On the correlated frailty model for bivariate current status data with applications in infectious disease epidemiology

Niel Hens[1], Andreas Wienke[2]
Marc Aerts[1] and Geert Molenberghs[1]

[1] Center for Statistics, Hasselt University, Belgium
[2] Martin-Luther-Universität Halle-Wittenberg, Halle, Germany

Warwick, April 15, 2008

Outline

Airborne Infections

- Examples: measles, mumps, varicella, parvovirus B19, ...
- Compartmental models: SIR, MSIR, MSEIR, ...
- Basic reproduction number R_0

  $R_0 > 1 \rightarrow$ epidemic

  $R_0 < 1 \rightarrow$ eradication

- Infectious disease control - vaccination

  Critical Vaccination Coverage: $1 - 1/R_0$

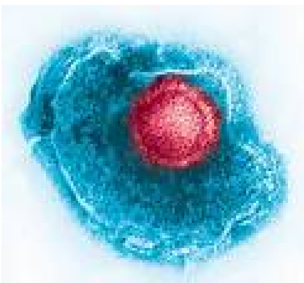
Airborne Infections

- Crucial parameter: transmission rate: 1st, 2nd moment
- Surrogate: force of infection and the associated heterogeneity
- Problem: ‘current status’ data rather than ‘time to event’ data
- Solution?
  - Estimating the force of infection from current status data: Muench (1934); Grenfell and Anderson (1985); Keiding (1991), ...
  - Estimating heterogeneity: Farrington et al. (2001); Sutton et al. (2006): shared frailty

Estimating the FOI from Serological Data

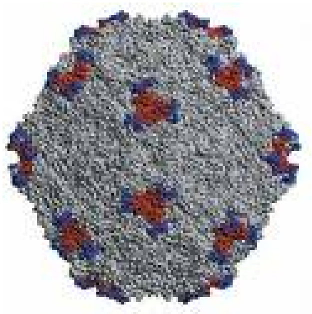
- Varicella Zoster Virus and Parvovirus B19
- As a proxy for other airborne infections
- No vaccination yet (Europe)
- Other diseases (CMV, EBV, ...)
- Data: Belgium, England & Wales, Finland, Italy and Poland
- Age-range: 0-20 years

Varicella Zoster Virus (VZV)



- **Varicella**
  - Primary VZV infection results in chickenpox
  - Transmission: direct or aerosol contact
  - When infected, infectious for about 7 days
  - Incubation period of two weeks
  - Reactivation later in life ($10 - 20\%$): herpes zoster or shingles
  - Disease burden: zoster: 25% is in constant pain

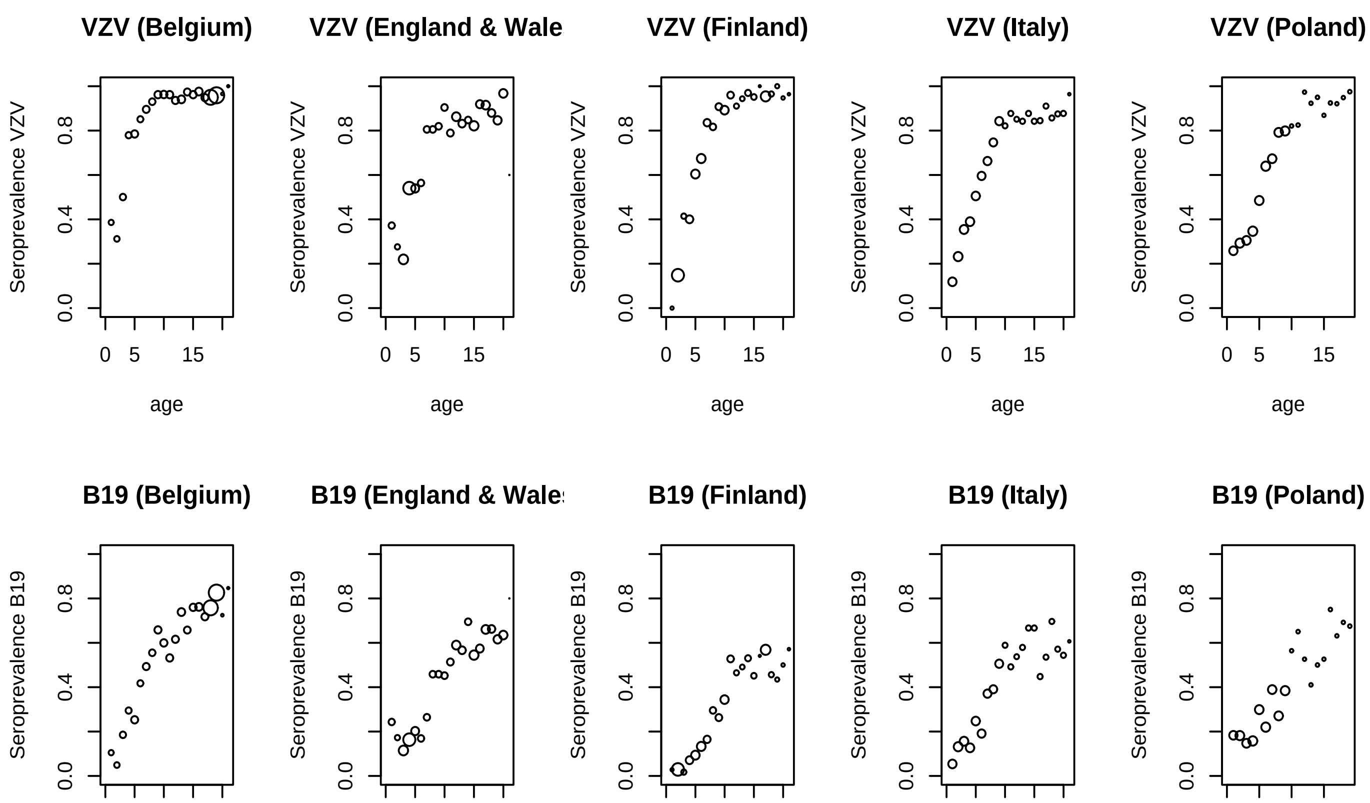
Parvovirus B19 (B19)



Baby with the typical "slapped-cheek" rash, which is characteristic of fifth disease.

- **B19**
  - B19 infection causes the so-called ‘fifth disease’, a mild rash illness (‘slapped-cheek’ rash)
  - Transmission: respiratory droplets
  - Infectious during the incubation period (± 14 days)
  - Disease burden: for pregnant women there is a potential for the newborn to have severe anemia, possibly leading to miscarriage.

Belgium, England & Wales, Finland, Italy, Poland



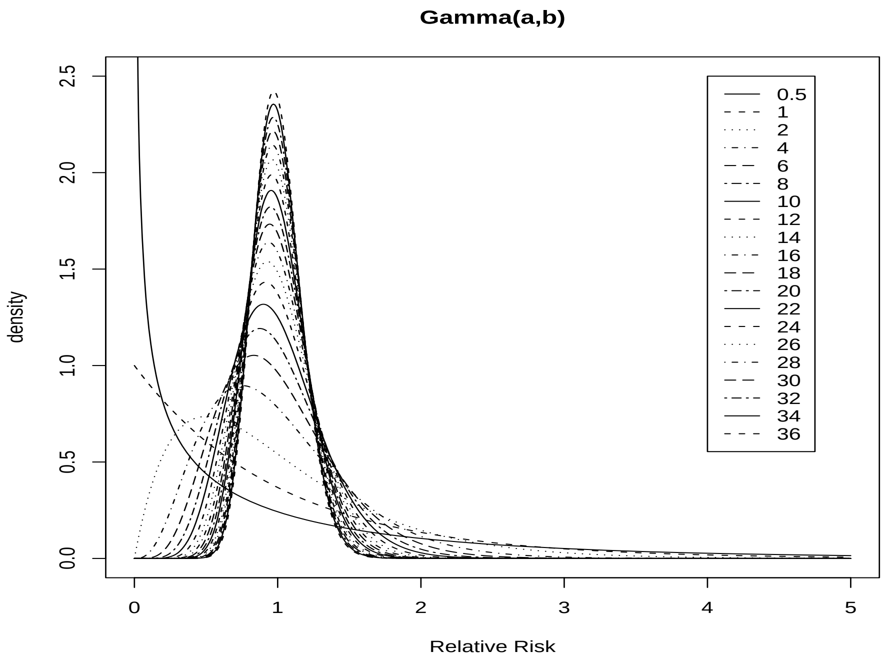
Frailties

Coutinho et al. (1999)

Persistent differences among individuals in their susceptibility, propensity, or relative risk with regard to the acquisition of infections.

- Individuals have different frailties
- The most frail individuals will experience the effect of the event earlier
- Vaupel et al. (1979); Aalen (1988): $\lambda(a, \omega) = \omega\lambda(a, 1)$ with ω a nonnegative *mixing variable*
- Often $\mathrm{E}(\omega) = 1$ is chosen, e.g. $\omega \sim \Gamma(\theta, 1/\theta)$

$\omega \sim \Gamma(\theta, 1/\theta)$



$$f(x) = \frac{1}{\Gamma(\theta)}\theta^{\theta}x^{\theta-1}e^{-\theta x}, \qquad \theta > 0, x > 0$$

Time to Infection

- Assume we have two infections with infection times $T_i, i = 1, 2$
- Denote the CDF
$$F_i(t_i) = P(T_i \leq t_i)$$
- Denote the survival function; proportion susceptible
$$S_i(t_i) = \exp\left(-\int_0^{t_i} \lambda_i(u)du\right) = \exp(-H_i(t_i))$$
- The infection hazard; force of infection
$$\lambda(t_i) = -\frac{d}{dt_i}\log(S_i(t_i)) = \frac{f_i(t_i)}{S_i(t_i)}$$

Univariate Frailty

- Assume frailty distributions $Z_i, i = 1, 2$
- Conditional survival function for infection i with frailty $Z_i, i = 1, 2$
$$S_i(t_i|Z_i) = e^{-\int_0^{t_i} \lambda_i(Z_i,u)du}$$
- Proportional hazards assumption $\lambda_i(Z_i, u) = Z_i\lambda_{i0}(u)$
$$S_i(t_i|Z_i) = e^{-\int_0^{t_i} Z_i\lambda_{i0}(u)du}$$
- The unconditional survival function
$$S_i(t_i) = p_i\left(\int_0^{t_i} \lambda_{i0}(u)du\right)$$
with p_i the Laplace transform of $Z_i, i = 1, 2,$

Correlated Frailty

- Assuming conditional independence $T_1 \perp T_2 | Z_1, Z_2$

$$S(t_1, t_2 | Z_1, Z_2) = S_1(t_1|Z_1) \times S_2(t_2|Z_2)$$

with (Z_1, Z_2) following a bivariate frailty distribution

- Yashin et al. (1995): correlated gamma frailty: scale 1, variances σ_i^2, correlation ρ

$$\begin{aligned} S(t_1, t_2) &= [S_1(t_1)]^{1-\frac{\sigma_1}{\sigma_2}\rho} \times [S_2(t_2)]^{1-\frac{\sigma_2}{\sigma_1}\rho} \\ &\times [S_1^{-\sigma_1^2}(t_1) + S_2^{-\sigma_2^2}(t_2) - 1]^{-\frac{\rho}{\sigma_1\sigma_2}} \end{aligned}$$

- $S_i(t_i) = (1 + \sigma_i^2 \tilde{H}_i(t_i))^{\frac{-1}{\sigma_i^2}}$

Identifiability

1. $\sigma_i = 0$: no frailty
2. $\sigma_i = \sigma > 0; \rho = 0$: univariate frailty

   $\longrightarrow$ Elbers and Ridder (1982); Heckman (1984); Hougaard (1986)
3. $\sigma_i = \sigma > 0; \rho = 1$: shared frailty

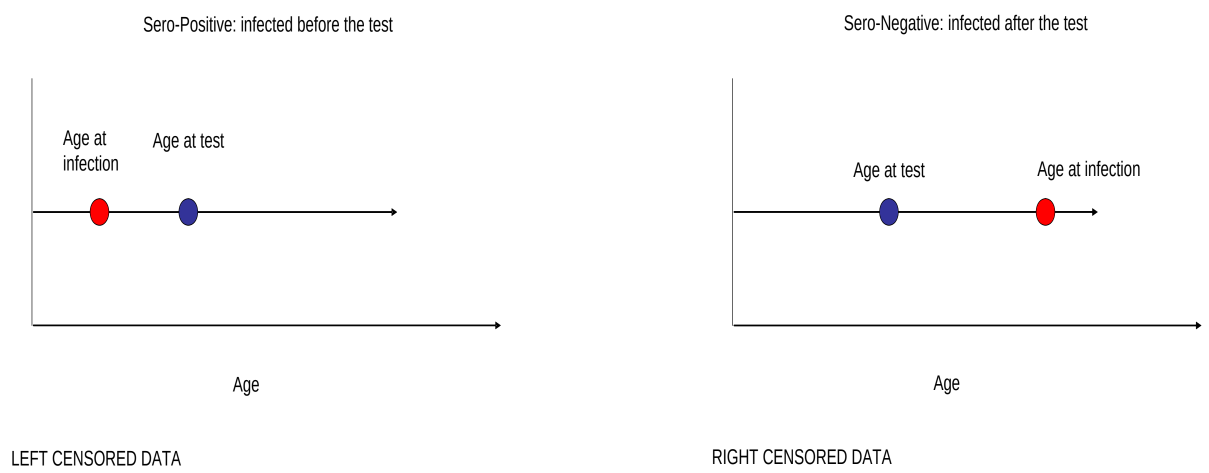
   $\longrightarrow$ Honoré (1993)
4. $\sigma_1, \sigma_2 > 0; 0 \leq \rho \leq 1$: correlated frailty

   $\longrightarrow$ Yashin et al. (1995)

- Estimation: ML, EM and MCMC

Current Status Data

- Serological studies: current status data



Current Status Data

- Denote Y_i the binary current status variable for infection i

$$\pi_i(a) = P(Y_i = 1|a) = P(T_i \leq a)$$

- Seroprevalence $\tilde{\pi}_i(a) = P(\tilde{Y}_i \leq a)$
- Assume no diagnostic test uncertainty $\pi_i(a) = \tilde{\pi}_i(a)$
- The force of infection

$$\lambda_i(a) = \frac{f_i(a)}{S_i(a)} = \frac{\pi_i'(a)}{1 - \pi_i(a)}$$

Current Status Data

- The correlated frailty expression now becomes

$$S(a,a) = [S_1(a)]^{1-\frac{\sigma_1}{\sigma_2}\rho} \times [S_2(a)]^{1-\frac{\sigma_2}{\sigma_1}\rho} \times [S_1^{-\sigma_1^2}(a) + S_2^{-\sigma_2^2}(a) - 1]^{-\frac{\rho}{\sigma_1\sigma_2}}$$

- The correlated frailty simplifies to:
  - Extended shared frailty $\rho = 1$

$$S(a,a) = [S_1(a)]^{1-\frac{\sigma_1}{\sigma_2}} \times [S_2(a)]^{1-\frac{\sigma_2}{\sigma_1}} \times [S_1^{-\sigma_1^2}(a) + S_2^{-\sigma_2^2}(a) - 1]^{-\frac{1}{\sigma_1\sigma_2}}$$

  - Shared frailty $\rho = 1, \sigma_1 = \sigma_2 = \sigma$

$$S(a,a) = [S_1^{-\sigma^2}(a) + S_2^{-\sigma^2}(a) - 1]^{-\frac{1}{\sigma^2}}$$

$\rightarrow$ Farrington et al. (2001); Sutton et al. (2006)

Current Status Data

- Note that

$$S_i(a) = \left(1 + \sigma_i^2 \int_0^a \lambda_{i0}(u)du\right)^{-1/\sigma_i^2}$$

where $\lambda_{i0}(u)$ is the baseline force of infection.

- Alternatively

$$S_i(a) = \exp\left(-\int_0^a \lambda_i(u)du\right)$$

where $\lambda_i(u)$ is the force of infection.

Analysis

- Piecewise constant FOI: $[0, 3), [3, 6), [6, 12)[12, 20) : \lambda_{ij}$
- Gamma frailties
- 4 models:
  - NF: univariate model without frailty
  - SF: bivariate model shared frailty, correlation one
  - ESF: bivariate model extended shared frailty, correlation one
  - CF: bivariate model correlated frailty
- SAS NLMIXED, MLa

Results: Belgium

par	NF	SF	ESF	CF
	estimate (se)	estimate (se)	estimate (se)	estimate se
$\tilde{\lambda}_{11}$	0.060 (0.012)	0.060 (0.012)	0.060 (0.012)	0.060 (0.012)
$\tilde{\lambda}_{12}$	0.124 (0.026)	0.125 (0.026)	0.125 (0.026)	0.125 (0.026)
$\tilde{\lambda}_{13}$	0.087 (0.019)	0.085 (0.019)	0.085 (0.019)	0.085 (0.019)
$\tilde{\lambda}_{14}$	0.072 (0.020)	0.072 (0.020)	0.072 (0.020)	0.072 (0.020)
$\tilde{\lambda}_{21}$	0.264 (0.029)	0.265 (0.029)	0.262 (0.029)	0.262 (0.029)
$\tilde{\lambda}_{22}$	0.450 (0.073)	0.455 (0.073)	0.453 (0.073)	0.453 (0.073)
$\tilde{\lambda}_{23}$	0.179 (0.044)	0.173 (0.044)	0.179 (0.043)	0.179 (0.043)
$\tilde{\lambda}_{24}$	0 (-)	0 (-)	0 (-)	0 (-)
σ_1		0.401 (0.082)	2.178 (24.467)	1.734 (5.040)
σ_2		0.401 (0.082)	6.165 (69.339)	4.877 (14.089)
ρ				1 (-)
-2loglik	2823.9	2816.2	2810.0	2810.0

Note: ‘rhobit’-link function was used to estimate ρ

Result

- Unidentifiability of the correlated gamma frailty for current status data
  - 'rhobit'-link
  - univariate monitoring times
- Is the extended shared frailty identifiable?

σ_1	$\hat{\sigma}_2$
2	0.27(0.07)
4	0.45(0.11)
6	0.65(0.16)
8	0.85(0.21)
10	1.06(0.26)

- Hougaard (2000): "having two random effects for the same source of variation implies that it will be difficult or impossible to separate the random effects"

Shared Frailty: BE, UK, FI, IT, PL

- We take the FOI as piecewise constant over different age-categories
- Enrolment ages (Source OECD, statistics on education):

	Country				
Enrolment level	BE	EW	FI	IT	PL
Pre-school	[0,3[	[0,3[	[0,3[	[0,3[	[0,3[
Pre-primary	[3,6[	[3,5[	[3,7[	[3,6[	[3,7[
Primary	[6,12[	[5,11[	[7,13[	[6,11[	[7,13[
Secondary and tertiary	12+	11+	13+	11+	13+

- Selecting the most parsimonious model using BIC

Shared Frailty: BE, UK, FI, IT, PL

Source	Parameter	Belgium	E & W	Finland	Italy	Poland
FOI B19	Infants	0.068	0.068	0.013	0.068	0.068
	Pre-primary	0.110	0.032	0.032	0.032	0.032
	Primary	0.092	0.092	0.092	0.092	0.092
	Secondary	0.069	0.017	0.017	0.017	0.017
FOI VZV	Infants	0.287	0.148	0.090	0.148	0.148
	Pre-primary	0.356	0.167	0.356	0.167	0.167
	Primary	0.222	0.222	0.222	0.222	0.222
	Secondary	0	0	0	0	0
Heterogeneity	Shape	8.283	1.984	8.283	8.283	1.984

Sensitivity Analysis

- Changing the functional relationship: B-splines, parametric modelling
- Changing the frailty distribution: log-normal, log-mixture of normals
- Using different copulas: Clayton's copula, Gumbel, . . .

Discussion

- Time to event - current status
- Survival setting - generalized linear mixed models
- Inevitable loss of information
- Unidentifiability of the correlated frailty
- Apparent (un)identifiability of the shared frailty

Further research

- Simulation study: what do we loose?
- Singular information matrix? Rotnitzky et al. (2000)

References

Aalen, O. (1988). Heterogeneity in survival analysis. *Statistics in Medicine*, 7:1121–1137.

Coutinho, F., Massad, E., Lopez, L., Burattini, M., Struchiner, C., and Azevedo-Neto, R. (1999). Modelling heterogeneities in individual frailties in epidemic models. *Mathematial and Computer Modelling*, 30:97–115.

Elbers, C. and Ridder, G. (1982). True and spurious duration dependence: The identifiability of the proportional hazard model. *Review of Economic Studies*, 49:403–409.

Farrington, C., Kanaan, M., and Gay, N. (2001). Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. *Applied Statistics*, 50:251–292.

Grenfell, B. T. and Anderson, R. M. (1985). The estimation of age-related rates of infection from case notifications and serological data. *Journal of Hygiene*, 95(2):419–36.

Heckman, J. ans Singer, B. (1984). The identifiability of the proportional hazard model. *Review of Economic Studies*, 51(2):231–241.

Honoré, B. (1993). Identification results for duration models with multiple spells. *Review of Economic Studies*, 60:241–246.

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73:387–396.

Hougaard, P. (2000). *Analysis of multivariate survival data.* New York: Springer.

Keiding, N. (1991). Age-specific incidence and prevalence: A statistical perspective (with discussion). *JRSS-A*, 154:371–412.

Muench, H. (1934). Derivation of rates from summation data by the catalytic curve , vol. 29, no. 185 (mar., 1934), pp. 25-38. *Journal of the American Statistical Association*, 29:25–38.

Rotnitzky, A., Cox, D., Bottai, M., and Robins, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli*, 6(2):243–284.

Sutton, A., Gay, N., Edmunds, W., Hope, V., Gill, O., and Hickman, M. (2006). Modelling the force of infection for hepatitis b and hepatitis c in injecting drug users in England and Wales. *BMC Infectious Diseases*, 6:93.

Vaupel, J., Manton, K., and Stallard, E. (1979). The impact of heterogeneity in individual frailty in the dynamics of mortality. *Demography*, 16(3):439–454.

Yashin, A., Vaupel, J., and Iachine, I. (1995). Correlated individual frailty: An advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies*, 5:145–159.