Module 5:
Introduction to Stochastic Epidemic Models with Inference

Instructors:

Tom Britton,
Dept. Mathematics, Stockholm University

Ira Longini,
Dept. Biostatistics, University of Florida

Jonathan Sugimoto,
Vaccine and Infectious Dis Div, FHCRC

CENTER FOR INFERENCE & DYNAMICS OF INFECTIOUS DISEASES

UF UNIVERSITY of FLORIDA

csquid
CENTER FOR STATISTICS & QUANTITATIVE INFECTIOUS DISEASES

Stockholm University

Some Books of Interest for This Course
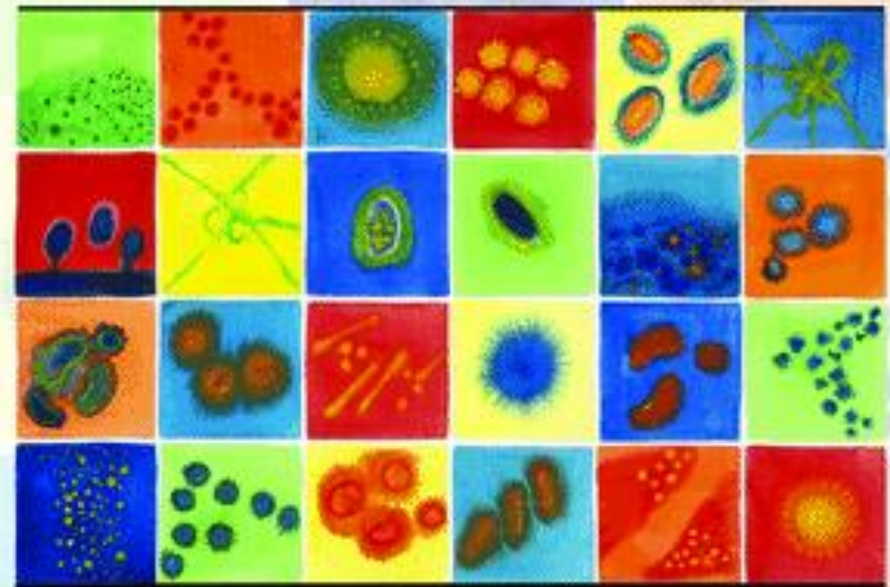
Lecture Notes in Statistics

151

Håkan Andersson   Tom Britton

Stochastic Epidemic Models and Their Statistical Analysis

Springer

Mathematical Tools for Understanding Infectious Disease Dynamics

PRINCETON SERIES IN THEORETICAL AND COMPUTATIONAL BIOLOGY

ODO DIEKMANN, HANS HEESTERBEEK, & TOM BRITTON

CENTER FOR INFERENCE & DYNAMICS OF INFECTIOUS DISEASES

UF UNIVERSITY of FLORIDA

csquid
CENTER FOR STATISTICS & QUANTITATIVE INFECTIOUS DISEASES

Stockholm University

Halloran • Longini
Struchiner

M. Elizabeth Halloran • Ira M. Longini, Jr. • Claudio J. Struchiner

Design and Analysis of Vaccine Studies

M. Elizabeth Halloran
Ira M. Longini, Jr.
Claudio J. Struchiner

Widespread immunization has many different kinds of effects in individuals and populations, including in the unvaccinated individuals. The challenge is in understanding and estimating all of these effects. This book presents a unified conceptual framework of the different effects of vaccination at the individual and at the population level. The book covers many different vaccine effects, including vaccine efficacy for susceptibility, for disease, for post-infection outcomes, and for infectiousness. The book includes methods for evaluating indirect, total and overall effects of vaccination programs in populations. Topics include household studies, evaluating correlates of immune protection, and applications of casual inference. Material on concepts of infectious disease epidemiology, transmission models, casual inference, and vaccines provides background for the reader. This is the first book to present vaccine evaluation in this comprehensive conceptual framework.

This book is intended for colleagues and students in statistics, biostatistics, epidemiology, and infectious diseases. Most essential concepts are described in simple language accessible to epidemiologists, followed by technical material accessible to statisticians.

Elizabeth Halloran and Ira Longini are professors of biostatistics at the University of Washington and the Fred Hutchinson Cancer Research Center in Seattle. Claudio Struchiner is professor of epidemiology and biostatistics at the Brazilian School of Public Health of the Oswaldo Cruz Foundation in Rio de Janeiro. The authors are prominent researchers in the area. Halloran and Struchiner developed the study designs for dependent happenings to delineate indirect, total, and overall effects. Halloran has made contributions at the interface of epidemiological methods, causal inference, and transmission dynamics. Longini works in the area of stochastic processes applied to epidemiological infectious disease problems, specializing in the mathematical and statistical theory of epidemics. Struchiner has contributed to understanding the role of transmission in interpreting vaccine effects.

EPIDEMIOLOGY

> springer.com

Design and Analysis of Vaccine Studies

Design and Analysis
of Vaccine Studies

🍁 Springer

CENTER FOR
INFERENCE & DYNAMICS
OF INFECTIOUS DISEASES

UF UNIVERSITY of FLORIDA

csquid
CENTER FOR STATISTICS & QUANTITATIVE INFECTIOUS DISEASES

Stockholm University

The Mathematical Theory of Infectious Diseases
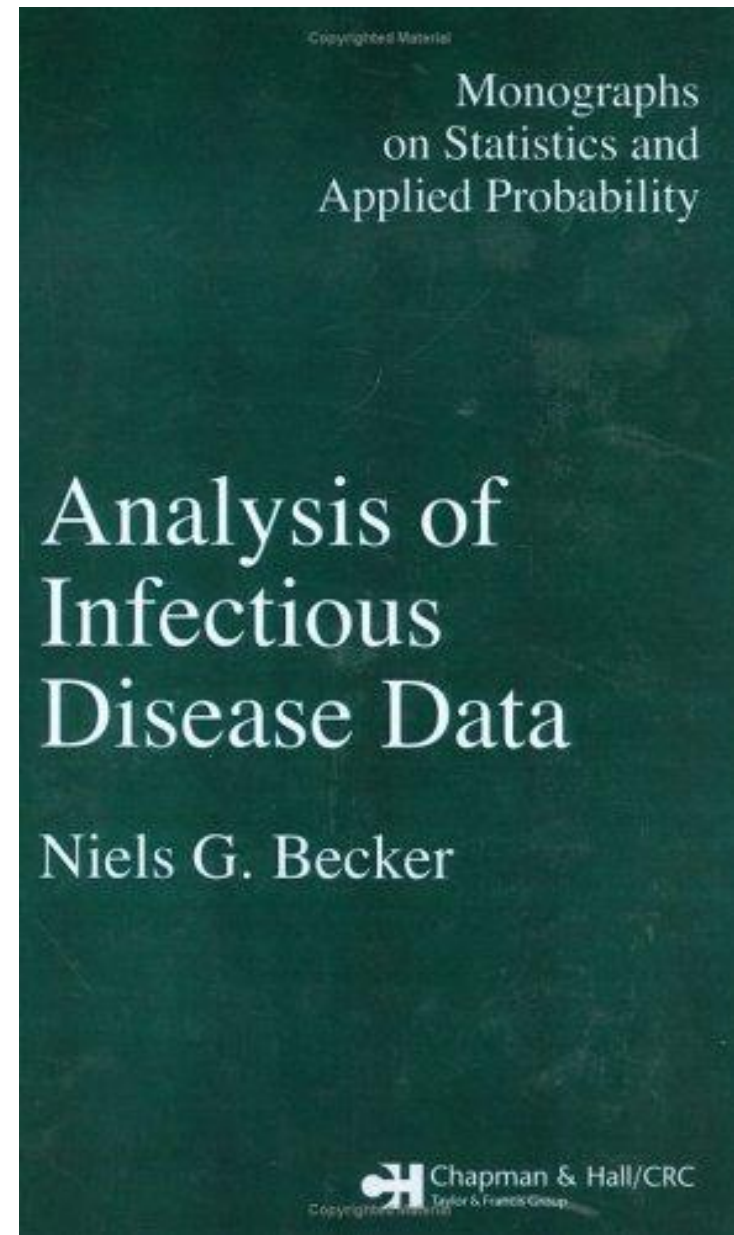## and its Applications

Norman T. J. Bailey, M.A., D.Sc.

Unit of Health Statistical Methodology, World Health Organization,
Geneva. Formerly Professor of Biomathematics, Cornell University
Graduate School of Medical Sciences, and Member of
the Sloan-Kettering Institute for Cancer Research

Second edition

Monographs
on Statistics and
Applied Probability

Analysis of Infectious Disease Data

Niels G. Becker

1975, 1ˢᵗ addition 1957

CENTER FOR
INFERENCE & DYNAMICS
OF INFECTIOUS DISEASES

UF UNIVERSITY *of* FLORIDA

csquid
CENTER FOR STATISTICS & QUANTITATIVE INFECTIOUS DISEASES

Stockholm
University

TranStat

Software for the inference on infectious disease data

TranStat

To analyze data from outbreaks of acute infectious disease, MIDAS scientists at the University of Washington and the Fred Hutchinson Cancer Research Center have developed TranStat, a tool for data entry, storage, and rapid analysis. It is being used to test for the presence of human-to human transmission (or animal-to-animal transmission in veterinary settings) and to estimate the epidemiological characteristics of the disease, such as secondary attack rates and the local reproductive number.

Purpose. The key to controlling a pandemic is early detection, containment, and mitigation. The TranStat tool was developed to enable field personnel and researchers to enter and revise data from local outbreaks.[1] From these data, TranStat provides a means of testing for the presence of human-to-human (or animal-to-animal) transmission. If this transmission is detected, estimates of the household-specific and neighborhood-specific secondary attack rates and local reproductive number are provided.

Data Input. TranStat uses information on the

- outbreak,
- population at risk,
- exposure events, and

Information on the population at risk includes sex; age; neighborhood and household of residence; dates of illness onset for cases; dates of hospitalization, if any; and dates of receiving treatments for cases or prophylaxis for noncases, if any.

Exposure details for each person include the neighborhoods and households visited and the dates of the visits.

The estimated distributions of the infectious or incubation periods are specified in terms of minimum and maximum number of days and probabilities for each particular duration (e.g., day, hour, week).

Method. A discrete-time maximum likelihood model is used to estimate the time-specific probabilities of transmission within and between households, from which

Inference on infectious diseases modules after this one

- Module 9: Simulation-based Inference for Epidemiological Dynamics, July 18 – 20

- Module 10: Statistics and Modeling with Novel Data Streams, July 18 – 20

- Module 11: MCMC II, Methods for Infectious Disease Studies, July 20 – 22

- Module 16: Spatial Statistics in Epidemiology and Public Health, July 25-27

CENTER FOR
INFERENCE & DYNAMICS
OF INFECTIOUS DISEASES

UF | UNIVERSITY of FLORIDA

csquid
CENTER FOR STATISTICS & QUANTITATIVE INFECTIOUS DISEASES

Stockholm University

Lectures

July 13:

1. Introduction to stochastic epidemic models; notation, properties, examples, IL, TB

2. Important properties: R_0, growth rate, generation intervals, etc., TB

July 14:

3. Inference on stochastic epidemic models, TB

4. Modeling using networks and other heterogeneities, TB

5. Different models for vaccine mechanisms, IL

6. Real time analysis of infectious disease outbreaks using TranStat, JS

Lectures

July 15:

7. Stochastic models for small groups such as households, IL

8. Inference for large groups such as cities, TB

9. Stochastic models for arboviruses, IL

10. Design and analysis of cluster randomized vaccine trials for emerging infectious disease epidemics: The case of ring vaccination for Ebola, IL

Some Infectious Diseases Under Study

- Influenza

- Novel Coronavirus, SARS-CoV, MERS-CoV

- Ebola

- Cholera

- Dengue

- Chikungunya

- Zika

- Others, polio, pertussis, hand-foot-and-mouth (EV71)

- Agent X

Some Infectious Diseases Under Study (Continued)
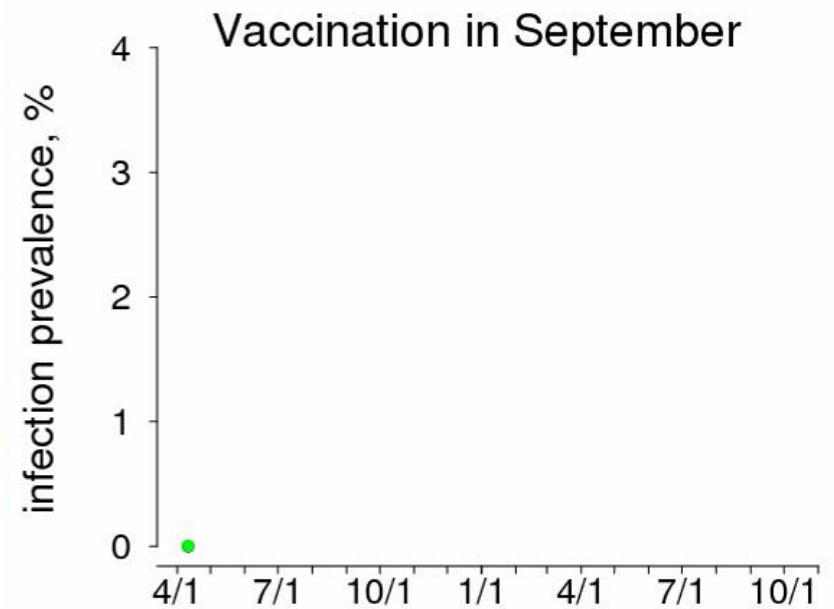
- Bioterrorist agents
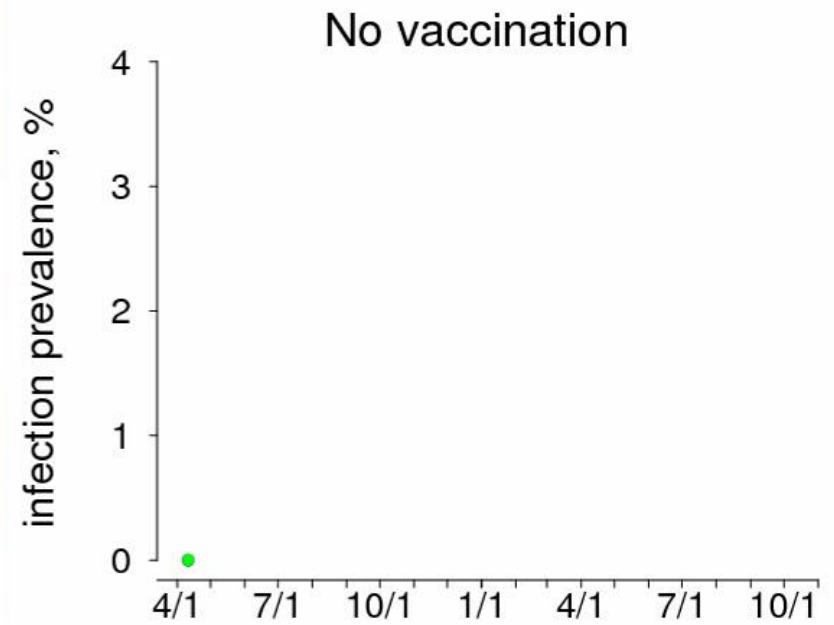 - Smallpox
  - Anthrax
  - Agent X

Some Examples

Pandemic H1N1, 2009-2010
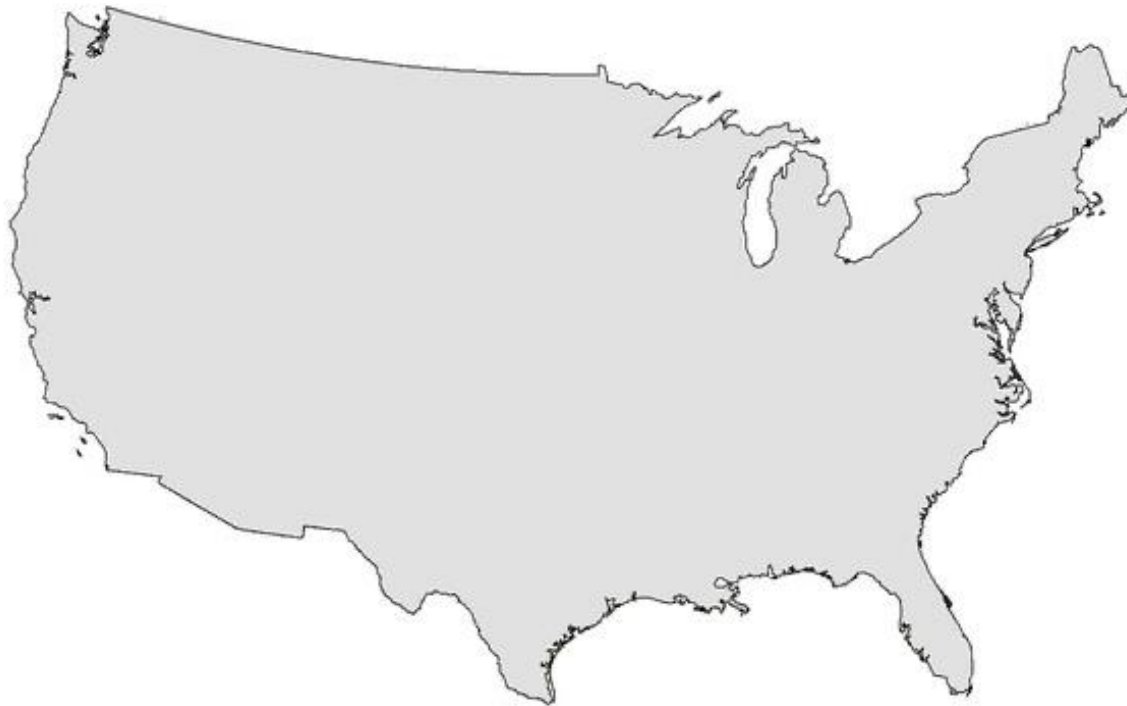
Stochastic, Compartmental, Patch
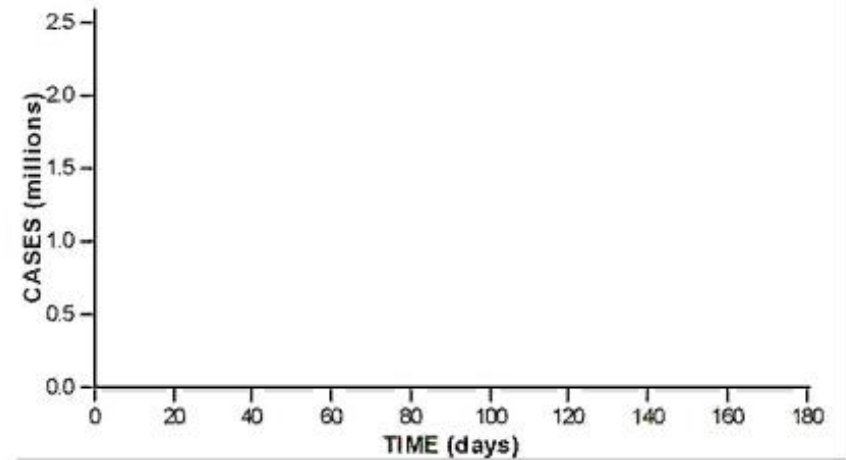
Pandemic H1N1 With Vaccination

Spread the within US

Agent-based, RF
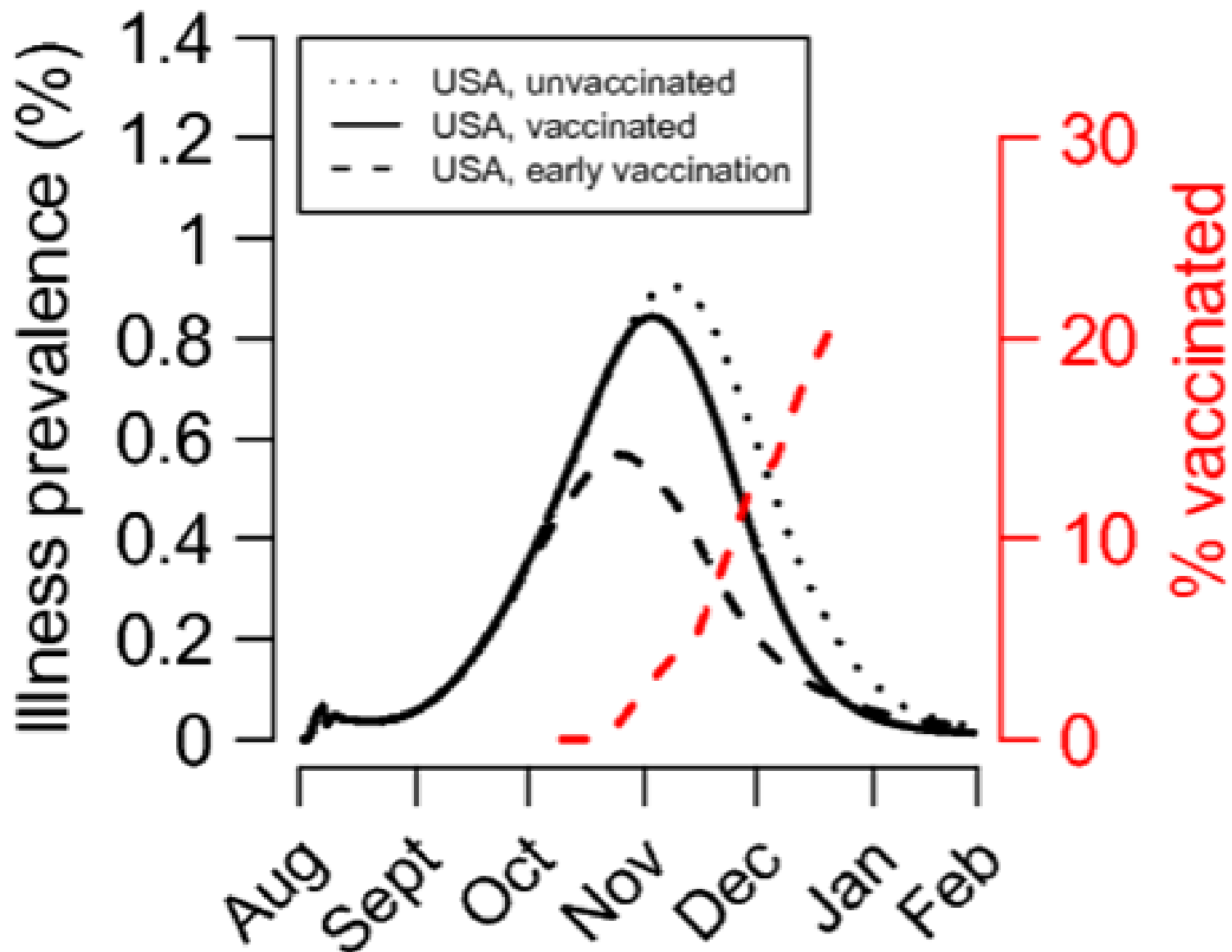

Day 001

Intervention(s): None

Illness Prevalence (%): 0.01 - 0.50   0.51 - 1.00   1.01 - 1.50   1.51+
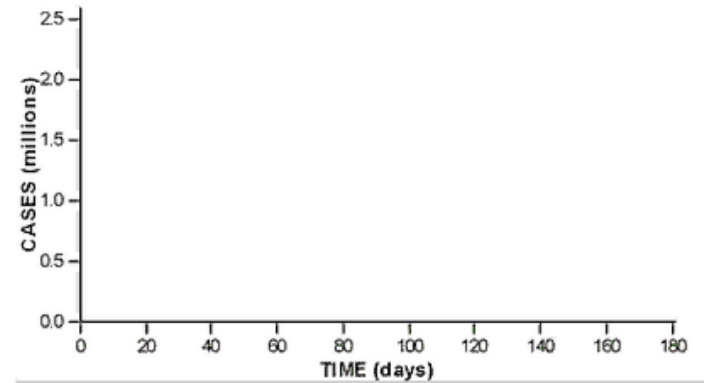
Simulated Mass Vaccination in US

What Could Have Been Done
Agent-based, RF

Day 001

Intervention(s): None

Illness Prevalence (%): 0.01 - 0.50   0.51 - 1.00   1.01 - 1.50   1.51+

Intervention(s): Vaccination
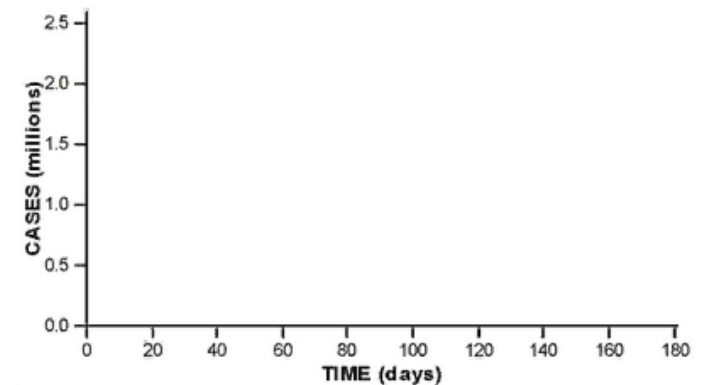
Dengue Vaccines in Thailand

Agent-based, RF
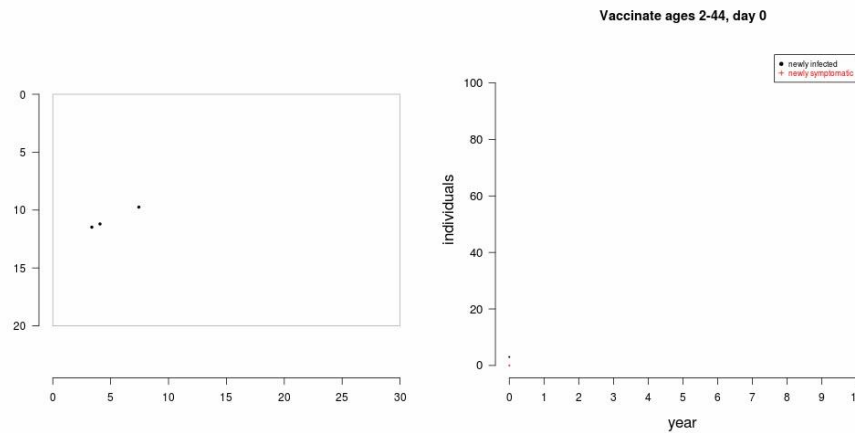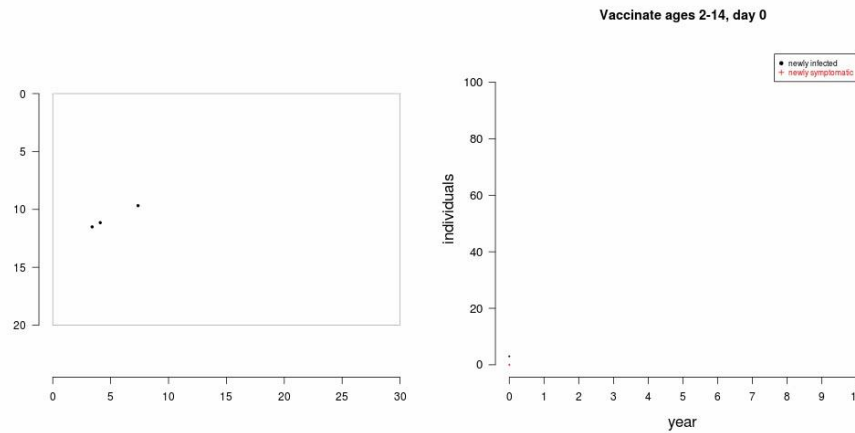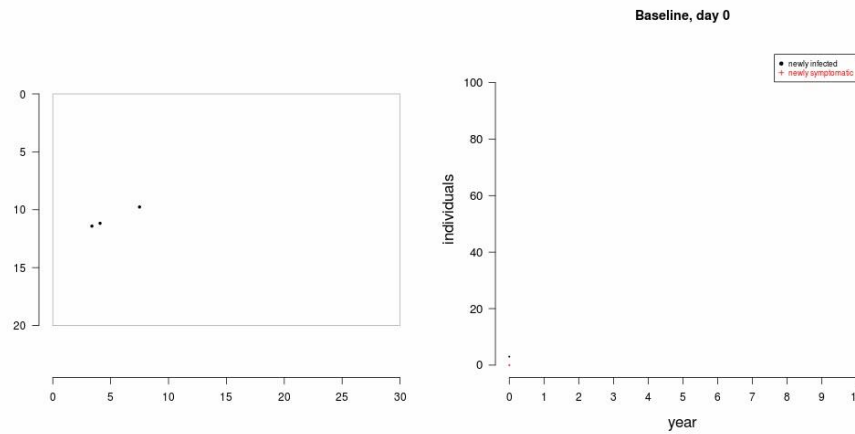
Ebola vaccine trials

- Phase III Ebola ring vaccination efficacy trial in Guinea – VSV vaccine estimated to have 100% efficacy.

- Stochastic transmission models have been used to help estimate vaccine efficacy and effectiveness

- Ring vaccination is used to eliminate Ebola in human populations, e.g., smallpox eradication

On with the course!

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

L1, Introduction to stochastic epidemic models

Tom Britton

July, 2016

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Stochastic models

Mathematical models describes some feature in a *simplified way*

The discrepancy between model and reality may be contained in "random part" in model

Very important in small populations

Stochastic models enable uncertainty estimates (i.e. standard errors) when estimating parameters

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Background

We want to model the spread of transmittable disease in a community of individuals

At a given time-point an individual may be *Susceptible*, infected but not yet infectious (*Latent* or *Exposed*), *Infectious*, or recovered and immune (*Removed*)

Different class of epidemic models: SIR, SEIR, SIS, SIRS, ...

Main focus: SIR (childhood diseases, STDs, influenza, ...)

Short term outbreak vs endemic situation

Simplification for short term: fixed population, no waning immunity

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Notation

Some notation to be used

- $n = \#$ individuals ($n(t)$ if varying over time)
- $S(t) = \#$ "susceptibles" (susceptible individuals) at time t
- $I(t) = \#$ "infectives" (infectious individuals) at time t
- $R(t) = \#$ "removeds" (removed individuals) at time t
- $T =$ the time when the epidemic stops
- Z ($= R(T) - 1) = \#$ infected during the epidemic (excluding index case). Possible values: $0, 1, ..., n-1$.

We start with the simplest situation: all individuals are "identical" (with respect to disease spreading) and all pairs of individuals have contact at equal rates.

Homogeneous community that mixes uniformly

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

The Reed-Frost stochastic epidemic model

Short term outbreak (fixed community), homogeneous community, uniform mixing

An epidemic model (Reed-Frost, 1928)

- Assume 1 index case (externally infected) the rest $n - 1$ susceptible
- Anyone who gets infected infects other susceptibles independently with prob p and then recovers
- A recovered individual plays no further role in epidemic

The index case infects a random number $(\text{Bin}(n - 1, p))$ of individuals, they in turn infect an additional random number, and so on. Once no new individuals are infected the epidemic stops

Think in "generations"

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Exercise 1

Suppose $n = 3$ (one index case and 2 susceptibles) and $p = 0.2$

Possible values for Z: 0,1,2.

$P(Z = 0)$? For this to happen the index can't infect anyone

$P(Z = 1)$? For this to happen the index must infect EXACTLY one AND this individual cannot infect anyone further

$P(Z = 2)$? Either the index infects exactly one AND this individual infects the last one, OR the index infects both

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Exercise 1

Suppose $n = 3$ (one index case and 2 susceptibles) and $p = 0.2$

Possible values for Z: 0,1,2.

$P(Z = 0)$? For this to happen the index can't infect anyone

$P(Z = 1)$? For this to happen the index must infect EXACTLY one AND this individual cannot infect anyone further

$P(Z = 2)$? Either the index infects exactly one AND this individual infects the last one, OR the index infects both

$P(Z = 0) = (1 - p)^2 = 0.64$
$P(Z = 1) = \binom{2}{1} p(1 - p) \times (1 - p) = 0.256$
$P(Z = 2) = \binom{2}{1} p(1 - p) \times p + p^2 = 0.104$
or ... $P(Z = 2) = 1 - P(Z = 0) - P(Z = 1)$

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

What about larger communities?

General n, think in "generations"

Epidemic chains: $i \to 3 \to 2 \to 0$: the index infects 3, they infect 2 and these infect no further and the epidemic stops

$$P(Z = 0) = P(i \to 0) = (1 - p)^{n-1}$$
$$P(Z = 1) = P(i \to 1 \to 0) = \binom{n-1}{1} p^1 (1 - p)^{n-2} \times (1 - p)^{n-2}$$
$$P(Z = 2) = P(i \to 2 \to 0) + P(i \to 1 \to 1 \to 0) = ...$$
$$P(Z = 3) = P(i \to 3 \to 0) + P(i \to 2 \to 1 \to 0) + P(i \to 1 \to 2 \to 0) + P(i \to 1 \to 1 \to 1 \to 0) = ...$$

$P_n(Z = z)$ gets very complicated when $n \geq 10$ and $z \geq 5$.

What to do then?

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Approximations when n large

When n large then often p (=per individual transmission probability) is small.

Expected number of infectious contacts: $(n-1)p \approx np =: R_0$

R_0 = basic reproduction number

Next page: Histogram of final outbreak sizes from 10 000 simulations in a community of $n = 1000$ individuals (both $R_0 < 1$ and $R_0 > 1$)

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Histogram of final size: $R_0 = 0.8$

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Histogram of final size: $R_0 = 1.5$

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet
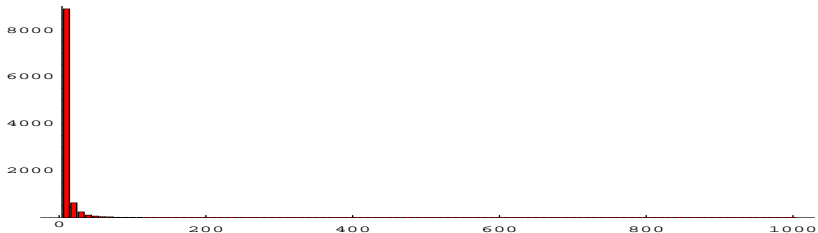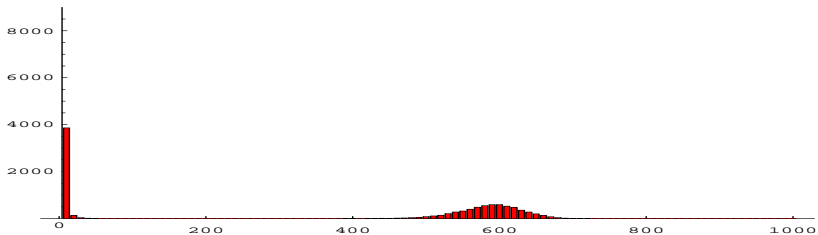
An approximation for the final size

$R_0 = 1$ is "threshold value"

We now derive an equation for τ heuristically

Assume n large and let $\tau = Z/n =$ final *fraction* infected

$$
\begin{align}
1 - \tau &= \text{proportion not infected} \tag{1}\\
&\approx \text{ probability not get infected} \tag{2}\\
&= \text{prob to escape inf from all infected} \tag{3}\\
&= (1 - p)^Z \tag{4}\\
&= \left(1 - \frac{R_0}{n}\right)^{n\tau} \tag{5}\\
&\approx e^{-R_0\tau} \qquad \text{(using that } (1 - x/n)^n \approx e^{-x}) \tag{6}
\end{align}
$$

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Approximation for final size

τ should hence (approximately) solve

$$1 - \tau = e^{-R_0 \tau}$$
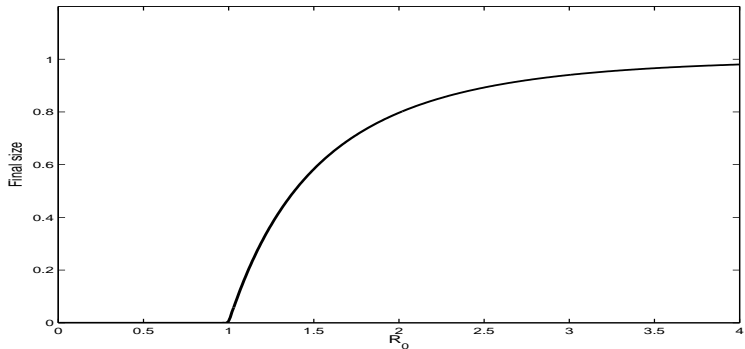
There are two solutions: $\tau = 0$ and (if $R_0 > 1$): $\tau = \tau^\star > 0$.

Exercise 2 Compute τ^* numerically when $R_0 = 1.5$, 3 and 6.

On next page is a plot of final size as function of R_0

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Plot of final outbreak size as function of R_0

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Approximation, cont'd

Seen from simulations: strong dichotomy: minor outbreak – major outbreak

P(major outbreak) $= 1 - P$(minor outbreak) can be determined using *branching process* theory

For Reed-Frost model: $P($ major outbreak$) = \tau^*$!!!

Normal distribution for major outbreak:

$$\sqrt{n}\left(\frac{Z}{n} - \tau^*\right) \approx Normal(0, \sigma^2)$$

σ^2 depends on model parameters

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

What about epidemic over time?

A related stochastic epidemic model (the "General stochastic epidemic") can be defined in continuous time:

- During the infectious period an individual has "infectious contacts" randomly in time at the average rate β, each time individual is chosen randomly

- A susceptible who receives an infectious contact becomes infectious and remains so for a exponentially distributed time with mean ν (other contacts have no effect)

$R_0 =$ expected number of infectious contacts $= \beta\nu$

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

What about epidemic over time?

When n is large the process $(S(t)/n, I(t)/n)$ is close to
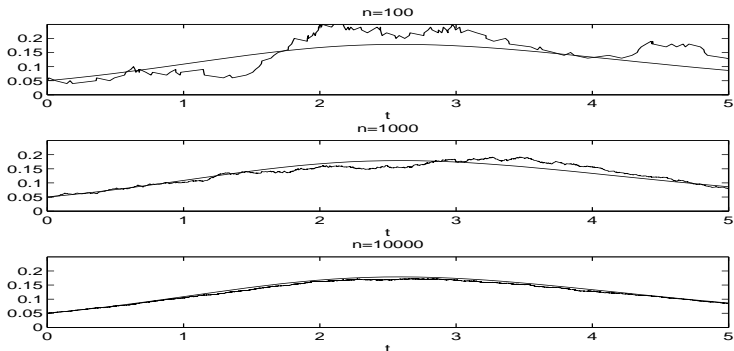deterministic limit $(s(t), i(t))$ which solves differential system

$$s'(t) = -\beta s(t)i(t) \tag{7}$$

$$i'(t) = \beta s(t)i(t) - \frac{1}{\nu}i(t) \tag{8}$$

$$r'(t) = \frac{1}{\nu}i(t) \tag{9}$$

Next page: plot of $I(t)/n$ for one (typical) simulated epidemic and
deterministic limit $i(t)$, for a few different n

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Plots of simulated stochastic epidemic and deterministic curve

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Summary

Exact distribution of Z (final size) is complicated

When n large two things may happen:

- either very few get infected ($Z/n \approx 0$), or
- a close to deterministic fraction $Z/n \approx \tau^*$ get infected

$R_0 = np = \beta\nu = $ expected number of infections by one individual during his/her infectious period

Second scenario only possible if $R_0 > 1$

$P($ major outbreak$) = \tau^*$ for Reed-Frost model

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Extensions

Random infectious force (e.g. length of infectious period): affects
P(outbreak) but hardly final size τ

Latent period: big effect on timing of epidemic peak and duration
of epidemic but no effect on final size (unless control measures are
initiated)

More than one index case: big effect on P(outbreak) but negligible
effect on final size τ in large outbreak

Exercise 3. If infectious period deterministic then P(major
outbreak)$= \tau^*$. If infectious period is exponentially distributed then
$P($ major outbreak)$= 1 - 1/R_0$. Compute the latter probability for
$R_0 = 1.5$, 3 and 6 and compare with Reed-Frost model.

Why "stochastic"?
Background and notation
A simple stochastic epidemic model
Some extensions

Stockholms
universitet

Extensions

Initial fraction of immunes. If there is a fraction r of initially immunes the same methodology can be used. The difference is that R_0 is replaced by $R_0(1 - r)$ since initially only the fraction $(1 - r)$ is susceptible. The final fraction infected *among the initally susceptible* then solves

$$1 - \tau = e^{-R_0(1-r)\tau}$$

Major outbreak possible only if $R_0(1 - r) > 1$

Exercise 4. Compute τ^* if initially only 50% were susceptible (and 50% were immune), for $R_0 = 1.5$, 3 and 6.

Exercise 5. What are the *overall* fractions infected during outbreak in later case?