sets on the infection history as well as on the behavior of a defined population would allow refining the model and estimating its parameters.

Furthermore, contact duration and intensity could be measured more precisely with other methods than contact diaries. Waber et al. [342] developed so-called wearable sociometric badges, which allow for a record with whom the equipped participants had conversation and with whom they stood in close spatial vicinity. Measuring the interaction of a defined population (this approach allows only measuring interaction between equipped participants) would offer a more precise picture of the epidemiologically relevant behavior of people than classical diary studies and would also allow correcting measurements gained by diary research.

Finally, for transmission systems of causative agents that are only transmitted via rather close physical contact, there might be a structural analogy to what Granovetter described in his research as the "strength of weak ties" [139]: There might be groups or clusters of highly and intensely connected individuals, whereas these groups themselves are interconnected via rather infrequent and non-intense ("weak") contacts. Within such cohesive groups diseases would spread quickly, but on a higher level, the infrequent and non-intense contacts might be structurally more important to channel an epidemic through an entire population (cf. discussion on betweenness centrality on pp. 41). This hypothesis makes clear that we need, also with respect to contact quality, a shift from the egocentric network perspective to population-level models.

Configuration of contacts between poultry farms

The work presented in Chapter 2 gives hints that the concrete configuration of contacts between poultry farms in Switzerland might influence the patterns of a future avian influenza outbreak. As both the number of contact partners of a farm and the reproduction rates on the farm level (i.e. the number of newly affected farms caused by an initial farm with avian influenza cases cf. [125, 212]) are rather low, we can expect that such a transmission system is predominantly governed by the contact structure and stochasticity.

With the multi-method approach described in Chapter 4, we were able to identify several epidemiologically relevant characteristics of the contact structure of Swiss poultry farms:

First, we found the distribution of the number of different contact partners and of the contact frequencies to be highly skewed. For both measures, commercial farms are mainly responsible for the long tail. Particularly some specialized commercial farms tend to have a high number of contact partners and/or high contact frequencies. If one accepts contact frequency to be an indicator for the overall contact quality, the opposite of what we found for individual conversational contacts (Chapter 3) might be true for organizational business contacts (Chapter 4): While in case of contacts between individuals, the average intensity decreases with the number of contact partners, such a relation is not reflected in the data on potentially contagious contacts between poultry farms. This also allows identifying farms, which might be particularly important for the dynamics of a future avian influenza outbreak in Switzerland.

Secondly, we were able to show that non-commercial farms play an important role in the net of between-farm contacts, contrary to what has typically been assumed in other studies. Although commercial farms have on average many more animals, more contact partners and higher contact frequencies than non-commercial farms, non-commercial farms can "outweigh" this by their mere number. Furthermore, we found out that commercial and non-commercial poultry husbandry are not separated systems as often assumed. Commercial and non-commercial

farms are interconnected by poultry sale, poultry shows and commonly used infrastructure. Moreover, there are even farms which have both professional and private livestock husbandry at the same location.

Various models of avian influenza spread with different underlying assumptions have been published. Based on our work, we recommended one of these competing modeling approaches. Furthermore, we are able to inform modelers about the sector's contact structure and we could show that non-commercial farms definitely should be included in models of avian influenza spread.

Needs for further research

Currently, the largest gap in relevant information about avian influenza transmission systems is the missing knowledge about contact quality. Research like that which was presented in Chapter 4 gives detailed information about the configuration of contacts between the different kinds of poultry farms, but weighing these contacts according to transmission risk seems currently impossible. For instance, we do not know whether sale/purchase relations between farms are more or less likely to lead to a newly infected farm than show visits. Furthermore, it is not clear how likely indirect connections between farms, e.g., via dead stock collection points or shared staff, lead to transmission when compared to direct contacts between the birds, e.g., at bird shows or when they are moved to another farm. Also the exact quantitative influence of good hygiene practice is not clear. Finally, we do not know whether the movement of 1000 birds from one farm to another has a higher transmission probability than the movement of just one bird, and, if so, what kind of function describes the relation between the number of birds and transmission probability.

Consequently, we see the most urgent need for research at this specific interface between microbiological risk assessment and the analysis of organizational behavior. One possibility to quantify the relative risks of transmission of different pathways and contact characteristics might be to simultaneously trace during an outbreak how the epidemic advances in a given population of farms and to measure the different kinds of potential pathways that exist between affected and non-affected farms. Such data might allow for calibrating models of epidemics to optimally fit the empirical data also with respect to contact quality.

Validation of high-resolution models of disease spread

An increasing number of publications deal with future influenza pandemics and how they could be fought. The customary approach to investigate this is to use mathematical or computer models. Thereby, a vast variety of model types are applied ranging from simple compartmental models [66] to highly sophisticated individual-based models [110]. A common problem of all of

these approaches is that it is inherently difficult to validate models of hypothetical epidemics or pandemics.

We rely on the prevailing view in the philosophy of science that validation is a rather argumentative process and, thus, cannot be sourced out to purely algorithmic methods. From a functionalist or pragmatic perspective, models are valid when they fulfill their purpose. The purpose of such models of epidemics is to accurately predict epidemic curves for the baseline and intervention scenarios. To judge this, we have to collect signs and evidence whether they do so sufficiently or not (cf. pp. 25). Such a sign is, for instance, if the model conforms in various aspects (which, ideally, were not fitted to measured data!) to empirically confirmed features. Another sign would be if one single model can be used to successfully reconstruct such empirically known features of epidemics for various outbreaks. To test models for pandemic influenza, we suggest challenging them with known data of recent seasonal outbreaks. If models stand the test in case of the more complex seasonal influenza, we have reason to believe that the model outcomes are also reasonable for pandemic scenarios.

We were able to show that an individual-based model including (i) the travel behavior of people, (ii) their assignment to specific households, work groups, classes, (iii) age-dependent numbers of contact partners and assortativeness of contacts, (iv) immunity patterns and (v) heterogeneous shedding rates as well as disease severities is capable of reproducing spatial and age-related patterns of a seasonal influenza epidemic. We did this for the 2003/2004 H3N2 epidemic in Switzerland (cf. Chapter 5).

Needs for further research

There are two main aspects in our own work, which should be improved by future work.

First, we have tested our model only for one influenza season that had optimal conditions for reconstruction. Further influenza seasons should follow to gain more confidence that the presented simulation model incorporates the relevant mechanisms that shape influenza spread. Cases that might be qualified are the 2000/2001 H1N1 season and the 2009 H1N1 season, because both had one clearly dominating strain.

Secondly, we did not differentiate the per-contact transmission probability according to age, number of contact partners or activity. According to our work presented in Chapter 3, it is plausible to assume that there are systematic differences in the transmission probability of various types of contact. However, there is currently no adequate data available which would allow parameterizing a model which would reflect such differences. Nonetheless, future work should assess how sensitive the simulation outcomes are to changed assumptions of contact

qualities despite the fact, that our model reproduces several aspects of influenza spread satisfactorily.

Furthermore, our work identified that there is a lack of continuously updated pre- and post-season serological data, which would perfectly complement the sentinel data recorded by the Swiss health authorities. Sentinel data offer valuable insights into the temporal and spatial *dynamics* of influenza epidemics. They allow for identifying when and where an epidemic began, when the peak was reached and when the epidemic phase was over. However, sentinel data are biased in so many ways that reliable estimates on the total outbreak size and the age distribution of infection cannot be made. Measuring the serological status before and after an influenza season would allow complementing the insights into the *dynamics* of the epidemic with good *static* data on the pre-existing immunity and the true extent of the epidemic in the various age groups.

Overall conclusion

This thesis provides some new building blocks to our understanding of how contact characteristics shape the spread of infectious diseases. All single contributions which build this thesis are good examples that neither biological knowledge nor a comprehension of the hosts' behavior leading to certain contact structures is – in isolation – sufficient for understanding disease spread. Instead, we find our point of view supported—that infectious disease spread should be approached with a systemic, problem-oriented perspective. Only with a thorough understanding of the biological, physical and chemical mechanisms that govern the transmission of infectious agents, can we determine (i) what kind of contacts are relevant for disease spread, (ii) how these contacts should be weighed to reflect transmission probabilities and (iii) which level of detail is functional for understanding those aspects of disease spread in which we are interested.

In recent years, a lot of work has been done to determine what makes models of disease spread "realistic" [102] or "appropriate" [349]. Science has learned a lot about the complex interplay between behavior, biological mechanisms and environmental conditions. The growing number of publications dealing with the role of biological and social factors for the understanding of disease spread indicates that we are in the midst of a period of knowledge generation for better epidemiological modeling. However, there is also a need for consolidation and systematization. In statistics, rules have been established as to what the preconditions are for a certain statistical tool to be applied. For example, if we want to calculate a linear regression model, we know that the residuals must follow the normal distribution and have to be homoscedastic as well as statistically independent so that the measures of the model make sense. On the other hand, we

also know that the regression equation is rather robust against violations of these preconditions and that "just" the tests for significance and the calculations of the error are wrong. A comparable, systematic guidance for the applied researcher is lacking for infectious disease spread. Future research should further outline systematically which details of the interplay between contact structure, biological mechanisms and environmental factors must be considered, under what conditions and, additionally, also when they can be neglected.

7 Acknowledgements

Conducting a long-term project in science is not possible without the intellectual, financial and practical support of other persons and institutions. In this final part of the thesis I acknowledge those persons and institutions, who contributed significantly to my research.

First, I thank my supervisors, Roland W. Scholz, Kay W. Axhausen (both ETH Zurich) and Jakob Zinsstag (Swiss Tropical and Public Health Institute, Basel), for making this research project possible and for supervising my work. I am grateful for the freedom I have enjoyed to tailor the project to my personal research interests. I have highly appreciated this enriching collaboration in a truly interdisciplinary environment.

Further, I thank the Swiss National Science Foundation, which funded my research project from the very beginning of my PhD studies. Further financial support came from the Hubert-Kudlich Foundation. Lena Fiebig (Swiss Tropical and Public Health Institute; now Robert Koch Institute, Berlin), who is one of the authors of two contributions, was funded by the Swiss Federal Veterinary Office. Furthermore, Lena Fiebig, Jennifer Saurina, Jan Hattendorf and Jakob Zinsstag (all Swiss Tropical and Public Health Institute) received funding from the "Constanze" research project for their research work in the context of this thesis.

Several persons contributed to the parts of this thesis in manifold ways. Their support was important for raising the quality of the research presented in this thesis. Subsequently, I will acknowledge contributions of these peers, colleagues and friends by chapter.

Chapter 1: Particularly extensive comments and suggestions were provided by Lena Fiebig, Daniel Lang (University of Lüneburg), Bernhard Ritter (University of Zurich) and Arnim Wiek (Arizona State University). Lena Fiebig reviewed the entire Chapter 1. Daniel Lang reviewed Section 1.1. Bernhard Ritter and Arnim Wiek reviewed Section 1.2. Furthermore, Pius Krütli, Corinne Moser, Andrea Ulrich and Stefan Zemp (all ETH Zurich) as well as Philippe Peter and Annabelle Puchta, read and commented on various parts of Chapter 1. Sandro Bösch (ETH Zurich) helped the design of the figures. The language was edited by the copyediting service EditMyEnglish.

Chapter 2: Myself, Lena Fiebig and Roland W. Scholz are the authors of this paper. Devon D. Brewer (Interdisciplinary Scientific Research, Seattle), Istvan Z. Kiss (University of Sussex), Peter de Haan and Fadri Gottschalk (both ETH Zurich) as well as Philippe Peter help with comments and suggestions at an early stage of this research. Jan Hattendorf and Esther Schelling (both Swiss Tropical and Public Health Institute) as well as four anonymous reviewers made valuable

comments on the manuscript. Sandro Bösch created the final layout of the figures. Stephanie Keller revised the language of this paper.

Chapter 3: The data underlying this paper was collected by Judith Maag and Ladina Muggler (both students at ETH Zurich) under my supervision. Mirjam Kretzschmar (RIVM, The Netherlands) kindly provided the original questionnaires of one of their contact studies. Nicola Low, Janneke Min-Heijne, Adrian Röllin (all University of Bern) and Fredrik Liljeros (Stockholm University) supported and improved this research at different stages with their valuable comments. The final manuscript was reviewed by Roland W. Scholz, Michael Stauffacher, Corinne Moser (all ETH Zurich) and Lena Fiebig as well as two anonymous reviewers. Sandro Bösch helped with the final layout of the figures. The language was edited by the copyediting service EditMyEnglish.

Chapter 4: The authors of this paper are Lena Fiebig, myself, Jennifer Saurina, Jan Hattendorf and Jakob Zinsstag. The coordinates for all locations were provided by Kay W. Axhausen and Veronika Killer (both ETH Zurich). Michael Binggeli and Heinz-Peter Schwermer (both Swiss Federal Veterinary Office) provided the Swisstopo base maps. Tyler O'Neill helped with the English editing. Valuable comments were provided by the anonymous reviewers.

Chapter 5: Myself, Michael Balmer, Jan Hattendorf, Kay W. Axhausen, Jakob Zinsstag and Roland W. Scholz are the authors of this manuscript. The Swiss sentinel data and the data of Table 5.1 were provided by the Swiss Federal Office of Public Health. The databases mentioned in Table 5.2 were provided by the Swiss Federal Statistical Office and the Swiss Federal Office for Spatial Development. Lena Fiebig, Sasa Parad and Justus Gallati (both ETH Zurich) reviewed the manuscript. Sandro Bösch helped with design of the figures. The language was edited by the copyediting service EditMyEnglish.

Chapter 6: Chapter 6 was edited by the copyediting service EditMyEnglish.

Finally, I thank my fiancée, Annabelle Puchta, who supported me with both practical help as well as loving patience. I also thank my parents Renate and Heinz Smieszek. Without their education, their continuous care and also their financial support, I would not have been able to find and go my own way as I did in the last decade.

8 Abbreviations

AGIS *AG*rar *I*nformation *S*ystem.

AIDS Acquired immune deficiency syndrome; disease of the immune system caused by →HIV.

CC Clustering coefficient (cf. Subsection 1.4.2).

CCLA Creative Commons Attribution License; http://creativecommons.org/licenses/by/2.5/

CI Confidence interval (95%, if not stated differently).

CFC Chlorofluorocarbon; a class of highly inert organic compounds that have been used as refrigerants, propellants and solvents. Nowadays the use of CFCs is highly regulated and restricted as they have been one of the main causes for stratospheric ozone depletion.

COPD Chronic obstructive pulmonary disease; refers to chronic bronchitis and emphysema; the disease is irreversible and worsens over time.

DALY Disability-adjusted life years; an index developed by the →WHO that integrates mortality and morbidity.

DDT Dichlorodiphenyltrichloroethane; a formerly widely used pesticide, which is regulated under the Stockholm Convention on persistent organic pollutants due to its adverse environmental effects.

DNA Deoxyribonucleic acid; a macromolecule that stores genetic information of living organisms and some viruses.

EID Emerging infectious diseases; infectious diseases whose incidence has increased significantly within the last 20 years or is likely to increase within the next few years.

H1N1 Specific influenza A virus subtype occurring in birds; the “H number” denotes a specific type of hemagglutinin; the “N number” stands for a specific neuraminidase. Both are surface glycoproteins of influenza viruses: the hemagglutinin helps in identifying and entering target cells, the neuraminidase helps to be released from a host cell.

H3N2 →H1N1

H5N1 →H1N1

H7N7 →H1N1

HI Hemagglutination-inhibiting; many virus (including influenza) bind to red blood cells, a process that is called hemagglutination. Strain-specific antibodies of the immune system can inhibit hemagglutination. Immunity against such virus strains can be tested by checking the quantity of antibodies in a serum. The highest dilution of serum that still inhibits hemagglutination is called the HI titer of the serum.

HIV Human immunodeficiency virus; causative agent of →AIDS.

HPAI Highly pathogenic avian influenza.

ILI Influenza-like illness; medical diagnosis of influenza or another illness with common symptoms.

IQR Inter-quartile range; range between the third and the first quartile.

ISO International Organization for Standardization.

MATSim Multi-Agent Transport Simulation Toolkit; http://www.matsim.org

MRSA Methicillin-resistant *Staphylococcus aureus*; a strain of Staphylococcus aureus with multiple resistances to antibiotics.

NGO Non-governmental organization.

OIE World Organisation for Animal Health, formerly known as Office international des epizooties

RNA Ribonucleic acid; a macromolecule that stores genetic information of most viruses.

R_0 Basic reproduction number (cf., e.g., Subsection 1.4.1).

SARS Severe acute respiratory syndrome; a respiratory disease in humans caused by the SARS coronavirus.

SD Standard deviation.

SEIR Stands for *S*usceptible-*E*xposed-*I*nfectious-*R*ecovered (or *R*emoved) and denotes a model type including these three stages an individual can take in an infection process.

SIR Stands for *S*usceptible-*I*nfectious-*R*ecovered (or *R*emoved) →SEIR (cf. Subsection 1.4.1).

SIS Stands for *S*usceptible-*I*nfectious-*S*usceptible →SEIR.

WAIFW "Who acquires infection from whom"-matrix (cf. Subsection 1.4.1).

WHO World Health Organization.

XML Extensible Markup Language; a standard for encoding documents in a structured way.

9 References

1. **Multi-Agent Transport Simulation Toolkit | MATSim** [http://matsim.org]
2. **Agency.** In *The Merriam-Webster dictionary*, 5th edition. Springfield, MA: Merriam-Webster; 1997.
3. **Cellular Automata/Neighborhood** [http://en.wikibooks.org/wiki/Cellular_Automata/Neighborhood]
4. **Validity.** In *Encyclopaedia Britannica Online*, 31 Dec 2009 edition: Encyclopaedia Britannica; 2009.
5. Albert R, Barabási A-L: **Statistical mechanics of complex networks.** *Rev Mod Phys* 2002, **74:**47-97.
6. Alexandersen S, Brotherhood I, Donaldson AI: **Natural aerosol transmission of foot-and-mouth disease virus to pigs: minimal infectious dose for strain O_1 Lausanne.** *Epidemiol Infect* 2002, **128:**301-312.
7. Allison VD: **Crowd diseases.** *Public Health* 1958, **72:**43-53.
8. Anderson RM, May RM: *Infectious diseases of humans: dynamics and control.* Oxford, UK: Oxford University Press; 1991.
9. Aparicio JP, Pascual M: **Building epidemiological models from R-0: an implicit treatment of transmission in networks.** *Proc R Soc Lond B* 2007, **274:**505-512.
10. Arbeitsgemeinschaft Influenza (AGI): *Abschlussbericht der Influenzasaison 2003/04.* Berlin: Robert Koch Institut; 2004.
11. Arino J, Brauer F, van den Driessche P, Watmough J, Wu J: **A model for influenza with vaccination and antiviral treatment.** *J Theor Biol* 2008, **253:**118-130.
12. Armitage P, Meynell GG, Williams T: **Birth-death and other models for microbial infection.** *Nature* 1965, **207:**570-572.
13. Armitage P, Spicer CC: **The detection of variation in host susceptibility in dilution counting experiments.** *J Hyg (Lond)* 1956, **54:**401-414.
14. Ashby WR: *An introduction to cybernetics.* London: Chapman & Hall; 1957.
15. Atkinson M, Wein L: **Quantifying the routes of transmission for pandemic influenza.** *Bull Math Biol* 2008, **70:**820-867.
16. Baccini P, Bader H-P: *Regionaler Stoffhaushalt: Erfassung, Bewertung und Steuerung.* Heidelberg: Spektrum Akademischer Verlag; 1996.
17. Baehr JS: **A priori and a posteriori.** In *Internet Encyclopedia of Philosophy* (Fieser J, Dowden B eds.), 18 Oct 2006 edition; 2006.
18. Baker A: **Simplicity.** In *The Stanford Encyclopedia of Philosophy* (Zalta EN ed., fall 2008 edition; 2008.
19. Balmer M, Axhausen KW, Nagel K: **Agent-based demand-modeling framework for large-scale microsimulations.** *Transport Res Rec* 2006, **1985:**125-134.
20. Balmer M, Horni A, Charypar D, Meister K, Ciari F, Axhausen KW: **Schlussbericht. Wirkungen der Westumfahrung Zürich: Analyse mit einer agentenbasierten Mikrosimulation.** (Institute for Transport Planning and Systems ed., vol. 550. Zurich: ETH Zurich; 2009.

21. Balmer M, Rieser M, Meister K, Charypar D, Lefebvre N, Nagel K: **MATSim-T: architecture and simulation times.** In *Multi-agent systems for traffic and transportation engineering.* Edited by Bazzan AL, Klügl F. Hershey, PA: Information Science Reference; 2009: 57-78
22. Bansal S, Grenfell BT, Meyers LA: **When individual behaviour matters: homogeneous and network models in epidemiology.** *J R Soc Interface* 2007, **4:**879-891.
23. Bar-Yam Y: *Dynamics of complex systems.* Boulder, CO: Westview Press; 1997.
24. Barbour A, Mollison D: **Epidemics and random graphs.** In *Stochastic processes in epidemic theory.* Edited by Gabriel J-P, Lefèvre C, Picard P. New York, NY: Springer; 1990: 86-89
25. Barlas Y, Carpenter S: **Philosophical roots of model validation: two paradigms.** *Syst Dynam Rev* 1990, **6:**148-166.
26. Barton AH: **Bringing society back in. Survey research and macro-methodology.** *Am Behav Sci* 1968, **12:**1-9.
27. Bavelas A: **A mathematical model for group structure.** *Hum Organ* 1948, **7:**16-30.
28. Beasley VR: **'One Toxicology', 'Ecosystem Health' and 'One Health'.** *Vet Ital* 2009, **45:**97-110.
29. Bell DC, Atkinson JS, Carlson JW: **Centrality measures for disease transmission networks.** *Soc Networks* 1999, **21:**1-21.
30. Bernoulli D: **Essai d'une nouvelle analyse de la mortalité causée par la petite Vérole, & des avantages de l'Inoculation pour la prévenir.** *Mém Math Phys Acad Roy Sci Paris* 1766**:**1-45.
31. Beutels P, Shkedy Z, Aerts M, Van Damme P: **Social mixing patterns for transmission models of close contact infections: exploring self-evaluation and diary-based data collection through a web-based interface.** *Epidemiol Infect* 2006, **134:**1158-1166.
32. Bildungsdirektion Kanton Zürich: **Bildungsstatistik Kanton Zürich** [http://www.bista.zh.ch/allg/default.aspx]
33. Bjornson A, Mellencamp M, Schiff G: **Complement is activated in the upper respiratory tract during influenza virus infection.** *Am Rev Respir Dis* 1991, **143:**1062-1066.
34. Bläuer Herrmann A, Capezzali E, Farine A, Kreis S, Lässig V, Murier T, Perrenoud S, Saucy F: *Arbeitsmarktindikatoren 2009.* Neuchâtel, CH: Office fédéral de la statistique; 2009.
35. Boccaccio G: *The Decameron of Giovanni Boccaccio.* London, UK: Folio Society; 2007.
36. Boender GJ, Hagenaars TJ, Bouma A, Nodelijk G, Elbers AR, De Jong MC, van Boven M: **Risk maps for the spread of highly pathogenic avian influenza in poultry.** *PLoS Comput Biol* 2007, **3:**e71.
37. Boivin G, Hardy I, Tellier G, Maziade J: **Predicting influenza infections during epidemics with use of a clinical case definition.** *Clin Inf Dis* 2000, **31:**1166-1169.
38. Bondy JA, Murty USR: *Graph theory.* London: Springer; 2008.
39. Bonmarin I, Levy-Bruhl D: **Apport des modélisations des épidémies dans la décision de santé publique: exemple de la pandémie grippale.** *Méd Mal Infect* 2007, **37:**S204-S209.
40. Boots M, Mealor M: **Local interaction select for lower pathogen infectivity.** *Science* 2007, **315:**1284-1286.
41. Boots M, Sasaki A: **'Small worlds' and the evolution of virulence: infection occurs locally and at a distance.** *Proc R Soc Lond B* 1999, **266:**1933-1938.

42. Bootsma MC, Diekmann O, Bonten MJ: **Controlling methicillin-resistant Staphylococcus aureus: quantifying the effects of interventions and rapid diagnostic testing.** *Proc Natl Acad Sci U S A* 2006, **103:**5620-5625.
43. Bortz J: *Statistik für Sozialwissenschaftler.* 5th edn. Berlin: Springer; 1999.
44. Bouma MJ, Dye C: **Cycles of malaria associated with El Niño in Venezuela.** *JAMA* 1997, **278:**1772-1774.
45. Bouma MJ, Dye C, van der Kaay HJ: **Falciparum malaria and climate change in the northwest frontier province of Pakistan.** *Am J Trop Med Hyg* 1996, **55:**131-137.
46. Bouma MJ, van der Kaay HJ: **The El Niño Souther Oscillation and the historic malaria epidemics on the Indian subcontinent and Sri Lanka: an early warning system for future epidemics?** *Trop Med Int Health* 1996, **1:**86-96.
47. Brankston G, Gitterman L, Hirji Z, Lemieux C, Gardam M: **Transmission of influenza A in human beings.** *Lancet Infect Dis* 2007, **7:**257-265.
48. Brewer GD: **The challenges of interdisciplinarity.** *Policy Sci* 1999, **32:**327-337.
49. Britton T, Nordvik MK, Liljeros F: **Modelling sexually transmitted infections: The effect of partnership activity and number of partners on Ro.** *Theor Popul Biol* 2007, **72:**389-399.
50. Brunswik E: *The conceptual framework of psychology.* Chicago, IL: University of Chikago Press; 1950.
51. Brunswik E: *Perception and the representative design of psychological experiments.* Berkeley, CA: University of California Press; 1956.
52. Bruschke C, Vallat B: **OIE standards and guidelines related to trade and poultry diseases.** *Rev Sci Tech* 2008, **27:**627-632.
53. Bundesamt für Gesundheit (BAG): **Saisonale Grippe 2007/08: Epidemiologie, Virologie, Impfstoffversorgung und -zusammensetzung.** *BAG Bulletin* 2008, **24/08:**419-423.
54. Bundesamt für Gesundheit (BAG): **Das schweizerische Sentinella-Meldesystem** [http://www.bag.admin.ch/k_m_meldesystem/00736/00817/index.html?lang=de]
55. Bundesamt für Landwirtschaft: **Agrarinformationssystem AGIS** [http://www.blw.admin.ch/themen/00006/00232/index.html?lang=de]
56. Bundesversammlung der Schweizerischen Eidgenossenschaft: **Tierseuchengesetz (TSG) vom 1. Juli 1966, Stand 13. Juni 2006** [http://www.admin.ch/ch/d/sr/c916_40.html]
57. Burrell GA, Seibert FM: **Experiments with small animals and carbon monoxide.** *Ind Eng Chem* 1914, **6:**241-244.
58. Camus A: *Die Pest.* Zurich, CH: Büchergilde Gutenberg; 1949.
59. Capua I, Dalla PM, Mutinelli F, Marangon S, Terregino C: **Newcastle disease outbreaks in Italiy during 2000.** *Vet Rec* 2002a, **150:**565-568.
60. Capua I, Mutinelli F, Pozza MD, Donatelli I, Puzelli S, Cancellotti FM: **The 1999-2000 avian influenza (H7N1) epidemic in Italy: veterinary and human health implications.** *Acta Trop* 2002b, **83:**7-11.
61. Carlisle R: *Scientific American inventions and discoveries: all the milestones in ingenuity - from the discovery of fire to the invention of the microwave oven.* Hoboken, NJ: John Wiley & Sons; 2004.
62. Carmines EG, Zeller RA: *Reliability and validity assessment.* Thousand Oaks, CA: Sage Publications; 1979.

63. Carpenter C, Sattenspiel L: **The design and use of an agent-based model to simulatie the 1918 influenza epidemic at Norway House, Manitoba.** *Am J Hum Biol* 2009, **21:**290-300.
64. Carrat F, Vergu E, Ferguson NM, Lemaitre M, Cauchemez S, Leach S, Valleron A-J: **Time lines of infection and disease in human influenza: a review of volunteer challenge studies.** *Am J Epidemiol* 2008, **167:**775-785.
65. Cauchemez S, Valleron A-J, Boëlle P-Y, Flahault A, Ferguson NM: **Estimating the impact of school closure on influenza transmission from Sentinel data.** *Nature* 2008, **452:**750-755.
66. Chowell G, Ammon CE, Hengartner NW, Hyman JM: **Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: Assessing the effects of hypothetical interventions.** *J Theor Biol* 2006, **241:**193-204.
67. Chowell G, Hengartner NW, Castillo-Chavez C, Fenimore PW, Hyman JM: **The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda.** *J Theor Biol* 2004, **229:**119-126.
68. Chowell G, Miller MA, Viboud C: **Seasonal influenza in the United States, France, and Australia: transmission and prospects for control.** *Epidemiol Infect* 2008, **136:**852-864.
69. Christakos G, Olea RA, Yiu H-L: **Recent results on the spatiotemporal modeling and comparative analysis of Black Death and bubonic plague epidemics.** *Pub Health* 2007, **121:**700-720.
70. Christley RM, Pinchbeck GL, Bowers RG, Clancy D, French NP, Bennet R, Turner J: **Infection in social networks: using network analysis to identify high-risk individuals.** *Am J Epidemiol* 2005, **162:**1024-1031.
71. Ciari F, Balmer M, Axhausen KW: **A new mode choice model for a multi-agent transport simulation.** In *8th Swiss Transport Research Conference*. Ascona, CH; 2008.
72. Ciofi degli Atti ML, Merler S, Rizzo C, Ajelli M, Massari M, Manfredi P, Furlanello C, Scalia Tomba G, Iannelli M: **Mitigation measures for pandemic influenza in Italy: An individual based model considering different scenarios.** *PLoS One* 2008, **3:**e1790.
73. Corey RB, Pauling L: **Molecular models of amino acids, peptides, and proteins.** *Rev Sci Instrum* 1953, **24:**621-627.
74. Couch RB, Douglas RG, Fedson DS, Kasel JA: **Correlated studies of a recombinant influenza-virus vaccine. 3. Protection against experimental influenza in man.** *J Infect Dis* 1971, **124:**473-480.
75. Crosby AW: *America's forgotten pandemic: the influenza of 1918.* 2nd edn. Cambridge, UK: Cambridge University Press; 2003.
76. Da Costa N, French S: *Science and partial truth: a unitary approach to models and scientific reasoning*. Oxford (UK): Oxford University Press; 2003.
77. Das TK, Savachkin AA, Zhu Y: **A large-scale simulation model of pandemic influenza outbreaks for development of dynamic mitigation strategies.** *IIE Transactions* 2008, **40:**893-905.
78. Daszak P, Cunningham AA, Hyatt AD: **Anthropogenic environmental change and the emergence of infectious diseases in wildlife.** *Acta Trop* 2001, **78:**103-116.

79. Davis BD, Dulbecco R, Eisen HN, Ginsberg HS: *Microbiology.* New York: Harper & Row; 1980.
80. de Jong JC, Claas EC, Osterhaus AD, Webster RG, Lim WL: **A pandemic warning?** *Nature* 1997, **389:**554.
81. Dent JE, Kao RR, Kiss IZ, Hyder K, Arnold M: **Contact structures in the poultry industry in Great Britain: exploring transmission routes for a potential avian influenza virus epidemic.** *BMC Vet Res* 2008, **4:**27.
82. Department for Environment Food and Rural Affairs (DEFRA): **Outbreak of highly pathogenic H5N1 avian influenza in Suffolk in January 2007** [http://news.bbc.co.uk/2/shared/bsp/hi/pdfs/20_04_07_defra_bird.pdf]
83. Diez Roux AV, Aiello AE: **Multilevel analysis of infectious diseases.** *J Infect Dis* 2005, **191:**S25-S33.
84. Dobson AP, Carper ER: **Infectious diseases and human population history.** *Biosci* 1996, **46:**115-126.
85. du Prel J-B, Puppe W, Gröndahl B, Knuf M, Weigl JAI, Schaaff F, Schmitt H-J: **Are meteorological parameters associated with acute respiratory tract infections?** *Clin Inf Dis* 2009, **49:**861-868.
86. Duerr H-P, Brockmann SO, Piechotowski I, Schwehm M, Eichner M: **Influenza pandemic intervention planning using *InfluSim*: pharmaceutical and non-pharmaceutical interventions.** *BMC Infect Dis* 2007, **7:**76.
87. Duerr HP, Schwehm M, Leary CC, De Vlas SJ, Eichner M: **The impact of contact structure on infectious disease control: influenza and antiviral agents.** *Epidemiol Infect* 2007, **135:**1124-1132.
88. Duizer E, Koopmans M: **Tracking foodborne viruses: lessons from noroviruses.** In *Emerging foodborne pathogens.* Edited by Motarjemi Y, Adams M. Boca Raton (FL): CRC Press; 2006: 77-110
89. Eames I, Tang JW, Li Y, Wilson P: **Airborne transmission of disease in hospitals.** *J R Soc Interface* 2009, **6:**S697-S702.
90. Eames KTD: **Modelling disease spread through random and regular contacts in clustered populations.** *Theor Popul Biol* 2008, **73:**104-111.
91. East IJ, Hamilton SA, Garner G: **Identifying areas of Australia at risk of H5N1 avian influenza infection from exposure to migratory birds: a spatial analysis.** *Geospat Health* 2008a, **2:**203-213.
92. East IJ, Hamilton SA, Sharp LA, Garner MG: **Identifying areas of Australia at risk for H5N1 avian influenza infection from exposure to nomadic waterfowl moving throughout the Australo-Papuan region.** *Geospat Health* 2008b, **3:**17-27.
93. Edmunds WJ, Gay NJ, Kretzschmar M, Pebody RG, Wachmann H: **The pre-vaccination epidemiology of measles, mumps and rubella in Europe: implications for modelling studies.** *Epidemiol Infect* 2000, **125:**635-650.
94. Edmunds WJ, Kafatos G, Wallinga J, Mossong JR: **Mixing patterns and the spread of close-contact infectious diseases.** *Emerg Themes Epidemiol* 2006, **3:**10.

95. Edmunds WJ, O'Callaghan CJ, Nokes DJ: **Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections.** *Proc R Soc Lond B* 1997, **264:**949-957.
96. Eichner M, Schwehm M, Duerr H-P, Brockmann SO: **The influenza pandemic preparedness planning tool InfluSim.** *BMC Infect Dis* 2007, **7:**17.
97. Eichner M, Schwehm M, Wilson N, Baker MG: **Small islands and pandemic influenza: potential benefits and limitations of travel volume reduction as a border control measure.** *BMC Infect Dis* 2009, **9:**160.
98. Eisenhart C, Wilson PW: **Statistical methods and control in bacteriology.** *Bacteriol Rev* 1943, **7:**57-137.
99. Epstein JM: *Nonlinear dynamics, mathematical biology, and social science.* Reading, MA: Addison-Wesley; 1997.
100. Erdös P, Rényi A: **On random graphs. I.** *Publicationes Mathematicae* 1959, **6:**290-297.
101. Eubank S: **Network based models of infectious disease spread.** *Jpn J Infect Dis* 2005, **58:**S9-S13.
102. Eubank S, Guclu H, Kumar VSA, Marathe MV, Srinivasan A, Toroczkal Z, Wang N: **Modelling disease outbreaks in realistic urban social networks.** *Nature* 2004, **429:**180-184.
103. Evans MR, Meldrum R, Lane W, Gardner D, Ribeiro CD, Gallimore CI, Westmoreland D: **An outbreak of viral gastroenteritis following environmental contamination at a concert hall.** *Epidemiol Infect* 2002, **129:**355-360.
104. Eynard F, Koch D, Birrer A, Boubaker K, Matter HC, Raeber P-A, Influenza Working Group (AGI): *Swiss influenza pandemic plan: Strategies and measures in preparation for an influenza pandemic.* January 2009 edn. Bern, CH: Bundesamt für Gesundheit (BAG); 2009.
105. Farman J: **Halocarbons, the ozone layer and the precautionary principle.** In *Late lessons from early warnings: the precautionary principle 1896-2000.* Edited by Harremoës P, Gee D, MacGarvin M, Stirling A, Keys J, Wynne B, Guedes Vaz S. Copenhagen: European Environmental Agency; 2001
106. Fasina FO, Sirdar MM, Bisschop SP: **The financial cost implications of the higliy pathogenic notifiable avian influenza H5N1 in Nigeria.** *Onderstepoort J Vet Res* 2008, **75:**39-46.
107. Feil M, Balmer M, Axhausen KW: **Generating comprehensive all-day schedules: expanding activitiy-based travel demand modelling.** In *European Transport Conference.* Leeuwenhorst; 2009.
108. Feinstein AH, Cannon HM: **A hermeneutical approach to external validation of simulation models.** *Simulat Gaming* 2003, **34:**186-197.
109. Ferguson NM: **Capturing human behavior.** *Nature* 2007, **446:**733.
110. Ferguson NM, Cummings DA, Fraser C, Cajka JC, Cooley PC, Burke DS: **Strategies for mitigating an influenza pandemic.** *Nature* 2006, **442:**448-452.
111. Ferguson NM, Cummings DAT, Cauchemez S, Fraser C, Riley S, Meeyai A, Iamsirithaworn S, Burke DS: **Strategies for containing an emerging influenza pandemic in Southeast Asia.** *Nature* 2005, **437:**209-214.

112. Ferrari MJ, Bjornstad ON, Dobson AP: **Estimation and inference of Ro of an infectious pathogen by a removal method.** *Math Biosci* 2005, **198:**14-26.
113. Fiebig L, Smieszek T, Zinsstag J: **Social patterns and communicable diseases: What epidemiology (still) can learn from SNA.** In *Applications of Social Network Analysis (ASNA)*. Zurich: University of Zurich; 2008.
114. Fields S, Johnston M: **Cell biology. Whither model organism research?** *Science* 2005, **307:**1885-1886.
115. Fishman JA, Rubin RH: **Infection in organ-transplant recipients.** *N Engl J Med* 1998, **338:**1741-1751.
116. Fox JP, Cooney MK, Hall CE, Foy HM: **Influenza virus infections in Seattle families, 1975-1979. II. Pattern of infection in invaded households and relation of age and prior antibody to occurrence of infection and related illness.** *Am J Epidemiol* 1982, **116:**228-242.
117. Fox JP, Hall CE, Cooney MK, Foy HM: **Influenza virus infections in Seattle families, 1975-1979. I. Study design, methods and the occurrence of infections by time and age.** *Am J Epidemiol* 1982, **116:**212-227.
118. Freeman LC: **A set of measures of centrality based on betweenness.** *Sociometry* 1977, **40:**35-41.
119. Frigg R, Hartmann S: **Models in science.** In *The Stanford Encyclopedia of Philosophy* (Zalta EN ed., summer 2009 edition; 2009.
120. Galea S, Ahern J: **Invited commentary: considerations about specificity of associations, causal pathways, and heterogeneity in multilevel thinking.** *Am J Epidemiol* 2006, **163:**1079-1082.
121. Gani R, Hughes H, Fleming D, Griffin T, Medlock J, Leach S: **Potential impact of antiviral drug use during influenza pandemic.** *Emerg Infect Dis* 2005, **11:**1355-1362.
122. Gani R, Leach S: **Transmission potential of smallpox in contemporary populations.** *Nature* 2001, **414:**748-751.
123. Garber L, Hill G, Rodriguez J, Gregory G, Voelker L: **Non-commercial poultry industries: surveys of backyard and gamefowl breeder flocks in the United States.** *Prev Vet Med* 2007, **80:**120-128.
124. Garnett GP, Holmes EC: **The ecology of emergent infectious disease.** *Biosci* 1996, **46:**127-135.
125. Garske T, Clarke P, Ghani AC: **The transmissibility of highly pathogenic avian influenza in commercial poultry in industrialized countries.** *PLoS One* 2007, **2:**e349.
126. Gauch HG: *Scientific method in practice.* Cambridge, UK: Cambridge University Press; 2003.
127. Germann TC, Kadau K, Longini IM, et al.: **From the cover: mitigation strategies for pandemic influenza in the United States.** *Proc Natl Acad Sci U S A* 2006, **103:**6146-6151.
128. Gerndt H: **Methodologische Untersuchung der Simulationsmethode: ein Ansatz für eine umfassende Definition.** *Doctoral dissertation.* TU Berlin, 1978.
129. Ghani AZ, Garnett GP: **Risk of acquiring and transmitting sexually transmitted diseases in sexual partner networks.** *Sex Transm Dis* 2000, **27:**579-587.
130. Ghedin E, Sengamalay NA, Shumway M, Zaborsky J, Feldblyum T, Subbu V, Spiro DJ, Sitz J, Koo H, Bolotov P, Dernovoy D, Tatusova T, Bao Y, St George K, Taylor J, Lipman DJ,

Fraser CM, Taubenberger JK, Salzberg SL: **Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution.** *Nature* 2005, **437:**1162-1166.

131. Giere R: **How models are used to represent reality.** *Phil Sci* 2004, **71 Supplement:**S742-S752.
132. Gilbert N, Troitzsch KG: *Simulation for the social scientist.* Buckingham: Open University Press; 1999.
133. Glass LM, Glass RJ: **Social contact networks for the spread of pandemic influenza in children and teenagers.** *BMC Public Health* 2008, **8:**61.
134. Godfrey K: *Compartmental models and their application.* London, UK: Academic Press; 1983.
135. Gottschalk F, Sonderer T, Scholz RW, Nowack B: **Modeled environmental concentrations of engineered nanomaterials (TiO_2, ZnO, Ag, CNT, fullerenes) for different regions.** *Environ Sci Technol* 2009, **43:**9216-9222.
136. Govaert TME, Dinant GJ, Aretz K, Knottnerus JA: **The predictive value of influenza symptomatology in elderly people.** *Fam Pract* 1998, **15:**16-22.
137. Grabkowsky B: **Wie sicher ist Ihr Betrieb vor der Geflügelpest?** [http://www.animal-health-online.de/drms/gef/analyse_ispa.pdf]
138. Granovetter M: **Threshold models of collective behavior.** *Amer J Soc* 1978, **83:**1420-1443.
139. Granovetter M: **The strength of weak ties: a network theory revisited.** *Sociol Theor* 1983, **1:**201-233.
140. Gregg MB: **The epidemiology of influenza in humans.** In *Annals of the New York Academy of Sciences: Airborne Contagion. Volume* 353. Edited by Kundsin RB. New York, NY: New York Academy of Sciences; 1980
141. Grenfell BT, Bolker BM: **Cities and villages: infection hierarchies in a measles metapopulation.** *Ecol Lett* 1998, **1:**63-70.
142. Grenfell BT, Harwood J: **(Meta)population dynamics of infectious diseases.** *TREE* 1997, **12:**395-399.
143. Guimarães Jr. PR, Argollo de Menezes M, Baird RW, Lusseau D, Guimarães P, dos Reis SF: **Vulnerability of a killer whale social network to disease outbreaks.** *Phys Rev E* 2007, **76**.
144. Haag D: **Models for the representation of ecological systems? The validity of experimental model systems and of dynamical simulation models as to the interaction with ecological systems.** *Doctoral dissertation.* University of Hohenheim (Germany), 2000.
145. Haas CN: **Estimation of risk due to low doses of microorganisms: a comparison of alternative methodologies.** *Am J Epidemiol* 1983, **118:**573-582.
146. Haas CN, Rose JB, Gerba CP: *Quantitative microbial risk assessment.* New York, NY: John Wiley & Sons; 1999.
147. Hales S, Kovats S, A. W: **What El Niño can tell us about human health and global climate change.** *Global Change Hum Health* 2000, **1:**66-77.
148. Hales S, Weinstein P, Souares Y, Woodward A: **El Niño and the dynamics of vectorborne disease transmission.** *Environ Health Perspect* 1999, **107:**99-102.
149. Hales S, Weinstein P, Woodward A: **Dengue fever epidemics in the south pacific: driven by El Niño Southern Oscillation.** *Lancet* 1996, **348:**1664-1665.

150. Hamer WH: **The Milroy lectures on epidemic disease in England - the evidence of variability and of persistency of type.** *Lancet* 1906**:**733-739.
151. Han K, Zhu X, He F, Liu L, Zhang L, Ma H, Tang X, Huang T, Zeng G, Zhu B-P: **Lack of airborne transmission during outbreak of pandemic (H1N1) 2009 among tour group members, China, June 2009.** *Emerg Infect Dis* 2009, **15:**1578-1581.
152. Hauser R, Breidenbach E, Jost U, Bachmann I: **Risikobasierte Massnahmen beim Auftreten von HPAI H5N1 bei Wildvögeln - Präventive Massnahmen in Risikogebieten** [http://www.bvet.admin.ch/gesundheit_tiere/00315/00317/02600/index.html?lang=de]
153. Hauser R, Schwermer H, Breidenbach E: **Risikoabschätzung der Einschleppung von Aviärer Influenza in den Bestand der in Gefangenschaft gehaltenen Vögel der Schweiz** [http://www.bvet.admin.ch/gesundheit_tiere/00315/00317/02600/index.html?lang=de]
154. Hays JN: *Epidemics and pandemics: Their impacts on human history*. Santa Barbara, CA: ABC-CLIO; 2005.
155. Heesterbeek H: **The law of mass-action in epidemiology: a historical perspective.** In *Ecological paradigms lost: routes of theory change*. Edited by Cuddington K, Beisner BE. Amsterdam: Elsevier Academic Press; 2005: 81-105
156. Hegenbart R: *Wörterbuch der Philosophie*. München: Humboldt; 1984.
157. Heikkinen T, Silvennoinen H, Peltola V, Ziegler T, Vainionpää R, Vuorinen T, Kainulainen L, Puhakka T, Jartti T, Toikka P, Lehtinen P, Routi T, Juven T: **Burden of influenza in children in the community.** *J Infect Dis* 2004, **190:**1369-1373.
158. Hens N, Ayele GM, Goeyvaerts N, Aerts M, Mossong J, Edmunds JW, Beutels P: **Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries.** *BMC Infect Dis* 2009, **9:**187.
159. Hersh R: *What is mathematics, really?* Oxford, UK: Oxford University Press; 1997.
160. Hethcote HW, Yorke JA: *Gonorrhea transmission dynamics and control*. Berlin: Springer; 1984.
161. Heymann A, Chodick G, Reichman B, Kokia E, Laufer J: **Influence of school closure on the incidence of viral respiratory diseases among children and on health care utilization.** *Pediatr Infect Dis J* 2004, **23:**675-677.
162. Heymann DL: *Control of communicable disease manual*. 18th edn. Washington, D.C.: American Public Health Association; 2004.
163. Hofmann MA, Renzullo S, Baumer A: **Phylogenetic characterization of H5N1 highly pathogenic avian influenza viruses isolated in Switzerland in 2006.** *Virus Genes* 2008, **37:**407-413.
164. Holland JH: **Studying complex adaptive systems.** *J Syst Sci Complexity* 2006, **19:**1-8.
165. Holm MV, Blank PR, Szucs TD: **Trends in influenza vaccination coverage rates in Germany over five seasons from 2001 to 2006.** *BMC Infect Dis* 2007, **7:**144.
166. Holsinger KE: **Comment: the blunting of Occam's Razor, or to hell with parsimony.** *Can J Zool* 1981, **59:**144-146.

167. Hopp M, Foley J: **Worldwide fluctuations in dengue fever cases related to climate variability.** *Clim Res* 2003, **25:**85-94.
168. Imboden DM, Koch S: *Systemanalyse: Einführung in die mathematische Modellierung natürlicher Systeme.* Berlin: Springer; 2003.
169. Johnson S: *Emergence: the connected lives of ants, brains, cities, and software.* New York, NY: Scribner; 2001.
170. Judson OP: **The rise of the individual-based model in ecology.** *TREE* 1994, **9:**9-14.
171. Kant I: *Kritik der reinen Vernunft.* Hamburg: Felix Meiner Verlag; 1971, orig. 1781/1787.
172. Karlqvist A: **Going beyond disciplines: the meaning of interdisciplinarity.** *Policy Science* 1999, **32:**379-383.
173. Keech M, Beardsworth P: **The impact of influenza on working days lost. A review of the literature.** *Pharmacoeconomics* 2008, **26:**911-924.
174. Keeling MJ: **The effects of local spatial structure on epidemiological invasions.** *Proc R Soc Lond B* 1999, **266:**859-867.
175. Keeling MJ: **Extensions to mass-action mixing.** In *Ecological paradigms lost: routes of theory change.* Edited by Cuddington K, Beisner BE. Amsterdam: Elsevier Academic Press; 2005: 107-142
176. Keeling MJ, Grenfell BT: **Disease extinction and community size: Modeling the persistence of measles.** *Science* 1997, **275:**65-67.
177. Keeling MJ, Grenfell BT: **Individual-based perspectives on R-0.** *J Theor Biol* 2000, **203:**51-61.
178. Keeling MJ, Rohani P: *Modeling infectious diseases in humans and animals.* Princeton, NJ: Princeton; 2008.
179. Keiser J, Caldas de Castro M, Maltese MF, Bos R, Tanner M, Singer BH, Utzinger J: **Effect of irrigation and large dams on the burden of malaria on a global and regional scale.** *Am J Trop Med Hyg* 2005, **72:**392-406.
180. Kermack WO, McKendrick AG: **A contribution to the mathematical theory of epidemics.** *Proc R Soc Lond A* 1927, **115:**700-721.
181. Kernen T: **Survey among Swiss poultry holders in the Lake of Constance region in relation to a potential spread of avian influenza.** *MSc thesis.* University of Basel, 2008.
182. Khan AS, Tshioko FK, Heymann DL, Le Guenno B, Nabeth P, Kerstiens B, Fleerackers Y, Kilmarx PH, Rodier GR, Nkuku O, Rollin PE, Sanchez A, Zaki SR, Swanepoel R, Tomori O, Nichol ST, Peters CJ, Muyembe-Tamfum JJ, Ksiazek TG: **The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995. Commission de Lutte contre les Epidemies a Kikwit.** *J Infect Dis* 1999, **179 Suppl 1:**S76-86.
183. Kilpatrick AM, Chmura AA, Gibbons DW, Fleischer R, Marra PP, Daszak P: **Predicting the global spread of H5N1 avian influenza.** *Proc Natl Acad Sci U S A* 2006, **103:**19368-19373.
184. Kish L: **Chance, statistics, and statisticians.** *J Am Stat Assoc* 1978, **73:**1-6.
185. Kiss IZ, Green DM, Kao RR: **Infectious disease control using contact tracing in random and scale-free networks.** *J R Soc Interface* 2006, **3:**55-62.
186. Klein JT: *Interdisciplinarity: history, theory, and practice.* Detroit, MI: Wayne State University Press; 1990.

187. Kleindorfer GB, O'Neill L, Ganeshan R: **Validation in simulation: various positions in the philosophy of Science.** *Manage Sci* 1998, **44:**1087-1099.
188. Klopfleisch R, Wolf PU, Wolf C, Harder T, Starick E, Niebuhr M, Mettenleiter TC, Teifke JP: **Encephalitis in a stone marten (*Martes foina*) after natural infection with highly pathogenic avian influenza virus subtype H5N1.** *J Comp Pathol* 2007, **137**.
189. Koopman JS: **Modeling infection transmission - the pursuit of complexities that matter.** *Epidemiology* 2002, **13:**622-624.
190. Koopman JS: **Modeling infection transmission.** *Annu Rev Public Health* 2004, **25:**303-326.
191. Koopman JS: **Infection transmission science and models.** *Jpn J Infect Dis* 2005, **58:**S3-S8.
192. Koopman JS: **Mass-action and system analysis of infection transmission.** In *Ecological paradigms lost: routes of theory change.* Edited by Cuddington K, Beisner BE. Amsterdam: Elsevier Academic Press; 2005: 143-155
193. Koppinen J: **How to finance animal health?** *Aust Vet J* 2005, **83:**190.
194. Kroon FP, van Dissel JT, de Jong JC, van Furth R: **Antibody response to influenza, tetanus and pneumococcal vaccines in HIV-seropositive individuals in relation to the number of CD4+ lymphocytes.** *AIDS* 1994, **8:**469-476.
195. Kuhn TS: **Objectivity, value judgment, and theory choice.** In *The essential tension: selecte studies in scientific tradition and change.* Edited by Kuhn TS. Chicago, IL: University of Chicago Press; 1977: 320-339
196. Kuiken T, Rimmelzwaan GF, van Riel D, van Amerongen G, Baars M, Fouchier R, Osterhaus AD: **Avian H5N1 influenza in cats.** *Science* 2004, **306:**241.
197. Kuperman M, Abramson G: **Small world effects in an epidemiological model.** *Phys Rev Lett* 2001, **86:**2909-2912.
198. Lange W, Vogel GE: *Influenza: Klinik, Virologie, Epidemiologie, Therapie und Prophylaxe.* Berlin: ABW Wissenschaftsverlag; 2004.
199. Legrand J, Grais RF, Boelle PY, Valleron AJ, Flahault A: **Understanding the dynamics of Ebola epidemics.** *Epidemiol Infect* 2007, **135:**610-621.
200. Lemieux C, Brankston G, Gitterman L, Hirji Z, Gardam M: **Questioning aerosol transmission of influenza.** *Emerg Infect Dis* 2007, **13:**173-174.
201. Li S, Eisenberg JNS, Spicknall IH, Koopman JS: **Dynamics and control of infections transmitted from person to person through the environment.** *Am J Epidemiol* 2009, **170:**257-265.
202. Li Y, Leung GM, Tang JW, Yang X, Chao CYH, Lin JZ, Lu JW, Nielsen PV, Niu J, Qian H, Sleigh AC, Su H-JJ, Sundell J, Wong TW, Yuen PL: **Role of ventilation in airborne transmission of infectious agents in the built environment - a multidisciplinary systematic review.** *Indoor Air* 2007, **17:**2-18.
203. Liljeros F, Edling CR, Amaral LAN, Stanley HE, Aberg Y: **The web of human sexual contacts.** *Nature* 2001, **411:**907-908.
204. Liljeros F, Giesecke J, Holme P: **The contact network of inpatients in a regional healthcare system. A longitudinal case study.** *Math Popul Stud* 2007, **14:**269-284.
205. Lloyd AL: **Realistic Distributions of Infectious Periods in Epidemic Models: Changing Patterns of Persistence and Dynamics.** *Theor Popul Biol* 2001, **60:**59-71.
206. Lloyd AL, May RM: **Spatial heterogeneity in epidemic models.** *J Theor Biol* 1996, **179:**1-11.

207. Long FA: **Interdisciplinary problem-oriented research in the university.** *Science* 1971, **171:**961.
208. Longini IM, Halloran ME, Nizam A, et al.: **Containing pandemic influenza with antiviral agents.** *Am J Epidemiol* 2004, **159:**623-633.
209. Lotka AJ: *Elements of physical biology*. Baltimore, MD: Williams & Wilkins; 1925.
210. Lyytikäinen T, Kallio ER, Sahlström L, Virtanen T: **Consequences of networks in epidemiological predictions of FMD.** In *Society for veterinary epidemiology and preventive medicine (SVEPM) proceedings 2009*. Edited by Peeler EJ, Alban L, Russell A, and the SVEPM Executive Committee. London, UK; 2009: 254-266
211. Mallik MAB, Tesfai K, Pancholy SK: **Formation of carcinogenic nitrosamines in soil treated with pesticides, and in sewage amended with nitrogen compounds.** *Proc Okla Acad Sci* 1981, **61:**31-35.
212. Mannelli A, Busani L, Toson M, Bertolini S, Marangon S: **Transmission parameters of highly pathogenic avian influenza (H7N1) among industrial poultry farms in northern Italy in 1999-2000.** *Prev Vet Med* 2007, **81:**318-322.
213. Markie P: **Rationalism vs. empiricism.** In *The Stanford Encyclopedia of Philosophy* (Zalta EN ed., fall 2008 edition; 2008.
214. Mathews JD, Chesson JM, McCaw JM, McVernon J: **Understanding influenza transmission, immunity and pandemic threats.** *Influenza Other Respi Viruses* 2009, **3:**143-149.
215. Mayring P: *Qualitative Inhaltsanalyse - Grundlagen und Techniken*. Weinheim, DE: Beltz UTB; 2003.
216. McAvaney BJ, Covey C, Joussaume S, Kattsov V, Kitoh A, Ogana W, Pitman AJ, Weaver AJ, Wood RA, Zhao Z-C: **Model evaluation.** In *Climate change 2001: the scientific basis Contribution of working group I to the third assessment report of the intergovernmental panel on climate change*. Edited by Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA. Cambridge, UK: Cambridge University Press: 471-524
217. McKendrick AG: **The rise and fall of epidemics.** *Paludism* 1912, **1:**54-66.
218. McKenzie V, J., Townsend AR: **Parasitic and infectious disease responses to changing global nutrient cycles.** *EcoHealth* 2007, **4:**384-396.
219. Meyers LA: **Contact network epidemiology: Bond percolation applied to infectious disease prediction and control.** *Bull Am Math Soc* 2007, **44:**6.
220. Meynell GG, Stocker BAD: **Some hypotheses on the aetiology of fatal infections in partially resistant hosts and their application to mice challenged with *Salmonella paratyphi-B* or *Salmonella typhimurium* by intraperitoneal injection.** *J Gen Microbiol* 1957, **16:**38-58.
221. Mikolajczyk RT, Akmatov MK, Rastin S, Kretzschmar M: **Social contacts of school children and the transmission of respiratory-spread pathogens.** *Epidemiol Infect* 2008, **136:**813-822.
222. Miller R, C.: **Varieties of interdisciplinary approaches in the social sciences: a 1981 overview.** *Issues Integrative Stud* 1982, **1:**1-37.

223. Mills CE, Robins JM, Lipsitch M: **Transmissibility of 1918 pandemic influenza.** *Nature* 2004, **432:**904-906.
224. Molina MJ, Rowland FS: **Stratospheric sink for cholofluoromethanes: chlorine atomic-catalysed destruction of ozone.** *Nature* 1974, **249:**810-812.
225. Molinari NAM, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E: **The annual impact of seasonal influenza in the US: Measuring disease burden and costs.** *Vaccine* 2007, **25:**5086-5096.
226. Molyneux DH, Ostfeld RS, Bernstein A, Chivian E: **Ecosystem disturbance, biodiversity loss, and human infectious disease.** In *Sustaining life: How human health depends on biodiversity*. Edited by Chivian E, Bernstein A. Oxford, UK: Oxford University Press; 2008: 287-323
227. Monto AS, Gravenstein S, Elliott M, Colopy M, Schweinle J: **Clinical signs and symptoms prediction influenza infection.** *Arch Intern Med* 2000, **160:**3243-3247.
228. Monto AS, Koopman JS, Longini I: **Tecumseh study of illness. XIII. Influenza infection and disease, 1976-1981.** *Am J Epidemiol* 1985, **121:**811-822.
229. Monto AS, Sullivan KM: **Acute respiratory illness in the community. Frequency of illness and the agents involved.** *Epidemiol Infect* 1993, **110:**145-160.
230. Morgan MS: **Imagination and imaging in model building.** *Phil Sci* 2004, **71:**753-766.
231. Morris M: **Epidemiology and social networks: modeling structured diffusion.** *Sociol Method Res* 1993, **22:**99-126.
232. Morris M (Ed.). **Network epidemiology: a handbook for survey design and data collection**. Oxford, UK: Oxford University Press; 2004.
233. Morrison LG, Yardley L: **What infection control measures will people carry out to reduce transmission of pandemic influenza? A focus group study.** *BMC Public Health* 2009, **9**.
234. Moser MR, Bender TR, Margolis HS, Noble GR, Kendal AP, Ritter DG: **An outbreak of influenza aboard a commercial airliner** *Am J Epidemiol* 1979, **110:**1-6.
235. Mosler H-J, Tobias R: **Simulation und Modellierung.** In *Enzyklopädie der Psychologie, Band1: Grundlagen, Paradigmen und Methoden der Umweltpsychologie*. Edited by Lantermann E-D, Linneweber V. Göttingen: Hogrefe; 2008: 899-928
236. Moss WJ, Griffin DE: **Global measles elimination.** *Nat Rev Microbiol* 2006, **4:**900-908.
237. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ: **Social contacts and mixing patterns relevant to the spread of infectious diseases.** *PLoS Med* 2008, **5:**381-391.
238. Mossong J, Muller CP: **Estimation of the basic reproduction number of measles during an outbreak in a partially vaccinated population.** *Epidemiol Infect* 2000, **124:**273-278.
239. Müller-Herold U, Morosini M, Schucht O: **Choosing chemicals for precautionary regulation: a filter series approach.** *Environ Sci Technol* 2005, **39:**683-691.
240. Müller-Herold U, Smieszek T, Peter P, Scheringer M, Morosini M: **A simple measure for precautionary assessment of organic chemicals with respect to global cold condensation.** *Ecol Model* 2006, **194:**266-273.

241. Müller R: **Zur Geschichte des Modelldenkens und des Modellbegriffs.** In *Modelle - Konstruktion der Wirklichkeit.* Edited by Stachowiak H. München: Wilhelm Fink; 1983: 17-86
242. Murray JD: **Simple epidemic models and practical applications.** In *Mathematical biology I: An introduction.* 3 rd edition. Edited by Murray JD. New York, NY: Springer; 2002: 319-327
243. Nagelkerke NJD, Moses S, de Vlas SJ, Bailey RC: **Modelling the public health impact of male circumcision for HIV prevention in high prevalence areas in Africa.** *BMC Infect Dis* 2007, **7:**16.
244. Newman MEJ: **Spread of epidemic disease on networks.** *Phys Rev E* 2002, **66:**016128.
245. Nicas M, Sun G: **An integrated model of infection risk in a health-care environment.** *Risk Analysis* 2006, **26:**1085-1096.
246. Nicholson KG, Wood JM, Zambon M: **Influenza.** *Lancet* 2003, **362:**1733-1745.
247. Nishiura H: **Time variations in the transmissibility of pandemic influenza in Prussia, Germany, from 1918-19.** *Theor Biol Med Model* 2007, **4:**20.
248. Nishiura H, Brockmann SO, Eichner M: **Extracting key information from historical data to quantify the transmission dynamics of smallpox.** *Theor Biol Med Model* 2008, **5:**20.
249. Nishiura H, Chowell G, Safan M, Castillo-Chavez C: **Pros and cons of estimating the reproduction number from early epidemic growth rate of influenza A (H1N1) 2009.** *Theor Biol Med Model* 2010, **7:**1.
250. Nishiura H, Wilson N, Baker MG: **Quarantine for pandemic influenza control at the border of small island nations.** *BMC Infect Dis* 2009, **9:**27.
251. Nordvik MK, Liljeros F: **Transmission of sexually transmitted infections.** *Sex Transm Dis* 2006, **33:**342-349.
252. Normile D: **Avian influenza. Are wild birds to blame?** *Science* 2005, **310:**426-428.
253. O'Neill RV: **A review of linear compartmental analysis in ecosystem science.** In *Compartmental analysis of ecosystem models.* Edited by Matis JH, Patten BC, White GC. Fairland, MD: International Co-operative Publishing House; 1979: 3-27
254. Ohkusa Y, Sugawara T: **Simulation model of pandemic influenza in the whole of Japan.** *Jpn J Infect Dis* 2009, **62:**98-106.
255. Oreskes N, Shrader-Frechette K, Belitz K: **Verification, validation, and confirmation of numerical models in the earth sciences.** *Science* 1994, **263:**641-644.
256. Ortúzar JdD, Willumsen LG: *Modelling transport.* 3rd edn. Chichester, UK: John Wiley & Sons; 2001.
257. Oswald A: *Meister der Miniaturen: Architekturmodelle.* 1st edn. Berlin: DOM Publishers; 2008.
258. Ou C, Wu J: **Spatial spread of rabies revisited: Influence of age-dependent diffusion on nonlinear dynamics.** *SIAM J Appl Math* 2006, **67:**138-163.
259. Oyok T, Odonga C, Mulwani E, Abur J, Kaducu F, Akech M, Olango J, Onek P, Turyanika J, Mutyaba I, Luwaga HRS, Bisoborwa G, Kaguna A, Omaswa FG, Zaramba S, Okware S, Opio A, Amandua J, Kamugisha J, Mukoyo E, Wanyana J, Mugero C, Lamunu M, Mugaga M, Kiyonga C: **Outbreak of ebola Hemorrhagic Fever - Uganda, Augsut 2000-January 2001.** *MMWR* 2001, **50:**73-77.

260. Paget J, Marquet R, Meijer A, van der Velden K: **Influenza activity in Europe during eight seasons (1999-2007): an evaluation of the indicators used to measure activity and an assessment of the timing, length and course of peak activity (spread) across Europe.** *BMC Infect Dis* 2007, **7:**141.
261. Pantelic J, Sze-To GN, Tham KW, Chao CYH, Khoo YCM: **Personalized ventilation as a control measure for airborne transmissible disease spread.** *J R Soc Interface* 2009, **6:**S715-S726.
262. Pascual M, Dobson AP: **Seasonal patterns of infectious diseases.** *PLoS Med* 2005, **2:**e5.
263. Pastor-Satorras R, Vespignani A: **Epidemic spreading in scale-free networks.** *Phys Rev Lett* 2001, **86:**3200-3203.
264. Peltola V, Reunanen T, Ziegler T, Silvennoinen H, Heikinnen T: **Accuracy of clinical diagnosis of influenza in outpatient children.** *Clin Inf Dis* 2005, **41:**1198-1200.
265. Persad G, Wertheimer A, Emanuel EJ: **Principles for allocation of scarce medical interventions.** *Lancet* 2009, **373:**423-431.
266. Planck M: **Über das Gesetz der Energieverteilung im Normalspectrum.** *Verh Dtsch Phys Ges* 1900, **2:**237-245.
267. Public Health Agency of Canada OoLS: **Material safety data sheet: Staphylococcus aureus** [http://www.phac-aspc.gc.ca/msds-ftss/msds143e.html]
268. Pujol JM, Eisenberg JE, Haas CN, Koopman JS: **The effect of ongoing exposure dynamics in dose response relationships.** *PLoS Comput Biol* 2009, **5**.
269. Råberg L, Grahn M, Hasselquist D, Svensson E: **On the adaptive significance of stress-induced immunosuppression.** *Proc R Soc Lond B* 1998, **265:**1637-1641.
270. Rabinowitz P, Gordon Z, Chudnov D, Wilcox M, Odofin L, Liu A, Dein J: **Animals as sentinels of bioterrorism agents.** *Emerg Infect Dis* 2006, **12:**647-652.
271. Rabinowitz P, Scotch M, Conti L: **Human and animal sentinels for shared health risks.** *Vet Ital* 2009, **45:**23-34.
272. Rakowski F, Gruziel M, Krych M, Radomski JP: **Large scale daily contacts and mobility model - an individual-based countrywide simulation study for Poland.** *JASSS-J Artif Soc S* 2010, **13:**13.
273. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC: **The genomic and epidemiological dynamics of human influenza A virus.** *Nature* 2008, **453:**615-620.
274. Rapport DJ: **What constitutes ecosystem health?** *Perspect Biol Med* 1989, **33:**120-132.
275. Rapport DJ: **Epidemiology and ecosystem health: natural bridges.** *Ecosystem Health* 1999, **5:**174-180.
276. Rashdall H: *The universities of Europe in the middle ages: Salerno - Bologna - Paris* Oxford, UK: Clarendon Press; 1895.
277. Ray KJ, Porco TC, Hong KC, Lee DC, Alemayehu W, Melese M, Lakew T, Yi E, House J, Chidambaram JD, Whitcher JP, Gaynor BD, Lietman TM: **A rationale for continuing mass antibiotic distributions for trachoma.** *BMC Infect Dis* 2007, **7:**91.
278. Read JM, Eames KTD, Edmunds WJ: **Dynamic social networks and the implications for the spread of infectious disease.** *J R Soc Interface* 2008, **5:**1001-1007.
279. Reed KD, Meece JK, Henkel JS, Shukla SK: **Birds, migration and emerging zoonoses: west nile virus, lyme disease, influenza A and enteropathogens.** *Clin Med Res* 2003, **1:**5-12.

280. Reeves WC, Hardy JL, Reisen W, Milby MM: **Potential effect of global warming on mosquito-borne arboviruses.** *J Med Entomol* 1994, **31:**323-332.
281. Reynolds O: **An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels.** *Phil Trans R Soc Lond* 1883, **174:**935-982.
282. Riley EC, Murphy G, Riley RL: **Airborne spread of measles in a suburban elementary school.** *Am J Epidemiol* 1978, **107:**421-432.
283. Riley S, Fraser C, Donnelly CA, Ghani AC, Abu-Raddad LJ, Hedley AJ, Leung GM, Ho LM, Lam TH, Thach TQ, Chau P, Chan KP, Leung PY, Tsang T, Ho W, Lee KH, Lau EMC, Ferguson NM, Anderson RM: **Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions.** *Science* 2003, **300:**1961-1966.
284. Robinson SE, Christley RM: **Identifiying temporal variation in reported births, deaths and movements of cattle in Britain.** *BMC Vet Res* 2006, **2:**11.
285. Rodriguez CV, Rietberg K, Baer A, Kwan-Gett T, Duchin J: **Association between school closure and subsequent absenteeism during a seasonal influenza epidemic.** *Epidemiology* 2009, **20:**787-792.
286. Rubin LG: **Bacterial colonization and infection resulting from multiplication of a single organism.** *Rev Infect Dis* 1987, **9:**488-493.
287. Rudnick SN, Milton DK: **Risk of indoor airborne infection transmission estimated from carbon dioxide concentration.** *Indoor Air* 2003, **13:**237-245.
288. Rüegg W: *A history of the university in Europe.* Cambridge: Cambridge University Press; 1992.
289. Russell S: **The economic burden of illness for households in developing countries: A review of studies focusing on malaria, tuberculosis, and human immunodeficiency virus/acquired immunodeficiency syndrome.** *Am J Trop Med Hyg* 2004, **71:**147-155.
290. Sabidussi G: **The centrality index of a graph.** *Psychometrika* 1966, **31:**581-603.
291. Sandberg IW: **On the mathematical foundations of compartmental analysis in biology, medicine, and ecology.** *IEEE T Circ Syst* 1978, **CAS-25:**273-279.
292. Santa Fe Institute (SFI): **Santa Fe Institute** [http://www.santafe.edu]
293. Sauerbrei A, Schmidt-Ott R, Hoyer H, Wutzler P: **Seroprevalence of influenza A and B in German infants and adolescents.** *Med Microbiol Immunol* 2009, **198:**93-101.
294. Saurina J, Fiebig L, Zinsstag J, Schelling E: **Disease awareness of the poultry keepers in Switzerland and their access to information concerning highly pathogenic avian influenza.** *Schw Arch Tierheilkd* in press.
295. Schafer A, Victor DG: **Global passenger travel: implications for carbon dioxide emissions.** *Energy* 1999, **24:**657-679.
296. Schlich R, Axhausen KW: **Habitual travel behaviour: evidence from a six-week travel diary.** *Transportation* 2003, **30:**13-36.
297. Scholz RW: *Environmental literacy in science and society: From knowledge to decisions.* Cambridge: Cambridge University Press; forthcoming.
298. Scholz RW, Tietje O: *Embedded case study methods: Integrating quantitative and qualitative knowledge.* Thousand Oaks, CA: Sage Publications; 2002.

299. Schotthoefer AM, Labak KM, Beasley VR: ***Ribeiroia ondatrae*** **cercariae are consumed by aquatic invertebrate predators.** *J Parasitol* 2007, **93:**1240-1243.
300. Schrag SJ, Wiener P: **Emerging infectious diseases: what are the relative roles of ecology and evolution?** *Trends Ecol Evol* 1995, **10:**319-324.
301. Schwabe CW: *Veterinary medicine and human health.* 3rd edn. Baltimore, MD: Williams & Wilkins; 1984.
302. Schweitzer F: *Brownian agents and active particles. Collective dynamics in the natural and social sciences.* Berlin: Springer; 2003.
303. Schweitzer F: **Coordination of decisions in a spatial model of Brownian Agents.** In *The complex dynamics of economic interaction.* Edited by Gallegati M, Kirman AP, Marsili M. Berlin: Springer; 2004: 303-318
304. Schweizerischer Bundesrat: **Verordnung über vorsorgliche Sofortmassnahmen zur Verhinderung der Einschleppung der Klassischen Geflügelpest vom 21. Oktober 2005** [http://www.iffwil.ch/pdf/dl.php?pdf=Vogelgrippe_Verordnung.pdf]
305. Schweizerischer Bundesrat: **Obligatorische Registrierung aller Pferde-, Fisch-, Bienen- und Geflügelhaltungen** [http://www.news.admin.ch/message/?lang=de&msg-id=28490]
306. Seale H, Leask J, Po K, MacIntyre CR: **"Will they just pack up and leave?" - attitutes and intended behaviour of hospital health care workers during an influenza pandemic.** *BMC Health Serv Res* 2009, **9**.
307. Selhorst T: **Improvement of emergency vaccination strategies against rabies in red fox (*Vulpes vulpes*) populations using a combination of cellular automata and evolutionary algorithm.** In *The Fourth International Workshop on Environmental Applications of Machine Learning (EAML 2004); Bled, Slovenia*. Edited by Dzeroski S, Zenko B, Debeljak M. Jozef Stefan Institute, Ljubljana; 2004: 55-56.
308. Sertsou G, Wilson N, Baker M, Nelson P, Roberts MG: **Key transmission parameters of an institutional outbreak during the 1918 influenza pandemic estimated by mathematical modelling.** *Theor Biol Med Model* 2006, **3:**38.
309. Shannon CE: **A mathematical theory of communication.** *Bell Syst Tech J* 1948, **27:**379-423, 623-656.
310. Sharma VP: **DDT: the fallen angel.** *Curr Sci* 2003, **85:**1532-1537.
311. Shimbel A: **Structural parameters of communication networks.** *Bull Math Biophy* 1953, **15:**501-507.
312. Shirley MDF, Rushton SP: **The impacts of network topology on disease spread.** *Ecol Complexity* 2005, **2:**287-299.
313. Smieszek T: **A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease spread.** *Theor Biol Med Model* 2009, **6:**25.
314. Smieszek T, Fiebig L, Scholz RW: **Models of epidemics: when contact repetition and clustering should be included.** *Theor Biol Med Model* 2009, **6:**11.
315. Snijders TAB: **The statistical evaluation of social network dynamics.** *Sociol Methodol* 2001, **31:**361-395.
316. Soper HE: **The interpretation of periodicity in disease prevalence.** *J R Stat Soc* 1929, **92:**34-61; 62-73.

317. Sørensen JH, Mackay DKJ, Jensen CO, Donaldson AI: **An integrated model to predict the atmospheric spread of foot-and-mouth disease virus.** *Epidemiol Infect* 2000, **124:**577-590.

318. Southgate VR: **Schistosomiasis in the Senegal River Basin: Before and after the construction of the dams at Diama, Senegal and Manantali, Mali and future prospects.** *J Helmintol* 1997, **71:**125-132.

319. Stachowiak H: **Erkenntnisstufen zum systematischen Neopragmatismus und zur allgemeinen Modelltheorie.** In *Modell - Konstruktionen der Wirklichkeit.* Edited by Stachowiak H. München: Wilhelm Fink; 1983: 87-146

320. Stanislaw H: **Tests of computer simulation validity: what do they measure?** *Simulat Gaming* 1986, **17:**173-191.

321. Stegeman A, Bouma A, Elbers ARW, de Jong MCM, Nodelijk G, de Klerk F, Koch G, van Boven M: **Avian influenza A viurs (H/N/) epidemic and effectiveness of control measures.** *J Infect Dis* 2004, **190:**2088-2095.

322. Steiner M, Vermeulen LC, Mullahy J, Hayney MS: **Factors influencing decisions regarding influenza vaccination and treatment: A survey of healthcare workers.** *Infect Cont Hosp Ep* 2002, **23:**625-627.

323. Stewart JJ: **Parameters of influenza aerosol transmission.** *Comment Theor Biol* 2002, **7:**445-457.

324. Swoboda WW: **Disciplines and interdisciplinarity: a historical perspective.** In *Interdisciplinarity and higher education.* Edited by Kockelmans JJ. University Park, PA: Pennsylvania State University Press; 1979

325. Szendrói B, Csányi G: **Polynomial epidemics and clustering in contact networks.** *Proc R Soc Lond B* 2004, **271:**S364-S366.

326. Szucs TD, Müller D: **Influenza vaccination coverage in five European countries - a population-based cross-sectional analysis of two consecutive influenza seasons.** *Vaccine* 2005, **23:**5055-5063.

327. Tang JW: **The effect of environmental parameters on the survival of airborne infectious agents.** *J R Soc Interface* 2009, **6:**S737-S746.

328. Tanner M, Zinsstag J: **"One health" - The potential of closer collaboration between human and animal health.** *Berl Munch Tierarztl* 2009, **122:**410-411.

329. Teller P: **Twilight of the perfect model.** *Erkenntnis* 2001, **55:**393-415.

330. Tellier R: **Review of aerosol transmission of influenza A virus.** *Emerg Infect Dis* 2006, **12:**1657-1662.

331. Teorell T: **Kinetics of distribution of substances administered to the body. I - the extravascular routes of administration.** *Arch Int Pharmacodyn Thér* 1937, **57:**205-225

332. Thomas ME, Bouma A, Ekker HM, Fonken AJ, Stegeman JA, Nielen M: **Risk factors for the introduction of high pathogenicity avian influenza virus into poultry famrs during the epidemic in the Netherlands in 2003.** *Prev Vet Med* 2005, **69:**1-11.

333. Tong SL, Hu WB, McMichael AJ: **Climate variability and Ross River virus transmission in Townsville region, Australia, 1985-1996.** *Trop Med Int Health* 2004, **9:**298-304.

334. Truscott J, Garske T, Chis-Ster I, Guitian J, Pfeiffer D, Snow L, Wilesmith J, Ferguson NM, Ghani AC: **Control of a highly pathogenic H5N1 avian influenza outbreak in the GB poultry flock.** *Proc Biol Sci* 2007, **274:**2287-2295.
335. van Fraassen BC: *The scientific image.* Oxford (UK): Oxford University Press; 1980.
336. van Steenbergen JE, van den Hof S, Langendam MW, van de Kerkhof JHTC, Ruijs WLM: **Measles Outbreak - Netherlands, April 1999 - January 2000.** *MMWR* 2000, **49:**299-303.
337. Vanderpasa J, Louisa J, Reynders M, Mascarta G, Vandenberg O: **Mathematical model for the control of nosocomial norovirus.** *J Hosp Inf* 2009, **71:**214-222.
338. Vickers DM, Osgood ND: **A unified framework of immunological and epidemiological dynamics for the spread of viral infections in a simple network-based population.** *Theor Biol Med Model* 2007, **4:**49.
339. Volterra V: **Variazioni e fluttuazioni del numero d'individui in specie animali conviventi.** *Mem R Accad Naz dei Lincei S VI* 1926, **2:**31-113.
340. von Bertalanffy L: **General system theory.** *Main Curr Mod Though* 1955, **11:**75-83.
341. Vrtic M, Fröhlich P, Axhausen KW: **Schweizerische Netzmodelle für Strassen- und Schienenverkehr.** In *Jahrbuch 2002/2003 Schweizerische Verkehrswirtschaft.* Edited by Bieger T, Lässer C, Maggi R. St Gallen: Verkehrswissenschaftliche Gesellschaft; 2003: 119-140
342. Waber BN, Olguín Olguín D, Kim T, Mohan A, Ara K, Pentland A: **Organizational engineering using sociometric badges.** In *NetSci 2007.* New York City, NY; 2007.
343. Wainwright J, Mulligan M: **Modelling human decision-making.** In *Environmental modelling: finding simplicity in complexity.* Edited by Wainwright J, Mulligan M. Chichester, UK: John Wiley & Sons; 2004: 225-244
344. Wallinga J, Levy-Bruhl D, Gay NJ, Wachmann CH: **Estimation of measles reproduction ratios and prospects for elimination of measles by vaccination in some Western European countries.** *Epidemiol Infect* 2001, **127:**281-295.
345. Wang J, McMichael AJ, Meng B, Becker NG, Han W, Glass K, Wu J, Liu X, Liu J, Li X, Zheng X: **Spatial dynamics of an epidemic of severe acute respiratory syndrome in an urban area.** *Bull World Health Organ* 2006, **84:**965-968.
346. Ward MP, Maftei D, Apostu C, Suru A: **Geostatistical visualisation and spatial statistics for evaluation of the dispersion of epidemic highly pathogenic avian influenza subtype H5N1.** *Vet Res* 2008, **39:**22.
347. Wasserman S, Faust K: *Social network analysis: methods and applications.* Cambridge, MA: Cambridge University Press; 1994.
348. Watts DJ, Strogatz SH: **Collective dynamics of 'small-world' networks.** *Nature* 1998, **393:**440-442.
349. Wearing HJ, Rohani P, Keeling MJ: **Appropriate models for the management of infectious diseases.** *PLoS Med* 2005, **2:**e174.
350. Webster RG, Peiris M, Chen HL, Guan Y: **H5N1 outbreaks and enzootic influenza.** *Emerg Infect Dis* 2006, **12:**3-8.
351. Weinert F: **Theories, models and constraints.** *Stud Hist Phil Sci* 1995, **30:**303-333.
352. Weinhold B: **Conservation medicine: combining the best of all worlds.** *Environ Health Perspect* 2003, **111:**A525-A529.

353. Weiss G: **Prologue.** In *Multiagent systems: a modern approach to distributed artificial intelligence*. Edited by Weiss G. Cambridge, MA: The MIT Press; 2000: 1-26
354. Wells WF: *Airborne contagion and air hygiene. An ecological study of droplet infections*. Cambridge, MA: Harvard University Press; 1955.
355. Wernecke J: *Denken im Modell: Theorie und Erfahrung im Paradigma eines pragmatischen Modellbegriffs*. Berlin: Duncker & Humbolt; 1994.
356. Wheeler CM: **Control of typhus in Italy 1943-1944 by use of DDT.** *Am J Public Health* 1946, **36:**119-129.
357. Wiener N: *Cybernetics or control and communication in the animal and the machine*. Cambridge, MA: MIT Press; 1948.
358. Willer R, Willer D: **Exploring dynamic networks: hypotheses and conjectures.** *Soc Networks* 2000, **22:**251-272.
359. Winkler K-HA, Chalmers JW, Hodson SW, Woodward PR, Zabusky NJ: **A numerical laboratory.** *Physics Today* 1987**:**28-37.
360. Wolfram S: *A new kind of science*. Champaign, IL: Wolfram Media; 2002.
361. Woodruff RE, Guest C, Gamer M, al. E: **Predicting Ross River virus epidemics from regional weather data.** *Epidemiology* 2002, **13:**384-393.
362. Wooldridge M: **Intelligent agents.** In *Multiagent systems: a modern approach to distributed artificial intelligence*. Edited by Weiss G. Cambridge, MA: The MIT Press; 2000: 27-77
363. Woolhouse ME, Dye C, Etard JF, Smith T, Charlwood JD, Garnett GP, Hagan P, Hii JL, Ndhlovu PD, Quinnell RJ, Watts CH, Chandiwana SK, Anderson RM: **Heterogenities in the transmission of infectious agents: implications for the design of control programs.** *Proc Natl Acad Sci U S A* 1997, **94:**338-342.
364. World Health Organisation: **Disease Outbreak News** [http://www.who.int/csr/don/archive/disease/severe_acute_respiratory_syndrome/en/]
365. World Health Organisation (WHO): *Influenza research at the human and animal interface. WHO Working Group on influenza research, epidemic and pandemic alert and response WHO/CDS/EPR/GIP/2006.3*. Geneva, CH: WHO; 2006.
366. World Health Organisation (WHO): *The global burden of disease: 2004 update*. Geneva: WHO; 2008.
367. World Health Organisation (WHO): **Cumulative number of confirmed human cases of avian influenza A/(H5N1) reported to WHO.** [http://www.who.int/csr/disease/avian_influenza/country/cases_table_2009_08_11/en/index.html]
368. World Organisation of Animal Health (OIE): **Highly pathogenic avian influenza. Animal disease data** [http://www.oie.int/eng/maladies/fiches/a_A150.htm]
369. Wright PF, Ross KB, Thompson J, Karzon DT: **Influenza A infections in young children. Primary natural infection and protective efficiency of live-vaccine-induced or naturally acquired immunity.** *N Engl J Med* 1977, **296:**829-834.
370. Yan X, Zou Y: **Optimal and sub-optimal quarantine and isolation control in SARS epidemics.** *Math Comput Model* 2008, **47:**235-245.

371. Yang J, Yang F, Huang F, Wang J, Jin Q: **Subclinical infection with the novel influenza A (H1N1) virus.** *Clin Inf Dis* 2009, **49:**1622-1623.

372. Zambon M, Hays J, Webster A, Newman R, Keene O: **Diagnosis of influenza in the community. Relationship of clinical diagnosis to confirmed virological, serological or molecular detection of influenza.** *Arch Intern Med* 2001, **161:**2116-2122.

373. Zaric GS: **Random vs. nonrandom mixing in network epidemic models.** *Health Care Manag Sci* 2002, **5:**147-155.

374. Zhou SZ: **A seasonal influenza theory and mathematical model incorporating meteorological and socio-behavioral factors.** *J Trop Meteorol* 2009, **15:**1-12.

375. Zietz BP, Dunkelberg H: **The history of the plague and the research on the causative agent Yersinia pestis.** *Int J Hyg Envir Heal* 2004, **207:**165-178.

376. Zilversmit DB, Entenman C, Fishler MC: **On the calculation of turnover time and turnover rate from experiments involving the use of labeling agents.** *J Gen Physiol* 1943, **26:**325-331.

377. Zinsstag J, Roth F, Orkhon D, Chimed-Ochir G, Nansalmaa M, Kolar J, Vounatsou P: **A model of animal-human brucellosis transmission in Mongolia.** *Prev Vet Med* 2005, **69:**77-95.

Curriculum Vitae

Timo Smieszek

Date of birth	January 31st, 1980
Citizenship	German
E-Mail	timo.smieszek@env.ethz.ch

Education

2006 – present	PhD studies at ETH Zurich, Institute for Environmental Decisions Natural and Social Science Interface (Prof. R.W. Scholz).
July 2008	BiostatEpi: Summer School on Modern Methods in Biostatistics and Epidemiology. Cison di Valmarino, Italy.
June 2007	Summer school on complex systems research at the New England Complex Systems Institute. Cambridge, MA, USA.
Jan. 2006	MSc ETH, ETH Zurich. Thesis: *How people communicate about new technologies and ideas: Modeling decision coordination to simulate the diffusion of innovations*.
2000 – 2006	Studies of Environmental Sciences at ETH Zurich (focus on environmental physics and anthroposphere).
June 1999	Abitur (~high school diploma; focus on physics and history) at the Klettgau Gymnasium Tiengen. Waldshut-Tiengen, Germany.

Professional experience

2006 – present	Research assistant at ETH Zurich.
Oct. 2004 – Jan. 2005	Internship XL Insurance Environmental. London, United Kingdom.
July 2003 – Sept. 2003	Internship XL Insurance Global Risk. Winterthur, Switzerland.
Feb. 2003 – Mar. 2003	Internship Federal Environmental Agency. Berlin, Germany.

Awards

2006	ETH medal (award for outstanding master's thesis)
1999	Franz Schnabel medal (award for very good students in history)

List of Publications

Peer-reviewed papers in ISI journals

Smieszek T: **A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease spread.** *Theor Biol Med Model* 2009, 6:25.

Fiebig L, Smieszek T, Saurina J, Hattendorf J, Zinsstag J: **Contacts between poultry farms, their spatial dimension and their relevance for avian influenza preparedness.** *Geospat Health* 2009, 4: 79-95.

Smieszek T, Fiebig L, Scholz RW: **Models of epidemics: when contact repetition and clustering should be included.** *Theor Biol Med Model* 2009, 6:11.

Smieszek T: **Unsicherheit, Werthaltungen und Handlungsblockaden.** *GAIA* 2006, 15: 251-254.

Hansmann R, Bernasconi P, Smieszek T, Loukopoulos P, Scholz RW: **Justifications and self-organization as determinants of recycling behavior: The case of used batteries.** *Resour Conserv Recy* 2006, 47: 133-159.

Müller-Herold U, Smieszek T, Peter P, Scheringer M, Morosini M: **A simple measure for precautionary assessment of organic chemicals with respect to global cold condensation.** *Ecol Model* 2006, 194:266-273.

Non-reviewed publications

Smieszek T, Mieg HA: *Bewertung von Umweltgütern: Contingent valuation, conjoint analysis und andere Bewertungsmethoden im kritischen Vergleich*. Zurich: ETH Zurich, Professur für Mensch-Umwelt-Beziehungen; 2003. doi:10.3929/ethz-a-004652577.

Conference papers

Fiebig L, Smieszek T, Zinsstag J: **Social patterns and communicable diseases: What epidemiology (still) can learn from SNA.** *In Applications of Social Network Analysis (ASNA).* Zurich: University of Zurich; 2008.

Smieszek T, Peter P, Müller-Herold U: **A simple assessment model for precautionary pre-selection of environmental chemicals with respect to global cold condensation.** In *The Fourth European Conference on Ecological Modelling (ECEM 2004); Bled, Slovenia*: Edited by Dzeroski S, Zenko B, Debeljak M. Jozef Stefan Institute, Ljubljana; 2004: 55-56.