Stockholms universitet

L8, Estimation uncertainty + Herd immunity

Tom Britton

July, 2020

Stockholms universitet

Repetition: Inference from large outbreaks

From lecture 3: basic reproduction number R_0 and critical vaccination coverage v_c were estimated by:

$$\hat{R}_0 = -\ln(1-\tilde{\tau})/\tilde{\tau}$$

$$\hat{v}_c = 1 - \frac{\tilde{\tau}}{-\ln(1-\tilde{\tau})}$$

if outbreak takes place in a fully susceptible homogeneous community resulting in a fraction $\tilde{\tau}$ getting infected during the outbreak

How about uncertainty?

Uncertainty of previous estimate

Intuition: The larger community (and more getting infected) the less uncertainty

It was mentioned that final number infected $n\tilde{\tau} = Z$ in case of a major outbreak is normally distributed with mean $n\tau^*$ and standard deviation $\sqrt{n\sigma^2}$ where σ^2 depends on model parameters and shown two slides ahead

This result can be used to show that $\hat{R}_0$ and $\hat{v}_c$ are normally distributed with correct means (i.e. R_0 and v_c respectively) and standard errors to be derived using δ-method

Stockholms universitet

The δ-method

Suppose random variable X has mean $\mu = E(X)$ and variance $V(X)$

Then the δ-method gives the following approximation for the mean and variance of $f(X)$, where $f(x)$ is a "nice function":

$$E(f(X)) \approx f(\mu) \qquad V(f(X)) \approx (f'(\mu))^2 V(X)$$

The approximation holds better the smaller variance X has (i.e. smaller $V(X)$)

The δ-method for $V(\hat{R}_0)$

Probabilists have proven that the asymptotic variance of $\tilde{\tau}$ equals:

$$V(\tilde{\tau}) \approx \frac{1}{n} \frac{\tau(1-\tau)}{(1-(1-\tau)R_0)^2} \left(1 + c_v^2(1-\tau)R_0^2\right)$$

where τ and R_0 are the true parameter values related by $R_0 = -\ln(1-\tau)/\tau$, and c_v is the coefficient of variation of the infectious period.

We now apply the δ-method on $\hat{R}_0 = -\ln(1-\tilde{\tau})/\tilde{\tau}$, we hence have the function $f(x) = -\ln(1-x)/x$

After some algebra we get $V(\hat{R}_0) \approx \frac{1}{n\tau(1-\tau)} \left(1 + c_v^2(1-\tau)R_0^2\right)$

For a standard error estimate we take square roots and replace unknown quantities with there estimates/observed values. The result, also for $\hat{v}_c$, is given by:

Uncertainty of previous estimate

$$s.e.(\hat{R}_0) = \sqrt{\frac{1 + c_v^2(1-\tilde{\tau})\hat{R}_0^2}{\tilde{\tau}(1-\tilde{\tau})}/n}$$

$$s.e.(\hat{v}_c) = \sqrt{\frac{1 + c_v^2(1-\tilde{\tau})\hat{R}_0^2}{\hat{R}_0^4\tilde{\tau}(1-\tilde{\tau})}/n}$$

$c_v^2 = V(I)/(E(I))^2$= squared coefficient of variation of infectious period of individuals (variance divided by the squared mean)

Larger n gives smaller standard deviation (as expected)!

Uncertainty of previous estimate

c_v^2 cannot be estimated from final outbreak size – possibly known from before

If not one has to insert a "conservative" bound. E.g. $c_v^2 = 1$: very rarely is standard deviation larger than mean

Exercise 25 Suppose that 239 out of 651 individuals in an isolated village were infected during an outbreak. Estimate R_0 and v_c and give 95% confidence interval for the estimates. Consider both the case when all individuals have the same length of infectious period (so no variation) and the case where its standard deviation is equal to the mean.

Exercise 26 Do the same thing assuming 2390 out of 6510 got infected.

Stockholms universitet

More detailed data

Suppose that disease incidence is observed during outbreak – not only final number

Intuition: more detailed data should improve estimation

Answer: yes, in a couple of ways:

- estimate of R_0 and v_c becomes more complicated, but standard errors are (moderately) smaller
- enables estimation of more parameters: exponential growth rate ρ, latent and infectious period distributions, ...
- possible to detect deviations from model: changing behavior, non-homogeneity, ...

If also information about contacts are available: "transmission probability upon contact" can be estimated

Multitype epidemics

Suppose final size of a multitype epidemic observed: $\tilde{\tau}_1, \ldots, \tilde{\tau}_k$, $\tilde{\tau}_i =$ observed proportion infected among i-types

Also assumed that community fractions $\pi_1, \ldots, \pi_k$ known.

We want to estimate R_0 which is largest eigenvalue of next generation matrix M

First estimate M. Impossible!! Data has dimension k and M has dimension k^2.

$\Longrightarrow$ M and R_0 cannot be estimated consistently!

Multitype epidemics, cont'd

Why? We can observe who was infected but not who "caused" the infections

Susceptibility easier to estimate than infectivity!

$\Longrightarrow$ only possible to obtain bounds on R_0: lower bound assuming all infections caused by least infected type – upper bound assuming all infections caused by most infected type

Inference in networks

Inference can be performed without an outbreak: estimation of network properties: $E(D)$, $V(D)$, clustering c, ...

R_0, potential outbreak size τ and v_c can then be estimated as a function of transmission probability p

Typical conclusion: Outbreaks are only possible for a disease having higher transmission probability than $p = 0.13$

Or: An STD with $p = 0.08$ can only become endemic in core-groups with average number of partners higher than $E(D) = 4.2$ per year

Inference in more complicated models

More complicated model $\Longrightarrow$ harder inference and more detailed data need

Inference of spread of infections extra hard:

- There are strong dependencies because infections are not independent events (likelihood complicated)
- Many things unobserved: infectious contacts, latent period, infectious period, ...

Inference with more detailed data gives higher precision

Illustration

Suppose an infected infects each susceptible independently with prob p

Data = epidemic chain: $1 \to 2 \to 2 \to 0$

Initially 1 index and 9 susceptible

Likelihood: $L(p) =$
$\binom{9}{2} p^2 (1-p)^7 \cdot \binom{7}{2} \left(1-(1-p)^2\right)^2 \left((1-p)^2\right)^5 \cdot \binom{5}{0} \left((1-p)^2\right)^5$

Maximum-likelihood (ML) estimate $\hat{p}$ maximizes $L(\cdot)$:
$\Longrightarrow$ quite easy for a computer

If we instead only know that 5 out of 10 were infected likelihood is much more complicated (a sum over all possible chains)

Stockholms universitet

Alternative approach for complicated models

Basic idea: If likelihood complicated for available data we can "pretend" as if we had more detailed data, estimate parameters under this assumption, recompute some likely more detailed data, re-estimate parameters, ...

This is underlying idea in both EM-algorithm and recently very popular *MCMC*

MCMC: here parameters are treated as outcomes of random variables (Bayesian framework) and even very complicated likelihoods (posterior probabilities) can be evaluated numerically with arbitrary high precision

MCMC: Very computer intensive. Treated specifically in other Modules

Preventive measures: homogeneous case – initial phase

Common way of expressing R_0 (Anderson & May, 1991):

$$R_0 = p * k * \ell$$

p is probability of transmission given a "contact" by an infective

k is the rate of "contacts" per unit of time

ℓ is average duration of infectious period

Suppose **preventive measures** (put in place very early) reduce $p * k * \ell$ by a factor c ($c(t)$ if time-varying)

$\Longrightarrow$ new *effective* reproduction number equals $R_E^{(Hom)} = (1 - c)R_0$

No outbreak possible if $R_E^{(Hom)} \leq 1$ which is equivalent to $c \geq 1 - 1/R_0$

Stockholms universitet

Preventive measures and immunity: homogeneous case

If $R_E^{(Hom)} \geq 1$ the epidemic grows and immunity builds up: only infectious contacts with not yet infected individuals result in infection:

$R_E^{(Hom)}(t) = R_0(1-c)s(t)$, where $s(t)$ is *fraction* susceptible

If initially $R_E^{(Hom)} = R_E^{(Hom)}(0) > 1$ then $R_E^{(Hom)}(t)$ decays and for t large enough $R_E^{(Hom)}(t) < 1$ (because $s(t)$ becomes small) and then the epidemic starts declining

Currently $R_E(t) < 1$ in all (?) countries of Europe but not US

Terminology: some use "effective reproductiion number" for R_E and others for $R_E(t)$ (i.e. also including immunity). $R_E(t)$ also denoted "current" or "daily" reproduction number

Preventive measures and immunity: heterogeneous case

Let $\mathcal{I}(t)$ represent the composition of individuals that get infected around time t

The effective reproduction number at t (assuming all types of individual reduce spreading by the same fraction c) is then given by

$$R_E^{(Het)}(t) = R_0^{(\mathcal{I}(t))}(1-c)s_{\mathcal{I}(t)}(t)$$

$R_0^{(\mathcal{I}(t))}$ is the average number of infectious contacts (before prevention) that individuals getting infected around t have

$s_{\mathcal{I}(t)}(t)$ denotes fraction still susceptible among individuals contacted by the $\mathcal{I}(t)$-individuals

Preventive measures and immunity: heterogeneous case

Crude (?) approximation: $s_{\mathcal{I}(t)}(t) \approx s(t)$ (more true with varying social activity – less with assortative mixing)

$$\implies R_E^{(Het)}(t) \approx R_0^{(\mathcal{I}(t))}(1-c)s(t)$$

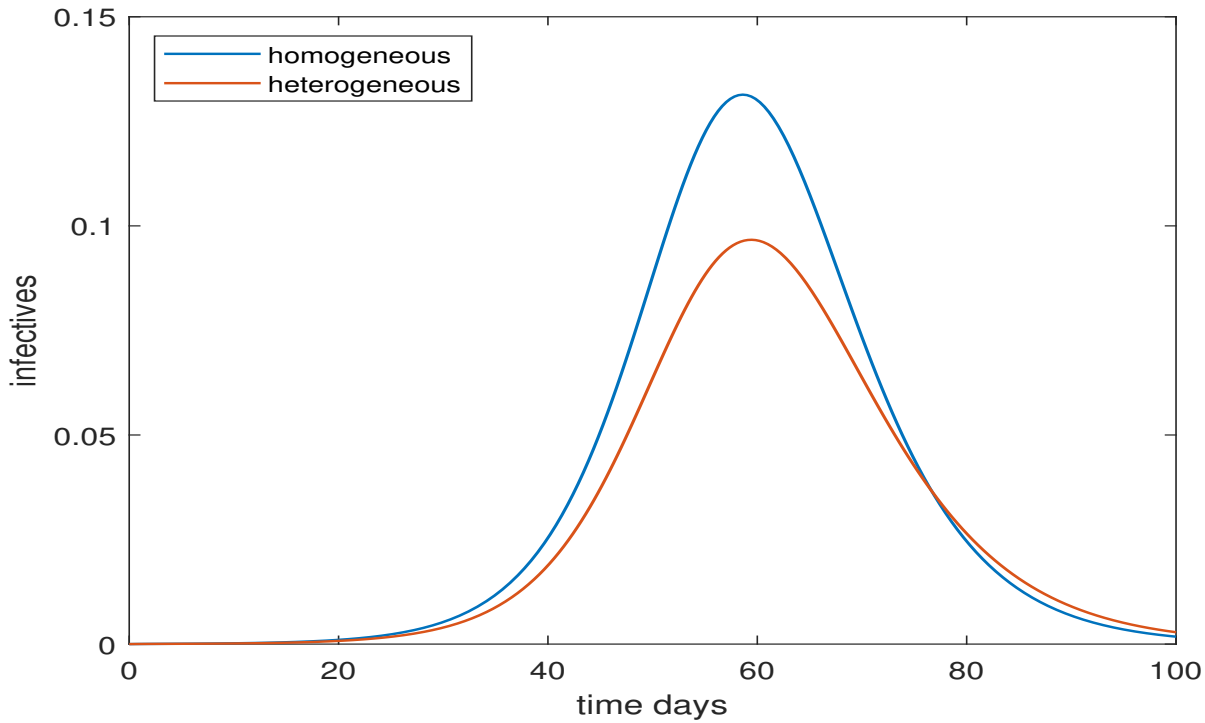
However: nearly **always** is $R_0 = R_0^{(\mathcal{I}(0))} > R_0^{(\mathcal{I}(t))}$!

Reason: Socially active individuals get infected early, later infected have fewer social contacts

$\implies$ fewer will get infected in heterogeneous case (true also without preventive measures!)

Stockholms universitet

Incidence over time: homogeneous vs heterogeneous



Figur: Incidence over time for a homogeneous model and heterogeneous with age and activity structure. Both have $R_0 = 2.5$ and same $g(s)$.

Also smaller final size: 72% vs 89%

Stockholms universitet

Herd immunity

When $R_E(t) < 1$ the epidemic declines and dies out, so those not yet infected are (soon) protected
(for example $R_E(t) = R_E^{(Het)}(t) = R_0^{(\mathcal{I}(t))}(1-c)s_{\mathcal{I}(t)}(t)) < 1$)

$\Longrightarrow$ given effect of current preventive measures c and given epidemic up until now, there is **sufficient immunity** for epidemic to die out and hence (soon) protecting susceptibles

Herd immunity: refers to the situation without preventive measures: are we safe if we go back to "normality" by setting $c = 0$?

Related question: How much back towards normality can we go (how much can c be reduced) and still have $R_E(t) \leq 1$?

Stockholms universitet

Classical Herd immunity (for vaccination)

Classical question: What fraction h needs to be immunized (by means of vaccination) beforehand, in order to avoid an outbreak without any preventive measures?

Answer when vaccinating uniformly (Anderson & May 1980's, or earlier?): No outbreak if $R_E = R_0(1-h) < 1$. Equivalent to $h \geq h_C = 1 - 1/R_0$ (true for very wide class of epidemic models)

Critical vaccination coverage: $h_C = 1 - 1/R_0$

Answer when vaccinating "optimally": fewer needs to be vaccinated (what fraction depends on model)

Illustration (Pastor-Satorras & Vespignani, 2001): For a scale free social network $h_C \approx 100\%$ when vaccinating uniformly but $h_C < 1\%$ if vaccinating optimally

Disease-induced Herd immunity

For details, see Britton et al, 2020, *Science*

Relevant Herd immunity question for Covid-19 (first time ever!): How many must have been infected during a mitigated epidemic outbreak in order to avoid a second epidemic outbreak once all preventive measures are lifted ($c = 0$)?

Scientific scenario: Consider the Covid-19 outbreak in a country with mitigation/lockdown and gradual exit towards normality

Scientific question: When will herd immunity be reached (for no restrictions, $c = 0$) assuming R_0 is known (e.g. $R_0 = 2.5$)?

Stockholms universitet

Disease-induced Herd immunity

For details, see Britton et al, 2020, *Science*

Relevant Herd immunity question for Covid-19 (first time ever!): How many must have been infected during a mitigated epidemic outbreak in order to avoid a second epidemic outbreak once all preventive measures are lifted ($c = 0$)?

Scientific scenario: Consider the Covid-19 outbreak in a country with mitigation/lockdown and gradual exit towards normality

Scientific question: When will herd immunity be reached (for no restrictions, $c = 0$) assuming R_0 is known (e.g. $R_0 = 2.5$)?

Answer: When fraction infected equals $h_C = 1 - 1/R_0 = 60\%$?

Stockholms universitet

Main result: Disease-induced herd immunity is lower!

This answer is correct if immunization is uniformly distributed in community (as in vaccination)

But this is **NOT** correct when immunity is achieved from disease spreading

Main result: Disease-induced herd immunity is lower!

This answer is correct if immunization is uniformly distributed in community (as in vaccination)

But this is **NOT** correct when immunity is achieved from disease spreading

Correct answer: Disease-induced herd immunity h_D will occur at a **substantially lower level**, perhaps around 40-45% if $R_0 = 2.5$

Gabriela Gomes et al (2020) indpendently show similar result for a different model: h_D could be as low as 10-20%

Heuristic explanation

In vaccination programs vaccinees are selected "randomly", so immunity is distributed uniformly in the community

But during a disease outbreak immunization is not distributed uniformly – highly active/social individuals are more likely to be infected

$\Longrightarrow$ Immunity is more "efficiently" distributed (still not optimal – cf "optimal vaccination policies")

Earlier knowledge: Well-known that after an outbreak immunity is more efficiently distributed (Diets & Schensle, Anderson & May, Bansal et al, Ferrari et al, ...)

But: No one seem to have realized that this is now "useful" when mitigation/suppression reduces spreading to lower levels ...

Model for illustration

Deterministic multitype model

- 6 age groups and mixing according to Wallinga et al 2006
- Individuals of each age group are divided into 3 "activity levels"
- 50% *Normal* activity, 25% have *Low* (half) activity and 25 % have *High* (double) activity
- Mimics network characterstics a bit

R_0 = dominant eigenvalue of next generation matrix

Final size equations exist

Stockholms universitet

Prevention

Prevention/restrictions: Suppose all mixing rates are reduced by a factor $c = 1 - \alpha$, so $R_E = \alpha R_0$

So if $\alpha < 1/R_0$ epidemic stops

Situation 1: Restrictions from start to end ($\rightarrow$ final size equations)

Our question: What is the smallest α that gives herd immunity after the outbreak is over? What is the overall disease-induced immunity level h_D for this α_*?

By this is meant: Suppose the outbreak with prevention α takes place. Then preventions are lifted. Is the population at risk for a second wave?

Herd immunity levels

Tabell: Disease-induced herd immunity level h_D and classical herd immunity level $h_C = 1 - 1/R_0$ for different population structures, for $R_0 = 2.0$, 2.5 and 3.0. Numbers correspond to percentages.

Population structure	$R_0 = 2.0$		$R_0 = 2.5$		$R_0 = 3.0$	
	h_D	h_C	h_D	h_C	h_D	h_C
Homogeneous	50.0	50.0	60.0	60.0	66.7	66.7
Age structure	46.0	50.0	55.8	60.0	62.5	66.7
Activity structure	37.7	50.0	46.3	60.0	52.5	66.7
Age & Activity structure	34.6	50.0	43.0	60.0	49.1	66.7

Stockholms universitet

Prevention: Situation 2 – restrictions and exit during outbreak

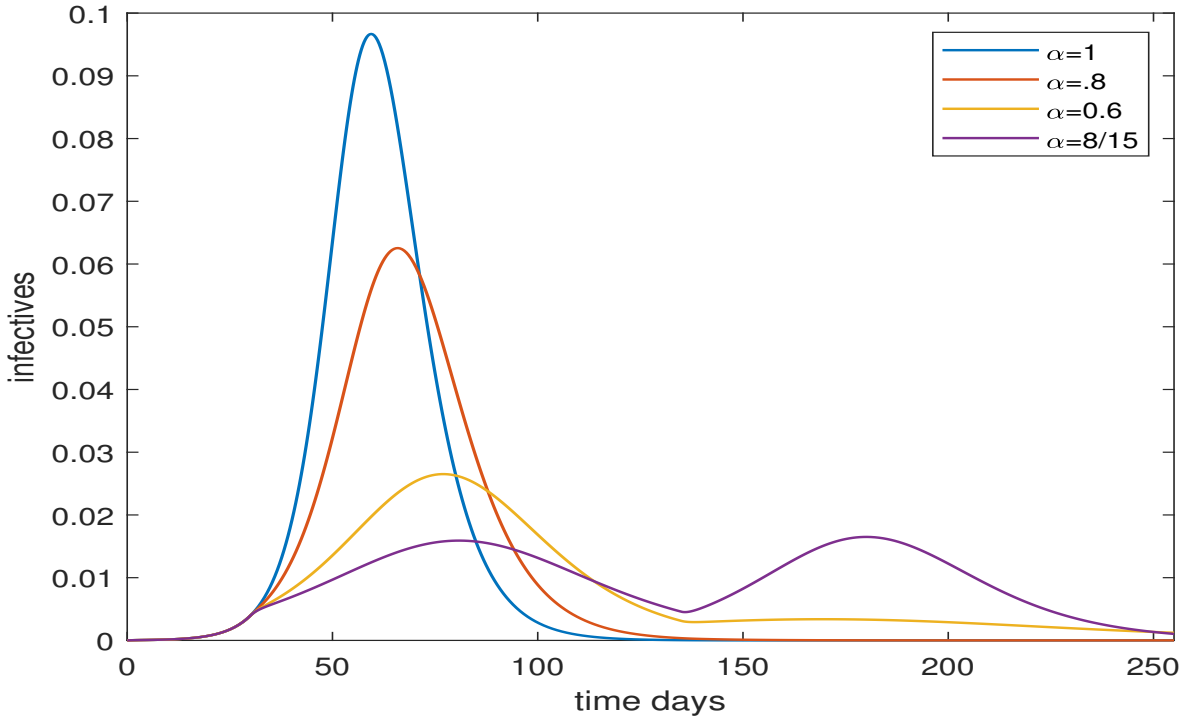
$\Longrightarrow$ We need to model time evolution of epidemic

Model: Deterministic SEIR

- February 15: start
- March 15: Restrictions put in place (4 different α)
- June 30: All restrictions lifted

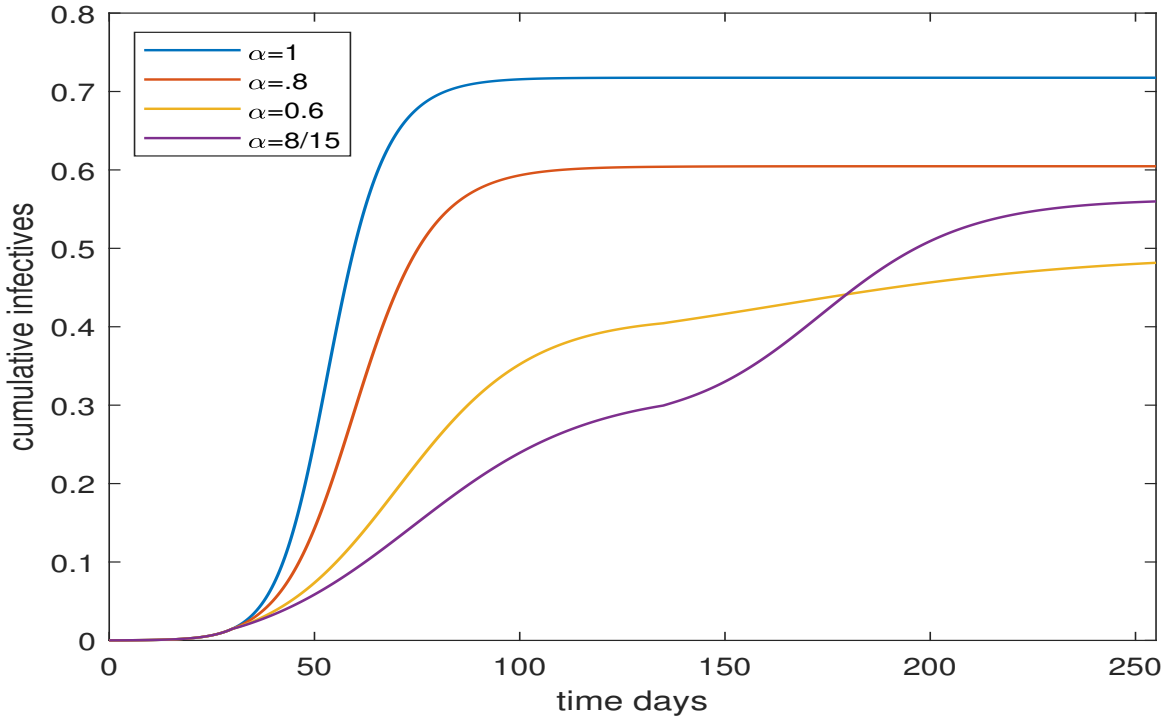
Stockholms universitet

Incidence over time



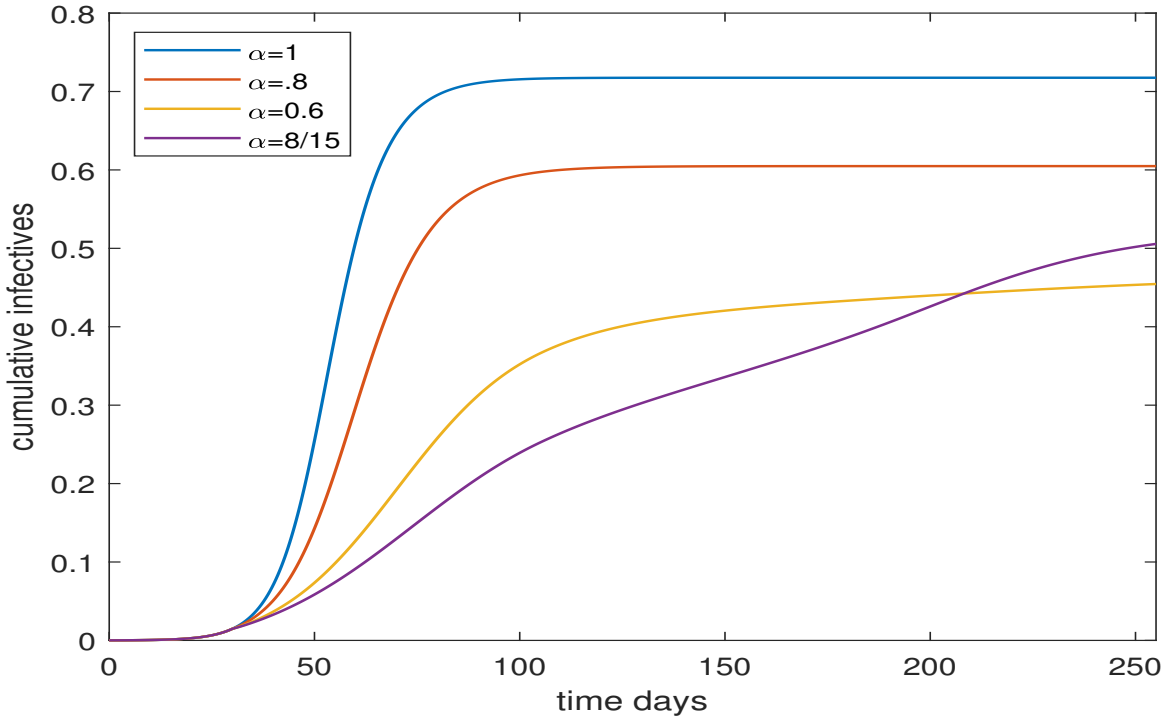
Figur: Incidence over time for the age+activity structure with $R_0 = 2.5$. Four different preventive levels inserted March 15 (day 30) and lifted June 30 (day 135). The blue, red, yellow and purple curves corresponds to no, light, moderate and severe preventive measures, respectively.

Cumulative fraction infected over time



Figur: Plot of the cumulative fraction infected over time for age+activity structure and $R_0 = 2.5$. Four different preventive levels inserted March 15 and lifted June 30. The blue, red, yellow and purple curves corresponds to no, light, moderate and severe preventive measures, respectively.

Cumulative: Gradual exit during summer



Figur: Same as above but: Preventive measures inserted March 15 and lifted gradually between June 1 and August 30. The blue, red, yellow and purple curves corresponds to no, light, moderate and severe preventive measures, respectively.

Stockholms universitet

Disease-induced Herd immunity: conclusions

Main result: Disease-induced herd immunity h_D is substantially lower than classical $h_C = 1 - 1/R_0$

How much lower? Needs to be investigated (Gabriela Gomes studied a model with continuously variable susceptibility)

Additional heterogeneities: Household, schools, work places, spatial, ... Most (all?) of these will make difference bigger!

"Non-proportional" restriction/prevention: isolation of elderly, school closing, ... Some will make difference bigger, others unclear

If socially active change behavior more $\Longrightarrow$ difference becomes smaller

Stockholms universitet

Over-all summary

General advice: Complement more advanced statistical analysis with simple model analysis. If similar conclusions: reassuring. If very different: mistake or understanding needed

Some important messages

- Prior (partial) immunty makes big difference for estimates
- Inference for emerging epidemics is hard
- Heterogeneities usually makes R_0 larger but not necessarily bigger outbreak!

Important but not treated:
– Changing behaviour over time
– Selection bias
– Asymptomatics and other under-reporting