Multiple imputation approach for interval-censored time to HIV RNA viral rebound within a mixed effects Cox model

Yovaninna Alarcón-Soto[*][,1], **Klaus Langohr** [1], **Csaba Fehér** [2,3], **Felipe García** [2,4], and **Guadalupe Gómez** [1]

[1] Universitat Politècnica de Catalunya/ BARCELONATECH, C/ Jordi Girona, 1–3, 08034 Barcelona, Spain

[2] Hospital Clínic de Barcelona, IDIBAPS, Barcelona, Spain

[3] IRB Barcelona, Barcelona, Spain

[4] University of Barcelona, Barcelona, Spain

We present a method to fit a mixed effects Cox model with interval-censored data. Our proposal is based on a multiple imputation approach that uses the truncated Weibull distribution to replace the interval-censored data by imputed survival times and then uses established mixed effects Cox methods for right-censored data. Interval-censored data were encountered in a database corresponding to a recompilation of retrospective data from eight Analytical Treatment Interruption (ATI) studies in 158 HIV-positive combination antiretroviral treatment (cART)-suppressed individuals. The main variable of interest is the time to viral rebound, which is defined as the increase of serum viral load to detectable levels in a patient with previously undetectable viral load, as a consequence of the interruption of cART. Another aspect of interest of the analysis is to consider the fact that the data come from different studies based on different grounds and that we have several assessments on the same patient. In order to handle this extra variability, we frame the problem into a mixed effects Cox model that considers a random intercept per subject as well as correlated random intercept and slope for pre-cART viral load per study. Our procedure has been implemented in R using two packages: `truncdist` and `coxme`, and can be applied to any data set that presents both interval-censored survival times and a grouped data structure that could be treated as a random effect in a regression model. The properties of the parameter estimators obtained with our proposed method are addressed through a simulation study.

Key words: HIV studies; Interval censoring; Mixed effects Cox model; Multiple imputation; Survival analysis

1 Introduction

Most statistical methods developed for the analysis of survival data assume that the event that defines time origin is known and allow the event of interest ε that determines failure and, hence, the survival time, to be right-censored. In many situations, however, the event of interest ε cannot be observed and it is only known to have occurred within two random times, say L and R. In this set-up, we say that the time to ε, T, is interval-censored.

Interval-censored data often arises in medical or health studies that entail periodic follow-ups, and many clinical trials and longitudinal studies fall into this category (Sun, 2007). In such situations, interval-censored data may arise in several ways. For instance, an HIV-infected patient is examined weekly to check if his/her viral load exceeds a certain threshold. Suppose that in a first measurement, at time t_1, it does not exceed the threshold and it does in a second measurement at time t_2. Hence, all that is known is that the viral load exceeded the threshold within the interval $(t_1, t_2]$, but the exact time of viral rebound is unknown.

---

*Corresponding author: e-mail: yovaninna.alarcon@upc.edu, Phone: +34-934-016-947, Fax: +34-934-015-855

We have encountered interval-censored data while studying the immunological response of HIV positive patients by means of different parameters of viral rebound dynamics within eight different Analytical Treatment Interruption (ATI) studies (Leal et al., 2017).

We start introducing the main concepts related to HIV-ATI studies to facilitate reading and understanding of the study that motivates this paper. The Human Immunodeficiency Virus (HIV) is a virus that infects cells of the immune system, destroying or impairing their function. Infection with the virus results in a progressive deterioration of the immune system, leading to immune deficiency, and the immune system is considered deficient if it is no longer able to fulfil its role of fighting infection and disease (World Health Organization, 2017). HIV-infected patients are generally treated with a combination of antiretroviral drugs known as combination Antiretroviral Therapy (cART) in order to maximally suppress the HIV virus and stop the progression to the Acquired Immune Deficiency Syndrome (AIDS).

The Analytical Treatment Interruption (ATI) is a controlled interruption of the cART in HIV-positive patients and appears in different interventional or observational studies in this field. The objective of this interruption is the evaluation of the immunological response of the patients, described by different parameters related to the viral rebound dynamics (Treasure et al., 2016). In the case of our data set, depending on the study, a single ATI episode or several episodes are studied. Studies including more than one ATI episode are based on the 'autovaccination' theory, according to which repeated encounters of the immune system with the antigenic stimulus (the virus) will be able to increase the specific immune response, leading to a major control of the posterior viral load (Graziani & Angel, 2015). The endpoint of interest in these studies is the time until viral rebound, which is defined as the first time that an HIV-infected patient, with previously undetectable serum viral load, surpasses the threshold of 20 copies/mL (or 1.30 in $\log_{10}$ scale). Viral load values lower than 20 copies/mL were considered undetectable when calculating time to rebound.

Previous to the start of cART, most of the patients in our data set had a high viral load, referred as pre-cART viral load. As soon as they start cART, the viral load drops down to undetectable levels. In our case, at the beginning of the first ATI episode (week 0), all patients presented undetectable levels of viral load. From that moment, the viral load was measured once every week and the corresponding ATI episode was stopped as soon as the viral load was detectable again; see Figure 1.

To study viral rebound dynamics, we were interested in characterizing the percentiles of the time to viral load rebound based on the corresponding distribution function, in establishing differences between male and female patients, in determining the importance of pre-cART VL in suffering a viral rebound, and in assessing relative risks by means of hazard ratios based on a Cox proportional hazards model (Cox, 1972). To provide rigorous answers to all these issues, several nonparametric and parametric methods for interval-censored data have been developed and the literature is already abundant, see Gómez et al. (2009) for a thorough overview. Finkelstein (1986) was among the first authors to adapt the Cox model to interval-censored data. Nowadays, parameter estimation in the presence of interval-censored data can be carried out in R (R Core Team, 2018) by means of the `icenReg` package (Anderson-Bergman, 2017).

However, the ATI data set not only has interval-censored times to viral rebound but these are also quite heterogeneous because they correspond to eight different studies that are based on different grounds. Moreover, several patients underwent more than one treatment interruption. A proper data analysis should take this into account. Since the variability of measurements in different individuals is usually larger than the variability between measurements in the same individual (Bland & Altman, 1994), a possible way to account for this extra variability could be to frame the problem into a mixed effects Cox model where a grouping variable corresponding to the subject would be added as random effect.

In the case of data coming from different studies, we might have two different scenarios. If there is homogeneity of the observed effect for some particular covariate, a single measure would be adequate to describe the general results. If, on the contrary, heterogeneity of effects is found, we should add a random effect of this covariate per study to capture this heterogeneity and carefully interpret the results. Our ATI data set falls in the second situation because of the different treatments used (combination of different drugs or different vaccines) and the different recruitment criteria for patients (such as early or late stage of
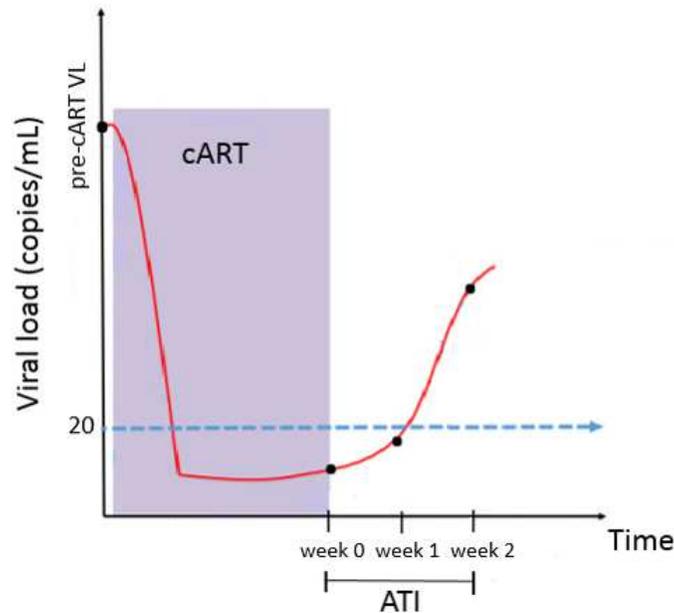
Figure 1 Viral load dynamics during (cART) and the first ATI episode. The last measure before starting the cART is denoted by pre-cART viral load (VL). In this case, the ATI stopped at week 2 because of detection of viral rebound.

HIV infection, CD4 count, viral load threshold, among others). Therefore, it is very likely to have some degree of variation (heterogeneity) among these studies, and for this reason, a random intercept and slope for pre-cART VL per study is considered.

Some other reasons to consider the inclusion of random effects are similar to those considered by Yamaguchi et al. (2002) and Senn (1998), who in the analyses of multicenter trials used random effects to model the center's effect variability. As Senn discusses, if we are interested in making inferences about patients from a given study, the fixed effect approach leaves little alternative but to use the results from that study only. The random effects approach will allow us to combine information with the given study with information from all the studies in a way which is more appealing and useful.

In the context of survival analysis, a mixed (fixed and random) effects Cox model with right-censored data has been presented by Therneau & Grambsch (2000) and its fit is accomplished with the `coxme` package of the same author (Therneau, 2018b). However, to the best of our knowledge, a mixed effects Cox model with interval-censored data has not been studied. Hence, our objective consists of developing a mixed effects Cox model with interval-censored data in order to correctly model the heterogeneity in our data set attributable to the repeated measures per patient and the different inclusion criteria per study.

A natural solution to the difficulties of direct estimation based on interval-censored data is to use an algorithm based on treating the interval-censored observations as missing data and imputing values for them, thus creating right-censored and exact data (Bebchuk & Betensky, 2000). Interval-censored data are actually incomplete data, not missing data, because the observed interval provides some information about the variable of interest (Sun, 2007). Nevertheless, we can still treat the underlying, unobserved true interval-censored failure times as missing and replace them by imputed times conditional on the observed information. Using the methodology and software already developed for right-censored data (Therneau, 2018b), our proposal is based on a multiple imputation approach using the truncated Weibull distribution to replace the censoring intervals by imputed survival times. Our idea is similar to the one of Satten et al.

(1998) in the case of the Cox model without random effects, who used imputation methods to replace the interval-censored survival times by imputed values. The authors propose the use of a parametric model for the baseline hazard in order to generate imputed failure times; following, a rank-based procedure based on the imputed failure times is used to estimate the regression coefficients.

Multiple imputation is a statistical technique to handle missing data that takes advantage of the flexibility in modern computing. With it, each missing value is replaced by two or more imputed values in order to represent the uncertainty about which value to impute (Rubin, 2004). According to the method for 'repeated imputation' inference, each of the simulated complete data sets is analysed by standard methods, and the results are combined to produce estimates and confidence intervals that incorporate missing-data uncertainty. Multiple imputation methods for related censoring problems regarding HIV data have been developed by Muñoz et al. (1989) and Taylor et al. (1990). Dorey et al. (1993) applied multiple imputation to interval-censored data corresponding to threshold-crossing time in some trials. Threshold-crossing times, somehow similar to viral load rebound, are common in medicine when patients move to a new risk category after crossing a threshold on some prognostic variable and because patient's examinations occur only periodically, the exact time of crossing the threshold is only known to fall within a specified interval.

The article is organized as follows: in Section 2, we present the relevant notation and preliminaries used throughout the paper. Following, in Section 3, we present our multiple imputation-based approach to fit a mixed effects Cox model in the presence of interval-censored data. In Section 4, the methodology is applied to the ATI data set, followed by the section dedicated to the study of the properties of the fixed parameter estimators under different settings via a simulation study (Section 5). Finally, the main findings of this work are summarized in Section 6 and, so far, unresolved topics are discussed. Information on the implementation of the parameter estimation in R is presented in Appendix A and more details on the main features of each study in ATI data set are given in Appendix B.

2 Notation and preliminaries

Let T be the time until the event of interest, ε, which in the ATI studies corresponds to viral rebound. T is a non-negative random variable whose distribution function at time t, $F(t) = \mathrm{P}(T \leq t)$, corresponds to the cumulative probability of reaching viral rebound before time t. The hazard function defined by $\lambda(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathrm{P}(t \leq T < t + \Delta t | T \geq t)$, represents the instantaneous risk of viral rebounding and it is the function on which the model we propose will be based on.

The observable data, based on a sample of n patients, consists, for the ith individual ($i = 1, \ldots, n$), of the random intervals $(L_i, R_i]$ during which the viral rebound occurred and the vector of covariates $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{ir})'$, including the study label, gender, and pre-cART viral load. Interval-censored data include right-censored times as a particular case with $R_i = \infty$, which in our data base corresponds to patients whose viral load has not rebounded by the end of the study.

To obtain a non-parametric estimation of the survival function, $S(t) = 1 - F(t)$, under interval censoring, one of the most popular methods is Turnbull's estimator (Turnbull, 1976). For this purpose, we define the so-called Turnbull intervals, denoted by $\mathcal{I} = \{(q_1, p_1], (q_2, p_2], \ldots, (q_m, p_m]\}$, where $\mathcal{I}$ are those intervals where all the mass of any non-parametric maximum likelihood estimator (NPMLE) will be concentrated. To obtain these intervals, let $\mathcal{L} = \{L_i, 1 \leq i \leq n\}$ and $\mathcal{R} = \{R_i, 1 \leq i \leq n\}$ be the set of left and right endpoints respectively. We need to derive all the distinct intervals $(q_j, p_j]$ such that $q_j \in \mathcal{L}$, $p_j \in \mathcal{R}$, and that there is no other left or right endpoint between q_j and p_j. The NPMLE for the survival function decreases inside the set $\mathcal{I}$ and is constant outside of them. Specifically, denoting by $w_j = \mathrm{P}(q_j < T \leq p_j) = S(q_j) - S(p_j)$ the weight of the jth Turnbull's interval, the NPMLE for $S(t)$ is given by

$$\hat{S}_n(t) = \begin{cases} 1, & \text{if} \quad t \le q_1 \\ 1 - (\hat{w}_1 + \cdots + \hat{w}_j), & \text{if} \quad p_j \le t \le q_{j+1}, \quad 1 \le j \le m-1 \\ 0, & \text{if} \quad t \ge p_m \end{cases} \tag{1}$$

and is not specified within $(q_j, p_j]$, for $1 \le j \le m$. Observe that Turnbull's estimate of the survival function has a special shape with horizontal stretches in non-Turnbull's intervals and rectangular boxes indicating areas of equal likelihood in Turnbull's intervals.

Let $\lambda(t|\boldsymbol{x})$ be the hazard function at time t for an individual with risk vector $\boldsymbol{x}$. The basic model (Cox, 1972) is given by

$$\lambda(t|\boldsymbol{x}) = \lambda_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x}) = \lambda_0(t)\exp\Big(\sum_{k=1}^{s} \beta_k x_k\Big) \tag{2}$$

where $\lambda_0(t)$ is an arbitrary baseline hazard function and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_s)'$ is a parameter vector. Once the model (2) has been established and assuming that the primary interest lies in the role played by the fixed effects, the estimate of $\boldsymbol{\beta}$ must be addressed allowing that $\lambda_0(t)$ is arbitrary. Given that the baseline function $\lambda_0(t)$ is not specified, to study the influence of the covariates in the survival times, a modification of the classical theory of maximum likelihood is needed. With this goal Cox proposed the partial likelihood function,

$$L_P(\boldsymbol{\beta}) = \prod_{j=1}^{k} \frac{\exp\{\boldsymbol{\beta}'\boldsymbol{x}_{(j)}\}}{\sum_{l \in R(t_{(j)})} \exp\{\boldsymbol{\beta}'\boldsymbol{x}_{(j)}\}}, \tag{3}$$

where $R(t_{(j)})$ is the risk set at time $t_{(j)}$ and $t_{(1)} < t_{(2)} < \cdots < t_{(k)}$ are the k distinct failure times. $L_P(\boldsymbol{\beta})$ is based on that part of the data that does not carry information about $\lambda_0(t)$. The partial likelihood function is especially useful when it is much more simple than the complete likelihood function and it is a good remedy when the general method of maximum likelihood is not adequate due to the presence of many nuisance parameters (Gómez & Cadarso-Suárez, 2017).

The Cox proportional hazards model, as described in (2), can be enhanced through the incorporation of random effect terms to account for within-cluster homogeneity in outcomes. This model, called mixed effects Cox model, was developed by Therneau & Grambsch (2000) and implemented for right-censored data. The hazard function of a mixed effects Cox model for an individual i is given by

$$\lambda_i(t; \mathbf{X}, \mathbf{Z}) = \lambda_0(t)\exp(\mathbf{X_i}\boldsymbol{\beta} + \mathbf{Z_i}\mathbf{b}) \quad \text{with } \mathbf{b} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta})), \tag{4}$$

where $\mathbf{X_i}$ and $\mathbf{Z_i}$ are the ith rows of the design matrices corresponding to the fixed and random effects, and $\boldsymbol{\beta}$ and $\mathbf{b}$ are the vectors of the fixed and random effects coefficients, respectively. In addition, as above, λ_0 is the unspecified baseline hazard function and the distribution of the random effects is assumed to follow a Gaussian distribution with mean zero and covariance matrix $\sum$, which in turn depends on a vector of parameters $\boldsymbol{\theta}$.

The partial likelihood function corresponding to the mixed effects Cox model is similar to that of the partial likelihood of the standard Cox model given in (3) and the expression of its logarithm for any fixed values of $\boldsymbol{\beta}$ and $\mathbf{b}$, is

$$\text{LPL}(\boldsymbol{\beta}, \boldsymbol{b}) = \log\{\text{PL}(\boldsymbol{\beta}, \mathbf{b})\} = \sum_{i=1}^{n} \int_{0}^{\infty} \Big[\mathbf{Y_i(t)}\eta_i - \log\big\{\sum_j \mathbf{Y_j(t)}\exp(\eta_j)\big\}\Big] \tag{5}$$

where $\eta_i = \mathbf{X_i}\boldsymbol{\beta} + \mathbf{Z_i}\mathbf{b}$ is the linear score for subject i and $Y_i(t)$ is an indicator variable that takes value 1, if individual i is at risk at time t and 0 otherwise; for further details, see Therneau (2018b).

A straightforward adaptation of the partial likelihood function (5) to allow for the presence of interval-censored data is not possible, because with such data, it is not feasible to identify the exact ranking of the failure times and, consequently, the indicator variables $Y_i(t)$ cannot be determined for all t. Our proposal to overcome that problem is to use multiple imputation to replace the interval-censored survival times by exact and right-censored imputed values. The details are presented in the next section.

3   Parameter estimation in the mixed effects Cox model

The basic idea of our proposal consists of replacing the censoring intervals $(L, R]$ that contain the unknown survival times by imputed values based on a truncated Weibull distribution. The mixed effects Cox model in (4) can then be fitted to the imputed exact and right-censored data. These steps will be repeated several times in order to account for the uncertainty of the imputation step, which is ignored in the case of single imputation. Our proposal can be summarized by the following three steps.

Step 1: Imputation of interval-censored survival times.

Censoring intervals are replaced by imputed times following two steps. First, an accelerated failure time model is fitted to the whole data set, considering the covariates of interest, and the maximum likelihood estimators of the corresponding parameters are derived. Second, for each individual's censoring interval, its corresponding truncated Weibull distribution is obtained, and a random value is generated from that distribution. In the case of right-censored observations, no imputation is performed. Hence, the resulting data set consists of uncensored (imputed) and right-censored survival times.

Step 2: Fit of the mixed effects Cox model and analysis.

The mixed effects Cox model (4) described in the previous section is fitted with the data resulting from Step 1 and the parameter estimates of interest are obtained.

Step 3: Pooling the results.

Given a pre-specified integer M, Steps 1 and 2 are repeated M times. Following, the parameters of interest are estimated as the average values of the estimates obtained in each of the M repetitions of Step 2.

Below we explain in more detail each of these steps.

3.1   Imputation of interval-censored survival times

The first step of our proposal consists of replacing the censoring intervals $(L, R]$ by imputed values based on a truncated Weibull distribution, where the truncation is induced by the respective intervals of the individuals. The motivation to use the Weibull distribution is twofold: on one hand, it is a common parametric model for survival data due the simplicity of the survival function and the flexibility of the hazard function, which is either constant, monotonically increasing or monotonically decreasing. On the other hand, it shares the assumption of proportional hazards with the Cox model. For that reason, given the Cox model in (2) and assuming that the baseline hazard function follows a Weibull distribution with cumulative distribution function $G(t) = 1 - \exp(-\lambda t^\alpha)$, survival times can be generated using the following expression (Bender et al., 2005):

$$T = \left(- \frac{\log(U)}{\lambda \exp(\boldsymbol{\beta}' \boldsymbol{x})} \right)^{1/\alpha}, \tag{6}$$

where U follows a uniform distribution on the interval from 0 to 1.

An estimation of $\boldsymbol{\beta}$, λ, and α in (6) are obtained after fitting the following accelerated failure time model –equivalent to the Cox model under the Weibull assumption–:

$$Y = \log T = \mu + \boldsymbol{\gamma}' \boldsymbol{X} + \sigma W,$$

where W follows the extreme value distribution. Standard software is used to obtain maximum likelihood estimators of μ, $\boldsymbol{\gamma}$ and σ and, from these, maximum likelihood estimators of the parameters in (6) are obtained by means of the following transformations:

$$\hat{\lambda} = \exp(-\hat{\mu}/\hat{\sigma}), \quad \hat{\alpha} = 1/\hat{\sigma}, \quad \hat{\boldsymbol{\beta}} = -\hat{\boldsymbol{\gamma}}/\hat{\boldsymbol{\sigma}}.$$

For each individual's censoring interval an uncensored imputed survival time is randomly generated using its corresponding truncated Weibull distribution with parameters $\hat{\lambda}$, $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$. No imputation is needed in the case of right-censored observations.

Regarding truncated distributions and according to Nadarajah & Kotz (2006), let X be a random variable representing the truncated version of some distribution function $G(\cdot)$, over the interval $(L_i, R_i]$ of each individual with $0 < L_i < R_i < \infty$. It is straightforward to check that the distribution function of X is given by

$$F_{X_i}(x) = \frac{G(\max(\min(x, R_i), L_i)|\hat{\boldsymbol{\vartheta}}) - G(L_i|\hat{\boldsymbol{\vartheta}})}{G(R_i|\hat{\boldsymbol{\vartheta}}) - G(L_i|\hat{\boldsymbol{\vartheta}})},$$

and its corresponding inverse by

$$F_{X_i}^{-1}(p) = G^{-1}(G(L_i|\hat{\boldsymbol{\vartheta}}) + p(G(R_i|\hat{\boldsymbol{\vartheta}}) - G(L_i|\hat{\boldsymbol{\vartheta}}))).$$

Hence, to obtain a random value of X, for every censoring interval $(L_i, R_i]$, a random uniform number u_i is generated and an imputed value $x_i = F_X^{-1}(u_i)$ is derived.

Given an original sample of n individuals with $n - r$ interval-censored and r right-censored survival times, step 1 yields a sample of size n with $n - r$ imputed uncensored and the r right-censored survival times. In the next step, a mixed effects Cox model can be fitted to this sample as we explain below.

3.2 Fit of the mixed effects Cox model

In this step, the mixed effects Cox model (4) explained in Section 2 is fitted using the previously imputed values. The objective here is to estimate the vector of fixed effects regression parameters $\boldsymbol{\beta}$ and the vector of parameters $\boldsymbol{\theta}$ for the covariance matrix $\boldsymbol{\Sigma}$ of the random effects. In what follows, we sketch the main ideas of Therneau's method (Therneau, 2018a).

The MLE for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ is based on an integrated penalized partial likelihood (IPL)

$$\text{IPL}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{(2\pi)^{q/2}|\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{1/2}} \int \text{PL}(\boldsymbol{\beta}, \boldsymbol{b}) \exp\{-\boldsymbol{b}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{b}/2\} \, db \qquad (7)$$

where $\boldsymbol{b}$ is the vector of random effects coefficients, as presented in (4) and q corresponds to the number of random effects. When the variance of the random effect is zero, this collapses to the ordinary Cox partial likelihood.

Since expression (7) is not a tractable integral and in order to perform computations under this likelihood, we rewrite the logarithm of the integrand of equation (7), that is, $\text{LPPL}(\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\theta}) = \text{LPL}(\boldsymbol{\beta}, \boldsymbol{b}) - (1/2)\boldsymbol{b}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{b}$, as a second-order Taylor series about $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}})$ as follows

$$\text{LPPL}(\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\theta}) \approx \text{LPPL}(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \hat{\boldsymbol{b}}(\boldsymbol{\theta})) - (1/2)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}, \boldsymbol{b} - \hat{\boldsymbol{b}}(\boldsymbol{\theta}))' H(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}, \boldsymbol{b} - \hat{\boldsymbol{b}}(\boldsymbol{\theta}))$$

where the Hessian H is evaluated at $(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \hat{\boldsymbol{b}}(\boldsymbol{\theta}))$. When $\boldsymbol{\theta}$ and hence $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ are fixed, the relevant values of $\boldsymbol{\beta}$ and $\boldsymbol{b}$ that maximize $\mathrm{LPPL}(\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\theta})$ are easily obtained using essentially the same methods as an ordinary Cox model.

As we are only interested in the values at $\hat{\boldsymbol{\beta}}$, the last term collapses to $(\mathbf{0}, \boldsymbol{b} - \hat{\boldsymbol{b}})' H (\mathbf{0}, \boldsymbol{b} - \hat{\boldsymbol{b}}) = (\boldsymbol{b} - \hat{\boldsymbol{b}})' H_{bb} (\boldsymbol{b} - \hat{\boldsymbol{b}})$, where H_{bb} is the portion of the Hessian corresponding to the random effects. When we replace the body of the integral in (7) with this approximation, then the result is an integral that can be solved in closed form. For further details see Therneau (2018a).

Basically, in this second step by fitting the mixed effects Cox model presented in (4), we obtain an estimation of the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{b}$ as well as of the corresponding covariance matrices.

3.3 Pooling the results

We need to repeat the previous steps (Steps 1 and 2) M times, where M is a pre-specified integer. The M complete-data analyses corresponding to the M imputations under the mixed effects Cox model result in M repeated complete-data statistics, and these are combined to form one repeated-imputation inference that appropriately adjusts for interval-censored data under the model used to create the repeated imputations.

The estimated parameter vectors and their corresponding covariance matrices obtained in the M repetitions of steps 1 and 2 are denoted by $\hat{\boldsymbol{\beta}}_m, \hat{\boldsymbol{b}}_m, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}_m}, \hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta})_m$, $m = 1, \ldots, M$.

Given a particular fixed effects parameter β_i, the repeated-imputation estimate $\hat{\beta}_{i,\mathrm{MI}}$ is the average over the M estimates of this parameter, that is, $\hat{\beta}_{i,\mathrm{MI}} = \left(\sum_{m=1}^{M} \hat{\beta}_{i,m} \right)/M$. In addition, multiple imputation also provides a simple formula to estimate the variance of $\hat{\beta}_{i,\mathrm{MI}}$ (Rubin, 2004), namely,

$$\widehat{\mathrm{Var}}(\hat{\beta}_{i,\mathrm{MI}}) = U_{i,\mathrm{MI}} + \left(1 + \frac{1}{M}\right) B_{i,\mathrm{MI}},$$

where $U_{i,\mathrm{MI}} = \left(\sum_{m=1}^{M} \widehat{\mathrm{Var}}(\hat{\beta}_{i,m}) \right)/M$ is the within-imputation variance and B_{MI} is the between-imputation variance given by $B_{i,\mathrm{MI}} = \left(\sum_{m=1}^{M} (\hat{\beta}_{i,m} - \hat{\beta}_{i,\mathrm{MI}})^2 \right)/(M - 1)$. The B_{MI} term is inflated by a factor $1/M$ to take into account the finite number of imputations.

The same procedure is applied for the estimation of the random effects as well as for the elements of the covariance matrices.

3.4 Software issues

We have accomplished the imputation process in `R` using the `truncdist` contributed package (Novomestky & Nadarajah, 2016). This package includes the function `rtrunc` to impute the values per subjects. This function generates n random deviates that are drawn from the specified truncated distribution.

To fit the mixed effects Cox model, we used the `R` package called `coxme` (Therneau, 2018b). The central computational strategy implemented in this package is an outer and an inner loop. The outer loop searches over the parameters $\boldsymbol{\theta}$ of the variance matrix for a maximum of the IPL (7) and does it in 3 steps. For each trial value of $\boldsymbol{\theta}$ in this search, the first step is to calculate $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})$. The second step is to solve the penalized Cox model $\mathrm{LPL}(\boldsymbol{\beta}, \boldsymbol{b}) - (1/2)\boldsymbol{b}'\boldsymbol{\Sigma}^{-1}\boldsymbol{b}$ to get the solution vector $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b}})$. The iterative Newton-Raphson solution to this problem is the inner loop. The third step is to use the Laplace approximation to compute the $\log$ IPL, using the results of step 2. The implemented `R` code to follow the algorithm is presented in Appendix A.

4 Effect of gender and pre-cART VL on the time to HIV RNA viral rebound considering multiple random effects

The ATI data set (Leal et al., 2017) that motivated this study corresponds to a recompilation from 229 ATI episodes belonging to 158 different patients. The main virologic outcome of interest in these studies is

the time to viral rebound, defined as the time between treatment interruption and the first detectable serum viral load (VL), which is measured as the number of HIV copies in a millilitre of blood. For the purpose of statistical analyses, it is commonly log-transformed because of its right-skewed distribution.

Following, we present a description of some of the variables in the data set. Therein, we refer to the eight studies by Study 1 to Study 8. More detailed information on the studies, including the inclusion criteria and the interventions, is provided in Appendix B. In addition, we will refer to the last viral load before the first initiation of cART by pre-cART VL.

4.1 Descriptive analysis of the ATI dataset

Table 1 presents the gender distribution and a numeric description of the log pre-cART VL separately for each of the eight studies as well as overall. Therein, N_{ATI} denotes the number of ATI periods per study and N the number of patients. As shown, 158 patients were involved in the 8 studies with a total of 229 ATI periods. Fifty nine patients were exposed to an immunomodulating intervention of some kind and 61% of the patients were men. Notice that in the case of both Study 6 and 7, we did not obtain the information on the gender of 17 patients. Concerning the pre-cART VL of the patients, the overall median of the log base 10-transformed pre-cART VL was 4.37 with similar values in all but one study (Study 4), within which the median log pre-cART VL was clearly lower (3.19). The reason for this resides in the inclusion criteria for Study 4 as explained in Appendix B.

Table 1 Description of the eight studies in ATI data set.

Study	n_{ATI}	n	Gender Male (Fem)	Missing	Log pre-cART VL Mean (SD)	Median (IQR)	Min Max	Missing
Overall	229	158[a]	96 (28)	34	4.37 (0.7)	4.31 (3.98 - 4.86)	2.33 6.00	3
Study 1 (García et al., 2005)	32	16	14 (2)	0	4.14 (0.64)	4.16 (3.88 - 4.57)	3.10 5.17	0
Study 2 (García et al., 2013)	70	35	27 (8)	0	4.78 (0.56)	4.69 (4.38 - 5.16)	3.20 5.74	0
Study 3 (ClinicalTrials.gov Identifier: NCT 02767193)	18	18	18 (0)	0	4.12 (0.57)	4.19 (3.95 - 4.48)	2.33 4.76	0
Study 4 (García et al., 2004)	11	11	5 (6)	0	3.21 (0.37)	3.19 (3.08 - 3.40)	2.61 3.80	0
Study 5 (García et al., 2003)	20	20	15 (5)	0	4.50 (0.49)	4.48 (4.11 - 4.89)	3.80 5.40	0
Study 6 (Mothe et al., 2015)	28	28	10 (1)	17	4.46 (0.77)	4.28 (3.96 - 4.87)	3.33 6.00	3
Study 7 (Fagard et al., 2003)	33	33	11 (5)	17	4.43 (0.52)	4.45 (4.08 - 4.70)	2.84 6.00	0
Study 8 (García et al., 1999, 2001)	17	10	7 (3)	0	4.59 (0.56)	4.40 (4.31 - 4.89)	3.89 5.70	0

[a]The sum of the column is not equal to 158 because of patients belonging to more than one study.

Concerning the time until viral rebound, all but one of the 158 patients suffered a viral rebound during the respective follow-up times of the ATI studies and since VL was determined weekly, all these times

were interval-censored. The exception corresponds to a right-censored observation of a patient in Study 5, whose viral load was below 20 copies/mL after four weeks when it was measured the last time. A graphical representation of all censoring intervals is shown in Figure 2, wherein, it can be observed that all interval lengths are multipliers of one week. This is due to the fact that the medical follow-up visits in all included studies were programmed with exact multiples of seven days following a strict protocol.
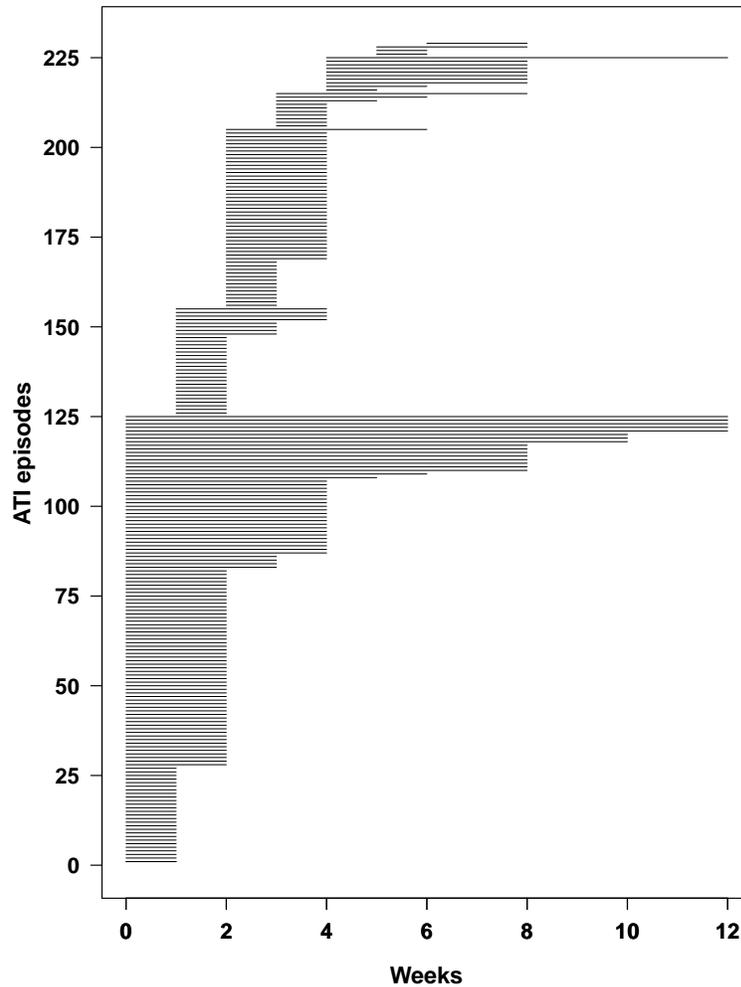
Figure 2   Lengths of the ordered interval-censored times (weeks) until viral rebound of the 229 ATI episodes. The average length of the censoring intervals is 2.6 weeks.

Additionally, in Figure 3, we provide the Turnbull estimates of the distribution functions of the time until viral rebound, obtained from Formula (1) by $1 - S_n(t)$. Therein, we can observe that the estimated probabilities of a viral rebound within the first two and four weeks are close to 0.6 and 0.9, respectively. Moreover, the separate Turnbull estimates of $F(t)$ in Studies 1 through 8 reflect a notable heterogeneity among the ATI studies: for example, the estimated probabilities of a viral rebound within the first two weeks vary from less than 0.2 (Study 4) to more than 0.8 (Study 1).
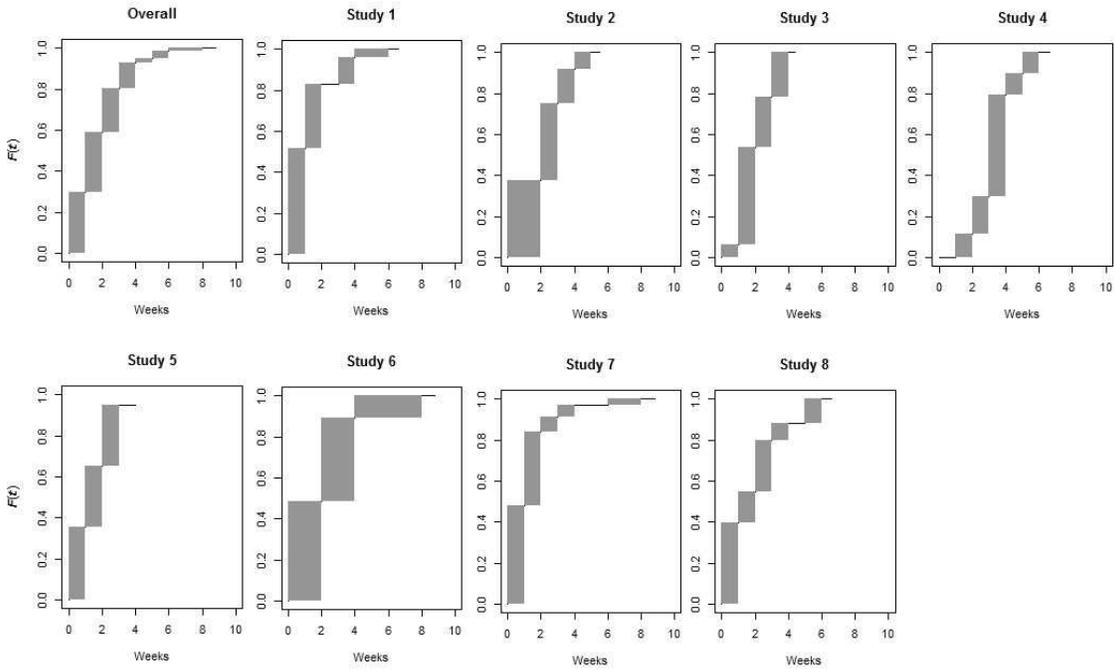
Figure 3 Non-parametric estimation of the distribution function of the time until viral rebound. The first graph (upper left corner) shows the estimation based on the pooled sample, the remaining graphs correspond to Studies 1 through 8.

4.2 Fit of the mixed effects Cox model

For the purpose of studying the possible effect of gender and log pre-cART VL on the time until viral rebound taking into account the within-subject and within-study correlation, we fitted the mixed effects model presented in (4):

$$\lambda_i(t; \mathbf{X}, \mathbf{Z}) = \lambda_0(\mathbf{t}) \exp(\mathbf{X_i}\boldsymbol{\beta} + \mathbf{Z_i}\mathbf{b}) \quad \text{with } \mathbf{b} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{8}$$

where $\mathbf{X} \in \mathbb{R}^{\mathbf{229 \times 2}}$ is the design matrix of the fixed effects Gender and log pre-cART VL and $\boldsymbol{\beta} \in \mathbb{R}^2$ is the fixed effects parameter vector. In the case of the log pre-cART VL, we decided to subtract 4, which is equivalent to a VL of 10000 HIV copies in a millilitre of blood and close to the overall median (4.31), since the value 0 was outside the range of the variable (Table 1). Concerning the random effects, we considered a random intercept per patient as well as correlated random intercept and slope for log pre-cART VL within each study. Hence, $\mathbf{b} = (\boldsymbol{b_1'}, \boldsymbol{b_2'}, \boldsymbol{b_3'})' \in \mathbb{R}^{\mathbf{158+8+8}}$ and $\mathbf{Z} \in \mathbb{R}^{\mathbf{229 \times 174}}$. In addition, the covariance matrix of $\mathbf{b}$ is given by

$$\Sigma = \begin{pmatrix} \sigma_{\boldsymbol{b_1}}^2 & 0 & 0 \\ 0 & \sigma_{\boldsymbol{b_2}}^2 & \sigma_{\boldsymbol{b_2},\boldsymbol{b_3}} \\ 0 & \sigma_{\boldsymbol{b_2},\boldsymbol{b_3}} & \sigma_{\boldsymbol{b_3}}^2 \end{pmatrix}.$$

Notice that the inclusion of a random intercept per study implies study-specific baseline hazard functions $\lambda_{0,1}(t), \ldots, \lambda_{0,8}(t)$.

For the model fit, we used the 3-step algorithm proposed in Section 3. In the first step of the algorithm, we replaced the interval-censored times until viral rebound by imputed times obtained randomly from the

truncated Weibull distribution, and in Step 2, Model (4) was fitted. Steps 1 and 2 were repeated $M = 15$ times providing the parameter estimates corresponding to the fixed effects presented in Table 2.

Table 2 Estimation of the fixed effects parameters and the standard deviation of the random effects of Model (4) using the three-step imputation method.

	$\hat{\beta}$	se($\hat{\beta}$)	$\widehat{\text{HR}}$	95%CI
Gender (Female vs Male)	0.50	0.26	1.65	[0.99, 2.74]
Log pre-cART VL	0.60	0.23	1.83	[1.16, 2.86]
Random effects	$\hat{\sigma}_{b_1}$	$\hat{\sigma}_{b_2}$	$\hat{\sigma}_{b_3}$	$\widehat{\text{Corr}}_{b_2,b_3}$
	0.35	0.34	0.34	-0.47

According to the results obtained, the instantaneous risk of viral rebound among female patients is 1.65 times larger than the instantaneous risk among male patients with the same pre-cART VL. However, since the standard error is relatively large, the corresponding 95% confidence interval of the hazard ratio does include 1 and, hence, we actually cannot claim that women are at larger risk for viral rebound than men among the population of HIV-infected persons at a confidence level of 0.95. Concerning pre-cART VL, the results obtained are clear: the larger the pre-cART VL, the larger the risk of suffering a viral rebound. The adjusted hazard ratio of 1.83 implies that a unit increase of the log pre-cART VL, that is, a 10-fold increase of the VL, increases the instantaneous risk of a viral rebound by factor 1.83.

Regarding the random effects, the estimated standard deviation of the random intercept per subject, $\hat{\sigma}_{b_1}$ can be interpreted as the unexplained variation between individuals after controlling for the explanatory variables in the model. The value of the estimated standard deviation of the random intercept per study, $\hat{\sigma}_{b_2}$, reflects the heterogeneity among the eight studies with respect to the inclusion criteria. The standard deviation $\hat{\sigma}_{b_3}$ quantifies the variability of the slopes of pre-cART VL among the eight studies. The study-specific hazard ratios associated to pre-cART VL varies from 1.28 (Study 2) to 2.49 (Study 8; values not shown). Moreover, the negative value of the estimated correlation (-0.47) between random intercept and random slope of pre-cART VL, implies that the smaller the baseline hazard function per study, the larger the effect of the pre-cART VL.

5 Simulation study

The objective of the following simulation study was to explore the properties of the estimation method presented in Section 3 in terms of bias and mean squared error of the fixed effects estimator $\hat{\beta}$ in a setting very similar to the one of the data set at hand.

5.1 Simulation settings and data generation

Data sets were generated based on Model (8) under a total of 36 different scenarios shown in Table 3.

Common to all scenarios, which were motivated by the data set presented in Section 4, were the distributions of the fixed effects variables X_1 (Binomial(1, 0.23)) and X_2 ($N(0, 0.7)$) as well as the parameter values $\beta_1 = 0.5$ and $\beta_2 = 0.6$. Regarding the random effects, the variable Study was generated from a multinomial distribution with equal probabilities for each of the eight studies, i.e., $p = 1/8$. In the case of Studies 1, 2, and 8, we duplicated the ATI episodes per patient, in order to reproduce the scenario of the real data set. The values of the random effects b_1, b_2, and b_3 were generated from a multivariate normal distribution with mean 0 and

Table 3   Settings of the simulation study

Sample size	$n \in \{100, 200, 300\}$
Assessment probability	$p \in \{0.2, 0.33, 0.5\}$
Number of imputations	$M = 15$
Distribution of T	Weibull, Gompertz
Percentage of right-censored observations	0 and 10
Distribution of X_1	Bin(1, 0.23)
Distribution of X_2	$N(0, 0.7)$
$(\beta_1, \beta_2)'$	$(0.5, 0.6)'$
Distribution of $\boldsymbol{b_1}, \boldsymbol{b_2}, \boldsymbol{b_3}$	$MVN(0, \Sigma)$ (see (9))

$$\Sigma = \begin{pmatrix} 0.35^2 & 0 & 0 \\ 0 & 0.35^2 & -0.47 \cdot 0.35 \cdot 0.35 \\ 0 & -0.47 \cdot 0.35 \cdot 0.35 & 0.35^2 \end{pmatrix}. \tag{9}$$

To generate survival times from the mixed effects proportional hazards model in (8), we used the inverse probability method described by Bender et al. (2005). According to the authors, a random survival time from Model (8) can be generated using the following equation:

$$T = \Lambda_0^{-1}\Big(-\frac{\log(U)}{\exp\left(\boldsymbol{X\beta} + \boldsymbol{Zb}\right)}\Big),$$

where U follows a uniform distribution in the interval (0, 1) and Λ_0 is the cumulative baseline hazard function. Survival times can be generated assuming the conditional distribution of T given $\boldsymbol{X}$ and $\boldsymbol{Z}$ follows a Weibull or Gompertz distribution since both share the assumption of proportional hazards (Bender et al., 2005).

In the case of the Weibull distribution with shape and scale parameters $\kappa > 0$ and $\rho > 0$, the baseline hazard function is $\lambda_0 = \kappa\rho(\rho t)^{\kappa-1}$. Hence, $\Lambda_0(t) = (\rho t)^{\kappa}$, and following the inverse probability method, a realization of T is obtained by computing

$$t = \frac{\{-\log(u) \cdot \exp\left(-\boldsymbol{X\beta} - \boldsymbol{Zb}\right)\}^{1/\kappa}}{\rho},$$

where u is a realization of U.

Regarding the Gompertz distribution with shape and scale parameters $\alpha \in (-\infty, \infty)$ and $\rho > 0$, the baseline hazard function is $\lambda_0 = \rho\exp\{\alpha t\}$. Hence, $\Lambda_0(t) = (\rho/\alpha) \cdot (\exp(\alpha t) - 1)$, and following the inverse probability method, a realization of T is obtained by computing

$$t = \frac{1}{\alpha} \cdot \log\Big(1 - \frac{\alpha\log(u)}{\rho\exp\left(-\boldsymbol{X\beta} - \boldsymbol{Zb}\right)}\Big).$$

Notice that in the case of $\alpha < 0$ there is a point mass at infinity since the argument of the log function could then be negative. However, this possibility was ruled out in the simulation study as $\alpha = 1.5$ was used to generate the data.

Given all survival times T_i, $i = 1, \ldots, n$, per data set, the censoring intervals $(L_i, R_i]$ were generated assuming noninformative censoring and following the procedure described in Gómez et al. (2009). All integers from 1 through 12, the maximum upper limit of the censoring intervals in our data set (Figure 2)

were considered possible assessment times using three different assessment probabilities for each of the 12 values: $p \in \{0.2, 0.33, 0.5\}$. That is, the assessment times were generated according to a Bernoulli distribution with parameter p and for each survival time T_i generated, the interval $(L_i, R_i]$ was obtained as the smallest interval of assessment times among all those including T_i.

In addition to the underlying distribution and the assessment probabilities, the different scenarios were determined by the sample size per data set ($n \in \{100, 200, 300\}$) and the percentage of right-censored observations (none and 10%, respectively). The number of imputations was kept the same for all settings ($M = 15$).

5.2 Evaluation criteria

Given any of the 36 simulation settings, we generated $D = 1000$ data sets, for each of which β_1 and β_2 were estimated by means of the method presented in Section 3. Thus, we obtained $\hat{\boldsymbol{\beta}}_d = (\hat{\beta}_{1,d}, \hat{\beta}_{2,d})'$ for $d = 1, \ldots, D$. Based on these estimations and given the true parameter vector $\boldsymbol{\beta_0} = (0.5, 0.6)'$, we calculated the mean, variance, bias, and mean squared error (MSE) of $\hat{\beta}_1$, and $\hat{\beta}_2$ as follows:

$$\bar{\hat{\beta}}_i = \frac{1}{D} \sum_{d=1}^{D} \hat{\beta}_{i,d}, \quad i = 1, 2,$$

$$\widehat{\mathrm{Var}}(\hat{\beta}_i) = \frac{1}{D-1} \sum_{d=1}^{D} (\hat{\beta}_{i,d} - \bar{\hat{\beta}}_i)^2, \quad i = 1, 2, \tag{10}$$

$$\widehat{\mathrm{Bias}}(\hat{\beta}_i) = \bar{\hat{\beta}}_i - \beta_{0,i}, \quad i = 1, 2,$$

$$\widehat{\mathrm{MSE}}(\hat{\beta}_i) = \widehat{\mathrm{Var}}(\hat{\beta}_i) + \widehat{\mathrm{Bias}}(\hat{\beta}_i)^2, \quad i = 1, 2.$$

The bias is a measure for the accuracy of the estimators, whereas the MSE can be used as a measure for the precision.

Concerning the estimation of the variance, $\mathrm{Var}(\hat{\beta}_{i,d})$ could also be estimated within each of the D runs and their mean could serve as well as an estimator of $\mathrm{Var}(\hat{\beta}_i)$, $i = 1, 2$. However, Formula (10) is generally more appropriate, since the mean of the D variance estimates usually underestimates the true variance $\mathrm{Var}(\hat{\beta}_i)$.

The whole simulation process was programmed in R (R Core Team, 2018) using the function `multimp` (see Appendix A) and is included in the Supporting Information.

5.3 Simulation results

The results of the simulations in terms of bias and MSE are shown in Tables 4 (no right-censored data) and 5 (10% of right-censored observations). In general, for the settings under study, it can be observed that our proposed method captures the real parameters in a proper way. Irrespective of the conditional distribution of T or the percentage of right-censored observations, the estimated bias of both $\hat{\beta}_1$ and $\hat{\beta}_2$ can be considered small.

Concerning the conditional distribution of the survival times, hardly any difference is observed between the Weibull and the Gompertz distribution with respect to the bias even though the former is always used for the imputation step of our method. Contrary to that, the MSE is generally larger in case of the Gompertz distribution indicating a somewhat smaller precision in that case. Only slight differences are observed comparing the settings without and with 10% of right-censored data. In the case of no right-censored observations, the bias is generally a bit lower, whereas the MSE is slightly bigger in most cases.

Table 4 Fixed parameters estimators without right-censored observations and 15 imputations ($\beta_1 = 0.5, \beta_2 = 0.6$)

	Weibull		Gompertz	
	Bias	**MSE**	**Bias**	**MSE**
$n = 100$				
Assessment probability: $p = 0.2$				
$\hat{\beta}_1$	0.060	0.246	0.045	0.293
$\hat{\beta}_2$	-0.001	0.078	-0.022	0.089
Assessment probability: $p = 0.33$				
$\hat{\beta}_1$	0.035	0.132	-0.007	0.135
$\hat{\beta}_2$	0.005	0.054	-0.030	0.064
Assessment probability: $p = 0.5$				
$\hat{\beta}_1$	0.001	0.090	-0.004	0.092
$\hat{\beta}_2$	-0.009	0.051	-0.021	0.049
$n = 200$				
Assessment probability: $p = 0.2$				
$\hat{\beta}_1$	-0.023	0.073	-0.028	0.092
$\hat{\beta}_2$	-0.029	0.039	-0.045	0.042
Assessment probability: $p = 0.33$				
$\hat{\beta}_1$	-0.034	0.049	-0.035	0.061
$\hat{\beta}_2$	-0.036	0.032	-0.044	0.038
Assessment probability: $p = 0.5$				
$\hat{\beta}_1$	-0.023	0.038	-0.032	0.041
$\hat{\beta}_2$	-0.032	0.030	-0.054	0.031
$n = 300$				
Assessment probability: $p = 0.2$				
$\hat{\beta}_1$	-0.038	0.046	-0.046	0.054
$\hat{\beta}_2$	-0.045	0.028	-0.064	0.036
Assessment probability: $p = 0.33$				
$\hat{\beta}_1$	-0.040	0.034	-0.051	0.039
$\hat{\beta}_2$	-0.045	0.027	-0.043	0.027
Assessment probability: $p = 0.5$				
$\hat{\beta}_1$	-0.040	0.027	-0.032	0.029
$\hat{\beta}_2$	-0.043	0.024	-0.052	0.027

The sample size does have the expected impact on the precision of $\hat{\beta}_1$ and $\hat{\beta}_2$: the MSE decreases as n increases no matter the width of the intervals or the conditional distribution of the survival times. In terms of bias, generally, no big differences are observed between the different sample sizes. However, with a sample size of $n = 100$, the bias seems to depend on the assessment probability: the larger the assessment probability and, hence, the smaller the censoring intervals, the more accurate the estimator. This tendency is, generally, not observed for $n = 200$ and $n = 300$. Moreover, the bias, even though generally small, is almost always negative in the case of $n = 200$ and $n = 300$ indicating, hence, a slight underestimation of both parameters. By contrast, with a sample size of 100, the bias is most often positive.

Table 5 Fixed parameters estimators considering 10% right-censored observations and 15 imputations ($\beta_1 = 0.5, \beta_2 = 0.6$)

	Weibull		Gompertz	
	Bias	**MSE**	**Bias**	**MSE**
n = 100				
Assessment probability: $p = 0.2$				
$\hat{\beta}_1$	0.088	0.315	0.071	0.342
$\hat{\beta}_2$	0.027	0.090	-0.015	0.091
Assessment probability: $p = 0.33$				
$\hat{\beta}_1$	0.019	0.140	0.011	0.143
$\hat{\beta}_2$	0.019	0.061	-0.018	0.062
Assessment probability: $p = 0.5$				
$\hat{\beta}_1$	0.013	0.097	0.001	0.098
$\hat{\beta}_2$	0.007	0.055	-0.005	0.053
n = 200				
Assessment probability: $p = 0.2$				
$\hat{\beta}_1$	0.001	0.085	-0.022	0.091
$\hat{\beta}_2$	-0.024	0.045	-0.030	0.047
Assessment probability: $p = 0.33$				
$\hat{\beta}_1$	-0.017	0.052	-0.010	0.065
$\hat{\beta}_2$	-0.019	0.036	-0.028	0.035
Assessment probability: $p = 0.5$				
$\hat{\beta}_1$	-0.017	0.043	-0.015	0.046
$\hat{\beta}_2$	-0.023	0.030	-0.033	0.031
n = 300				
Assessment probability: $p = 0.2$				
$\hat{\beta}_1$	-0.030	0.049	-0.037	0.057
$\hat{\beta}_2$	-0.033	0.033	-0.054	0.036
Assessment probability: $p = 0.33$				
$\hat{\beta}_1$	-0.034	0.036	-0.030	0.037
$\hat{\beta}_2$	-0.036	0.027	-0.041	0.029
Assessment probability: $p = 0.5$				
$\hat{\beta}_1$	-0.027	0.028	-0.013	0.029
$\hat{\beta}_2$	-0.023	0.025	-0.031	0.026

6   Discussion

Our collaboration as data scientists with clinicians and virologists from Hospital Clínic of Barcelona, IRB Barcelona, and University of Barcelona led to the analysis of the eight different Analytical Treatment Interruption studies in chronic HIV-positive combination Antiretroviral Therapy (cART)-suppressed individuals. The main clinical question to be studied was whether gender and pre-cART VL were risk factors on the time until viral rebound in HIV-infected patients with previously undetectable viral load, taking into account the heterogeneity between the different studies and the fact that different patients had different

number of ATI episodes. The first challenge we encountered was that the times until viral rebound were interval-censored. Other difficulties were that the studies had different inclusion criteria (for example, with respect to pre-cART VL values), and that some individuals participated in more than one study or were exposed to more than one ATI episode.

For these reasons, the analysis of this data set with the mixed effects Cox model seemed to be the natural choice. However, to the best of our knowledge, this model had not been studied with interval-censored data previously. Hence, our proposal is a first step to close the gap between the mixed effects Cox model (Therneau & Grambsch, 2000) and the Cox model with interval-censored data (Finkelstein, 1986).

The method proposed is based on multiple imputations in order to replace the censoring intervals by imputed values to simplify the data structure to uncensored and possibly right-censored survival times. For this step, we propose to generate random survival times from a truncated Weibull distribution. As an alternative to multiple imputation, single imputation methods imputation could have been applied, such as midpoint imputation replacing the censoring interval $(L_i, R_i]$ by its midpoint $(L_i + R_i)/2$. However, midpoint imputation is only reasonable when the time period between consecutive visits (or measurements) is short leading to approximately unbiased estimations. But even in this case, the standard error of the estimator would be underestimated since single imputation methods ignore the imputation uncertainty (Kim, 2003) and do not take into account the variability of the censoring interval. In contrast, multiple imputation does not attempt to estimate each missing value through simulated values but rather to represent a random sample of the missing values. This process results in valid statistical inferences that properly reflect the uncertainty due to missing data (Yuan, 2010).

According to the results of the simulation study presented in Section 5, the estimation method proposed has desirable properties in terms of accuracy (small bias) and precision (low MSE) of the estimators of the fixed effects parameters. A relatively large MSE was only observed in the case of the smallest sample size ($n = 100$) and smallest assessment probabilities ($p = 0.2$) considered (see Tables 4 and 5). We have to admit, however, that the simulation study did only comprise 36 different scenarios determined by sample size, assessment probability, condition distribution of T given $\boldsymbol{X}$ and $\boldsymbol{Z}$, and the percentage of right-censored observations. Further simulation studies would be desirable in order to explore the properties of our estimation method under additional settings or with mixed effects Cox models with a different number of fixed and random effects.

An apparent limitation of our estimation methods seems to be the fact that the imputation step is based exclusively on the (truncated) Weibull distribution. However, the Weibull distribution is a flexible distribution in the sense that its hazard function can have different shapes (constant, monotonically increasing, or monotonically decreasing). For this reason, its use must not necessarily be a limitation even though the underlying survival time distribution may be a different distribution. Actually, the simulation results in the case of the Gompertz distribution of T (small bias and low MSE) seem to confirm this supposition.

An important aspect of model fitting that has been beyond the scope of the present work is the comparison of nested mixed effect Cox models in the presence of interval-censored data. Nested mixed effects Cox models with right-censored data can be compared by means of the values of the log partial likelihood function after integrating out the random effects (Therneau, 2018b), whose difference multiplied by minus 2 follows a chi-squared distribution under the null hypothesis that the nested model does not improve the model fit. In our case, for example, we could be interested in testing, whether the inclusion of the different random effects actually improve the model fit. The application of the Likelihood Ratio Test proposed by Therneau (2018b) with our estimation method, however, is not straightforward because of the multiple imputations. One possible *ad hoc* solution to this question could be to compute, separately for the models to be compared, the corresponding values of the integrated log partial likelihood for each of the M model fits obtained with each imputation. Following, the mean difference over the M model fits could be calculated and compared to the quantiles of the corresponding chi-squared distribution. Further studies should address this topic in order to develop guidelines for researchers for how to compare nested mixed effects Cox models with interval-censored data.

Concerning the clinical results obtained, as could be expected, the higher the last viral load before the first initiation of cART (pre-cART VL), the larger the instantaneous risk of a viral rebound. In our data set, female patients were at larger instantaneous risk for viral rebound than male patients with the same pre-cART VL, however, using a 95% confidence level, we cannot claim that this is also valid among the population of HIV-infected patients. We did not study any further variables, but our R function `multimp` could be easily adapted to the estimation of more than two fixed parameters. The same could be done to consider more random effects.

We did also check whether the inclusion of the patient identifier and the (correlated) intercept and slope of pre-cART VL among studies as a random effects in the Cox model modifies the results obtained from a Cox model that ignored such random effects. To fit this model, we used the `icenReg` package (Anderson-Bergman, 2017), which enables the parameter estimation in the Cox model in the presence of interval-censored data. The differences observed were: $\hat{\beta}_1 = 0.51$ versus $\hat{\beta}_1 = 0.50$ in the case of variable Gender, and $\hat{\beta}_2 = 0.36$ versus $\hat{\beta}_2 = 0.60$ in the case of pre-cART VL. That is, ignoring the random effects ATI Study and patient identifier, the estimated hazard ratio associated with pre-cART VL would have been 1.43 and, hence, notably smaller than the estimated hazard ratio 1.83 reported in Section 4. These differences highlight the importance to consider random effects when fitting a Cox model in the presence of interval-censored data and a grouped data structure.

In Table 2, we also report the estimated standard deviations of the random effects. However, as Gleiss et al. (2018) point out, the values presented lack the direct comparability with the contribution of fixed effects. The authors are working on this issue addressing the explained variation in shared frailty models, which are a particular case of the mixed effects model, namely when a random intercept per random effect is considered.

Even though we have applied our method only to one data set, our method is valid for any data set that presents both interval-censored survival times and a grouped data structure that could be treated as a random effect in a regression model. Nonetheless, a fit of the mixed effects Cox model with interval-censored data that did not require multiple imputations would be desirable. For this purpose, the expression of the likelihood function of the Cox model with interval-censored data presented by Finkelstein (1986) would need to be extended to consider random effects and tools of the area of optimization such as the ones presented by Langohr & Gómez (2005) could be useful to achieve this goal.

Conflict of Interest *The authors have declared no conflict of interest.*

A   R code used to fit the mixed effects Cox model

```
multimp <- function(data, Tl, Tu, dist = "weibull", nimp, uncens,
                    var1, var2, rand1, rand2){
  require(truncdist)
  require(coxme)
  require(actuar)
```

```
aft <- survreg(Surv(t1, t2, cens, type="interval") ~ var1 + var2, data)
lambda <- exp(-aft$coef[[1]]/aft$scale)
alpha <- 1/aft$scale

  mat <- matrix(0, ncol = nimp, nrow = nrow(data))
  colnames(mat) <- as.vector(paste("M", 1:nimp, sep = ""))

  for(i in 1:nrow(data)){
    mat[i,] <- with(data,
                      rtrunc(nimp, spec = dist,a = Tl[i], b = Tu[i],
shape = alpha,
scale = (lambda * exp(-(aft$coef[[2]] *
data$var1[i]/sca
+ aft$coef[[3]] *
data$var2[i]/sca)))^(-1/alpha)))
  }
  mat[data$Tu == Inf, ] <- rep(data$Tl[data$Tu == Inf], nimp)
  dat <- data.frame(data, mat)

  S <- vector("list", nimp)
  mfit <- vector("list", nimp)
  beta1 <- rep(0, nimp)
  varbeta1 <- rep(0, nimp)
  beta2 <- rep(0, nimp)
  varbeta2 <- rep(0, nimp)
  std1 <- rep(0, nimp)
  std2 <- rep(0, nimp)
  std3 <- rep(0, nimp)
  corr1 <- rep(0, nimp)

  for(i in 1:nimp){
    S[[i]] <- with(dat, Surv(dat[,(ncol(dat) - nimp + i)],
                             uncens))
    mfit[[i]] <- coxme(S[[i]] ~ var1 + var2 + (1|rand1)
                      + (1 + var2 | rand2), dat)
beta1[i] <- as.numeric(fixef(mfit[[i]])[1])
varbeta1[i] <- as.numeric(vcov(mfit[[i]])[1])
beta2[i] <- as.numeric(fixef(mfit[[i]])[2])
varbeta2[i] <- as.numeric(vcov(mfit[[i]])[4])
    std1[i] <- sqrt(mfit[[i]]$vcoef[[1]])
    std2[i] <- sqrt(mfit[[i]]$vcoef[[2]][1])
    std3[i] <- sqrt(mfit[[i]]$vcoef[[2]][4])
    corr1[i] <-  mfit[[i]]$vcoef[[2]][2]
  }
```

```
b1 <- round(mean(beta1),3)
wi1 <- mean(varbeta1)
bi1 <- (sum((beta1-b1)^2))/(nimp - 1)
varb1 <- wi1 + (1 + (1/nimp)) * bi1
sb1 <- round(sqrt(varb1), 4)
hr1 <- round(exp(b1), 2)
ci1 <- round(hr1 * exp(qnorm(c(0.025, 0.975)) * sb1), 2)

b2 <- round(mean(beta2), 3)
wi2 <- mean(varbeta2)
bi2 <- (sum((beta2 - b2)^2))/(nimp - 1)
varb2 <- wi2 + (1 + (1/nimp)) * bi2
sb2 <- round(sqrt(varb2), 4)
hr2 <- round(exp(b2), 2)
ci2 <- round(hr2 * exp(qnorm(c(0.025, 0.975)) * sb2), 2)

sd1 <- round(mean(std1), 2)
sd2 <- round(mean(std2), 2)
sd3 <- round(mean(std3), 2)
cor <- round(mean(corr1), 2)


vars <- c(b1, sb1, hr1, ci1, b2, sb2, hr2, ci2, sd1, sd2, sd3, cor)
names(vars) <- c('Coef', 'se(coef)', 'HR', 'Lower 95%', 'Upper 95%',
                 'Coef', 'se(coef)', 'HR', 'Lower 95%', 'Upper 95%',
                 'sd1', 'sd2', 'sd3', 'Corr')
return(vars)

}
```

B   Additional information on the ATI data set

The eight ATI studies differ from each other with respect to the types of intervention defined by different therapeutic vaccines or drug combinations during the combination Antiretroviral Therapy (cART) and the inclusion criteria, which depended on the CD4 cell count, the plasma viral load (VL), previous years during cART, or the stage of infection; see Table 6. In addition, the number of patients varied from one study to the other as well as the geographical recruitment area. Following, some more detailed information is provided on each study.

In Study 1 (García et al., 2005), the first ATI episode was done to harvest autologous HIV virus to create the therapeutic vaccine. The therapeutic vaccine was administered just before the second ATI. Patients are coming from a previous study with non-advanced chronic HIV-1 infection.

Study 2 (García et al., 2013) has a similar design as that of Study 1. In this study, there are 2 ATI episodes, the first one is to harvest the autologous virus and the second is the post-vaccine stop.

Study 3 (`https://clinicaltrials.gov/ct2/show/NCT02767193`) corresponds to the ongoing new version of the dendritic cell-based vaccine trial. The first ATI episode was done, as in the previous studies, to harvest autologous HIV virus to create the therapeutic vaccine, which was administered just before the second ATI. Patients in this trial had to be on stable cART for at least one year and the average of all measurements of CD4 cells during the year before starting cART.

Study 4 (García et al., 2004) was carried out to evaluate the effect of mycophenolate mofetil (MMF) on the immunologic control during the ATI. In particular, the first ATI episode was done to evaluate the effect of MMF over the viral load dynamics. MMF is a well characterized drug widely used in renal transplantation because of its ability to selectively inhibit lymphocyte division and it may inhibit HIV replication. Patients were chronic HIV-1 infected persons at very early stages and were treated with cART for 12 months.

Part of the objectives of Study 5 (García et al., 2003) was to study the effect of concomitant hydroxyurea treatment (HU), which was administered to a group of patients but not in the first ATI episode. The study group were patients with chronic HIV infection from the Spanish EARTH-2 study. This study 5 was a designed study to explore the effect of controlled and repeated interruptions over the immune response against the virus. They were based on the 'autovaccination' theory, as explained in the introduction. The first ATI episode does not have any specific 'importance', but the joint set of ATI episodes (which corresponds to the studied intervention in this case).

Study 6 (Mothe et al., 2015) corresponds to patients participating in a randomized phase I HIV vaccine trial with recombinant modified vaccinia Ankara-based and Gag-Pol-Nef polyprotein with or without a drug to reactivate latent HIV in 3 centers. The only ATI episode of this study was done 8 weeks after the last dose of MVA-B and the viral rebound dynamics were assessed during the first 12 weeks after cART interruption. The ATI was done to evaluate the effect of the vaccine in the control of viral load during the absence of cART. Participants were chronically HIV-infected individuals recruited at three HIV units in Barcelona and Madrid (Spain).

In Study 7 (Fagard et al., 2003), cART was interrupted for 2 weeks, restarted for 8 weeks. After 4 such cycles, treatment was indefinitely suspended 40 weeks after study entry. The ATI rationale in this study is to try to prove the 'autovaccination hypothesis', that is that reexposure to HIV during treatment interruptions may stimulate the HIV-specific immune response and lead to low viremia after withdrawal of cART.

Study 8 (García et al., 1999, 2001) involved patients with chronic HIV-1 infection in very early stages who started a twice daily three-drug regimen. cART was discontinued after one year of treatment and effective virologic response. The ATI rationale in this study is the same as in the Study 5.

Table 6 Inclusion criteria and treatment for the different studies of the ATI data set.

Study	Patients	Treatment (apart from cART)	CD4 counts (cells/mm^3)	Viral Load (copies/mL)
Study 1	16	- Dendritic cell-based HIV-vaccine (n=12) - Placebo (n=4)	>500 (pre-cART)	>5000 (pre-cART) <20 for at least 104 weeks while receiving cART
Study 2	35	- Dendritic cell-based HIV-vaccine (n=24) - Placebo (n=11)	>450 (baseline)	<37 (enrollment)
Study 3	18		>350 (previous years) >450 (at enrollment)	Undetectable at least 6 months before inclusion
Study 4	11	- Received MMF (n=7) - Did not (n=4)	>500	200-5000 (baseline)
Study 5	20		>500 (pre-cART)	>5000 (pre-cART)
Study 6	28	- Ankara-based vaccine and Gag-Pol-Nef polyprotein with a drug to reactive latent HIV (n=19) - Placebo (n=9)	>450	Not specified
Study 7	33		>740	Undetectable for a median of 21 months
Study 8	10		>500 (last 3 months)	>10000

References

Anderson-Bergman, C. (2017). icenReg: Regression models for interval censored data in R. *Journal of Statistical Software*, *81*, 1–23.

Bebchuk, J. D., & Betensky, R. A. (2000). Multiple imputation for simple estimation of the hazard function based on interval censored data. *Statistics in Medicine*, *19*, 405–419.

Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, *24*, 1713–1723.

Bland, J. M., & Altman, D. G. (1994). Correlation, regression, and repeated data. *British Medical Journal*, *308*, 896.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*, 187–220.

Dorey, F. J., Little, R. J., & Schenker, N. (1993). Multiple imputation for threshold-crossing data with interval censoring. *Statistics in Medicine*, *12*, 1589–1603.

Fagard, C., Oxenius, A., Günthard, H., Garcia, F., Le Braz, M., Mestre, G., ... others (2003). A prospective trial of structured treatment interruptions in human immunodeficiency virus infection. *Archives of Internal Medicine*, *163*, 1220–1226.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, 845–854.

García, F., Climent, N., Guardo, A. C., Gil, C., León, A., Autran, B., ... others (2013). A dendritic cell–based vaccine elicits t cell responses associated with control of HIV-1 replication. *Science Translational Medicine*, *5*, 166ra2–166ra2.

García, F., Lejeune, M., Climent, N., Gil, C., Alcamí, J., Morente, V., ... Fumero, E. (2005). Therapeutic immunization with dendritic cells loaded with heat-inactivated autologous HIV-1 in patients with chronic HIV-1 infection. *Journal of Infectious Diseases*, *191*, 1680–1685.

García, F., Plana, M., Arnedo, M., Brunet, M., Castro, P., Gil, C., ... Martorell, J. (2004). Effect of mycophenolate mofetil on immune response and plasma and lymphatic tissue viral load during and after interruption of highly active antiretroviral therapy for patients with chronic HIV infection: a randomized pilot study. *Journal of Acquired Immune Deficiency Syndromes*, *36*, 823–830.

García, F., Plana, M., Arnedo, M., Ortiz, G. M., Miró, J. M., Lopalco, L., ... Gatell, J. M. (2003). A cytostatic drug improves control of HIV-1 replication during structured treatment interruptions: a randomized study. *AIDS*, *17*, 43–51.

García, F., Plana, M., Ortiz, G. M., Bonhoeffer, S., Soriano, A., Vidal, C., ... Pantaleo, G. (2001). The virological and immunological consequences of structured treatment interruptions in chronic HIV-1 infection. *AIDS*, *15*, F29–F40.

García, F., Plana, M., Vidal, C., Cruceta, A., Pantaleo, G., Pumarola, T., ... Gatell, J. M. (1999). Dynamics of viral load rebound and immunological changes after stopping effective antiretroviral therapy. *AIDS*, *13*, F79–F86.

Gleiss, A., Gnant, M., & Schemper, M. (2018). Explained variation in shared frailty models. *Statistics in Medicine*, *37*, 1482–1490.

Gómez, G., & Cadarso-Suárez, C. (2017). El modelo de riesgos proporcionales de Cox y sus extensiones. Impacto en estadística y biomedicina. *Gaceta de la Real Sociedad Matemática Española*, *20*, 513—538.

Gómez, G., Calle, M. L., Oller, R., & Langohr, K. (2009). Tutorial on methods for interval-censored data and their implementation in R. *Statistical Modelling*, *9*, 259–297.

Graziani, G. M., & Angel, J. B. (2015). Evaluating the efficacy of therapeutic HIV vaccines through analytical treatment interruptions. *Journal of the International AIDS Society*, *18*, 20497.

Kim, J. S. (2003). Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*, 489–502.

Langohr, K., & Gómez, G. (2005). Likelihood maximization using web-based optimization tools. *The American Statistician*, *59*, 192–202.

Leal, L., Lucero, C., Plan, M., Climent, N., Martinez, E., Castro, P., . . . Garcia, F. (2017). Viral outcomes after treatment interruptions to evaluate a functional cure [Abstract] [Computer software manual]. CROI Foundation/IAS-USA. (In CROI 2017. Conference on Retroviruses and Opportunistic Infections)

Mothe, B., Climent, N., Plana, M., Rosàs, M., Jiménez, J. L., Muñoz-Fernández, M. Á., . . . León, A. (2015). Safety and immunogenicity of a modified vaccinia ankara-based HIV-1 vaccine (MVA-B) in HIV-1-infected patients alone or in combination with a drug to reactivate latent HIV-1. *Journal of Antimicrobial Chemotherapy*, *70*, 1833–1842.

Muñoz, A., Wang, M.-C., Bass, S., Taylor, J. M., Kingsley, L. A., Chmiel, J. S., . . . Multicenter AIDS Cohort Study Group (1989). Acquired immunodeficiency syndrome (AIDS)-free time after human immunodeficiency virus type 1 (HIV-1) seroconversion in homosexual men. *American Journal of Epidemiology*, *130*, 530–539.

Nadarajah, S., & Kotz, S. (2006). R programs for truncated distributions. *Journal of Statistical Software*, *16*, 1–8.

Novomestky, F., & Nadarajah, S. (2016). truncdist: Truncated random variables [Computer software manual]. (R package version 1.0-2)

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Rubin, D. B. (2004). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

Satten, G. A., Datta, S., & Williamson, J. M. (1998). Inference based on imputed failure times for the proportional hazards model with interval-censored data. *Journal of the American Statistical Association*, *93*, 318–327.

Senn, S. (1998). Some controversies in planning and analysing multi-centre trials. *Statistics in Medicine*, *17*, 1753–1765.

Sun, J. (2007). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer Science & Business Media.

Taylor, J. M., Muñoz, A., Bass, S. M., Saah, A. J., Chmiel, J. S., & Kingsley, L. A. (1990). Estimating the distribution of times from HIV seroconversion to AIDS using multiple imputation. *Statistics in Medicine*, *9*, 505–514.

Therneau, T. M. (2018a). *Coxme and the Laplace Approximation.* Retrieved from `http://cran.r-project.org/web/packages/coxme/vignettes/laplace.pdf`

Therneau, T. M. (2018b). coxme: Mixed effects cox models [Computer software manual]. (R package version 2.2-10)

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag.

Treasure, G. C., Aga, E., Bosch, R. J., Mellors, J. W., Kuritzkes, D. R., Para, M., . . . Li, J. Z. (2016). Relationship among viral load outcomes in hiv treatment interruption trials. *Journal of Acquired Immune Deficiency Syndromes (1999)*, *72*, 310.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 290–295.

World Health Organization. (2017). *HIV/AIDS.* `http://www.who.int/features/qa/71/en/`. (Accessed: 2018-09-20)

Yamaguchi, T., Ohashi, Y., & Matsuyama, Y. (2002). Proportional hazards models with random effects to examine centre effects in multicentre cancer clinical trials. *Statistical Methods in Medical Research*, *11*, 221–236.

Yuan, Y. C. (2010). Multiple imputation for missing data: Concepts and new development (version 9.0). *SAS Institute Inc, Rockville, MD*, *49*, 1–11.