

# The statistical analysis of serological and social contact data to inform infectious disease modelling

**Niel Hens**

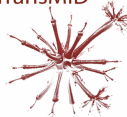


Interuniversity Institute for Biostatistics  
and statistical Bioinformatics



European Research Council  
Established by the European Commission

TransMID



[www.simid.be](http://www.simid.be)



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



ESOC  
Epidemiology and Social Medicine  
University of Antwerp



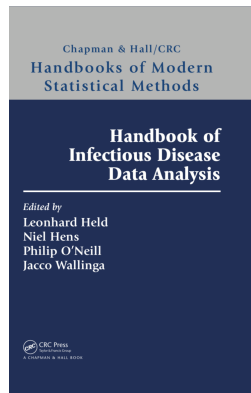
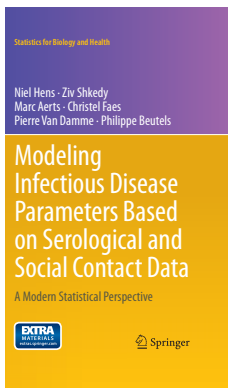
[www.simid.be](http://www.simid.be)  
[www.socialcontactdata.org](http://www.socialcontactdata.org)

journées Statistiques et Santé, CNAM, Paris 2019

# Background

- SIMID-initiative (since 2008):
  - CenStat, I-BioStat, Data Science Institute, UHasselt
  - CHERMID, Vaxinfectio, UAntwerp
- Unifying statistical and mathematical models:  
Sir Ronald Ross (1857-1932)
- Research Topics:
  - maternal immunity
  - vaccine-induced immunity
  - serological data analysis
  - social contact patterns
  - epidemic modelling
  - demographic modelling
  - cost effectiveness analysis

# Background



Hens et al. (2012), Held et al. (2019)

# Outline

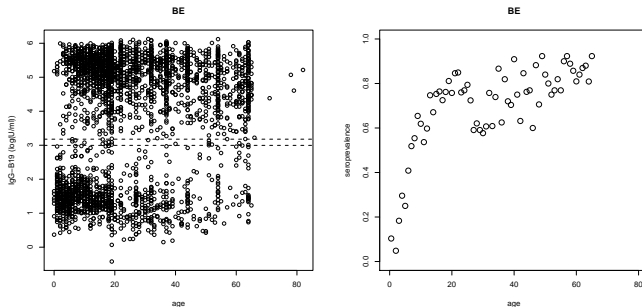
- 1 Introduction
  - Serological data
  - Estimation framework
  - The traditional 'WAIFW' approach
- 2 Estimating inf. disease parameters using serological and social contact data
  - The social contact hypothesis
  - Dimensions of uncertainty
  - Application to VZV
- 3 An excursion to questions that can be answered by using these data
  - Examples of use and analysis of social contact data
  - Examples of use and analysis of serological data
- 4 Using frailty models for the analysis of multivariate serological data
  - Setting the scene
  - The mass action principle
  - Bivariate correlated frailty models
  - Infection processes
  - Data
  - Discussion & conclusion

# Outline

- 1 Introduction
  - Serological data
  - Estimation framework
  - The traditional 'WAIFW' approach
- 2 Estimating inf. disease parameters using serological and social contact data
  - The social contact hypothesis
  - Dimensions of uncertainty
  - Application to VZV
- 3 An excursion to questions that can be answered by using these data
  - Examples of use and analysis of social contact data
  - Examples of use and analysis of serological data
- 4 Using frailty models for the analysis of multivariate serological data
  - Setting the scene
  - The mass action principle
  - Bivariate correlated frailty models
  - Infection processes
  - Data
  - Discussion & conclusion


## Serological data

- cross-sectional set of residual blood samples (hospital laboratories, blood donors, ...)
- tested for infection-specific IgG antibodies using ELISA
- antibody level  $>$  cut-off value  $\rightarrow$  seropositive



Example: parvovirus B19 in Belgium

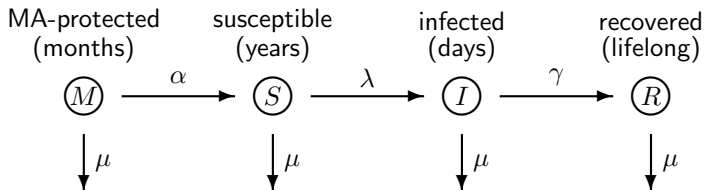
# Setting the scene: varicella zoster virus

	varicella zoster virus (VZV) 
clinical presentation	primary infection → varicella reactivation → herpes zoster
transmission	direct or aerosol
infectious period	about 7 days
vaccination	no active immunisation in most of Europe
serological surveys	BE (2002)

# Estimation framework

- compartmental models to describe infectious disease dynamics
- simple example:

## MSIR model



- dynamical system over age  $a$  and time  $t$



# Estimation framework

- MSIR model and variants can be used to analyze serological surveys
  - estimate:
    - age-specific force of infection
    - basic reproduction number  $R_0$
    - critical immunization level
    - ...
  - plan and monitor vaccination programmes or intervention strategies
- often requires estimation of age-dependent transmission rates
  - essential to determine the pattern of person-to-person spread of an infection in a large population
- requires making assumptions



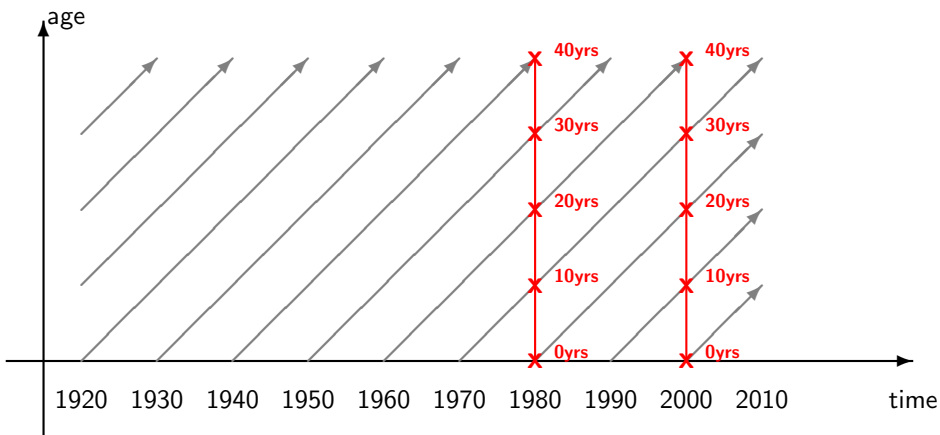
# The MSIR model

- Compartments and time scales.
- Each differential equation represent the change (over time and age) in the compartment.

$$\left\{ \begin{array}{l} \frac{\partial M(a,t)}{\partial a} + \frac{\partial M(a,t)}{\partial t} = -\alpha(a)M(a,t), \\ \frac{\partial S(a,t)}{\partial a} + \frac{\partial S(a,t)}{\partial t} = \alpha(a)M(a,t) - (\lambda(a,t) + \mu(a))S(a,t), \\ \frac{\partial I(a,t)}{\partial a} + \frac{\partial I(a,t)}{\partial t} = \lambda(a,t)S(a,t) - (\nu + \mu(a))I(a,t), \\ \frac{\partial R(a,t)}{\partial a} + \frac{\partial R(a,t)}{\partial t} = \nu I(a,t) - \mu(a)R(a,t), \end{array} \right.$$

where  $N(a,t) = M(a,t) + S(a,t) + I(a,t) + R(a,t)$  and  $M(0,T) = B(t)$ , the number of births all susceptible to infection.

# The Lexis diagram



# MSIR model

## MSIR model: assuming endemic equilibrium

$$\left\{ \begin{array}{lcl} \frac{dM(a)}{da} & = & -\{\alpha(a) + \mu(a)\}M(a), \\ \frac{dS(a)}{da} & = & \alpha(a)M(a) - \{\lambda(a) + \mu(a)\}S(a), \\ \frac{dI(a)}{da} & = & \lambda(a)S(a) - \{\gamma(a) + \mu(a)\}I(a), \\ \frac{dR(a)}{da} & = & \gamma(a)I(a) - \mu(a)R(a) \end{array} \right.$$

- $\mu(a)$  = mortality rate
- $\alpha(a)$  = rate of losing maternal antibodies
- $\lambda(a)$  = rate at which a susceptible of age  $a$  acquires infection  
→ force of infection (FOI)
- $\gamma(a)$  = recovery rate

# Mass action principle

## MAP

$$\lambda(a) = D \int_0^{\infty} \beta(a, a') \lambda(a') S(a') da'$$

- $\beta(a, a')$  = transmission rate, i.e. per capita rate at which an individual of age  $a'$  makes an effective contact with a person of age  $a$ , per year
- $D$  = mean duration of infectiousness

# Estimation from serological data

- solve MSIR ODEs and derive expressions for
  - fraction of susceptibles  $s(a)$  → solve MAP:

$$\lambda(a) = D \int_0^{\infty} \beta(a, a') \lambda(a') s(a') da'$$

- fraction of seropositives  $r(a)$  → evaluate loglikelihood:

$$\begin{aligned} \ell\ell &= \sum_{i=1}^n \tilde{w}_i \{ y_i \log[r(a_i)] + (1 - y_i) \log[1 - r(a_i)] \} \\ y_i &= \begin{cases} 1, & \text{if seropositive} \\ 0, & \text{if seronegative} \end{cases} \end{aligned}$$

# Estimation from serological data

- assume type I maternal antibodies  $\rightarrow$  age  $A$
- MSIR - fraction of susceptibles:

$$s(a) = \exp \left( - \int_A^a \lambda(u) du \right) \approx 1 - r(a)$$

- move to discrete age framework  $\rightarrow$  solve MAP iteratively

$$\beta(a, a') = \beta_{ij} \text{ if } a \in [a_{[i]}, a_{[i+1]}) \text{ and } a' \in [a_{[j]}, a_{[j+1]})$$

- $\beta_{ij}$  'Who-Acquires-Infection-From-Whom' (WAIFW) matrix

# Estimation from serological data

- estimate parameters determining  $\beta_{ij}$  from serological data using ML-estimation
- estimate  $R_0$  = dominant eigenvalue of the next generation matrix with elements:

$$D \left( \int_{a_{[i]}}^{a_{[i+1]}} N(a) da \right) \beta_{ij}$$



## Traditional approach

- Anderson and May (1992): mixing patterns
  - impose mixing pattern on  $\beta_{ij}$
  - constrain # distinct elements
  - based on prior knowledge of social mixing behaviour

	age class 1	age class 2	age class 3	age class 4
	↓	↓	↓	↓
age class 1 →	$\beta_1$	$\beta_4$	$\beta_4$	$\beta_4$
age class 2 →	$\beta_4$	$\beta_2$	$\beta_4$	$\beta_4$
age class 3 →	$\beta_4$	$\beta_4$	$\beta_3$	$\beta_4$
age class 4 →	$\beta_4$	$\beta_4$	$\beta_4$	$\beta_4$

# WAIFW approach applied to VZV

previously used WAIFW structures for VZV: 6 age-categories:

$$\begin{aligned}
 W_1 &= \begin{pmatrix} \beta_1 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\ \beta_6 & \beta_2 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\ \beta_6 & \beta_6 & \beta_3 & \beta_6 & \beta_6 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_4 & \beta_6 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_5 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \end{pmatrix} & W_2 &= \begin{pmatrix} \beta_1 & \beta_1 & \beta_3 & \beta_4 & \beta_5 & \beta_6 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 & \beta_6 \\ \beta_3 & \beta_3 & \beta_3 & \beta_4 & \beta_5 & \beta_6 \\ \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_5 & \beta_6 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \end{pmatrix} \\
 W_3 &= \begin{pmatrix} \beta_1 & \beta_1 & \beta_1 & \beta_1 & \beta_1 & \beta_1 \\ \beta_2 & \beta_2 & \beta_2 & \beta_2 & \beta_2 & \beta_2 \\ \beta_3 & \beta_3 & \beta_3 & \beta_3 & \beta_3 & \beta_3 \\ \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_4 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \end{pmatrix} & W_4 &= \begin{pmatrix} \beta_1 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\ \beta_6 & \beta_2 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\ \beta_6 & \beta_6 & \beta_3 & \beta_6 & \beta_6 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_4 & \beta_6 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_5 & \beta_6 \\ \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_5 \end{pmatrix}
 \end{aligned}$$

issues:

- direct estimation or model based - regular matrices
- selection of best WAIFW-matrix using AIC

## WAIFW approach applied to VZV

with (Belgian situation)

- $N = 9\,943\,749$ ,  $L = 80$ ,  $D = 7/365$
- age categories  $[0.5,2)$ ,  $[2,6)$ ,  $[6,12)$ ,  $[12,19)$ ,  $[19,31)$ ,  $[31,80)$

using direct estimation based on binomial likelihood

Model	$\hat{R}_0$	AIC
$W_1$	8.831	1375.7
$W_2$	3.512	1372.8
$W_3$	4.213	1372.8
$W_4$	8.807	1375.7

# WAIFW approach applied to VZV

- Farrington and Whitaker (2005): continuous surface

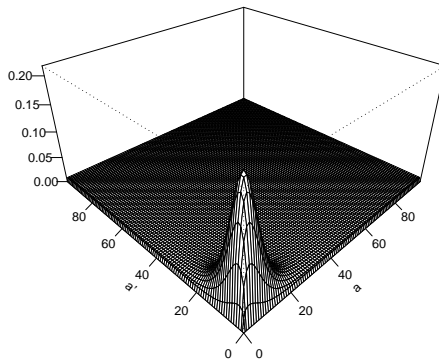
$$\beta(a, a') = \kappa(\gamma(u) \times b(v|u) + \delta),$$

where

$$\gamma(u; \mu, \nu) = c^{-1} u^{\nu-1} \exp\left(-\frac{\nu u}{\sqrt{2}\mu}\right),$$

$$b(v|u; \alpha, \beta) = \frac{(u+v)^{\alpha-1} (u-v)^{\beta-1}}{u^{\alpha+\beta-2}}$$

# A Continuous WAIFW Applied to VZV



# Traditional approach

- Anderson and May (1992): mixing patterns  
→ disadvantages:
  - low dimensional matrices
  - non-realistic discontinuities
  - choice age classes: ad hoc
- Farrington & Whitaker (2005): continuous contact surface
- both methods rely on strong parametric assumptions
- Wallinga et al. (2006): use data on social contacts to inform estimation of age-dependent transmission rates

# Outline

- 1 Introduction
  - Serological data
  - Estimation framework
  - The traditional 'WAIFW' approach
- 2 Estimating inf. disease parameters using serological and social contact data
  - The social contact hypothesis
  - Dimensions of uncertainty
  - Application to VZV
- 3 An excursion to questions that can be answered by using these data
  - Examples of use and analysis of social contact data
  - Examples of use and analysis of serological data
- 4 Using frailty models for the analysis of multivariate serological data
  - Setting the scene
  - The mass action principle
  - Bivariate correlated frailty models
  - Infection processes
  - Data
  - Discussion & conclusion

## Social contact approach

for infections transmitted primarily through non-sexual social contacts:

rates of conversational/physical contact

↕ relation?

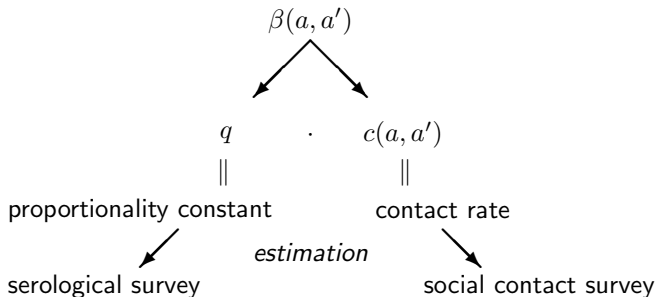
transmission rates  $\beta(a, a') \Leftrightarrow$  WAIFW-matrix





# Social contact hypothesis

Social contact hypothesis (Wallinga et al., 2006)



# Social contact survey

## Alternative approach:

Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents.

## Objectives

- Disentangle contact behaviour from transmission process
- Get insights in predictiveness of social contact data
- Get new insights in the transmission process



Edmunds et al. (1997)

Beutels et al. (2006)

Mossong et al. (2008)

Hens et al. (2009)

McCaw et al. (2010)

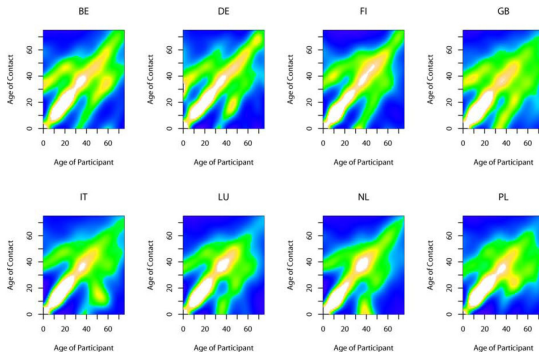
Horby et al. (2011)

...

- └ Estimating inf. disease parameters using serological and social contact data

- └ The social contact hypothesis

## EU mixing patterns



- common structure
- note the converging off-diagonals: parents get older

# Dimensions of uncertainty

Social contact hypothesis (Wallinga et al., 2006)

$$\begin{array}{ccc} & \beta(a, a') & \\ & \swarrow \quad \searrow & \\ q & \cdot & c(a, a') \end{array}$$

model selection uncertainty:

①  $c(a, a') \rightarrow$  what type of contact?

②  $q = \text{constant} \rightarrow$  assumption too strong?

## Applied to VZV in Belgium

- modeling assumptions:
  - MSIR compartmental model (lifelong immunity)
  - endemic equilibrium
  - type I maternal antibodies and type I mortality
- ML-estimation:  $\hat{q} \rightarrow \hat{R}_0$

## Type of contact?

- select types of contact with high transmission potential for VZV
- which one induces the best fit to the observed serological profile?

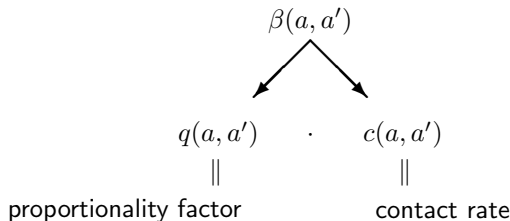
model	parameter	type of contact
$C_1$	$q_1$	all contacts
$C_2$	$q_2$	close contacts
$C_3$	$q_3$	close contacts > 15 minutes
$C_4$	$q_4$	close contacts and non-close contacts > 1 hour
$C_5$	$q_5$	close contacts > 15 minutes and non-close contacts > 1 hour

model	$\hat{q}_k$	95% CI	$\hat{R}_0$	95% CI	AIC
$C_1$	0.132	[0.103,0.175]	15.69	[12.34,21.41]	1386.618
$C_2$	0.160	[0.126,0.208]	10.24	[8.21,13.68]	1379.581
$C_3$	0.173	[0.133,0.221]	8.68	[6.89,11.34]	1374.958
$C_4$	0.145	[0.113,0.188]	11.73	[9.41,15.95]	1380.354
$C_5$	0.156	[0.119,0.204]	10.40	[8.05,14.10]	1376.068

## $q = \text{constant?}$

- $q$  might depend on age-specific characteristics related to susceptibility and infectiousness
- contacts reported in diaries are just proxies of at-risk events by which infection can be transmitted

### age-dependent proportionality



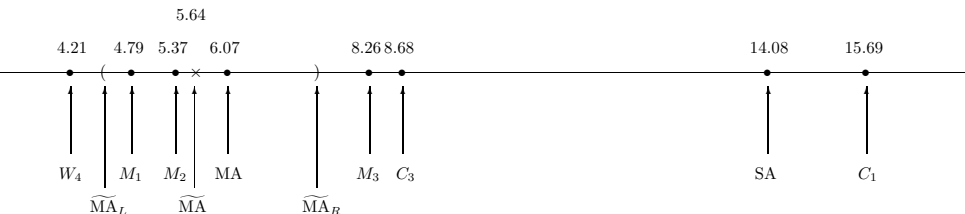
various options: back to issues with identifiability  $\rightarrow$  model averaging

- Estimating inf. disease parameters using serological and social contact data

- Application to VZV

# Multi-model inference

Goeyvaerts et al. (2010):





# Summary

- Dimensions of uncertainty:
  - type of contact underlying transmission of disease
  - parametric model relating the contact rates to the transmission rates (e.g. constant/age-dependent proportionality)
- model selection uncertainty: different models may induce similar fit, while entailing different estimates of  $R_0$
- to overcome this problem: multi-model inference  
→ model averaged estimate for  $R_0$
- VZV in Belgium: improvement of fit by modeling transmission as the product of two age-specific variables: the age-specific contact rate and an age-specific proportionality factor

# Outline

- 1 Introduction
  - Serological data
  - Estimation framework
  - The traditional 'WAIFW' approach
- 2 Estimating inf. disease parameters using serological and social contact data
  - The social contact hypothesis
  - Dimensions of uncertainty
  - Application to VZV
- 3 An excursion to questions that can be answered by using these data
  - Examples of use and analysis of social contact data
  - Examples of use and analysis of serological data
- 4 Using frailty models for the analysis of multivariate serological data
  - Setting the scene
  - The mass action principle
  - Bivariate correlated frailty models
  - Infection processes
  - Data
  - Discussion & conclusion

- └ An excursion to questions that can be answered by using these data

- └ Examples of use and analysis of social contact data

## A historical perspective

- Social contacts as **proxies of transmission events** of airborne infections
- **Rapoport and Horvath (1961)**: first social surveys to construct networks for studying the spread of infection
- Several studies have been conducted since: Edmunds et al. (1997); Wallinga et al. (2006); Mossong et al. (2008); Salathé et al. (2010); Read et al. (2012); Danon et al. (2013); Kucharski et al. (2014); Kwok et al. (2014); Eames et al. (2015); Dodd et al. (2016), ...
- **Wallinga et al. (2006)**: conversational contacts predictive for age-specific proportion of persons immune against mumps in Utrecht in 1986 and against pandemic influenza in Cleveland in 1957.

social contact hypothesis:  $q(a, a') = q$

- └ An excursion to questions that can be answered by using these data

- └ Examples of use and analysis of social contact data

# Systematic review

Hoang et al. (2019):

REVIEW ARTICLE

OPEN

## A Systematic Review of Social Contact Surveys to Inform Transmission Models of Close-contact Infections

*Thang Hoang<sup>a</sup>, Pietro Coletti<sup>a</sup>, Alessia Melegaro<sup>b</sup>, Jacco Wallinga<sup>c,d</sup>, Carlos G. Grijalva<sup>e</sup>,  
John W. Edmunds<sup>f</sup>, Philippe Beutels<sup>g</sup>, and Niel Hens<sup>a,g</sup>*

- diary-based approach & face-to-face interviews
- data sharing initiative: [www.socialcontactdata.org](http://www.socialcontactdata.org) & socialmixr-package (R software)

- └ An excursion to questions that can be answered by using these data
- └ Examples of use and analysis of social contact data

# Estimating contact rates

- General guidelines
  - Hens and Wallinga (Wiley, StatsRef 2019)
  - Wallinga, van de Kasstele and Hens (HIDDA, 2019)
- Smoothing approaches
  - van de Kasstele et al. (2017): smoothing INLA
  - Camarda and Hens (2013), Vandendijck et al. (2018): clever choice of axes to smooth over

- └ An excursion to questions that can be answered by using these data

- └ Examples of use and analysis of social contact data

## Behavioural change: regular holiday periods

Luca et al. (2018):

De Luca et al. *BMC Infectious Diseases* (2018) 18:29  
DOI 10.1186/s12879-017-2934-3

BMC Infectious Diseases

RESEARCH ARTICLE

Open Access



### The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for Belgium

Giancarlo De Luca<sup>1</sup>, Kim Van Kerckhove<sup>2</sup>, Pietro Coletti<sup>2</sup>, Chiara Poletto<sup>1</sup>, Nathalie Bossuyt<sup>3</sup>, Niel Hens<sup>2,4</sup> and Vittoria Colizza<sup>1,5\*</sup> 

- Holiday periods have a significant impact on transmission
- Organising holiday periods could mitigate the transmission of infectious diseases

└ An excursion to questions that can be answered by using these data

└ Examples of use and analysis of social contact data

# Behavioural change: illness

Santermans et al. (2017):

Mathematical Biosciences 285 (2017) 43–54



## Structural differences in mixing behavior informing the role of asymptomatic infection and testing symptom heritability



Eva Santermans<sup>a,\*</sup>, Kim Van Kerckhove<sup>a</sup>, Amin Azmon<sup>b</sup>, W. John Edmunds<sup>c</sup>,  
Philippe Beutels<sup>d</sup>, Christel Faes<sup>a</sup>, Niel Hens<sup>a,d</sup>

<sup>a</sup>Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Belgium

<sup>b</sup>Novartis Pharma AG, Oncology Business Unit/General Medical Affairs, Novartis Campus, Basel, Switzerland

<sup>c</sup>Centre for the Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom

<sup>d</sup>Centre for Health Economics Research and Modelling Infectious Diseases, Vaccine & Infectious Disease Institute, University of Antwerp, Belgium

- Illness has an impact on activity (contacts & mobility)
- Different disease dynamics

- └ An excursion to questions that can be answered by using these data

- └ Examples of use and analysis of social contact data

# Inferring networks from social contact data

Goeyvaerts et al. (2018):

## PROCEEDINGS B

[rspb.royalsocietypublishing.org](http://rspb.royalsocietypublishing.org)

### Research



**Cite this article:** Goeyvaerts N, Santermans E, Potter G, Torneri A, Van Kerkhove K, Willem L, Aerts M, Beutels P, Hens N. 2018 Household members do not contact each other at random: implications for infectious disease modelling. *Proc. R. Soc. B* **285**: 20182201. <http://dx.doi.org/10.1098/rspb.2018.2201>

## Household members do not contact each other at random: implications for infectious disease modelling

Nele Goeyvaerts<sup>1,1,‡</sup>, Eva Santermans<sup>1,‡</sup>, Gail Potter<sup>2</sup>, Andrea Torneri<sup>3</sup>, Kim Van Kerkhove<sup>1</sup>, Lander Willem<sup>3</sup>, Marc Aerts<sup>1</sup>, Philippe Beutels<sup>3</sup> and Niel Hens<sup>1,3</sup>

<sup>1</sup>Interuniversity Institute for Biostatistics and Statistical Bioinformatics, UHasselt, Hasselt, Belgium

<sup>2</sup>The Emmes Corporation, Rockville, MD, USA

<sup>3</sup>Centre for Health Economics Research and Modelling Infectious Diseases, Vaccine and Infectious Disease Institute, University of Antwerp, Antwerp, Belgium

ES, 0000-0002-8537-2625

- Household members do not mix at random
- Combining household and egocentric data proves to be useful



- └ An excursion to questions that can be answered by using these data

- └ Examples of use and analysis of social contact data

# Outline

- 1 Introduction
  - Serological data
  - Estimation framework
  - The traditional 'WAIFW' approach
- 2 Estimating inf. disease parameters using serological and social contact data
  - The social contact hypothesis
  - Dimensions of uncertainty
  - Application to VZV
- 3 An excursion to questions that can be answered by using these data
  - Examples of use and analysis of social contact data
  - Examples of use and analysis of serological data
- 4 Using frailty models for the analysis of multivariate serological data
  - Setting the scene
  - The mass action principle
  - Bivariate correlated frailty models
  - Infection processes
  - Data
  - Discussion & conclusion

# Mixture modelling avoiding the use of thresholds

Bollaerts et al. (2012)

*Statistical Modelling* 2012; **12**(5): 441–462

## Estimating the population prevalence and force of infection directly from antibody titres

K Bollaerts<sup>1</sup>, M Aerts<sup>2</sup>, Z Shkedy<sup>2</sup>, C Faes<sup>2</sup>, Y Van der Stede<sup>4</sup>, P Beutels<sup>3</sup>  
and N Hens<sup>2,3</sup>

<sup>1</sup>Scientific Institute of Public Health, Brussels, Belgium

<sup>2</sup>Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Hasselt University & Katholieke Universiteit Leuven, Diepenbeek, Belgium

<sup>3</sup>Centre for Health Economics Research and Modeling Infectious Diseases, Centre for the Evaluation of Vaccination, Vaccine & Infectious Disease Institute, University of Antwerp, Antwerp, Belgium

<sup>4</sup>Veterinary and Agrochemical Research Centre, Brussels, Belgium

- Outperforms using sensitivity and specificity
- The force of infection underestimated using dichotomous data

- └ An excursion to questions that can be answered by using these data

- └ Examples of use and analysis of serological data

# Back calculation

Borreman et al. (2016)



RESEARCH ARTICLE

## Estimating Time of Infection Using Prior Serological and Individual Information Can Greatly Improve Incidence Estimation of Human and Wildlife Infections

Benny Borremans<sup>1\*</sup>, Niel Hens<sup>2,3</sup>, Philippe Beutels<sup>2</sup>, Herwig Leirs<sup>1</sup>, Jonas Reijnders<sup>1,4</sup>

1 Evolutionary Ecology Group, University of Antwerp, Antwerp, Belgium, 2 Centre for Health Economics Research & Modelling Infectious Diseases (CHERMID), Vaccine & Infectious Disease Institute (VAXINFECTIO), University of Antwerp, Antwerp, Belgium, 3 Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BIOSTAT), Hasselt University, Diepenbeek, Belgium, 4 Department of Engineering Management, University of Antwerp, Antwerp, Belgium

\* bennyborremans@gmail.com



- Back calculation requires knowing the response after vaccination/infection
- When known can improve incidence estimation

# Inferring immunological processes from serological data

Goeyvaerts et al. (2011)

*Biostatistics* (2011), **12**, 2, pp. 283–302  
doi:10.1093/biostatistics/kxq059  
Advance Access publication on September 14, 2010



## **Model structure analysis to estimate basic immunological processes and maternal risk for parvovirus B19**

NELE GOEYVAERTS\*

*Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Hasselt University,  
Agoralaan 1 Gebouw D, B3590 Diepenbeek, Belgium  
nele.goeyvaerts@uhasselt.be*

- Model structure forensics
- Discerning between models will not always be possible - model averaging

# Testing the social contact hypothesis

Santermans et al. (2015)



The social contact hypothesis under the assumption of endemic equilibrium: Elucidating the transmission potential of VZV in Europe



E. Santermans<sup>a,\*</sup>, N. Goeyvaerts<sup>a,b</sup>, A. Melegaro<sup>c</sup>, W.J. Edmunds<sup>d</sup>, C. Faes<sup>a</sup>, M. Aerts<sup>a</sup>,  
P. Beutels<sup>b,e</sup>, N. Hens<sup>a,b</sup>

- Relying on the effective reproduction number as a sanity check
- Acknowledging data selection uncertainty

└ An excursion to questions that can be answered by using these data

└ Examples of use and analysis of serological data

# Estimating MMR vaccine coverage

Wood et al. (2015)

---

ORIGINAL ARTICLE

---

## Estimating Vaccine Coverage from Serial Trivariate Serologic Data in the Presence of Waning Immunity

*James G. Wood,<sup>a</sup> Nele Goeyvaerts,<sup>b,c</sup> C. Raina MacIntyre,<sup>a</sup> Robert I. Menzies,<sup>d,e</sup>  
Peter B. McIntyre,<sup>d,e</sup> and Niel Hens<sup>b,c</sup>*

- Exploiting the multivariate nature of MMR seroepidemiology and trivalent vaccine
- Important to acknowledge time-varying vaccine-induced antibody levels

# Outline

- 1 Introduction
  - Serological data
  - Estimation framework
  - The traditional 'WAIFW' approach
- 2 Estimating inf. disease parameters using serological and social contact data
  - The social contact hypothesis
  - Dimensions of uncertainty
  - Application to VZV
- 3 An excursion to questions that can be answered by using these data
  - Examples of use and analysis of social contact data
  - Examples of use and analysis of serological data
- 4 Using frailty models for the analysis of multivariate serological data
  - Setting the scene
  - The mass action principle
  - Bivariate correlated frailty models
  - Infection processes
  - Data
  - Discussion & conclusion

# Introduction

- Airborne infections (influenza, measles, varicella, ...)
- Transmission parameters estimated from serological data, e.g. force of infection  $\lambda(a)$ , basic reproduction number  $R_0$ , ...
- Force of infection relates to transmission rates through the so-called “mass action principle”; infectious and susceptible individuals are assumed to mix completely in the population
- Estimation of (age-dependent) transmission rates: social mixing patterns

## Goal

Incorporating individual heterogeneity in acquisition of infections in the mass action principle



# Introduction: the mass action principle

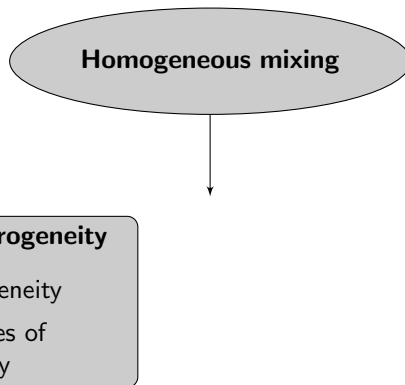
## Homogeneous mixing

Force of infection:

$$\lambda = \beta I^*$$

- $\beta$ : transmission rate
- $I^*$ : number of infected individuals
- $R_0 = \beta ND$  is the basic reproduction number

# Introduction: mass action principle



## Mass action principle - Heterogeneous mixing

- Age-dependent force of infection  $\lambda(a)$  obtained by (Anderson and May, 1991):

$$\lambda(a) = ND \int_0^{\infty} \beta(a, a') \lambda(a') S(a') \phi(a') da'$$

where

- $\beta(a, a')$ : effective contact function
- $S(a')$ : proportion of susceptible individuals of age  $a'$
- $\phi(a')$ : is the age-specific population density

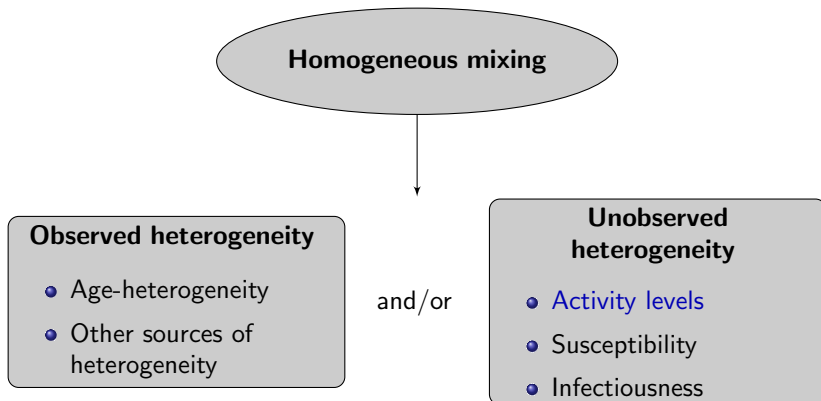
Note that  $ND\lambda(a')S(a')\phi(a') \approx I^*(a')$ .

- Other types of observed heterogeneity, e.g. gender: focus on age

## Mass action principle - Individual heterogeneity

- Individuals differ in
  - propensity to make contacts with others
  - susceptibility to infection
  - infectiousness after infection
- Sources of heterogeneity are typically **unobserved** or **latent**  
see e.g. Coutinho et al. (1999) & Farrington et al. (2001)

# Mass action principle



## Mass action principle - Heterogeneity in Activity Levels

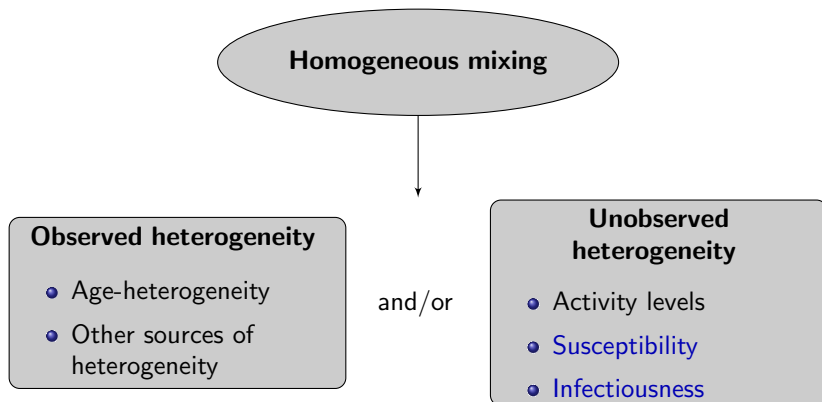
- Conditional force of infection  $\lambda(a, Z)$  obtained by

$$\lambda(a, Z) = ND \int_0^\infty \int_0^\infty \beta(a, Z; a', Z') \lambda(a', Z') S(a' | Z') \phi(a') f(Z') dZ' da'$$

where

- $Z$  and  $Z'$ : **individual frailty terms** distributed according to  $f(\cdot)$
- $\beta(a, Z; a', Z')$ : **augmented effective contact function**
- $\beta(a, Z; a', Z') = Z\beta_0(a, a')Z'$ : **proportionality assumption**
- $D$ : **average infectious period**
- $N$ : **population size**
- **Shared frailty interpretation** for infections sharing transmission routes
- **Basic reproduction number  $R_0$ :**  
 $\{1 + \text{Var}(Z)\} \times \text{dom. eigenval. of } \beta_0^*(a, a') = ND\phi(a)\beta_0(a, a')$

# Mass action principle



- └ Using frailty models for the analysis of multivariate serological data

- └ The mass action principle

## Heterogeneity in Susceptibility and Infectiousness

- Heterogeneity in susceptibility ( $Z_1$ ) and infectiousness ( $Z_2$ ):  
physiological characteristics of infection process
- Hazard of infection  $\lambda(a, Z_1, Z_2) = Z_1 \lambda_0(a)$  implies frailty model  
w.r.t. heterogeneity in susceptibility only
- Basic reproduction number  $R_0$ :  
 $\{1 + \text{Cov}(Z_1, Z_2)\} \times$  dominant eigenvalue of  $\beta_0^*(a, a')$



- └ Using frailty models for the analysis of multivariate serological data

- └ The mass action principle

## Combined model

- Sources of heterogeneity combined in one frailty model
- Augmented effective contact function

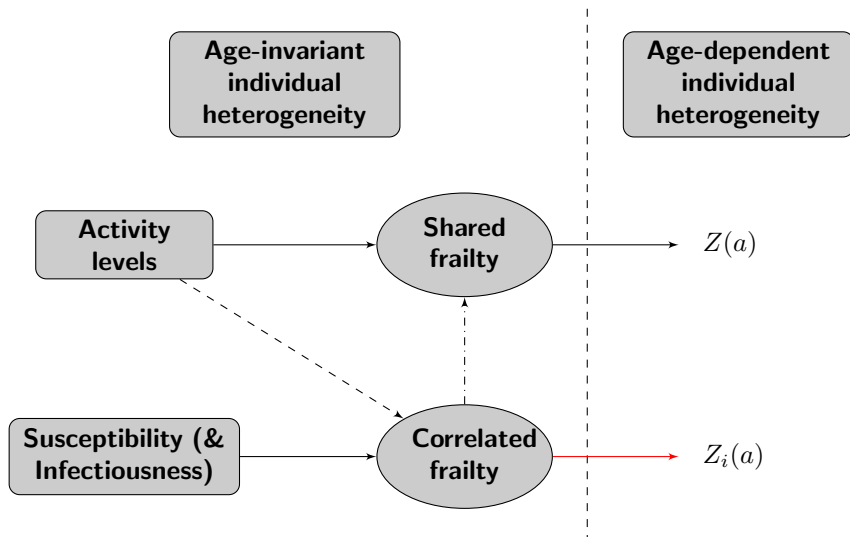
$$\beta(a, Z, Z_1, Z_2; a', Z', Z'_1, Z'_2) = ZZ_1\beta_0(a, a')Z'Z'_2$$

- Force of infection: (multiplicative) frailty model w.r.t. heterogeneity in activity levels & susceptibility

$$\lambda(a, Z, Z_1, Z_2) = ZZ_1\lambda_0(a)$$

- In general, basic reproduction number  $R_0$  equals  $\{1 + \text{Var}(Z)\} \times \{1 + \text{Cov}(Z_1, Z_2)\} \times$  dominant eigenvalue of  $\beta_0^*(a, a')$
- Extension to age-dependent frailties (activity levels)

# Frailty models for bivariate data



## Data application: PVB19 and VZV serology

- Bivariate serological survey data on **parvovirus B19** (PVB19) and **varicella-zoster virus** (VZV) from Belgium anno 2002
- PVB19 causes range of diseases, e.g. **fifth disease** (transmission by **infected respiratory droplets**)
- Primary infection with VZV results in **chickenpox**, maybe reactivated resulting in **herpes zoster** (through **direct close contact with lesions** or **aerosol contact by saliva and sneezing**)
- $n = 3379$  **serological profiles** for the infections under study
- **Bivariate current status data** with  $a$  representing the **age at sampling time** of an individual  $j$  (infections  $i = 1, 2$ ), and

$$Y_{ij} = \begin{cases} 0, & \text{if seronegative,} \\ 1, & \text{if seropositive.} \end{cases}$$

- **Type I interval censored** (current status) data

# Data application: PVB19 and VZV

- Assumptions:

- Social contact hypothesis:  $\beta_0(a, a') = q(a, a'|c)c(a, a')$
- Constant (infection-specific) proportionality factor:  $q_i(a, a'|c) \equiv q_i$
- Gamma frailty distributions with unit mean & frailty variances  $\gamma_{.i}$
- Direct likelihood approach

- Infection processes:

- PVB19: immunizing process (1) or recurrent infection process (2)
- VZV: immunizing infection process

- Models:

- Age-invariant shared frailty models
- Age-dependent shared frailty models: 1C and 2C:  $\gamma_{hi}, h = 1, 2$

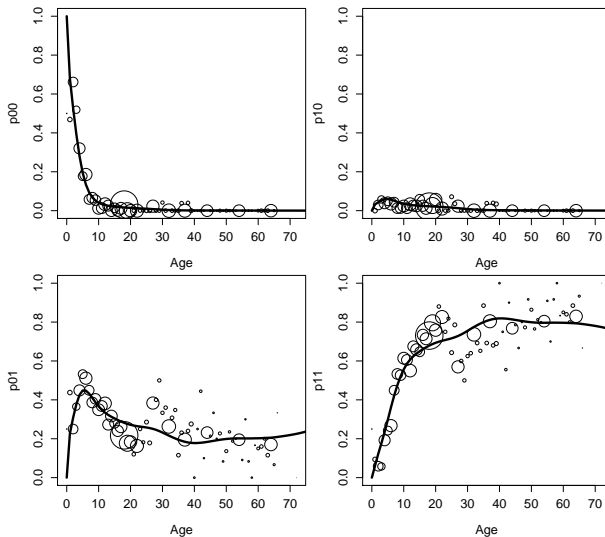
# Bivariate age-invariant shared frailty models

Model				$\hat{R}_0$		AIC	BIC
SGF-1	$q_1$	0.072	[0.069, 0.075]	3.60	[3.35, 3.88]	4937.14	4955.51
	$q_2$	0.200	[0.188, 0.214]	11.64	[10.59, 12.82]		
	$\gamma$	0.152	[0.118, 0.188]				
	$\rho_{12}$	1.000	-				
SGF-2	$q_1$	0.071	[0.068, 0.074]	3.18	[2.97, 3.43]	4869.83	4894.33
	$\sigma$	0.011	[0.008, 0.015]				
	$q_2$	0.173	[0.163, 0.183]	8.98	[8.22, 9.83]		
	$\gamma$	0.032	[0.002, 0.065]				
	$\rho_{12}$	1.000	-				

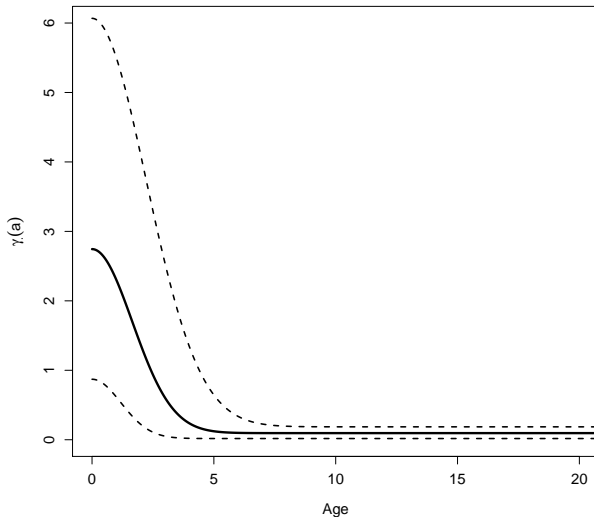
# Bivariate age-dependent shared frailty models

Model				$\hat{R}_0$		AIC	BIC
ADSGF-1-1C	$q_1$	0.072	[0.069, 0.076]	3.60	[3.22, 3.99]	4939.14	4963.64
	$q_2$	0.200	[0.183, 0.221]	11.64	[9.99, 13.49]		
	$\gamma_{\cdot 1}$	0.152	[0.100, 0.210]				
	$\phi_{\cdot 1}$	0.000	[0.000, 0.009]				
ADSGF-1-2C	$q_1$	0.066	[0.062, 0.071]	3.74	[3.15, 4.87]	4912.08	4942.70
	$q_2$	0.235	[0.191, 0.299]	15.65	[11.38, 24.08]		
	$\gamma_{\cdot 1}$	2.918	[1.524, 5.004]				
	$\gamma_{\cdot 2}$	0.233	[0.156, 0.323]				
	$\phi_{\cdot 1}$	0.316	[0.246, 0.425]				
ADSGF-2-1C	$q_1$	0.065	[0.061, 0.070]	2.90	[2.64, 3.49]	4862.93	4893.56
	$\sigma$	0.012	[0.009, 0.016]				
	$q_2$	0.158	[0.141, 0.179]	8.19	[7.15, 10.46]		
	$\gamma_{\cdot 1}$	1.470	[0.415, 3.498]				
	$\phi_{\cdot 1}$	0.330	[0.209, 0.530]				
ADSGF-2-2C	$q_1$	0.066	[0.063, 0.071]	3.30	[2.79, 4.45]	4859.26	4896.01
	$\sigma$	0.011	[0.007, 0.015]				
	$q_2$	0.193	[0.156, 0.257]	11.27	[8.11, 18.90]		
	$\gamma_{\cdot 1}$	2.419	[0.839, 4.960]				
	$\gamma_{\cdot 2}$	0.095	[0.017, 0.186]				
	$\phi_{\cdot 1}$	0.303	[0.226, 0.423]				

# Seroprevalence of PVB19 and VZV



## Age-varying shared frailty variance





# Discussion & conclusion

- Sources of heterogeneity:
  - Untestable assumptions regarding **heterogeneity in infectiousness**
  - **Shared versus correlated frailty models**: sensitivity analysis
- **Age-dependent frailty models**:
  - Age-dependent frailty models improve model fits
  - Specific parametric decay function
- Combined model for **overdispersion** and **individual heterogeneity**:  
→ communicating vessels
- Book work in progress: Abrams, S., Wienke, A., Unkel, S., Hens, N.  
**Frailty Models for Infectious Disease Epidemiology @Wiley**

# Acknowledgements

## Collaborators

- All co-authors and colleagues @CenStat, I-BioStat and @CHERMID, Vaxinfectio
- All collaborating centres nationally and internationally

## Financially

- These projects were partially funded by the Chair of Evidence-based Vaccinology at the University of Antwerp, sponsored by a gift from Pfizer and GSK
- These projects were partially funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 682540 - TransMID)

## Using frailty models for the analysis of multivariate serological data

## Discussion &amp; conclusion

- Anderson, R. and May, R. (1992). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford.
- Beutels, P., Shkedy, Z., Aerts, M., and Van Damme, P. (2006). Social mixing patterns for transmission models of close contact infections: exploring self-evaluation and diary-based data collection through a web-based interface. *Epidemiology and Infection*, 134:1158–1166.
- Bollaerts, K., Aerts, M., Shkedy, Z., Faes, C., Van der Stede, Y., Beutels, P., and Hens, N. (2012). Estimating the population prevalence and force of infection directly from antibody titers. *Statistical Modeling*, 12:441–462.
- Borremans, B., Hens, N., Beutels, P., Leirs, H., and Reijnders, J. (2016). Estimating Time of Infection Using Prior Serological and Individual Information Can Greatly Improve Incidence Estimation of Human and Wildlife Infections. *PLoS Computational Biology*, 12(5):1–18.
- Camarda, G. and Hens, N. (2013). Modelling social contact data: a smoothing constrained approach. *Proceedings of the 13th IWSM*.
- Danon, L., Read, J. M., House, T. A., Vernon, M. C., and Keeling, M. J. (2013). Social encounter networks: characterizing great britain. *Proceedings. Biological sciences*, 280:20131037.
- Dodd, P. J., Looker, C., Plumb, I. D., Bond, V., Schaap, A., Shanaube, K., Muyoyeta, M., Vynnycky, E., Godfrey-Faussett, P., Corbett, E. L., Beyers, N., Ayles, H., and White, R. G. (2016). Age- and sex-specific social contact patterns and incidence of mycobacterium tuberculosis infection. *American journal of epidemiology*, 183:156–166.
- Eames, K., Bansal, S., Frost, S., and Riley, S. (2015). Six challenges in measuring contact networks for use in modelling. *Epidemics*, 10:72–77.
- Edmunds, W., O'Callaghan, C., and Nokes, D. (1997). Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proceedings of the Royal Society B: Biological Sciences*, 264:949–957.
- Farrington, C. and Whitaker, H. (2005). Contact surface models for infectious diseases: estimation from serologic survey data. *Journal of the American Statistical Association*, 100:370 – 379.
- Goeyvaerts, N., Hens, N., Aerts, M., and Beutels, P. (2011). Model structure analysis to estimate basic immunological processes and maternal risk for parvovirus B19. *Biostatistics*, 12(2):283–302.
- Goeyvaerts, N., Hens, N., Ogunjimi, B., Aerts, M., Shkedy, Z., Van Damme, P., and Beutels, P. (2010). Estimating infectious disease parameters from data on social contacts and serological status. *Journal of the Royal Statistical Society Series C*, 59:255–277.
- Goeyvaerts, N., Santermans, E., Potter, G., Torneri, A., Van Kerckhove, K., Willem, L., Aerts, M., Beutels, P., and Hens, N. (2018). Household members do not contact each other at random: implications for infectious disease modelling. *Proceedings. Biological sciences*, 285:20182201.
- Hens, N., Goeyvaerts, N., Aerts, M., Shkedy, Z., Damme, P. V., and Beutels, P. (2009). Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. *BMC Infectious Diseases*, 9:5.
- Hoang, T., Coletti, P., Melegaro, A., Wallinga, J., Grijalva, C. G., Edmunds, J. W., Beutels, P., and Hens, N. (2019). A systematic review of social contact surveys to inform transmission models of close-contact infections. *Epidemiology (Cambridge, Mass.)*, 30:723–736.

## Using frailty models for the analysis of multivariate serological data

## Discussion &amp; conclusion

- Horby, P., Pham, Q. T., Hens, N., Nguyen, T. T. Y., Le, Q. M., Dang, D. T., Nguyen, M. L., Nguyen, T. H., Alexander, N., Edmunds, W. J., Tran, N. D., Fox, A., and Nguyen, T. H. (2011). Social contact patterns in Vietnam and implications for the control of infectious diseases. *PLoS One*, 6(2):e16965.
- Kucharski, A. J., Kwok, K. O., Wei, V. W. I., Cowling, B. J., Read, J. M., Lessler, J., Cummings, D. A., and Riley, S. (2014). The contribution of social behaviour to the transmission of influenza a in a human population. *PLoS Pathog*, 10(6):e1004206.
- Kwok, K. O., Cowling, B. J., Wei, V. W. I., Wu, K. M., Read, J. M., Lessler, J., Cummings, D. A., Peiris, J. S. M., and Riley, S. (2014). Social contacts and the locations in which they occur as risk factors for influenza infection. *Proceedings. Biological sciences*, 281:20140709.
- Luca, G. D., Kerckhove, K. V., Coletti, P., Poletto, C., Bossuyt, N., Hens, N., and Colizza, V. (2018). The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for belgium. *BMC infectious diseases*, 18:29.
- McCaw, J. M., Forbes, K., Nathan, P. M., Pattison, P. E., Robins, G. L., Nolan, T. M., and McVernon, J. (2010). Comparison of three methods for ascertainment of contact information relevant to respiratory pathogen transmission in encounter networks. *BMC Infect Dis*, 10:166.
- Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Scalia Tomba, G., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., and Edmunds, J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine*, 5(3):381–391.
- Rapoport, A. and Horvath, W. J. (1961). A study of a large sociogram. *Behavioral science*, 6:279–291.
- Read, J. M., Edmunds, W. J., Riley, S., Lessler, J., and Cummings, D. A. T. (2012). Close encounters of the infectious kind: methods to measure social mixing behaviour. *Epidemiol Infect*, pages 1–14.
- Salathé, M., Kazandjieva, M., Lee, J. W., Levis, P., Feldman, M. W., and Jones, J. H. (2010). A high-resolution human contact network for infectious disease transmission. *Proc Natl Acad Sci U S A*, 107(51):22020–22025.
- Santermans, E., Goeyvaerts, N., Melegaro, A., Edmunds, W. J., Faes, C., Aerts, M., Beutels, P., and Hens, N. (2015). The social contact hypothesis under the assumption of endemic equilibrium: Elucidating the transmission potential of vzv in europe. *Epidemics*, 11:14–23.
- Santermans, E., Van Kerckhove, K., Azmon, A., John Edmunds, W., Beutels, P., Faes, C., and Hens, N. (2017). Structural differences in mixing behavior informing the role of asymptomatic infection and testing symptom heritability. *Mathematical biosciences*, 285:43–54.
- van de Kastele, J., van Eijkeren, J., and Wallinga, J. (2017). Efficient estimation of age-specific social contact rates between men and women. *Ann. Appl. Stat.*, 11(1):320–339.
- Wallinga, J., Teunis, P., and Kretzschmar, M. (2006). Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American Journal of Epidemiology*, 164:936–944.
- Wood, J. G., Goeyvaerts, N., MacIntyre, C. R., Menzies, R. I., McIntyre, P. B., and Hens, N. (2015). Estimating vaccine coverage from serial trivariate serologic data in the presence of waning immunity. *Epidemiology*, 26(3):381–389.