On modelling heterogeneity in the acquisition of infectious diseases

Niel Hens[1,2]
Steven Abrams[1]

Smiddy 2013
Zürich

[1] I-Biostat, Hasselt University, Hasselt, Belgium,
[2] CHERMID, University of Antwerp, Antwerp, Belgium,

niel.hens@uhasselt.be - niel.hens@ua.ac.be

Individual heterogeneity

- Coutinho et al. (1999): first to explicitly account for individual heterogeneity in the acquisition of infectious diseases
- Farrington et al. (2001): shared gamma frailty model for bivariate serological data (Measles and Mumps, UK)
- Hens et al. (2009): more flexible correlated gamma frailty model outperforms shared frailty model at the cost of assuming parametric hazard (Hepatitis A and B, Belgium)
- Traditional shared and correlated gamma frailty models based on the assumption of lifelong immunity after recovery
- Aim: integrate mechanistic models and traditional frailty models to encompass disease dynamics for non-immunizing infections, comprising potential reinfections

Univariate frailty model: notations

- Consider univariate current status data (y, a), where a represents the age of an individual and

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

- Type I interval censored (current status) data
- Z denotes individual's frailty term with respect to single infection
- Proportional hazards assumption: $\lambda(a, Z) = Z\lambda_0(a)$
- $\lambda_0(a)$: age-dependent baseline force of infection
- Assumption of endemic equilibrium: unit of time a instead of t

Univariate frailty model

- Conditional survival function $S(a|Z)$ using proportional hazards:

$$S(a|Z) = \exp\left(-\int_0^a \lambda(u,Z)du\right) = \exp\left(-Z\int_0^a \lambda_0(u)du\right)$$

- Unconditional survival function $S(a)$:

$$S(a) = \mathbf{L}\left(M_0(a)\right),$$

with $M_0(a) = \int_0^a \lambda_0(u)du$ and $\mathbf{L}(.)$ the Laplace transform of Z

Maximum likelihood estimation

- Loglikelihood contribution for univariate current status data (y, a):

$$ll(y, a|\boldsymbol{\beta}, \boldsymbol{\psi}) = y \log\left(1 - S(a|\boldsymbol{\beta}, \boldsymbol{\psi})\right) + (1 - y) \log\left(S(a|\boldsymbol{\beta}, \boldsymbol{\psi})\right)$$

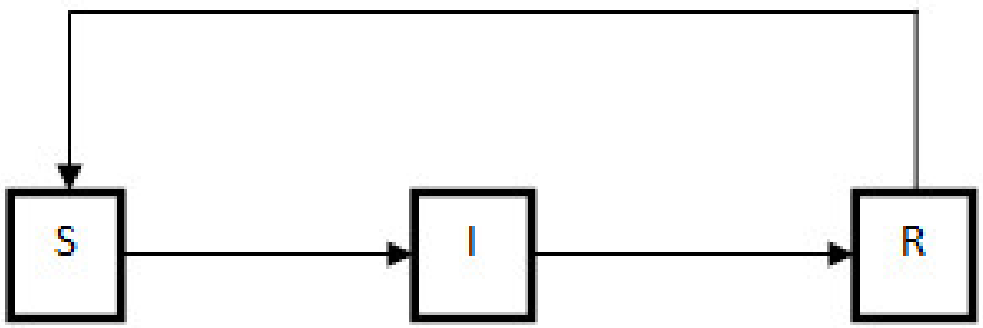
- $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$: vectors of unknown parameters associated with baseline force of infection and frailty distribution, respectively

Extensions

- This is then easily extended to the setting of
  - a shared frailty (Farrington et al. 2001)
  - a correlated frailty (Hens et al. 2009)
  - a shared frailty with age-dependent shape parameter (Farrington et al. 2012, 2013)
- Goals:
  - extending these methods for non-immunizing infections
  - integrating correlated and age-dependent frailties

Univariate SIRS frailty model

- Previous expressions for $S(a|Z)$ and $S(a)$ not valid for non-immunizing infections
- Formulas derived based on mathematical transmission models
- Non-immunizing infection with SIRS transmission dynamics



Mathematical transmission model

- Set of ordinary differential equations (ODEs) in time homogeneous setting:

$$\begin{aligned}
\frac{dS(a|Z)}{da} &= -\lambda(a,Z)S(a|Z) + \sigma(a)R(a|Z), \\
\frac{dI(a|Z)}{da} &= \lambda(a,Z)S(a|Z) - \gamma I(a|Z), \\
\frac{dR(a|Z)}{da} &= \gamma I(a|Z) - \sigma(a)R(a|Z).
\end{aligned}$$

- S, I, R: age-specific proportions of susceptible, infected and recovered individuals, respectively
- λ, σ, γ: force of infection, replenishment rate and recovery rate, respectively

Solution of the ODEs

- Solving the set of ODEs using $S(a|Z) \approx 1 - R(a|Z)$ yields

$$S(a|Z) = \exp\left(-\int_0^a \{\lambda(u,Z) + \sigma(u)\}\, du\right) + \int_0^a \sigma(u) \exp\left(-\int_u^a \{\lambda(v,Z) + \sigma(v)\}\, dv\right) du$$

- Integral part in $S(a|Z)$ no closed-form expression
- Unconditional survival function $S(a)$ derived under proportional hazards assumption by taking expectation with respect to Z

Unconditional survival function

- Unconditional survival $S(a)$:

$$S(a) = \mathbf{L}\left(M_0(a)\right)\exp\left(-\int_0^a \sigma(u)du\right) + \int_0^a \sigma(u)\mathbf{L}\left(M_0(a)-M_0(u)\right)\exp\left(-\int_u^a \sigma(v)dv\right)du$$

- Numerical integration techniques required to approximate the integral part in expression for $S(a)$

Identifiability

A parametric baseline hazard

To ensure identifiability we need to use parametric baseline hazards for both infections. We will therefore use a mechanistic model based on the mass action principle (Farrington et al. 2001).

Social contact hypothesis

We extend this model by using data from social contact surveys providing an empirical basis for underlying mixing patterns.

The mass action principle

- Short infectious period D:

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

with population size N, augmented contact function $\beta(a,Z;a',Z')$, and $\phi(a') = \exp(-\int_0^{a'} \mu(u)du)$, with $\mu(a)$: age-dependent mortality rates

- multiplicative decomposition (Farrington et al., 2001)

$$\beta(a,Z;a',Z') = ZZ'\beta_0(a,a'),$$

which implies proportional hazards assumption with respect to the force of infection

Social contact hypothesis

- Furthermore, baseline contact function $\beta_0(a, a')$ consists of two components (social contact hypothesis):

$$\beta_0(a, a') = q(a, a'|c)c(a, a'),$$

  with $q(a, a'|c)$ proportionality factor and $c(a, a')$ age-dependent contact rates
- Estimating $c(a, a')$:
  - Data on social mixing in Belgium based on POLYMOD survey
  - Large-scaled European prospective survey between May 2005 and September 2006 on textbfcontact behaviour
  - Annual contact rates $c(a, a')$ estimated using bivariate smoothing approach (Goeyvaerts et al., 2010)

Solving the mass action principle

- Mass action principle does not exhibit a closed-form solution
- Turning to discrete age-intervals, a piecewise constant force of infection can be estimated using an iterative procedure (Kanaan and Farrington, 2005)
- Basic reproduction number R_0 is estimated as $\left(1+\sigma_f^2\right)$ times the dominant eigenvalue of the function:

$$\beta_0^*(a,a') = \frac{ND}{L}\exp\left(-\int_0^a \mu(u)du\right)\beta_0(a,a')$$

Data: VZV and B19

- 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 reactived resulting in herpes zoster (through direct close contact with lesions or aerosol contact by saliva and sneezing)
- $n = 2974$ complete serological profiles for both infections

Data application

- Assumptions:
  - Type I mortality rates: $\mu(a) = 0$ if $a \leq L$ and $\mu(a) = \infty$ otherwise
  - Absence of maternally derived antibodies, no disease-related mortality
  - Constant proportionality factor $q(a, a'|c) \equiv q$
  - Gamma frailty distribution with unit mean and variance σ_{if}^2, $i = 1,2$ for infection 1,2, respectively
- Parameters:
  - $N = 9943749$, $L = 80$ years
  - PVB19: $D = 6$ days, VZV: $D = 7$ days
- Univariate frailty models
- Replenishment rate $\sigma(a)$ assumed to be constant (SIRS-SIR models) or dichotomous (SIRSext-SIR models) (cut-off: 35 yrs)

Results: Univariate gamma frailty models

- Bivariate serological data while assuming independence: product of univariate likelihoods

Model				$\hat{R}_0$		$\hat{R}$		AIC	BIC
SIR-SIR	q_{10}	0.086	[0.079, 0.094]	5.27	[4.47, 6.22]	1.831	[1.568, 2.142]	4506.27	4530.26
	σ^2_{1f}	0.435	[0.316, 0.560]						
	q_{20}	0.169	[0.159, 0.179]	8.40	[7.89, 8.92]	1.149	[1.137, 1.162]		
	σ^2_{2f}	3.0e-6	[3.0e-6, 3.0e-6]						
	ρ_{12}	0.000	-						
SIRS-SIR	q_{10}	0.071	[0.068, 0.074]	3.03	[2.91, 3.15]	1.059	[1.054, 1.064]	4481.84	4511.82
	σ	0.011	[0.008, 0.015]						
	σ^2_{1f}	3.0e-6	[3.0e-6, 3.0e-6]						
	q_{20}	0.169	[0.159, 0.179]	8.40	[7.90, 8.93]	1.149	[1.137, 1.162]		
	σ^2_{2f}	3.0e-6	[3.0e-6, 3.0e-6]						
	ρ_{12}	0.000	-						
SIRSext-SIR	q_{10}	0.072	[0.069, 0.074]	3.05	[2.93, 3.17]	1.069	[1.060, 1.077]	4477.00	4512.99
	σ_1	0.017	[0.012, 0.023]						
	σ_2	0.008	[0.005, 0.012]						
	σ^2_{1f}	3.0e-6	[3.0e-6, 3.0e-6]						
	q_{20}	0.169	[0.159, 0.179]	8.40	[7.90, 8.93]	1.149	[1.137, 1.162]		
	σ^2_{2f}	3.0e-6	[3.0e-6, 3.0e-6]						
	ρ_{12}	0.000	-						

Results: Bivariate shared gamma frailty models

- Bivariate shared frailty model extended as well to encompass SIRS transmission dynamics

Model				$\hat{R}_0$		$\hat{R}$		AIC	BIC
SIR-SIR	q_{10}	0.073	[0.069, 0.077]	3.59	[3.27, 3.90]	1.278	[1.188, 1.368]	4537.28	4555.27
	q_{20}	0.209	[0.189, 0.232]	12.07	[10.47, 13.74]	1.516	[1.370, 1.664]		
	σ_f^2	0.158	[0.103, 0.210]						
	ρ_{12}	1.000	-						
SIRS-SIR	q_{10}	0.072	[0.068, 0.075]	3.17	[2.94, 3.43]	1.106	[1.052, 1.178]	4477.98	4501.97
	σ	0.011	[0.007, 0.014]						
	q_{20}	0.177	[0.162, 0.196]	9.15	[8.07, 10.53]	1.221	[1.140, 1.333]		
	σ_f^2	0.036	[3.5e-6, 0.086]						
	ρ_{12}	1.000	-						
SIRSext-SIR	q_{10}	0.072	[0.069, 0.075]	3.13	[2.96, 3.38]	1.093	[1.058, 1.167]	4474.39	4504.38
	σ_1	0.016	[0.010, 0.022]						
	σ_2	0.008	[0.005, 0.012]						
	q_{20}	0.173	[0.161, 0.192]	8.82	[8.03, 10.20]	1.189	[1.136, 1.301]		
	σ_f^2	0.021	[3.4e-6, 0.072]						
	ρ_{12}	1.000	-						

Results: Bivariate correlated gamma frailty models

- Bivariate correlated frailty model extended as well to encompass SIRS transmission dynamics

Model				$\hat{R}_0$		$\hat{R}$		AIC	BIC
SIR-SIR	q_{10}	0.086	[0.079, 0.094]	5.26	[4.47, 6.20]	1.827	[1.567, 2.135]	4505.62	4535.61
	q_{20}	0.180	[0.163, 0.200]	9.40	[8.21, 10.92]	1.246	[1.147, 1.376]		
	σ^2_{1f}	0.433	[0.314, 0.558]						
	σ^2_{2f}	0.048	[2.7e-6, 0.099]						
	ρ_{12}	0.332	[0.002, 0.499]						
SIRS-SIR	q_{10}	0.072	[0.068, 0.075]	3.17	[2.95, 3.43]	1.106	[1.054,1.178]	4481.98	4517.96
	σ	0.011	[0.007, 0.014]						
	q_{20}	0.177	[0.162, 0.197]	9.15	[8.07, 10.54]	1.221	[1.141, 1.337]		
	σ^2_{1f}	0.036	[4.8e-6, 0.086]						
	σ^2_{2f}	0.036	[4.8e-6, 0.086]						
	ρ_{12}	1.000	[0.999, 1.000]						
SIRSext-SIR	q_{10}	0.071	[0.068, 0.074]	3.08	[2.92, 3.34]	1.077	[1.042, 1.154]	4478.53	4520.51
	σ_1	0.017	[0.010, 0.022]						
	σ_2	0.009	[0.005, 0.012]						
	q_{20}	0.173	[0.162, 0.193]	8.82	[8.09, 10.29]	1.188	[1.138, 1.304]		
	σ^2_{1f}	0.021	[4.8e-6, 0.073]						
	σ^2_{2f}	0.021	[4.8e-6, 0.073]						
	ρ_{12}	1.000	[0.999, 1.000]						

Age-dependent frailty model

- The age-dependent shared gamma frailty (ADSGF) (Farrington et al., 2012, 2013)

$$Z_i(a) = \prod_{j=1}^{k} \left[1 + (Z_{ij} - 1)\, h_{ij}(a)\right], \qquad 0 \leq h_{ij}(a) \leq 1,$$

where

$$h_{ij}(a) = \exp\left(-(\phi_{ij} a)^l\right), \qquad \phi_{ij} \geq 0.$$

If $l = 2$ and $Z_{ij} = Z_j$, $j = 1, 2, \ldots, k$ are independent gamma distributed random variables with unit mean and frailty variance σ_j^2, it follows:

$$\begin{aligned} \sigma_i^2(a) = \mathrm{Var}\left(Z_i(a)\right) &= \prod_{j=1}^{k} \mathrm{E}\left[\left(1 + (Z_j - 1)\, h_{ij}(a)\right)^2\right] - \prod_{j=1}^{k} \left(\mathrm{E}\left[1 + (Z_j - 1)\, h_{ij}(a)\right]\right)^2 \\ &= \prod_{j=1}^{k} \left(1 + h_{ij}(a)^2 \sigma_j^2\right) - 1 \end{aligned}$$

Age-dependent frailty model

- Example I: One component ($k = 1$) and an infection-invariant exponential decay rate ($\phi_{ij} \equiv \phi_j$) (ADSGF-1):

$$\sigma_i^2(a) = h_{i1}(a)^2 \sigma_1^2,$$

where

$$h_{i1}(a) = \exp\left(-(\phi_1 a)^2\right)$$

- Example II: Two-component multiplicative models using $h_{i2}(a) = 1, \forall a$. The decay rates are assumed to differ, denoted by ϕ_{1j} and ϕ_{2j} (ADSGF-2).

$$\sigma_i^2(a) = h_{i1}(a)^2 \sigma_1^2 \left(1 + \sigma_2^2\right) + \sigma_2^2$$

The frailty variance decreases from σ_1^2 to σ_2^2 as $a \to \infty$.

Age-dependent frailty model

- The piecewise constant shared gamma frailty (PCSGF) model (Paik, 1994):

$$Z_i(a) = \sum_{j=1}^{k} I_j(a) Z_{ij},$$

where $I_j(a)$ equals one if $a \in \left[a_{[j]}, a_{[j+1]}\right)$, and zero otherwise.

Age-dependent correlated gamma frailty models

- The age-dependent correlated gamma frailty (ADCGF) models combine the multiplicative model proposed by Farrington et al. (2012) with the additive decomposition introduced by Yashin (1995):

$$\begin{aligned} Z_i(a) &= \left[1 + (Z_{i1} - 1)\, h_{i1}(a)\right], \\ Z_{i1} &= \sigma_i^2 \left(Y_0^* + Y_i^*\right), \\ h_{i1}(a) &= \exp\left(-\left(\phi_{i1} a\right)^l\right), \end{aligned}$$

  where the components Y_l^* are independent gamma distributed random variables.

- For identifiability reasons: $\phi_{11} = \phi_{21} \equiv \phi_1$ (ADCGF).

Data: Hepatitis A & B

- Bivariate serological survey data on hepatitis A (HAV) and Hepatitis B (HBV) from Belgium anno 1993-1994
- The main transmission route for hepatitis A is foodborne or faeco-oral
- For hepatitis B it is sexual or bloodborne
- $n = 3787$ complete serological profiles for both infections
- Analyzed in Hens et al. (2009) using a correlated frailty model

Age-dependent correlated gamma frailty models

Gompertz Baseline

	SGF	CGF	PCSGF	ADSGF-1	ADSGF-2	ADCGF-1	ADCGF-2
a_1	0.012 (0.001)	0.007 (0.001)	0.028 (0.008)	0.073 (0.026)	0.136 (0.091)	0.119 (0.061)	0.127 (0.079)
b_1	0.037 (0.005)	0.105 (0.018)	0.010 (0.007)	-0.011 (0.006)	-0.020 (0.009)	-0.018 (0.008)	-0.019 (0.009)
a_2	0.002 (3E-4)	0.002 (4E-4)	0.002 (4E-4)	0.003 (0.001)	0.003 (0.001)	0.003 (0.001)	0.003 (0.001)
b_2	-0.000 (0.007)	0.002 (0.007)	-0.002 (0.008)	-0.006 (0.007)	-0.009 (0.008)	-0.006 (0.007)	-0.008 (0.008)
σ_1	0.723 (0.086)	1.632 (0.028)	3.698 (0.683)	5.771 (0.816)	6.448 (1.014)	6.332 (0.451)	6.362(0.452)
σ_2	0.723 (0.086)	1.135 (0.093)	2.429 (0.544)	5.771 (0.816)	6.448 (1.014)	5.649 (0.503)	6.040(0.548)
σ_3	-	-	0.001 (3E-4)	-	-	-	-
σ_4	-	-	7.962 (6.787)	-	-	-	-
ϕ_1	-	-	-	0.034 (0.005)	0.026 (0.007)	0.027 (0.006)	0.026 (0.007)
ϕ_2	-	-	-	-	0.045 (0.011)	-	0.038 (0.020)
ρ	1.000 (-)	0.696 (0.056)	1.000 (-)	1.000 (-)	1.000 (-)	0.931 (0.097)	0.948 (0.063)
$-2ll$	5687.0	5653.4	5605.3	5617.3	5614.5	5614.6	5614.3

Conclusion

Non-immunizing infections:

- SIRS-SIR frailty models outperform traditional SIR frailty models
- Shared SIRS-SIR gamma frailty models perform best based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)
- Frailty variance seriously overestimated when assuming lifelong immunity for PVB19 compared to waning immunity assumption
- Upper bound for correlation coefficient in correlated gamma frailty model elevated
- Consequently, estimates for R_0 and R inflated under SIR dynamics

Conclusion

Age-dependent frailties:

- Correlated versions seem to outperform shared versions
- Overall the shared piecewise-constant age-dependent version provides the best fit

Epilogue

- Combine social contact hypothesis with age-dependent frailty approach for non-immunizing infections
- Investigate performance of models in context of other infections
- Extend SIRS frailty models to analyse serial seroprevalence data
- Use frailty-dependent replenishment rates $\sigma(a, Z)$
- Taking imperfect testing into account using principles of direct estimation (Hens et al., 2012)

- Research project financed by grant BOF11NI31 (Hasselt University)

- Coutinho, F. A. B., Massad, E., Lopez, L. F., Burattini, M. N., Struchiner, C., and Azevedo-Neto, R. (1999). Modelling heterogeneities in individual frailties in epidemic models. *Mathematical and Computer Modelling*, 30:97-115.
- Farrington, C. P., Kanaan, M. N., and Gay, N. J. (2001). Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. *Applied Statistics*, 50:251-292.
- Farrington, C. P., Whitaker, H. J., Unkel, S., and Pebody, R. (2012). Correlated infections: quantifying individual heterogeneity in the spread of infectious diseases. *American Journal of Epidemiology*, preprint.
- Goeyvaerts, N., Hens, N., Ogunjimi, B., Aerts, M., Shkedy, Z., Van Damme, P., and Beutels, P. (2010). Estimating transmission parameters and the basic reproduction number using social contact data and serological data on varicella zoster virus in Belgium. *Journal of the Royal Statistical Society Series C*, 59:255-277.
- 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*, 12:283-302.
- Hens, N. Wienke, A., Aerts, M., and Molenberghs, G. (2009). The correlated and shared gamma frailty model for bivariate current status data: An illustration for cross-sectional serological data. *Statistics in Medicine*, 27:2785-2800.
- Kanaan, M. N. and Farrington, C. P. (2005). Matrix models for childhood infections: a Bayesian approach with applications to rubella and mumps. *Epidemiology and Infection*, 133:1009-1021.