On the Correlated Gamma Frailty Model for Bivariate Current Status Data

Niel Hens

23 August 2010

Andreas Wienke
Institute of Medical Epidemiology, Biostatistics and Informatics,
Medical Faculty, Martin Luther University, Halle-Wittenberg

Outline

Motivation

- infectious disease epidemiology
- key parameters:
  - force of infection or infection hazard

    the rate at which a susceptible individual acquires an infection

    $\rightarrow$ known to be age-dependent because of age-related mixing of individuals in a population
  - reproduction number

    the average number of secondary cases produced by one typical infectious individual in a population

    $\rightarrow$ known to depend on transmission heterogeneity and thus the mixing of individuals in a population
- Farrington et al. (2001) used a shared gamma frailty to estimate both for two infections transmitted through similar routes: mumps and rubella

Data

- pre-vaccination cross-sectional multisera data on VZV-B19-HAV-CMV from Belgium, 2002:
  - varicella (chickenpox)
  - parvovirus B19 (fifth disease)
  - hepatitis A (non-foodborne, small fraction vaccinated)
  - cytomegalovirus
- bivariate analyses
- resulted in 2815 to 3255 individuals aged 6 months to 71 years
- response (ignoring diagnostic uncertainty): current status data on past infection (ELISA-tests)

Overview

- estimating the force of infection from current status data: Muench (1934); for an overview see Hens et al. (2010)
- quantifying the association in acquisition of two infections:
  - Farrington et al. (2001): shared gamma frailty applied to two infections with the same transmission route: mumps and rubella
  - Hens et al. (2008): bivariate Dale & baseline category logits model
- How do we quantify dependence between two infections?
- When using frailties, can we disentangle heterogeneity from correlation?

Correlated Frailty

- Assume we have two infections with infection times $T_i, i = 1, 2$.
- Assume 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

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

- $S_i(t_i) = (1 + \sigma_i^2 \tilde{H}_i(t_i))^{\frac{-1}{\sigma_i^2}}$
- necessary condition: $0 \leq \rho \leq \min(\sigma_1/\sigma_2, \sigma_2/\sigma_1)$

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 or infection hazard is given by

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

Current Status Data

- The correlated frailty expression becomes

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

- where

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

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

- or, alternatively,

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

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

- The shared frailty is given by $\rho = 1, \sigma_1 = \sigma_2 = \sigma$.

Identifiability & Estimation

- time to event & right censored data:
  - identifiability: Yashin et al. (1995)
  - estimation: ML, EM and MCMC (Wienke et al., 2005)
- current status data:
  - identifiability - necessary condition: parametric baseline hazard

    $\rightarrow$ theory on parameter redundancy (Catchpole and Morgan, 1997)
  - estimation: ML

Simulations

- Scenario 1: unbiasedness under different censoring schemes
  - correlated gamma frailty model using a Gompertz baseline hazard ($\lambda_{i0}(t) = a_i \exp(b_i t), i = 1, 2$, Hens et al. (2009))
    - uncensored time to event
    - right censored data
    - current status data
- Scenario 2: misspecification of the frailty distribution
  - true mechanism: correlated frailty
  - assumed mechanism: univariate, shared, correlated (equal var)
- Scenario 3: misspecification of the baseline hazard
  - true mechanism: exponential and Weibull
  - assumed mechanism: exponential, Weibull and Gompertz

Scenario 1: unbiasedness under different censoring schemes

parameter	true value	uncensored time to event mean (e.s.e.)	right censored data mean (e.s.e.)	current status data mean (e.s.e.)
σ_1	1.600	1.604 (0.113)	1.621 (0.466)	1.694 (1.854)
σ_2	1.000	0.999 (0.068)	1.056 (0.214)	1.179 (0.920)
ρ	0.500	0.501 (0.035)	0.540 (0.169)	0.636 (0.257)
a_1	0.006	0.006 (0.001)	0.006 (0.001)	0.006 (0.001)
b_1	0.020	0.020 (0.002)	0.022 (0.010)	0.045 (0.420)
a_2	0.008	0.008 (0.001)	0.008 (0.001)	0.008 (0.001)
b_2	0.030	0.030 (0.003)	0.032 (0.007)	0.048 (0.228)

Table: averaged parameter estimates and empirical standard errors

Scenario 2: frailty misspecification

par	true value	correlated frailty mean (e.s.e.)	common variance CF mean (e.s.e.)	shared frailty mean (e.s.e.)	univariate frailty mean (e.s.e.)
σ_1	1.600	1.694 (1.854)	1.185 (0.429)	0.769 (0.051)	1.962 (2.219)
σ_2	1.000	1.179 (0.920)	1.185 (0.429)	0.769 (0.051)	1.107 (0.941)
ρ	0.500	0.636 (0.257)	0.679 (0.219)	1.000 (-)	0.000 (-)
a_1	0.006	0.006 (0.001)	0.006 (0.001)	0.006 (0.001)	0.006 (0.005)
b_1	0.020	0.045 (0.420)	0.013 (0.009)	0.007 (0.004)	0.062 (0.135)
a_2	0.008	0.008 (0.001)	0.008 (0.001)	0.008 (0.001)	0.008 (0.001)
b_2	0.030	0.048 (0.228)	0.039 (0.019)	0.024 (0.003)	0.047 (0.047)

Table: averaged parameter estimates and empirical standard errors

Scenario 3: hazard misspecification

scheme	AIC	σ_1	σ_2	ρ
EE	5045	1.601 (0.057)	0.998 (0.042)	0.500 (0.050)
EG	5049	1.601 (0.126)	1.010 (0.087)	0.508 (0.053)
EW	5049	1.619 (0.459)	1.078 (0.272)	0.559 (0.104)
WW	3797	1.632 (0.307)	1.031 (0.145)	0.525 (0.081)
WG	3816	1.128 (0.113)	0.601 (0.078)	0.535 (0.072)
WE	3815	1.241 (0.041)	0.581 (0.041)	0.468 (0.036)

Table: First letter in the scheme denotes the generating function W: Weibull or E: Exponential, the second letter the fitted function with W: Weibull, G: Gompertz and E: Exponential.

Data Analysis: VZV-B19-HAV-CMV

comparison	ρ (s.e.)	best model (AIC)
VZV-B19	0.364 (0.096)	Weibull - correlated frailty
VZV-CMV	0.030 (0.068)	Gompertz - univariate frailty
VZV-HAV	0.004 (0.095)	Gompertz - univariate frailty
B19-CMV	0.240 (0.100)	Gompertz - univariate frailty
B19-HAV	0.094 (0.610)	Gompertz - univariate frailty
CMV-HAV	0.999 (0.050)	Gompertz - shared frailty

Table: VZV-B19-HAV-CMV two-by-two analyses: correlation estimates based on the best correlated frailty model (AIC) and best frailty model (AIC).

Discussion

- Infectious disease epidemiology:
  - What does the correlation in acquisition of two infections tell us?
  - Transmission routes:
    - varicella and parvovirus B19: close non-sexual contacts
    - hepatitis A: fecal-oral or foodborne
    - cytomegalovirus and hepatitis B: sexual or via blood

    so far:
    - varicella and parvovirus B19: same transmission route
  - What about hepatitis A and CMV? - fluids?

Discussion

- Statistics:
  - correlated frailty: disentangling heterogeneity and association
  - unbiasedness for current status data
  - identifiable using a parametric baseline hazard
  - misspecification of the frailty or baseline hazard leads to biased estimates
- Further research:
  - sufficient conditions to ensure identifiability
  - semi-parametric estimate for the baseline hazard
  - extensions beyond the bivariate correlated gamma frailty

Acknowledgements

- Marc Aerts, Philippe Beutels, Tom Cattaert, Jianxing Chen, Paddy Farrington, Elasma Milanzi and Geert Molenberghs
- POLYMOD, a European Commission project funded within the Sixth Framework Programme, Contract number: SSP22-CT-2004-502084
- Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), project number 060081
- IAP research network nr P6/03 of the Belgian Government (Belgian Science Policy).
- German Research Council, project number WI 3288/1-1.

References

Catchpole, E. and Morgan, B. (1997). Detecting parameter redundancy. *Biometrika*, 84:187–196.

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.

Hens, N., Aerts, M., Faes, C., Shkedy, Z., Lejeune, O., Damme, P. V., and Beutels, P. (2010). Seventy-five years of estimating the force of infection from current status data. *Epidemiol Infect*, 138(6):802–812.

Hens, N., Aerts, M., Shkedy, Z., Theeten, H., Damme, P. V., and Beutels, P. (2008). Modelling multisera data: the estimation of new joint and conditional epidemiological parameters. *Stat Med*, 27(14):2651–2664.

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. *Stat Med*, 28(22):2785–2800.

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.

Wienke, A., Arbeev, K., Locatelli, I., and Yashin, A. (2005). A comparison of different bivariate correlated frailty models and estimation strategies. *Mathematical Biosciences*, 198:1–13.

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.