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

Niel Hens

www.simid.be
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 → seropositive

Example: parvovirus B19 in Belgium
## Setting the scene: varicella zoster virus

<table>
<thead>
<tr>
<th></th>
<th>varicella zoster virus (VZV)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>clinical presentation</strong></td>
<td>primary infection → varicella reactivation → herpes zoster</td>
</tr>
<tr>
<td><strong>transmission</strong></td>
<td>direct or aerosol</td>
</tr>
<tr>
<td><strong>infectious period</strong></td>
<td>about 7 days</td>
</tr>
<tr>
<td><strong>vaccination</strong></td>
<td>no active immunisation in most of Europe</td>
</tr>
<tr>
<td><strong>serological surveys</strong></td>
<td>BE (2002)</td>
</tr>
</tbody>
</table>
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 represents the change (over time and age) in the compartment.

\[
\begin{align*}
\frac{\partial M(a,t)}{\partial a} + \frac{\partial M(a,t)}{\partial t} &= -\alpha(a)M(a,t), \\
\frac{\partial S(a,t)}{\partial a} + \frac{\partial S(a,t)}{\partial t} &= \alpha(a)M(a,t) - (\lambda(a,t) + \mu(a))S(a,t), \\
\frac{\partial I(a,t)}{\partial a} + \frac{\partial I(a,t)}{\partial t} &= \lambda(a,t)S(a,t) - (\nu + \mu(a))I(a,t), \\
\frac{\partial R(a,t)}{\partial a} + \frac{\partial R(a,t)}{\partial t} &= \nu I(a,t) - \mu(a)R(a,t),
\end{align*}
\]

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

The Lexis diagram is a visual tool used in social history and demography to represent the life histories of a group of individuals. It is particularly useful for visualizing the distribution of age and time, allowing for the analysis of when certain events occurred and how they relate to the life stages of individuals. The diagram typically uses a time axis and an age axis, with events marked along the lines that connect these axes. This helps in understanding the timing and sequence of events in relation to the life stages of individuals.
MSIR model

MSIR model: assuming endemic equilibrium

\[
\begin{align*}
\frac{dM(a)}{da} &= -\{\alpha(a) + \mu(a)\} M(a), \\
\frac{dS(a)}{da} &= \alpha(a) M(a) - \{\lambda(a) + \mu(a)\} S(a), \\
\frac{dI(a)}{da} &= \lambda(a) S(a) - \{\gamma(a) + \mu(a)\} I(a), \\
\frac{dR(a)}{da} &= \gamma(a) I(a) - \mu(a) R(a)
\end{align*}
\]

- \(\mu(a)\) = mortality rate
- \(\alpha(a)\) = rate of losing maternal antibodies
- \(\lambda(a)\) = rate at which a susceptible of age \(a\) acquires infection \(\rightarrow\) 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 \( s(a) \) \( \rightarrow \) solve MAP:
    \[
    \lambda(a) = D \int_0^\infty \beta(a, a') \lambda(a') S(a') da'
    \]

- fraction of seropositives \( r(a) \) \( \rightarrow \) evaluate loglikelihood:
  \[
  \ell \ell = \sum_{i=1}^{n} \tilde{w}_i \{ y_i \log[r(a_i)] + (1 - y_i) \log[1 - r(a_i)] \}
  \]
  \[
  y_i = \begin{cases} 
    1, & \text{if seropositive} \\
    0, & \text{if seronegative}
  \end{cases}
  \]
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 $\beta_{i,j}$ from serological data using ML-estimation
- estimate $R_0 = \text{dominant eigenvalue of the next generation matrix with elements:}$

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

<table>
<thead>
<tr>
<th>age class 1</th>
<th>age class 2</th>
<th>age class 3</th>
<th>age class 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>$\beta_4$</td>
<td>$\beta_4$</td>
<td>$\beta_4$</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>$\beta_2$</td>
<td>$\beta_4$</td>
<td>$\beta_4$</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>$\beta_4$</td>
<td>$\beta_3$</td>
<td>$\beta_4$</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>$\beta_4$</td>
<td>$\beta_4$</td>
<td>$\beta_4$</td>
</tr>
</tbody>
</table>
WAIFW approach applied to VZV

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

\[
W_1 = \begin{pmatrix}
\beta_1 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\
\beta_6 & \beta_2 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\
\beta_6 & \beta_6 & \beta_3 & \beta_6 & \beta_6 & \beta_6 \\
\beta_6 & \beta_6 & \beta_6 & \beta_4 & \beta_6 & \beta_6 \\
\beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_5 & \beta_6 \\
\beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6
\end{pmatrix}
\]

\[
W_2 = \begin{pmatrix}
\beta_1 & \beta_1 & \beta_3 & \beta_4 & \beta_5 & \beta_6 \\
\beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 & \beta_6 \\
\beta_3 & \beta_3 & \beta_3 & \beta_4 & \beta_5 & \beta_6 \\
\beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_5 & \beta_6 \\
\beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_6 \\
\beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6
\end{pmatrix}
\]

\[
W_3 = \begin{pmatrix}
\beta_1 & \beta_1 & \beta_1 & \beta_1 & \beta_1 & \beta_1 \\
\beta_2 & \beta_2 & \beta_2 & \beta_2 & \beta_2 & \beta_2 \\
\beta_3 & \beta_3 & \beta_3 & \beta_3 & \beta_3 & \beta_3 \\
\beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_4 \\
\beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \\
\beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6
\end{pmatrix}
\]

\[
W_4 = \begin{pmatrix}
\beta_1 & \beta_1 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\
\beta_1 & \beta_2 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\
\beta_2 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\
\beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\
\beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 \\
\beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6 & \beta_6
\end{pmatrix}
\]

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

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{R}_0$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_1$</td>
<td>8.831</td>
<td>1375.7</td>
</tr>
<tr>
<td>$W_2$</td>
<td>3.512</td>
<td>1372.8</td>
</tr>
<tr>
<td>$W_3$</td>
<td>4.213</td>
<td>1372.8</td>
</tr>
<tr>
<td>$W_4$</td>
<td>8.807</td>
<td>1375.7</td>
</tr>
</tbody>
</table>
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

\[ \beta(a, a') \leftrightarrow \text{WAIFW-matrix} \]
Social contact hypothesis

Social contact hypothesis (Wallinga et al., 2006)

\[ \beta(a, a') \]

- Proportionality constant
- Contact rate

- Serological survey
- Social contact survey

\[ q \cdot c(a, a') \]

Estimation
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)
...
EU mixing patterns

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

Social contact hypothesis (Wallinga et al., 2006)

\[ \beta(a, a') \]

\[ q \cdot c(a, a') \]

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?

<table>
<thead>
<tr>
<th>model</th>
<th>parameter</th>
<th>type of contact</th>
<th>$\hat{q}_k$</th>
<th>95% CI</th>
<th>$\hat{R}_0$</th>
<th>95% CI</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$</td>
<td>$q_1$</td>
<td>all contacts</td>
<td>0.132</td>
<td>[0.103,0.175]</td>
<td>15.69</td>
<td>[12.34,21.41]</td>
<td>1386.618</td>
</tr>
<tr>
<td>$C_2$</td>
<td>$q_2$</td>
<td>close contacts</td>
<td>0.160</td>
<td>[0.126,0.208]</td>
<td>10.24</td>
<td>[8.21,13.68]</td>
<td>1379.581</td>
</tr>
<tr>
<td>$C_3$</td>
<td>$q_3$</td>
<td>close contacts $&gt; 15$ minutes</td>
<td>0.173</td>
<td>[0.133,0.221]</td>
<td>8.68</td>
<td>[6.89,11.34]</td>
<td>1374.958</td>
</tr>
<tr>
<td>$C_4$</td>
<td>$q_4$</td>
<td>close contacts and non-close contacts $&gt; 1$ hour</td>
<td>0.145</td>
<td>[0.113,0.188]</td>
<td>11.73</td>
<td>[9.41,15.95]</td>
<td>1380.354</td>
</tr>
<tr>
<td>$C_5$</td>
<td>$q_5$</td>
<td>close contacts $&gt; 15$ minutes and non-close contacts $&gt; 1$ hour</td>
<td>0.156</td>
<td>[0.119,0.204]</td>
<td>10.40</td>
<td>[8.05,14.10]</td>
<td>1376.068</td>
</tr>
</tbody>
</table>
\( q = \text{constant?} \)

- \( q \) 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

\[
\begin{align*}
\beta(a, a') & \quad q(a, a') \cdot c(a, a') \\
\text{proportionality factor} & \quad \| \quad \text{contact rate}
\end{align*}
\]

various options: back to issues with identifiability \( \rightarrow \) model averaging
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
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.

  \[
  \text{social contact hypothesis: } q(a, a') = q
  \]
Systematic review

Hoang et al. (2019):

- diary-based approach & face-to-face interviews
- data sharing initiative: www.socialcontactdata.org & socialmixr-package (R software)
Estimating contact rates

- **General guidelines**
  - Hens and Wallinga (Wiley, StatsRef 2019)
  - Wallinga, van de Kassteele and Hens (HIDDA, 2019)

- **Smoothing approaches**
  - van de Kassteele et al. (2017): smoothing INLA
  - Camarda and Hens (2013), Vandendijck et al. (2018): clever choice of axes to smooth over
Behavioural change: regular holiday periods

Luca et al. (2018):

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

Santermans et al. (2017):

Illness has an impact on activity (contacts & mobility)
Different disease dynamics
Inferring networks from social contact data

Goeyvaerts et al. (2018):

- Household members do not mix at random
- Combining household and egocentric data proves to be useful
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 Bollaerts¹, M Aerts², Z Shkedy², C Faes², Y Van der Stede⁴, P Beutels³ and N Hens²,³

¹Scientific Institute of Public Health, Brussels, Belgium
²Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Hasselt University & Katholieke Universiteit Leuven, Diepenbeek, Belgium
³Centre for Health Economics Research and Modeling Infectious Diseases, Centre for the Evaluation of Vaccination, Vaccine & Infectious Disease Institute, University of Antwerp, Antwerp, Belgium
⁴Veterinary and Agrochemical Research Centre, Brussels, Belgium

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

Borremans et al. (2016)

- 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)

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)

- Relying on the effective reproduction number as a sanity check
- Acknowledging data selection uncertainty
Estimating MMR vaccine coverage

Wood et al. (2015)

- 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

Homogeneous mixing

Observed heterogeneity
- Age-heterogeneity
- Other sources of heterogeneity
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

Homogeneous mixing

- Observed heterogeneity
  - Age-heterogeneity
  - Other sources of heterogeneity

- Unobserved heterogeneity
  - Activity levels
  - Susceptibility
  - Infectiousness
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

Homogeneous mixing

Observed heterogeneity
- Age-heterogeneity
- Other sources of heterogeneity

Unobserved heterogeneity
- Activity levels
- Susceptibility
- Infectiousness
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 $R_0$:
  \[ \{1 + \text{Cov}(Z_1, Z_2)\} \times \text{dominant eigenvalue of } \beta_0^*(a, a') \]
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 \( \{1 + \text{Var}(Z)\} \times \{1 + \text{Cov}(Z_1, Z_2)\} \times \text{dominant eigenvalue of } \beta_0^*(a, a') \)

- **Extension to age-dependent frailties** (activity levels)
Frailty models for bivariate data

**Age-invariant individual heterogeneity**

- Activity levels
- Susceptibility (& Infectiousness)

**Shared frailty**

**Correlated frailty**

**Age-dependent individual heterogeneity**

\[ Z(a) \]

\[ Z_i(a) \]
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

<table>
<thead>
<tr>
<th>Model</th>
<th>( q_1 )</th>
<th>( q_2 )</th>
<th>( \gamma )</th>
<th>( \rho_{12} )</th>
<th>( \hat{R}_0 )</th>
<th>( \text{AIC} )</th>
<th>( \text{BIC} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SGF-1</td>
<td>0.072</td>
<td>0.200</td>
<td>0.152</td>
<td>1.000</td>
<td>3.60</td>
<td>[3.35, 3.88]</td>
<td>4937.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.64</td>
<td>[10.59, 12.82]</td>
<td></td>
</tr>
<tr>
<td>SGF-2</td>
<td>0.071</td>
<td>0.011</td>
<td>0.173</td>
<td>0.032</td>
<td>8.98</td>
<td>[8.22, 9.83]</td>
<td>4869.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.18</td>
<td>[2.97, 3.43]</td>
<td></td>
</tr>
</tbody>
</table>
### Bivariate age-dependent shared frailty models

<table>
<thead>
<tr>
<th>Model</th>
<th>$q_1$</th>
<th>$q_2$</th>
<th>$\gamma_1$</th>
<th>$\phi_1$</th>
<th>$\hat{R}_0$</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADSGF-1-1C</td>
<td>0.072</td>
<td>0.200</td>
<td>0.152</td>
<td>0.000</td>
<td>3.60</td>
<td>[3.22, 3.99]</td>
<td>4939.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$q_1$</td>
<td></td>
<td></td>
<td>[0.069, 0.076]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$q_2$</td>
<td></td>
<td></td>
<td>[0.183, 0.221]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\gamma_1$</td>
<td></td>
<td></td>
<td>[0.100, 0.210]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\phi_1$</td>
<td></td>
<td></td>
<td>[0.000, 0.009]</td>
<td></td>
</tr>
<tr>
<td>ADSGF-1-2C</td>
<td>0.066</td>
<td>0.235</td>
<td>2.918</td>
<td>0.316</td>
<td>3.74</td>
<td>[3.15, 4.87]</td>
<td>4912.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$q_1$</td>
<td></td>
<td></td>
<td>[0.062, 0.071]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$q_2$</td>
<td></td>
<td></td>
<td>[0.191, 0.299]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\gamma_1$</td>
<td></td>
<td></td>
<td>[1.524, 5.004]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\gamma_2$</td>
<td></td>
<td></td>
<td>[0.156, 0.323]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\phi_1$</td>
<td></td>
<td></td>
<td>[0.246, 0.425]</td>
<td></td>
</tr>
<tr>
<td>ADSGF-2-1C</td>
<td>0.065</td>
<td>0.158</td>
<td>1.470</td>
<td>0.330</td>
<td>2.90</td>
<td>[2.64, 3.49]</td>
<td>4862.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$q_1$</td>
<td></td>
<td></td>
<td>[0.061, 0.070]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$q_2$</td>
<td></td>
<td></td>
<td>[0.141, 0.179]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\gamma_1$</td>
<td></td>
<td></td>
<td>[0.415, 3.498]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\gamma_2$</td>
<td></td>
<td></td>
<td>[0.209, 0.530]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\phi_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSGF-2-2C</td>
<td>0.066</td>
<td>0.193</td>
<td>2.419</td>
<td>0.303</td>
<td>3.30</td>
<td>[2.79, 4.45]</td>
<td>4859.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$q_1$</td>
<td></td>
<td></td>
<td>[0.063, 0.071]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$q_2$</td>
<td></td>
<td></td>
<td>[0.156, 0.257]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\gamma_1$</td>
<td></td>
<td></td>
<td>[0.839, 4.960]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\gamma_2$</td>
<td></td>
<td></td>
<td>[0.017, 0.186]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\phi_1$</td>
<td></td>
<td></td>
<td>[0.226, 0.423]</td>
<td></td>
</tr>
</tbody>
</table>
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)


