Extending frailty models applied to infectious disease epidemiology

Extending frailty models applied to infectious disease epidemiology

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

[1] Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Diepenbeek, Belgium
[2] Centre for Health Economics Research and Modeling Infectious Diseases, Vaccine and Infectious Disease Institute, University of Antwerp, Wilrijk, Belgium

E-mail for correspondence: steven.abrams@uhasselt.be

Abstract: It has been shown that individual heterogeneity in the acquisition of infectious diseases has a large impact on the estimation of important epidemiological parameters such as the (basic) reproduction number. Therefore frailty modelling has become increasingly popular in infectious disease epidemiology. However, so far, using frailty models, it was assumed infections confer lifelong immunity after recovery, an assumption which is untenable for non-immunizing infections. Our work concentrates on refining the existing frailty models to encompass infection processes with reinfections and waning immunity. Shared gamma frailty models, which are frequently used in practice, and correlated gamma frailty models that have proven to be a valuable alternative are considered. We show that naively assuming lifelong immunity in frailty models introduces substantial bias in the estimation of the basic and effective reproduction number. We illustrate our work using Belgian cross-sectional serological data on parvovirus B19 (PVB19) and varicella zoster virus (VZV). Whereas it is typically assumed that lifelong immunity holds for VZV, more recently, empirical evidence for PVB19 indicates waning of immunity after infection, leading to potential reinfections with the virus.

Keywords: shared and correlated gamma frailty models; social contact rates; SIRS transmission model; mass action principle; serological data.

1 Introduction

In recent years, frailty modelling has become increasingly popular in survival analysis to model multivariate event times. Even more so, as individuals differ greatly in their risk of acquiring infections, frailty models found their way into the field of infectious disease epidemiology. Farrington et al. (2001) considered the shared gamma frailty model in the context of bivariate current status data. However, due to its severe limitations, the more flexible correlated frailty model was used by Hens et al. (2009), at the cost of assuming a parametric baseline hazard. From an epidemiological point

of view, frailty models rely on the assumption of lifelong immunity after recovery which becomes untenable for non-immunizing infections. Furthermore, as individual heterogeneity inflates estimates for the basic reproduction number, a correct assessment of heterogeneity, and therefore a correct specification of the infection process, is of utmost importance to obtain reliable estimates for this quantity. In our work, we focus on shared and correlated frailty models for non-immunizing infections. The methodology is illustrated using Belgian current status data on parvovirus B19 (PVB19) and varicella zoster virus (VZV) collected between 2001 and 2003. In addition, a parametric baseline hazard of infection is derived from the mass action principle in which transmission of the pathogen is related to social contact data obtained from the Belgian POLYMOD survey.

2 Materials and methods

Consider bivariate current status data (y_1, y_2, a) with y_i the observed immunological status with respect to infection i = 1,2 and a the age of the subject at the cross-sectional sampling time. The binary random variables Y_i, given age a, follow a binomial distribution with probability of being seropositive equal to $\pi_i(a) = 1 - S_i(a)$, and $S_i(a)$ is the proportion susceptible of age a. The age-dependent seroprevalence for both infections can be modelled using frailty models, thereby estimating model parameters $\boldsymbol{\theta}$ while maximizing the multinomial loglikelihood with contribution

$$\begin{aligned} ll(y_1, y_2, a|\boldsymbol{\theta}) &= y_1 y_2 \log\left(1 - S_1(a|\boldsymbol{\theta}) - S_2(a|\boldsymbol{\theta}) + S_{12}(a|\boldsymbol{\theta})\right) + \\ & \quad y_1(1-y_2)\log\left(S_2(a|\boldsymbol{\theta}) - S_{12}(a|\boldsymbol{\theta})\right) + \\ & \quad (1-y_1)y_2\log\left(S_1(a|\boldsymbol{\theta}) - S_{12}(a|\boldsymbol{\theta})\right) + \\ & \quad (1-y_1)(1-y_2)\log\left(S_{12}(a|\boldsymbol{\theta})\right), \end{aligned}$$

From this point onwards, dependence on the model parameters $\boldsymbol{\theta}$ is suppressed from notation. Let Z_i represent a frailty with unit mean and variance σ_{if}^2. For infections in endemic equilibrium and without loss of natural immunity, the susceptible proportion of age a with frailty Z_i is given by

$$S_i(a|Z_i) = \exp\left(-\int_0^a Z_i \lambda_{i0}(u)du\right) = \exp\left(-Z_i M_{i0}(a)\right), \quad i = 1, 2$$

under the proportional hazards assumption (PHA). The unconditional survival functions equal $S_i(a) = \mathbf{L}_i(M_{i0}(a))$, expressed in terms of the Laplace transform $\mathbf{L}_i$ of Z_i and the integrated baseline hazard function $M_{i0}(a)$. Solving the system of ordinary differential equations associated with the mathematical SIRS compartmental model yields:

$$\begin{aligned} S_i(a) &= \exp\left(-\int_0^a \sigma_i(u)du\right)\mathbf{L}_i(M_{i0}(a)) \; + \\ & \quad \int_0^a \sigma_i(u)\exp\left(-\int_u^a \sigma_i(v)dv\right)\mathbf{L}_i(M_{i0}(a) - M_{i0}(u))du. \end{aligned}$$

when individuals are allowed to flow back from the recovered to the susceptible state at a replenishment rate $\sigma_i(a)$. The bivariate unconditional survival function $S_{12}(a)$ is derived assuming conditional independence of the infection times, given the frailty terms Z_i. In the shared gamma frailty setting, $Z_1 = Z_2 \equiv Z$, where $Z \sim \Gamma(1/\sigma_f^2, 1/\sigma_f^2)$. In the correlated frailty model, we have $Z_1 = \sigma_{1f}^2 (Y_0^* + Y_1^*)$, $Z_2 = \sigma_{2f}^2 (Y_0^* + Y_2^*)$, where $Y_l^* \sim \Gamma(k_l, 1)$ (l = 0,1,2) are independent random variables. The gamma frailty distribution is preferred due to its mathematical convenience and closed-form expression for the Laplace transform.

The time homogeneous mass action principle, which is briefly described here, links the available information on social contact behaviour to the baseline hazard $\lambda_{i0}(a)$. In the presence of individual frailty terms the mass action principle can be rendered as follows (Farrington et al., 2001):

$$\lambda_i(a, Z_i) = \frac{ND_i}{L} \int_0^\infty \int_0^\infty \beta_i(a, Z_i; a', Z_i') \lambda_i(a', Z_i') S_i(a'|Z_i') \phi(a') f_i(Z_i') da' dZ_i'$$

where f_i is the density function of Z_i', $\beta_i(a, Z_i; a', Z_i')$ equals the per capita rate at which an infectious individual of age a' and frailty Z_i' makes an effective contact with a susceptible individual of age a and frailty Z_i, and $\phi(a')$ represents the probability of being alive at age a'. In addition, N, D_i and L are the population size, the mean duration of infectiousness for infection i and the life expectancy, respectively. Under the PHA, $\beta_i(a, Z_i; a', Z_i') = Z_i Z_i' \beta_{i0}(a, a')$ and $\lambda_i(a, Z_i) = Z_i \lambda_{i0}(a)$. Moreover, $\beta_{i0}(a, a')$ is decomposed into a proportionality factor $q_i(a, a'|c)$, representing transmission potential upon a contact, and $c(a, a')$, the annual per capita rate at which individuals of age a' contact individuals of age a. An iterative procedure is used to solve the mass action principle and to derive the baseline hazard of infection thereof. The basic reproduction number R_{i0}, i = 1,2, is defined as $(1+\sigma_{if}^2)$ times the dominant eigenvalue of the next generation matrix (Diekmann et al., 1990).

3 Data application

Three shared gamma frailty models are fitted to the serology from PVB19 and VZV. Despite potential reinfections with PVB19, VZV infections are assumed to confer lifelong immunity since accounting for more complex infection dynamics did not improve model fit. The model relying on the assumption of lifelong immunity for both infections is denoted by M1. In addition, model M2 allows for replenishment of the susceptible compartment at a constant rate σ_1 solely for PVB19. Finally, model M3 simply extends model M2 by introducing an age-dependent dichotomous replenishment for PVB19 based on a cut-off value of 35 years.

The results in Table 1 indicate that the models with SIRS dynamics for PVB19 (M2 and M3) outperform the traditional SIR model (M1) based

on AIC-values. Furthermore, the frailty variance is seriously overestimated in model M1 which is reflected as well in the estimated basic reproduction numbers $\hat{R}_{i0}$. Misspecification of the underlying infection process for one infection also influences the estimated reproduction number for the other one.

TABLE 1. ML estimates with regard to PVB19 ($i = 1$) and VZV ($i = 2$) with 95% bootstrap-based CI and corresponding AIC-values.

Model				$\hat{R}_{i0}$		AIC
M1	q_{10}	0.073	[0.069, 0.077]	3.59	[3.27, 3.90]	4537.28
	q_{20}	0.209	[0.189, 0.232]	12.07	[10.46, 13.74]	
	σ_f^2	0.158	[0.102, 0.210]			
M2	q_{10}	0.072	[0.068, 0.075]	3.17	[2.94, 3.43]	4477.98
	σ	0.011	[0.007, 0.014]			
	q_{20}	0.177	[0.162, 0.196]	9.15	[8.07, 10.53]	
	σ_f^2	0.036	[5.4e-7, 0.086]			
M3	q_{10}	0.072	[0.069, 0.075]	3.13	[2.95, 3.38]	4474.39
	σ_1	0.016	[0.010, 0.022]			
	σ_2	0.008	[0.005, 0.012]			
	q_{20}	0.173	[0.161, 0.191]	8.82	[8.01, 10.13]	
	σ_f^2	0.021	[3.6e-7, 0.071]			

4 Discussion

We showed that the use of traditional frailty models results in biased estimates of important epidemiological parameters when incorrectly relying on the assumption of lifelong immunity. Henceforth, frailty models comprising more general infection processes should be considered instead when evidence against natural immunity exists.

References

Diekmann, O. et al. (1990). On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, **28**, 365 – 382.

Farrington, C. P. et al. (2001). Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. *Applied Statistics*, **50**, 251 – 292.

Hens, N. et al. (2009). The correlated and shared gamma frailty model for bivariate current status data: An illustration for cross-sectional serological data. *Statistics in Medicine*, **27(14)**, 2785 – 2800.