DEPARTMENT OF MEDICAL EPIDEMIOLOGY AND BIOSTATISTICS
Karolinska Institutet, Stockholm, Sweden

COMPUTER AIDED INFECTIOUS DISEASE EPIDEMIOLOGY – BRIDGING TO PUBLIC HEALTH

Martin Camitz

Karolinska Institutet

Stockholm 2010

Published by Karolinska Institutet. Printed by Universitetsservice AB, Solna

ISBN 978-91-7457-105-9

Till dig, mamma

ABSTRACT

This thesis explores the junctions of mathematical and computer modeling of infectious disease epidemics, the basis of such research and the communication of results. With increasing frequency we turn to computers and software for any type of research problem encountered. Computer modeling is a blessing with many hidden trapdoors. Skipping mathematical modeling, resorting to code immediately, is ill advised. Validity, uncertainty, bugs and old mathematical truths must all be taken under careful consideration. The same duality is present in the communication of the results from computer models to the public, to decision makers and to peers.

These topics are discussed in the context of four contributing papers.

The first paper describes a computer model of an infectious disease epidemic in Sweden. Using Swedish travel data we were able to demonstrate a way of successfully restricting travel to delay the spread of disease.

The second paper discusses a known fallacy common to many epidemic models, often overlooked when mathematical models are simulated on computers. It is demonstrated that it must be considered also with more complex models. The model in Paper I is used to exemplify the problem.

The third study takes the parsimonial considerations of the first two papers to another level, proposing static models for use in epidemic modeling. Understanding, an eluding - especially in computer models - but essential component in all models, is benefited.

The fourth study explores the epidemiology of sexual networks. Using survey datasets we show that with high probability, the sexually active population is largely connected, in a so called giant component, rendering the Swedish population an ideal isotope for sexually transmitted pathogens.

LIST OF PUBLICATIONS

I. Camitz, M., F. Liljeros
The effect of travel restrictions on the spread of a moderately contagious disease
BMC Med 2006,4:32.
[Reprint]

II. Camitz, M., A. Svensson
The effect of time distribution shape on a complex epidemic model
Bull Math Biol 2009,71(8):1902-13.
[Reprint]

III. Camitz, M.
StatFlu - a static modeling tool for pandemic influenza hospital load for decision makers
Euro Surveil. 2010;15(35):pii=19256
[Reprint]

IV. Liljeros, F., Giesecke, J., Camitz. M
How sexually active are the Swedes?
[Finished manuscript]

CONTENTS

LIST OF ABBREVIATIONS

STI	Sexually transmitted infection
IDE	Infectious disease epidemiology
R_0	Basic reproduction number
R_t	Replacement number

1 FOREWORD – CROSS DOMAIN

This thesis is about two interfaces, the overlapping regions between three domains. The first is between two types of modeling - mathematical and modeling. The second is between computer modeling and the target audience - the public, decision makers and other scientists, in the fields of medicine and/or more theoretically oriented engineers and mathematicians.
The following questions are contributed to:

- Are travel restrictions effective?
- What considerations should be taken when modeling epidemics, in general and specifically with computers, in terms of
    - simplicity and complexity?
    - understanding and validity?
    - data?
- Given the difficulty of validating an epidemic model, how can we communicate the results to
    - decision makers?
    - scientists?
    - the public?
- Why are sexually transmitted diseases endemic in Sweden and many other countries?

Some readers will welcome an opportunity to thoroughly scrutinize my work of the past years; most readers will be able to grasp the theory and results at least superficially; all readers can derive pleasure or benefit from reading about infectious disease epidemiology and computer modeling. The language and format throughout is intended to encourage all of you.
This foreword, briefly summarizing the papers, is followed by and introduction, as is customary. This is followed by a chapter on infectious disease epidemiology, outlining the key differences to the non-infectious sister field to which many of the readers may be more accustomed. Chapter 4 briefly summarizes the history of epidemic modeling and serves as an introduction to some important concepts in the field. Concepts of modeling in general and epidemic modeling in particular, follow in Chapter 5.
The remainder of the preface I have opted to organize paper-wise, rather than collate reoccurring sections within. It would perhaps be suitable to have a chapter for each of the questions asked in the list above, but I feel that the issues are best illuminated and exemplified when the discussion is intertwined with the description of each of the papers. Some points are lost without at least an idea of what the papers are about. The issues are therefore included within the chapters dedicated to the papers individually, sometimes as named sections and sometimes as a brief mention.

1.1 MODELING WITH COMPUTERS

Computer modeling is hardly an unconventional approach, in this day and age. To any given problem presented or question asked, more often than not, we turn for a solution to software engineers and their calculating machines. What we have come to find is that, much less than before, problems have to be adapted to suit the tool. Much more than before, the modeler will try to answer the question that was actually posed, not a simplified or idealized version thereof.

Computers were originally dreamt up by mathematicians to solve mathematical problems that were infeasible to solve by hand or impossible to solve analytically. Soon economists, engineers and sociologists caught on to solve their problems. Being accustomed to what can be accomplished, problems are now posed directly with the computer in mind, often skipping tedious mathematical modeling. But that is both a relief and a concern.
Speaking as en engineer, skipping the mathematics part, is not real engineering. Turning the problem over and over before you attempt to break it with the most readily available means has obvious benefits. Inconsistencies can be found, expected outcome can be assessed and used to *validate* the model and *verify* the output. If too much laboring is lost when we enthusiastically feed the problem into our machines, then we run the same risk as before the teraflops era, of finding that what comes out is a solution to another problem than the one originally stated, or, not a problem at all. Computers are blind, so is the software and neither can replace solid mathematical modeling entirely.
The secondary aspect from which to view this collection of papers is that of communication. This concerns message, goal, target audience and not the least, the journals in which the papers are published. …to be considered when reading the thesis and papers.
On a scale from mathematics to software engineering, from simplistic to realistic, from broad to narrow target audience, the models and papers described in this thesis cover a reasonably wide range over the middle of the spectrum. They will serve well to illuminate the issues.

1.2 THE PAPERS

Before we start with the background knowledge required to appreciate the papers, let me first briefly introduce them.
The first paper describes a simulation of an infectious disease over the country of Sweden. It is an example of a simple mathematical model describing the dynamics of disease spread and using a computer program to simulate it. The core of the model is the way people travel and move over the country.
We run the many thousands of simulations, each starting with a single infected person in Stockholm. The infection spreads from person to person within the municipality and, by way of the traveling model we set up, to other municipalities around the country. By carrying out many simulations we are sure to capture sporadic effects that might prove significant.
We show that by limiting the travel distance we can achieve a significant delay in the spread of the disease. However, as others have consistently shown, travel restrictions alone are not able to stop an outbreak in its tracks.[1, 2]
On a scale from simplistic versus complex, as a *multi-scaled* model, this model places it somewhere in the middle. We will discuss how much information was included in the model and whether or not we were overly scrutinous, particularly concerning the travel data.
The second paper can be seen as an amendment to the first in that it describes an important trait, in our model and others, that is oftentimes overlooked - not without consequences for the result. Epidemics may be delayed or accelerated and other dependent variables will be affected. Depending on the problem at hand, the questions

to be answered, this may be perfectly in order, since the mean time will still be respected. In the model in Paper I, we were mostly interested in mean results. For calculating the total number of infected people at the end of an outbreak, for instance, the simplification works.

Are computers to blame? Partly, yes. The computer program, the code, conceals certain facts that would be easily spotted by a capable mathematician on reviewing the underlying equations. The issue should be understood and considered for any result more intricately involving the timing of the epidemic and related processes. If not before, this is achieved by trying to model with mathematics before code.

The problem addressed in Paper II is just one of the many traps set by seeing computers as an easy way out. I am not saying that this particular problem is exclusive to computer models, only that a new breed of scientists means that old lessons have to be relearned. Computers allow us to solve a new, bigger class of problems and it is very tempting to try an answer questions for which the underlying mathematical model is not suited. Computers do not substitute traditional mathematical ground work. Another question is of what use the rather complex result of a computer model is. A solution to a simple mathematical model can hold an equally simple truth, one that is both beautiful and powerful. Can we understand the output without really perceiving the intricacies of a computer simulation? If so, when applied to the real world, what insight can be lent?

The point of the argument I cannot bring further than with an idea such as *static modeling*. Static modeling wholeheartedly embraces the simplicicist stance, the keywords being simplicity, transparency and interpretability. The model in Paper III, predicting hospital load during an influenza pandemic, is simple. The model proposed takes historical data on past flu pandemics and epidemics and transfers them to settings of today. Computers play a vital part in displaying the output, accounting for uncertainty, quickly visualizing the effect of changing certain parameters as well as solving some of the equations numerically. This is not to conceal the fact that the core mathematical model is so simple that it can be analyzed by anyone with pen and paper using simple algebra.

Finally, Paper IV concerns sexually transmitted disease and asks the question just how closely knitted the sexually active population is. We do so by modeling the population and social contacts between people as a network. We set out to find the *giant sexual component*.

Physicists have contributed to infectious disease epidemiology for quite some time mainly through their interest in network theory[3-5] and statistical mechanics[6, 7], the two fields actually having a few notable points of contact. Some very fundamental questions can be answered by very simple means, just by asking who can be reached by whom and how fast.

If it is the case a giant component exists in the Swedish sexual network, then that would mean that most sexually active Swedes would be connected to each other via their partners, their partners' partners and so on - an ideal habitat for certain sexually transmitted infections, in other words.

The existence of a giant component can be demonstrated using some superficial knowledge about sexual contacts and a few cleverly applied theorems. The model is simple and idealized in order to be tractable. To support our claim, we wanted to simulate a few networks to see what they looked like, thus computers come to our aid

also in Paper IV. We add a tad more complexity to strengthen a particular assumption in the analytical part. The two models work hand in hand very nicely.

1.2.1 Target

Although the first two papers concern almost the same model, they are quite different in both style and intended audience. While the first is a policy paper published in an open access journal and has decision makers in its list of targeted audience, the other is more theoretical, intended only for the enjoyment of other modelers and those interested in the field.

Paper III makes its appearance in a medical journal. It describes software, the proposed usage of which is amongst decision makers and epidemiologists primarily, but can also be understood by the general public. In spite of this, it is the one paper of the four that is most firmly grounded in the engineering field. The programming was preceded by laborious mathematical modeling, numerical modeling as well as a literature review of the field and other studies.

The message of paper IV is clear enough. As for the recipient, this paper targets public health informers, but only as the middle man. The insufficiently enlightened general public should listen attentively to this argument for a more responsible sex life.

2 INTRODUCING MODELERS

The capture of the mechanism of the universe in the language of mathematics is a true art. Mathematical modeling and it is essential in understanding many things around us. The hydrogen bond between water molecules, for example, is what gives rise to the formation of the beautiful hexagonal patterns we see emerge as water freezes to crystals and is essential for the existence of life. It has captured the imagination of scientists through centuries and withstood most attempts of modeling. As enthralling as it is to look upon a snowflake, for a certain breed of people, the equations describing their growth are even more beautiful than the real thing. In any case, one must agree that enclosing the richness of nature in the rigorous structure of mathematics, enabling us to explain and predict, is one of the pinnacles of human achievement.

Chemical bonds are one thing. To model epidemics is to model humans and their erratic behavior. The obstacles imposed by the complexities of a conscious mind that itself is able to do mathematics, is what makes sociology seem inapproachable by mathematics. "Physics would be a lot harder if atoms could think." as Nobel laureate Murray Gell-Mann said.[8] Yet, in sociology, it is not the brain that is to be modeled, but the statistical properties of humans as a group. Richard Swartz[9]:

> *...it is easy to forget that our world becomes unmanageable without generalizations, atomized [as it is] in an infinite number of special cases that rarely, if ever, allow themselves to be summed to a sensible whole."*

Although he was talking about bigotry, I think this rather sums up why mathematical modeling always has part to play in sociology. If not, Auguste Comte, one of the founders of sociology, spoke about "social physics" before the precedent term was coined[10] and, underlining the foresight in this coinage, Anatol Rapoport developed *social network* models in the 1940s. For further reading, consult Coleman[11], Karlsson[12], Rashevsky[13] or Simon[14].

Often, indeed, people *are* modeled as were they particles. Helbing[15] found that an obstacle, like a column, suitably placed in front of an emergency exit will facilitate the evacuation of a crowd in case of an emergency. In fact, the simplest models demonstrating epidemic spread though a human population, are simple in comparison to Helbing's.

See if you can spot this one[16]:

> *There's definitely, definitely, definitely no logic*
> *To human behaviour*
> *But yet so, yet so irresistible*

2.1 ENTER COMPUTERS

Most definitely: exactly at the point where computers enter, mathematics looses some, if not most, of its elegance. The exquisite texture of symbols and elegance of compact formalism is replaced by variables spelled out in plain English in page after page of code glaring brightly from a screen. As if this were not enough, variables are assigned values, the sort of thing mathematicians go to great lengths to avoid. Mathematics has after all been developed over centuries to formalize human thought. Programming

languages have been developed in a very short span of time to formalize the comparatively rather gnomish instruction set of a processor.
Then again many mathematicians would rather not have their equations have anything to do with reality either and computers may be seen as a step towards this. The process is too useful to be hard-neckedly ignored. The remarkable development in the field of computer hardware and programming tools have, in addition to offering new insights in phenomenon replicable in the lab, enabled scientists to put mathematical models to the test by computer simulation, in an environment that, in other cases, is as close to reality as we can hope to achieve. Inescapably, ethics often prevent us from performing controlled experiments on human subjects, in a lab or otherwise. This is the case for sociology most of the time and certainly for infectious disease epidemiology all of the time. Real events of course allow us to gather data but it is impossible to control for variables in a way that would completely satisfy empirically.
It is the physicists that have promoted research into computer modeling. They have recently found that there is little work for them to do in theoretical physics. Advancement in that area can be entrusted to handful of theoretical geniuses. Instead physicists have had to expand in to other fields, applying statistical mechanics, wherever they are welcome and most other places as well. No area of research has been considered out of reach for applied physics. Some went into sociology, spawning the modern research into network theory.
The heart of the issue is that the equations of the mathematical model require solving. Some are hard to solve analytically, some are impossible. The benefit of a computer is that it enables the numeric solution of hard or impossible equations. That's what attracts mathematicians. First and foremost, however, it enables the modeler to propose models that he or she knows to be intractable to start with. That's what attracts the physicist. The explosion of new models proposed in all fields of science can surely not be a bad thing, even though the average quality may have declined some what in later years.
Models may be proposed for the exclusive processing of computers. When it comes to sociological models, from which epidemic models inherit, computer models have an edge. They can represent individuals explicitly and assign each a different behavioral model.[17] Mathematical models cannot.[18] Micro-simulation models or agent based models are motivated thus. Although I have put substantial effort into micromodels,[19] I will touch upon it only briefly in this thesis.

2.2 ENTER DATA

In epidemic modeling, physicists and engineers are now doing what mathematicians have been doing the past century. Neither are necessarily epidemiologists. In order to be an epidemiologist and not just a modeler, you have to get your hands dirty with the data. Until then, everything you do is just theory.
Epidemiologists are expert statisticians and data analysts. That means searching, finding, analyzing, cleaning and aggregating data. Data never fits the model, the model always has to be adapted to the data, theory and method developed, margin of error estimated – never the other way around. It's raw, it's erroneous, it's biased and people always underestimate the work going into en epidemiology paper with the words "available data".

All the results in this thesis have been preceded by months of working with data. Paper I and II build on enquiry data from SIKA[20] regarding travel of all kinds by all kinds of people to and thro in Sweden. In the process I amused myself with creating a map of Sweden as it would look if the cities were moved in order to minimize fuel consumption.[21] Paper III uses a multitude of data sources, mostly published aggregated data, but also population data[22] and data from the Hospital Discharge Register[23]. Paper IV uses the data from a Swedish sex enquiry[24].
This thesis will also discuss considerations of processing and applying data.
Chapter 4 will continue the introduction to epidemic modeling. An introductory chapter on *infectious disease epidemiology* (IDE) cannot be postponed, however.

3 INFECTIOUS DISEASE EPIDEMIOLOGY

Diseases have throughout most of the history of human kind been believed to be a punishment for their sins from deities.[25] This has contributed to the species-preservational practice of shunning and stigmatizing patients, before certain diseases were recognized to be contagious. The Christian church considered leprosy patients as unclean.[26] There was awareness early on in the western world, China and Africa, that avoiding smallpox patients was a sensible thing to do.[27] 16th century French knew to avoid suspected bearers of the Italian disease and vice versa, the disease in question known since 1543 as Syphilis, the name of the shepherd who offended Apollo in a poem by Girolamo Fracastoro.[28]

In others cases, there is no ground for any kind of social intervention, if the disease in question is not contagious, substance abuse for instance. Infectious and non-infectious diseases have thus been separated in history since eons. Typically in the former branch one finds those diseases that are the most terrible, most feared and that time and time again have altered the course of history - the bubonic plague, smallpox, syphilis and pandemic influenza etc. Hippocrates used the terms *epidemic* and *endemic* [*] to differentiate those diseases that occurred from time to time in large numbers, to those that were always in the population.

It is worthwhile to consider the origin of infectious diseases. Bacteria and viruses are living organisms, obeying the same rules of selection as the rest of us. With that in mind it is entirely possible to think of a time without infectious diseases. Evolution would preserve those species that specialized to a large, moving population which in the early days of humanity did not include us. Denis Mollison writes in the introduction to *Epidemic models: their structure and relation to data*[29]:

> *The spectacular success of humans in dominating the world's ecology has meant that they – and their domestic animals and their crops – provide an unprecedented rich resource for parasites. Not surprisingly parasites have evolved, and continue to evolve to exploit this resource.*

However, the distinction between infectious and non-infectious disease *epidemiology* should preferably not be drawn particular diseases as such, but rather methodology of understanding they're causes. The *clinical* study of either type of disease, is after all, rather similar. The epidemiological fields are markedly dissimilar.

Epidemiology in general is the study of that which causes illness in a population. A typical definition is[30]:

> *The study of the relationships of various factors determining the frequency and distribution of diseases in the human population.*

We read the words *frequency, distribution* and *population*. Population means large numbers and large numbers means statistics, connotations to frequency and distribution being almost excessive. The definition implies that epidemiology is not so much a clinical field as a subfield of statistics and/or public health. In fact the clinical element which does exist is too often ignored.[31] Epidemiologists, the non-infectious disease

[*] Today the usage of the term endemic is not exclusive to non-infectious diseases. Malaria, for instance, is an infectious disease that is always present in many parts of the world.

kind, at any rate, try to group the population according to characteristics in order to find differences between those who are cases and those who are not. *Causes* and *effects* are discerned and *exposures* measured put in relation to *cases.* Usually the method is different forms of *statistical regression*: *linear, logistic, cox* etc, making epidemiology akin to a science of the likes of sociology, to name just one. The power of these methods applied to the data that have be collected in cohorts, some for several decades, have been enormously successful in advancing medicine and public health, providing firm scientific ground for what we should and should not do to avoid cancer for instance. In the press, it seems, reports about an X times increased risk of B for people being C, eating D and/or engaging in E are never more than a few days apart.
The US Surgeon General William Stewart is often accused of momentously announcing that it was time to close the book on infectious disease epidemiology[*]. Unfair attribution or not, this quote resonates of the general attitude at the time. Improved sanitation, vaccines and antibiotics had all but extinguished threats like smallpox and polio, with malaria next in line.[35] It seemed that what little we had yet to learn in IDE was of little significance and that resources could be put to better use in for instance non-IDE. Smoking had recently been established as the primary cause of lung cancer and the promise of similar unveilings from population and patient data banks was irresistible for scientists.
Of course, no small amount of western chauvinism is detected in this way of thinking. In the undeveloped world, many plagues that the west had rid itself of, continued to persist and cripple nations.[36]
40 years and some 230 newly emerged infectious diseases later[37]: most antibiotics rendered virtually useless[38] and we now have the likes of Legionnaire's disease and AIDS with no vaccine in sight. New hemorrhagic fevers and respiratory infections emerge every now and then as well as new infectious agents like prions to complement the comparatively well studied bacteria and viruses. Mostly the new infections are of zoonotic origin emerge in the parts of the world where monitoring is the weakest the reprisal of our western chauvinism.[37]
Nevertheless, infectious disease epidemiology remains a sub branch of epidemiology along with environmental epidemiology, genetic epidemiology etc. Non-IDE remains the flagship, still receiving most of the funding.

3.1 DEFINITION REVISITED

As we have seen, history itself warrants a division of the disciplines of epidemiology. The way epidemiologists study their particular diseases is in many ways similar, but in more ways fundamentally different. The key difference is illustratively captured in the typical definition of epidemiology above.
The keyword is *factor,* which when concerning the study of infectious diseases is commonly also a *case*. The incidence of cancer patients does not decline if the effectiveness of patient treatment is improved upon. Treating tuberculosis patients, on the other hand, will affect the risk of future patients falling ill. A cancer patient is not a

[*] The quote has all the trademark of a popularized misquote. There are only secondary sources and Stewart himself has no recollection. Supposedly the quote is from a conference in 1967,[32] but the published transcript gives no confirmation. The Public Health Service has looked into this and can neither confirm nor deny the quote. [33] In any case, it is so widely spread and marks the era very suitably, rather like Marie Antionette and post-construction about cupcakes.[34]

factor. A tuberculosis factor is. Non-infectious disease epidemiologists in bewilderment find that it is not possible to separate *case* and *exposure*, rendering useless all regression models - all that I know of, at any rate.
Like the mother roe deer banishes her kids of last year, the very definition of epidemiology painfully expresses the need for a division of fields. In studying IDE, we need models of another kind. For example, in order to glance at the infinite range of possible random events that play such a great part in determining the outcome of an outbreak, stochastic models are commonplace in IDE in a way that is not immediately required in non-IDE.
Many joint ventures still remain, of course. Firstly, all the methods of non-IDE are still available for IDE in dealing with *non-communicable* diseases, significantly food borne diseases, where the disease in question is not primarily transmitted between humans. The distinction in this case between exposure and case is rather more clear. Genetic and other factors of our upbringing and environment are still important for the susceptibility of infectious disease[*]. This is plainly seen in the west when comparing to other parts of our world. Poverty is a major determinant for infectious diseases.[36, 37, 39]
Although never a direct cause in IDE, environment and other facilitating factors are always there,[40] so it cannot be seen unfit to do some reverse thinking. For some time evidence has been amounting that many of the diseases we see as non-infectious, really are caused be viruses or bacteria. Cancer immediately springs to mind. 18 % of all cancer cases are attributable to infections, 8 % in the developed world, 26 % in the developing world.[41] These figures are likely underestimated.[42] All the facts are not in. Stomach ulcer is another such disease, a discovery earning the Nobel Prize in 2007, but the list may also include diabetes and obesity.[43-46] This should be incitement enough for minds from both fields to get together and do some interdisciplinary thinking and perhaps even shed some more light on peculiarities like *pie chart models* and *biologic interaction.*[47, 48]

3.2 THE MAIN DIFFERENCES

Giesecke[31] suggested five traits that primarily separate the discipline of infectious disease epidemiology from its non-infectious sister. The first has already been discussed. The second has been touched upon.

- A case may be an exposure.
- Contact patterns play a major role.
- Sub-clinical infections influence epidemiology.
- Immunity
- There is sometimes a need for urgency.

In non-infectious disease epidemiology one seeks the causes of a disease in what we eat, how we live or in the genome.[48] Sometimes these factors are exclusive, sometimes they combine and interact. The role of the epidemiologist is to determine those factors, explain the mechanisms and finally quantify them and any interactions. In infectious disease epidemiology the analogous process is called *contact tracing*. While the above

[*] The reciprocal fact that genes could be seen as a form of disease agent from parent to child should probably be ignored in this discussion.

factors may be - and some clearly are - factors also for infections, the proximal cause of all cases is always another case. The question of causality, discussions about which always venture into the philosophical, is thereby somewhat simplified.
In practice the difference is far from marginal. Doubling the exposure is expected to double the number of cases, provided there is no interaction. All other things equal, in a country with twice as many cars as another, we would expect twice as many car accidents. To first approximation at any rate, this is linear growth. Granted, epidemiology would not be a science if reality were this easy. There may be limiting as well as accelerating associations. Complications abound.
However, growth patterns in non-IDE have no correspondence to, for example, a measles epidemic. A single infected index case can spawn an exponential growth in incidence. Factors, for instance global immunity, factors which at first glance appear random, may pivot the outbreak, either quenching it or accelerating it many-fold.
For certain diseases, when all factors have been ruled out, there are still individuals who will not become cases no matter what the exposure, suggesting immunity.
Immunity is at the heart of infectious disease dynamics. The equations (4-1) on page 14 demonstrate this. Ultimately, immunity - or rather, the depletion of susceptibles - is what kills the epidemic. In the case of measles and many other diseases, the interplay of immunity and prevalence is thought to be a major mechanism behind keeping it *endemic* and *periodic*.
In non-IDE, on the other hand, immunity is but a complicating statistic.

3.2.1 Need for speed pertaining to policy

The last of the points above may now have become apparent. The recent outbreak of the novel flu should convince anybody of the difference a speedy response will make in a way that is not an issue in non-IDE. That is why developments in influenza vaccine production, shortening by a few weeks the time from the discovery of a new strain until the vaccine is in production, is a very worthwhile endeavour.[49]
The implications stretch into policy. Many countries, including Sweden, have vast systems for collecting data and are very good at it. Some would have infectious disease control snugly fitted into this system of keeping logs and registering. Admittedly, this is an integral part of infectious disease control, but keeping tabs, for the most part, only establishes the chain of events after they have taken place. This is the way we view statistics in non-IDE: in retrospect. The need-for-speed principal emphasizes how much more there is to IDE and infectious disease control. An agency is needed to act, when it happens and preferably before it happens. Data is needed in real time.[50]

3.2.2 More differences

A pivotal circumstance is social contact patterns and *contact rate*. It is of course intuitively reasonable that the number of people we meet everyday, on average, is very influential for disease spread. Modelers easily show that not only the average, but also the distribution, is important, i.e. some people have many contacts, others have few. So called *social hubs*, *core groups* and *super-spreaders* are studied for natural reasons as they can explain outbreaks in cases where other models are too simple.[3, 51-53] Although it is an enjoyable field of study, it is not only for recreational purposes that epidemic modelers have become so interested in social networks, that is, the population as structured in *nodes* and *links*. Studying the properties of *complex networks* has proved invaluable. These are some of the topics for later chapters.

Another factor the non-infectious disease epidemiologists need not concern themselves with is *subclinical* (*asymptomatic*) cases. A case that shows no symptoms is not a case, but the fact that she may be an exposure i.e. she may be infectious regardless of symptoms, means that it is of the utmost relevance for an infectious disease epidemiologist to study the prevalence and infectiousness of sub-clinical cases. This gives epidemiology a sinister element stemming from ignorance, rather like black matter in astrophysics. Often vaccination leads to an increase in the proportion of sub-clinical cases. The plot thickens.

3.3 A FINAL THOUGHT

I'd like to end this chapter with some free theorizing about what an infectious disease really is, completely speculative on my part. Apologies for my exuberance, but this is where I reward myself for writing this thesis.

Given that infectious diseases evolve and adapt, same as any species subject to Darwin's laws of selection: One strategy to ensure the continued existence of the species is to avoid detection by predators. Non-symptomatic infection is to infectious diseases what camouflage is to stick insects. Furthermore, it is not immediately beneficial for a population to destroy its own habitat; at least not before it has had the chance to spread into new ones. The ultimate infectious disease is not a disease at all, having reached the point where it is not considered parasitical in any host, a *commensal* organism, one that neither helps nor harms us.

How many diseases that are not diseases thrive in our population? 90 % of the cells in the adult human are microbes, after all.[54] Those that help us could very well have been diseases at one point in our history. 8 % of our DNA is estimated to originate from ancient retroviruses[55] just as the mitochondria, vital in aerobic metabolism in all eukaryotes, is though of as having bacterial origin.[56]

We do not know the infectivity of these silent invaders since we tend to focus our energy on that which presents a threat.

With this in mind, fighting a particular disease with treatment and isolation is a doubly beneficial activity, not only for the direct effects of decimating the population of infectious agents, but also by constraining the evolutionary paths into more benign organisms, ultimately into non-somaticity. Chlamydia might be headed this way.[57]

At any rate, thinking along these lines presents another argument in support for intensifying our efforts against emerging diseases. The threat against humanity from infectious diseases lies not in the ones we know about, not even in the ones we don't know about but from the one's that are about to emerge.

4 EPIDEMIC MODELING AND ITS HISTORY

Daniel Bernoulli, the Swiss mathematician, was the first to calculate the immunity level of smallpox in the population and hence the vaccination level required to deter small pox epidemics.[58] He formulated and solved his mathematical model in 1760. Real progress in the field, however, would have to wait until much later when it was realized how much common ground epidemiology and demographics shared.[59] Many concepts from demographics were introduced in the early 20th century. Surprisingly, many concepts would still be developed independently within the two fields.[59]

Ross, Hamer and others[60, 61] were interested in malaria, Italian for *bad air*, which was a major nuisance to infra structure projects in the colonized world, for example during the building of the Panama Canal.[62] Since it is tick borne it was conceptually harder to model than for example measles. Still it was Ross who would first discover the critical elements that signify infectious disease modeling, the rate of contact between infectious and susceptible.

Typical of infectious diseases is that they can lie dormant in the population or other reservoirs, held back by the acquired immunity of the persons they would otherwise infect. Occasionally there may be local outbreaks that die out quickly. Suddenly, when the circumstances are right, in particular, when the general susceptibility has fallen under a certain level, the disease erupts and incidence levels climb, not gradually, but with near exponential growth. Exponential growth, wherever it is observed in nature, from rodent spawning to nuclear fission, indicates that the growth is not accelerated by external factors but rather fuelled by itself, as is the case for an epidemic.

More analogies transfer directly from other exponential growth processes. Corresponding to uranium and carrots in the above examples, epidemics are nourished by a plentiful supply of susceptible hosts. The element of criticality, that is, the sudden development from dormant to eruptive, as a consequence of a very small change in the environment, is called the *critical mass* in nuclear science and the *tipping point* when it comes to marketing shoes.[63] In infectious disease epidemiology criticality is captured in *epidemic threshold* theory, first developed by Kermack and McKendrick.[64, 65] With various mathematical models we can determine the level of susceptibility required for an outbreak, be it 5 % or 20 %.

The epidemic threshold is central to the upcoming chapters and indeed the entire thesis. Recommended reading for those interested in the history of epidemic modeling is the Brief History of R_0 -part of Heesterbeek's paper *A Brief History of* R_0 *and a Recipe for its Calculation*.[59]

4.1 THE CLASSIC SIR MODEL

Mathematicians made their first entrance in infectious disease epidemiology the 1950s, a sudden interest being raised after Bailey published a classic book on mathematic modeling of diseases[66]. Before that Kermack and McKenrick, demographers, introduced what is known as the classic SIR-model.[65] All models today, in one way or another, take as their point of origin Kermack and McKendrick's model. Though it is

very simple it encapsulates some very important features common to many infectious diseases, including a simple definition of R_0*.
The SIR-model introduces the three stages *Susceptible*, *Infectious* and *Removed*. The last of these is understood to denote immune, recovered, removed, resistant or deceased, whichever your fancy, in either case, neither infectious nor susceptible. Individuals in the model may pass through these stages in order, usually starting off as susceptible. Once infected by others, they move on to the infectious stage.
The *SIR* in *SIR-model,* of course, abbreviate these three stages, but different interpretations of the composing letters are abundant as well as pure misconceptions e.g. "systemic inflammatory response".[67, 68]
A version of the classic SIR model of disease spread can be written thus[69]:

$$\begin{aligned} \frac{dS}{dt} &= -\beta I \frac{S}{N} & S(0) &= S_0 \geq 0, \\ \frac{dI}{dt} &= \beta I \frac{S}{N} - \gamma I & I(0) &= I_0 \geq 0, \\ \frac{dR}{dt} &= \gamma I & R(0) &= R_0 \geq 0. \end{aligned} \tag{4-1}$$

The *contact rate*, which we only just mentioned in the previous chapter, appears here as β. People move from the susceptible to the infectious stage at a rate proportional to the average number of contacts of a susceptible individual with infectious people, $\beta I/N$ sometimes known as *standard incidence*. N is the total population. Originally the classic SIR-model was stated using the well known *mass action principal* with the infection coefficient, βI , but this is in my view inadequate for reasons nicely explained in Hethcote[69]. The coefficient to the next transition between I and R, describes the resting time within the infectious stage, $1/\gamma$ being the average infectious period.
Perhaps the most urgent extension to this model is to introduce the latency stage, commonly and equally inexplicably, known as the *Exposed* stage and assigned the letter *E*. Hence we have the *SEIR-model* - in Paper I call it the *SLIR-model, L* for *Latent*. In some models it is irrelevant to model an R-stage, people stay infectious longer than the time scale of interest: *SI*. In others, people return to a susceptible state after a period of temporary immunity: *SIS*. In an *MSEIR-model,*[69] mothers may pass on the infection to their newborns. Different disease representations occur in great numbers in literature.

4.2 PARTICLES IN A GAS

Going back to the original SIR model, the assumption is made that people in the population make contact with one another in random fashion, equally likely to meet anyone else at any time. This is called *homogeneous mixing.* It can be compared to a gas where particles move about and collide with each other completely randomly. Historically, the SIR-model and other compartmental models are linked to chemical kinetics. In fact the term *mass action* is borrowed from chemistry and defines what I mean by particles in a gas: colliding and reacting to form new substances. The mass action-principle dictates the speed of the reaction and what fractions of the constituting

* This mysterious symbol, mentioned twice already, will soon be explained, but much later than in most textbooks.

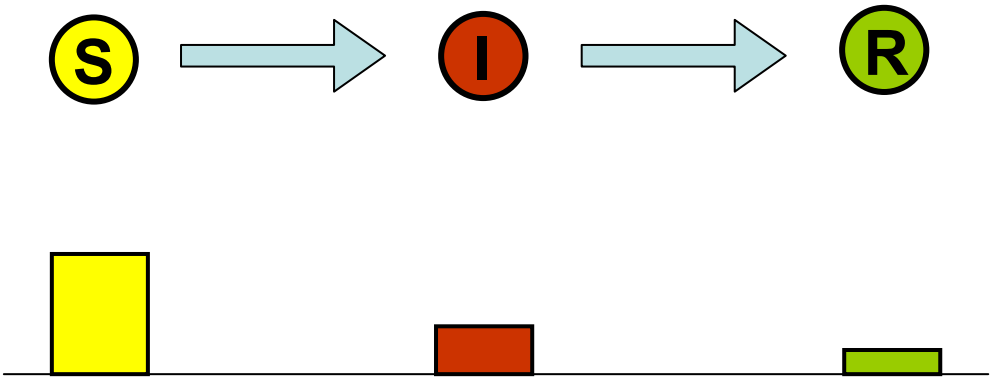

Figure 4-1 SIR-model schematics
The sketch shows the three states Susceptible, Infected and Recovered. Arrows indicate the possible state transitions. Beneath each state is a bar indicating the number of people in each state. The sketch represents a snapshot of an epidemic that is well into its lifetime but possibly even past its peak. There are more people in the infected state than in the recovered state. As long as there are infected left susceptible will continue transition to the infected state. At some point the number of susceptible will have diminished to a point where the rate of infection is lower than the rate of recovery. At this point the peak is reached and the red bar will start to shrink. The added height of all three bars must always correspond to the total population.

materials are used up once equilibrium is reached - these variables and others being independent of the system size. The analogy to epidemics is not perfect but you can see how McKendrick was thinking, he himself coming from a chemical background.[59]
Homogeneous mixing is a simple approximation. Due to the random and transient nature of each assumed person-to-person contact, the approximation is completely unrealistic in almost every situation.[70] Nevertheless, it is surprisingly powerful. What is more, the wealth of tools from physics and chemistry provided for analyzing what is after all a very well studied phenomenon, makes the method as a first approach very attractive.
Measles is one of the easiest diseases to model due to, to name but a few reasons, the fact that it does not require an intermediate host vector for transmission; a high probability of clinical recognition; a large number of recorded cases; and most important of all, a high rate of infection. From the perspective of the pathogen, the measles virus, the individuals comprising its biotope resembles very much a gas. From the infected it will jump to other family members with almost the same ease as the friend, the co-worker, the teacher or the cashier at the mall.
Standard incidence is one development to the classic SIR-model. Some circumstances may require more structured contacts models to work well. A development to the basic assumptions is to partition the population in different ways, by age, social boundaries or geographical boundaries. In Paper I, we used the homogeneous mixing within the borders of the Swedish municipalities, a so called *meta-population* model.

4.3 THE THRESHOLD CONCEPT

The equations governing the passage of individuals through the three stages can be solved quite readily, yielding a simple formula for the *epidemic threshold* in terms of density of susceptibles. We will take a look at this formula to summarize some important concepts in epidemic models.

$$\frac{S_{threshold}}{N} = 1 - \frac{1}{\alpha\beta}. \quad (4\text{-}2)$$

We'll save β for the next chapter; but α is the average infectious period $1/\gamma$ above, the time that an infected person, on average, transmits the disease to others. In other words, for the solutions to Equations (4-1) where the number of susceptibles is larger or equal to $S_{threshold}$, the time dependent variable *I* will initially grow. The longer the infectious period α, the larger the numerator and the smaller the right term of the equation. In other words, the degree of susceptibles, as a percentage of the total population, required for an outbreak, decreases. This makes sense since the more time you have to infect someone else, the more people you are likely to infect and the easier it is for an epidemic to initiate and sustain itself.

As the epidemic progresses, the number or susceptibles available diminishes, the infectious have fewer and fewer to infect and the epidemic comes in demise. The mechanism is called *herd immunity*.

Smallpox has been estimated to have a threshold susceptibility of 70-80 %.[36] Mankind was able to eradicate smallpox by bringing down the level of susceptibility using vaccination. Polio (82-87 %) may similarly be in reach of eradication whereas measles at 90-95 % is likely to be impossible.

One of the implications of modeling with a constant decay rate, γ, is that infectious times will be exponentially distributed meaning that while some people are infectious for a very long time, most will be so for unrealistically short times. The concern is the topic for Paper II. Depending on the analysis, a more realistic (fixed or non-skewed) infectious time is required.[71]

4.4 CONTACT RATE

If α is the infectious period, what then is β? If we were talking about a real particle system, we would be looking for something that encapsulates the willingness of the particles to bump and collide, determined by the thermodynamic parameters like temperature and pressure. We should probably abstain from taking the analogy as far as to say that the increased *contact rate* in the tropics is directly related to temperature, but it is true that the contact rate varies between regions and countries. β is thus the contact rate, the average number of *sufficient* contacts made per time unit. Not only does β vary geographically, it also differs from disease to disease, implicit in the interpretation of *sufficient* - sexual intercourse for some diseases, contaminating a door knob for others.

The SIR-model and the homogeneous mixing assumption capture contacts in the simplest possible way as it dictates that everybody will have equal opportunity to bump into everybody else, hence the analogy with the random movement in a gas. For some people this may seem deeply unsatisfying. Especially if "some people" are equipped with modern computers, they may be inspired to develop their own models with particular emphasis on more realistic contact setups.

4.5 FINALLY R_0

Another way of viewing the epidemic threshold, this time from the perspective of an aspiring infective agent, ready to invade a completely susceptible population, is R_0.

Also known as the *basic reproduction rate*. Solving the SIR model in (1), one actually finds that $R_0 = \alpha\beta$. In words it is defined as:

> *The average number of secondary infections caused by a typical infectious individual in an otherwise susceptible population.*

This is close enough to Anderson and May's definition taken to be the gold standard, where the word *typical* is implicit.[36] When R_0 is above 1 an outbreak is *possible.* In other words the circumstances are right that if introduced, the infective agent *may* spread in the population. If R_0 is below 1, an outbreak *cannot* occur. There may possibly be a local outbreak, but in the population for which R_0 is defined, the circumstances - the contact rate, infectivity, what have you - are not sufficient.
In my experience confusion often arises with R_0 here in the void between model and reality, but to be on the safe side it is best to say that when $R_0 > 1$, and outbreak *can* occur and not the assertive, *will* occur. As always, the difference depends on the model.
R_0 is a quantity no paper in epidemic modeling feels complete without at least some mention of. In my humble opinion, the roll of R_0 is an unfortunate consequence of this field having been dominated by mathematicians for so long. Hence, out of the four papers in this thesis, three ignore it. R_0 is an enlightening concept fetched from demography and is easy to grasp for mathematicians, but leads to many misconceptions with both modelers and laymen.
Often it is just a case of missing out parts of the definition. For instance, R_0 is defined for a completely susceptible population so you cannot use vaccination to lower R_0. The mistake is made in the abstract of this heavily cited piece by Gray *et al*.[72] One often sees in literature, like Everitt's handbook for clinicians and medical students,[73] for example, that when R_0=1 the disease is in an endemic state meaning that it neither dies out nor fully becomes an outbreak. While this is theoretically true, it is nevertheless a purely hypothetical case, a mathematical curiosity that has no correspondence to anything familiar in the real world, and certainly not what we know as an endemic. In simple models where endemic dynamics are incorporated, the infectious and susceptible proportions rise and fall periodically over time, in counter phase. Please review Hethcote[69] for an example. In this case the time dependant version of R_0, often denoted R_t or just R, oscillates around 1 in tune with the periodicity of the endemic. R_0 is as always constant, and what's more, if we are to have an outbreak at all, larger than 1.
Just a note: R_t: it is often known as the *replacement number*. Sometimes you will see *reproduction number* (note, no "basic") but I think this is too confusing.
Other misinterpretations often arise from the words typical and average. R_0 is an average value, defined in its context – population and/or geographic location – and there are always local variations which is often not included in models. The typical individual is sublimely different from the average individual, the latter having approximately one breast and an infinite menstrual period. The typical individual may exist but "it" can only be uniquely selected if everybody is indistinguishable in the relevant properties. Again chemistry sets the agenda.
Calculating R_0 is another matter still, as it is invariably model specific.[59, 71, 74, 75] In mathematical models, specifying R_0 with a closed formula is important. Sometimes but not always, this entails going to the limit of an infinite population where you can have one infected and still have a completely susceptible population. Infinity minus any finite number happens to evaluate to infinity. But if you calibrate a model to a specific

outbreak of a specific disease, for instance,[76] then you will with certainty get different values for all models. In other words. R_0 tells you more about the model than the disease and the context in which thrives.

Almost invariably, the specific calculation of R_0 relies on parameters that cannot be obtained, parameter values being notoriously hard to come by in IDE. Most of the time, proxy variables too are scarce, for instance. [77] If the parameters are too many, for instance,[78] this might also cast doubt on the estimate, parameter values being notoriously uncertain in IDE. For these reasons R_0 is rarely calculated from formula using parameter estimates. A variety of techniques are available to determine R_0 from outbreak data.[59, 79]

For computer models, following the definition may be very hard. Even in computer models, people are not always indistinguishable. Moreover, we cannot have an infinite population in a computer but have to make do with a very large one. To complicate matters even further, the threshold value may be larger than 1 when considering heterogeneity in contact patterns and stochastic models.[80, 81]

All of these issues and the misconceptions make the widely popular R_0 very problematic today. The conceptual value of the epidemic threshold, however, whether in terms of R_0 or a critical population immunity level, Section 4.3, *The threshold concept*, cannot be underestimated. The concept implies, as well as it explains in simple cases, that epidemics initially grow near exponentially, gradually losing pace before a peak is reached and a certain fraction of the population is infected, always less than the whole. Then it gradually declines and dies out. In extension the threshold concept is integral in providing support for herd immunity i.e. the prevention of epidemics by the reduction of susceptibility. There is a given level of susceptibility below which an epidemic is not possible. Below it - in other words, if vaccination is unsuccessful - an epidemic is possible. This depends on the disease and of course depending on the disease it may even be possible for eradication, like smallpox and polio, rather than just containment. I will use this opportunity to stress the importance of global vaccination. We vaccinate to protect not only ourselves but the entire community. One less vaccinated may make a world of difference.

5 WHAT IS A MODEL?

We have come quite far without in so many words developing what a model is. There are several concerns with our line of research from scientific and philosophical point of view. I will, if I may, attempt to put this thesis in a philosophical, epistemological and semantic context.

A model is an idealized representation of a system or process. A model airplane perhaps would be the first thing a child thinks of and this would entirely fit the definition. In such general terms, a picture of a glass of water is a model, as would be the model our brain makes for ourselves when we see the glass of water, real or depicted, from the input coming from our perceptions, as electrical impulses passing through our nerves. The model can be reproduced merely by thinking of a glass of water. "Our brain runs first class simulation software", as Richard Dawkins put it.[82]

For practical purposes, in science, what we generally converge on to be a model is an abstraction, a description in a suitable language that makes sense of physical phenomena, processes or objects, that serves to convey knowledge to others and that enables us to make predictions of our world. We can discern three, perhaps self-evident, worthwhile goals of a good model: *explanation, understanding* and *prediction*. In *validating* a model, it is desirable for it to be able to *explain* past occurrences and *predict* future ones, but if it doesn't offer much in the way of *understanding* then what better is it than a glass of water?

One such suitable language that has proved convenient in representing the universe is of course mathematics. Even in the realm of quantum physics or at speeds approaching the speed of light, where neither our perceptions, common sense, imagination nor the simulation software of our brains, have been adapted to convey what is going on, mathematics still does not fail us. Surely, mathematics would stand the test also where epidemiology is concerned?

5.1 MODELS AND COMMUNICATION: A DISCONCERTING EXAMPLE

There are problems with mathematical and computer modeling also concerning communication in the context of *validity*. Introducing one of my favorite models, the often quoted Moore's law. Moore's law describes the long-term trend in computer hardware, hypothesizing that computing power is doubled every two years[*].[83, 84] It was stated by Intel co-founder Gordon E. Moore for the density of components on computer chips and can be applied to memory size and megapixels in digital cameras, to name only two examples. Moore's law has become industry standard so it's rather a self fulfilling prophecy, which is partly why I like it.[85] Originally it didn't offer any understanding of the driving force behind the development, merely prediction. Now it *is* both driving force and explanation. The issue is raised of a model's raison d'être. What in terms of understanding is offered from this model?

The main point I want to make here, however, has to do with communication and the language of mathematics. When stated verbally, Moore's law is easily understood and conveyed. What's more, it is understood by the recipient as a rule of thumb, an approximate law.

[*] Originally it was stated as doubling every one year and if you factor in the increasing performance of transistors, the coefficient in terms of processor *performance* should be closer to 18 months.

Moore's law can also be stated as

$$K = C \cdot 2^{at}$$

or even

$$K = C_1 \cdot e^{bt}.$$

Immediately it becomes unintelligible to the population at large and, to a great many others with the theoretical capacity to grasp it, not worth their while. But above all, the mathematical language has granted it, to put it in the words of Giesecke, "an air of exactitude which [it] doesn't deserve".[31] The time coefficient *b* in the equation above, when fitted to benchmark tests over the past decades, is 0.46. Writing the equation with this value, as in

$$K = C_1 \cdot e^{bt} \text{ for } b = 0.46,$$

makes the crime even more severe with Moore's law appearing something of a law of nature.

Quite often the opposite is true. Models start of with some sort of verbal, hypothetical reasoning. By necessity, by the time they are formalized in mathematical terms they have become so simplified that they hold less truth than verbal ones. Epidemic models are not exempt.

The recent financial crisis has put the spotlight on modelers or "quants", and their day-to-day life. Financial models has a very problematic relationship to reality. In the same way as epidemic models they attempt to describe human behavior – that of the buyers and sellers on the markets. When a model has been generally adopted, however, it tends to affect the behavior of the very same crowd it is modeling. Taking this into account would require a meta-model and so on. Without any further insight into how these models work, it is plain to see how a self fueling chain reaction could emerge and topple the markets.

The ones that commission the models are the ones that make the hard decisions. There is of course a communication issue when the models are so complicated that the decision makers can hardly be expected to understand the benefits or hazards. At the same time, the quants may not be directly connected to the "real world" or have a stake in the outcome of their work, other than their career, of course. An oversight can be disastrous. Contrary to what people might think, I believe that the experience and gut feeling of the men and women working the market is really what stabilizes it.

5.2 WHY MODEL?

If you stretch the definition, as I did above, modeling is an integral part of all scientific research after the observational stage. It is not enough to simply observe that viruses reproduce between cells or humans. Eventually, you will need a model to explain why, communicate your understanding and use it, for example, to design drugs. It can be stated in conversational terms or very specific and it's usefulness in different applications depend on the language. In some instances modeling is our only resort for empirical study as conducting live experiments are impossible, perhaps due to inadequacies of measuring equipment, or, as in the case of epidemiology, ethical constraints. We cannot experiment on the population as much as we'd like and so we resort to models.

This in point of fact, why we model in infectious disease epidemiology (IDE). Usually all three of the following prevails: limited observational data, incidence data of poor quality and transmission routes cannot be inferred. Our measuring equipment is severely impaired. As much as we would like, we cannot introduce a new strain of a disease into population in order to carry out empirical studies. That would overcome our measurement issues, but it's still a bad idea. Modeling is essential in IDE as perhaps in no other field.

In fact, we don't know the infectivity of most diseases. What is the chance of getting the flu from sitting down over coffee with an infected friend for one hour? One well known opportunity for hands on measurement occurred on a commercial airliner that was forced to remain on ground for three hours with the ventilation system malfunctioning. One passenger had recently been infected with influenza and the researchers were able to count the number of people he infected and look up where they were seated.[86, 87] Such opportunities are extremely rare.

The difficulty in measuring infectivity is apparent but is not only inherent of influenza. Measles is so infective that it is virtually impossible to determine who was infected by whom. It could be a family member but may as well have been someone walking by an open window. SARS is dormant for the moment, perhaps it does not exist anymore, however vital it is to study in preparation for the emergence of the next deadly respiratory disease. Avian flu hasn't emerged as a disease able to spread among humans, so that's anyone guess.

Of course, once the infectivity is established, we need to converge on what exactly constitutes an average contact among friends, family, colleagues and people we don't know. How many times a day do we engage in such contacts and how does it compare to the one hour coffee in time and intensity? We'll talk more about sexually transmitted infections (STIs) in upcoming chapters but perhaps this is one area where the activities surrounding an infection event should be comparably easy to establish. However, the incubation time and opportune transmission window of HIV usually leads to problems determining the transmission route. Add to that our shallow knowledge about what goes on in the bedroom, the length and intensity of sexual relations and concurrent relationships.[88] After more than a decade of research the medical community is still uncertain about mechanisms of the protective effects of circumcision.[89, 90]

The war against infectious diseases is going to be long and it is one which can only be won through better understanding and *communication* of understanding to decision makers and the public. Modelers have for a long time been concerned mainly with this. We would use this understanding for the design of effective intervention strategies to stop or mitigate epidemics. If there is one area where epidemic models excel, it is when it comes to the design and testing of such control measures. This must be the main goal for epidemic modeling, to which I'd like to see this thesis as a contribution.

5.3 ARE MODELS ANY GOOD?

With quite simple epidemic models we have, at least conceptually, been able to understand the basic dynamics of disease transmission, how a few key parameters can drive an epidemic. Models have been extremely successful in supporting some of the basic hypotheses concerning epidemic spread, empowering such civilization advances as vaccination policies.

Key issues remain. We have not yet been able to explain satisfyingly why Chlamydia remains endemic or why AIDS does not spread like wildfire*. What's more, of the three benchmarks by which we can evaluate a model, explanation, understanding and prediction, the last eludes most epidemic models. At the time of writing the first draft the novel influenza (swine-flu) was fully ablaze. Even with tens of thousands of infected, it was anyone's guess what the final toll would be.[19, 92-94] Operational scenarios in Sweden ranged from 400 000 to 5 million.[19] Even though, or quite frankly, just because, people are active in epidemic modeling as never before, one can ask whether any of it is justifiable.

5.3.1 Validation

The example with Moore's Law and financial models above raises the important issue of communication. This section, although provocatively phrased in general terms is about how we can validate a given model. Ignoring the fact that Moore's Law and many financial models turned out to be self-fulfilling, they are easy to verify by comparing to reality. The law can be matched to the real development, as is the case with many models describing real world phenomena. Their validity seems unquestionable.

The issue of validation is rather more problematic with epidemic models. We have very few opportunities to test epidemic models. The above mentioned flu pandemic had by the time of writing the second draft, proved most models inadequate, at any rate.

I implied earlier that a model for prediction that offers no understanding is little better than a glass of water. An important insight into modeling can be gained however, by realizing that modeling depends completely by the intended function of the model. Most often, for practical purposes, a model for prediction is distinct from a model for understanding. In the end this all comes down to meta-knowledge or epistemology. What is meant by understand something? That's the brain successfully building a model in our head. Nesterov[95]:

> *Generally, a model is a simplified embodiment of the aggregate properties of the modeled object. This simplification also implies the subject of modeling i.e., the subjective choice of a certain totality of properties of the modeling study object.*

This means that there are different models of the same object. Which is the best is a matter of taste and preference, but primarily perspective, the research question and data. Is there a way to determine if a model is right or wrong? The subject is called *model validation*. The reductionist/logical positivist schools,[96] including empiricism among others, would say that anything short of accurate comparison against empirical facts is to be abandoned. This would clearly disqualify the whole of epidemiology and sociology as well as most of medicine. The validity is always in dispute, perhaps the trait that IDE has most in common with non-IDE .Both fields require the widest definition of scientific method to thrive. The would-be data against which to validate model output is flawed, incomplete, complicated with factors not of interest, rendering them questionable in terms of empiricism.[97] Subsequently, we reuse this data to deduce

* This might strike the reader as a peculiar statement given the millions that are infected, but the truth is that given what we know about the transmissibility of the disease, HIV spreads slower than expected.[91]

parameters for modeling. We also use it to validate the model outcome as well as functions within the model, what's called internal validation,[98].
Internal validation, more common with complex computer models, may be summarized by the rule, "right behavior for the right reasons". Before Lorentz and Einstein, the concept of the "aether" was widespread and accepted, a substance serving as a medium for the propagation of light and electro-magnetic waves. The model was perfectly valid in the domain of classical physics providing perfect predictions, at least up until the Michelson-Moreley experiment, and offered[*] both explanation and understanding. The equations it provided, notably Maxwell's equations are of course still used to a large extent but the model as such proved internally incorrect.

5.4 MODELING'S PLACE IN PHILOSOPHY

What this means is that not only do we have a problem judging the correctness of our model but also the incorrectness of the model, what's called falsifiability. An empiricist would sneer. Thankfully, the acknowledgement that logical positivism failed in being a useful foundation for scientific research, the post-positivist stance was born.[96] The issue is put to the point in the following joke. Two empiricists are observing the country side from the window of a train. One remarks on the sheep grousing on the pasture, saying, "Look, the sheep have had their fleece cropped." To this other immediately corrects, "Yes, on one side, at least."
The joke is enlightening in many ways but the core of the dilemma is hard to approach. Surely there must be some value to empiricism? I think the joke tells us that, though holding many virtues that are important for every scientist to be aware of, empiricism as a single guiding light will get us nowhere. No laws or axioms have ever been produced from the social sciences. With that in mind I find it amusing that the most diligent proponents of positivism are to be found in the social sciences.[96]
The post-positivist philosophy of science may perhaps be a more useful one to abide by. But even post-positivists would feel uncomfortable with epidemic models, still preferring the controlled environment of a laboratory. Perhaps this is where infectious disease epidemiology and epidemiology comes in closest contact with each other. The validity of the models of either are always in dispute and both fields require the widest definition of scientific method to thrive.
Fortunately we are not without philosophical support. Model validation is in fact strongly related with the *justification of theories*, and this is still an unresolved question of philosophy of science.[98] Our line of work fits snugly in the void. Barlas and Carpenter[99]:

> *No particular representation is superior to others in any absolute sense, although one could prove to be more effective. No model can claim absolute objectivity, for every model carries in it the modeler's worldview. Models are not true or false, but lie on a continuum of usefulness.*

[*] I should say "offers". Many would reinstate the theory if only for pedagogical purposes. It has come to my understanding the aether-school undergoing somewhat of a revival as some concepts in electrodynamics are easier to understand assuming aether. Such wildcards are common in physics. As for explanation, the theory did well in its domain. No one would replace Newton's classical model completely, after all.

Furthermore, House[100] offers that the object of science is to attempt to explain even though we fail to predict.
With the support of these arguments, the modeler would deviate from the demands of right/wrong and venture into the more lenient requirements of usefulness. This does not preclude modelers from their responsibility of validity, it just means that it's not entirely clear cut. The process of validation requires testing, interpretation and, not least, communication until enough *confidence* in the model is achieved.
For an in depth discussion about modeling in conjunction with the philosophy of science I recommend Barlas and Carpenter: *Philosophical roots of model validation: Two paradigms.*[99]
The question of correctness and validation pertains to one of the themes of this thesis of complexity versus simplicity, and to the upcoming section about sensitivity analysis, see Section 6.4. One way of overcoming the demand of giving a correct answer is by covering the range of uncertainty with answers, in a sense, spraying the target with fist full of pebbles.

5.5 SIMPLIFICATION AND COMPLEXITY

Any computer simulation is doomed to succeed.[101]

This is a popular quote attributed to Rodney Brooks, constantly reiterated amongst modelers. A modeler needs constant reminder that he is precisely as good as his tools, always running the risk of ending up with a model that is "better" than reality.
You should strive for simplicity. Obviously if you can make accurate predictions with a simple model, this is to be preferred. At heart we believe that the occurrences around us are based on simple mechanisms, and although humans are complicated beings, social research starts from this view point.[100] From this line of thought springs the term *parsimony* aka Ockham's razor. All things equal, choose the model that is most parsimonial.[102] Something of a philosophical axiom.
Simplification and idealization is a foundation for understanding. So even if what is being modeled is through and through very complex, if a particular aspect with some accuracy is captured by the simple model, this is to be preferred, according to Ockham's razor, since the possibility to analyze and in extension, understand, makes it warranted. Perhaps different models are required to capture different aspects, each one inadequate to completely describe the system. Einstein's laws of motion are more accurate than those of Newton, but at low speeds the latter's are sufficient and certainly offer more in the way of understanding. Ask any high school student.
Modeling in epidemiology is perhaps harder than in most fields. The problem is, as I have mentioned, that the validation, both internal and external, is difficult due to the scarcity of data, and the quality of it, where it exists. When discussing simplicity, it is often expressed as a balance between simplicity and validity. Sometimes even between accuracy and complexity .[103] If anything, both are equally weighed and in my experience, the situations where you actually realize this balance are rare.
One difference between mathematical modelers and computer modelers, is that mathematics semantically imposes restraints on the complexity. Computers let the modeler get quite carried away if he or she is so inclined.
Computer or mathematics aside, the *simplicicist* side boasts few assumptions and interpretable results which serve to underpin hypotheses about the real world but fall

short whenever prediction is discussed. The *complexicist* side claims, in theory, limitless adaptation to the real world, but relies heavily on a mass of assumptions which cannot be verified.

So what does all this mean in practice? Achieving simplicity, parsimony or idealizations means making *assumptions* about the environment. So the simpler model, the more severe and numerous the assumptions. Making assumptions is really the core of modeling. We can hypothesize simple rules for human behavior, in particular contact behavior. We then say, "Under the assumption X we predict Y"

One common assumption we have already covered can profitably replace X above, namely, homogeneous mixing. We claim to be supported the "law of many" i.e. statistical properties of the population. If on average people behave a certain way, it doesn't matter much if one or other deviated. Again an assumption. In many cases, the oddballs are what really matters, certain people with very high rates of contact, for example.

Barlas and Carpenter ask a valid question[99] which, although pertaining to *system dynamics*, is valid also in a more general scope. Are models scientific? The answer is: It depends on your definition of science. If anyone thought that science was absolute, at least it's open to debate. The relativistic view of science allows for it.

However, there is a crucial point I cannot find in Barlas and Carpenter or anywhere else: the issue of repeatability. The increasing complexity and the wealth of assumptions that can be crammed into a computer model, means that a bird's-eye view of your work is dispensable. At best you have an intuitive feel. The next scientist will not share that intuitive feel - at best have his/her own.

Epidemic models of today are hardly repeatable. They are vastly complex with thousands of lines of code that are difficult to test and validate internally, let alone externally. With anything less than copying the code, your fellow scientist cannot duplicate experiments. To be able to duplicate an experiment this is one of the foundations of peer-reviewed science and is even considered a definition of science on some level. This is a question worthy of a response but I haven't found one. Meanwhile I urge all modelers to seriously consider this issue and strive to make their experiments repeatable and their papers descriptive. You should be able to rebuild the model from the description in the paper.

6 MORE CONCEPTS IN EPIDEMIC MODELING

This section contains some further considerations and queries in epidemic modeling. Some pertain especially to the contributing papers and serve as a further introduction before the actual papers are described in the following chapters.

6.1 STOCHASTIC MODELS

The randomness human behavior, the society we have built around us as well as completely haphazard natural occurrences, make exact predictions a largely meaningless activity. A cashier at IKEA may have been infected with a new strain of measles but just as he fell ill had decided to seclude himself in the archipelago instead of going to work. Chance occurrences have a terrifyingly large effect on epidemics. *Stochastic* means random. The opposite, *deterministic*, means predetermined. This can be understood in the context of the definition for R_0 in previous chapters, see Section 4.5, Finally R_0. We noted that an R_0 above 1 means that *can* occur. The stochastic modeler will calculate the probability of an outbreak given an R_0 and a set of initial conditions. For a deterministic modeler the *can* is replaced with a *will*.
A model that is stochastic by design, will give a distribution of possible outcomes, e.g. Colizza[104] and indeed Paper I and II of this thesis.[105] This in contrast to a purely *deterministic* model, e.g. Rvachev and Longini.[106] If the model is computer simulated, as are papers I and II, then many runs must be carried out in order to get an idea of the range. Calculating any type of statistic with acceptable precision usually requires an order further of simulations. In Paper I the main scenarios are run 1000 times each.

6.1.1 Some misunderstandings

As a point of order, I should like to be clear about a few concepts. A deterministic model always produces the same result from a given set of initial conditions. In stark contrast, a stochastic model lets chance influence the result. In practice deterministic means formulating a set of *differential equations* and solving them *analytically*.
Stochastic modeling entails a *stochastic process*, in which a series of events follow one another based a rulebook of probabilities, usually *branching*, i.e. one event triggering several to follow.
Between the two types of modeling, the latter is comparably young. The stochastic process can, with benefit, be simulated with a computer. In simpler cases it can be translated into a system of *differential equations* - same art, different genre - for the *analytical solving* of the so inclined mathematician. The possibility to do so means that the science can be advanced and firmly based on solid, stringent mathematics.
The *analytical* solving of differential equations from either source can with benefit be done with a *symbolic mathematics* package on a computer. As if this wasn't confusing enough, we can add the fact that differential equations can also be solved *numerically* with a computer, when the analytical method is too burdensome or impossible. Or by hand, of course. *Numerical analysis* preceded computers by about 2000 years and forebode them by 400. Anything done by computer can theoretically be done by hand, although the size of feasible problems may be limited.
Numerical solving is not the same thing as *simulating* the actual process but should of course lead to the same conclusions. The word simulating precedes the computer and

simply means realizing the probabilities involved, as a throw of a dice is to simulate the discrete uniform distribution of numbers one to six.
I bring this up because of misconceptions I often come across: that stochastic necessarily implies a computer and even that deterministic would necessarily imply pen and paper. Both have clearly developed alongside computer processing power. Most mathematical modeling of epidemics today is by means of stochastic modeling and thoroughly analyzing them can be done with different techniques that do not necessarily rely on the use of computers.
Stochastic models, in a way that deterministic models can't, offer a richness in the distribution of results. The stochasticity means that many random events can occur that determine the outcome. In epidemic modeling that means that we can set R_0 to 5 and with a stochastic model still find that certain runs dying out quickly. This is to be expected, as, recall, R_0 above 1 indicates that an outbreak is *possible*. A deterministic model, such as the classic SIR-model does not give you this information. An R_0 above 1 invariably produces an outbreak. The distinction is of course also a cause for further misconceptions regarding R_0 itself.
The model in Paper I and II is a stochastic model, "solved" by *simulation*. As I said above, that does not mean the model definition is void of differential equations. And the fact that I use a computer to get results does not mean it is impossible to analyze it with a pen and paper. In fact, as is done in Paper I, the equations give you the probability of extinction in a simple case of the model.
Since differential equations may be hard not only to solve but also to actually state, stochastic models also do not, in the same way, restrain the creativity of the modeler in producing ever more complex models. As should have become clear by now, this is both a blessing and a curse.

6.2 MEAN RESULTS

The mean outcome is a useful quantity when analyzing the data from 1000 simulations of a stochastic model. It is the *expected* outcome of the model but that doesn't mean it will ever occur, even in theory. The expected value of a dice throw is 3.5. It means that averaged over many runs, this is the mean outcome. Incidentally the expected outcome is precisely what a corresponding deterministic model will give you.
Some time after the model is proposed, published, with a prediction in expected size and an R_0, perhaps an outbreak really occurs. Our epidemic model will have, at best, been able to predict the probability of roughly that size of an outbreak occurring. "Hopefully" the outbreak falls within a range of possible outcomes predicted by the model, more or less likely, and not disqualify the model, not precluding false negatives. There may be certain ways in which we can calculate the *likelihood*[107] of the model producing this outcome, as a form of validation, see Section 5.3.1. The closer the real final size to the projected mean outcome, the higher the likelihood, but really, one occurrence is not enough data to make such a calculation meaningful.
There has also been devised methods in which the predictability of models might be assessed and hence under what circumstances variables are most accurately predicted. Colizza *et al*. introduced more statistical mechanics into IDE, using a measure of entropy to account for heterogeneity in prevalence over the world over time and suggested that the network structure be responsible for most of it. They also suggested a measure of variability between runs and in extension predictability of simulations as a

whole, a welcome contribution. It is one thing to account for uncertainty, another to quantify it.
The featured object of enquiry, that of prediction, does not have to be abandoned completely in favor of understanding and explaining, for there are other measurements that might be made on an epidemic that could make the likelihood more reliable and hence the model more useful. We might use the daily incidence. In fact every transmission event from one person to another can be used as parameter input in an epidemic model, if it can be accurately measured. Failing that, an educated guess about the probability person A being the one who infected person B, known as contact tracing, see Section 3.2 page 10.
Using contact tracing and directly feeding the results in this way to the model is what real time modeling is all about.[79] We could use such a model as the epidemic is underway and this might give us invaluable information about the optimal course of action in order to impede the process.
Modelers are currently developing tools for use in an epidemic situation to predict events weeks in advance, possibly only days. The emergence of the novel flu pandemic[108] demonstrated that this type of resource is being assembled everywhere around the world. In a matter of weeks results were produced for use in speedy policy making.[19, 92-94] Dishearteningly, I have information that the usefulness of these attempts are questionable.[109]
The product resembles weather forecasting. Our models may be sufficiently accurate to make good predictions provided we have 90 % of the info: the infectivity, the number of people already infected, where they are. As is discussed in Paper II, the initial phase of the epidemic, the very first few infected in an area, largely determine the outcome of the epidemic.[105, 110] This is true of simple and complex models alike, as well as of the real world. Once the initial phase has been passed, we have made the transition from a stochastic domain to a deterministic (statistical) domain where deviances are insignificant in comparison with the underlying trend.

6.3 ASSUMPTIONS REVISITED

When we talk about assumptions in epidemic models, we are usually referring to assumptions of human mobility and heterogeneous contact patterns. Increasing the complexity of the models we may add age dependencies[111]; structures such as families and workplaces on the lower levels; regions and countries on the higher levels.[6]
Halloran and Longini[112] assumes 2000 individuals divided into only two communities. In Paper I we model 9 million Swedes into their natural division of their 300 municipalities.
Another approach may be to assume that human interaction takes place on social networks. We define people as *nodes* or *vertices* and connect them with *links* or *edges*. We may model the network explicitly and simulate[113] or use the statistical properties measured for social networks and other mathematical results to directly calculate variables of interest.[114] Networks may be artificially generated to mimic the contact structure of a population and indeed the growth of such structures. In this way much has been learnt about epidemic spread in general.
The network approach is especially appropriate for sexually transmitted diseases. In Paper IV we use survey data to deduce certain values from an implicit network

structure. We also explicitly build networks in a computer to simulate disease spread on. I will therefore discuss networks in some detail in conjunction with Paper IV.
As these assumptions and the models become more complex, mathematicians will lose their motivation to solve the problems. While equations are always important describing all of the model or certain components within, it is no longer possible or even desirable to find closed-form solutions. A computer is nowadays an essential tool for the epidemic modeler. Today, micro-models, or individual based models, where each individual has its own explicit representation in the computer, have been accepted into the modeling flora.[103, 115] Contacts are made according to arbitrarily complex heuristics and day-to-day movement alike.
The wealth of parameters that are available in describing human behavior, contact structure, the disease in question and all the topics discussed previously, even leaving out the subtleties that cannot be quantified even poorly; all this leads to a very tricky question. Should we keep the model simple, simple to understand and to analyze, or should we try to capture as many aspects of the epidemic and its playing field as possible? Simple versus complex. In epidemic modeling, there are really no bounds to how far we can go in our assumptions.

6.3.1 Sensitivity analysis

"Rubbish in and rubbish out", is another popular quote from modeling. This tells us what goes without saying, that for every assumption you make you have to be pretty sure that it holds up. If there is an uncertainty parameter value involved, for instance contact rate, this will propagate and probably amplify to the result. If there is a high uncertainty, what you can do is to vary the parameter through the parameter *range*. This is called *sensitivity analysis* since it determines how the outcome is dependent on what we know of the parameter. In conjunction with micro-modeling, an appropriate term in the context, *exploratory modeling*, is sometimes used.[116]
With more parameters, it may not be enough to explore the range of parameters one at a time. It may be advised to perform sensitivity analysis on several or all simultaneously, that is, for each possible combination of values of each parameter. The *parameter space* will quickly become impossibly large. To add to the problem, complex computer models usually have a longer running time and are usually stochastic, in other words, requiring the combined outcomes of several runs. There are advanced methods to address this[117, 118] that can treat, if not cure.
Simple models will always face dissatisfaction about aspects of biology or sociology not being covered. Complex models on the other hand inevitably face the objections of large uncertainty and insufficient sensitivity analysis. Faced with unanswerable questions about the validity of a model, many find the simplistic approach appealing, offering at least some level of transparency.

6.3.2 Further assumptions

Further assumptions to be made are infinite populations, a sometimes acceptable assumption, providing many tools for solving equations. Meltzer[119] rather peculiarly introduces into computer simulations a *queue* of individuals ready to be infected at the rate determined. There is not even a theoretical depletion of susceptibles throughout the epidemic. The discrete nature of real populations has implications.[71] Exponential, fixed or other distributions of latency/infectious periods: exponential infectious times leads to an underestimation of R_0 and ultimately epidemic impact.[71, 120] Lloyd[121, 122]

demonstrated that realistic sojourn times introduce instabilities in oscillating endemic models.
We could spend pages on different assumptions that can be made, really each model and each paper has their own, although there are some common classes, most of them outlined here. The message is this: For the mathematician the goal provided by assumptions i.e. simplification, is mathematical tractability; for the computer modeler finite computing time; and for them both, understanding and explanation. These goals should always be weighed against the validity of the model and that is a very intricate handicraft not to be put down.

6.4 UNCERTAINTY AND COMPARATIVE RESULTS

Thank you for bearing with me through this dismal chapter. We have finally some light in the tunnel. There *is* vital information we can gain from epidemic models, in addition to just better understanding. We can now assemble the supports we need for Paper I.
Say a car manufacturer has to decide which of several competing materials for brake discs to continue development on. They might try to model the brake system, including the tire and road, and do simulations on a computer. They can test different parameters, different conditions like precipitation and temperature, not to mention speed. The impact of many factors will be uncertain or just guesses and this places the reliability of the results under some skepticism. The absolute results of these simulations, therefore, in terms of breaking distance, operating temperature etc. are not to be taken at face value. But we can compare the results. If one of the materials consistently performs better than the other then the manufacturer might be inclined to go ahead with that one.
The process is analogous to comparing distances with faulty ruler or time periods with a watch that is slow or off by a few seconds. Such a watch could tell you which of two was the fastest runner, but not whether any of them beat any records. This type of error is called *systematic error*. The prerequisite for a successful comparative result is not so much a low systematic error, rather a low *random error*, corresponding to a high *precision* in your measurements. Modeling can be a substitute for measuring and the same principles apply. Computer models happen to be very good at avoiding random error.
Needless to say, either error must not be dependent on the object of measurement. That would be dependent error or *bias*. Using different watches for different runners is ill advised. Computer models happen to be very good at avoiding measurement bias as well.
The same reasoning is used to justify very complicated models of epidemics under very uncertain circumstances with heavy assumptions. Usually, most variables and our assumptions come from experience or are just guessing work.[115] We think of our assumptions as *baseline* conditions that *cancel out* when we compare our results. Whenever possible, we try to test the full range of possible or plausible values for uncertain variables to see the affect on the results. We may also make minor modifications of our model to see if we get the same results. This is known as sensitivity analysis, outlined in Section 6.3.1.
We are of course very interested in the possible final size of a future epidemic to appropriately dimension facilities and vaccination stock. But take for example pandemic influenza. Nobody can say for certain what such a virus will look like and how dangerous it will be. The infectiousness that will determine how many will be

infected, is anyone's guess. The absolute results from epidemic models may only be reliable under very specific circumstances for instance if we have calibrated it to data from previous outbreaks of the same disease i.e. not very reliable at all.
However, decision makers may rely on the fact that comparative results *can* be used. This enables them to draw conclusions about something potentially more important namely the best course of action to halt or impede an ongoing outbreak. We may, for example, provide some enlightenment evaluating different strategies for vaccination. Is it responsible to vaccinate everybody, given side effects? Will vaccination of cases and their families be enough to contain the outbreak? Can we vaccinate workplaces, only women or a certain age span? A perspective on the last question is found in the current recommendation in Sweden for HPV-vaccination (Human Papiloma Virus): girls aged 10-12.
To illustrate, as a premonition of the forthcoming chapter, in Paper I we modeled a fictive "moderately contagious disease", our ill concealed inspiration collected from the recent outbreaks of SARS (Severe Acute Respiratory Syndrome). Since we didn't have enough data on SARS in Sweden – it never reached our country – we could not in good conscience mention SARS in the paper. In a sense, what we modeled was an altogether fictive disease. But we did simulate scenarios of varying infectivity such that we were confident that the parameters of a SARS outbreak in Sweden would lie somewhere in between the values we tested.
Our angle of approach was whether and how a certain type of travel restriction strategy would be effective in delaying the spread. In all our tested scenarios we showed that there was a significant reduction in the speed at which the disease dispersed in the population.
As far as sensitivity analysis goes, this is a concept almost entirely developed for complex models, implicitly computer modeling. I'd like to point out that comparative results and sensitivity analysis should not only be seen as methods to overcome the weaknesses of computer modeling compared to pure mathematical modeling, but as strengths. Though purely mathematical models provide specific answers, sometimes in distributions of probable outcomes, the range of possible events is never so clear and rich as when simulated thousands of times using random initial conditions.

6.4.1 Intervention and control

Given what we've learnt so far about mean results, comparative results and assumptions, the question posed in Section 5.2, Why model?, may be answered. As far as epidemic models are concerned, in practical use, epidemic models are at their best when evaluating intervention strategies. As seen in the previous section, this is possibly the only time we can claim with any persuasiveness that our results hold water. When comparing different scenarios, uncertainties to first approximation cancel out, leaving us with the difference between what would happen with an intervention and what would happen without. All the issues of validity and internal validity remain, of course.
Decision makers and the general public are more interested in how we are to control an epidemic - stop, contain or impede - than a hypothetical mean value of the number of infected or deceased. In my view the most important contribution of epidemic models have been to convince the world of the benefits of national and global vaccination programs. Even the simplest models show the imperativeness for parents to vaccinate their children against measles, despite probable or hypothetical adverse effects and despite that "everyone else is already vaccinated".[123]

Besides vaccination, many other types of interventions can be considered and modeled such as quarantine/isolation, screening/treatment, closure of work, schools and temples, facemasks, hygienic measures, environmental control (humidity, temperature and air quality) and, not least, travel restrictions.[124, 125] The last of these is of course is the subject of Paper I. Studies may differ in views of which is the best coarse of action, but all of them[126-128] underline the critical need for preparedness and speedy response, mandating a substantial crisis preparedness body.
Table 6-1 gives a more extensive list and some references for further reading.
Of all interventions, vaccination is the queen, royalty merited by history. When modeling is concerned, vaccination actually changes the initial state of the model by moving a portion from the susceptible class to the recovered (immune) class, leaving R_0 unchanged. Most interventions mentioned are designed to reduce the contact rate and thereby affecting R_0. Table 6-1 also lists the mechanisms through which the interventions act.

Table 6-1 Intervention strategies and their mechanisms

Intervention	Mechanism	References
Vaccination/prophylax	Herd immunity (Infectivity, Infectiusness time, Susceptibility*)	112, 113, 129-135
Quarantine/isolation/seclusion	Contact rate	77, 113, 136, 137
School/workplace/church closing	Contact rate	93, 138
Environmental actions	Contact rate (adequacy of contacts)	40, 139
Travel restrictions	Contact rate/pattern	2, 104, 140
Treatment	Infectivity/infectiousness time	126, 128, 132, 133, 141
Facemasks	Infectivity (susceptibility)	113, 142
Hygien/disinfection	Susceptibility	142
Treatment/destruction of vectors/reservoirs	Infectivity	61, 143
Fighting poverty*	Infectivity/susceptibility/contact rate, emerging infections	36, 37, 39

Furthermore there are many ways to vaccinate a population with varying applicability considering circumstances such as before or during the epidemic, for treatment or protection, time constraints and other resources, effectiveness of the drug, availability, cost and side-effects. To prevent or decrease the risk of an epidemic you may vaccinate the entire population or only those at high risk of infection and death.
During an epidemic the best option may be to spend resources on contact tracing and vaccinating those in contact with the index case.[132] Even though they may have been infected already, a timely shot could cure the subject before symptoms erupt or relieve symptoms after they erupt and in that case hopefully also stop/lessen infectiousness.

* Failing providing complete immunity, vaccines often provide partial immunity and/or reduced infectivity. Reduced infectivity can be accomplished administering the vaccine before or after the infection event.[129]

This is called *ring vaccination* or *culling*[144] and is in effect optimized mass vaccination, realizing that a dose is only effective or necessary when the subject is at risk of being infected. If the contacts of an infected are speedily traced and vaccinated this will hopefully break enough links before they are exploited by the epidemic, hinging it back into extinction.
Considering only the cost of vaccine, this is obviously much more effective since only a fraction of the population is vaccinated.[115] Also consider that a full scale mass-vaccination is time consuming and places heavy requirements on logistics. Ring vaccination may be a quicker response depending on the distribution systems available.
Considering the profound effects that travel has in modern epidemics, restricting travel seems like a sensible approach.[2, 7, 104, 140, 141, 145, 146] A disease can not spread if people stay at home. Travel is probably a fundamental for diseases emerging in the first place.[147, 148] Sanctioning travel only when absolutely necessary, may be all it takes to quench an outbreak.
I can't help but to finish off this section with a most important message. If vaccination is the queen intervention, fighting poverty is the king. Although this is not an immediate option when outbreak threatens, there can be no doubt that it is the most effective policy in the long run.[36, 37, 39] There is a mutual association between disease and poverty. Therefore all efforts to eliminate poverty in the developing world will undeniably benefit not only those areas, but in extension the wealthy parts of the world as well; and not only the disease prevalence and risk, but also the risk for new emerging diseases and strains.

7 AIMS

The goal of the contributing studies were to:

- Assess the validity for the model of Hufnagel *et al.*[7] in Sweden.
- Assess the feasibility of travel restrictions as a way to delay epidemic.
- Contribute to computer modeling methodology.
- Contribute to communication of epidemic modeling results.
- Demonstrate the existence of a giant component in the Swedish sexual network.

8 PAPER I - TRAVEL RESTRICTIONS AS A WAY TO SLOW THE SPREAD OF SARS

A disease would spread very slowly or not emerge at all if people didn't move about.[7, 37, 140, 145-148] In the days of the plague, disease spread like ripples on a pond. Towns were hit roughly in the order of their distance from the original source. Today you can make it across the globe in a day. The more people travel, the more quickly the disease will spread. Travelling can in fact in itself be the tipping point that transforms a minor outbreak, one that perhaps wouldn't even erupt on a local scale, into one of pandemic proportions. Medium and large scale models will need to address this in some way, either by randomly dispersing the disease across an area or taking into account the full network of different types of transport.

Whether or not travel restrictions are effective in preventing or delay the spread of infectious disease has been a matter of some debate. In Paper I we present a model, the results of which indicate a reduction in the speed of transmission if a certain type of travel restrictions is applied. But to stop it in its tracks seems impossible.[1, 2]

In theory it is simple. As long as a population is contained, transmission will not spread beyond. Travelers may be infected and infect other populations beyond the confines of the initial population. The more frequent and distant travelling that is admitted, the more probable the event of transmission and large scale epidemic. R_0 is raised and homogenized on a global scale. Many models have indicated what seems entirely self-evident.[106, 145, 149, 150] Some have discussed the possibility that many recently emerged diseases would not have existed were it not for human kind's increased rate of travelling.[37, 147, 148]

This text on the models in Paper I and II builds on what has previously been said about modeling in general and epidemic modeling in particular. All the concepts and methods introduced carry over to this work and I will explain how in detail in this chapter. In fact, the travel restrictions model is very near to a simple classic SIR-model. The only modification we have made is to divide the population into municipalities and simulate them separately and concurrently, adding means of disease propagation between the municipalities, simulating travel. We simulate this in a computer stochastically. The homogeneous mixing assumption is applied within the confines of the municipalities: everybody has equal opportunity to be infected as the next fellow municipal resident.

8.1 A SHORT REVIEW OF TRAVELING IN EPIDEMIC MODELS

Sattenspiel and Deitz[151] modeled mobility among regions, tying together everything from complete isolation to mass migration in a neat mathematical framework versatile enough to act as a foundation for epidemic modeling. They demonstrated its use by verifying it against a measles outbreak in Dominica, though as a practical model the data required to satisfy it is quite specific.

Going global and perhaps overreaching in terms of picking a tough disease to model, Rvachev and Longini ignited the travel models-era with their ubiquitous *A mathematical model for the global spread of influenza*, perhaps the most cited paper in the genre. I think they had more modest ambitions for the paper. They skimmed airline data that had become available and extended the classic SIR-model to a global context by submitting the MPU (minimum publishable unit). The paper illustrated the spread of

influenza in separate epidemic curves in cities all over the world, all being fed by one another.
The model is still hailed as a milestone in computer modeling but for me it has little to do with computer modeling. It took a computer - today, an iPhone would have sufficed - to solve the equations but it's a completely deterministic model that doesn't exploit any possibilities of computer modeling. It gives the exact date for the arrival of influenza in your city without any margin of error. More than demonstrating what could be done with computers, it demonstrated what could be done with travel data if only scientists put a little effort into finding it. As such it has deserved merit. An epidemic modeler is not an epidemiologist until he/she starts searching for and processing real data.
As for exploiting the possibilities of computer modeling Longini's more recent paper, Longini *et al.*,[126] in the commotion of the avian flu, has a lot more to offer. While we expressed concern about the simplicity of the Rvachev/Longini-model, now we are allowed to crack down on the far reaching assumptions in the Longini *et al.*-model. The paper was published in Science the same day as Ferguson *et al.* targeted the exact same issue in Nature. The battle was furious. In my view Longini *et al.* came out on top but the comparison was a vain exercise in weighing assumptions and hypothesis. The H5N1-avian flu pandemic thankfully hasn't occurred yet.
This was a parenthesis because neither Longini nor Ferguson neither had computer power nor data to support a global simulation. Going back to Rvachev and Longini, they had a few followers like Grais *et al*.,[145] for example, advancing the paper with better and more up to date data. Flahault *et al*.[152] just copied the whole thing, used another dataset specific to Europe and solved a subset of the global problem already published by Rvachev and Longini. As for Flahault *et al*.'s case for airline travel and HIV[149], it's just too far fetched.
Brownstein *et al*.[1] made the uncontroversial claim that airline travel had anything to do with the inter-regional spread of influenza in the U.S., using time-series analysis. A three page critique of the paper is available on demand from yours truly. Still time series analysis is a welcome approach.
As this thesis suggests, don't forget mathematics. Going back to fundamentals in metapopulation systems, and the math introduced by Rvachev/Longini and also of Ball[153], Colizza, Vespignani *et al*.[6, 154] explore the subject in depth. The mathematics of either of these is way beyond my capacity to comprehend, though. I do however recommend reading Colizza, Vespignani *et al*.[140], summarized on page 27, for those not disheartened by just a little bit of math. And, of course, Hufnagel *et al*.[7], to be detailed shortly.

8.1.1 Travel restrictions

Now that we know about the effects of travel and in particular airline travel, what do we do about it? The papers reviewed so far all suggest that given the impact of travel in general and airline travel in particular, on the global and local spread of disease, restricting travel ought to be very fruitful in endeavoring to control the spread.
The paper which inspired Paper I, Hufnagel *et al*. [7], did venture into a complicated control application, although I mainly used the basic modeling part. How to or even whether or not it was possible, to control an epidemic with travel restrictions was not established when I started off. One year into my research, Cooper *et al.*[2] said it was not.

8.2 THE HUFNAGEL-MODEL

The model builds on the paper by Hufnagel *et al.* [7] in so far as the dispersion equations found in the paper, the method of estimating the so called *travel matrix* and some of the parameters used. A travel matrix can be pictured as a distance table one sometimes sees in atlases, featuring a daily bulk load of people traveling between origin and destination in place of distances. While Hufnagel *et al*.'s model spanned the globe, we limited ourselves to Sweden where we had detailed travel data.
Hufnagel *et al*. simulated a SARS outbreak using a simple SEIR-model* in the catchment areas of international airports worldwide corresponding to our municipalities. In a SEIR model every individual can be in one of the following states: susceptible (S), latent (E), infectious (I), and recovered (R). These local processes were linked together by the international aviation network, which enabled the disease to be transmitted along flight routes.
The main interest of Hufnagel *et al*.was to validate their model approach. It was fitted to the real SARS outbreak of 2002-03 by tweaking a single parameter γ. All that was needed to run the model was a point of origin, current aviation data and the value of γ in order to capture the infectivity of the disease in question. The latter was estimated by very simple means, simply as the quote of infected in Hong Kong and those outside. To me, for all its apparent complexity, it is beautifully simple. The startling accuracy to which it fit the actual outbreak is impressive, though probably a fluke. Of all countries that were suggested to be afflicted only Japan was a false positives. The frequently mentioned Colizza, Vespignani *et al*.[140] offers an explanation for their accuracy by analyzing the *predictability* in a nice way.
Earlier we mentioned real time models that are used during the actual outbreak to aid in putting into force precise preventative measures. New information is continuously entered and the model is rerun for up to date predictions. Hufnagel *et al*.'s model in my opinion has proved its worthiness in this application in the event of a similar outbreak of respiratory disease in the future.

8.3 RESTRICTING DISTANCE

The goal with our version of the model was simply to assess the feasibility of the model in Sweden given the traffic data *we* had access to; then to find what it would take to inhibit the spread. Already at the outset, aviation data was considered insufficient in resolution to capture the complete picture of travelling going on within our borders. Also the catchment areas of our very few airports are quite large and hard to estimate. Our travel data covered Sweden and was very detailed, giving us the required resolution at municipal level. It also enabled us to suggest a slightly modified version of travel restrictions than is commonly found, namely, restricting distance. If we removed all the daily trips over a certain distance, how would that affect the transmission dynamics?
As for the parameter γ, SARS never reached Sweden and so this value could not be estimated a posteriori. This in addition to the fact that the *type* of travelling introduced - car, bike, foot etc. - would have affected this value. In other words we had to be careful to test a range of possible values to support our claims, pertaining to our discussion about sensitivity analysis. Incidentally this is the reason we make no mention of SARS

* In Hufnagel *et al*.'s paper as well as in mine, the equivalent term SLIR is used.

in the paper. With the best of intentions we could not claim that we were modeling SARS, a hype-word in those days. Its mention would have certainly added citations to our paper. In my thesis I have no inhibitions against speaking freely about the ill concealed original intent.
Hence, the working title featuring "SARS" was changed to "SARS-like" and finally "moderately contagious disease". On a note of interest, "contagious disease" is an obsolete term, according to my supervisor, Johan Giesecke,[31] superseded by the more correct "infectious disease". According to his book the contagious suggests an infectiousness somewhere between moderate and high. The contradictory anachronism which was the title, underwent several revisions without him intervening, however. The final revision, from *highly* to *moderately* contagious, was on suggestion by a reviewer. By that time the prevailing view in the infectious disease community had shifted. Although the SARS-outbreak was very swift, it was understood that a comparatively high dose was required for infection.

8.4 A STOCHASTIC MODEL

The SARS-model is a stochastic model, see Section 6.1, Stochastic models. The number of infected may differ from one simulation to the next, as may the geographic distribution of the infected, even though we're using the same model and the same initial configuration. The only thing that changes from one run to the next is the random numbers used to generate the events. This means, in full agreement with IDE theory, there is also always a chance that the outbreak doesn't pick up enough momentum to become a full scale epidemic. The disease instead meets early.
The randomness comes into play in the simulation when deciding who gets infected and who does not. Given a certain number of infected in each municipality, we can calculate a probability for one remaining susceptible becoming infected. This probability is determined by the model and the disease parameters, including the travel matrix. In principle, we throw a minutely faceted dice for each susceptible and check the result against the probability. This is the basic difference between a stochastic and deterministic model. If it were a deterministic model we would be satisfied with the expected - or mean - outcome. Indeed, if we perform a large number of simulations, the average outcome should compare to the outcome of the corresponding deterministic model.
In practice, the outcome is determined quite early in the model. Whether the run will peak at 50 days and or at 120 days is determined in the first few days. The importance of the initial phase is logical if you think about it. Once you have several thousand individuals infected you will have a high probability of infection. For each individual you throw the dice. If the outcome is well defined, if we're expecting 100 additional infected, the result may turn out to be 99 or 101. The difference will be of little consequence for the production of further infected generations.
At the start of the simulation, however, the number of infected are few and the probability very low. Of the millions of dice throws we perform, say we expect 2,5 additional infected. Realized, 2 or 3 infected seems approximately equally likely with 1 and 4 not far behind. For the course of epidemic the difference between one and four infected is enormous, however.
Thus, in principle, most of everything is determined in the initial phase of the simulation. That being as it may, since we simultaneously simulate the model in some

300 municipalities, there are many initial phases happening over the course of the simulation, contributing to the erratic geographic behavior exhibited. Outbreaks may die out within the confines of municipality only to fire up again due to infection from a traveler from a municipality where the epidemic is well under way. In Paper II we see why this is of particular importance.
The compliance level was introduced to determine how robust the results are to people ignoring the ban. In conjunction with this discussion we can add a mention about a *fear factor*, on par with the one used in Paper III. In the event of a wide scale epidemic of a life threatening disease it is not far fetched to conceive of a spontaneous decrease in travel, the official recommendation notwithstanding.

8.5 DATA

The travel data we used was from a survey executed over three years, from 1999-2001.[20] The dataset is comprised of the collected responses of a detailed questionnaire completed by 17 000 respondents together with interviewers. The response rate was 73 %. The respondents were divided equally over the survey period and asked to record their travel activities of their assigned day. All mediums of travel were covered, from airplanes to boots, for whatever purpose. Having disability service, the cost of the trip and the number of members in your household were some of the variables recorded, giving you an idea of the detail of this survey. Most importantly for us, the origin and destination was recorded.

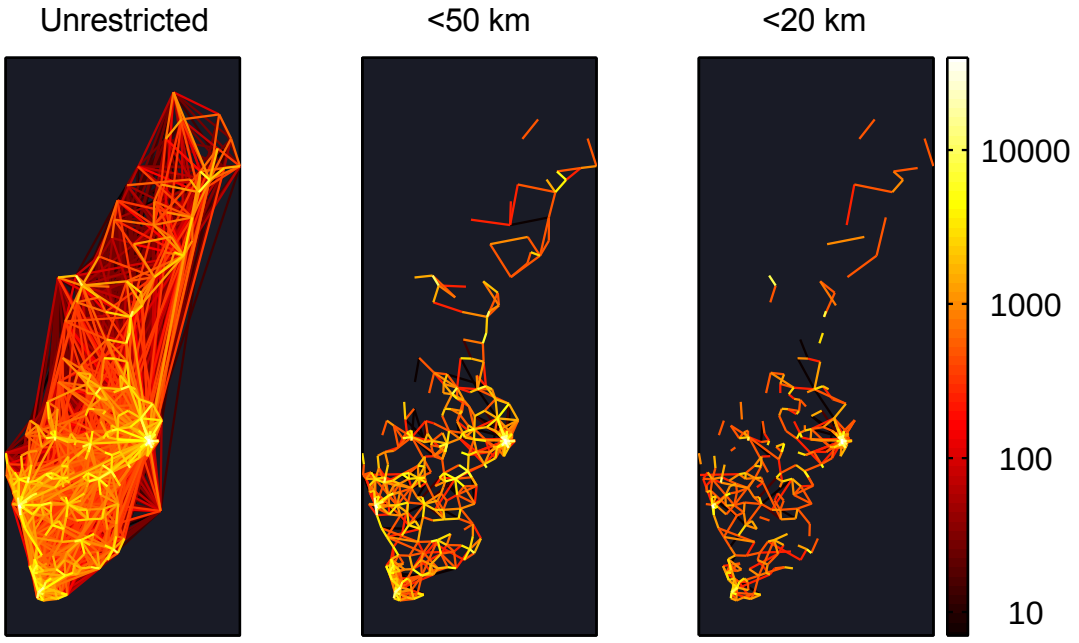

Figure 8-1 Travel matrix plots
The full travel matrix compared to the restricted matrices, with trips over 50 and 20 km respectively, curbed. The travel intensity is on a logarithmic colored scale.

Prior to the assigned date, the respondents were asked to log long distance travelling over 60 days, with somewhat less detail. This measure was taken to attain a proper sample of long distance travelling which was poorly represented in the daily travel log. We actually used only the long distance data in our paper. Even so we had to be careful. Even though the number of respondents is appreciable, as is the 35 000 trips they amassed, it is low compared to the possible travel routes between Swedish

municipalities, 289·289=83 521. Less travelled routes may not show up in our data, but the real problem is when they do. Due to the weighting, a sporadic travel between two unlikely destinations will over represent the actual travelling going on. We considered all single trips like this to be outliers, removed them and compensated by weighting up more populous routes. Thereby we lose those routes where the single trip really represents actual but rare travelling, ultimately under estimating the disease spread by travelling. Considering our goal, to show the consequences of travel restrictions, we felt it was better to be on the low side. Also, there is the matter of the parameter γ, which in the high end of its range will compensate, at least globally.

8.5.1 Further ideas

In a subsequent project we subjected the data to further analysis, smoothing and careful cleaning, however using a courser geographic division of 81 workforce uptake areas. We were also able to make use of the daily log by marking sporadic and common trips, effectively doubling the number of data points. Using the daily logs, a few common routes that were just under the threshold distance for the long range data was found to have been underestimated previously. We also included such data as travelling with family, age distributions, duration and purpose of travel.

The basic assumption behind marking sporadic trips is that we consider a route between two regions that shows up once in our data as an equally likely route of equal distance that does not show up. From this assumption we model sporadic trips occurring randomly over the whole country based on distance. Finally sporadic and common trips are weighed together.

The consequence of including sporadic travel is that infection may pop up in unexpected places, creating local epidemics which ultimately contributes to the nation wide epidemic. This may contribute to the final size and also serve to lengthen the epidemic. It may also provide data for vaccination planning.

We used this travel model in an agent-based flu simulation model, Mikrosim.[19] During an influenza pandemic we may expect both visits to relatives and work related visits to decrease.

8.6 RESULTS

It turns out travel restriction does have an effect on delaying an outbreak, according to our study. We showed a drop in the number of people so far infected after 60 days simulation in all scenarios, the *cumulative incidence*; this even when we simulated a reduction in compliance level. With compliance as low as 70 % there was still an appreciable reduction in incidence.

The main mechanism behind this can be seen clearly in the plots. They show that the epidemic, which starts off in Stockholm, does not manage to spread very far geographically with imposed travel restrictions. This has local consequences also, as blossoming epidemics in other municipalities will normally spread back to Stockholm via traveling, thereby sustaining and strengthening a progressing epidemic. This reciprocal maintaining effect is curbed with travel restrictions. In numbers, the 18 000 that are infected in Stockholm on the 60th day without restrictions, are reduced by a third with 50 km-restrictions and by another third with 20 km-restrictions.

In the paper the term simulation intervals (SI) is used. These ranges can be thought of as 95 % confidence intervals but strictly speaking they are not. Real confidence

intervals based on a sample population by definition reflect the confidence in an estimate in relation to the "true" value, the most likely event in the real world. The notion of a true value in an epidemic model is rather misplaced. What the simulation intervals really hope to capture is the "true" mean value of an infinite number of runs. They answer the question if the results of two scenarios are to be interpreted as genuinely different or not. For practical purposes, one can consider the interval to be a measure of the variance. In Paper II the reviewers requested that the simulation intervals be replaced with a max-min range, in our opinion a poor alternative, since these values are so random that they may deserve confidence intervals of their own. In the end we complied.

8.7 A COMMUNICATION ISSUE

Paper I is published in a medical journal, directed to the public health community and written in such a way that any medical professional is able to assess the contents. Indeed it was first submitted to a daughter journal of the same publisher specialized in public health issues. At the decision maker level the intended message may be vague. The suggestion that restricting travel is a viable option is put forward with as much political correctness as we could muster. What is passed on is additional knowledge about the mechanisms of disease threatening our society.

There is a story behind the policy implications of the study. The idea and results so far produced was presented by my supervisor and co-author at a meeting at a government agency in Sweden. The presentation was reproached by a very severe remark from one of the assembled, in essence, "It is unconstitutional to impede the movement of our citizens, hence such action need not be researched", implying that our research was not only irrelevant, but unjustified.

This story vividly illustrates that the gap between the research and the executive community is deeper than just breadth and communication. It's a whole different mind set. Perhaps this is as it should be, but the story underlines that all effort to mediate effective communication is to be fervently applauded. In this particular case, a compromise would have been most inadvisable.

The gist of the suggestion was that besides legal and ethical considerations of the research itself, only research of which the *implications* of the results is not expected to trespass beyond the confines of the law, should be permitted. The absurdity of this principal is of course only matched in the actual realization where it has been allowed to enter the code of law it references, constituting a delightful idiosyncrasy. I'm referring of course to the "prohibition of thought", which effectively prevented research into nuclear fission in Sweden for a quarter of a century. The law was lifted in parliament only four years ago. I shudder when I think about the insult to the science community and the harm done to Sweden as a whole.

The direct consequence of the comment for Paper I was the inclusion of the compliance scenarios, in itself softening the potential interpretation and impact of the results. It could be pictured that a government would *recommend* citizens to refrain from moving more than a specified distance. The population would of course not be expected to follow the recommendation to the full extent. Even so the delay in spread would be sustained, despite compliance rates as low as 70%.

Academia can learn something from the converse of this story. There is sometimes a zeal to get involved in the practicalities and the details occupying the mind of decision

makers, at least considering them, even though they will be subject to simplification as any other part of modeling. The simplicicist is of course quick to prune any assumption deemed unnecessary. The complexicists relishes in the opportunity to include yet another variable, introducing a multitude of provisions and subcases. In all modesty I propose my Paper III as a good realization of this balance, a very simple model with pedagogical benefits, at same time dealing with the important applications of the decision making level.

9 PAPER II -- THE DISTRIBUTION OF INFECTIOUSNESS AND LATENCY TIMES AND THEIR USE IN EPIDEMIC MODELS

In Paper II we turn our attention to a flaw in the model described in the first paper. Depending on the purpose of the model, the flaw would be serious and invalidate our results. In the context of Paper I, this is one of those acceptable simplifications that modelers make, above all because the results were comparative, see Section 6.4. We were trying to demonstrate the effect of travel restrictions in a hypothetical disease scenario. In many cases, however, this simplification is not acceptable.
The flaw has to do with our assumption of *exponentially distributed* latency and infectiousness time periods. The time an infected individual spends in the latent or infectious state was not fixed in out model, but picked from a distribution, with a mean, in our reference scenario, about 4.8 days for latency and 5.3 days for infectiousness.

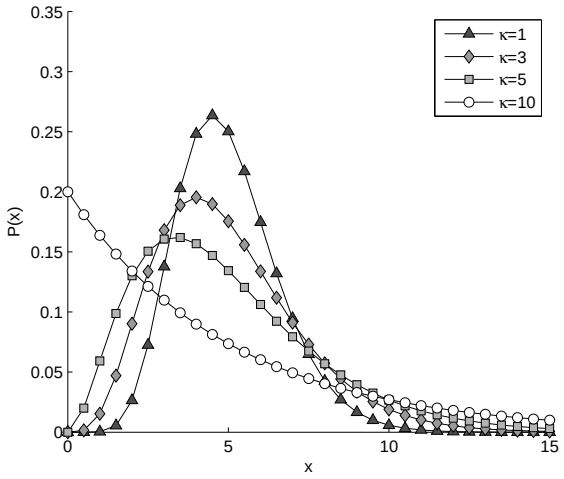

Figure 9-1 The Gamma distribution
Plots of the probability density of a few example Gamma distributions, of which the Exponential distribution (circles) is a special case.

A probability distribution, or just distribution, can be seen as a bucket from which to pick random values. In this case the values represent a stretch of time - time as in incubation time and as in time that a person is infectious. Usually when one thinks of random numbers, one thinks of a uniform probability distribution where a certain number in a certain range is as likely to be picked as any other. Without any deliberate design, the SIR-model is inherently based on a distribution that is not uniform but one where low values are favored to high values. The distribution is the Exponential distribution, shown with empty circles in Figure 9-1.
Depending on the modeling issue at hand, this may be perfectly in order, since the mean time will always be respected. For calculating the total number of infected people at the end of an outbreak, for instance, the simplification works.

If the distributions for latency and infectious time periods were, say, Gamma distributions instead of Exponential distributions* - the choice is not coincidental and the reason will established in a moment - with a form perhaps as the curve with dark triangles in Figure 9-1, then nobody would raise an eyebrow. Values chosen from this distribution will be around 4.8 days on average with a reasonable variance of both higher and lower times. We do not expect latencies much shorter than or much larger than this, a few days at the most.

Exponentially distributed time periods is not a reasonable assumption in any case where the variance cannot be disregarded, where the results are absolutely dependent on these times. One is entitled to ask, what property of an epidemic is *not* dependant on time? Theoretically, at least, there will be no consequences for the final size - traditionally of particular interest among mathematicians - and very little for comparative results like vaccination policies or travel restrictions, regardless of the model.[155]

But there will be a problem with results that concerns speed, time and anything that is not a special case of the equations – e.g. when the peak will be reached or when the last person will be infected. Lloyd, [156] for instance, demonstrated that the dynamics and disease persistence was affected when using realistic sojourn times. Additionally, the problem is larger for stochastic models for which the outcomes are highly dependent on the initial random phase of the outbreak.[64, 157-159] The first few infected will have infectiousness times and latency times that do not average as designed until they have reached a certain number. In effect it's like changing the parameters randomly for each round of simulation and this is after all the root and reason for stochastic epidemic modeling. After the influential initial phase the times periods will average out and the epidemic proceed in an almost deterministic fashion and the distribution of time periods less important.

Even though the ultimate results in Paper I are comparative and do not heavily depend on the variance, why not just avoid the issue, just to be on the safe side? As mentioned, the choice of Exponential distributions was not deliberate, not really a choice at all. They are an inherent consequence of the particular design, explained in the next section. We decided to write Paper II because it turns out that this is something that is commonly overlooked despite being well known. Paper II shows the consequences for a complex spatial model intractable by regular stochastic methods, made feasible in the PC era.

9.1 THE EXPONENTIAL DISTRIBUTION

The curve with empty circles in Figure 9-1 shows a plot of the probability density function of the Exponential distribution as a special case of the Gamma distribution, the circumstances of this relationship to be explained later. The Exponential distribution has a single parameter equal to the *expectation value* (the mean). It is highly skewed, with high densities for short times and a long tail.

Since the median is lower than the expectation value, most time periods sampled from an Exponential distribution will be shorter than the mean. With an expectation value of

* This is inconsistent language on my part, since the Exponential distribution *is* a Gamma distribution. I hope what I mean to say is clear, when I put the two in opposition: Gamma distributions means the set of all Gamma distributions excluding all Exponential distributions, that is to say, all Gamma distributions with form parameter $\kappa>1$.

the latency time of 5 days, 18 % of the sampled time periods will be shorter than 1 day. Such short latency times can safely be considered unrealistic.
There are many reasons for the widespread use of Exponential distributions. Above all it is a simplification issue. The Exponential distribution is one of the most fundamental, occurring frequently in nature and the universe and is easily manipulated mathematically. For the computer modeler, exponentially distributed values can be generated from uniform random values transformed with a simple formula, whereas other distributions often require an iterative algorithm that is harder to program and comes with increased run time.
The mathematical modeler appreciates the easy manipulation of the Exponential distribution which, depending on the model, makes certain parameters of interest drop out in closed form. Using other distributions, the same parameters might be intractable and require a computer to crack. As mentioned, mathematicians generally use pen and paper as far as possible at the expense of sometimes oversimplifying the model. To their credit they are mostly aware of the consequences of their simplifications and the applicability of their results. Many computer modelers are not.
Final size for instance, is a parameter that *can* be trusted despite this simplification, again depending on the model. Our simulations show that if we wanted to accurately predict the incidence at a certain point in time then we would overestimate the value due to this simplification.
The Exponential distribution is inherently "memory-less." This means that predictions of the future state of the epidemic in terms of number of latent and infectious individuals, etc., is based solely on the current state and not on any prior history. Thus, a model with exponentially distributed time periods makes it possible to base the analysis of the model on *Markovian*[160] properties, which will greatly simplify both simulations and analytic derivations.
Turn your attention to the classic SIR model. People move from one compartment to the next, driven by a rate. It's entirely natural to think in terms of rate. Assume X persons per day move from infectiousness to recovered. If the number of infectious is doubled, so is the rate X. On average everybody is infectious Y days, regardless of the number of infectious. The proof that this assumption leads to exponentially distributed times can be found in any text book on queue theory and/or Markov chains, e.g. Norris.[160]
The fallacy of the assumption can be seen when you realize that this is equivalent to saying that at every point in time there is a constant chance of recovering. No, there isn't! The probability depends on when you are infected, of course. If you were just infected, there is a zero chance of me recovering straight away. If you think about it, the fallacy originates from the homogeneous mixing principle, or at least, the line of thought associated with viewing each and every one as an indistinguishable particle. If they were, it wouldn't matter which individual you picked out to recover as we would have no way of telling when that individual was infected.
And in case you were wondering, yes, if elementary particles were distinguishable, the universe would collapse.

9.2 IMPLEMENTING THE GAMMA DISTRIBUTION

A few authors have proposed that the Gamma distribution be used instead.[64, 161, 162] Using this or another less skewed distribution will prevent the predominance of short

time periods characterizing the Exponential distribution. The effect will be fewer people in the infectious and latent stages in the initial phase of the epidemic, in turn affecting results where the initial phase is of consequence.
The Gamma distribution, denoted $\Gamma(\kappa,\theta)$, has two parameters, a *shape* parameter κ and a *scale* parameter θ. For κ=1 the Gamma distribution is in fact identical to the Exponential distribution. Keeping the expectation value constant, with increasing κ, the Gamma distribution becomes increasingly symmetric, or equivalently, less skewed. For a suitable choice of κ the density function can be made to resemble latency/infectious time distributions of empirical studies.
The Gamma distribution can actually be realized with an uncomplicated extension of a SIR-model and it works well with a more complex model, as in Paper I. It can easily be shown that the sum of several exponentially distributed values is in fact Gamma distributed. The trick is the to let each infected pass through not only one latency stage but several, the number of stages depending on the desired shape of the distribution and the appropriate variance and skewness. The parameter κ will namely be the integer you choose. The same applies for infectiousness time, using several infectiousness stages before the unfortunate infected finally recovers. The idea is seen in Figure 9-2.

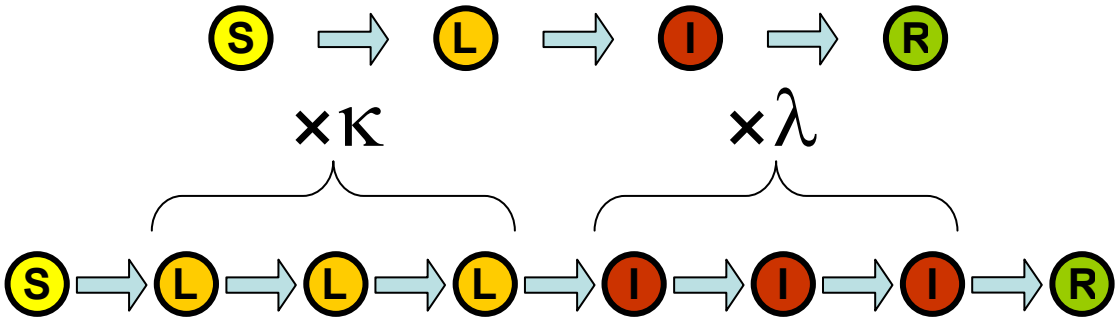

Figure 9-2 Adjusting the distribution shapes by adding model stages
Instead of only latency stage, the simulated people must here pass through three and the same for the infectious stage. Both κ and λ will be three in this case, already an improvement, see Figure 9-1.

The method is not new, it was proposed by Bailey[163] in 1964 and thoroughly analyzed by Anderson and Watson.[64] The added stages are is artificial construct. They have no epidemiological significance and serve only to change the appearance of the time distribution. We can achieve an arbitrarily symmetric time distribution with a minimal alteration to our SLIR-model.
What's going on behind the scenes is that the individuals (or particles) in the model are made just a little bit more distinguishable. You can get just a tad more information about how long someone's been infectious by the stage they're in, and presto, the resulting time distribution loses its memory magic.

9.3 RESULTS

The setup was almost the same as in Paper I, simulating for 60 days, but we limited our self to the base value of parameter γ, and arbitrarily rounding of some of the parameters, just to idealize the situation. First off we used only one stage for both latency and infectiousness, thereby in principle reproducing the simulations from Paper I. Then we increased the number of latency and infectious stages in turn to 3. Finally we set them both to three.

For comparison with previous theoretical results, we made experiments with only one municipality and no travel. We compared the number of extinction runs, in effect the probability of extinction, with Anderson and Watson[64] and found them matching.
By comparing the outcome for different κ for both latency times and infectious times in the full scale runs, we can show that ignoring the shape of the time distribution devalues the results. Increasing the number of stages of latency (more realistic) has a strong negative effect on the number of infected and the geographical reach produced after 60 days indicating that the epidemic is delayed. Increasing the number of stages in the infectiousness stage has the opposite effect, increasing the incidence.
Comparing the runs from the one-municipality run to the full run indicates that the added complexity emphasizes the expected effect.
To see why the realistic distributions have this effect, remember that events early on in the epidemic to a very disproportionate degree determine the outcome of the epidemic. Consider first latency. On *average* everybody is latent five days, but they only support the background trend. It is the ones that are latent 3 days, or even 1 day that make a difference because they are the ones that quickly become infectious and are able to make their mark early on in the epidemic. As mentioned earlier, the 1-day-latent people make up an unrealistic fraction of 18 % of the total number falling ill, if the time periods are exponentially distributed. If they are Gamma distributed, on the other hand, with a form parameter $\kappa = 3$, 1-day-latent people are much more uncommon, amounting to only about 2 %.
It will be as if you boosted the epidemic with hundreds of ill people from day one instead of just the single one in Stockholm. Once the epidemic is underway the distribution will not make much of a difference since, after all, the mean is the mean and provides the expected. But the damage is already done. To compensate for the short latency people, a significant amount of people will have unrealistically long latency times, but these will only become infectious once the short latency people have already brought the epidemic from the unruly initial phase into the more smooth and well-defined deterministic phase. Many of them stay latent right till the end of the simulation.

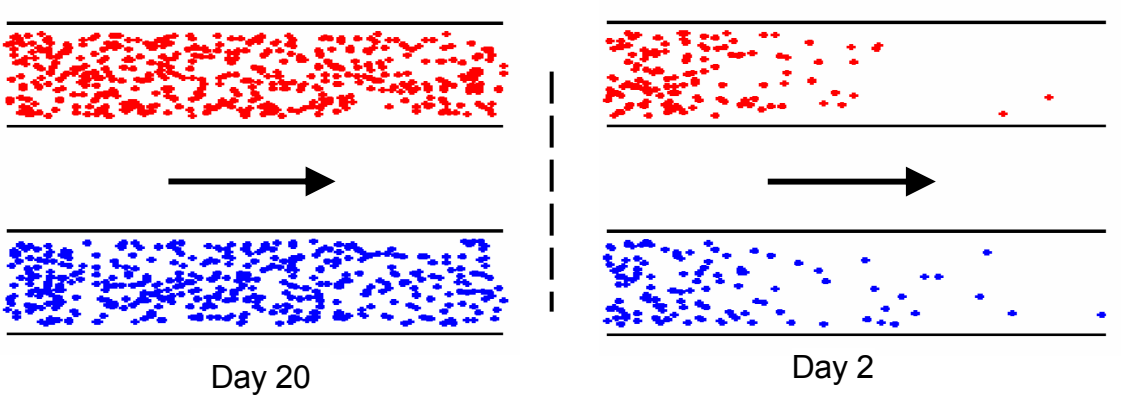

Figure 9-3 A "tunnel perspective" of exponential and Gamma-distributed time periods
Consider the latency stage as tunnels, Gamma- (realistic) tunnel (top), exponential (unrealistic) tunnel (below). An equal number of people both tunnels every day. How long each person spends in the tunnel corresponds to his/her latency times. This is 5 days on average. Some are fast and some are slow, but they are all smudged out in the tunnel, as is shown left. But this is 20 days into the simulation. Look what's happing on day 2 (right) when the first hundred or so have entered the tunnel. While the exponential people are more spread out in the tunnel, the Gamma people are still crowding the entrance. Exponential people with unrealistically short latency times are starting to exit the tunnel and they will ignite the epidemic to come, more forcibly and a lot earlier than in the Gamma people.

Another model for explanation is provided in Figure 9-3.
The opposite effect for the infectiousness times can be grasped in the same way. In the early stages of the epidemic, with exponentially distributed time periods, an unrealistic fraction of 18 % will attempt infection of their peers for only one day. In contrast, the people who have more ample time to cause havoc in the early phase of the epidemic when it counts the most, are more numerous if the time periods are Gamma distributed. The conclusion is, the increased complexity of the meta-population model and the geographical and travel components do not topple the paradigm previous authors have shown. The lowered predictability offered by computer happily does nothing to change matters. The conclusion is: old rules still apply.

10 PAPER III – STATFLU – A DECISION MAKING TOOL FOR PANDEMIC INFLUENZA PREPAREDNESS

As notes Becker,[164] very few published works reference empirical observations. Epidemic modeling is a playground for people whose interests are directed at finding results about the models themselves. But as long as no data is available, the model cannot be validated and what interest is left is purely theoretical. Hard-line mathematicians usually have no objection to this, seeing applied mathematics as a branch of engineering, not mathematics. Too many chances are missed to bridge the gap to policy makers. The work in Paper III, from start to finish, has this in mind.

10.1 COMPLEXICISTS VERSUS SIMPLICICISTS

Mathematicians and physicists often prefer simple, parsimonious models due to the tractability of certain results. Often the model is adapted or a special case is invoked so that a particularly simple expression falls out. More complex models evolve slowly under these constraints.

The accessibility of computers means that the model can be adapted to empirical data almost arbitrarily without attention to the feasibility of analytical solutions. Paper II describes one pot hole of this development. The discussion over complex versus simple, now more than ever, is about validity and applicability. Whether empirical data exists or not may be a secondary issue to some, but at least now there is more opportunity to adapt the model to better fit to data, for validity, both external and internal.

This is not a discussion taken in literature as far as I've seen, but there seems to be a consensus that there is a risk of things getting out of hand. Models overflow with assumptions on behavior. There may be too much reliance in models, and since validation is so difficult, for reasons already discussed, models are often left untested. Given that you are a responsible modeler with a keen eye towards validation, where to place yourself on this balance is as much a choice of personal preference as anything else. It may be satisfying to opt for as much simplicity as possible, and certainly, for purposes of understanding and explanation, simple mechanisms are what scientists seek to find. Even so, special cases abound and you may be lured to a complex model.

10.1.1 Dynamic versus static

StatFlu was designed with these considerations in mind. It is very satisfying to distance yourself from *dynamic* modeling altogether. The term dynamic modeling may be my own, invented to contrast them from *static* models, also possibly my own term.[165] I am not implying *system dynamics* as in Balci[166], as this term is associated with a very specific design of models in computer science. In a dynamic model, whether deterministic or stochastic, there is a path to be trod from initial state to the final result. In a static model, the results are already available; we just have to project them onto the desired frame. The static model is data centered and not model centered.

Exemplifying by dice roll, a dynamic model would consist of a series of experiments to find how often a 1 is expected to show up when rolling an 8-sided dice. A static model would apply what we know about 6-sided die and extend it to make predictions about 8-sided die.

The models in Paper I and II are *dynamic* in the sense that changes in the circumstances of the spread may change the outcome in a non-linear way. In particular, adding one extra initially infected person may increase the size and speed of the outbreak drastically.
Static models, in contrast, merely convert data, rather like a fiscal report converts and summarizes data from the invoice system. The benefit is decreased complexity. The model proposed takes historical data on past flu pandemics and epidemics and transfer them to settings of today. The possibility of doing so is associated with a great deal of uncertainty and this is intrinsically inserted into the model.
In StatFlu and other like models, historic data about influenza epidemics and pandemics is used and transferred to current settings. Most of the guessing work - assumptions about the proportion of infected in the population, i.e. the attack rate, and so on - is left to the user in StatFlu. This will allow for experimentation and a deeper understanding of the issues and parameters involved. StatFlu will display, as a function over time, the number of currently occupied hospital beds as well as the number of excess visits to the primary care system.
Depending on the case and audience, dynamic models are sometimes more convincing than static, sometimes the other way around. This for the same reason that a simple model may be more convincing than a complex one and vice versa. Whatever your position in the matter, after working with dynamic models for a time the transparency of static models is extremely refreshing.

10.2 COMMUNICATION ISSUES

A large portion of research is funded by government. In Sweden of 2009 this percentage was 26 %.[167] There is no need for a justification of why researchers and politicians should communicate. It is simply a question of society getting its money's worth.
The more important question is how. The organization Science & Public works to find ways to bridge the gap. In a survey they have carried out they found that 94% of Swedish politicians have a large or very large trust in academic researchers.[168] Of the articles with scientific content (16 %) in the party associated periodicals, only 18 % cited scientific articles. Only 20 % of the Danish MPs read journal articles in the natural science field.[169]
At the same time it must be said that it is not productive to put forward any kind of suggestion without discrimination. Researchers must retain some sort of credibility, even though they are allowed a substantial freedom in their pursuits. The line should be drawn somewhere on the near side of suggesting that slaughtering all domestic poultry in South-East Asia is a beneficial option for controlling of the bird-flu situation. Were it only for the infeasibility of this operation I would consider effort spent on this modeling question worthwhile. The answer, however, is quite obvious to start with. Yes it would. What gains could be achieved by expunging only a proportion of the stock is perhaps interesting, but all the more obvious is that whatever the gain, the damage would exceed. For either reason, as researchers we should concern ourselves with questions and answers that either are or are foreseeable to be, in demand by the decision makers and the public.

Mostly, anyway. The Ig® Nobel Prizes[170] are awarded to contradict this statement and to highlight that research for gains that are not immediately apparent, is also immensely important.

StatFlu was designed for and funded by The Swedish Board of Health and Welfare to predict hospital and primary care load in Sweden in the event of a flu pandemic, as part of the pandemic preparedness plan. In other words, the target audience of this work is decision makers in a very direct sense. This kind of model has been produced before.[171-173] Some are freely available as computer applications. A more detailed report has been published on these models, see Camitz: *Översikt över statiska influensamodeller för pandemiberedskap*.[174]

The data centered view point lends transparency to the model. It is immediately clear what data is used, what could be used to validate the model internally and how to validate the model when, heaven forbid, the results come in from a real pandemic. This is of course appealing to decision makers.

Since results are produced by those who use the software and not me, the scientist, careful consideration had to be taken, particularly regarding how the results are interpreted. Ideally, one would like to test the application on a panel of test users, in software development so called *end user testing*. In StatFlu, attention has been put into displaying the uncertainty in the results to the user, resulting in new developments in the graph output. The danger is focusing too much on the mean outcome, in our case, in terms of hospital load. The mean outcome is only interesting for comparing scenarios. The mean outcome of a dice throw is 3.5, but don't place any bets on it. The mean outcome and the most likely event are distinct in general. At the casino, you should prepare yourself of the possibility of getting both ones and sixes with equal probability. The graphs produced by StatFlu, an example seen in Figure 10-1, shows that the extreme values are improbable, so the dangers associated with interpretation and focusing too much on the light parts of the graph is still an issue. For pandemic preparedness it is of course crucial to design for the worst case scenario or at least be prepared for the possibility.

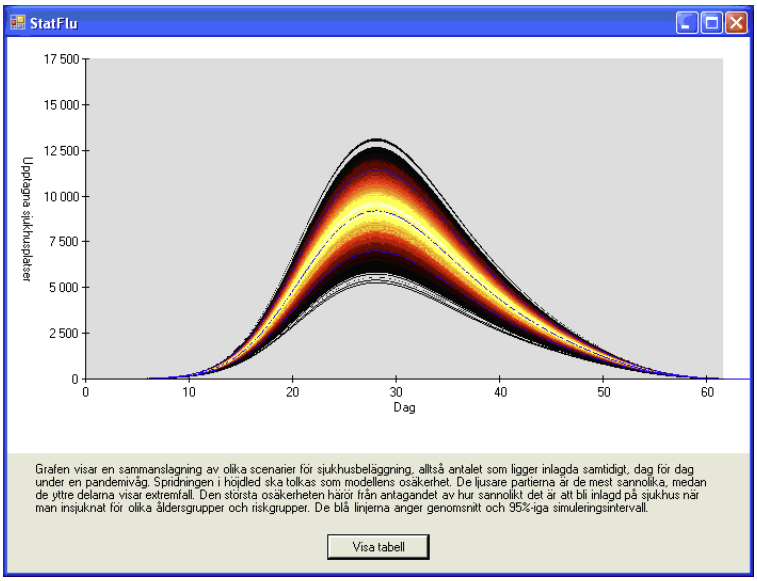

Figure 10-1 Output from statflu for hospital load

The plot is for the whole country using Swedish risk group estimates. We set the attack rate to 25 %, the duration to 90 days and the average duration of stay to 10 days. At its peak 28 days in, the epidemic will likely produce somewhere between 7500 and 10 000 simultaneously occupied hospital beds.

10.3 BACKGROUND

The earliest example of a static model in use in the service of pandemic flu planning as far as I am aware is Meltzer *et al*. [172, 175], a study to evaluate the economic impact of a pandemic flu outbreak in the United States. To this project there has apparently been an update, [176] published after Paper IV. Essential parameters for age group specific attack rates were collected from various studies of outbreaks of seasonal and pandemic flu. What's interesting is that these parameters were not seen as fixed. Rather, each was associated with a range of possible values, the centre value being the most likely. This is to account for the uncertainty inherent in the values and the uncertainty that results from applying them to a different context. When calculating the results, the distributions were sampled thousands of times producing an uncertainty range for the output as well.

This is known as Monte Carlo-sampling,[177, 178] akin to scenario analysis where you would typically set up a number of scenarios for comparison with each other along with some kind of estimate of the probability of the occurrence of each. Instead you set up thousands of scenarios, weighing your choices with the associated probability and analyzing everything in batch.

Meltzer *et al*. calculated mean economic impact based on mean hospital admittance and mortality, each with 90 % confidence bounds. They also compared results with and without the use of vaccination.

A notable and widely used implementation of a static modeling approach available on the web is FluSurge,[179] released from the U.S. CDC, to which Meltzer is associated. FluSurge is used to project the total hospital load over the duration of the pandemic. This software, its forerunner without time projection FluAid,[180] as well as slightly adapted versions thereof, have been used by authors in published articles predicting hospital load in several regions and countries.[181-187]

A similar setup as in Meltzer *et al*. was made in France using by Doyle *et al*.[171], having slightly different background variables and with a focus on hospital admittance and mortality. Many parameters were taken directly from Meltzer *et al*'s study, as also I have done. Also in this study the authors compared scenarios with and without the use of intervention programs, in this case vaccination and anti-viral pharmaceuticals.

Van Genugten *et al*.[165, 188] used detailed national data collected from seasonal influenzas. Their approach slightly different from the former two's use of Monte Carlo simulation, opting instead for scenario analysis i.e. a handful scenarios are presented and analyzed separately.

It should also be noted at this point that static models' contribution to understanding the effect of vaccination and anti-viral pharmaceuticals is questionable. Usually, the effectiveness of the drug is quoted and used simply as a reduction of the final outcome.[171-173] As with any intervention strategy, they have the potential of completely knocking out the disease as well as failing miserably. This is in part due to chance, but more specifically, a prevented due to the targeted mechanism: a vaccinated case will not spread the disease further. Therefore the effect of one dose in the transmission chain is potentially much greater than one saved case. This is discussed more in depth in Chapter 3, Infectious disease epidemiology.

Not taking into account the full dynamics in disease spread when considering pharmacotherapeutical effects is not permissible in my view, so static models are not well equipped to deal with these issues. Both Meltzer *et al*. and Doyle *et al*. provide

estimates of incidence reduction following either vaccination or anti-viral drugs, though Wallinga *et al*. are safe from incrimination by combining the static model of van Genugten *et al*. with a dynamic simulation built on top of the original.[188, 189]
The StatFlu project was initiated to build on these previous examples. I developed a stand alone application that can be used as a prediction tool and a pedagogical tool for decision makers at all levels. The system is now in use in Sweden but is readily adaptable to other regions or countries. It's free to use and available from the project homepage.[190]
The static part is the model's use of historic data and applying it to the selected country's demographics. The data describes simply how likely it is to be hospitalized given infection with flu. Then we can estimate, given a certain number of infected, how many hospital beds will simultaneously be occupied. Complications arise when stratifying for age-groups and persons who are at higher risk of hospitalizations due to chronic illnesses, for example. Most of the data used is the same as in Meltzer *et al*.[172], mostly originating from studies of the Hong Kong flu in Oregon. Some data from Working Group on Influenza Pandemic Preparedness and
Emergency Response has not been published and was used taken directly from Meltzer *et al*. Where possible Swedish data was used, particularly for the estimation of primary care load, which is very specific to the target country.

10.4 TIME

StatFlu estimated the number of simultaneously occupied hospital beds. That is the variable of interest. The total number of occupied beds is emphatically not equal to the number of afflicted patients seeking care which cannot directly be used to dimension care capacity. A longer average stay means the turnaround is lower, also making the load heavier. In the extreme case where the required hospitalization time is very short, then the number of occupied beds at every point in time will approach the distribution of admissions.
Either way, we have established that time is a factor. It should and does have a role to play in the StatFlu-model. This is one of the major enhancements of this model compared to others. Ignoring the element of time is bordering on deceptive.[171, 172]
Those requiring hospital care during an outbreak will not storm the hospital all at once. There will be a distribution of admissions over time and the heaviest burden will be at the peek of the epidemic. At every point in time the number depends on two additional variables, the duration of the outbreak and the average length of hospital treatment. A short outbreak will concentrate the load, making the peak higher.
Reasonable adjustments were made to add a time dimension to some of the models described earlier. Bonmarin *et al*.[191] published a follow-up calculation to the French study, assuming the shape of the time function would be similar to that of previous seasonal influenza outbreaks as gathered from French sentinel data. Van Genugten *et al*. included a time plot in the original study where the estimated attack rate was distributed over a Gaussian (normal) curve. The software FluSurge also plots the output on a time axis though it is not clear what the rational behind their choice of algorithm is.
Though perhaps less than ideal, Van Genugten *et al*.'s selection of the Normal distribution is the direct reason for my choice of the same distribution.

Table 10-1 Parameters in StatFlu and their implementation

Variable	Description	Source	Uncertainty	Implementation/ treatment
Gross attack rate	Fraction of population infected	User-specified	Hypothetical	User-specified 5-50%
Duration of epidemic	From first infected to last	User-specified	Hypothetical	User-specified 10-150 days
Population	Population in age groups 0-19, 20-64, >64, by region	Population register[22]	High certainty	Fixed, specified in text file
Duration of hospital visit	Average length of treatment at hospital	User-specified	Attainable, partly hypothetical	User-specified for all ages 1-14 days
Age group-dependent relative risk of infection		User-specified	Hypothetical	Specified for age group relative to the other age groups
Size of risk group	Fraction of age group at elevated risk for complications	Provided by [192]	Definition-dependent, attainable in theory	ca. 2% for the whole population; specified in text file
Risk of hospitalization	Risk per age and risk group of hospitalization given infection	Provided by [172,193-195] and expert opinion, see Table 3 in [172]	Uncertain, dependent on risk group definition	Sampled from beta-distribution; hard-coded
For primary care load only				
Primary care visits	Yearly primary care visits per region under normal circumstances	Provided by [196]	High certainty	Fixed, editable in text file
Hospitalizations associated with influenza-like illness	Hospital patients coded with influenza	Provided by [192]	Highly uncertain, coding-dependent	Fixed, editable in text file
Risk of primary care visit	Risk per age and risk group of hospitalization given infection	Provided by [193,197,23] and expert opinion, see Table 3 in [172]	Uncertain, dependent on risk group definition	Sampled from beta-distribution, hard-coded
Fear factor	Deterrence from primary care due to pandemic	User-specified	Hypothetical	User-specified 0-40%

10.5 DATA CONSIDERATIONS

The use of data in StatFlu has been carefully balanced, objectively but arbitrarily, by me. Perhaps less uncertainty and more intuitiveness and certainly more objectivity could be gained by simply allowing experts to guess the settings, rather than going to exaggerated lengths to use real data.
When designing StatFlu the approach was to use as much real data as was available. The following priority order was used:

1. Swedish empirical data
2. Empirical data
3. Data provided by expert panel
4. User input/educated guess

Table 10-1 accounts for all the parameters StatFlu, their origin and how they are treated and applied in the model. There is of course no consensus on how these parameters should be treated and which source to be used. Which parameters are varied with sensitivity analysis and scenario analysis differed between Doyle *et al*. and Meltzer *et al*.
Many times the data used in Meltzer *et al*. is based on expert opinion. This is standard practice in *pharmacoeconomics*[198] a field no more blessed by an abundance of real data than any other. There are many ways of employing the educated guesses of parameters from a panel. StatFlu is designed with some of those methods in mind, in particular Bayesian methods described in Dittus, Roberts *et al*.[199], Clark[200]; and Cooper, Sutton *et al.*[201] In particular, these considerations affected the distributions used internally in the model.
Needless to say, research using expert opinion is sensitive to the choice of participants and particularly vulnerable when real data becomes available to verify model input.[202] Bayesian procedures are designed particularly to approach this problem. On this note, the use of your own parameters in StatFlu exploits the opinion of the expert in you, if you consider yourself such. Indeed, as mentioned, studies have been published based on FluSurge and the input of experts.
For the case of StatFlu and other static models, the contrary point might be argued. Rather than using the Oregon data for hospitalization risk, given the presumably large discrepancies in both time and setting, mightn't it be better to use an expert panel all the way, filtering the Oregon data by common sense and experience? This objection is equally valid directed at Meltzer *et al*'s original use of the same data. The U.S. is not Oregon and the 90s are not the 60s, see Chapter 8 on traveling. Equally, given our very low estimate for risk group size, would not anyone's opinion be better? A weighed average of cited values used in other countries may be an alternative. The StatFlu response to these thoughts is to include these parameters as user settings in the application, as well.
The choice between the third and fourth point in the numbered list above, was made, as has already been observed, by evaluating the "type" of uncertainty. Attack rate is uncertain, essentially a guess. The number of models that provide this number on the output side has not changed this fact and in StatFlu this value is provided by the user. Perhaps more of the uncertain parameter values should be user specified, or at the very least, the source of them should be easily shifted. Risk group size, for instance. As is described in the discussion of Paper III, the risk group size was determined from the

Swedish Discharge Register[203] but this source turned out a comparatively low value. The analysis and the very definition of risk groups is subjective, as is seen by comparing the aforementioned paper. Ultimately the risk should pertain to having to seek hospital care when sick. Also, there is the ICD-code[204] problem, the *international classification of diseases and symptoms* using which the source data is coded. There are misclassifications and physicians may use different codes for the same disease. Influenza especially is hard to classify using the ICD-code book.
Fortunately, the risk group size can be changed by modifying the contents of a text file and this is the intended m.o. if StatFlu is used in other countries besides Sweden. Recompilation of the software is not needed, but it still might not be an on-the-fly operation.
Another suggestion brought forward by a reviewer was the possibility of having a "negative" fear factor. There might of course be an increase load to primary care facilities if medication is distributed there. This will be included in future versions of the model.
There is also the matter of hospital length stay. Although theoretically attainable, I can find no data to support it. In part, this is dependent on the severity of the disease. Mortality should also be factored in. High mortality due to a severe infection would perhaps shorten the average treatment time, all things equal. This is not explicitly handled in the model.
Of importance is also what kind of care and resources is required and this has been left out of the model, but it can be specified indirectly though the length of the hospital visit. This is of more importance in estimating the economic cost and should after all be a simple algebraic operation.

10.6 AN ENGINEERING PAPER

Paper III may be seen as engineering paper, describing, as it does, a software tool, even though it was submitted for reading by the European infectious disease epidemiological community in Euro Surveillance. This is a journal published by the European Centre for Disease Prevention and Control (ECDC). Where readers may be accustomed to finding results in an article, instead they will find a comparison with StatFlu and FluSurge over which StatFlu claims superiority. Essentially, the result of the project is the software itself, along with a review of the research in the field. The review in brief form can be read in the paper and a more in-depth version can be found in the aforementioned report.[174]
It is worthwhile in the context of a thesis to include the software requirements specification, assembled with the participation of the product users and the design work of the application, which is based on the review. [174] A framework for systematically defining the requirements is the *technique trade-off* method.[205] This method has previously been used to evaluate a dynamic transmission model.[206] Usually the end-user and/or an expert panel is consulted to evaluate the various design issues. Due to resource constraints the panel in this case was limited to one person which is less than optimal. A summary follows. Perhaps a full end-user evaluation, *acceptance testing*, of the software will be performed in the near future.
It was decided early on a static model approach with the already mentioned benefits, so this consideration was never included in the specification process and only static models were included in the review. The trade-offs have also been mentioned. The

static model approach disregards the dynamics of the epidemic and consequently cannot evaluate the effect of therapeutic pharmaceuticals and other interventions. Such interventions must be implicit in the attack rate assumption.

Functional requirements

Input parameter specification

The input parameters vary in uncertainty, due both to origin and applicability. If possible, real data should be used. If unavailable, a distribution reflecting the views of an expert panel may be used and considerations in the model for this approach is to be researched. For purely hypothetical parameters, the user should determine the values in order to fully appreciate the uncertainty.
The trade-off is partly between usability and control. Advanced users may wish for control over certain inputs e.g. risk group size, while others may be intimidated by a cluttered workspace. This could be addressed via an advanced settings dialog, but the cost may be high. Advanced inputs should be modifiable at least by altering the contents of a text file. An advanced user can then set values that are acceptable to colleagues throughout his/her department.

Application interface

The interface must be intuitive and self-descriptive with a “wizard” style work flow. The graphic output should be fast and provide the desired output in graph form clearly visualizing the uncertainty.

A use-case (adapted)

The primary actor is a regional government employee wishing to get a grip on the hospital load under pandemic circumstances. The user starts up the application, chooses his region and clicks *Next*. She enters an *attack rate* of *25 %,* a *duration* of 90 days, 5 days as the *mean length of a hospital stay* and clicks *Next*. She then decides on the relative risk between age groups. She chooses from a list of presets, *Meltzer A*, with which she is familiar, but modifies it slightly, believing that the elder population group will have only half the risk of infection as the adult group. She clicks next and sets the fear factor to *10 %* before displaying the graph output for hospital occupancy and primary care visits. The graph is displayed in a new window. Without closing it she returns to the main window and adjusts the length of stay to 10 days, then opens a new output window. She compares both outputs side by side. She clicks *Tabular format* and is provided with more information in figures. The user exits the application.

Mathematical model

The mathematical model should be sufficiently versatile to handle total hospital admittance, duration, hospital visit length in a closed form time dependent equation. Age differences, differences in risk groups and other parameters should lead to a grouping of intermediate results. Preferably, all equations should be analytically solvable, if not, numerically with computing time on the seconds-scale.
The trade-off of a complex model is transparency. At a minimum it must be possible to describe the way input variables are treated by the model so that users are not surprised by the results of their actions. The model should be deterministic. The sensitivity analysis should be the only source of stochasticity.

Numerical model

Numerical calculations must be optimized to allow for near real time usage. Caching or pre-calculation of intermediate values and distributions should be employed.

Uncertainty model

The sensitivity analysis should be based on 1 000 to 10 000 input scenarios with parameters varied over the parameter distribution, resulting in a distribution of hospital admittances. Final results may be binned before the solution of the mathematical model. The frequency of each bin should be reflected in color of the epidemic curve output. The distribution suggested is a beta-distribution to exploit subjective opinions.

There is no trade-off using the Beta distribution as it can be manipulated to resemble arbitrary distributions.
In conjunction with the Monte Carlo based sensitivity analysis, a scenario analysis is provided for those parameters that the user specifies. Scenarios should be compared with ease, side-by-side in the workspace.
There is a risk of the user interpreting the mean outcome as the result without considering the full range of outcomes, particularly the extremes, despite the model being specifically designed to target this problem.
A trade-off of showing the full range and a probability distribution is that not enough emphasis is given to the extremes. At the same time the extremes are more dependent on the number of simulation which is unreasonable.

Age group risk

Our users may be familiar to the work of previous authors and/or accustomed to FluSurge. The risk of being infected depending on age groups should be available to the users as predefined presets alongside the ability to set their own arbitrary values. These should be in a population independent form for transferability to other countries and regions. This means that the risks between age groups can only be shown/specified as relative to each other.
A trade-off is to ease of use. For the user it is simpler to think in terms of absolute risks.

Non-functional prototyping

All aspects of the internal functionality and validity of the model as well as the graphical output must be thoroughly tested in several prototyping levels. Optimized numerical solutions must be tested for accuracy against robust methods in Matlab under a full range of parameter scenarios.

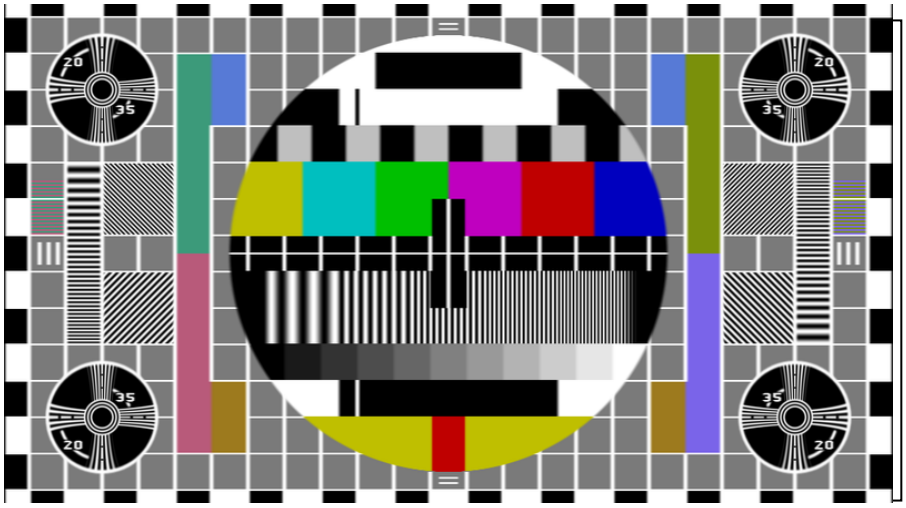

Figure 10-2 Yours truly, demonstrating StatFlu in Swedish news show Aktuellt.
In the upheaval during the Novel Flu in 2009, I was interviewed on TV. I had an idea that combining thesis, dissertation and TV appearance would be a big boost to my ego. Unfortunately, the copyright charge to reproduce a snapshot would have been 800 kr, which I found to be too steep. So instead I placed this test image from Wikipedia, published as Public Domain.
The snapshot intended I recorded and captured myself, very lo-res. The report itself is paid for by taxpayers, licensees and of course my time. Two of these sources incidentally funded the work portrayed. Many individuals, profit and non-profit organizations have contributed to this thesis, free of charge. Paper I is published Open Access meaning that anyone can freely reprint, resell or rework it. I have also used images under the Creative Commons license. But not everybody thinks alike.

11 PAPER IV – THE GIANT SEXUAL COMPONENT

Modeling sexually transmitted infections, STIs, requires special considerations. Perhaps most importantly, homogenous mixing, always the default approach among epidemic modelers, does not work well.[36, 51, 207, 208] Homogeneous mixing models, i.e. the particles-in-a-gas concept detailed in previous chapters, have been unable to explain the slower than expected spread in the beginning of the HIV epidemic or why Chlamydia is endemic. The failure of the homogenous mixing assumption is entirely reasonable, especially if you already harbored doubt about its applicability even when it came to describing measles outbreaks. For one thing, most humans have a preference towards either of the sexes, so equal likelihood of transmission regardless of whom we bump into is conceivable to be a flawed model. To put it short, more than any type of contact between humans, sexual contact is structured.

Having said that, homogeneous mixing was never designed to fully capture human interaction. It is after all only a model, an idealization of human mass-behavior that has proved itself useful. It may not be entirely clear why such models, which hold up to prediction in other cases, do not work when put up to the test of STIs.[5, 36, 209]

By understanding networks we hope to gain understanding of the dynamics of network based epidemic models, and by understanding that we might gain insight into what drives STIs and keep them endemic in the world population today.[3, 114]

Paper IV explores a special entity borrowed from graph theory, the *giant component.*[210] Essentially the work hopes to contribute to understanding STIs by establishing firmly that a majority of the population is connected - sexually. This is something that is intuitive for most people who have a certain grasp of percolation. Thankfully, this is true of many of those who have read Malcolm Gladwell's bestseller The Tipping Point.[63] Network physicists and Gladwell's readers notwithstanding, it is a good thing to try and prove the apparent in order gain support for a message sent to the public health community, and in extension, the general public. The ideal would be to get such an "appealing" message across that it would be spread by word of mouth. What an enticing thought: to exploit the social networks that sustain epidemics for spreading information that counters the very same epidemics.

11.1 WHAT IS A NETWORK?

A network is a structure of entities referred to as *nodes*, aka *vertices*, connected to each other by *links,* aka *edges*. Today's computer savvy people will benefit from embarking from the notion of the internet in order to grasp the general concept of a network. Understand a computer or physical piece of hardware to be a node, and the conduits between them as edges. The structure makes up a network. On top of this physical network a number of abstract layers have been added, the World Wide Web being the most familiar. It is a network of web pages connected by hyperlinks. As a representation of your social self you have your Facebook profile and from it you are connected to your friends, in turn connected to their friends, some of which are connected back to you. There is even a Facebook app that plots the immediate neighborhood as a network on your profile.

The Facebook network is according to all definitions, a social network, spanning the globe as indeed all social networks have been found to do. As a matter of fact, quite a

number of network papers have sprung from the analysis of online communities such as Facebook.[211-213]
The above are just a few of the networks that are found all around us and that have been extensively studied, from metabolic and genetic networks to trade and aviation networks.[214-217] One example of a travel network can be found in Paper I and II.
It is not surprising that network science has received so much attention from scientists of all fields. What captures most scientists' interest is the statistical properties of very large networks of thousands or millions of nodes. The theoretical bulk of knowledge is now quite large, but in many aspects network science is still a young science.

11.2 A FEW NETWORK CONCEPTS

To understand networks we need to have some concept of the emergence and growth of networks. Under what conditions does one node connect to another and how do those conditions influence the topology of the network? The search for a universal law of network structure and emergence has only just started.
Scientists try to answer the question by studying random network models. From 1959 over next three years, Hungarians Erdős and Rényi published a series of papers describing the properties of a simple network generated by assigning a probability, p, whereby a fixed set of nodes would connect to each other.[218-220] If this probability is 100 % , then all of the possible links between the nodes are present. For a network with 10 nodes, $10 \cdot 9/2=45$, edges are possible, an edge from each node to all the others, counted only once in the case of an *undirected* network.
Although perhaps not prominent in the real world, the statistical properties of this type of network have served as a starting point for further exploration. Again the reality is adapted to suit the map, pertaining to our ongoing discussion about simplicity versus complexity, and as always, there is great insight to be gained from starting small. As we shall shortly see, the particular assumption generating the Erdős-Rényi-random network, has a tie to one that has already been discussed.
Strange things happen when p is varied. When p is low, so is the propensity of forming a connected network. Nodes will be isolated. Only small connected components of a few nodes will emerge. As the probability to form links is increased, the components interconnect to form larger components. For large p, almost all the nodes are connected forming a single dominating component, a so-called *giant component*. Only very few islands remain separate from the giant component and these are orders in magnitude smaller. The strange thing is that there is no gradual emergence of this giant component, of which there is only, equally strange, one. p is a threshold parameter, a critical value. When $p > 1/n$, n being the number of nodes, there is a giant component. Below the threshold value, there is not.
The forming of the giant component can be seen in the mathematics and very readily by simple computer simulations. The process is rather like when a fluid freezes and has borrowed the name *percolation*.
Seen from an epidemiological standpoint the existence of a giant component is of course detrimental to our health. Viewing the global community as a random network leads to many conclusions about the spread of diseases – how a high probability to interconnect implies a high chance of an outbreak; how the evolution of travel and aviation has shortened the path length between individual nodes and reshaped the epidemiological landscape.

The attentive reader will recall the discussion of another threshold value, namely R_0 which determined the epidemic potential of a pathogen. The relationship is of course not incidental as the random network is the network equivalent of homogenous mixing. Recall that homogenous mixing implies that everyone has equal chance of contacting everyone else. In network terms, each node has the same probability of connecting to all others. And in the same way that homogeneous mixing is an oversimplification in some cases, so of course is the Erdős-Rényi random graphs. The story is rather more complicated.

11.2.1 Small world

We cannot write about networks and skip the small world effect. We have all heard either of Bacon numbers[221] which gives the shortest collaborative path from any actor to Kevin Bacon –in a relaxed version from everyone in the world; or Milgram's famous experiment in which packages were distributed with the instructions to send them back to the lab via your personal network of social contacts.[222] Initially we are surprised that the shortest social contact path between any two people in the USA is around 6, on average. How could infectious diseases not thrive under such beneficial conditions? A hypothetical disease with 100 % infectiousness would cover the US in only 6 generations.

Watts and Strogatz[223] later made history when they proposed a network growth model that exhibited a small-world effect. Small-world networks exhibit a high degree of *clustering* of nodes i.e. heterogeneity in contacts: you are likely to be connected to your contacts' contacts. This alone does not account for the patterns of spread we expect from epidemics today. More on this in Section 11.4.1.

11.2.2 Scale-free networks

A great number of papers emerged at the turn of the century showing that a number of networks found in nature and society were *fat-tailed* indicating that they own a property of networks known as being *scale-free*.[224, 225] What is indicated by this term is a special property of the *degree distribution*. The degree of a node is simply the number of links

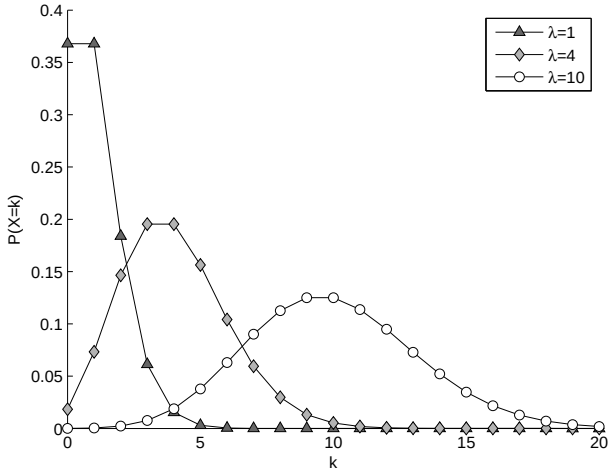

Figure 11-1 The Poisson distribution

The plot shows a few versions of the Poisson distribution with different expectation (read mean) values λ. Random graphs have Poisson distributed degrees meaning that the mean degree is also the most common degree – together with the mean degree less one. Most nodes tend to have degrees around the mean, ensuring an uneventful network. Extremes are rare.

it hosts. The mean degree in a random graph is $z = p(n-1)$, where n is the number of nodes in the network. Distributions are discussed in Chapter 9. The distribution of degrees in a random network is a Poisson distribution which has its peak around the mean value, see Figure 11-1.
What is commonly seen in the real world, however, is a *power-law* degree distribution as in Figure 11-2, essentially describing the case when a few nodes have a comparatively high degree, but the vast majority of nodes have a moderate degree. The mean degree in this case becomes rather uninteresting. The shape of the distribution does not depend of the size of the network, hence scale-free.
The World Wide Web, for instance, appears to be scale free. We know that a few sites are extremely well connected. Wikipedia, as is common these days, will be the link-beneficiary once this thesis is published online. There is a large amount of moderately connected sites from newspapers down to popular blogs, while the vast majority of sites remain poorly connected, my own blog being a good example. The vast number of poorly connected nodes is referred to as the fat tail or the long tail, the latter expression recently kidnapped by Anderson[226] in describing how companies these days must adapt to selling many diverse products, each which few consumers. There is only room for so many iPhones on the market.

11.3 NETWORKS IN EPIDEMIOLOGY

Epidemics simulated on small-world networks spread homogeneously in a wavelike pattern such as is recorded from the bubonic plague and diseases in the animal kingdom, like tularemia[7], and seen on Hollywood blockbusters. These tendencies were seen also with the Spanish flu in 1918-19 although local pockets soon emerged creating new centers of transmission.[47] Traveling disrupts the wave pattern, connecting geographically distant territories, probably making the world more scale-free. The dichotomy between small world and scale free manifests itself in IDE as the bubonic plague contra SARS.
One property of networks of particular interest is the resilience against random

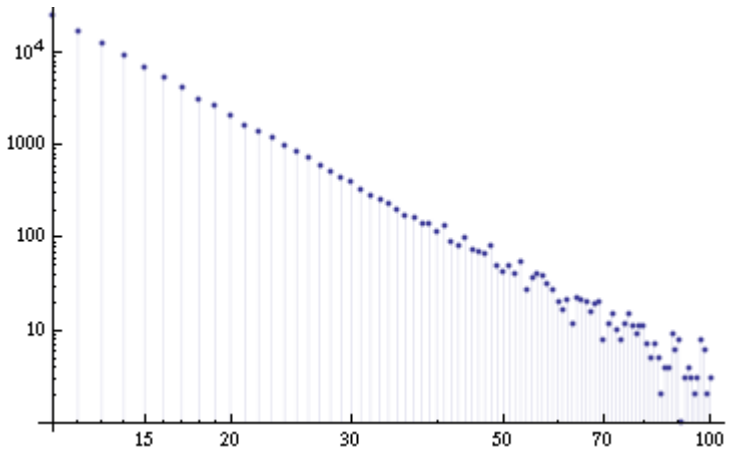

Figure 11-2 Power law distribution plotted on a log-log scale
Note the scale: this is not a linear association. A value of 13 is ten times more likely as 22 and 100 times more likely than 40. Values of over 100 are very uncommon. If the plot depicted the degree distribution of a network we would say of it: typical small world, lower degree always the more common and a few really high degree nodes. Although the mean degree can be calculated, the interpretation is not really well-defined.

attacks.[227, 228] The internet thrives perfectly well even though random hardware equipment constantly goes out of order.[229] The traffic is simply diverted with hardly any loss of efficiency. On the other hand, if routers and servers were attacked intelligently, targeting the hubs of the network, the disruption would be very severe.[230] The internet is very susceptible to targeted attacks due precisely to its scale-free nature. This fact carries over to infectious disease epidemiology in an analogous way. The vaccination of a few highly connected individuals – scientists use the word hubs also here – will benefit in combating, more so than a comparably large vaccination campaign of the random public. Put in the language of R_0, even though the true R_0 may be low, the unfortunate infection of a hub may lead to a fast and severe outbreak. As recalled, R_0 has everything to do with contact degree, though the calculation of R_0 in a network setting is very different than the simple formula using the mean contact rate, given in previous section, applicable only under the assumption of homogenous mixing.

Work with networks and epidemics in particular and resilience in general has been done by others.[3, 113, 114, 231] Liljeros, in an unpublished paper,[232] performed a test on the network of Swedish companies hoping to find that closing down a few major workplaces might lead to a more sparsely connected network not able to sustain epidemic growth. The trend of research, however, seems to be downplaying the role of hubs and super-spreaders[233].

11.3.1 Sexual networks

Sexual networks have also been studied in IDE with pertinence to STIs.[234, 235] There are many questions that remain unanswered about STIs, for instance, how Chlamydia persists in the population with such a stable prevalence. Could the answer lie in as yet undiscovered properties of networks? Largely due to a new strain and recent higher risk behavior, Chlamydia is once again on the rise in Sweden.[236] The fact that Chlamydia Tracheomatis can be subclinical (see Section 3.2.2, page 12) and remain undiscovered by many carriers is one reason.

Common for all STIs, is that they spread very slowly. This is of course due to the low contact rate, on a time scale longer than for diseases like influenza or measles. Contact rate corresponds to node degree in sexual networks and is of course important, but perhaps not in the linear way explained by simple SIR-models.[237] Recalling Section4, Epidemic modeling and its history, readers will know that for an outbreak to be sustained the infectiousness time must be longer to compensate. Still, researchers find it baffling that limiting the spread of STIs has turned out so very difficult, despite effective protection being readily and cheaply available and easy to administer.[90, 161] Analyzing the contact patterns seems to be a necessary and beneficial approach. For example, Morris and Kretzschmar,[88] demonstrated the importance of concurrent relationships when comparing the HIV epidemic in Africa and South East Asia using simulated networks that evolved over a five year period. Pastor-Satorras and Vespignani,[3, 4] showed the negative impact of preferential attachment in sexual contact networks, that is, the tendency for highly connected individuals to seek new partners among other highly connected individuals.

The latter phenomenon will lead to the emergence of so-called *core groups*,[52] long since identified of vital importance by public health professionals. Core groups are groups of high degree nodes connected to each other. If you think about a celebrity

cocktail party, you know where to look to find a core group*. According to the core group theory, an infection released randomly will soon find its way to a core group, where it may remain endemic due to a locally favorable environment – multiple and intense contacts - even though in the population as a whole, R_0 is lower than the threshold value 1.[238, 239]

11.3.2 The giant sexual component

Though the core-group theory is à la mode with public health workers, it is not a powerful message to send to the youths of today, frequently neglecting to protect themselves.[240] The core groups exhibiting high risk behavior, after all, are very few. Most people have a modestly promiscuous lifestyle.

The fact that your partner probably has more contacts than you do may cause some concern however. What we try to show in Paper IV is that everybody, very permanent singles and monogamous couples excluded, are linked together in a giant component of the sexual contact network. Given what we know about the small-world effect, this might be a sufficiently strong motivation for safe sex. An STI can span the network in only a few generations, and you don't want to be part of the chain.

Under what circumstances will a giant component be present in the sexual network? Remember the sexual network is not a random network and there is no *p* to be varied. If anyone were to suggest an acceptable model for the forming and evolution of sexual networks, it is guaranteed to be more complicated than that.

11.4 THE PAPER

What can we know about the real sexual network based on degree distribution? Would it be possible to construct a reasonably accurate network with only information about the degree distribution?

Alas, in Paper IV, we do not offer a network model to withstand the heavier kind of scrutiny. We offer *a* model and we do it to support an argument about the giant component based on node degree alone. This is the main theme of the paper. In random networks a giant component either exists or it does not. Intuitively, this is probably true also of sexual networks and if so, then our argument is strong.

A rather large complication of our simple argument is the *temporal* aspect. Relationships have duration and sequence. So unlike networks most thoroughly studied in network science, perhaps our network should not conform to conventional assumptions. Our network is temporal (or *dynamic*) rather than static, as in Figure 11-3.

Adhering to protocol, I offer here a summary of the methods and results. I do however recommend reading the Paper I instead. It is the most recent paper of the four and not much can be added. Since it hasn't been submitted yet I certainly haven't had any feedback to share. It is written in a, for a thesis, suitably leisurely style which I can only plagiarize here.

11.4.1 Methods and results

As always, we are held back by our data. There is an abundance of enquiries of sexual behavior, habits and experience. This appears to tickle our curiosity to the utmost. Most enquiries include questions about the number of sexual contacts per individual, or

* No intended sexual connotations to cocktail parties.

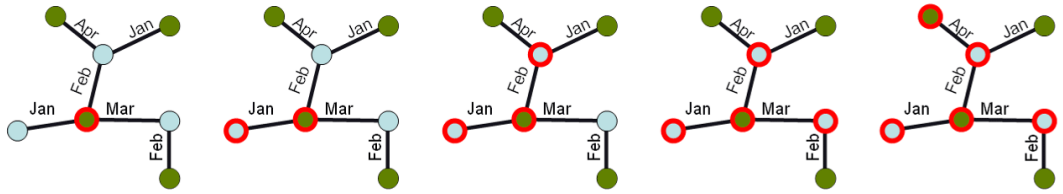

Figure 11-3 A temporal network
A hypothetical temporal heterosexual network and the spread of an STI starting from the centre node in January. The shades represent sex and the bold ring signifies infected. The relationships in this example are dated i.e. only exist the indicated month. The disease propagation therefore is limited by progression of time. Some of the nodes cannot be infected since their link to an infected node does not exist at the time of the attempted transmission.

degree, if network jargon is used. But surprisingly few probe the rate of contacts, the intensity, or the concurrency. We would do much better if these parameters were known.

The fact of the matter is, whenever you are dealing with networks the data requirements are very high and hence the quality of data available almost always insufficient. To get public opinion about a political party or to assess the prevalence of smoking related disease it is enough to sample a few thousand of the population. But it is easy to see that people in a random sample are unlikely to be in contact with each other. To construct a network that properly reflects the real situation, you pretty much need data on every node in it and the responses must be accurate and true to definition.[241] To estimate statistical properties requires a significant portion of the network to be known. Usually this entails a more refined sampling technique, dependant on what you are after.[242]

At the very least, the degree distribution will tell you how many partners your partners have, taken on average. Following a counterintuitive argument known as the *ripple effect*, this number is likely to be higher than your own.[243] That means to say, your partners have more partners than you do. Expanding on this argument to an arbitrary number of steps, plus taking into account the dichotomous nature of the heterosexual network, as well as a property known as *mutuality*, we demonstrate that a network *would* have a giant component provided that the mutuality is sufficiently high.

Mutuality expresses in a number how likely it is for two or more of your partners to share a common partner – other than yourself, of course. A high mutuality means that the network is tightly bound, *clustered*. Despite a high average degree, the components of the network do not spread out to include as many as they potentially could. Hence a high mutuality works against the forming of a giant component.

A low mutuality, on the other hand instead means the network is tree like, branching out in all directions. When a new link is formed from a component, it is likely that it will be with a node not already in the component. The components will seek to interconnect. Hence a low mutuality will work in favor of giant component formation.

Our data suggests that a low mutuality is probable and a high mutuality, conversely, is improbable. We have data from two sets from the same population. The first set reflects the number of partners over one year. The second, over a whole lifetime. Both data sets seem to favor giant components.

11.4.2 Simulated networks

Another caveat is the core-group theory, discussed earlier. The analytical approach assumed that connections are random. If there is a tendency for heavily linked individuals to attach only to other heavily linked individuals, then this also would undermine our argument. As much as we would like to expand our mathematical model to consume this new variable, we realized the limits of our abilities and, thematic to this thesis, resorted to computers where no bounds of intractable equations obstruct.

We ran into a lot of complications when building our networks but the principle is easy to explain. Start with a set of male nodes and a set of female nodes. Assign degrees to each according to a known distribution. Then attempt to connect them, constrained by the assigned degrees. When no more links can be added, check that there are no, or at least acceptably few, nodes with connections "left over".

Further more, we could apply dates to our links so that when one relationship ceases, another begins. If this is done carefully, a realistic temporal network can be achieved, meaning, minimizing the time a node is not connected to anybody. This means that, necessarily, some relationships are long, some are short.

When a node attempts to connect to another, this is choice can be uniformly random or it the randomness can be affected by a variable of choice e.g. node degree. This would result in increasing *assortativity,* if we made it increasingly likely for high degree nodes to favor other high degree nodes when connecting.

Programming this scheme into the computer, we generated thousands of networks with varying sizes and assortativity. We measured mutuality, but more to the point, we looked for and recorded the size of the giant component.

Our conclusion from this enquiry was that the giant component quickly emerges at around 12-months of sampling. If the data collection period were more than twelve months, then we feel that a significant giant component would exist regardless of the parameters defining the network. 12 months is after all, from a Chlamydia perspective, a rather short period of time. This is a disease which can be latent for many years.

11.5 FUTURE RESEARCH

One consideration that did not make it into the Paper I is concurrency. Using the assumptions from Morris and Kretzschmar,[88] that a varying proportions of the population engage in concurrent relationships, we could have made things very easy for ourselves, certainly producing a giant component in all runs. Keep this in mind when reading the paper. Concurrency is very widespread, not only in sub-saharan Africa, but in the western world as well, particularly among the young. Concurrency is probably a key variable in the chaotic incidence curves of STIs.[244]

Ideally we wouldn't have to think about issues as concurrency and assortativity. This is information that is available by asking people. The trick is measuring it. It may require something more than simple multiple choice enquiries. One way to go about it is to use a timeline. Respondents would register the sexual partners over a period of time using X's indicating a single encounter or a bar for a lengthier relationship. It is perceivable that using a timeline would minimize recollection bias in the number of relationships, the timings of them and any other variables simultaneously collected. It is known that in sex surveys, men exaggerate the number of contacts. When using the timeline, this discrepancy between the sexes, it seams, is eliminated.[24]

Yes, the timeline has been used and the results published, almost 20 years ago. The original timeline was a part of a survey with about 1000 youths in the county of Gotland, in Sweden. The respondents jotted X:s and bars with pencil.
Work has recently been complete to bring the timeline into the 21st century for the benefit of today's internet savvy youths. The timeline is fully interactive and can be inserted into any website, for instance, an internet survey.
I hope to see the timeline fully tested soon to see differences in variables compared with other surveys. Ultimately, I'd like to see it used continuously in full scale surveys in all sorts of communities. When this happens, we must remember to sample longer than 12 months, as was suggested in Paper IV – perhaps 18 or 24. Data from the timeline would certainly make building networks of the kind we did in Paper IV, a breeze.

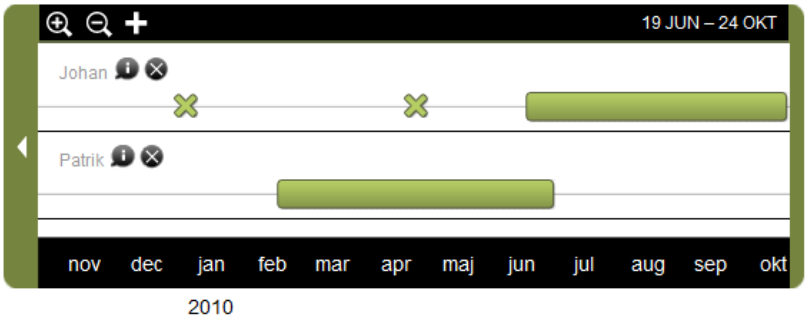

Figure 11-4 The Timeline control as seen on an internet survey
The respondent has specified two sexual partners over the time period with some concurrency. There was an encounter with Johan around New Year and end of April. The last occurred during the relationship with Patrik which lasted from the middle of February through June. The end of the relationship with Patrik overlapped the intensified relationship with Johan.

12 ACKNOWLEDGEMENTS

”Grad students are not bad people, they just made the wrong life choices.” REF! To my defense, I did not pursue PhD-status actively, although I did immediately accept when the possibility opened up. The person responsible and who I’ve since periodically both thanked and cursed for his confidence in me is Johan Giesecke who was the head of the department when I did my masters thesis at Epidemiologen. He is also my supervisor and incidentally a fellow physicist from KTH.

My PhD term has indebted me to a long list of people, but if I had to pick out only one person it would be Fredrik Liljeros, supervisor, guide, mentor and friend. With acute tactical skills, he made sure I made the right decisions when I was inclined to make the wrong ones - and I’ve been a slow learner. I admire his sensibility and his academic proficiency. As I gradually depart from academia, I hope I can count on him for sensible advice from time to time in the future.

When it comes to mathematics, statistics and method, I’ve relied on two of the best there are, professors of mathematical statistics, Åke Svensson and Tom Britton. Both have had something to do with all of my work, in particular, Åke is co-author of Paper II and Tom is responsible for setting me on course with my masters thesis and my first published piece, Paper I. Åke was also stationed part-time at Epi where his door was always open and his patience with inept engineers virtually limitless.

I have had a secret patron and benefactor in Anders Tegnell, commissioning projects in the fields of pandemic preparedness, blood transfusion risk and more. My PhD-term would have not worked out without his confidence and patience.

While we’re on the topic of financial providing, Dysonet has funded me for the travel restrictions papers and a few projects that have not yielded results. As side projects, S-gem has given me employment periodically for administration and web management. I’m also grateful to the Department of Sociology and SMI for the teaching I’ve done there, to SMI and Malin also for Tessy. Finally, I would like to acknowledge the Microsim project, with Lisa, Kalle and Baki. This thesis was to include one or two papers to account for the massive work we put into the system. Finally a few papers emerged, but not in time to be but referenced.

Due to the delay with Microsim there was a shortage of papers, at least when I started off on this thesis. Unknowingly, Philip Polgreen and Forrest Nelson saved the thesis by prompting me to write a paper about StatFlu and I owe them thanks for that.

And where would I be here without people asking me that most annoying question of all, ”When are you done?”. That would be Anna (best roomy, ever!!), Malin and Annika.

Thanks to the members of S-gem, Stockholm Group of Epidemic Modeling. During those meetings in Trosa and Mariefred, the gathered academic brilliance was inspiring on those early days of my researching career.

I’d like to thank those I’ve cooperated with in other research projects, Gustaf, Britt, Rut, Torsten, Martin, Lina, Linus, Xin, Veronika...

Thank you colleagues at SMI. If my lack of presence, even when present, has given you the impression I’m not enjoying myself at Epi, that’s just not so. I’ve always felt right at home. Where would you find more mommies than at Epi? Anna is the best roomy but close runners up were Hanna, Monika, Helena, Anette, Baki, Lisa, Kalle, Patricia, Maria, Paul…

Everyone knows that being a grad student is at least 30 % administration. With few exceptions I've never had to administrate anything but myself and my work and I find even that a terrible nuisance. That makes me appreciate my supervisor Fredrik, my bosses Katarina, Kasia, Yvonne, Sharon and Anette, and the administrators Louise and Salou at SMI. Thanks also to administrators at MEB, particularly Camilla.
I've made an effort to show my face once in a while at MEB but I'm afraid it's mostly been take-take relationship on my part. MEB is a department of the highest standard which I am proud to be a part of. A lot of people at MEB deserve thanks, the administrative staff, teachers (also at IMM) and my fellow postgraduate colleagues. The dedication deservedly goes out to Cecilia for long standing mom-service. You've got *a lot* of accumulated son-service to cash whenever you need it.
Jannike, you might think your hours and hours of proofreading finally made you a part of this thesis. You've had *everything* to do with it, always. Through it all, you've kept me healthy, sane and happy to be alive. And yet, because of you, this thesis is just a thesis.

13 LIST OF FIGURES

14 REFERENCES

1. Brownstein, J.S., C.J. Wolfe, and K.D. Mandl. *Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States*. PLoS Med. 2006;3(10):e401.
2. Cooper, B.S., et al. *Delaying the international spread of pandemic influenza*. PLoS Med. 2006;3(6):e212.
3. Pastor-Satorras, R. and A. Vespignani. *Epidemic spreading in scale-free networks*. Physical review letters. 2001;86(14):3200-3203.
4. Pastor-Satorras, R. and A. Vespignani. *Epidemic dynamics and endemic states in complex networks*. Physical Review E. 2001;63(6):66117.
5. Liljeros, F., C.R. Edling, and L.A.N. Amaral. *Sexual networks: implications for the transmission of sexually transmitted infections*. Microbes and Infection. 2003;5(2):189-196.
6. Colizza, V. and A. Vespignani. *Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations*. J Theor Biol. 2008 Apr 7;251(3):450-67.
7. Hufnagel, L., D. Brockmann, and T. Geisel. *Forecast and control of epidemics in a globalized world*. Proc Natl Acad Sci U S A. 2004 Oct;101(42):15124-15129.
8. Angel, J. *A Mathematical Sociologist's Tribute to Comte: Sociology as Science*. 2007 [cited 2009 2009-03-19]; Available from: http://www.asanet.org/footnotes/feb07/fn9.html.
9. Swartz, R. *Fördomen kan bli ett intellektuellt terrorvapen*. SvD. 2009 2009-03-18.
10. Wikipedia. *Sociology*. [cited; Available from: http://en.wikipedia.org/wiki/Sociology.
11. Coleman, J.S. *Introduction to Mathematical Sociology*. New York: Free Press of Glencoe; 1964.
12. Karlsson, G. *Social Mechanisms*. New York: Free Press of Glencoe; 1958.
13. Rashevsky, N. *Mathematical Bilogy of Social Behavior*. Chicago: Chicago University Press; 1951.
14. Simon, H.A. *Models of man: social and rational: mathematical essays on rational human behavior in a social setting*. Wiley New York; 1957.
15. Helbing, D., I. Farkas, and T. Vicsek. *Simulating dynamical features of escape panic*. Nature. 2000;407(6803):487-490.
16. Björk and N. Hooper, *Human Behaviour*. [Music] 1993, One Little Indian. p. CD, 12"-vinyl, kassett.
17. Epstein, J.M. and R. Axtell. *Growing artificial societies: social science from the bottom up*. The MIT Press; 1996.
18. Chattoe, E. *4 Why Are We Simulating Anyway? Some Answers from Economics*. Social science microsimulation. 1996.
19. Brouwers, L., et al. *Economic consequences to society of pandemic H1N1 influenza 2009 – preliminary results for Sweden*. Euro Surveill. 2009;14(37).
20. Statens Institut för Kommunikationsanalys, S.S.C., *RES 2001. Den nationella reseundersökningen*. [PDF] 2002.
21. Camitz, M., *The effect of travel restrictions on the spread of a highly contagious disease in Sweden*, in *Theoretical Biological Physics,* 2005, Royal Institute of Technology: Stockholm.
22. Statistics Sweden, *Statistik databasen*, Statistics Sweden.

23. Centre for Epidemiology, *Swedish Hospital Discharge Diagnosis Register, 1987-2006*. [Database] 2006, National Board of Health and Welfare: Stockholm.
24. Giesecke, J., et al. *Sexual behaviour related to the spread of sexually transmitted diseases--a population-based survey*. Int J of STD & AIDS. 1992;3(4):255.
25. Uknown. The Bible. Exodus
26. Jopling, W.H. *Leprosy stigma*. Leprosy review. 1991;62(1):1-12.
27. Nelson, K.E. and C. Williams. *Infectious Disease Epidemiology: Theory and Practice*. Jones and Barlett Publications; 2008.
28. Fracastoro, G., *Syphilis sive morbus gallicus*, 1530.
29. Mollison, D. *Epidemic models: their structure and relation to data*. Cambridge Univ Pr; 1995.
30. Dorland, W.A.N. *Dorland's Medical Dictionary*. 26th edition ed. Philadelphia: Saunders, W B Co; 1981.
31. Giesecke, J. *Modern infectious disease epidemiology*. Arnold London; 2002.
32. Garret, L. *The coming plague: Newly Emerging Diseases in a World Out of Balance*. 5th edition ed. NY: Farrar, Strous & Giroux Publ; 1994.
33. *FAQs*. [cited 2009-01-09]; Available from: http://lhncbc.nlm.nih.gov/apdb/phsHistory/faqs.html.
34. Olausson, P. *Faktoider - Försanthållna osanningar, halvsanningar och missuppfattningar*. Bokförlaget Forum, Stockholm; 2007.
35. Diamond, J.M. *Guns, germs, and steel: The fates of human societies*. WW Norton & Company; 1999.
36. Anderson, R.M. and R.M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, USA; 1991.
37. Jones, K.E., et al. *Global trends in emerging infectious diseases*. Nature. 2008;451(7181):990-993.
38. Struwe, T. and B. Olsson-Liljequist, *A report on Swedish antimicrobial utilization and resistance in human medicine.*, in *SWEDRES* 2008.
39. Weiss, R.A. and A.J. McMichael. *Social and environmental risk factors in the emergence of infectious diseases*. Nature Medicine. 2004;10:S70-S76.
40. Tang, J.W. *The effect of environmental parameters on the survival of airborne infectious agents*. Journal of The Royal Society Interface. 2009;6(Suppl 6):S737.
41. Parkin, D.M. *The global health burden of infection-associated cancers in the year 2002*. Int J of Cancer. 2006;118(12):3030-3044.
42. Persson, C., *Roles of Helicobacter Pylori infection, host genetic variation and other environmental exposures in gastric carcinogenesis*, in *Intstitutionen för Medeicinsk Epidemiolgi och Biostatistik* 2009, Karolinska Institutet: Solna.
43. Dhurandhar, N.V., et al. *Human Adenovirus Ad-36 Promotes Weight Gain in Male Rhesus and Marmoset Monkeys 1 2*. Journal of Nutrition. 2002;132(10):3155-3160.
44. Eising, S., et al. *Type 1 diabetes risk analysis on dried blood spot samples from population-based newborns: design and feasibility of an unselected case-control study*. Paediatric and Perinatal Epidemiology. 2007;21(6):507.
45. Horwitz, M.S., et al. *Diabetes induced by Coxsackie virus: initiation by bystander damage and not molecular mimicry*. Nature Medicine. 1998;4(7):781-785.
46. Christakis, N.A. and J.H. Fowler. *The spread of obesity in a large social network over 32 years*. New England Journal of Medicine. 2007;357(4):370.

47. Smieszek, T., *Models of epidemics: How contact characteristics shape the spread of infectious diseases*, 2010, ETH: Zürich.
48. Rothman, K.J., S. Greenland, and T.L. Lash. *Modern epidemiology*. Lippincott Williams & Wilkins; 2008.
49. Treanor, J. *Influenza vaccine--outmaneuvering antigenic shift and drift*. New England Journal of Medicine. 2004;350(3):218.
50. Morner, M. *Något att hålla i när det blåser*. Smittskydd. 2009.(3)
51. Timo, S. *A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease spread*. Theoretical Biology and Medical Modelling. 2009;6.
52. Brunham, R. *Core group theory: a central concept in STD epidemiology*. Venereology. 1997;10:34-39.
53. Kiss, I.Z., D.M. Green, and R.R. Kao. *Infectious disease control using contact tracing in random and scale-free networks*. Journal of The Royal Society Interface. 2006;3(6):55.
54. (HMP), H.M.P. *Human Microbiome Project (HMP)*. 2010 [cited 2010 2010-10-20]; Available from: http://nihroadmap.nih.gov/hmp/.
55. Feschotte, C. *Virology: Bornavirus enters the genome*. Nature. 2010;463(7277):39-40.
56. Andersson, G.E., et al. *On the origin of mitochondria: a genomics perspective*. Philosophical Transactions B. 2003;358(1429):165.
57. Giesecke, J. [Personal Communication]. 2005.
58. Bernoulli, D. *Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir*. Mémoires de Mathématiques et de Physique, Académie Royale des Sciences, Paris. 1760:1–45.
59. Heesterbeek, J.A.P. *A brief history of R_0 and a recipe for its calculation*. Acta Biotheoretica. 2002;50(3):189-204.
60. Hamer, W.H. *Epidemic disease in England*. Lancet. 1906;1:733–739.
61. Ross, R. *Report on the Prevention of Malaria in Mauritius*. Waterlow, London; 1908.
62. Ross, R. *The prevention of malaria*. Murray; 1910.
63. Gladwell, M. *The tipping point: How little things can make a big difference*. Little, Brown and Company; 2000.
64. Anderson, D. and R.A.Y. Watson. *On the spread of a disease with gamma distributed latent and infectious periods*. Biometrika. 1980;67(1):191-198.
65. Kermack, W.O. and A.G. McKendrick. *A contribution to the mathematical theory of epidemics*. Proceedings of the Royal Society. 1927;115:700-721.
66. Bailey, N.T.J. *The mathematical theory of epidemics*. Hafner; 1957.
67. Hufnagel, L., D. Brockmann, and T. Geisel. *Forecast and control of epidemics in a globalized world*. Proceedings of the National Academy of Sciences of the United States of America. 2004 Oct;101(42):15124-15129.
68. Rangel-Frausto, M.S., et al. *The natural history of the systemic inflammatory response syndrome (SIRS). A prospective study*. Jama. 1995;273(2):117-123.
69. Hethcote, H.W. *The mathematics of infectious diseases*. SIAM review. 2000;42(4):599-653.
70. Bansal, S., B.T. Grenfell, and L.A. Meyers. *When individual behaviour matters: homogeneous and network models in epidemiology*. Journal of The Royal Society Interface. 2007;4(16):879.
71. Keeling, M.J. and B.T. Grenfell. *Individual-based perspectives on R(0)*. J Theor Biol. 2000 Mar 7;203(1):51-61.

72. Gray, R.H., et al. *Stochastic simulation of the impact of antiretroviral therapy and HIV vaccines on HIV transmission; Rakai, Uganda*. Aids. 2003;17(13):1941.
73. Everitt, B.S. *Medical Statistics from A to Z: A Guide for Clinicians and Medical Students*. 2 edition ed.: Cambridge University Press; 2006.
74. Heffernan, J.M., R.J. Smith, and L.M. Wahl. *Perspectives on the basic reproductive ratio*. Journal of the Royal Society Interface. 2005;2(4):281.
75. Nåsell, I. *The threshold concept in stochastic epidemic and endemic models*. Epidemic Models: Their Structure and Relation to Data. Cambridge: Cambridge University Press; 1995. p. 71–83
76. Riley, S., et al. *Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions*. Science. 2003 Jun;300(5627):1961-1966.
77. Lipsitch, M., et al. *Transmission dynamics and control of severe acute respiratory syndrome*. Science. 2003;300(5627):1966-1970.
78. Chowell, G., et al. *Model parameters and outbreak control for SARS*. Emerg Infect Dis. 2004;10(7):1258-1263.
79. Kretzschmar, M. and J. Wallinga. *Mathematical Models in Infectious Disease Epidemiology*. In: K. Dietz, et al., editors. Modern Infectious Disease Epidemiology: Springer New York. p. 209-221
80. Nasell, I. *The threshold concept in stochastic epidemic and endemic models*. Epidemic Models: Their Structure and Relation to Data. 1995:71–83.
81. Mollison, D. *Dependence of epidemic and population velocities on basic parameters*. Mathematical Biosciences. 1991;107(2):255-288.
82. Dawkins, R. *The God Delusion*. 1st ed.: Houghton Mifflin Company; 2006.
83. Moore, G.E. *Cramming More Components onto Integrated Circuits*. Proceedings of the IEEE. 1998;86(1).
84. Moore, G.E. *Cramming more components onto integrated circuits*. Electronics. 1965;38(8).
85. staff, I. *Gordon Moore Says Aloha to Moore's Law*. The Inquirer. 2005(13 April).
86. Mangili, A. and M.A. Gendreau. *Transmission of infectious diseases during commercial air travel*. The Lancet. 2005;365(9463):989-996.
87. Moser, M.R., et al. *An outbreak of influenza aboard a commercial airliner*. Am J Epidemiol. 1979;110(1):1.
88. Morris, M. and M. Kretzschmar. *Concurrent partnerships and the spread of HIV*. AIDS. 1997;11(5):641.
89. Weiss, H.A., M.A. Quigley, and R.J. Hayes. *Male circumcision and risk of HIV infection in sub-Saharan Africa: a systematic review and meta-analysis*. Aids. 2000;14(15):2361.
90. Caldwell, J.C. and P. Caldwell. *The African AIDS epidemic*. Scientific American. 1996;274(3):62-68.
91. Caldwell, J.C. *Rethinking the African AIDS epidemic*. Population and development review. 2000;26(1):117-135.
92. Balcan, D., et al. *Seasonal transmission potential and activity peaks of the new influenza A (H 1 N 1): a Monte Carlo likelihood analysis based on human mobility*. BMC Med. 2009;7(1):45.
93. Sypsa, V. and A. Hatzakis. *School closure is currently the main strategy to mitigate influenza A(H1N1)v: a modeling study*. Euro Surveill. 2009 Jun 18;14(24).
94. Vaillant, L., et al. *Epidemiology of fatal cases associated with pandemic H1N1 influenza 2009*. Euro Surveill. 2009;14(33).

95. Nesterov, A.V. *Philosophy of modeling*. Automatic Documentation and Mathematical Linguistics. 2007;41(3):73-79.
96. Robson, C. *Real world research: A resource for social scientists and practitioner-researchers*. Blackwell Pub; 2002.
97. Mollison, D. *The structure of epidemic models*. Epidemic Models: their structure and relation to data. 1995:17-33.
98. Barlas, Y. *Formal aspects of model validity and validation in system dynamics*. System dynamics review. 1996;12(3).
99. Barlas, Y. and S. Carpenter. *Philosophical roots of model validation: Two paradigms*. System Dynamics Review. 1990;6(2).
100. House, E.R. *Realism in research*. Educational Researcher. 1991;20(6):2.
101. Brooks, R., *Interview*, 1992, MIT.
102. Baker, A., *Simplicity*, in *In The Stanford Encyclopedia of Philosophy* 2008, Zalta.
103. Brouwers, L., *Microsimulation Models for Disaster Policy Making*, in *Department of Computer & Systems Sciences* 2005, Stockholm University/Royal Institute of Technology.
104. Colizza, V., et al. *Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions*. PLoS Med. 2007;4(1):95.
105. Andersson, H. and T. Britton. *Stochastic epidemic models and their statistical analysis*. Springer Verlag; 2000.
106. Rvachev, L.A. and I.M. Longini Jr. *A mathematical model for the global spread of influenza*. Mathematical Biosciences. 1985;75(1):3-22.
107. Wikipedia. *Likelihood*. [cited; Available from: http://en.wikipedia.org/wiki/Likelihood.
108. World Health Organization. *Swine influenza* 2009 [cited 2010-10-01]; Available from: http://www.who.int/mediacentre/news/statements/2009/h1n1_20090425/en/index.html.
109. Giesecke, J. [Personal Communication]. 2010.
110. Scalia Tomba, G., et al. *Some model based considerations on observing generation times for communicable diseases*. Mathematical Biosciences.223(1):24-31.
111. Wallinga, J., P. Teunis, and M. Kretzschmar. *Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents*. Am J Epidemiol. 2006;164(10):936.
112. Halloran, M.E., et al. *Containing bioterrorist smallpox*. Science. 2002 Nov 15;298(5597):1428-32.
113. Pourbohloul, B., et al. *Modeling control strategies of respiratory pathogens*. Emerg Infect Dis. 2005;11(8).
114. Newman, M.E.J. *Spread of epidemic disease on networks*. Physical Review E. 2002;66(1):16128.
115. Brouwers, L., et al. *Micro-simulation of a smallpox outbreak using official register data*. Euro Surveill.15(35).
116. Bankes, S. *Exploratory modeling for policy analysis*. Operations Research. 1993;41(3):435-449.
117. DeVuyst Paul, V. and A. Eric. *Sensitivity analysis revisited: A quadrature-based approach*. Journal of Policy Modeling. 1997;19(2):175-185.
118. Iman, R.L. and W.J. Conover. *Small sample sensitivity analysis techniques for computer models. with an application to risk assessment*. Communications in Statistics-Theory and Methods. 1980;9(17):1749-1842.

119. Meltzer, M.I., et al. *Modeling Potential Responses to Smallpox as a Bioterrorist Weapon*. Emerg Infect Dis. 2001;7(6):959-969.
120. Wearing, H.J., P. Rohani, and M.J. Keeling. *Appropriate models for the management of infectious diseases*. PLoS Med. 2005;2(7):621.
121. Lloyd, A., L. . *Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods* Procedings of the Royal Society. 2007 May 7, 2001;268(1470):985-993.
122. Lloyd, A.L. *Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods* Procedings of the Royal Society. 2007 May 7, 2001;268(1470):985-993.
123. Cliff, A.D. and P. Haggett. *Statistical modelling of measles and influenza outbreaks*. Statistical Methods in Medical Research. 1993;2(1):43.
124. Bootsma, M.C. and N.M. Ferguson. *The effect of public health measures on the 1918 influenza pandemic in U.S. cities*. Proc Natl Acad Sci U S A. 2007 May 1;104(18):7588-93.
125. Fraser, C., et al. *Factors that make an infectious disease outbreak controllable*. Proc Natl Acad Sci U S A. 2004;101(16):6146.
126. Longini, I.M., Jr., et al. *Containing pandemic influenza at the source*. Science. 2005 Aug 12;309(5737):1083-7.
127. Wallinga, J. and P. Teunis. *Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures*. Am J Epidemiol. 2004;160(6):509.
128. Ferguson, N.M., et al. *Strategies for containing an emerging influenza pandemic in Southeast Asia*. Nature. 2005 Sep 8;437(7056):209-14.
129. Couch, R.B., et al. *Induction of partial immunity to influenza by a neuraminidase-specific vaccine*. Journal of Infectious Diseases. 1974;129(4):411-420.
130. Kaplan, E.H., D.L. Craft, and L.M. Wein. *Emergency response to a smallpox attack: the case for mass vaccination*. Proc Natl Acad Sci U S A. 2002 Aug 6;99(16):10935-40.
131. Eubank, S., et al. *Modelling disease outbreaks in realistic urban social networks*. Nature. 2004;429(6988):180-184.
132. Porco, T.C., et al. *Logistics of community smallpox control through contact tracing and ring vaccination: a stochastic network model*. BMC Public Health. 2004 Aug 6;4:34.
133. Carrat, F., et al. *A 'small-world-like' model for comparing interventions aimed at preventing and controlling influenza pandemics*. BMC Med. 2006;4(1):26.
134. Gallos, L.K., et al. *Improving immunization strategies*. Phys Rev E Stat Nonlin Soft Matter Phys. 2007 Apr;75(4 Pt 2):045104.
135. Longini Jr, I.M., et al. *Optimal vaccine trial design when estimating vaccine efficacy for susceptibility and infectiousness from multiple populations*. Stat Med. 1998;17(10):1121-1136.
136. Lloyd, A.L. *The dependence of viral parameter estimates on the assumed viral life cycle: limitations of studies of viral load data*. Proceedings of the Royal Society B: Biological Sciences. 2001;268(1469):847.
137. Carrat, F., et al. *Time lines of infection and disease in human influenza: a review of volunteer challenge studies*. Am J Epidemiol. 2008;167(7):775.
138. Cauchemez, S., et al. *Estimating the impact of school closure on influenza transmission from Sentinel data*. Nature. 2008;452(7188):750-754.
139. Pantelic, J., et al. *Personalized ventilation as a control measure for airborne transmissible disease spread*. Journal of The Royal Society Interface. 2009;6(Suppl 6):S715.

140. Colizza, V., et al. *The role of the airline transportation network in the prediction and predictability of global epidemics*. Proc Natl Acad Sci U S A. 2006;103(7):2015.
141. Camitz, M. and F. Liljeros. *The effect of travel restrictions on the spread of a moderately contagious disease*. BMC Med. 2006;4:32.
142. Cowling, B.J., et al. *Facemasks and hand hygiene to prevent influenza transmission in households*. Annals of Internal Medicine. 2009;151(7):437.
143. Nelson, J.D. *The changing epidemiology of pertussis in young infants: the role of adults as reservoirs of infection*. Archives of Pediatrics and Adolescent Medicine. 1978;132(4):371.
144. Sattenspiel, L. and A.L. Lloyd. *The Geographic Spread of Infectious Diseases: Models and Applications*. 1st edition ed.: Princeton University Press; 2001.
145. Grais, R.F., J. Hugh Ellis, and G.E. Glass. *Assessing the impact of airline travel on the geographic spread of pandemic influenza*. European J of Epidemiol. 2003;18(11):1065-1072.
146. Klontz, K.C., et al. *An Outbreak of Influenza a/Taiwan/1/86 (H1n1) Infections at a Naval Base and Its Association with Airplane Travel*. Am J Epidemiol. [Article]. 1989 Feb;129(2):341-348.
147. Chen, L.H. and M.E. Wilson. *The role of the traveler in emerging infections and magnitude of travel*. Medical Clinics of North America. 2008;92(6):1409-1432.
148. Wilson, M.E. *Travel and the emergence of infectious diseases*. Journal of Agromedicine. 1996;3(1):51-66.
149. Flahault, A. and A.J. Valleron. *A method for assessing the global spread of HIV-1 infection based on air travel*. Mathematical Population Studies. 1992;3(3):161-171.
150. Klaus, S. *A structured epidemic model incorporating geographic mobility among regions*. Mathematical Biosciences. 1995;128(1-2):71-91.
151. Sattenspiel, L. and K. Deitz. *A structured epidemic model incorporating geographic mobility among regions*. Mathematical Biosciences. 1995;128(1-2):71-91.
152. Flahault, A., S. Deguen, and A.J. Valleron. *A mathematical model for the European spread of influenza*. European J of Epidemiol. 1994;10(4):471-474.
153. Ball, F. *Stochastic and deterministic models for SIS epidemics among a population partitioned into households*. Mathematical Biosciences. 1999;156(1-2):41-67.
154. Colizza, V., et al. *The modeling of global epidemics: Stochastic dynamics and predictability*. Bull Math Biol. 2006;68(8):1893-1921.
155. Lefevre, C. and P. Picard, *Collective epidemic processes: a general modeling approach to the final outcome of SIR infectious diseasesIn: Mollison D, editor. Epidemic Models*, 2000, Cambridge, United Kingdom: Cambridge University Press.
156. Lloyd, A.L. *Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics*. Theoretical Population Biology. 2001;60(1):59-71.
157. Asikainen, T. *Some results in the field of epidemic modeling and analysis of a smallpox outbreak*. Mathematical statistics, Stockholm University.
158. Daley, D.J. *Epidemic modelling: an introduction*. Cambridge University Press; 1999.
159. Svensson, Å. *A note on generation times in epidemic models*. Mathematical Biosciences. 2007;208(1):300-311.
160. Norris, J.R. *Markov Chains* Cambridge: Cambridge University Press; 1998.

161. Bailey, N.T.J. and J. Estreicher. *Epidemic prediction and public health control, with special reference to influenza and AIDS*. Proceedings, World Congress of the Bernoulli Society. 1987;1:507–516.
162. Longini Jr, I.M., et al. *Statistical analysis of the stages of HIV infection using a Markov model*. Stat Med. 1989;8(7):831-43.
163. Bailey, N.T.J. *Some stochastic models for small epidemics in large populations*. Applied Statistics. 1964;13(1):9-19.
164. Becker, N. *The uses of epidemic models*. Biometrics. 1979:295-305.
165. van Genugten, M.L.L., M.L.A. Heijnen, and J.C. Jager. *Scenario analysis of the expected number of hospitalisations and deaths due to pandemic influenza in the Netherlands*. [PDF]. 2003.
166. Balci, O., *Validation, verification, and testing techniques throughout the life cycle of a simulation study*, in *Proceedings of the 26th conference on Winter simulation* 1994, Society for Computer Simulation International: Orlando, Florida, United States.
167. Vinnova. *Svensk FoU i siffror 2009*. [cited 2010-07-01]; Available from: http://www.vinnova.se/sv/Om-VINNOVA/Svensk-FoU-i-siffror/.
168. Vetenskap & Allmänhet, V., *Politik och Vetenskap - Vetenskap & Allmänhets politikerstudie i korthet*, 2006:5.
169. Analyseinstitut for Forskning, D., *Politikere og forskning*, D. Analyseinstitut for Forskning, Editor 2000:4.
170. *The Ig® Nobel Prizes*. [cited 2010-10-15]; Available from: http://improbable.com/ig/.
171. Doyle, A., et al. Estimation de l'impact d'une pandémie grippale et analyse de stratégies Journal [Electronic article]. 2005: Available from: http://www.invs.sante.fr/publications/2005/pandemie_grippale_170205/rapport_pandemie_grippale.pdf.
172. Meltzer, M.I., N.J. Cox, and K. Fukuda. *The economic impact of pandemic influenza in the United States: priorities for intervention*. Emerg Infect Dis. 1999 Sep-Oct;5(5):659-71.
173. van Genugten, M.L.L., M.L.A. Heijnen, and J.C. Jager. *Pandemic influenza and Healthcare demand in the Netherlands: Scenario analysis*. Emerg Infect Dis. [Article]. 2003 May;9(5):531-538.
174. Camitz, M., *Översikt över statiska influensamodeller för pandemiberedskap*, 2006, Socialsyrelsen.
175. Meltzer, M.I., N.J. Cox, and K. Fukuda. *Modeling the economic impact of pandemic influenza in the United States: Implications for setting priorities for intervention* 1999 [cited; Available from: http://www.cdc.gov/ncidod/eid/vol5no5/melt_back.htm.
176. Molinari, N.A.M., et al. *The annual impact of seasonal influenza in the US: measuring disease burden and costs*. Vaccine. 2007;25(27):5086-5096.
177. Critchfield, G.C. and K.E. Willard. *Probabilistic Analysis of Decision Trees Using Monte-Carlo Simulation*. Medical Decision Making. 1986;6(2):85-92.
178. Doubilet, P., et al. *Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach*. Med Decis Making. 1985 Summer;5(2):157-77.
179. FluSurge, *FluSurge*. [Excel Spreadsheet] 2005, CDC.
180. FluAid, *FluAid*. [Windows Application], CDC.
181. Schopflocher, D.P., et al. *Pandemic influenza planning: Using the US centers for disease control FluAid software for small area estimation in the Canadian context*. Annals of Epidemiology. [Article]. 2004 Jan;14(1):73-76.

182. Menon, D.K., B.L. Taylor, and S.A. Ridley. *Modelling the impact of an influenza pandemic on critical care services in England*. Anaesthesia. [Article]. 2005 Oct;60(10):952-954.
183. Anderson, T.A., G.K. Hart, and M.A. Kainer. *Pandemic influenza-implications for critical care resources in Australia and New Zealand*. Journal of Critical Care. [Article]. 2003 Sep;18(3):173-180.
184. Wilson, N., O. Mansoor, and M. Baker. *Estimating the impact of the next influenza pandemic on population health and health sector capacity in New Zealand*. N Z Med J. 2005 Mar 11;118(1211):U1346.
185. Wilson, N., et al. *Modeling the impact of pandemic influenza on Pacific Islands*. Emerging Infectious Diseases. [Letter]. 2005 Feb;11(2):347-349.
186. The national Influenza Pandemic Planning Committee, *A Model Plan for Influenza Pandemic Preparedness*
2005, NDSC p. 25.
187. Influenza Pandemic Planning Committee of the Communicable Diseases Network Australia and New Zealand, *A framework for an australian influenza pandemic plan*
in *Technical Report Series No. 4*, Pubmed, Editor. [Socialstyrelsens riktlinjer för planering/tillsyn/metodutv] 2005, NDSC p. 25.
188. Wallinga, J., T.J. Hagenaars, and M. van Genugten. *Scenario analysis-estimating the effect of different interventions during an influenza pandemic*. Euro Surveill. 2004;8:19.
189. Mylius, S.D., et al. *Optimal allocation of pandemic influenza vaccine depends on age, risk and timing*. Vaccine. 2008;26(29-30):3742-3749.
190. Camitz, M. *StatFlu*. 2007 [cited 2010-10-31]; Available from: http://www.s-gem.se/statflu.
191. Bonmarin, I. and D. Lévy-Bruhl. *Estimation du nombre hebdomadaire d'admissions et de journées d'hospitalisation lors d'une pandémie grippale*. 2005.
192. Meltzer, M.I., N.J. Cox, and K. Fukuda. Modeling the economic impact of pandemic influenza in the United States: Implications for setting priorities for intervention Journal [Electronic article]. 1999: Available from: http://www.cdc.gov/ncidod/eid/vol5no5/melt_back.htm.
193. Mullooly, J.P. and W.H. Barker. *Impact of Type A influenza on children: A restrospective sudy*. Am J Public Health. 1982;72:1008-1016.
194. Schoenbaum, S.C., B.J. McNeil, and J. Kavet. *Swine-Influenza Decision*. New England Journal of Medicine. 1976;295(14):759-765.
195. Barker, W.H. and J.P. Mullooly. *Impact of epidemic Type A influenza on in a defined adult population*. Am J Epidemiol. 1980;112(6):789-813.
196. Welfare, T.N.B.o.H.a. *Health and Medical Care: Performances in Health Care*. Stockholm: The National Board of Health and Welfare, Sweden; 2005.
197. Glezen, W.P. *Emerging infections: Pandemic influenza*. Epidemiologic Reviews. 1996;18(1):64-76.
198. Evans, C. and B. Crawford. *Expert judgement in pharmacoeconomic studies: guidance and future use*. Pharmacoeconomics. 2000;17(6):545-553.
199. Dittus, R.S., S.D. Roberts, and J.R. Wilson. *Quantifying uncertainty in medical decisions*. J Am Coll Cardiol. 1989;14(3 Suppl A):23A-28A.
200. Clark, D.E. *Computational methods for probabilistic decision trees*. Comput Biomed Res. 1997 Feb;30(1):19-33.
201. Cooper, N.J., et al. *Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach*. Health Econ. 2004 Mar;13(3):203-26.

202. Cox, E.R., B. Motheral, and D. Mager. *Verification of a decision analytic model assumption using real-world practice data: implications for the cost effectiveness of cyclo-oxygenase 2 inhibitors (COX-2s)*. AMERICAN JOURNAL OF MANAGED CARE. 2003;9(12):785-796.
203. Centre for Epidemiology, N.B.o.H.a.W., *Swedish Hospital Discharge Diagnosis Register, 1987-2006*, 2006: Stockholm.
204. World Health Organization. *International Classification of Diseases (ICD)*. 2009 [cited 2010-10-18]; Available from: http://www.who.int/classifications/icd/en/.
205. Gutwin, C. and S. Greenberg, editors. *Design for individuals, design for groups: tradeoffs between power and workspace awareness*; 1998. ACM New York, NY, USA.
206. Timpka, T., et al., *Towards a simulation environment for modeling of local influenza outbreaks*, 2005, American Medical Informatics Association.
207. Smieszek, T. *A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease spread*. Theoretical Biology and Medical Modelling. 2009;6(1):25.
208. Nordvik, M.K., et al. *Spatial bridges and the spread of Chlamydia: the case of a county in Sweden*. Sexually Transmitted Diseases. 2007;34(1):47.
209. Morris, M. *Network Epidemiology: A Handbook for Survey Design and Data Collection*. Oxford University Press; 2004.
210. Bondy, A. and U.S.R. Murty. *Graph Theory*. 3rd Corrected Printing ed.: Springer; 2008.
211. Holme, P., C.R. Edling, and F. Liljeros. *Structure and time evolution of an Internet dating community*. Social Networks. 2004;26(2):155-174.
212. Rybski, D., et al. *Scaling laws of human interaction activity*. Proc Natl Acad Sci U S A. 2009;106(31):12640.
213. Lu, X., et al. *The Sensitivity of Respondent-driven Sampling* Journal of the Royal Statistical Society. [Unpublished Work]. In press 2010.
214. Newman, M.E.J., A.L. Barabasi, and D.J. Watts. *The structure and dynamics of networks*. Princeton University Press; 2006.
215. Newman, M.E.J. *The structure and function of complex networks*. Arxiv preprint cond-mat/0303516. 2003.
216. Guimera, R. and L.A.N. Amaral. *Functional cartography of complex metabolic networks*. Nature. 2005;433(7028):895-900.
217. Guimera, R., et al. *The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles*. Proc Natl Acad Sci U S A. 2005;102(22):7794.
218. Erdős, P. and A. Rényi. *On random graphs*. Publicationes Mathematicae. 1959;6:290-297.
219. Erdős, P. and A. Rényi. *On the evolution of random graphs*. Magyar Tud. Akad. Mat. Kutató Int. Közl. 1960;5:17-61.
220. Erdős, P. and A. Rényi. *On the strength of connectedness of a random graph*. Acta Mathematica Hungarica. 1961;12(1):261-267.
221. Wikipedia. *Six degrees of Kevin Bacon*. [cited 2010-06-01]; Available from: http://en.wikipedia.org/wiki/Bacon_number.
222. Milgram, S. *The small world problem*. Psychology today. 1967;2(1):60-67.
223. Watts, D.J. and S.H. Strogatz. *Collective dynamics of 'small-world' networks*. Nature. 1998;393(6684):440-442.
224. Albert, R., H. Jeong, and A.L. Barabási. *The diameter of the world wide web*. Arxiv preprint. 1999;cond-mat/9907038.

225. Liljeros, F., et al. *The web of human sexual contacts*. Nature. 2001;411(6840):907-908.
226. Anderson, C. *The Long Tail: Why the Future of Business Is Selling Less of More*. Hyperion; 2006.
227. Callaway, D.S., et al. *Network robustness and fragility: Percolation on random graphs*. Arxiv preprint cond-mat/0007300. 2000.
228. Albert, R., J. Jeong, and A.L. Barabasi. *Error and attack tolerance of complex networks*. The Structure and Dynamics of Networks. 2006:503.
229. Cohen, R. and K. Erez. *Resilience of the internet to random breakdowns*. Physical Review Letters. 2000;85(21):4626.
230. Cohen, R. and K. Erez. *Breakdown of the internet under intentional attack*. Physical Review Letters. 2001;86(16):3682.
231. Chen, Y., et al. *Finding a better immunization strategy*. Physical Review Letters. 2008;101(5):58701.
232. Liljeros, F. *Closing workplaces*. [Unpublished Work].
233. Jonkers, A.R.T., K.J. Sharkey, and R.M. Christley. *Preventable H5N1 avian influenza epidemics in the British poultry industry network exhibit characteristic scales*. Journal of The Royal Society Interface.7(45):695.
234. Colgate, S.A., et al. *Risk Behavior-Based Model of the Cubic Growth of Acquired Immunodeficiency Syndrome in the United States*. Proceedings of the National Academy of Sciences. 1989;86(12):4793-4797.
235. Britton, T., M.K. Nordvik, and F. Liljeros. *Modelling sexually transmitted infections: the effect of partnership activity and number of partners on R0*. Theoretical Population Biology. 2007;72(3):389-399.
236. Herrmann, B. *A new genetic variant of Chlamydia trachomatis*. Sexually Transmitted Infections. 2007;83(4):253.
237. Nordvik, M.K. and F. Liljeros. *Number of Sexual Encounters Involving Intercourse and the Transmission of Sexually Transmitted Infections*. Sexually Transmitted Diseases. 2006;33(6):342.
238. Yorke, J.A., H.W. Hethcote, and A. Nold. *Dynamics and Control of the Transmission of Gonorrhea*. Sexually transmitted diseases. 1978;5(2):51.
239. Liljeros, F. *Human Sexual Networks*. In: R.A. Meyers, editor. Encyclopedia of Complexity and Systems Science: Springer Reference; 2009
240. Bergström, M., *Kådiskollen*, 2009, RFSU, Sweden.
241. Helleringer, S. and H.P. Kohler. *Sexual network structure and the spread of HIV in Africa: evidence from Likoma Island, Malawi*. Aids. 2007;21(17):2323.
242. Cohen, R., S. Havlin, and D. Ben-Avraham. *Efficient immunization strategies for computer networks and populations*. Physical Review Letters. 2003;91(24):247901.
243. Newman, M.E.J. *Ego-centered networks and the ripple effect*. Social Networks. 2003;25(1):83-95.
244. Nordberg, T. and J. Levén, *Nordisk sex- och kondomundersökning*, 2009.