Journal Title
XX(X):1–12

DOI: 10.1177/ToBeAssigned
www.sagepub.com/

Dynamic prediction of recurrent events data by landmarking with application to a follow-up study of patients after kidney transplant

J.Z. Musoro[1] G.H. Struijk[2], R.B. Geskus[1], I.J.M. ten Berge[2], A.H. Zwinderman[1]

Abstract

This paper extends dynamic prediction by landmarking to recurrent event data. The motivating data comprised post kidney transplantation records of repeated infections and repeated measurements of multiple markers. At each landmark time point t_s, a Cox proportional hazards model with a frailty term was fitted using data of individuals who were at risk at landmark s. This model included the time-updated marker values at t_s as time-fixed covariates. Based on a stacked data set that merged all landmark data sets, we considered supermodels that allow parameters to depend on the landmarks in a smooth fashion. We described and evaluated four ways to parameterize the supermodels for recurrent event data. With both the study data and simulated data sets, we compared supermodels that were fitted on stacked data sets that consisted of either overlapping or non-overlapping landmark periods. We observed that for recurrent event data, the supermodels may yield biased estimates when overlapping landmark periods are used for stacking. Using the best supermodel amongst the ones considered, we dynamically estimated the probability to remain infection free between t_s and a prediction horizon t_{hor}, conditional on the information available at t_s.

Keywords

Dynamic prediction; Frailty models; Landmark; Multiple markers; Recurrent events.

[1]Department of Clinical Epidemiology, Biostatistics and Bioinformatic Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands
[2]Renal Transplant Unit, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands

Corresponding author:

,
Email: z.j.musoro@amc.uva.nl

Prepared using sagej.cls *[Version: 2015/06/09 v1.01]*

Introduction

In clinical practice it is commonplace to monitor markers of the disease state of patients at several time points, alongside information on non terminal events. Interest often focuses on characterizing the relationship between both quantities, and a typical quest is that of evaluating the prognostic value of the markers for the event time, as well as how updated marker information changes prognosis. This is valuable in understanding the disease process, and may help clinicians to make better decisions as regards to patient care. The motivating example considered in this paper comprise follow-up data of patients after kidney transplantation[??]. Patients experienced repeated opportunistic infections of different types and several immunlogical markers were used to monitor their immune status over time. Since infection complications are one of the leading causes of morbidity and mortality amongst such a patient population[?], it is important to be able to identify frailer patients who, for instance, might benefit from a modification of their immunosuppressive treatment or receive prophylaxes. Therefore tools that make use of updated patient information (e.g. updated marker values) to describe their future risk of infection are advocated.

Consider the time to an event T and a longitudinal marker Y. Let $Y^H(t)$ be the entire marker history until time t. Survival analysis methods for time-dependent covariates that model the hazard of the form

$$\lambda\left\{t|Y^H(t)\right\} = \lim_{dt\downarrow 0}\Pr\left\{t < T \leq t + dt|T > t, Y^H(t)\right\}/dt \tag{1}$$

have been used to associate the time-dependent marker history $Y^H(t)$ with the instantaneous hazard $\lambda(t)$. In order to compute survival probabilities based on the time-dependent marker values, knowledge of the distribution of marker development is required. One way to do this is to use joint models for $(Y(), T)$[??????]. Alternatively, predictive probabilities can be obtained by creating simple survival models at certain moments in time for individuals who are still at risk at those moments. This is the subject of landmarking or partly conditional modelling[???].

The landmarking paradigm offers a flexible and relatively simple way to characterize the association between a longitudinal marker and the time until an event. At each landmark s, the longitudinal marker value at a landmark time point t_s is handled as a time-fixed variable, hence facilitating the direct prediction of event risks. Statistical aspects of dynamic prediction of time-to-event data has received a lot of attention[???????]. An overview of existing methods for dynamic prediction based on landmarking has been provided by Van Houwelingen and Putter[?]. They restrict their overview to single-type events that can occur only once. Extensions to a competing risks scenario have also been addressed[???]. In this paper, we generalize the landmarking approach to a setting with recurrent events.

We focus on the repeated urinary tract infections and four immunological markers (CD4+ T cell, CD8+ T cell, natural killer(NK) cell, and B cell counts) from the post-kidney transplantation data set[??]. Our main interest was to dynamically predict the infections until defined prediction horizons t_{hor}'s, conditioning on information that was available at landmark time points (e.g. the updated marker values at t_s). At each landmark, since patients could experience more than one infection, a Cox proportional hazards model with a frailty term was fitted. Dependency of the hazards on markers was assumed to be via fitted marker values at the landmark time point t_s that were obtained using mixed effects models. Since some patients had a large time difference between the last marker measurement and the landmark time point, it was more plausible to use fitted values as opposed to the commonly used last observed marker value before t_s. This has the extra benefit of accounting for possible measurement error in the

original marker measurements. Specific to recurrent events data is the possibility to use observed event history as a characteristic of patients' state of health to predict the risk of future events. Hence the number of previous infections within a time window, prior to a given landmark time point was considered as a time-updated covariate as well.

Additionally, we fitted supermodels[??] to an enlarged data set that stacked all landmark data sets together. Usually, in settings with single-type events, the stacked data set comprises overlapping landmark periods[?]. For recurrent events data, it could be a problem to use overlapping landmark periods since there might be multiple counting of recurrent events over the different landmark data sets. Using both the study data and simulated data sets, we compared supermodels that were fitted to stacked data sets that comprised either overlapping or non-overlapping landmark periods. We considered four ways of parameterizing a supermodel in the context of recurrent events data. This was based on making different model assumptions to describe the correlation between the separate landmark frailty terms, as well as how parameters varied over landmark periods. We settled for parameterizations that can be implemented with existing software. Using our best supermodel amongst the ones considered, we estimated the probability to experience at least one infection until t_{hor}, conditional on the information available at t_s. Predictive performance was quantified using the area under the ROC curve (AUC).

The rest of this paper is organized as follows. A description of the motivating data is given in section 2. Section 3 provides a brief description of the landmarking approach for a setting with a single event type that can occur only once. Section 4 introduces notation and a description of the landmark models for recurrent event data. In a simulation study in section 5, we evaluate if the parameter estimates from a supermodel are affected by how the stacked data is created. Two ways of creating a stacked data set are considered; (i) a setting with overlapping landmark periods where information from the same patient is repeatedly used over different landmark data sets and (ii) a setting with non-overlapping landmark periods where there is no overlap of information over the different landmark data sets. In section 6, we present an application of our landmark models to the kidney transplantation data, and performed dynamic prediction of infection probabilities based on our best supermodel. The last section presents a discussion.

The kidney transplantation data

The data comprised post kidney transplantation records of 358 patients with end-stage renal disease[?]. At transplantation, covariates such as age, gender and duration of dialysis prior to transplant were recorded. The duration of follow-up after transplant varied between 10 days and 9.66 years (median=3.1 years). During this period, 161 patients experienced at least one urinary tract infection (UTI). The number of infections per subject ranged between 1 and 10 (mean=2.5), with 17% of the infected individuals having at least 4 re-infections. The remaining 197 patients were free from UTI's throughout follow-up. The immunological markers CD4+ T cell, CD8+ T cell, natural killer(NK) cell, and B cell counts were repeatedly measured to monitor patients' immune status over time. Marker measurements were taken at irregular time intervals, with the number of repeated measurements varying between 1 and 18 (mean $= 3.0$) per patient. All markers were reported as the absolute number of cells ($\times 10^9/L$) and shall be referred to as cd4, cd8, nk and bcl respectively. Individual profiles for the markers (Figure 1) exhibited varying patterns within patients, which could be attributed to short term fluctuations and measurement error. Also, patients varied in how their markers evolved over time.

Landmark basics for single events

Here were refer to a setting with a single event type that can occur only once, for instance death. At each landmark s, $s = 1, 2, ...Q$, only patients who are still at risk at landmark time point t_s are selected. So patients that had an event or were censored before landmark s are removed. Also, only the information available from t_s up till a desired prediction horizon t_{hor}, which represents a landmark window, is used: $t_{hor} = t_s + w$, where w is the width of the landmark window. Van Houwelingen[?] argued that it suffices to fit a simple Cox model, that is with no time-dependent parameter, of the form

$$\lambda_s(t|Y(t_s)) = \lambda_{s0}(t)\exp\{Y(t_s)\beta(s)\} \quad (2)$$

to data on the interval (t_s, t_{hor}), enforcing administrative censoring at t_{hor}, to construct valid prediction models for $P\{T > t_{hor}|Y(t_s), T > t_s\}$. In (2), the value of the time-dependent covariate at t_s is handled as a time-fixed covariate. Even when the validity of the proportional hazards assumption is questionable, it is still a convenient and useful way to obtain dynamic predictions at t_{hor} without having to fit models with complicated time varying effects[?]. Furthermore, since landmark models handle a time-dependent covariate as a time-fixed covariate, they can be easily extended to accommodate several time-dependent covariates.

If $Q = 100$, then fitting (2) for each of the landmark points separately yields 100 sets of landmark specific parameters which are too many and might be difficult in practice to communicate, for instance, to a clinical user. Parsimony can be achieved by defining a supermodel that combines a sequence of landmark models in the range of Q landmark points and postulating how $\beta(s)$ and $\lambda_{s0}(.)$ change over s. Typically a stacked data set is created by combining all the separate landmark data sets. Since the stacked data often comprise overlapping information from the same subjects over the different landmark data sets, correct standard errors can be obtained by correcting for clustering via the sandwich estimator of Lin and Wei[?]. Landmark models usually require few modelling assumptions, hence are robust and can be easily implemented in practice since no specialized software is required. See[???] for more details.

Landmark model for recurrent events

Notation

A patient i $(i = 1, ..., n)$ could have an infection observed M_i times; M_i can be zero. After the last infection patients were followed further in time, and were right censored at the last visit. Let $T_{i1}, T_{i2}, ..., T_{iM_i}$ represent the recurrent event times and let T_{iM_i+1} be the censoring time, taken from time of transplant onwards. We assumed that the censoring mechanisms were independent of the infection time distribution and that values from the different patients were independent. Let $\mathbf{Z}_i = (\mathbf{Y}_i, \mathbf{X}_i)$ be a vector of covariates of patient i. $\mathbf{Y}_i = (\mathbf{y}_{i1}, ..., \mathbf{y}_{iK})$ where $\mathbf{y}_{ik}$ is vector of length n_{ik}, comprising values of the longitudinal marker of type k $(k = 1, 2, ..., K)$ measured at time points $t_{ij,k}$ $(j = 1, 2, ..., n_{ik})$. $\mathbf{X}_i$ is a vector of baseline covariates.

Recall that at landmark s, t_s is the landmark time point and the prediction horizon $t_{hor} = t_s + w$, where w is the window over which the conditional probabilities will be estimated. We refer to the interval (t_s, t_{hor}) as the landmark period. A landmark data set $data_s$ contains only patients who are at risk at t_s. Thus, the $data_s$ for patient i comprises $(T_{im}(s), \mathbf{Z}_i(s))$. $\{T_{im}(s)\}_{m\in\mathcal{I}}$, where $\mathcal{I} = \{1, ..., I_{is}\}$, are the infection times for patient i occurring in the interval $(t_s, t_{hor}]$. I_{is} denotes the number of infections for

subject i in that interval, and could be equal to zero in the case of no infections. We define $T_{i0}(s) = t_s$ and $T_{iI_{is}+1}(s) = \min(t_{hor}, T_{i,M_i+1})$. $\mathbf{Z}_i(s)$ comprises all baseline covariates, the fitted values of the time varying markers $\hat{\boldsymbol{Y}}_i(t_s)$ at t_s, and the number of infections $h_i(s)$ observed in the previous landmark period.

Model definition

First we consider the analysis per landmark. We assume the relationship between the rate of infection and $\mathbf{Z}$ to be according to a frailty model. For a single landmark point s, we postulate

$$\lambda_s(t|\mathbf{Z}_i(s), \upsilon_{is}, \boldsymbol{\theta}(s)) = \lambda_{0,s}(t)\exp\left\{\mathbf{X}_i^{\intercal}\boldsymbol{\alpha}(s) + \hat{\boldsymbol{Y}}_i^{\intercal}(t_s)\boldsymbol{\beta}(s) + h_i(s)\delta(s) + \upsilon_{is}\right\}, \tag{3}$$

with $t_s < t \leq t_s + w$. $\lambda_{0,s}(t)$ is the landmark specific baseline hazard. $\boldsymbol{\alpha}(s)$, $\boldsymbol{\beta}(s)$ and $\delta(s)$ are respectively vectors of parameters for all baseline covariates $\mathbf{X}_i$, the marker values $\hat{\boldsymbol{Y}}_i(t_s)$ and number of infections prior to s. $\hat{\boldsymbol{Y}}_i(s)$ at t_s were estimated using univariate linear mixed effects (LME) models where, similar to Musoro *et al.*[?], we included a natural cubic spline function of time for the population trend because it improved model fit. Each LME model included age, gender and duration of dialysis prior to transplant as fixed effects and was fitted using the complete follow-up information for each subject in the kidney transplantation data set. $\boldsymbol{\theta}(s) = (\boldsymbol{\alpha}(s), \boldsymbol{\beta}(s), \delta(s))$ and υ_{is} is the frailty term for subject i which is assumed to be $\sim N(0, \sigma_s)$. Applying model (3) separately to every landmark allows every component to be landmark specific. The likelihood for $data_s$ is given by

$$L_s(\boldsymbol{\theta}(s)) = \prod_{i=1}^{n_s}\int_{\upsilon_{is}}\left\{\prod_{m=1}^{I_{is}+1} f_{1s}(t_{im}(s)|\mathbf{Z}_i(s), \upsilon_{is}, \boldsymbol{\theta}(s))\right\} \times f_{2s}(\upsilon_{is}|\sigma_s)d\upsilon_{is} \tag{4}$$

n_s is the number of individuals at risk at landmark s, f_{2s} is a mean zero normal density function for the random effect υ_{is} of patient i for landmark s, and $f_{1s}(t_{im}(s)|\mathbf{Z}_i(s), \upsilon_{is}, \boldsymbol{\theta}(s))$ is defined as

$$f_{1s}(.) = \begin{cases} \lambda_s\left(t_{im}(s)|\mathbf{Z}_i(s), \upsilon_{is}, \boldsymbol{\theta}(s)\right) \times S_s\left(t_{im}(s)|\mathbf{Z}_i(s), \upsilon_{is}, \boldsymbol{\theta}(s), t_{im-1}(s)\right) & \text{if } m \in \mathcal{I} \\ S_s\left(t_{im}(s)|\mathbf{Z}_i(s), \upsilon_{is}, \boldsymbol{\theta}(s), t_{im-1}(s)\right) & \text{if } m = I_{is}+1 \\ S_s\left(t_i(s)|\mathbf{Z}_i(s), \upsilon_{is}, \boldsymbol{\theta}(s)\right) & \text{if } \mathcal{I} = \emptyset \end{cases} \tag{5}$$

$S_s(.)$ is the probability to remain infection-free for a specific individual since a previous infection given its frailty, and is defined as

$$S_s(t|\mathbf{Z}_i(s), \upsilon_{is}, \boldsymbol{\theta}(s), t_{im-1}(s)) = \exp\left\{-\int_{t_{im-1}(s)}^{t}\lambda_s(u|\mathbf{Z}_i(s), \upsilon_{is}, \boldsymbol{\theta}(s))du\right\}. \tag{6}$$

Fitting frailty models to each landmark-specific data set resulted in sets of baseline hazards $\left\{\hat{\lambda}_{s,0}(t)\right\}$, coefficients of fixed effects $\{\hat{\boldsymbol{\alpha}}(s)\}$, coefficients of covariates that were updated at new landmarks $\left\{\hat{\boldsymbol{\beta}}(s)\right\}$ and $\left\{\hat{\delta}(s)\right\}$, and landmark specific frailty variances $\left\{\sigma_s^2\right\}$.

Alternatively, the use of supermodels has been advocated to introduce parsimony by allowing parameters to depend on s in a smooth fashion. This required the construction of a stacked data set that merged all the $data_s$'s. Below, we propose four possible ways of parameterizing such supermodels for the case of recurrent events data. First we specified a general parameterization, then considered some constraints to settle for parameterizations that could be implemented within existing software.

If we use a stacked data set consisting of non-overlapping landmark periods, the likelihood of the general parameterization can be written as

$$L_1 = \prod_{i=1}^{n} \int_{\upsilon_{i1}} ... \int_{\upsilon_{iQ_i}} \left[\left\{ \prod_{s=1}^{Q_i} \prod_{m=1}^{I_{is}+1} f_{1s}(t_{im}(s)|\mathbf{Z}_i(s), \upsilon_{is}, \boldsymbol{\theta}(s)) \right\} \times f_2(\upsilon_{i1}, ..., \upsilon_{iQ_i}|\boldsymbol{\Sigma}_i) \right] d\upsilon_{i1}, ..., d\upsilon_{iQ_i} \tag{7}$$

where the number of landmark points for individual i is $Q_i \leq$ the total number of landmark points Q since patients with shorter follow-up time leave the study as time goes on. $f_{1s}(.)$ can be written similarly as in (5) with baseline hazards that are specific to s. $\upsilon_{i1}, ..., \upsilon_{iQ_i}$ are landmark-specific random effects of individual i with $f_2(.)$ a multivariate normal density function and $\boldsymbol{\Sigma}_i$ is a $Q_i \times Q_i$ covariance matrix. Notice that the regression weights $\boldsymbol{\theta}(s)$ are also landmark-specific. If the stacked data used to maximize (7) comprise overlapping landmark periods, then (7) can be referred to as a pseudo-likelihood since there are data from the same subject that are common to different landmark data periods. Hence the $data_s$ are not mutually independent, not even conditionally on $\upsilon_{i1}, ..., \upsilon_{iQ}$.

It is not straightforward to maximize (7) within standard software. Furthermore, it may involve estimating too many parameters; for instance, selecting 40 landmarks (as in our real data application below) will require that we estimate 820 variance and covariance parameters between the υ_{is}'s, apart from the fixed effects. We can constrain (7) in several ways to facilitate implementation within standard software. First, when $\upsilon_{i1}, ..., \upsilon_{iQ_i}$ are assumed to be independent, (7) simplifies to

$$L_2 = \prod_{i=1}^{n} \prod_{s=1}^{Q_i} \int_{\upsilon_{is}} \left\{ \prod_{m=1}^{I_{is}+1} f_{1s}(t_{im}(s)|\mathbf{Z}_i(s), \upsilon_{is}, \boldsymbol{\theta}(s)) \right\} \times f_2(\upsilon_{is}|\sigma_s) \; d\upsilon_{is}. \tag{8}$$

Note that performing the per-landmark analysis is the same as maximizing (8) except that there is no smoothing of $\theta(s)$ over landmarks. On the other hand, if we assume a perfect correlation between the υ_{is}'s, such that $\upsilon_{i1} = \upsilon_{i2} = ... = \upsilon_{iQ_i}$, then (7) simplifies to

$$L_3 = \prod_{i=1}^{n} \int_{\upsilon_i} \left\{ \prod_{s=1}^{Qi} \prod_{m=1}^{I_{is}+1} f_{1s}(t_{im}(s)|\mathbf{Z}_i(s), \upsilon_i, \boldsymbol{\theta}(s)) \right\} \times f_2(\upsilon_i|\sigma) \; d\upsilon_i \tag{9}$$

A compromise between the models with (pseudo-) likelihood in (8) and (9) is a model with both a frailty

term (υ_{io}) shared by all landmarks as in (9), and landmark specific frailty terms (υ_{is}) that are mutually independent as in (8). The (pseudo-) likelihood for such a model is

$$L_4 = \prod_{i=1}^{n} \int_{\upsilon_{io}} \left\{ \prod_{s=1}^{Q_i} \int_{\upsilon_{is}} \prod_{m=1}^{I_{is}+1} f_1(t_{im}(s)|\mathbf{Z}_i(s), \upsilon_{io}+\upsilon_{is}, \boldsymbol{\theta}(s)) \times f_{2s}(\upsilon_{is}|\sigma_s)\ d\upsilon_{is} \right\} \times f_{2o}(\upsilon_{io}|\sigma_o) d\upsilon_{io} \tag{10}$$

The covariance between the frailty terms $\upsilon_{io}+\upsilon_{is}$ and $\upsilon_{io}+\upsilon_{is'}$ at landmarks s and s' is σ_o^2 and the correlation equals $\frac{\sigma_o^2}{\sqrt{(\sigma_o^2+\sigma_s^2)(\sigma_o^2+\sigma_{s'}^2)}}$.

Furthermore, parameters $\boldsymbol{\theta}(1), \boldsymbol{\theta}(2)..., \boldsymbol{\theta}(Q)$ can be constrained to be landmark invariant, $\boldsymbol{\theta}(1) = \boldsymbol{\theta}(2) = ... = \boldsymbol{\theta}(Q)$, or can be modeled as a smooth function of s which may be a spline or a polynomial function, the form of which can differ between the covariates.

Instead of maximizing (8), (9) and (10), we considered maximizing their corresponding integrated (pseudo-) partial likelihoods. This choice has the advantage that it can be implemented using existing software such as the `coxme` package[?] in the R statistical software.[?] Within `coxme` all parameters and random effects in a frailty model are estimated by maximizing an integrated Cox partial likelihood[????]. At landmark s, given υ_{is}, the contribution of the events of individual i to the Cox partial likelihood of $data_s$ is

$$PL_s^i(\boldsymbol{\theta}(s), \boldsymbol{\upsilon_s}) = \prod_{m=1}^{I_{is}} \left[\frac{\exp\left\{\mathbf{Z}_i^{\intercal}(s)\boldsymbol{\theta}(s) + \upsilon_{is}\right\}}{\sum_{j \in R(t_{im}(s))} \exp\left\{\mathbf{Z}_j^{\intercal}(s)\boldsymbol{\theta}(s) + \upsilon_{js}\right\}} \right] \tag{11}$$

where $\boldsymbol{\upsilon_s}$ is a vector of frailty terms of all patients and $R(t_{im}(s))$ denotes the risk set at $t_{im}(s)$. Integrating over the random effects distribution creates the following integrated partial likelihood (IPL)

$$IPL_s(\boldsymbol{\theta}(s)) = \int \prod_{i=1}^{n_s} PL_s^i(\boldsymbol{\theta}(s), \boldsymbol{\upsilon_s}) f_{2s}^*(\boldsymbol{\upsilon_s}|\boldsymbol{\Sigma_s}) d\upsilon_s \tag{12}$$

where $f_{2s}^*(.)$ is a multivariate normal density function for the υ_s's of all patients at landmark s and $\boldsymbol{\Sigma}_s$ is often a sparse covariance matrix of dimension $n_s \times n_s$. Since the integral in (12) is not tractable, a Laplace approximation is implemented within coxme to obtain parameter estimates. This has been described in more details by Therneau[?] and Ripatti[?]. The major difference between (12) and the corresponding IPL's for (8), (9) and (10) is with the dimension of covariance matrix in $f_{2s}^*(.)$

We used the terms "independent-frailty", "shared-frailty" and "nested-frailty" models to refer to the models that are fitted by maximizing the IPL's for (8), (9) and (10) respectively. Fitting an independent-frailty model within `coxme` requires that we assume that the landmark-independent υ_{is}'s show the same variation over the landmark time points, that is $\sigma_s = \sigma$. The disadvantage of this constrain is that, if the distribution of the frailty random effects would change over landmarks due to drop out of more frail patients, this might introduce some bias in the estimated parameters. Fitting the shared-frailty is more straightforward since it only requires specifying a shared frailty term for each patient over the landmarks. A nested-frailty model is implemented within `coxme` by adding a nested effect which estimates an overall

patient effect over the landmarks as well as a patient specific effect within the separate landmarks. For all three models, we allow for delayed entry at later landmark time points and include landmark as stratum variable in order to estimate separate baseline hazards per landmark.

Dynamic prediction of infection probabilities

We calculated the probability of experiencing at least one infection before t_{hor} for a new subject j given covariates $\mathbf{Z}_j(s)$ at t_s. For any time $u > t_s$ the conditional probability that subject j will be infection-free at least until u is

$$S_{js}(u|\mathbf{Z}_j(s), \mathcal{D}_{s-}) = \int_{\upsilon_j} S_{js}(u|\mathbf{Z}_j(s), \upsilon_j, \hat{\boldsymbol{\theta}}(s), t_s) \times f_2(\upsilon_j|\mathcal{D}_{s-}, \sigma) d\upsilon_j, \tag{13}$$

where $\mathcal{D}_{s-} = \left\{data_1^*, ..., data_{s-1}^*\right\}$ is the infection data for subject j which is available before t_s. Since υ_j is shared by all landmarks in the shared-frailty model, its expected value at t_s is changed by earlier observed infection history $(\mathcal{D}_{s-})$. The same applies to υ_{jo} in the nested-frailty model. $\hat{\boldsymbol{\theta}}(s)$ contains the parameter estimates from the supermodel. At every later landmark time point $t_{s'}$, updated information on $\hat{\boldsymbol{y}}_j(s')$ and $h_j(s')$ is available. $S_{js}(.)$ is dynamically updated to $S_{js'}(.)$ based on new information recorded until $t_{s'}$. $S_{js}(u|\mathbf{Z}_j(s), \upsilon_j, \hat{\boldsymbol{\theta}}(s), t_s)$ is calculated as in (6). Using Bayes' theorem, $f_2(\upsilon_j|\mathcal{D}_{s-}, \sigma)$ in (13) is written as

$$f_2(\upsilon_j|\mathcal{D}_{s-}, \sigma) = \frac{\mathcal{L}_j(\mathcal{D}_{s-}|\upsilon_j, \hat{\boldsymbol{\theta}}(s)) \times f_2(\upsilon_j|\sigma)}{\int_{\upsilon_j} \mathcal{L}_j(\mathcal{D}_{s-}|\upsilon_j, \hat{\boldsymbol{\theta}}(s)) \times f_2(\upsilon_j|\sigma) d\upsilon_j} \tag{14}$$

where $\mathcal{L}_j$ is the likelihood of the events before t_s for subject j. The integral in (14) was approximated using the Gauss-Hermite quadrature (GHQ) technique. For the case of the independent-frailty, no information on infection history $(\mathcal{D}_{s-})$ was required since the frailty terms were landmark specific. Hence, the survival probabilities at each t_{hor} can be computed in the same way as with a classical frailty model.

The predictive performance was evaluated on the basis of how well infected patients were distinguished from those who did not experience an infection within every landmark period. A leave-one-out cross-validation procedure was used to make predictions for a new individual j. So, $\hat{\boldsymbol{\theta}}(s)$ was replaced by $\hat{\boldsymbol{\theta}}_{-j}(s)$ estimated from our supermodel by using data that excludes individual j. We based our evaluations on the area under the ROC curve (AUC) based on the incident/dynamic ROC curve as defined by Heagerty *et al.*[?] .

Simulation Study

Study setup

In order to assess if fitting a supermodel to stacked data with overlapping or non-overlapping landmark periods would yield good parameter estimates, we performed a simulation study by generating data with repeated events. For simplicity and easy comparison with true values, we did not consider markers but

only one time fixed covariate $X \sim N(0, \sigma_x)$, with $\sigma_x^2 = 0.5$. X was assumed to be associated with the infection times via the following frailty model:

$$\lambda_i(t) = \lambda_0(t)\exp\left\{X\beta + \upsilon_i\right\}, \tag{15}$$

where $\beta = 1.6$ and the frailty was distributed as $\upsilon_i \sim N(0, \sigma_\upsilon)$, with $\sigma_\upsilon^2 = 0.1225$. Infection times were generated from (15), in which λ_0 was held constant at 10 to ensure that the number of repeated events within an individual was sufficiently large.

We performed 250 simulations using a sample size of 500. For each simulated data set, first a standard frailty model as in (15) was fitted to evaluate the data generation procedure and then we proceeded with the landmark analysis. Two scenarios were considered to construct the landmark data sets before stacking them.

i The landmark time points were selected such that landmark periods could overlap. This was done by choosing an equidistant grid of $Q = 20$ landmark points in the interval $[s_1 = 0.0$ years, $s_Q = 4$ years$]$. A prediction window of 1 year was used. Here, the stacked data set contained multiple infections from the same individual, which were sometimes common to different landmark periods.

ii Non-overlapping landmark periods were chosen by selecting a sequence of points between $[s_1 = 0.0$ years and $s_Q = 9.5$ years$]$ with half a year in between. A prediction window of 0.5 year was used. As opposed to the overlapping landmark setting, here there was no repetition of the same infections within an individual over the landmark periods.

For both settings, the censoring times were generated from a uniform distribution U[0,10]. We fitted separate models per landmark period, followed by fitting supermodels that assumed a shared frailty term for an individual over the landmark periods, a constant covariate effect, and with and without landmark as stratum variable.

Simulation study results

The results are summarized in tables 1 and 2. We represent the true parameter values as "True" and the square root of the mean squared error as "RMSE". The column "Mean" contains the mean over the 250 simulated data sets. The bias is the difference between the average estimate over all simulations and the true parameter value. As shown in table 2, the estimates of β and σ^2 from analyzing the simulated data sets using a standard frailty model (that is, the original data sets without stacking) were close to their true values. In table 1 we summarize findings from the per landmark analysis. For the separate landmark analysis, the estimated β's were close to the true value. On the other hand the estimated σ^2's showed a slightly larger downward bias compared to the estimate from the standard frailty model, which could be attributed to the smaller sample size of the separate landmark data sets.

For the setting with overlapping landmark periods, results from a supermodel that included landmark as stratum variable showed a very small upward bias for β and a substantial upward bias for σ^2 (table 2). Similar findings were observed from supermodels that assumed a common baseline hazard over landmarks (results not shown). For the setting with non-overlapping landmark periods, since the likelihoods for supermodels with and without landmark as stratum variable were equivalent to those from

analyzing the original data without stacking, we obtained identical estimates. We show the estimates from the supermodel that included landmark as stratum variable in table 2.

Application to the kidney transplantation data

First, for the setting with overlapping landmark periods, landmark time points were defined by choosing an equidistant grid of $Q = 40$ points in the interval $[s_1 = 0.083$ years, $s_Q = 3.50$ years $]$, representing the period between 1 month after transplant to 3.5 years post transplantation. A prediction window w of 1 year was chosen. For the setting with non-overlapping landmark periods, due to limited follow-up time, only $Q = 11$ points were selected in the interval $[s_1 = 0.083$ years and $s_Q = 5.083$ years$]$ with half a year in between. A prediction window of 0.5 year was used. For both settings w was chosen such that there was a sufficient number of repeated events after every landmark, and $h_i(s)$ was defined as the number of infections in the previous landmark period. In figure 2 the proportions of individuals experiencing 0, 1 or more than one UTI's at the different landmarks window periods are shown for the settings with overlapping (left panel) and non-overlapping (right panel) landmark periods.

The effects of all covariates were initially estimated by maximizing the integrated partial likelihood (12) on each landmark data set separately. Next we fitted the independent-frailty, shared-frailty and nested-frailty models by maximizing their various integrated (pseudo-)partial likelihoods. A second-order polynomial of landmark time was used to smooth the effect of the parameters over landmarks.

Results

In the setting with non-overlapping landmark periods, the independent-frailty model gave an Akaike information criterion (AIC) of 3238.14 which was lower than those of the shared-frailty (3316.13) and the nested-frailty (3241.70) models, as well as the sum of the AIC's over the analyses per landmark (3271.13). Figure 3 presents the parameter estimates from the per-landmark analysis, represented by the grey lines and grey confidence bars (non-overlapping landmark periods), and also the variance of the landmark specific random effect. The estimates seemed more or less stable across the landmark points for all covariates. Wider confidence intervals were observed at the later landmark points probably due to the decreasing sample sizes. The variance of the landmark specific random effect decreased over the landmark time points. Figure 3 also shows the smoothed estimates over the landmarks from the supermodels and their 95%-confidence intervals. We show only estimates for the independent-frailty (solid black lines) and shared-frailty (dashed black lines) models. Apart from cd8, it seemed that the quadratic effect of landmark time had no added value in smoothing the effect of parameters over the landmarks. However for both the supermodels, a multivariate Wald test showed that the interactions between the quadratic landmark time and the different covariates were significantly different from zero (p values$<$0.005). We also observed in figure 3 that both supermodels followed the estimates from the per-landmark analysis fairly well (except for the smoothed estimates of infection history from the shared-frailty model), and had much smaller confidence intervals than those from the per-landmark analysis.

Figure 4 presents the results from the setting with overlapping landmark periods. Smaller confidence intervals were obtained from the per-landmark analysis, compared to the setting with non-overlapping landmark periods, probably due to the use of a wider landmark period comprising more data. We also observe in figure 4 that estimates (in absolute values) from the shared-frailty model (dashed black lines) were often larger compared to those of the independent-frailty model (solid black lines), and were

sometimes not in correspondence with the estimates from the separate landmark analysis. Similar findings were observed with the nested-frailty model (results not shown), with slightly bigger differences between the estimates.

We observed that the estimates of infection history from the shared-frailty model were biased downward with respect to the estimates from the per-landmark analysis. This was more extreme for the setting with overlapping landmark periods (which decreased over landmarks) and was probably due to collinearity between event history and the shared frailty term.

Figure 5 shows AUC values calculated at the different t_{hor}'s of the 11 landmark periods based on the independent-frailty model. AUC values were on average around 0.7 over the different landmarks. This suggest that patients who experienced UTI's within the various landmark periods could be well distinguished from those who were infection free. Plots of the distribution of the estimated linear predictors LP ($\mathbf{Z}_i^{\intercal}(s)\boldsymbol{\theta}(s) + \upsilon_{is}$), for patients over the landmarks periods also showed that, on average, patients who were infected at least once (within a landmark period) had higher scores than those who were infection free (see supplementary figures 1and 2).

Discussion

We have extended dynamic prediction by landmarking to a setting with recurrent event data. This was motivated by data comprising post kidney transplantation records of patients with repeated urinary tract infections and repeated measurements of multiple markers. First we considered simple frailty models to estimate infection risk for each landmark separately. Here, time-updated markers were handled as time-fixed covariates, facilitating the prediction of event risk. In scenarios that include only a single longitudinal marker, Rizopoulos *et al.*[?] observed that there was a gain in predictive performance from using a joint modelling approach instead of landmarking, for a setting in which the joint model was correct. However, compared to joint models, landmark models require fewer modelling assumptions, hence are probably more robust. Furthermore, applying the landmarking approach is more straightforward, particularly when the analysis involves several longitudinal markers.

For the sake of parsimony and power, we also fitted supermodels using a stacked data format that merged all landmark data sets. Here, parameters could vary between landmarks in a smooth fashion. For certain choices of the parameterization of the supermodels, implementation is possible within available software. Based on the study data, for the setting with non-overlapping landmark periods, the best supermodel was the independent-frailty model which allowed for landmark specific frailties for each individual.

The landmark models presented in this paper assumed that within the same landmark period, the risk of the first infection is the same as the risk of the second and so on. However, for a new landmark period the risk of the first infection changes due to the updated covariates and new random-effect. Further research is needed for possible model extensions that would allow subsequent infections in the same landmark period to have a different hazard.

In a simulation study, the estimated parameters in a supermodel (with a shared frailty term) were affected by how the landmark periods were selected before stacking. When landmark periods were chosen such that there was overlap of information between the separate landmark data sets (overlapping landmark periods), the parameter estimates were biased upwards, especially the estimated variance of the shared frailty term. This could be attributed to the repeated counting of several events over the overlapping

landmark periods. For instance, a patient with 10 recurrent infections that are counted several times will be way more frail than another patient who was infection-free throughout, hence inflating the variation. This may also explain why with the real data example, for the setting with overlapping landmark periods, the estimates from the shared-frailty and the nest-frailty models were sometimes larger and not in correspondence with the per-landmark analysis. More accurate estimates were obtained in the simulation study when a supermodel was fitted to a stacked data set comprising non-overlapping landmark periods. Thus, as opposed to the setting with a single-type event that occurs only once, where the stacked data sets often comprise overlapping landmark periods, it may be better to use non-overlapping landmark periods for recurrent events data, particularly when fitting supermodels with a shared frailty term.

When using frailty models with left truncated data, it is important to deal with the information of left truncation into the distribution of the random effects[??]. In our data, there was left truncation at each landmark time point since only a subset of patients who were still under follow-up were used. It is possible that only the less frail patients are dynamically retained over the landmark periods, causing the composition of patients with respect to their random effects to differ over landmarks. Failure to correct for this would lead to biased estimates. However, for our data, this was not the case since there were very few deaths and end of follow-up was mainly due to administrative reasons.

All analysis was performed using the `coxme` package in the R statistical software since it was straightforward to fit supermodels with stratification on a large number of landmark time points (up till 40 landmarks for our data). We have also shown in a simulation study that the partial likelihood estimation procedure (via `coxme`) works well. With respect to computation time, apart from the nested-frailty model which ran for a few hours (on a computer with 2 Gb of RAM and a frequency of 3.16 GHz), all the other models took a couple of seconds to a few minutes to run. As a sensitivity check, we compared estimates from `coxme` to those from a full likelihood approach via the `frailtypack` package[??] in the R statistical software, for the per-landmark analysis and the independent-frailty model. A small number of landmarks was used since `frailtypack` allowed for a maximum of 6 strata. The estimates from both software packages were comparable. Other software packages for analyzing recurrent event data have been discussed in the literature[??].

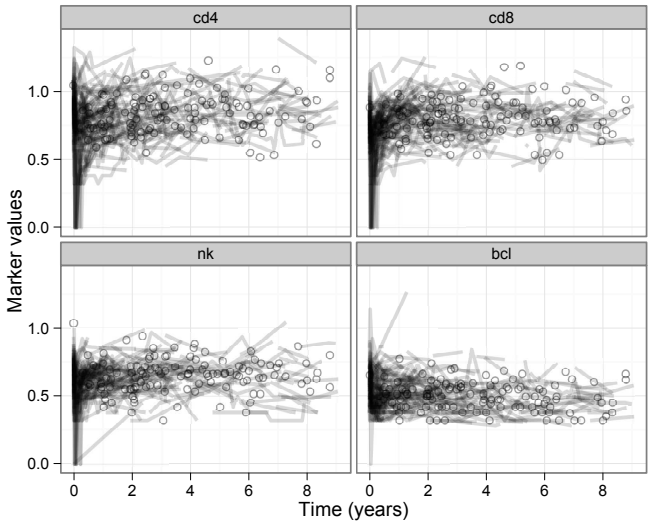


Figure 1. Individual profiles for cd4, cd8, nk and bcl. Circular points represent patients with a single marker measurement

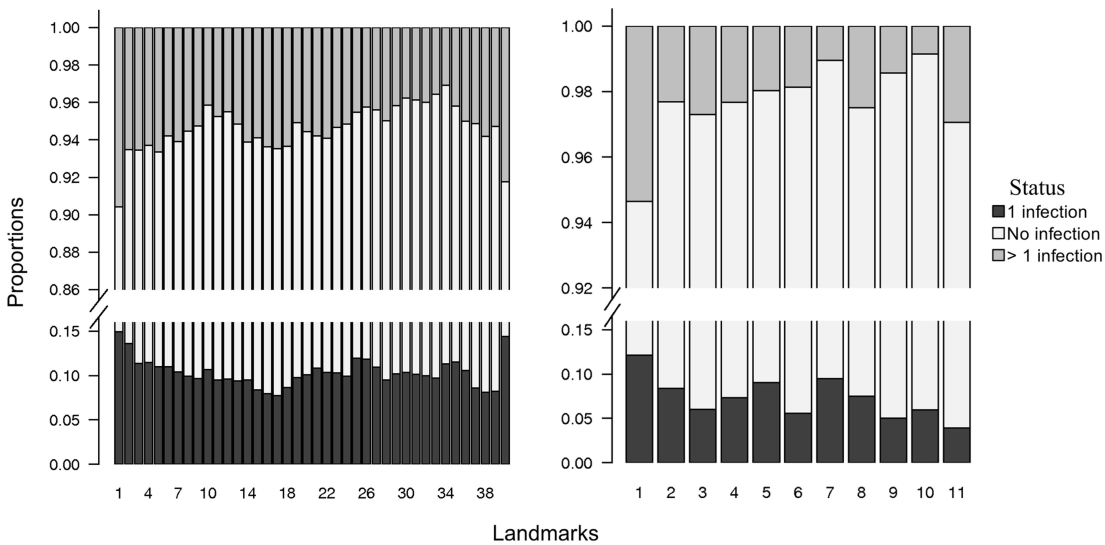


Figure 2. Proportion of individuals with 0, 1 and > 1 UTI's per landmark period for the settings with overlapping (left panel) and non-overlapping (right panel) landmark periods.

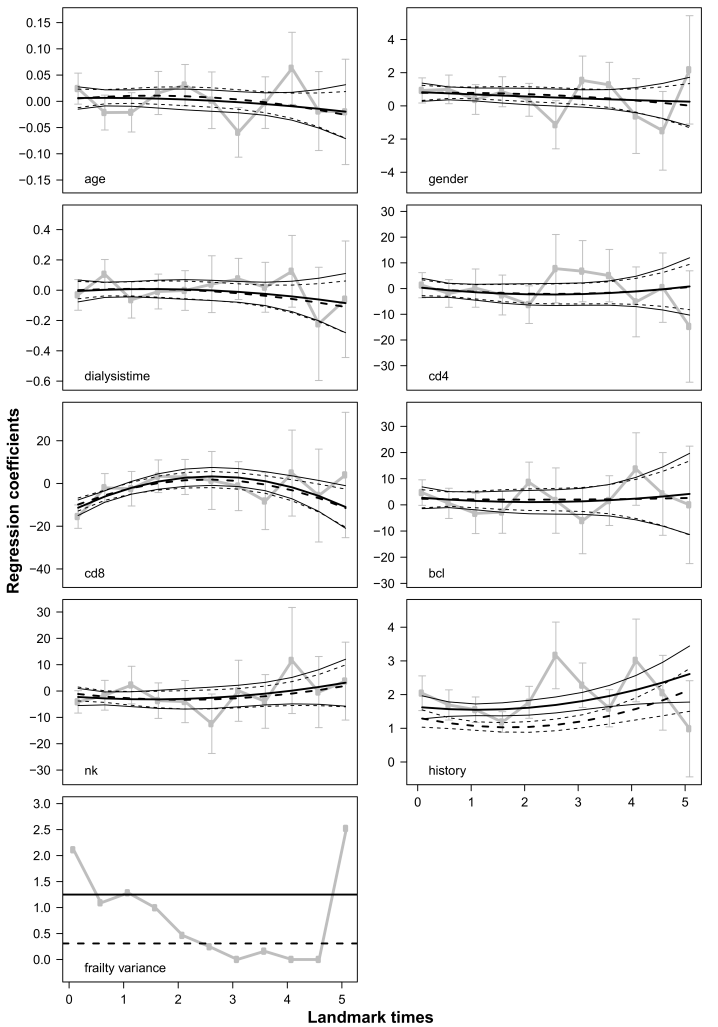


Figure 3. Estimated parameters of covariates from the analysis per landmark (grey), the independent-frailty model (solid black lines) and the shared-frailty model (dashed black lines) for the setting with non-overlapping landmarks.

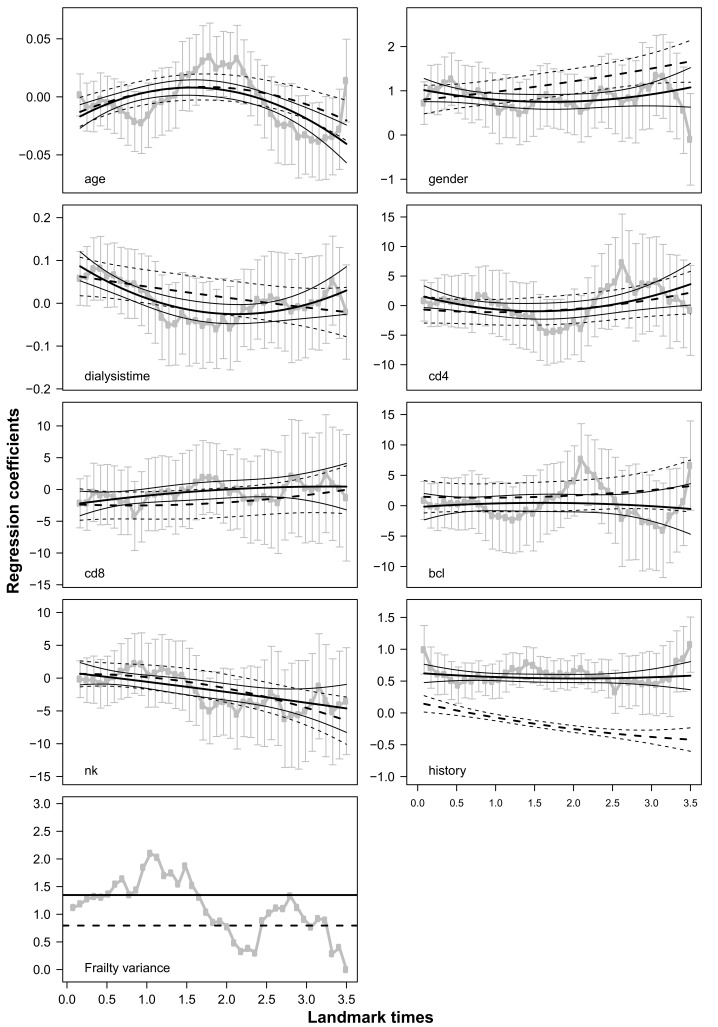


Figure 4. Estimated parameters of covariates from the analysis per landmark (grey), the independent-frailty model (solid black lines) and the shared-frailty model (dashed black lines) for the setting with overlapping landmarks.

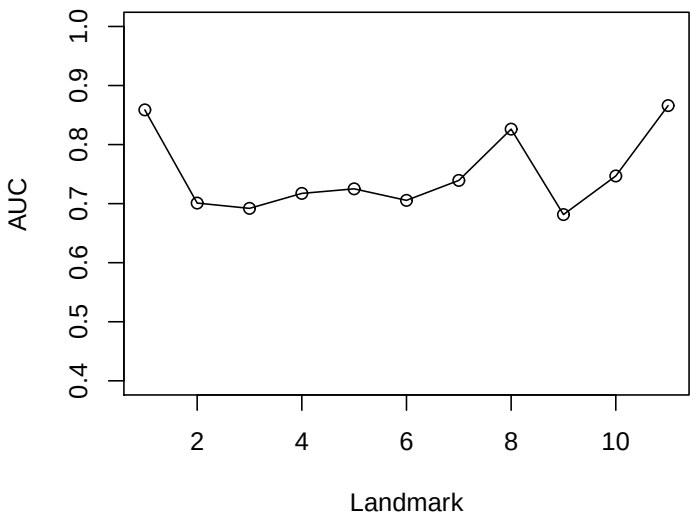


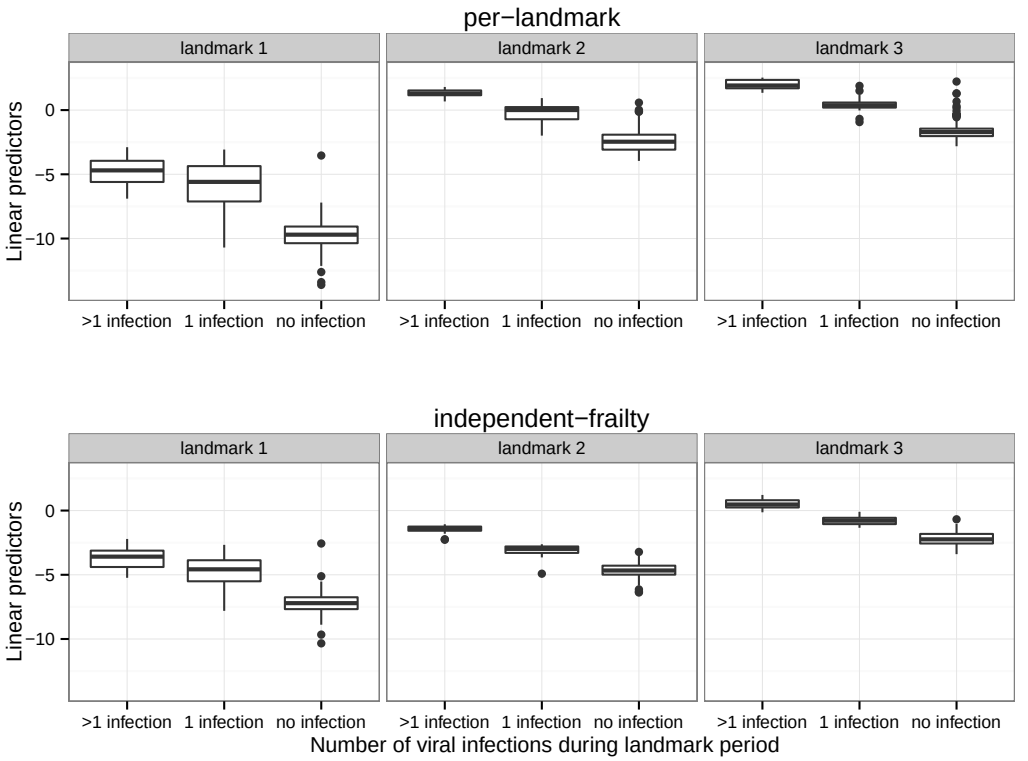
Figure 5. AUC per landmark point based on predicted probabilities using the independent-frailty model in the setting with non-overlapping landmark periods

Table 1. Simulation study results from the analysis per landmark based on 250 data sets (n = 500).

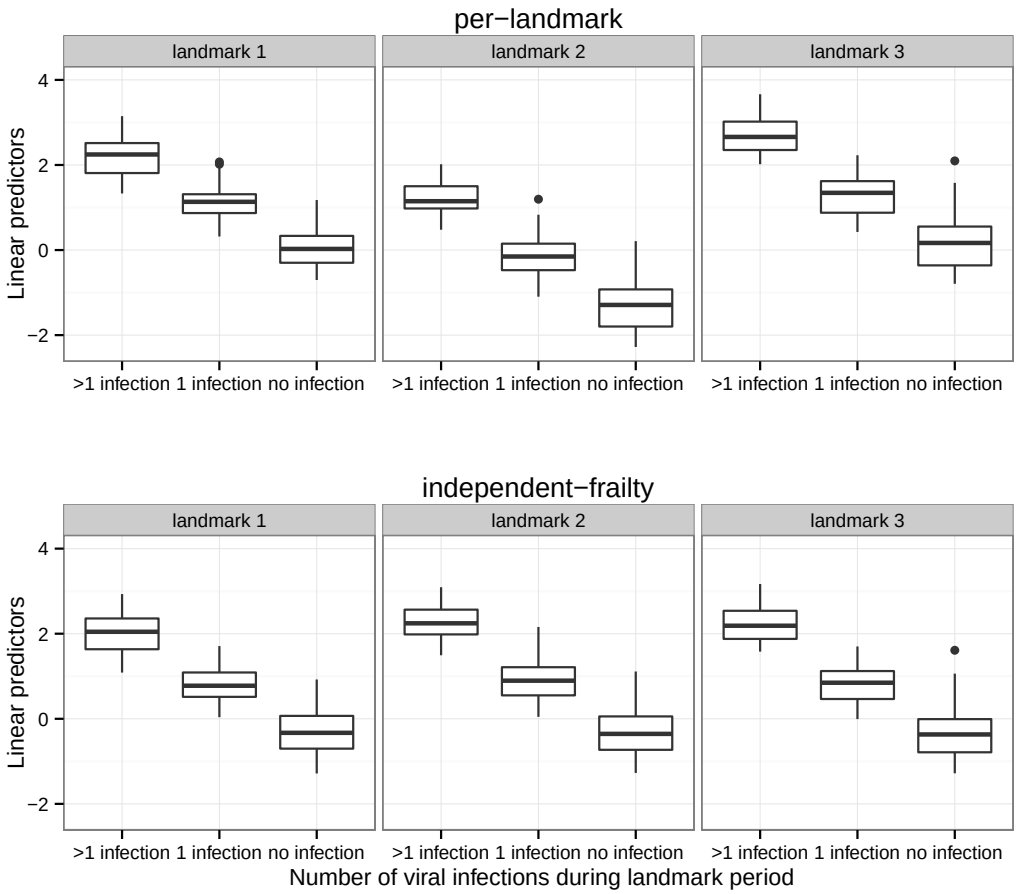
	β (True=1.6)			σ_v^2 (True=0.1225)		
Landmark	Mean	Bias	RMSE	Mean	Bias	RMSE
1	1.5965	-0.0035	0.0437	0.1197	-0.0028	0.0183
2	1.5880	-0.0120	0.0515	0.1206	-0.0019	0.0183
3	1.5880	-0.0120	0.0555	0.1196	-0.0029	0.0182
4	1.5853	-0.0147	0.0604	0.1210	-0.0015	0.0191
5	1.5862	-0.0138	0.0545	0.1203	-0.0022	0.0198
6	1.5891	-0.0109	0.0557	0.1199	-0.0026	0.0218
7	1.5898	-0.0102	0.0595	0.1211	-0.0014	0.0229
8	1.5788	-0.0212	0.0665	0.1214	-0.0011	0.0214
9	1.5887	-0.0113	0.0633	0.1206	-0.0019	0.0233
10	1.5820	-0.0180	0.0736	0.1201	-0.0024	0.0244
11	1.5891	-0.0109	0.0738	0.1216	-0.0009	0.0254
12	1.5942	-0.0058	0.0784	0.1208	-0.0017	0.0271
13	1.5858	-0.0142	0.0924	0.1183	-0.0042	0.0311
14	1.5943	-0.0057	0.0922	0.1173	-0.0052	0.0310
15	1.5872	-0.0128	0.1008	0.1173	-0.0052	0.0340
16	1.5908	-0.0092	0.1120	0.1178	-0.0047	0.0407
17	1.5965	-0.0035	0.1244	0.1121	-0.0104	0.0450
18	1.5927	-0.0073	0.1433	0.1208	-0.0017	0.0512
19	1.6073	0.0073	0.1817	0.1194	-0.0031	0.0645
20	1.6051	0.0051	0.3524	0.1193	-0.0032	0.1370

Table 2. Simulation study results based on 250 data sets (n = 500).

	Parameters	True	Mean	Bias	RMSE
Unstacked data (standard frailty model)					
	β	1.6	1.5905	-0.0095	0.0358
	σ^2	0.1225	0.1223	-0.0002	0.0095
Stacked data, with strata (overlapping intervals)					
	β	1.6	1.6121	0.0121	0.0389
	σ^2	0.1225	0.1588	0.0363	0.0381
Stacked data, with strata (non-overlapping intervals)					
	β	1.6	1.5905	-0.0095	0.0358
	σ^2	0.1225	0.1223	-0.0002	0.0095



Supplementary Figure 1. Distribution of linear predictors from the per-landmark analysis (upper panel) and the independent-frailty model (lower panel) for the setting with non-overlapping landmarks, for landmarks, 1, 2 and 3. This included only patients who were still under follow-up at various landmark periods. ”no infection”, ”1 infection”, and ”$>$ 1 infection” stand for individuals who did not experience any infection, had only 1 infection, and had more than 1 infection at each landmark period.



Supplementary Figure 2. Distribution of linear predictors from the per-landmark analysis (upper panel) and the independent-frailty model (lower panel) for the setting with overlapping landmarks, for landmarks, 1, 2 and 3. This included only patients who were still under follow-up at various landmark periods. ”no infection”, ”1 infection”, and ”$>$ 1 infection” stand for individuals who did not experience any infection, had only 1 infection, and had more than 1 infection at each landmark period.