Modelling time varying heterogeneity in recurrent infection processes: an application to serological data

Steven Abrams,

Hasselt University, Diepenbeek, Belgium

Andreas Wienke,

Martin Luther University Halle-Wittenberg, Halle, Germany

and Niel Hens

Hasselt University, Diepenbeek, and University of Antwerp, Wilrijk, Belgium

Summary. Frailty models are often used in survival analysis to model multivariate time-to-event data. In infectious disease epidemiology, frailty models have been proposed to model heterogeneity in the acquisition of infection and to accommodate association in the occurrence of multiple types of infection. Although traditional frailty models rely on the assumption of lifelong immunity after recovery, refinements have been made to account for reinfections with the same pathogen. Recently, Abrams and Hens quantified the effect of misspecifying the underlying infection process on the basic and effective reproduction number in the context of bivariate current status data on parvovirus B19 and varicella zoster virus. Furthermore, Farrington, Unkel and their co-workers introduced and applied time varying shared frailty models to paired bivariate serological data. In this paper, we consider an extension of the proposed frailty methodology by Abrams and Hens to account for age-dependence in individual heterogeneity through the use of age-dependent shared and correlated gamma frailty models. The methodology is illustrated by using two data applications.

Keywords: Age-dependent frailties; Basic reproduction number; Non-immunizing infections; Serology; Shared and correlated frailty models

1. Introduction

The analysis of time-to-event data expanded rapidly since the development of the proportional hazards model by Cox (1972) with applications in medicine, epidemiology, demography and many other fields. Frailty models have been used in survival analysis to model dependence and unobserved heterogeneity in multivariate time-to-event data explicitly and to address questions that are related to the association between multiple event times (see, for example, Clayton (1978), Hougaard (2000) and Duchateau and Janssen (2008)). The dependence between event times usually arises since individuals in the same group (family, cluster or litter) are related, or because multiple event times are recorded for the same subject. Multivariate frailty models provide an extension of the univariate frailty model (Vaupel *et al.*, 1979; Lancaster, 1979), taking

mutual dependence of event times into account by using conditional proportional hazards given latent frailty terms. To impose an association structure among the event times, two approaches are often considered. First, a shared frailty model uses a common frailty term for all individuals in the same group (cluster); see, for example, Hougaard (2000), Therneau and Grambsch (2000) and Duchateau and Janssen (2008). Second, the correlated frailty model extends the shared frailty model to allow for a more flexible correlation structure between the frailty terms (see, for example, Yashin *et al.* (1995), Xue and Brookmeyer (1996) and Wienke *et al.* (2002, 2003)).

Individual heterogeneity with respect to factors that may enhance or inhibit the transmission of infectious diseases affects the effectiveness of control strategies as has been understood for a long time (Anderson and May, 1991). When this heterogeneity is not appropriately accounted for, one potentially ends up with biased estimates for important epidemiological parameters. One of the key assumptions in shared and correlated frailty models is that of age or time invariant individual heterogeneity, implying that individuals have a constant 'frailty' level during their entire life, which is rather restrictive. Recently, Farrington *et al.* (2012) introduced time varying shared frailties for two immunizing infections. We present new methodology for infections conferring temporary humoral immunity on recovery (henceforth referred to as non-immunizing infections) taking into account the fact that individual frailties—and hence the correlations that they induce—may vary with time. This enables us to describe and quantify heterogeneities further and to study how these heterogeneities evolve over time. Although our novel frailty methodology is generally applicable to recurrent event processes, we focus here on recurrent infection processes using cross-sectional serological data. Cross-sectional serological data constitute type I interval-censored data, or current status data, for which the only knowledge about the event is whether it occurred before the observation time point or not, at least for infections conferring lifelong humoral immunity. In the case of pathogens for which recurrent infections during life are possible, an individual's immunological status merely indicates whether he or she experienced recent infection (giving rise to a humoral immunity response) and therefore whether the individual is susceptible to infection or not at the moment of data collection. Although we focus on current status data, the models that are presented here are applicable to recurrent event time data in the presence of any type of censoring, underlining the importance of our approach in the general survival context (not shown here). More specifically for the current status data setting, in contrast with a general recurrent event setting entailing information on multiple events per subject, recurrent infections are not directly observed but rather accounted for through the specification of non-immunizing infection dynamics (a theoretical mechanistic model). For infections in endemic equilibrium, serology can be used to infer the force of infection, even in the presence of individual heterogeneity and more complicated disease dynamics (Farrington *et al.*, 2001; Abrams and Hens, 2015). For a historical note on the estimation of the force of infection from serological data, we refer to Hens *et al.* (2010).

In summary, the methodology that is presented in this paper adds

(a) time varying shared frailty models for recurrent event time data and
(b) time varying correlated frailties for single-event time data

to the existing frailty literature. The paper is organized as follows. In Section 2, two different data applications are introduced. In Section 3, time varying frailty models are derived for current status data allowing for recurrent events in the past. In Section 4, the proposed methodology is applied to the data. Finally, advantages and disadvantages of our approach are summarized in Section 5.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

```
http://wileyonlinelibrary.com/journal/rss-datasets
```

2. Motivating examples

First, the methodology is illustrated on hepatitis A virus (HAV) and hepatitis B virus (HBV) serology, obtained from a seroepidemiological study undertaken in 1993–1994 in Flanders, Belgium. In total, 4026 blood samples were drawn from a study group that was representative of the Flemish population and tested for the presence of immunoglobulin G antibodies against HAV and HBV (Beutels *et al.*, 1997). On the basis of a prespecified cut-off level for the antibody titres, individuals are classified as either seropositive or seronegative for each disease. In addition to the immunological status with respect to a specific infection, age at sample collection was registered. As the true infection times are known only to lie between 0 and the age at sampling time for seropositive individuals or between the age at sampling time and ∞ for seronegative individuals, we are faced with current status data.

Second, we consider serological survey data on parvovirus B19, PVB19, and varicella zoster virus (VZV) infections in Belgium. Blood samples were collected for 3379 different individuals between 2001 and 2003 and tested for the presence of immunoglobulin G antibodies against PVB19 and VZV (Hens *et al.*, 2008). A detailed description regarding symptomatology and epidemiology of the aforementioned infections can be found in the on-line appendix A. In both examples only the (current) immunological status of an individual is known at the individual's age at the time of investigation. The number and time points of former infections, if applicable, are unknown, which distinguishes this type of data from classical recurrent event time data. Note, however, that VZV, HAV and HBV infections are assumed to confer lifelong immunity, and only reinfection dynamics for PVB19 are studied and accounted for in the respective data applications.

3. Materials and methods

3.1. Notation—univariate frailty model

Time-to-event data represent the time to a defined event such as death, failure or infection. The analysis of time-to-event data in survival analysis is often complicated by censoring. Therefore, let T_j^* represent the true event time for individual $j = 1, \ldots, n$ and consider right-censored data (T_j, Δ_j), where $T_j = \min(T_j^*, C_j)$ represents the observation time, C_j refers to the censoring time and Δ_j denotes the censoring indicator defined as

$$\Delta_j = \begin{cases} 1 & \text{when } T_j^* \leqslant C_j, \\ 0 & \text{when } T_j^* > C_j. \end{cases}$$

Therefore, the observed event time T_j resembles the true event time T_j^* only whenever it occurs before censoring. Generally, right-censored data are a special case of interval-censored data for which the true event time is known only to lie in some interval. In the case of right-censored data, the right end point of that interval corresponds to ∞. Finally, for current status (CS) data, the observation (monitoring) time T_j equals the censoring time C_j, and $\Delta_j = \mathbf{1}_{T_j^* \leqslant C_j}$ indicates whether the event of interest took place before C_j or not. In contrast with right-censored data, the exact value of the event times is always unknown.

Various mathematical compartmental models can be found in the infectious disease literature to describe the transmission of different pathogens in large populations. For the purpose of this paper we consider so-called susceptible–infected–recovered (SIR) and susceptible–infected–recovered–susceptible (SIRS) mechanic models, for infections conferring lifelong immunity or temporary immunity on recovery respectively. We further assume that infections are in endemic equilibrium, the population has reached a demographic equilibrium and the number of deaths
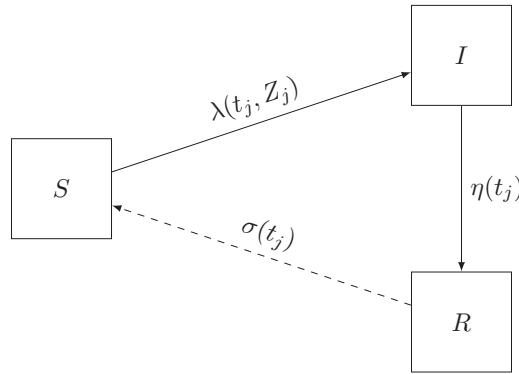
Fig. 1. Mathematical SIR(S) model: subject j flows from the susceptible compartment S to the infected and infectious state I at rate $\lambda(t_j, Z_j)$, and from I to the recovered state R at rate $\eta(t_j)$; in an SIR model, subject j remains in R for life; in an SIRS model, subject j moves to S at rate $\sigma(t_j)$ (- - - ▸)

and births is exactly balanced, entailing a constant population of size N. In Fig. 1, a flow diagram of an SIR model is presented. All individuals are born in compartment S. A subject j flows to the infectious state I at a time-dependent rate $\lambda(t_j, Z_j)$ conditional on a subject-specific frailty term Z_j, i.e. a non-negative random variable, which is often constrained to have unit mean for identifiability. After being infected with the pathogen under study, subject j recovers at a rate $\eta(t_j)$ and moves towards the recovered class R. In continuous time, the conditional proportion of subjects in state $S^{(0)}$ at time t_j is given by

$$S(t_j|Z_j) = \exp\left\{-\int_0^{t_j} \lambda(u, Z_j)\mathrm{d}u\right\}, \tag{1}$$

which is referred to as the conditional survival function in survival analysis. We use the commonly made assumption of proportional hazards, implying that the frailty acts multiplicatively on a baseline hazard: $\lambda(t_j, Z_j) = Z_j \lambda_0(t_j)$, where $\lambda_0(t_j)$ is termed the baseline hazard function. Consequently, the unconditional survival function can be derived by integrating out the frailty terms Z_j according to a specific frailty distribution with density $f(\cdot)$:

$$S(t_j) = E\{S(t_j|Z_j)\} = \int_0^\infty S(t_j|z_j) f(z_j)\mathrm{d}z_j = \mathbf{L}\{\Lambda_0(t_j)\}, \tag{2}$$

where $\mathbf{L}(s)$ denotes the Laplace transform of the random variable Z_j and $\Lambda_0(t_j) = \int_0^{t_j} \lambda_0(u)\mathrm{d}u$ is the cumulative baseline hazard function.

A straightforward extension of the SIR model to encompass recurrent events (i.e. reinfections) is the so-called SIRS model which is graphically depicted in Fig. 1; subjects flow from state R to the susceptible compartment S at replenishment rate $\sigma(t_j)$ (the broken arrow). In an SIRS setting, we can derive an expression for the conditional proportion of subjects who can experience the event at time t_j, which is denoted by $S(t_j|Z_j)$, assuming that the proportion of subjects in I is negligible compared with those in the other compartments (see appendix B in Abrams and Hens (2015)):

$$S(t_j|Z_j) = \exp\left[-\int_0^{t_j}\{\lambda(u, Z_j) + \sigma(u)\}\mathrm{d}u\right] + \int_0^{t_j}\sigma(u)\exp\left[-\int_u^{t_j}\{\lambda(v, Z_j) + \sigma(v)\}\mathrm{d}v\right]\mathrm{d}u,$$

where $\sigma(u)$ refers to the time-dependent replenishment rate at which subjects become at risk again for the event under consideration. Integrating out the random effects, the unconditional proportion of subjects in S can be rendered as

$$S(t_j) = \exp\left\{-\int_0^{t_j} \sigma(u)\mathrm{d}u\right\} \mathbf{L}\{\Lambda_0(t_j)\} + \int_0^{t_j} \sigma(u)\exp\left\{-\int_u^{t_j} \sigma(v)\mathrm{d}v\right\} \mathbf{L}\{\Lambda_0(t_j) - \Lambda_0(u)\}\mathrm{d}u.$$

When $\sigma(u) = 0, \forall u$, the conditional proportion of subjects in $S^{(0)}$ reduces to equation (2). The approach that is undertaken in this paper is closely related to (continuous time) Markov processes in multistate modelling. For a more detailed discussion on multistate models and Markov processes, the reader is referred to Andersen and Keiding (2002).

3.2. Shared and correlated frailty models

Bivariate frailty models enable us to model the association between bivariate event times. Bivariate event time data can arise either when two different events are studied in the same individual, or when the same event occurs in two different but related subjects. Throughout the paper, we formulate the methodology for two events affecting the same individual, albeit that the same reasoning holds in the other case. Consider bivariate right-censored data (T_{ij}, Δ_{ij}), $i = 1, 2$, $j = 1, \ldots, n$, such that the event-specific hazard function can be formulated as $\lambda_i(t_{ij}, Z_{ij}) = Z_{ij}\,\lambda_{i0}(t_{ij})$, where Z_{ij} represent event- and individual-specific frailty terms that are associated with subject $j = 1, \ldots, n$ and event $i = 1, 2$. The dependence structure is implied by the frailty structure imposed. Two different frailty structures are often considered in the survival literature: the shared frailty model

$$Z_{1j} = Z_{2j} \equiv Z_j$$

and the correlated frailty model

$$Z_{1j} = \gamma_1(Y_{0j}^* + Y_{1j}^*),$$
$$Z_{2j} = \gamma_2(Y_{0j}^* + Y_{2j}^*),$$

where γ_i denotes the frailty variance with respect to event i, and the additive components Y_{lj}^*, $l = 0, 1, 2$, are independent random variables with mean and variance equal to $\omega_l > 0$. Therefore, direct calculation of the frailty variances results in the equality $\gamma_i = 1/(\omega_0 + \omega_i)$ ascertaining unit frailty means. The additive decomposition of the frailty terms in the correlated frailty model was proposed by Yashin *et al.* (1995) and first applied to infectious disease modelling by Hens *et al.* (2009). The correlated frailty model extends the restrictive shared frailty model in the sense that the implied correlation between the event times is allowed to differ from 1. Needless to say, the correlated frailty model offers a more general approach to account for individual heterogeneity. However, the implied correlation coefficient ρ is restricted because of the additive decomposition as follows (Yashin *et al.*, 1995):

$$0 \leqslant \rho \leqslant \min\left\{\sqrt{\left(\frac{\gamma_1}{\gamma_2}\right)}, \sqrt{\left(\frac{\gamma_2}{\gamma_1}\right)}\right\}.$$

The shared frailty model is a special case of the correlated frailty model in which the event-specific components Y_{lj}^*, $l = 1, 2$, are identical, i.e. $Y_{1j}^* = Y_{2j}^*$. The time-dependent proportion of individuals who are susceptible to both events under study can be derived relying on the assumption of conditional independence of the event times given the frailties. More specifically, $S(t_{1j}, t_{2j}|Z_{1j}, Z_{2j}) = S_1(t_{1j}|Z_{1j})S_2(t_{2j}|Z_{2j})$, where $S_i(t_{ij}|Z_{ij})$ refers to the proportion of individuals at risk of infection i at time t_{ij} given frailty Z_{ij} for which formulae are presented in Section 3.1. For expressions regarding $S(t_{1j}, t_{2j}|Z_{1j}, Z_{2j})$ and $S(t_{1j}, t_{2j})$, the reader is referred to the on-line appendix C. In this paper, the gamma (frailty) distribution is considered because of its computational and analytical convenience. The usefulness of the gamma distribution is due to the closed form expression for the Laplace transform. Furthermore, it is well established

that in a time invariant bivariate setting the gamma distribution implies a constant association between two failure times of interest, and therefore it could serve as a useful starting point in the fitting process. Independent gamma-distributed components Y_{lj}^* with equal scale parameters result in infection-specific gamma frailties Z_{ij}. The Laplace transform corresponding to a gamma-distributed random variable with mean 1 and variance γ_i is $\mathbf{L}(s) = (1 + \gamma_i s)^{-1/\gamma_i}$, for $s > -1/\gamma_i$.

3.3. *Time varying frailty models*
So far, it has been assumed that individual- and infection-specific frailty terms Z_{ij} do not vary over time. However, time varying frailty models offer a way to allow individual frailty terms to vary with time. Suppose that the hazard at time t_{ij} under the proportional hazards assumption is $\lambda_i\{t_{ij}, Z_{ij}(t_{ij})\} = Z_{ij}(t_{ij})\lambda_{i0}(t_{ij})$, where $Z_{ij}(t_{ij})$ denotes a time varying frailty term.

3.3.1. *Piecewise constant model*
Paik *et al.* (1994) proposed a piecewise constant frailty model accommodating nested structures in the data. Without the complexity of nesting, a piecewise constant frailty model on disjoint intervals $[t_{[m]}, t_{[m+1]}[$ can be obtained as follows:

$$Z_{ij}(t_{ij}) = \sum_{m=1}^{M} I_m(t_{ij}) Z_{ijm}^*,$$

where $I_m(t_{ij}) = 1$ if and only if $t_{ij} \in [t_{[m]}, t_{[m+1]}[$, and $I_m(t_{ij}) = 0$ otherwise. The components Z_{ijm}^* are independent non-negative random variables with unit mean and variance γ_{im}. The time varying frailty variance function equals $\gamma_i(t_{ij}) = \sum_{m=1}^{M} I_m(t_{ij})\gamma_{im}$. A piecewise constant shared gamma frailty model (model TDPCSGF) is derived under the constraint $Z_{jm}^* = Z_{1jm}^* = Z_{2jm}^*$, $m = 1, \ldots, M$, and Z_{jm}^* follows a gamma distribution with unit mean and variance $\gamma_{\cdot m}$. Although this model for $Z_{ij}(t_{ij})$ is straightforward and simple, a disadvantage is the *ad hoc* choice of time intervals $[t_{[m]}, t_{[m+1]}[$ in which the frailties Z_{ijm}^* are assumed constant. In addition, the frailty components are assumed to be independent such that heterogeneity in subsequent time intervals is unrelated, which seems counterintuitive. Therefore, an alternative model was introduced by Farrington *et al.* (2012) and further exploited by Unkel *et al.* (2014) in the context of shared frailty models for immunizing infections. The piecewise constant frailty model is used to assess the goodness of fit of the parametric time-dependent frailty variance functions in the following models.

3.3.2. *Multiplicative model*
In line with the work by Farrington *et al.* (2012) and Unkel *et al.* (2014), a bivariate frailty model is proposed with time varying frailty terms given by

$$Z_{ij}(t_{ij}) = \prod_{m=1}^{M} \{1 + (Z_{ijm}^* - 1)h_{im}(t_{ij})\}, \qquad 0 \leqslant h_{im}(t_{ij}) \leqslant 1,$$

where Z_{ijm}^*, $m = 1, \ldots, M$, are independent random variables with unit mean and variance γ_{im}, and $h_{im}(\cdot)$ is a deterministic function describing the frailty evolution over time. One can easily verify that the frailty terms $Z_{ij}(t_{ij})$ have unit mean and time varying frailty variance

$$\gamma_i(t_{ij}) = \prod_{m=1}^{M} [\{1 + h_{im}^2(t_{ij})\gamma_{im}\} - 1].$$

Suppose that $h_{im}(t_{ij})$ is a deterministic exponential decay function defined as $h_{im}(t_{ij}) = \exp\{-(\phi_{im}t_{ij})^2\}$, where the exponential decay parameter $\phi_{im} > 0$. Although we could consider a more general decay function $\exp\{-(\phi_{im}t_{ij})^k\}$ with any non-negative integer value k, the choice of $k = 2$ provided the best model fits for the multiplicative models when applied to our motivating examples. Assuming that the frailty components Z^*_{ijm} follow a gamma distribution, the following two-component multiplicative time-dependent gamma frailty models can be considered (i.e. taking $M = 2$):

(a) two-component time-dependent shared gamma frailty model TDSGF-2C,

$$Z_{ij}(t_{ij}) = \{1 + (Z^*_{j1} - 1)h_{i1}(t_{ij})\}Z^*_{j2},$$
$$Z^*_{j1} \equiv Z^*_{1j1} = Z^*_{2j1}, \qquad Z^*_{j1} \sim \Gamma(\gamma_{.1}^{-1}, \gamma_{.1}^{-1}),$$
$$Z^*_{j2} \equiv Z^*_{1j2} = Z^*_{2j2}, \qquad Z^*_{j2} \sim \Gamma(\gamma_{.2}^{-1}, \gamma_{.2}^{-1}),$$
$$\gamma_i(t_{ij}) = \gamma_{.2} + h_{i1}^2(t_{ij})\gamma_{.1}(1 + \gamma_{.2}),$$
$$\rho(t_{1j}, t_{2j}) = \frac{\gamma_{.2} + h_{11}(t_{1j})h_{21}(t_{2j})\gamma_{.1}(1 + \gamma_{.2})}{\sqrt{[\{\gamma_{.2} + h_{11}^2(t_{1j})\gamma_{.1}(1 + \gamma_{.2})\}\{\gamma_{.2} + h_{21}^2(t_{2j})\gamma_{.1}(1 + \gamma_{.2})\}]}};$$

(b) two-component time-dependent correlated gamma frailty model TDCGF-2C,

$$Z_{ij}(t_{ij}) = \{1 + (Z^*_{ij1} - 1)h_{i1}(t_{ij})\}Z^*_{ij2},$$
$$Z^*_{ij1} = \gamma_{i1}(Y^*_{0j1} + Y^*_{ij1}), \qquad Y^*_{lj1} \sim \Gamma(\omega_{l1}, 1),$$
$$Z^*_{ij2} = \gamma_{i2}(Y^*_{0j2} + Y^*_{ij2}), \qquad Y^*_{lj2} \sim \Gamma(\omega_{l2}, 1),$$
$$\gamma_{i1} = (\omega_{01} + \omega_{i1})^{-1},$$
$$\gamma_{i2} = (\omega_{02} + \omega_{i2})^{-1},$$
$$\gamma_i(t_{ij}) = \gamma_{i2} + h_{i1}^2(t_{ij})\gamma_{i1}(1 + \gamma_{i2}),$$
$$\rho(t_{1j}, t_{2j}) = \frac{\omega_{02}\gamma_{12}\gamma_{22} + h_{11}(t_{1j})h_{21}(t_{2j})\,\omega_{01}\gamma_{11}\gamma_{21}(1 + \omega_{02}\gamma_{12}\gamma_{22})}{\sqrt{[\{\gamma_{12} + h_{11}^2(t_{1j})\gamma_{11}(1 + \gamma_{12})\}\{\gamma_{22} + h_{21}^2(t_{2j})\gamma_{21}(1 + \gamma_{22})\}]}}.$$

These two-component models simplify to a one-component time-dependent shared gamma frailty model TDSGF-1C and a one-component time-dependent correlated gamma frailty model TDCGF-1C if Z^*_{j2} and Z^*_{ij2}, $i = 1, 2$, follow degenerate distributions with unit mean ($\gamma_{.2} = 0$ and $\gamma_{12} = \gamma_{22} = 0$) respectively. In model TDSGF-1C, the event-specific frailty variances $\gamma_i(t_{ij})$ decrease from $\gamma_{.1}$ to 0 as $t_{ij} \to \infty$. Although the frailty variances differ for different decay rates ϕ_{ir}, the implied correlation $\rho(t_{1j}, t_{2j})$ is constant and equal to 1, which is in line with the time invariant shared frailty model. Furthermore, frailty variances in model TDSGF-2C decline from $\gamma_{.2} + \gamma_{.1}(1 + \gamma_{.2})$ to an asymptotic value $\gamma_{.2}$. Note that $h_{i2}(t_{ij}) \equiv 1$ for both events at all time points t_{ij} in the general formulation of the multiplicative model to limit the complexity of the models. Allowing the decay parameters to be different for both events, the correlation $\rho(t_{1j}, t_{2j})$ decreases over time. Time varying correlated frailty models with one or two components enable a more flexible (time varying) correlation as opposed to their shared counterparts. In models TDCGF-1C and TDCGF-2C, the time-dependent frailty variances evolve in the same way as in TDSGF-1C and TDSGF-2C respectively. In addition, model TDCGF-1C implies a time invariant positive correlation $\rho(t_{1j}, t_{2j})$ which deviates from 1 if ω_{11} and/or ω_{21} differ from 0.

First assume that for the events under study individuals experience each event at most once (SIR model) and time-dependent heterogeneity is assumed to evolve according to the most general model (TDCGF-2C). Hence, the univariate conditional proportion of (susceptible)

individuals at risk for event $i = 1, 2$ under the proportional hazards assumption is

$$S_i\{t_{ij}|Z_{ij}(t_{ij})\} = \exp\left\{-\int_0^{t_{ij}} Z_{ij}(u)\lambda_{i0}(u)\mathrm{d}u\right\}$$

$$= \exp\left(-\int_0^{t_{ij}} [\{1 + (Z_{ij1}^* - 1)h_{i1}(u)\}Z_{ij2}^*]\lambda_{i0}(u)\mathrm{d}u\right),$$

where $\lambda_{i0}(t_{ij})$ is the baseline hazard function at time t_{ij} for infection $i = 1, 2$. The unconditional susceptibility probabilities are obtained by integrating out the random frailty terms $Z_{ij}(t_{ij})$ according to the joint distribution of (Z_{ij1}^*, Z_{ij2}^*). Relying on independence of the frailty components, we have $f_{12}(Z_{ij1}^*, Z_{ij2}^*) = f_1(Z_{ij1}^*)f_2(Z_{ij2}^*)$, where $f_1(\cdot)$, $f_2(\cdot)$ and $f_{12}(\cdot, \cdot)$ represent the marginal density functions for Z_{ij1}^* and Z_{ij2}^*, and the joint density function for both random variables respectively. Let $\mathbf{L}_1$ and $\mathbf{L}_2$ denote the Laplace transforms of the gamma-distributed random variables Z_{ij1}^* and Z_{ij2}^* respectively. Consequently, we can write

$$S_i(t_{ij}) = \int_0^\infty \mathbf{L}_1\{z_{ij2}^*\Lambda_{i0}^*(t_{ij})\}\exp[-z_{ij2}^*\{\Lambda_{i0}(t_{ij}) - \Lambda_{i0}^*(t_{ij})\}]f_2(z_{ij2}^*)\,\mathrm{d}z_{ij2}^*,$$

where $\Lambda_{i0}(t_{ij})$ is the cumulative baseline hazard function with respect to infection $i = 1, 2$, and $\Lambda_{i0}^*(t_{ij}) = \int_0^{t_{ij}} h_{i1}(u)\lambda_{i0}(u)\mathrm{d}u$. Again, relying on the assumption of conditional independence of the event times given the (time varying) frailty terms, the bivariate unconditional proportion of individuals who are susceptible to both events can be derived. These unconditional probabilities are obtained by using a combination of analytical and numerical integration techniques (see the on-line appendix C).

Second, in a mathematical SIRS model with recurrent events, the expressions become more complicated. Under the assumption of frailty-independent event-specific replenishment rates $\sigma_i(t_{ij})$ at time t_{ij}, $i = 1, 2$, we have

$$S_i\{t_{ij}|Z_{ij}(t_{ij})\} = W_{i1}(t_{ij})\exp\left(-\int_0^{t_{ij}} [\{1 + (Z_{ij1}^* - 1)h_{i1}(u)\}Z_{ij2}^*]\lambda_{i0}(u)\mathrm{d}u\right)$$

$$+ \int_0^{t_{ij}} \sigma_i(u)W_{i2}(t_{ij}, u)\exp\left(-\int_u^{t_{ij}} [\{1 + (Z_{ij1}^* - 1)h_{i1}(v)\}Z_{ij2}^*]\lambda_{i0}(v)\mathrm{d}v\right)\mathrm{d}u,$$

where $W_{i2}(t_{ij}, u) \equiv \exp\{-\int_u^{t_{ij}} \sigma_i(v)\mathrm{d}v\}$ and $W_{i1}(t_{ij}) = W_{i2}(t_{ij}, 0)$. Taking the expectation with respect to the components Z_{ij1}^* and Z_{ij2}^* yields an expression for the marginal unconditional proportions of individuals who are susceptible to each event:

$$S_i(t_{ij}) = W_{i1}(t_{ij})\int_0^\infty \mathbf{L}_1\{z_{ij2}^*\Lambda_{i0}^*(t_{ij})\}\exp\{-z_{ij2}^*\Lambda_{i0}^{**}(t_{ij})\}f_2(z_{ij2}^*)\mathrm{d}z_{ij2}^* + \int_0^{t_{ij}} \sigma_i(u)W_{i2}(t_{ij}, u)$$

$$\times \left(\int_0^\infty \mathbf{L}_1[z_{ij2}^*\{\Lambda_{i0}^*(t_{ij}) - \Lambda_{i0}^*(u)\}]\exp[-z_{ij2}^*\{\Lambda_{i0}^{**}(t_{ij}) - \Lambda_{i0}^{**}(u)\}]f_2(z_{ij2}^*)\mathrm{d}z_{ij2}^*\right)\mathrm{d}u,$$

where $\Lambda_{i0}^{**}(t) = \Lambda_{i0}(t) - \Lambda_{i0}^*(t)$. For ease of presentation, the derivation of the joint probabilities is presented in the on-line appendix C. Identifiability of time varying correlated frailty models, at the cost of parametric baseline hazard and decay functions, is not pursued theoretically but is assessed by means of simulations (see the on-line appendix D), and relying on the general methodology of detecting parameter redundancy as proposed by Catchpole and Morgan (1997, 2001). A formal proof, and investigation of sufficient conditions for the models to be identifiable, is considered beyond the scope of this paper.

4. Results

In this section, the time varying frailty methodology is applied to bivariate CS data on

  (a) HAV and HBV, and
  (b) PVB19 and VZV,

as introduced in Section 2. In general, for the bivariate CS data $(t_{1j}, t_{2j}, \delta_{1j}, \delta_{2j})$, the random variable $(\Delta_{ij}|T_{ij} = t_{ij})$ follows a Bernoulli distribution with mean $\pi_i(t_{ij}) = E(\Delta_{ij}|T_{ij} = t_{ij})$. The individual log-likelihood contributions can be expressed in terms of joint susceptibility proportions $p_{\delta_{1j},\delta_{2j}}(\cdot,\cdot)$ as follows (Sun, 2006):

$$\begin{aligned}
\text{ll}_j(\boldsymbol{\theta}, \boldsymbol{\psi}|t_{1j}, t_{2j}, \delta_{1j}, \delta_{2j}) = {} & \delta_{1j}\delta_{2j} \log\{p_{11}(t_{1j}, t_{2j}|\boldsymbol{\theta}, \boldsymbol{\psi})\} + \delta_{1j}(1 - \delta_{2j}) \log\{p_{10}(t_{1j}, t_{2j}|\boldsymbol{\theta}, \boldsymbol{\psi})\} \\
& + (1 - \delta_{1j})\delta_{2j} \log\{p_{01}(t_{1j}, t_{2j}|\boldsymbol{\theta}, \boldsymbol{\psi})\} \\
& + (1 - \delta_{1j})(1 - \delta_{2j}) \log\{p_{00}(t_{1j}, t_{2j}|\boldsymbol{\theta}, \boldsymbol{\psi})\},
\end{aligned}$$

where $p_{\delta_{1j},\delta_{2j}}(t_{1j}, t_{2j}|\boldsymbol{\theta}, \boldsymbol{\psi}) = \Pr(\Delta_{1j} = \delta_{1j}, \Delta_{2j} = \delta_{2j}|t_{1j}, t_{2j}, \boldsymbol{\theta}, \boldsymbol{\psi})$ are the multinomial probabilities that are associated with the distribution of $(\Delta_{1j}, \Delta_{2j})$ and depending on the model parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ that are associated with the baseline hazard functions, and the joint frailty distribution. In general, we have

$$\begin{aligned}
p_{\delta_{1j},\delta_{2j}}(t_{1j}, t_{2j}|\boldsymbol{\theta}, \boldsymbol{\psi}) = {} & \delta_{1j}\delta_{2j} + (\delta_{2j} - 2\delta_{1j}\delta_{2j})S_1(t_{1j}, t_{2j}|\boldsymbol{\theta}, \boldsymbol{\psi}) + (\delta_{1j} - 2\delta_{1j}\delta_{2j})S_2(t_{1j}, t_{2j}|\boldsymbol{\theta}, \boldsymbol{\psi}) \\
& + (1 - 2\delta_{1j} - 2\delta_{2j} + 4\delta_{1j}\delta_{2j})S_{12}(t_{1j}, t_{2j}|\boldsymbol{\theta}, \boldsymbol{\psi}),
\end{aligned}$$

where δ_{1j} and δ_{2j} are equal to 0 when he or she is at risk for respectively the first and second event under consideration, and δ_{1j} and δ_{2j} equal 1 whenever the individual is not at risk. In both applications, the unit of time refers to the age a_{ij} of individual $j = 1, \ldots, n$ at the cross-sectional sampling time for infection $i = 1, 2$. Furthermore, univariate monitoring times are present such that $a_{1j} = a_{2j} = a_j$. In infectious disease epidemiology, the hazard function is often termed the infection hazard or force of infection and can be estimated from serological survey data. In both applications, all available serology is included by means of a direct likelihood approach; individuals with a missing immunological status for one of the infections under consideration contribute to the log-likelihood as follows:

$$\text{ll}_j(\boldsymbol{\theta}, \boldsymbol{\psi}|t_{ij}, \delta_{ij}) = \delta_{ij} \log\{\pi_i(t_{ij}|\boldsymbol{\theta}, \boldsymbol{\psi})\} + (1 - \delta_{ij}) \log\{1 - \pi_i(t_{ij}|\boldsymbol{\theta}, \boldsymbol{\psi})\}.$$

4.1. Hepatitis A and B infections

Time-dependent frailty models are fitted to the HAV and HBV serology. Gompertz infection-specific baseline hazard functions $\lambda_{i0}(a_j)$, $i = 1, 2$, are considered here, i.e. $\lambda_{i0}(a_j) = \xi_i \exp(\nu_i a_j)$, where $\xi_i > 0$ and $-\infty < \nu_i < \infty$. In addition, both infections are presumed to confer lifelong immunity, entailing SIR infection dynamics. For more details on the models that were fitted to the Flemish serological survey data on HAV and HBV infections, and their names, we refer to Table 1.

In Table 2, maximum likelihood estimates and standard error estimates for the model parameters are presented. Models are compared on the basis of the Akaike information criterion (AIC) and Bayesian information criterion (BIC). An improvement in model fit is observed when comparing age-invariant frailty models (SGF and CGF) with their age-dependent counterparts, underlining the importance of accommodating time varying heterogeneity. Leaving aside the piecewise constant model because of its different interpretation, shared frailty models with (TDSGF-1C*) and without unequal frailty variances (TDSGF-1C) provide the best fit among

Table 1. Overview of frailty models fitted to the Flemish seroprevalence data on HAV and HBV, and the Belgian seroprevalence data on PVB19 and VZV

Model	Infection process		Heterogeneity	Z^*_{ijm}
	HAV	HBV		
SGF	Immunizing	Immunizing	Time invariant	Shared
CGF	Immunizing	Immunizing	Time invariant	Correlated
TDSGF-1C	Immunizing	Immunizing	One-component time dependent with $\phi_{11}=\phi_{21}$	Shared
TDSGF-1C*	Immunizing	Immunizing	One-component time dependent	Shared
TDCGF-1C	Immunizing	Immunizing	One-component time dependent with $\phi_{11}=\phi_{21}$	Correlated
TDPCSGF	Immunizing	Immunizing	Piecewise constant time dependent	Shared
	PVB19	VZV		
SGF-1	Immunizing	Immunizing	Time invariant	Shared
TDSGF-1-1C	Immunizing	Immunizing	One-component time dependent	Shared
TDSGF-1-2C	Immunizing	Immunizing	Two-component time dependent	Shared
SGF-2	Non-immunizing	Immunizing	Time invariant	Shared
TDSGF-2-1C	Non-immunizing	Immunizing	One-component time dependent	Shared
TDSGF-2-2C	Non-immunizing	Immunizing	Two-component time dependent	Shared

all fitted multiplicative models based on their respective AIC and BIC values respectively, though the correlated frailty model (TDCGF-1C) yields a lower value for the log-likelihood compared with the shared models, at the cost of an additional parameter. Two-component shared frailty models (TDSGF-2C) with either equal or unequal decay rates for both infections did not improve the model fit (see the on-line appendix D). The time-dependent piecewise constant shared gamma frailty model is considered to assess the goodness of fit of the imposed variance functions or, equivalently, the exponential decay functions $h_{i1}(t_{ij})$. In model TDPCSGF, we selected four age groups [0, 12), [12, 25), [25, 65) and older than 65 years for which independent frailty terms are considered. Note that this model enables greater flexibility in modelling age-dependent heterogeneity; however, the choice of the age intervals is subjective and influences the model fit. The piecewise constant model (TDPCSGF) outperforms all other presented frailty models on the basis of AIC, albeit that model TDSGF-1C has a smaller BIC value. We clearly see that heterogeneity is highest at young ages. Variability decreases with time and increases again in the last age group, albeit that a limited amount of data is available therein (large standard error). Although slightly outperforming model TDSGF-1C*, little evidence against an exponential decay in heterogeneity is obtained from model TDPCSGF. In spite of its improved fit, the biological interpretation of model TDPCSGF is not straightforward and independence of the frailty components is unrealistic. Nevertheless, fitting model TDPCSGF is useful to investigate the time varying shape of the frailty variance. The estimated multinomial probabilities for TDSGF-1C* (the full curve) and TDPCSGF (the dotted curve) are shown in Fig. 2. The TDPCSGF fit deviates from the TDSGF-1C* fit at higher ages. In Fig. 3, age-dependent frailty variances for HAV (Fig. 3(a)) and HBV (Fig. 3(b)) are presented, signalling a faster decrease for hepatitis B.

Estimating the correlation between infection-specific frailty terms could reflect to what extent latent processes, such as social contact behaviour of people, susceptibility to infection and infectiousness after infection, drive the more general infection process, or could reveal transmission via similar transmission routes when the correlation is perfect. Shared frailty models are restrictive in that they assume a perfect correlation implying common activity levels for both

Table 2. Maximum likelihood estimates and standard errors for the model parameters with corresponding AIC and BIC values†

| Parameter	Estimates for the following models:					
	SGF	*CGF*	*TDSGF-1C,* $\phi_{11}=\phi_{21}$	*TDSGF-1C‡,* $\phi_{11}\neq\phi_{21}$	*TDCGF-1C,* $\phi_{11}=\phi_{21}$	*TDPCSGF*
ξ_1	0.012 (0.001)	0.007 (0.001)	0.077 (0.029)	0.139 (0.093)	0.151 (0.105)	0.029 (0.006)
ν_1	0.036 (0.005)	0.106 (0.017)	−0.012 (0.007)	−0.021 (0.009)	−0.022 (0.009)	0.009 (0.005)
ξ_2	0.002 (4×10^{-4})	0.002 (4×10^{-4})	0.003 (0.001)	0.003 (0.001)	0.003 (0.001)	0.002 (4×10^{-4})
ν_2	−0.003 (0.008)	−0.001 (0.014)	−0.009 (0.008)	−0.012 (0.009)	−0.008 (0.008)	−0.005 (0.008)
$\sqrt{\gamma_{11}}$	0.725 (0.086)	1.651 (0.176)	5.843 (0.829)	6.444 (1.013)	6.606 (1.020)	3.671 (0.606)
$\sqrt{\gamma_{21}}$	0.725 (0.086)	1.608 (2.272)	5.843 (0.829)	6.444 (1.013)	5.765 (0.831)	3.671 (0.606)
$\sqrt{\gamma_{\cdot2}}$	—	—	—	—	—	2.421 (0.504)
$\sqrt{\gamma_{\cdot3}}$	—	—	—	—	—	0.012 (0.160)
$\sqrt{\gamma_{\cdot4}}$	—	—	—	—	—	8.813 (7.856)
ϕ_{11}	—	—	0.034 (0.005)	0.026 (0.007)	0.025 (0.007)	—
ϕ_{21}	—	—	0.034 (0.005)	0.044 (0.011)	0.025 (0.007)	—
ρ	1.000 (—)	0.497 (0.702)	1.000 (—)	1.000 (—)	0.871 (0.080)	1.000 (—)
AIC	5824.90	5794.89	5756.01	5755.52	5757.04	*5749.01*
BIC	5856.41	5838.99	*5793.82*	5799.62	5807.44	5799.42

†Minima are indicated by italics.
‡Unequal frailty variances.

infections, i.e. the propensity to make contacts that are relevant for disease transmission, without accounting for unshared sources of heterogeneity, e.g. differences in susceptibility between infections. Here, little evidence against the hypothesis of a perfect correlation is present. Estimation of important epidemiological parameters such as the basic and effective reproduction number for HAV and HBV can be done by imposing mixing patterns via the so-called 'Who acquires infection from whom' approach. For more details thereon, we refer to Anderson and May (1991) and Hens *et al.* (2012). In Section 4.2, we show the estimation of the reproduction numbers for PVB19 and VZV when relying on social contact data.

4.2. Parvovirus B19 and varicella zoster virus

The baseline infection hazard $\lambda_{i0}(a_j)$ with respect to infection i is modelled by means of the time homogeneous mass action principle, which is a specific parametric hazard that is often used in infectious disease modelling. However, our approach is generally applicable to all types of parametric hazard, and therefore it is not mandatory for readers to grasp all the complexities of the mechanistic model. The mass action principle relates the infection hazard to social contact data through the specification of the so-called social contact hypothesis, assuming it to depend on contact rates and a constant proportionality factor q_i. A thorough description of the mass action principle can be found in the on-line appendix E.

An important transmission parameter describing the spread of an infection is the so-called basic reproduction number. The basic reproduction number R_0 is defined as the expected number of secondary cases produced by a single typical infectious individual during his or her entire infectious period when introduced into a fully susceptible population. Furthermore, R_0 represents an epidemiological threshold parameter in large populations for which a value larger than 1 implies that an infection can invade the population, whereas the infection will become extinct otherwise. More realistically, the effective reproduction number R_e reflects the actual
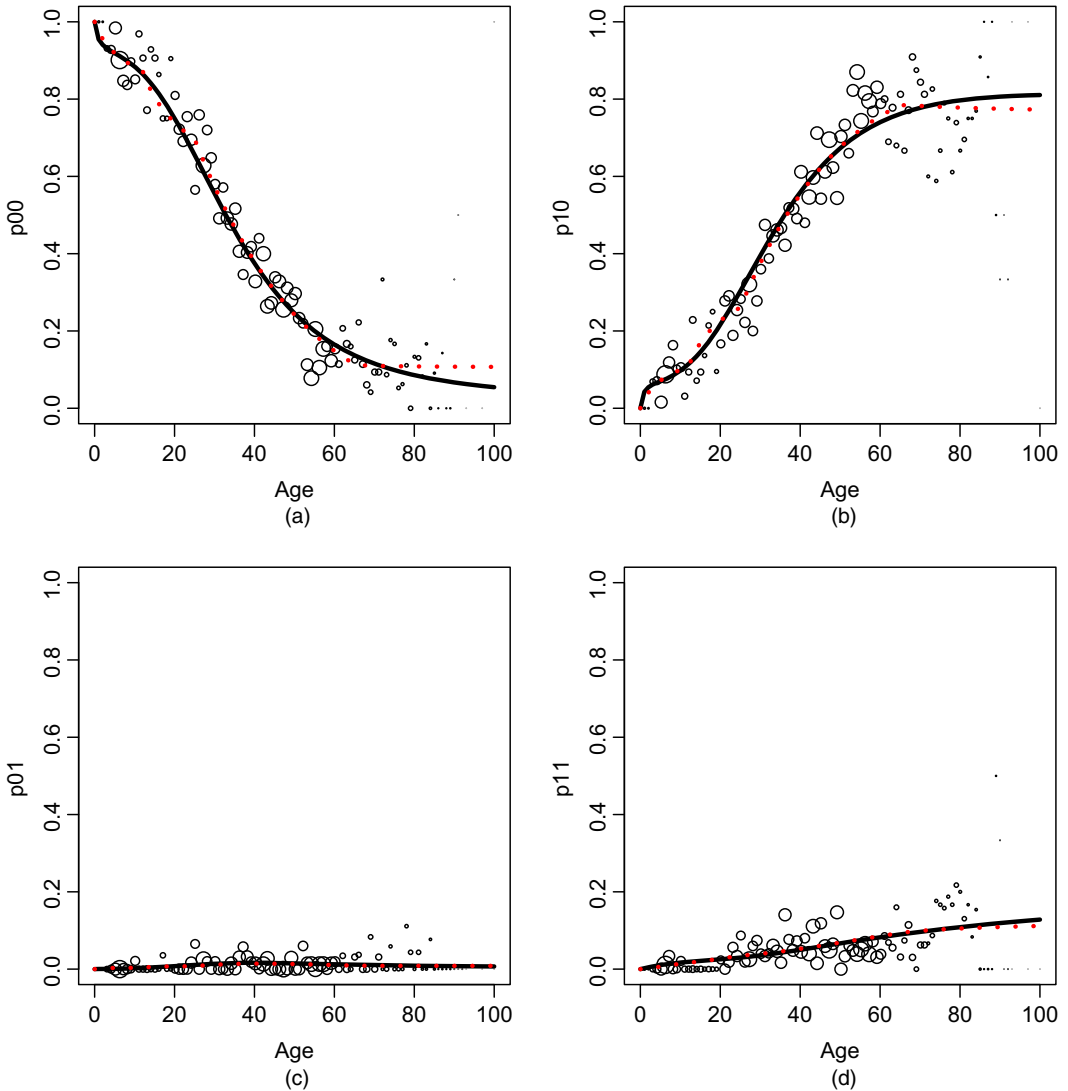
Fig. 2. Estimated multinomial probabilities $\hat{p}_{\delta_1,\delta_2}(\cdot)$ for models TDCGF-1C* (⸻) and TDPCSGF (•••••) when applied to the bivariate serological survey data on HAV and HBV infections in Flanders, Belgium, years 1993–1994: (a) $\hat{p}_{00}(a_j)$; (b) $\hat{p}_{10}(a_j)$; (c) $\hat{p}_{01}(a_j)$; (d) $\hat{p}_{11}(a_j)$

expected number of secondary cases in a population with pre-existing immunity. The effective reproduction number R_{ie} related to infection i, or alternatively R_{i0} when $S_i(a_j) = 1, \forall a_j$, can be calculated as the leading eigenvalue of the next generation matrix (see, for example, Farrington *et al.* (2013a) and the on-line appendix E). The results of fitting time varying bivariate shared frailty models to the Belgian serology on PVB19 and VZV are presented in Table 3. In addition to the parameter estimates and profile likelihood confidence limits, AIC and BIC values are shown for model comparison. Table 1 presents an overview of the models that were fitted to the Belgian serology on PVB19 and VZV.

Although the SGF-2 and TDSGF-2 frailty models cover more complex disease dynamics, the improved fit to the data reveals that the assumption of lifelong immunity for PVB19 seems
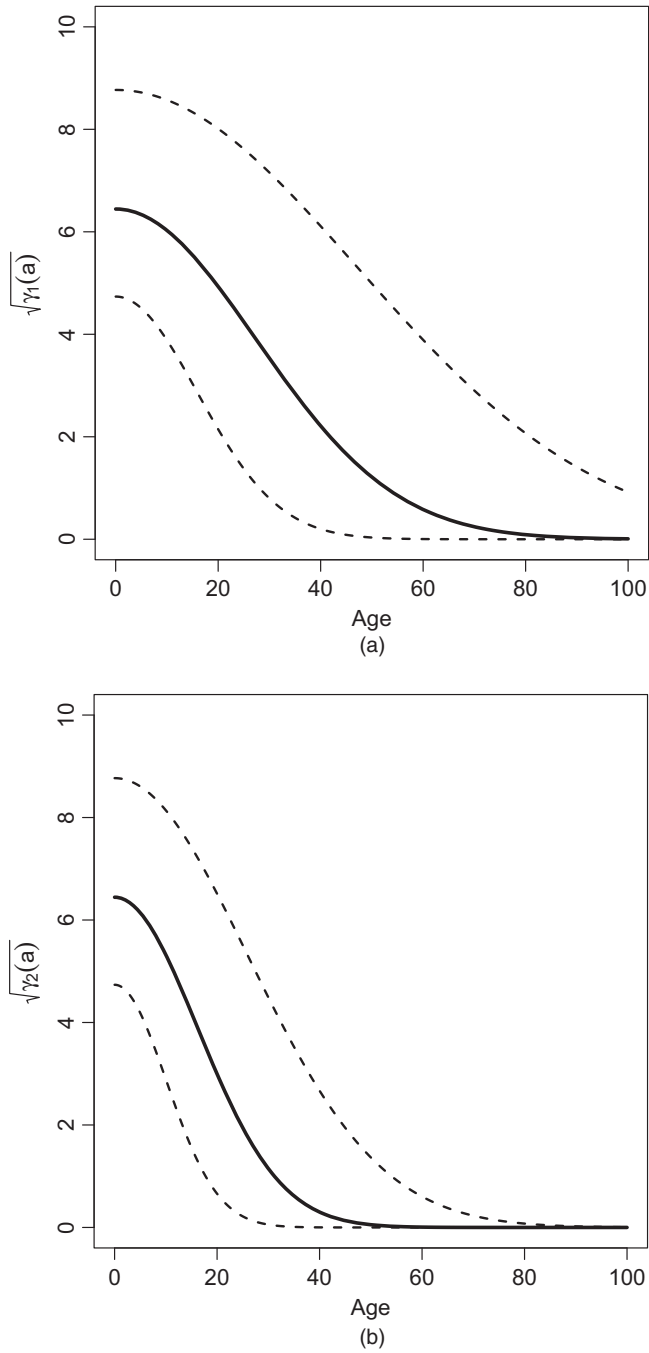
Fig. 3. Age-dependent frailty standard deviations (———) in model TDSGF-1C* and corresponding 95% confidence-limits (- - - - -) for (a) HAV and (b) HBV

Table 3. Maximum likelihood estimates for the model parameters as well as for the basic and effective reproduction numbers R_{i0} and R_{ie} (PVB19, $i = 1$; VZV, $i = 2$), with 95% profile likelihood confidence intervals in square brackets and AIC and BIC values†

Model		Estimate	$\hat{R}_{i0}$	$\hat{R}_{ie}$	AIC	BIC
SGF-1	q_{10}	0.072 [0.069, 0.075]	3.60 [3.35, 3.88]	1.268 [1.209, 1.334]	4937.14	4955.51
	q_{20}	0.200 [0.188, 0.214]	11.64 [10.59, 12.82]	1.488 [1.397, 1.591]		
	γ	0.152 [0.118, 0.188]				
TDSGF-1-1C	q_{10}	0.072 [0.069, 0.076]	3.60 [3.22, 3.99]	1.268 [1.169, 1.375]	4939.14	4963.64
	q_{20}	0.200 [0.183, 0.221]	11.64 [9.99, 13.49]	1.488 [1.348, 1.656]		
	$\gamma_{.1}$	0.152 [0.100, 0.210]				
	$\phi_{.1}$	0.000 [0.000, 0.009]				
TDSGF-1-2C	q_{10}	0.066 [0.062, 0.071]	3.74 [3.15, 4.87]	1.735 [1.286, 2.894]	4912.08	4942.70
	q_{20}	0.235 [0.191, 0.299]	15.65 [11.38, 24.08]	4.957 [2.417, 11.082]		
	$\gamma_{.1}$	2.918 [1.524, 5.004]				
	$\gamma_{.2}$	0.233 [0.156, 0.323]				
	$\phi_{.1}$	0.316 [0.246, 0.425]				
SGF-2	q_{10}	0.071 [0.068, 0.074]	3.18 [2.97, 3.43]	1.100 [1.051, 1.157]	4869.83	4894.33
	σ	0.011 [0.008, 0.015]				
	q_{20}	0.173 [0.163, 0.183]	8.98 [8.22, 9.83]	1.207 [1.141, 1.282]		
	γ	0.032 [0.002, 0.065]				
TDSGF-2-1C	q_{10}	0.065 [0.061, 0.070]	2.90 [2.64, 3.49]	1.142 [1.047, 2.565]	4862.93	*4893.56*
	σ	0.012 [0.009, 0.016]				
	q_{20}	0.158 [0.141, 0.179]	8.19 [7.15, 10.46]	1.890 [1.193, 4.225]		
	$\gamma_{.1}$	1.470 [0.415, 3.498]				
	$\phi_{.1}$	0.330 [0.209, 0.530]				
TDSGF-2-1C, unequal frailty variances	q_{10}	0.065 [0.060, 0.070]	2.94 [2.60, 4.97]	1.242 [1.046, 3.405]	4863.83	4900.57
	σ	0.013 [0.009, 0.021]				
	q_{20}	0.154 [0.133, 0.175]	7.98 [6.76, 11.83]	1.873 [1.160, 6.605]		
	$\gamma_{.1}$	1.646 [0.459, 6.443]				
	ϕ_{11}	0.239 [0.141, 0.648]				
	ϕ_{21}	0.377 [0.226, 0.677]				
TDSGF-2-2C	q_{10}	0.066 [0.063, 0.071]	3.30 [2.79, 4.45]	1.453 [1.083, 2.706]	*4859.26*	4896.01
	σ	0.011 [0.007, 0.015]				
	q_{20}	0.193 [0.156, 0.257]	11.27 [8.11, 18.90]	3.304 [1.473, 8.897]		
	$\gamma_{.1}$	2.419 [0.839, 4.960]				
	$\gamma_{.2}$	0.095 [0.017, 0.186]				
	$\phi_{.1}$	0.303 [0.226, 0.423]				
TDSGF-2-2C, unequal frailty variances	q_{10}	0.066 [0.063, 0.081]	3.40 [2.78, 6.17]	1.585 [1.072, 4.107]	4860.73	4903.60
	σ	0.012 [0.007, 0.020]				
	q_{20}	0.188 [0.151, 0.251]	10.98 [7.85, 19.47]	3.257 [1.416, 10.200]		
	$\gamma_{.1}$	2.554 [0.861, 5.994]				
	$\gamma_{.2}$	0.095 [0.016, 0.181]				
	ϕ_{11}	0.249 [0.160, 0.706]				
	ϕ_{21}	0.327 [0.228, 0.486]				

†The minima are indicated in italics.

questionable. Furthermore, the age-dependent frailty models that are described in this paper tend to outperform their age invariant counterparts. This leads to the conclusion that variability is indeed higher in the younger age groups and decreases with age until a constant frailty variance γ_2 is achieved (Fig. 4). A graphical illustration of the TDSGF-2-2C model fit can be found in the on-line appendix C.

5. Discussion

First, this paper presents an extension of traditional shared and correlated frailty models to
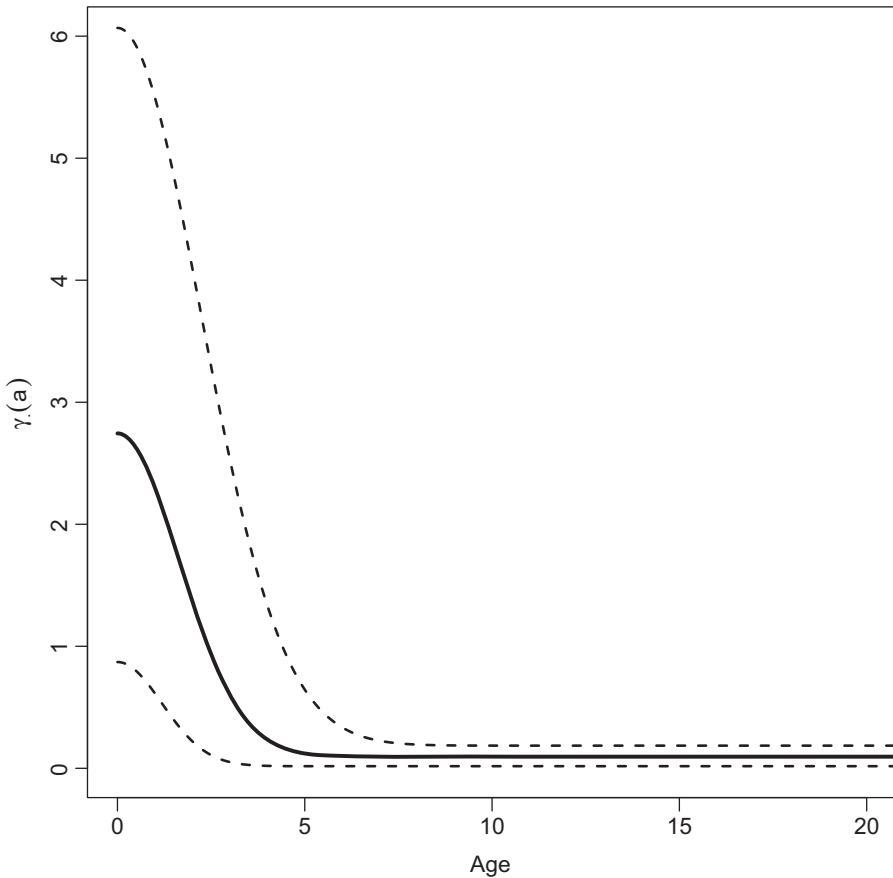
Fig. 4.   Age-dependent frailty variance (———) for model TDSGF-2-2C and corresponding 95% confidence limits (- - - - -) when applied to the bivariate serological survey data on PVB19 and VZV in Belgium, years 2001–2003

be applicable in the context of non-immunizing infections (i.e. infections conferring temporary humoral immunity on recovery) while allowing for time varying heterogeneity in their acquisition. The approach proposed combines frailty methodology for non-immunizing infections with time varying heterogeneity, at least while assuming a shared frailty with respect to multiple infections. The use of a time varying correlated frailty model, however, is complicated by the non-identifiability of the unshared components in the additive decomposition of the infection-specific frailty terms. Identifiability of the time invariant correlated gamma frailty model was formally established for bivariate time-to-event data in the presence of covariates by Iachine (2004), and a simulation-based assessment for CS data ensures identifiability at the cost of parametric baseline hazards (Hens *et al.*, 2009). Nevertheless, identifiability in the time varying setting for recurrent infections seems questionable. In addition to the aforementioned time varying frailty models, we propose novel multiplicative time varying correlated frailty models for infections conferring lifelong immunity. Expressions for the time-dependent correlation coefficients are derived for the two-component models, albeit that those can be readily generalized.

Age-dependent shared frailty models are fitted to bivariate serological data on childhood infections PVB19 and VZV, and these models were compared with age invariant models ei-

ther with or without accommodating potential PVB19 reinfections. The results support the hypothesis of potential PVB19 reinfections as argued by Abrams and Hens (2015). In addition, age-dependent shared frailty models outperform their age invariant counterparts. In general, the age-dependent shared gamma frailty models with potential reinfection processes for PVB19 (TDSGF-2-1C and TDSGF-2-2C) were found to be the best-fitting models to the data at hand, and estimates of the reproduction numbers for PVB19 and VZV were in line with those derived in the presence of age invariant individual heterogeneity (Abrams and Hens, 2015). Childhood infections are typically highly correlated within individuals in early childhood, which was observed as well for PVB19 and VZV, with the correlations persisting into adulthood only for infections having similar transmission routes (Farrington *et al.*, 2013b). Heterogeneity in childhood could be explained by confounding of different transmission routes and could be due to the nature of contacts made at young ages. At older ages, however, behaviour and environmental factors tend to change such that transmission routes become differentiated and as social factors such as school attendance intervene the heterogeneity drops either to 0 (for infections with different transmission routes) or to some constant (reflecting variability in behaviour associated with a common route of transmission). In addition to confounding of different transmission routes, demographic factors, such as differences in socio-economic status or ethnicity, could explain some of the unobserved individual heterogeneity at young ages, albeit that such information was absent in our data. In shared frailty models, frailty terms have the interpretation of activity levels, meaning that individuals with a high social activity level are more likely to acquire both infections. On top of that, susceptibility to infection is presumed to be equal for both infections. For infections with different routes of transmission and/or different levels of susceptibility, correlated frailties are more suited to describe heterogeneity as well as association in the acquisition of both infections.

In addition, age-dependent shared and correlated frailty models are applied to bivariate serological survey data on HAV and HBV. In this application, we relied on the assumption of lifelong immunity for both infections. Time varying frailty terms improve the model fit substantially, and a decreasing frailty variance with age is observed. Although the piecewise constant shared gamma frailty model TDPCSGF allows for varying frailty variances within different age groups, its biological interpretation is difficult and the assumption of mutual independence between frailty components is unrealistic. In this setting, one-component age-dependent correlated gamma frailty models are found to be identifiable at the cost of parametric baseline hazards and common decay parameters for both infections. Identifiability of the frailty models is investigated on the basis of the general methodology of detecting parameter redundancy (Catchpole and Morgan, 1997, 2001). A formal proof of identifiability in the context of immunizing infections is beyond the scope of this paper. Note that the underlying transmission model for HBV did not include a carrier state in which individuals remain infective for life. Because of the low prevalence of HBV in Flanders, we believe that this simplification, albeit unrealistic to some extent, does not have a large effect on the results that are presented in this paper.

The proposed time varying frailty models are based on an explicit parametric model for the frailty variables in terms of independent gamma frailty components. This was suggested by Farrington *et al.* (2012) due to empirical evidence of a decreased heterogeneity with time (age) for several childhood infections. However, further research is required to test the hypothesis of an increased variability in the acquisition of (childhood) infections earlier in life, especially when these frailty models are applied to infections that mainly affect adolescents and the elderly. Although the computationally convenient gamma frailty distribution is considered throughout this paper, other frailty distributions could be used without additional work whenever closed form Laplace transforms are available. Frailty distributions without a closed form Laplace

transform could be considered as well; however, multivariate numerical integration techniques are required to integrate out the random frailties, thereby increasing the computational burden. A shortcoming of the multiplicative time-dependent frailty models is a time dependence of the range of $Z_{ij}(t_{ij})$ which is required to maintain $Z_{ij}(t_{ij}) > 0$. Recently, Enki *et al.* (2014) proposed a new family of time varying shared frailty models, based on power transformations rather than linear models, overcoming the disadvantage of a time-dependent frailty range in the context of immunizing infections. Further research could entail improvements in modelling time-dependent heterogeneity for non-immunizing infections by using these frailty models, albeit that preliminary analyses yield similar results (which are not shown) in terms of model fit. Although copula and counting process models could be used instead of (time varying) frailty models for the analysis of bivariate survival data, these approaches are not pursued in this paper. The connection between frailty and copula models has been studied by, for example, Goethals *et al.* (2008). For more information about the counting process approach, we refer to Aalen *et al.* (2008).

Acknowledgements

References

Aalen, O. O., Borgan, Ø. and Gjessing, H. (2008) *Survival and Event History Analysis: a Process Point of View*. Berlin: Springer.

Abrams, S. and Hens, N. (2015) Modeling individual heterogeneity in the acquisition of recurrent infections: an application to parvovirus B19. *Biostatistics*, **16**, 129–142.

Andersen, P. K. and Keiding, N. (2002) Multi-state models for event history analysis. *Statist. Meth. Med. Res.*, **11**, 91–115.

Anderson, R. M. and May, R. M. (1991) *Infectious Diseases of Humans; Dynamics and Control*. Oxford: Oxford University Press.

Beutels, M., Van Damme, P., Aelvoet, W., Desmyter, J., Dondeyne, F., Goilav, C., Mak, R., Muylle, L., Pierard, D., Stroobant, A., Van Loock, F., Waumans, P. and Vranckx, R. (1997) Prevalence of hepatitis A, B and C in the Flemish population. *Eur. J. Epidem.*, **13**, 275–280.

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

Catchpole, E. and Morgan, B. (2001) Deficiency of parameter redundant models. *Biometrika*, **88**, 593–598.

Clayton, D. G. (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141–151.

Cox, D. R. (1972) Regression models and life-tables (with discussion). *J. R. Statist. Soc.* B, **34**, 187–220.

Duchateau, L. and Janssen, P. (2008) *The Frailty Model*. Berlin: Springer.

Enki, D. G., Noufaily, A. and Farrington, C. P. (2014) A time-varying shared frailty model with application to infectious diseases. *Ann. Appl. Statist.*, **8**, 430–447.

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 (with discussion). *Appl. Statist.*, **50**, 251–292.

Farrington, C. P., Unkel, S. and Anaya-Izquierdo, K. (2012) The relative frailty variance and shared frailty models. *J. R. Statist. Soc.* B, **74**, 673–696.

Farrington, C. P., Unkel, S. and Anaya-Izquierdo, K. (2013a) Estimation of basic reproduction numbers: individual heterogeneity and robustness to perturbation of the contact function. *Biostatistics*, **14**, 528–540.

Farrington, C. P., Whitaker, H. J., Unkel, S. and Pebody, R. (2013b) Correlated infections: quantifying individual heterogeneity in the spread of infectious diseases. *Am. J. Epidem.*, **177**, 474–486.

Goethals, K., Janssen, P. and Duchateau, L. (2008) Frailty models and copulas: similarities and differences. *J. Appl. Statist.*, **35**, 1071–1079.

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

Hens, N., Aerts, M., Shkedy, Z., Theeten, H., Van Damme, P. and Beutels, P. (2008) Modelling multi-sera data: the estimation of new joint and conditional epidemiological parameters. *Statist. Med.*, **27**, 2651–2664.

Hens, N., Shkedy, Z., Aerts, M., Faes, C., Van Damme, P. and Beutels, P. (2012) *Modeling Infectious Disease Parameters based on Serological and Social Contact Data: a Modern Statistical Perspective*. New York: Springer.

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. *Statist. Med.*, **27**, 2785–2800.

Hougaard, P. (2000) *Analysis of Multivariate Survival Data*. New York: Springer.

Iachine, I. A. (2004) Identifiability of bivariate frailty models. *Preprint 5*. Department of Statistics, University of Southern Denmark, Odense.

Lancaster, T. (1979) Econometric models for the duration of unemployment. *Econometrica*, **47**, 939–956.

Paik, M. C., Tsai, W.-Y. and Ottman, R. (1994) Multivariate survival analysis using piecewise gamma frailty. *Biometrics*, **50**, 975–988.

Sun, J. (2006) *The Statistical Analysis of Interval-censored Failure Time Data*. New York: Springer.

Therneau, T. and Grambsch, P. (2000) *Modelling Survival Data: Extending the Cox Model*. New York: Springer.

Unkel, S., Farrington, C. P., Whitaker, H. J. and Pebody, R. (2014) Time varying frailty models and the estimation of heterogeneities in transmission of infectious diseases. *Appl. Statist.*, **63**, 141–158.

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

Wienke, A., Christensen, K., Skytthe, A. and Yashin, A. I. (2002) Genetic analysis of cause of death in a mixture model with bivariate lifetime data. *Statist. Modllng*, **2**, 89–102.

Wienke, A., Holm, N., Christensen, K., Skytthe, A., Vaupel, J. and Yashin, A. I. (2003) The heritability of cause-specific mortality: a correlated gamma-frailty model applied to mortality due to respiratory diseases in Danish twins born 1870–1930. *Statist. Med.*, **22**, 3873–3887.

Xue, X. and Brookmeyer, R. (1996) Bivariate frailty model for the analysis of multivariate survival time. *Liftim. Data Anal.*, **2**, 277–290.

Yashin, A. I., Vaupel, J. W. and Iachine, I. A. (1995) Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. *Math. Popln Stud.*, **5**, 145–159.