

Who infected whom?

Estimation of infection trees, generation intervals and local network structure

Jacco Wallinga

Acknowledgements

- Hasselt University, Belgium
 - Teele Tamme
 - Sandra Waaijenborg
 - Han Ling
 - Ziv Shkedy
 - Niel Hens
- National Institute for Public Health and the Environment, Bilthoven, the Netherlands
 - Peter Teunis
 - Janneke Heijne

Questions in controlling emerging epidemics

- at the start of an epidemic
 - how effective should control measures be?
- during the epidemic
 - how effective are the implemented control measures?
 - are they effective enough to contain the outbreak?

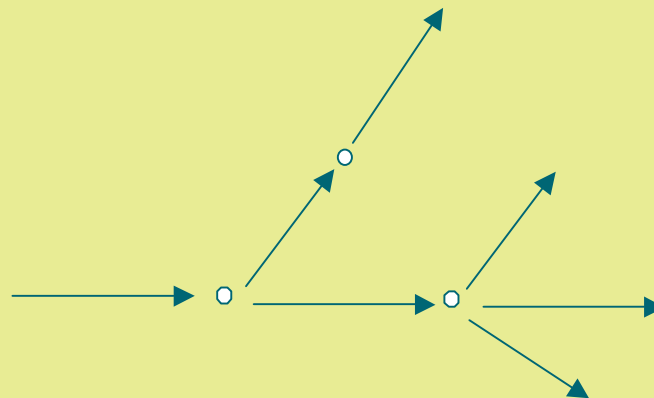


Reproductive number R

- R is defined as the number of secondary cases per primary case
- reproductive number R determines the required control effort
 - if R exceeds unity, the epidemic is out of control
 - in a homogeneously mixing population: the epidemic is controlled if a proportion $1-1/R$ of all infected cases is traced and isolated
- Estimation of R requires observing the generation interval τ
 - the duration between onset of symptoms of a secondary case and its primary case
 - characterize the generation interval distribution $g(\tau)$ by
 - mean μ
 - standard deviation σ

Observations in an (idealised) epidemic: Transmission tree

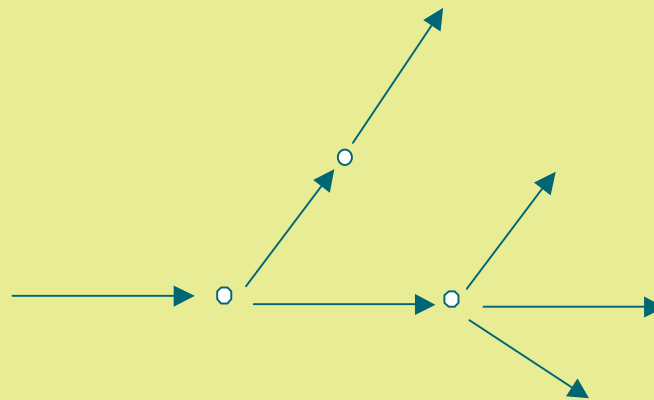
- node
 - position of node
 - edge
- case
- time of symptom onset
- transmission of infection



Estimation of R and τ from a transmission tree

- length of edge
- nr. outgoing edges

generation interval
reproduction number



- Grassly and Fraser, Nature Rev Microbiol 2008
- SIR model (panels a-c)
- SARS in Singapore (panels d-f)

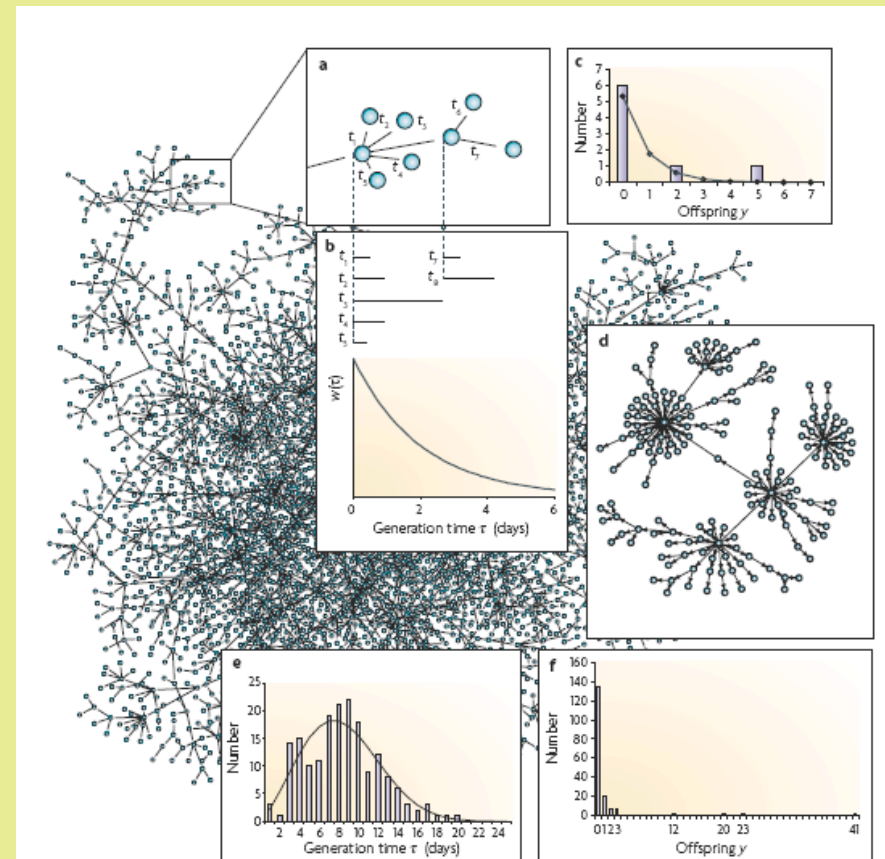
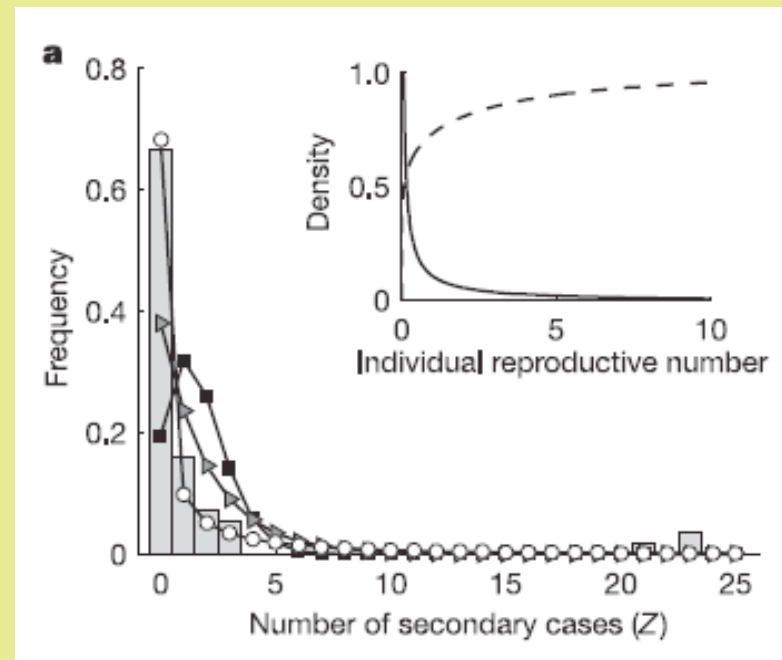


Figure 2 | Offspring and generation-time distribution for an epidemic. If each individual infects, on average, more than one additional individual then an epidemic can occur. A transmission tree is provided that shows who has infected who during a computer-generated epidemic in a closed population of 5,000 individuals. a | The expanded part of the

- Lloyd-Smith et al, Nature 2005
- SARS in Singapore



- Haydon et al. Proc. R. Soc. Lond. B 2003
- Foot-and-mouth outbreak in UK

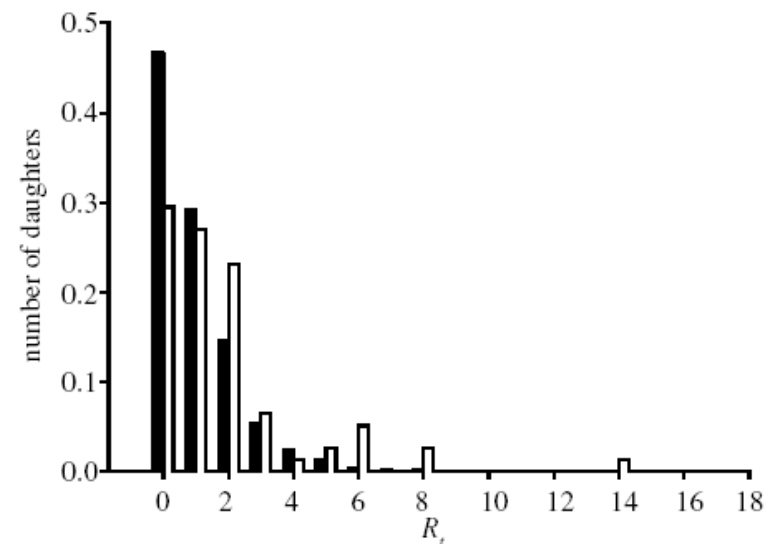
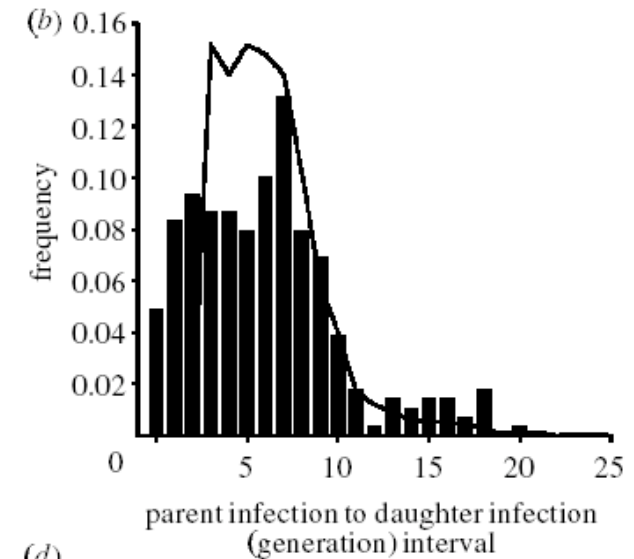


Figure 3. The distribution of numbers of daughters as deduced for all IPs up to 22 September for the pre-NMB (open bars) and post-NMB (filled bars) periods as inferred from the nearest-neighbour tree reconstruction.

Local contact network structure affects duration of generation intervals

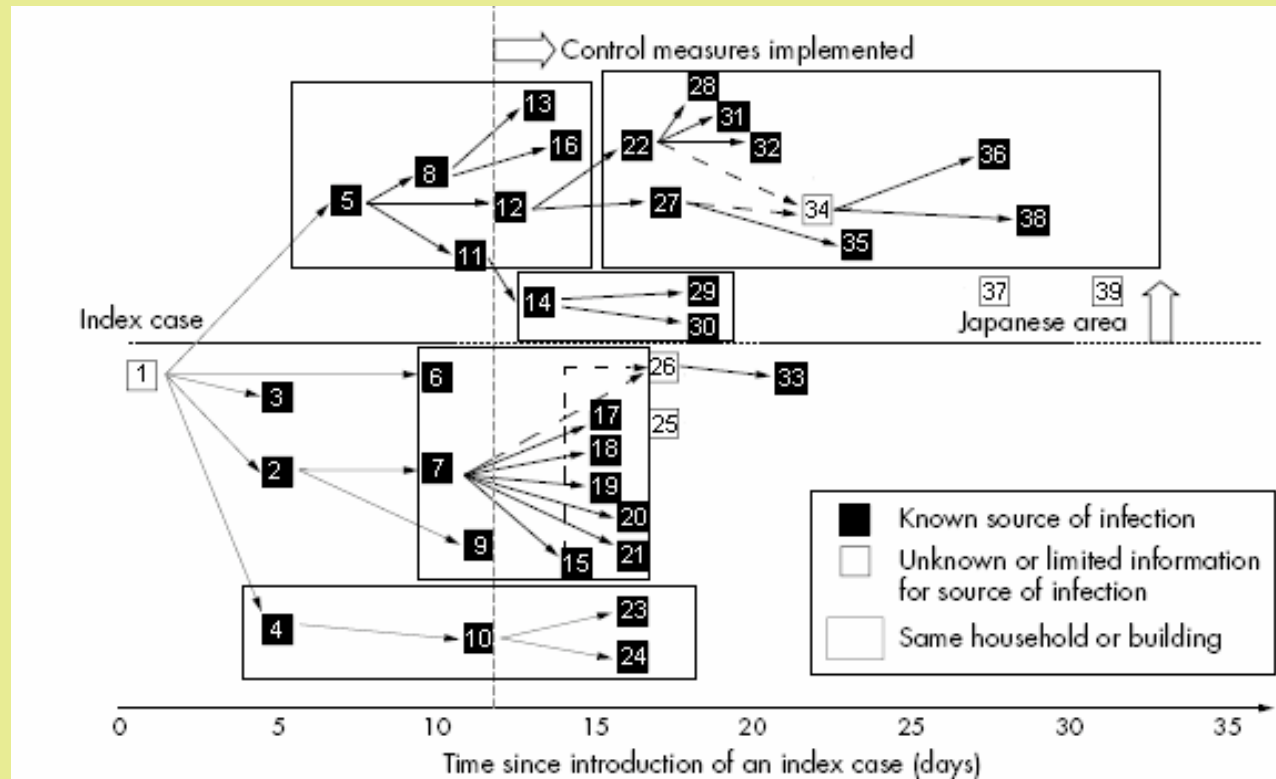
- If the contact network structure is a tree, the expected duration of the generation interval is constant during the epidemic
- If the contact network is highly clustered (many short loops), the expected duration of the generation interval should contract at the peak of the epidemic
 - Svensson. Math Biosci 2007; 208:300-311
 - Kenah et al. Math Biosci 2008; 213:71-79
- Can we use observed transmission trees to detect the local contact network structure?

Example: Pneumonic plague

- Bacterium *Yersinia pestis*
- Humans and animals (rodents) are hosts
- When transmitted from person-to-person: pneumonic plague
- Case fatality rate ~ 95%



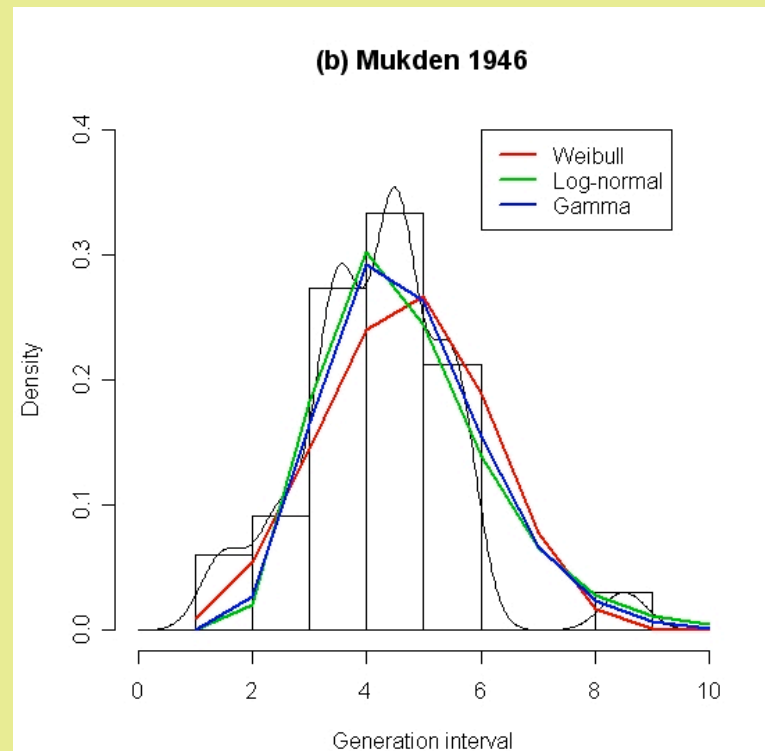
A completely observed transmission tree for pneumonic plague in Mukden, China, 1946.



Nishiura et al. J. Epidemiol. Community Health 2006; 60: 640-645

Generation interval distribution for pneumonic plague

- mean generation interval μ : 4.7 days
- standard deviation σ : 1.4 days
- distribution follows a Gamma distribution $g(\tau | \alpha, \beta)$



But some epidemiological links were missing

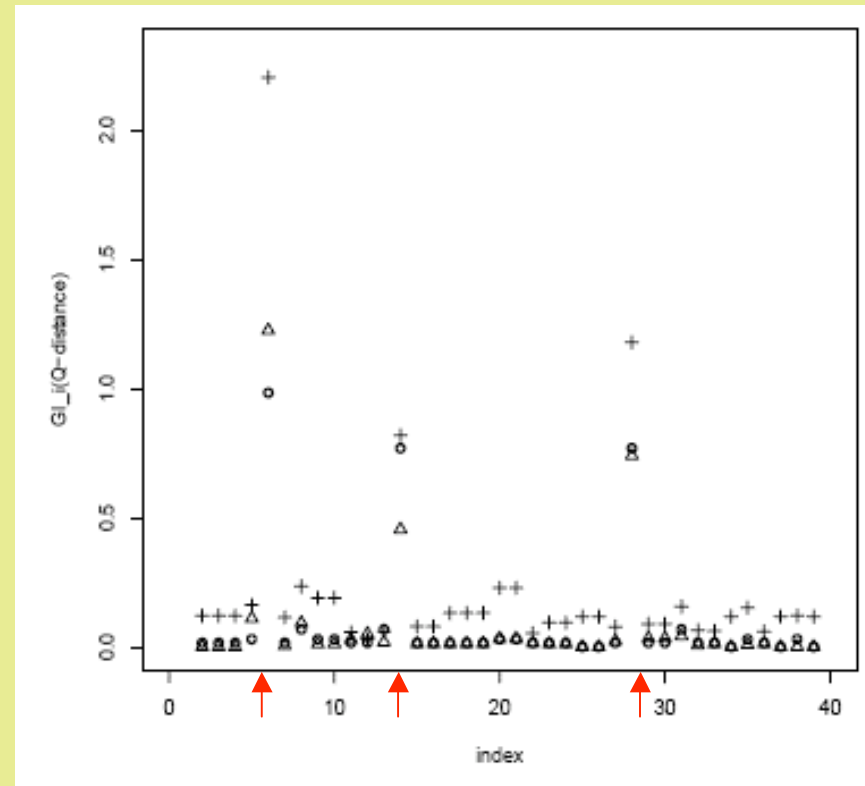
- Assume links are missing at random
- Use missing data techniques
 - relative likelihood of links (Likelihood)
 - Expectation-Maximization (EM-algorithm)
 - Expectation-Maximization with prior knowledge about possible contacts (PEM-algorithm)

Missing Link	Likelihood	EM-algorithm	PEM-algorithm
Case 25	14 (0.152)	14 (0.153)	15 (0.505)
	13 (0.152)	13 (0.153)	7 (0.228)
	12 (0.137)	12 (0.138)	
Case 26	14 (0.152)	14 (0.153)	7 (0.998)
	13 (0.152)	13 (0.153)	2 (0.002)
	12 (0.137)	12 (0.138)	
Case 34	30 (0.101)	30 (0.101)	27 (0.621)
	29 (0.101)	29 (0.101)	22 (0.379)
	28 (0.101)	28 (0.101)	
Case 37	35 (0.353)	35 (0.353)	35 (0.421)
	34 (0.318)	34 (0.319)	34 (0.387)
	33 (0.186)	33 (0.186)	32 (0.108)
Case 39	37 (0.374)	37 (0.376)	37 (0.470)
	36 (0.374)	36 (0.376)	36 (0.470)
	38 (0.210)	38 (0.208)	
Gamma Shape	11.57	11.72	11.02
Gamma Scale	0.41	0.40	0.43
Mean	4.70	4.70	4.75
Variance	1.91	1.88	2.04

Hens et al. Identifying missing and unlikely links in pneumonic plague outbreaks. In prep.

And some epidemiological links were misclassified

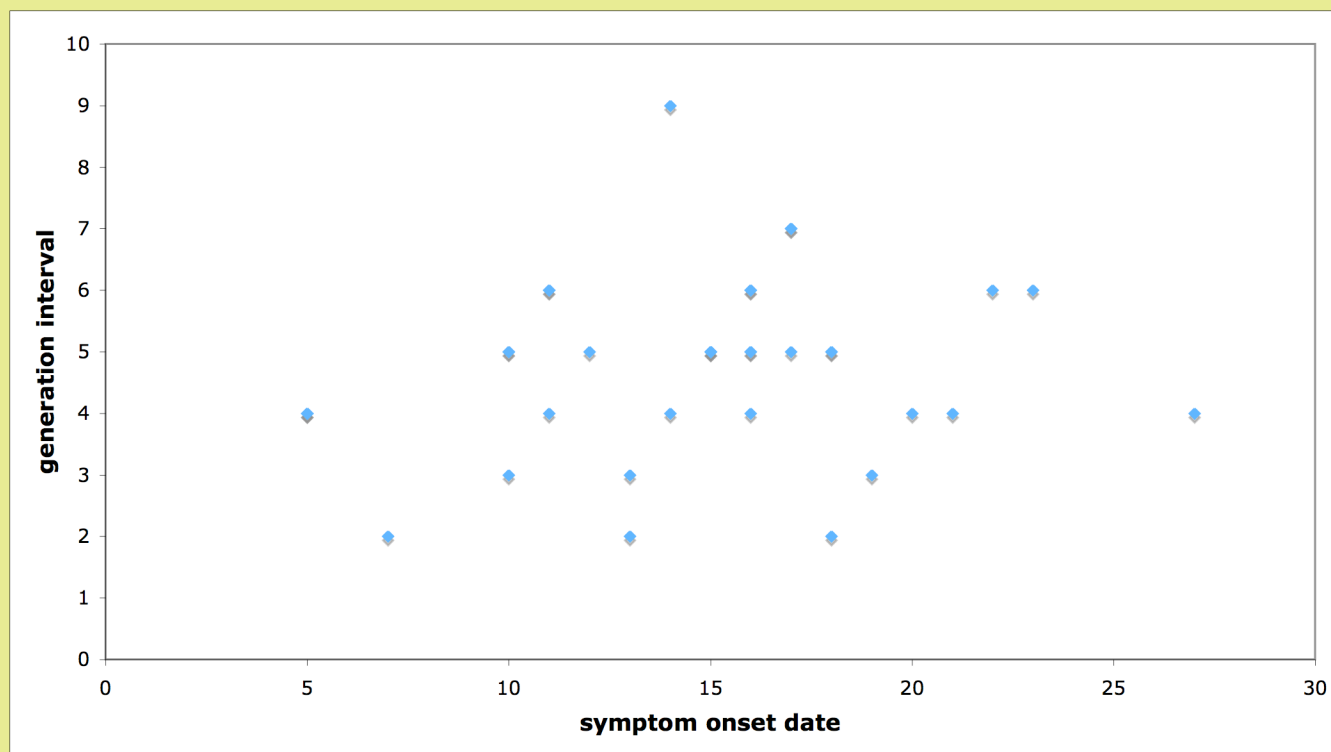
- Influence measure
 - sensitivity in outcome to one data point
- Links with a high influence measure are suspect
- Treat suspect links as missing



Hens et al. Identifying missing and unlikely links in pneumonic plague outbreaks. In prep.

Does the mean generation interval change?

- So far, no strong support for a decrease



Almost completely observed transmission trees

- reported transmission trees need verification and reconstruction
- characterization of the reproduction number is straightforward
 - at the individual level, number of outgoing links
 - at the population level, frequency distribution of outgoing links
- characterization of the generation interval distribution is straightforward
 - at the individual level, duration of a link
 - at the population level, frequency distribution of durations
- for pneumonic plague
 - the reproduction number was about 2 secondary cases per primary case before control
 - the mean generation interval was about 4.7 days
 - negligible contraction of generation interval during epidemic

Advantages of using transmission trees

- requires few arbitrary decisions and reduces to a standard approach for estimating reproductive number if we
 - assume exponential distribution of generation intervals
 - assume exponential growth of epidemic
 - observe mean generation interval μ and exponential growth rate r
 - take $R = 1 + \mu r$
- data requirements coincide with usual data collection for cases during outbreaks:
 - time of symptom onset
 - source of infection, if known
- potential for incorporating other information
 - molecular sequence of pathogen
 - characteristics of host (spatial location, gender, age, etc)

Limitations of using transmission trees

- requires some assumptions for reconstructing trees
 - links are missing at random
 - all infectious contacts are independent, that is, all cases are considered equally infectious
 - all cases are observed
 - generation interval has a stationary distribution

References

- Wallinga J, Teunis P. Different epidemic curves for Severe Acute Respiratory Syndrome reveal similar impacts of control measures. American Journal of Epidemiology 160: 509-516 (2004).
- Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc. R. Soc. B. 274: 599-604 (2007).
- Heijne JCM, Teunis P, Morroy G, Wijkman C, Oostveen S, Duizer E, Kretzschmar K, Wallinga J. Impact of enhanced hygiene measures on change in reproduction number during an outbreak of norovirus. Emerging Infectious Diseases xx: xx-xx (2008)

rivm

National Institute
for Public Health
and the Environment

jacco.wallinga@rivm.nl