Quantifying superspreading for COVID-19 using Poisson mixture distributions

C. Kremer[*1], A. Torneri[*2], S. Boesmans[3], H. Meuwissen[3], S. Verdonschot[3], K. Vanden Driessche[4,5], C. L. Althaus[6], C. Faes[1], N. Hens[1,2]

[1]I-BioStat, DSI, Hasselt University, Belgium
[2]CHERMID, VAXINFECTIO, University of Antwerp, Belgium
[3]Hasselt University, Belgium
[4]Antwerp University Hospital, Belgium
[5]Radboud UMC, The Netherlands
[6]University of Bern, Switzerland

Publication: https://doi.org/10.1038/s41598-021-93578-x

RSSB Annual Meeting
13 October 2021

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Health Economics & Modelling Infectious Diseases
Vaccine & Infectious Disease Institute
University of Antwerp

Background

R_0 = average number of secondary cases by an infected individual in a completely susceptible population

- considered constant among population members or specific population groups

Transmission potential = complex combination of host, pathogen, and environmental factors
→ individual variation in disease transmission

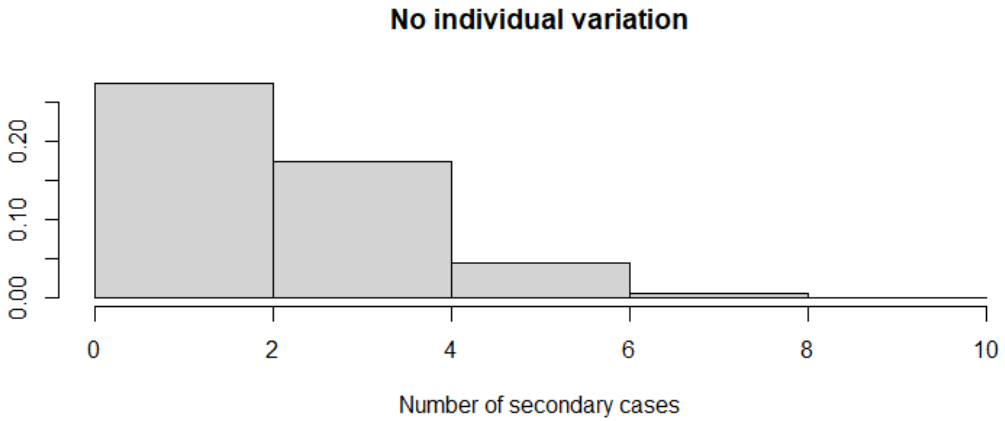
- affects outbreak probability and subsequent course

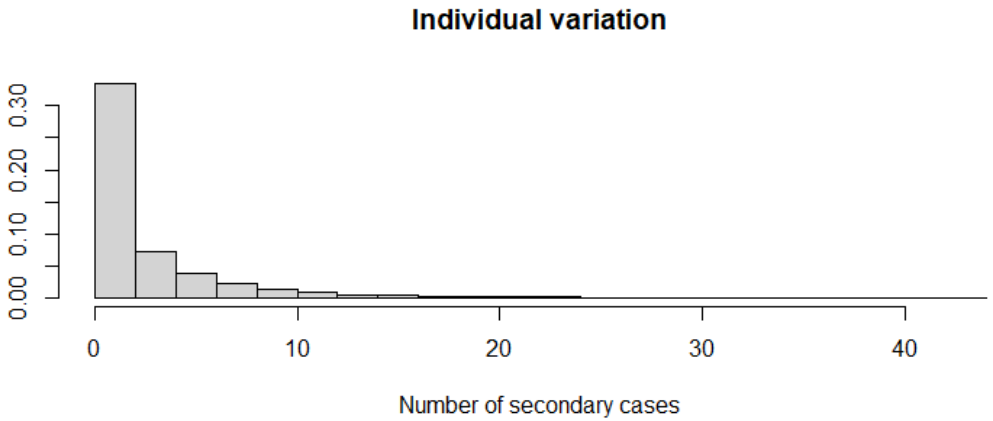
Superspreaders: infect substantially more individuals than others

- relatively small part of cases responsible for most of transmission
- many cases do not transmit disease

Population estimate of R_0 does not account for individual variation

→ Overdispersed offspring distribution with mean R_0





Lloyd-Smith *et al.* (2005): individual reproduction number ν as a random variable that represents the **expected** number of secondary cases

- Distribution with population mean R_0
  - encodes all variation in infectiousness
  - $\rightarrow$ SSE are realizations from the right-hand tail of the distribution of ν
- Stochastic effects in transmission modelled using a Poisson process
  - Most studies have assumed the rate $\nu \sim$ Gamma
  - $=$ negative binomial offspring distribution $\rightarrow$ heterogeneity quantified by dispersion parameter k
    - $k < 1$: substantial superspreading
    - lower k = increased heterogeneity
    - $\rightarrow$ easy comparison between studies

Individual variation in SARS-CoV-2 transmission

* (almost) all studies assume a negative binomial distribution
* k ranging from 0.1 to 0.6
* $p_{80\%}$ = proportion of cases responsible for 80% of transmission; ranging from 9% to 20%

$\rightarrow$ indicates important role of SSE in transmission

Other distributions?

- Brooks-Pollock *et al.* (2020): model the distribution of cluster sizes for TBC in the UK and NL
  - Poisson-lognormal better fit to UK data

→ importance of comparing different assumptions

What if the assumption of a negative binomial offspring distribution is not the best one?
→ Impact on estimates of heterogeneity?

Poisson mixtures

Number of secondary infections caused by each case is described by an offspring distribution

- Poisson contact process
- Rate following a continuous probability distribution
  - individual reproduction number ν

$\rightarrow$ Poisson mixture distribution: $Y \sim \text{Poisson}(\nu)$

Distribution for ν	Offspring distribution	Mean R	Variance σ^2
$\nu \sim \text{GG}(a,d,p)$	$Y \sim \text{PoGG}(a,d,p)$	$a\frac{\Gamma(\frac{d+1}{p})}{\Gamma(\frac{d}{p})}$	$a\frac{\Gamma(\frac{d+1}{p})}{\Gamma(\frac{d}{p})} + a^2\left[\frac{\Gamma(\frac{d+2}{p})}{\Gamma(\frac{d}{p})} - \left(\frac{\Gamma(\frac{d+1}{p})}{\Gamma(\frac{d}{p})}\right)^2\right]$
$\nu \sim \text{Ga}(\alpha,\beta)$	$Y \sim \text{NB}(\mu,k)$	$\mu = \frac{\alpha}{\beta}$	$\mu(1+\frac{\mu}{k}) = \frac{\alpha}{\beta}(1+\frac{1}{\beta})$
$\nu \sim \text{LogN}(\mu_{log},\sigma_{log})$	$Y \sim \text{PoLN}(\mu_{log},\sigma_{log})$	$e^{\mu_{log}+\frac{\sigma^2_{log}}{2}}$	$e^{\mu_{log}+\frac{\sigma^2_{log}}{2}} + \left[(e^{\sigma^2_{log}}-1)e^{2\mu_{log}+\sigma^2_{log}}\right]$
$\nu \sim \text{Weibull}(p,l)$	$Y \sim \text{PoWB}(p,l)$	$l\Gamma(1+\frac{1}{p})$	$l\Gamma(1+\frac{1}{p}) + l^2\left[\Gamma(1+\frac{2}{p}) - \left(\Gamma(1+\frac{1}{p})\right)^2\right]$

Simulation study

Objective: assess bias in estimates of R and its overdispersion

* generate data from each proposed mixture
  * varying levels of overdispersion
* estimate parameters using MLE
* obtain $\hat{R}_i$ and $\hat{\sigma}_i$

$\rightarrow$ Impact of assumed offspring distribution on proportion of cases responsible for certain proportion of transmission?

* i.e. importance of superspreading

Expected vs. realized proportions of transmission

Two methods:

1. **Lloyd-Smith et al. (2005):** calculate proportion responsible based on the distribution of ν $\rightarrow$ **expected** proportion of transmission based on inherent transmission potential
    * only depends on level of overdispersion, not R
2. **Endo et al. (2020):** take into account additional variation from Poisson process $\rightarrow$ **realized** proportion of transmission
    * depends on level of overdispersion as well as R

Same reasoning $\rightarrow$ CDF for disease transmission, $1 - F_{trans}(x)$ gives proportion of transmission due to cases with # secondary $> x$

1. Define proportion p of transmission we are interested in
2. x such that $1 - F_{trans}(x) = p$
3. Proportion responsible are those with # secondary $> x$

When based on discrete Poisson mixture $\rightarrow$ unlikely that there exists an integer x such that $1 - F_{trans}(x) = p$

- **Endo et al. (2020):** continuous approximation of discrete distribution
- Account for uncertainty around proportions
  - define x_1,x_2 such that $1 - F_{trans}(x_2) < p < 1 - F_{trans}(x_1)$
  - proportion responsible given by a range

Data examples

Fit Poisson mixtures to three **COVID-19 datasets** using MLE

- Empirical offspring distribution for
  - 290 cases in Hong Kong Adam et al. (2020)
  - 84 965 cases in India Laxminarayan et al. (2020)
  - 795 cases in Rwanda

→ Compare distributions in terms of AIC and GOF

- Akaike weights to quantify model selection uncertainty

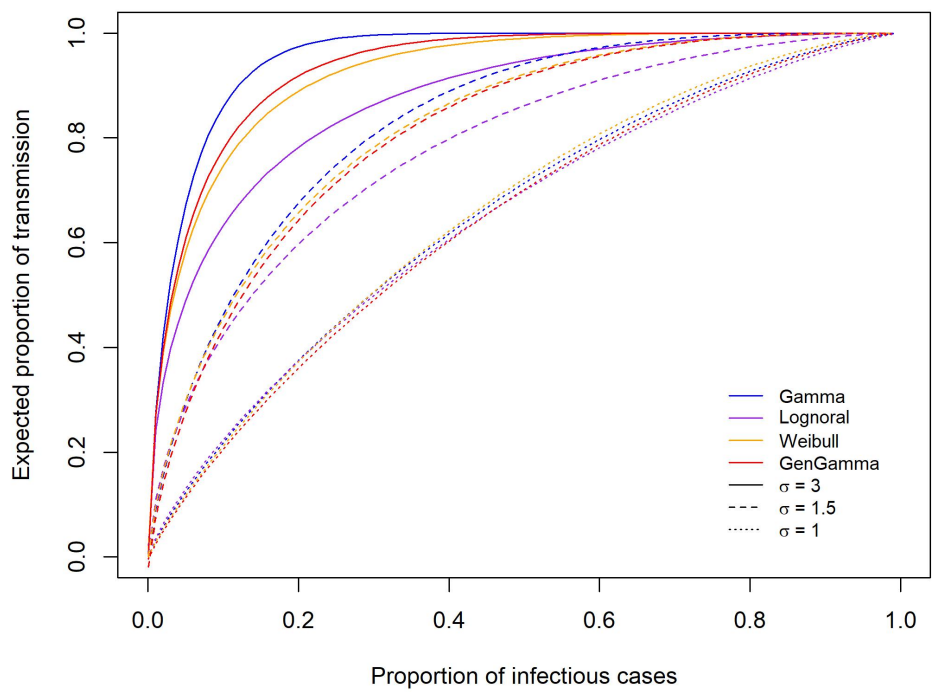
→ Different conclusions regarding superspreading potential, based on $p_{80\%}$?

Simulation study

As overdispersion ↗

- bias in variance estimates increases
- substantial difference between distributions

→ Inference of heterogeneity based on the 'wrong' distribution may be biased

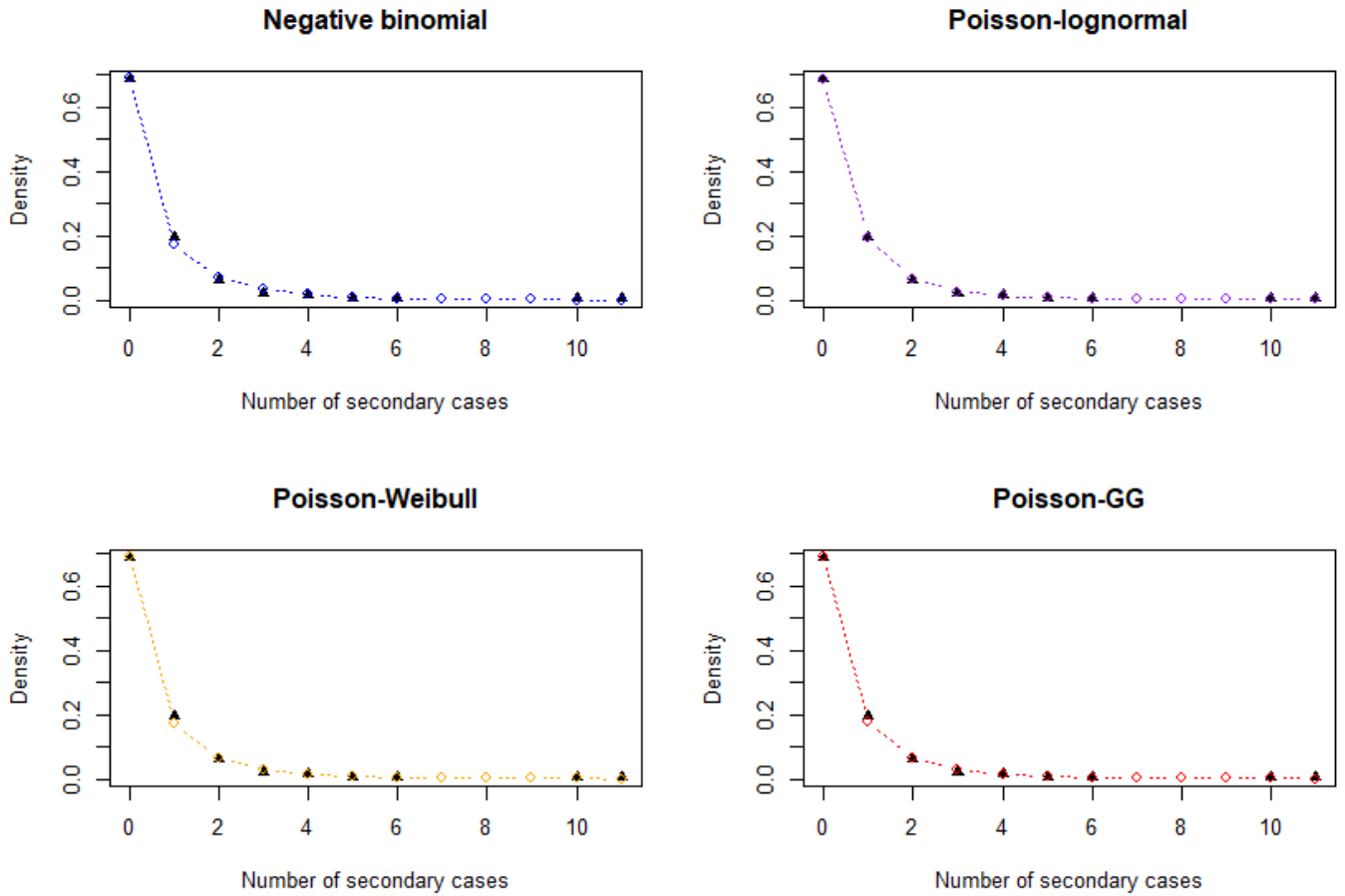


Data examples

Dataset	Distribution	R (95%CI)	$p_{80\%}$ (95%CI)	AIC	w_i
Hong Kong	NB	0.583 (0.448 - 0.718)	0.288 (0.208 - 0.345)	593.925	0.078
	POLN	0.587 (0.456 - 0.779)	0.332 (0.236 - 0.438)	590.009	0.551
	POWB	0.580 (0.445 - 0.745)	0.294 (0.223 - 0.358)	591.747	0.231
	POGG	0.580 (0.3789 - 0.724)	0.303 (0.279 - 0.325)	592.738	0.141
India	NB	0.484 (0.480 - 0.494)	0.319 (0.314 - 0.324)	163974.5	0.000
	POLN	0.484 (0.477 - 0.491)	0.373 (0.367 - 0.379)	162980.6	1.000
	POWB	0.483 (0.476 - 0.489)	0.322 (0.318 - 0.327)	163530.8	0.000
	POGG	0.484 (0.477 - 0.490)	0.333 (0.332 - 0.335)	163286.5	0.000
Rwanda	NB	0.259 (0.216 - 0.302)	0.323 (0.223 - 0.390)	1015.261	0.157
	POLN	0.260 (0.219 - 0.311)	0.389 (0.318 - 0.459)	1013.073	0.468
	POWB	0.259 (0.217 - 0.311)	0.331 (0.241 - 0.394)	1014.350	0.247
	POGG	0.259 (0.216 - 0.301)	0.344 (0.337 - 0.350)	1015.667	0.128

* Poisson-lognormal best fit based on AIC
* Considerable differences in expected $p_{80\%}$ when based on distribution of ν, compared to NB
    * ranges mostly overlapping when based on complete offspring distribution

Hong Kong data: NB, POGG, (POWB) don't adequately capture proportion of cases generating one secondary case → overestimating superspreading potential



Conclusions

Importance of model comparison

- NB often underestimates proportion of cases with only 1 secondary case
  - → overestimates importance of SSEs

- Studies can also be compared by $p_{80\%}$ instead of k
  - can be obtained for any distribution
  - more intuitive interpretation
  - be aware of the two different approaches!

Importance of correctly quantifying heterogeneity

- 'Superspreading potential' needs to be taken in account when modeling disease control / planning control strategies
  - In case of high overdispersion (i.e. low $p_{80\%}$), control measures should focus on limiting potential SSE
    - restricting large events, avoid crowding, ...
  - For higher $p_{80\%}$, additional control measures needed focusing on regular contacts

Future work

- Inference from final size data
- Disentangle heterogeneity coming from variation in contact rates vs. variation in viral shedding, to improve control measures
- Other distributions to describe contact process
  - Poisson process is likely a simplification

Acknowledgements

We thank Muhammed Semakula for sharing the data obtained from Rwandan contact tracing efforts.

This project has received funding from the European Union's Horizon 2020 research and innovation programme - project EpiPose (Grant agreement number 101003688).

The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation Flanders (FWO) and the Flemish Government department EWI.

Selected references

Adam, D.C., et al. (2020). Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nature Medicine*, https://doi.org/10.1038/s41591-020-1092-0.

Brooks-Pollock, E., et al. (2020). A model of tuberculosis clustering in low incidence countries reveals more transmission in the United Kingdom than the Netherlands between 2010 and 2015. *PLoS Computational Biology* 16, e1007687, https://doi.org/10.1371/journal.pcbi.1007687.

Burton A., et al. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine* 25, 4279-4292, 10.1002/sim.2673.

Endo, A., et al. (2020). Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Research*, https://doi.org/10.12688/wellcomeopenres.15842.3.

Lau, M.S.Y., et al. (2020). Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. *PNAS* 36; 22430-22435, https://doi.org/10.1073/pnas.2011802117.

Laxminarayan, R., et al. (2020). Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science*, 10.1126/science.abd7672.

Lloyd-Smith, J.O., et al. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature* 438;355-359, doi:10.1038/nature04153.

Susswein, Z. & Bansal, S. (2020). Characterizing superspreading of SARS-CoV-2: from mechanism to measurement. *medRxiv preprint*, https://www.medrxiv.org/content/10.1101/2020.12.08.20246082v1.