Working Paper Series, N. 1, January 2014

Integrated likelihoods in survival models for highly-stratified censored data

Giuliana Cortese
Department of Statistical Sciences
University of Padua
Italy

Nicola Sartori
Department of Statistical Sciences
University of Padua
Italy

Abstract: When inference is about a parameter of interest in presence of many nuisance parameters, standard likelihood methods often perform very poorly and may lead to severe bias. For stratified data, this problem is particularly evident in models with stratum nuisance parameters when the number of strata is relatively high with respect to the within-stratum size. Stratified data are very frequent in many applied settings, such as in cohort studies based on multi-center clinical trials, and are often incomplete, e.g., due to censoring. We consider stratified survival data in a parametric framework under the general assumption of noninformative independent censoring (both type I and random censoring schemes), and propose frequentist inference based on an integrated likelihood. When failure times have a Weibull distribution, simulation studies show that appropriately defined integrated likelihoods provide very accurate results in all circumstances, even in extreme settings where standard likelihoods lead to strongly misleading results. An application, which concerns treatments for a frequent disease in late-stage HIV-infected people, shows the proposed inferential method in Weibull regression models, and warns against different inferential conclusions when integrated and profile likelihoods are used.

Keywords: Noninformative censoring; Profile Likelihood; Right-censored Data; Stratum Nuisance Parameters; Weibull model.

Department of Statistical Sciences
University of Padua
Italy

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Cortese Giuliana
tel: +39 049 827 4124
gcortese@unipd.it
http://www.stat.unipd.it/~gcortese

Integrated likelihoods in survival models for highly-stratified censored data

Giuliana Cortese
Department of Statistical Sciences
University of Padua
Italy

Nicola Sartori
Department of Statistical Sciences
University of Padua
Italy

Abstract: When inference is about a parameter of interest in presence of many nuisance parameters, standard likelihood methods often perform very poorly and may lead to severe bias. For stratified data, this problem is particularly evident in models with stratum nuisance parameters when the number of strata is relatively high with respect to the within-stratum size. Stratified data are very frequent in many applied settings, such as in cohort studies based on multi-center clinical trials, and are often incomplete, e.g., due to censoring. We consider stratified survival data in a parametric framework under the general assumption of noninformative independent censoring (both type I and random censoring schemes), and propose frequentist inference based on an integrated likelihood. When failure times have a Weibull distribution, simulation studies show that appropriately defined integrated likelihoods provide very accurate results in all circumstances, even in extreme settings where standard likelihoods lead to strongly misleading results. An application, which concerns treatments for a frequent disease in late-stage HIV-infected people, shows the proposed inferential method in Weibull regression models, and warns against different inferential conclusions when integrated and profile likelihoods are used.

Keywords: Noninformative censoring; Profile Likelihood; Right-censored Data; Stratum Nuisance Parameters; Weibull model.

1 Introduction

Stratified or clustered data are very common in many settings, such as longitudinal and cohort studies and multi-center clinical trials, and are mostly due to the study design or to the naturally occurring hierarchies in the reference population (see, e.g., Carlin and Hodges (1999), and Ravani et al. (2010)). The major issue in clustered structures is that data within clusters may be correlated, and thus statistical models should account for this. The three general approaches to account for clustering are to introduce fixed effects or random effects into a model, or to correct inference by using sandwich variance estimators. When the number of strata (clusters) is high, fixed-effects models or use of sandwich variance estimators are the preferred approaches, since they require fewer assumptions. Presence of high-dimensional strata may often lead to violation of the assumptions on the random-effects distribution (e.g., assuming common variance for all strata).

Stratum or cluster structures are frequently encountered in studies which involve time-to-event data and censoring schemes. The current paper deals with parametric fixed-effects models for censored data, where fixed effects represent the strata and are considered as nuisance parameters, while the remaining parameters, common to all strata, are considered as parameters of interest. These models are useful when the intra-cluster correlation is not of interest. We consider noninformative independent censoring, which is a commonly used assumption in biomedical applications.

It is well known that standard likelihood inference for a parameter of interest could be seriously misleading in the presence of many nuisance parameters, relatively to the sample size. The main reason is that inference is in fact based on the profile likelihood, which is simply the likelihood in which the nuisance parameters are maximized out, for every fixed value of the parameter of interest. The profile likelihood is not a proper likelihood. Indeed, for instance, the corresponding score function is biased (Severini, 2000, Chap. 4). While this is not a problem in standard settings, this bias may grow with the dimension of the nuisance parameter and invalidate usual asymptotic results (Sartori, 2003). Many alternative pseudo likelihoods have been proposed to solve this problem, such as marginal and conditional likelihoods (see, for instance, Severini, 2000, Chap. 8). The problem with these pseudo likelihoods is that their existence depends on the model structure and, even when they exists, they may be difficult to compute. A more general alternative is to consider a modification of the profile likelihood, which takes into account the presence of nuisance parameters (Cox and Reid, 1987; Barndorff-Nielsen and Cox, 1994, Chap. 8). Modified profile likelihoods have been widely studied and it is well known that they often perform much better than the profile likelihood in models with many nuisance parameters, especially in stratified models where nuisance parameters as associated to the strata (Sartori, 2003; Bellio and Sartori, 2006). However, in the presence of censored data it is not clear how to compute the modified profile likelihood, especially under general censoring and in regression settings; an example based on Monte Carlo simulations is given in Pierce and Bellio (2006).

A recent alternative approach, which is the standard practice in Bayesian settings, is to summarize the proper likelihood by averaging with respect to some function of the nuisance parameters. In a frequentist setting, this method leads to a particular type of pseudo likelihood for the parameter of interest, called the integrated likelihood function (Severini, 2007, 2010, 2011). It has been shown that integrated likelihood functions may provide an accurate approximation to modified profile likelihood and, in some cases, may have better properties, e.g., in presence of small sample sizes (Examples 2 and 4 in Severini (2007)). Furthermore, integrated likelihoods have the advantage to be always computable and available. It is therefore of great interest to investigate the properties of inference based on integrated likelihoods in survival models for stratified censored data, especially under general censoring mechanisms. Some theoretical results are given in De Bin et al. (2013) for integrated likelihoods in models with many stratum nuisance parameters, but with no reference to survival models, nor to censored data.

The scope of the paper is then to investigate the performance of integrated likelihood functions for inference in survival models for stratified censored data. In the article, the inferential procedure based on integrated likelihoods is presented in the general setting of parametric survival models mentioned above. Furthermore, in order to show the practical use of integrated likelihood functions, the paper describes an application for a parametric model with failure times from a Weibull distribution, under both the assumptions of type I and random censoring. The regression setting is illustrated by means of a real data example about HIV-infected people.

Section 2 introduces the notation and describes the profile likelihood for stratified survival data. In Section 3 the integrated likelihood approach is presented for stratified survival models in a general setting under the assumption of noninformative independent censoring. Then, in Section 4 we illustrate the specific results of integrated likelihood functions for a stratified model with failure times from a Weibull distribution when type I censoring and random censoring are assumed. Results from Monte Carlo simulations studies for both complete data and right-censored data from a stratified Weibull model are described in Section 5. Section 6 shows the real-data application. General remarks and possible extensions are discussed in Section 7.

2 Background and profile likelihood

Let us assume a setting where data are stratified (or clustered) with $i = 1, \ldots, n$ strata and $j = 1, \ldots, k_i$ observations within each stratum i. The total sample size is $m = \sum_{i=1}^{n} k_i$. Suppose also that the observations are times to a certain event, T_{ij}, which may be observed over the period $(0, \tau)$.

Consider then a parametric model with stratified observations of the form

$$T_{ij} \sim p_{ij}(t_{ij}; \psi, \lambda_i), \tag{1}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, k_i$. In the following, for ease of notation, we assume $k_1 = \ldots = k_n = k$ so that all strata have the same size. Let the model depend on $\theta = (\psi, \lambda)$ where ψ is a parameter of interest taking values in Ψ, and $\lambda = (\lambda_1, \ldots, \lambda_n)$ is a n-dimensional nuisance parameter. Each within-stratum nuisance parameter λ_i is assumed to be scalar for simplicity, without compromising the validity of the results in the paper.

Let us define with $\mathbf{T}_i = (T_{i1}, \ldots, T_{ik})$ the vector of independent within-stratum random variables. Suppose that $\mathbf{T}_1, \ldots, \mathbf{T}_n$ are independent but not identically distributed because of the clustering structure and the possible presence of covariates, and have densities $p_i(\mathbf{t}_i; \psi, \lambda_i)$. Denote with $S_i(\mathbf{t}_i; \psi, \lambda_i) = \Pr(\mathbf{T}_i > \mathbf{t}_i)$ the survival function of $\mathbf{T}_i$, with $h_i(\mathbf{t}_i; \psi, \lambda_i) = p_i(\mathbf{t}_i; \psi, \lambda_i)/S_i(\mathbf{t}_i; \psi, \lambda_i)$ the hazard function, and with $H_i(\mathbf{t}_i; \psi, \lambda_i) = -\log S_i(\mathbf{t}_i; \psi, \lambda_i)$ the cumulative hazard function of $\mathbf{T}_i$.

Regression models that consider explanatory variables x_{ij} may also be considered without additional difficulties to the theoretical aspects presented in the paper. The application in Section 6 is an example of such regression models for randomly-censored data.

Complete data are available if all the realizations of T_{ij}, for $i = 1, \ldots, n$ and $j = 1, \ldots, k$, are observed. Typically time-to-event data are incomplete, i.e., observations are subject to right censoring. In the following, notation and introductory theory are provided, first, for complete data and then for incomplete right-censored data.

2.1 Complete data

In presence of complete data, the log likelihood function can be written as

$$\ell(\psi, \lambda) = \sum_{i=1}^{n} \ell^i(\psi, \lambda_i), \tag{2}$$

where the log likelihood contribution of the ith stratum is

$$\ell^i(\psi, \lambda_i) = \log[h_i(\mathbf{t}_i; \psi, \lambda_i)\ S_i(\mathbf{t}_i; \psi, \lambda_i)] = \log h_i(\mathbf{t}_i; \psi, \lambda_i) - H_i(\mathbf{t}_i; \psi, \lambda_i).$$

Note that, because of the independence assumption between the $\mathbf{T}_i$, and thus between strata, each contribution of stratum i to the likelihood, $\ell^i(\psi, \lambda_i)$, depends only on the corresponding within-stratum nuisance parameter λ_i. Therefore, the log likelihood is separable with respect to the nuisance parameters, for fixed ψ.

From the independence between $T_{i1}, \ldots, T_{ik}$, we have that $h_i(\mathbf{t}_i; \psi, \lambda_i) = \prod_{j=1}^{k} h_{ij}(t_{ij}; \psi, \lambda_i)$, $S_i(\mathbf{t}_i; \psi, \lambda_i) = \prod_{j=1}^{k} S_{ij}(t_{ij}; \psi, \lambda_i)$, and then the log likelihood function for the ith stratum may be written as

$$\ell^i(\psi, \lambda_i) = \sum_{j=1}^{k} \log[p_{ij}(t_{ij}; \psi, \lambda_i)] = \sum_{j=1}^{k} [\log h_{ij}(t_{ij}; \psi, \lambda_i) - H_{ij}(t_{ij}; \psi, \lambda_i)]. \tag{3}$$

Let $\hat{\theta} = (\hat{\psi}, \hat{\lambda})$ be the maximum likelihood estimator of $\theta = (\psi, \lambda)$. Standard likelihood inference for the parameter of interest ψ is based, either explicitly or implicitly, on the profile log likelihood function, which for stratified data can be given as the sum of n strata terms

$$\ell_P(\psi) = \sum_{i=1}^{n} \ell^i(\psi, \hat{\lambda}_{i\psi}) = \sum_{i=1}^{n} \ell^i_P(\psi). \tag{4}$$

The elements of $\hat{\lambda}_\psi = (\hat{\lambda}_{1\psi}, \ldots, \hat{\lambda}_{n\psi})$, the maximum likelihood estimates of λ for fixed ψ, are the solutions to the independent likelihood estimating equations for the strata

$$\ell^i_{\lambda_i}(\psi, \lambda_i) \equiv \frac{\partial}{\partial \lambda_i} \ell^i(\psi, \lambda_i) = 0.$$

The function $\ell_P(\psi)$ is then used for construction of point estimates and test statistics such as the likelihood ratio statistic $W = 2[\ell_P(\hat{\psi}) - \ell_P(\psi)]$ for inference on ψ, or the signed square root $R = \text{sgn}(\hat{\psi} - \psi)\sqrt{W}$ when ψ is scalar. For stratified data, the usual asymptotic properties of $\ell_P(\psi)$ are valid only when $n = o(k)$, which is not very common in such settings (Sartori, 2003).

2.2 Censored data

Let $\tilde{T}_{ij}$ be the failure times with densities given in (1) and C_{ij} be the censoring times. In presence of incomplete information, for each unit (i, j) the observed data is represented by the couple (T_{ij}, Δ_{ij}), where $T_{ij} = \min(\tilde{T}_{ij}, C_{ij})$ and $\Delta_{ij} = I(\tilde{T}_{ij} \leq C_{ij})$.

Let us assume independent and noninformative censoring, i.e., the censoring mechanism does not depend on the times to event nor on their distribution. Consequently, the distribution of the C_{ij} does not depend on the parameters (ψ, λ). Moreover, let us consider one of the two alternative censoring schemes:

(i) Type I censoring: the censoring times c_{ij} are fixed in advance. Constant values $c_1, \ldots, c_n$ may also be assumed across the n strata, so that $C_{ij} = c_i$ for $j = 1, \ldots, k$.

(ii) Random censoring: censoring times are random variables C_{ij} with density and survival functions $g_{ij}(\cdot)$ and $G_{ij}(\cdot)$, respectively.

Suppose that the censoring times have the same structure as the failure times and $(C_{i1}, \ldots, C_{ik})$, for $i = 1, \ldots, n$, are independent vectors of within-stratum random variables with densities $g_i(\cdot)$.

For fixed censoring as in point (i), the joint density of (T_{ij}, Δ_{ij}) has the simple form

$$f_{ij}(t_{ij}, \delta_{ij}; \theta) = \begin{cases} p_{ij}(t_{ij}; \theta)^{\delta_{ij}} S_{ij}(t_{ij}; \theta)^{1-\delta_{ij}} = h_{ij}(t_{ij}; \theta)^{\delta_{ij}} S_{ij}(t_{ij}; \theta), & \text{if } t_{ij} \in (0, c_{ij}] \\ 0, & \text{if } t_{ij} > c_{ij}\,. \end{cases}$$

In the random censoring scheme (ii) the distribution of C_{ij} must also be taken into account, leading to the form

$$f_{ij}(t_{ij}, \delta_{ij}; \theta, \nu) = p_{ij}(t_{ij}; \theta)^{\delta_{ij}}\ S_{ij}(t_{ij}; \theta)^{1-\delta_{ij}}\ G_{ij}(t_{ij})^{\delta_{ij}}\ g_{ij}(t_{ij})^{1-\delta_{ij}}$$

for the joint density of (T_{ij}, Δ_{ij}). Here and in the following, we can assume that the C_{ij} have either a nonparametric distribution $G_{ij}(\cdot)$, or a parametric distribution that depends on ν with survival and density functions denoted as $G_{ij}(\cdot; \nu)$ and $g_{ij}(\cdot; \nu)$ respectively. In the latter case, the observed censoring time c_{ij} is both a partially sufficient statistic for ν (sufficient statistic when θ is fixed) and a constant statistic for all θ (it does not depend on θ because of noninformative censoring). Thus, the (i, j) contribution to the likelihood function for θ may be based on the conditional density

$$f_{T_{ij}, \Delta_{ij}/C_{ij}=c_{ij}}(t_{ij}, \delta_{ij}; c_{ij}, \theta) = p_{ij}(t_{ij}; \theta)^{\delta_{ij}}\ S_{ij}(t_{ij}; \theta)^{1-\delta_{ij}}, \tag{5}$$

which does not depend on the parameter ν of the censoring distribution.

Therefore, if we assume also that $C_{i1}, \ldots, C_{ik}$ are independent, the full likelihood $L(\theta, \nu)$ is separable with respect to the parameters θ and ν, and can be factorized as $L(\theta, \nu) = L(\theta)\ L(\nu)$. The factor $L(\theta)$ can then be used as a proper likelihood for θ.

Under both fixed and random censoring assumptions, the log likelihood $\log L(\theta) = \ell(\theta) = \ell(\psi, \lambda)$ is still given as the sum in equation (2), where the contribution of the

ith stratum is

$$\ell^i(\psi, \lambda_i) = \sum_{j=1}^{k} [\delta_{ij} \log h_{ij}(t_{ij}; \psi, \lambda_i) - H_{ij}(t_{ij}; \psi, \lambda_i)]. \tag{6}$$

Since each $\ell^i(\psi, \lambda_i)$ depends only on the corresponding within-stratum nuisance parameter λ_i, the profile likelihood for ψ can be computed as a sum of the n strata terms as given in equation (4) for complete data.

3 Integrated likelihood in stratified survival data

In the following we introduce the integrated likelihood function for ψ (Severini, 2007), which has the form

$$\bar{L}(\psi) = \int_{\Lambda} L(\psi, \lambda)\pi(\lambda \mid \psi) d\lambda, \tag{7}$$

where $\pi(\lambda \mid \psi)$ is a nonnegative weight function for the nuisance parameter $\lambda \in \Lambda$. It is not required for $\pi(\lambda \mid \psi)$ to be a proper density, but its integral on the space Λ should have the same finite value given each ψ. Severini (2007) provides suggestions for the proper choice of $\pi(\lambda \mid \psi)$, so that the corresponding integrated likelihood has good frequentist properties.

To illustrate the idea behind integrated likelihood theory, let us consider the ideal situation where $L(\theta) = L(\psi)L(\lambda)$, i.e., the likelihood factorises and is then separable with respect to θ. In this case, integrated likelihoods can be completely independent of the selection of weight functions, provided this latter does not depend on ψ. This is because any choice of the weight function such that $\pi(\lambda \mid \psi) = \pi(\lambda)$ provides the same integrated likelihood. Approximately, a similar situation can be obtained when parameters are orthogonal, i.e., the element $i_{\psi\lambda}(\theta)$ of the Fisher information is null. Of course if separable likelihoods are encountered in practice, inference is based only on $L(\psi)$ and we do not need to resort to pseudo likelihoods such as integrated likelihoods. In general, this case is not frequent and often the model parameters are not orthogonal. Integrated likelihood theory achieves to find a new nuisance parameter that is orthogonal to ψ (in the meaning proposed by Severini (2007)). The method is based on obtaining an interest-respecting data-dependent reparameterization of $\theta = (\psi, \lambda)$, which leads to a nuisance parameter $\phi \equiv \phi(\psi, \lambda; \hat{\psi})$ dependent on the data only through $\hat{\psi}$. For the property of orthogonality, construction of the nuisance parameter ϕ requires that two basic properties are fulfilled:
(a) ϕ is strongly unrelated to ψ, i.e., its constrained estimator $\hat{\phi}_\psi$ should be approximately constant as a function of ψ, i.e., $\hat{\phi}_\psi = \hat{\phi} + O(n^{-1/2})O(|\psi - \hat{\psi}|)$ for small deviation of ψ from $\hat{\psi}$;
(b) the weight function for ϕ does not depend on ψ, so that ϕ and ψ are independent under this function.
Finally, the integrated likelihood with respect to $\pi(\phi)$ defined over Φ is given by

$$\bar{L}(\psi) = \int_{\Phi} \tilde{L}(\psi, \phi)\pi(\phi) d\phi,$$

where $\tilde{L}(\psi,\phi)$ is the likelihood reparameterized in (ψ,ϕ). This integrated likelihood is then approximately score-unbiased to order $O(n^{-1})$.

In practice, the two major steps to compute integrated likelihoods consist of finding the parameter ϕ and choosing an opportune weight function $\pi(\phi)$ as given in points (a) and (b) above. Define with $\ell_\lambda(\cdot)$ the partial derivative of the likelihood function with respect to λ. It has been proved by Severini (2007) that property (a) is verified when $\phi \equiv \phi(\psi,\lambda;\hat{\psi})$ is the solution to the equation

$$E\{\ell_\lambda(\psi,\lambda);\hat{\psi},\phi\} \equiv E\{\ell_\lambda(\psi,\lambda);\psi_0,\lambda_0\} \mid_{(\psi_0,\lambda_0)=(\hat{\psi},\phi)} = 0, \tag{8}$$

where $(\psi,\lambda,\hat{\psi})$ are considered as fixed values. The solution ϕ has been defined as the zero-score-expectation parameter, since it recalls the likelihood property of score-unbiasedness.

When dealing with stratified data, under the assumptions previously made, it is sufficient to solve n independent equations

$$E\{\ell_{\lambda_i}(\psi,\lambda_i);\hat{\psi},\phi_i\} \equiv E\{\ell_{\lambda_i}(\psi,\lambda_i);\psi_0,\lambda_{i0}\} \mid_{(\psi_0,\lambda_{i0})=(\hat{\psi},\phi_i)} = 0, \tag{9}$$

where $\ell^i_{\lambda_i}(\psi,\lambda_i)$ are the n independent score functions for the nuisance parameters and the solutions ϕ_i are the elements of the zero-score-expectation parameter $\phi = (\phi_1,\ldots,\phi_n)$. Note that each equation and the corresponding solution ϕ_i depend only on the parameter λ_i.

For property (b), it has been shown that for any weight function fulfilling this property, results from integrated likelihoods are approximately unchanged. It is then suggested that often the uniform function $\pi(\phi) = 1$ is an appropriate choice (for a more general discussion about how to select weight functions see, e.g., Severini (2010)).

Finally, inference based on the integrated log likelihood $\bar{\ell}(\psi) = \log \bar{L}(\psi)$ can be performed by means of, e.g., the likelihood ratio statistics $\bar{W} = 2[\bar{\ell}(\hat{\psi}) - \bar{\ell}(\psi)]$ for inference on ψ, or the signed square root $\bar{R} = \mathrm{sgn}(\hat{\psi}-\psi)\sqrt{\bar{W}}$ when ψ is scalar (Severini, 2010).

3.1 Integrated likelihood for stratified survival data

Under the assumption of noninformative independent censoring (both type I and random censoring schemes), recall from Section (2) that the likelihood function for (ψ,λ) is

$$L(\psi,\lambda) = \prod_{i=1}^{n} L^i(\psi,\lambda_i) = \prod_{i=1}^{n}\left[\prod_{j=1}^{k} h_{ij}(t_{ij};\psi,\lambda_i)^{\delta_{ij}}\; S_{ij}(t_{ij};\psi,\lambda_i)\right],$$

with $\ell^i(\psi,\lambda_i) = \log L^i(\psi,\lambda_i)$.

We need to reparameterize this likelihood as a function of the zero-score expectation parameter ϕ. Since, in our setting of stratified data, nuisance parameters belong to distributions of independent variables, it is suitable to choose a weight function of the form $\pi(\phi) = \pi(\phi_1) \cdots \pi(\phi_n)$. The resulting integrated likelihood is a product of n independent integrals, and is given by

$$\bar{L}(\psi) = \prod_i \int_{\Phi_i} \tilde{L}^i(\psi, \phi_i) \pi(\phi_i) d\phi_i,$$

where $\tilde{L}(\psi, \phi) = \prod_i \tilde{L}^i(\psi, \phi_i)$ and $\tilde{L}^i(\psi, \phi_i)$ is the likelihood contribution of the ith stratum reparameterized in (ψ, ϕ_i).

In the following, the procedure to find the zero-score expectation parameter $\phi = (\phi_1, \ldots, \phi_n)$ is presented for right-censored data, first, under type I censoring, and then assuming random censoring. Complete data are presented as a special case when all $\delta_{ij} = 1$.

Let us define the vector of independent variables $\mathbf{\Delta}_i = (\Delta_{i1}, \ldots, \Delta_{ik})$. Assume without loss of generality that the random censoring times are equally distributed within each stratum, i.e., $C_{i1} = \ldots = C_{ik} = C_i$.

(i) *Type I censoring.* The C_i can be considered as random variables with probability mass function $\Pr(C_i = c_i) = 1$. The fixed censoring times are known in advance for all subjects, and are equal for all observations within stratum i, i.e., $c_{i1} = \ldots = c_{ik} = c_i$ for $i = 1, \ldots, n$. Consequently, the zero-score expectation parameter is found as a solution to the equations

$$E_{\mathbf{T}_i, \mathbf{\Delta}_i | C_i = c_i} \left[\ell^i_{\lambda_i}(\psi, \lambda_i);\ \psi_0, \lambda_{i0}, c_i \right] |_{(\psi_0, \lambda_{i0}) = (\hat{\psi}, \phi_i)} = 0, \qquad i = 1, \ldots, n,$$

where the expected values are taken with respect to the conditional random variables $(\mathbf{T}_i, \mathbf{\Delta}_i) | C_i = c_i$. Thus, since the pairs $(T_{i1}, \Delta_{i1}), \ldots, (T_{ik}, \Delta_{ik})$ were assumed to be independent, the equations reduce to

$$\sum_{j=1}^{k} \sum_{\delta_{ij}=0}^{1} \left[\int_0^\infty \ell^{ij}_{\lambda_i}(\psi, \lambda_i)\ f_{T_{ij}, \Delta_{ij} / C_i = c_i}(t, \delta_{ij}; c_i, \psi_0, \lambda_{i0})\ dt \right] = 0, \tag{10}$$

with $f_{T_{ij}, \Delta_{ij} / C_i = c_i}(\cdot)$ given in (5).

Furthermore, let us write

$$\ell^i_{\lambda_i}(\psi, \lambda_i) = \sum_j \ell^{ij}_{\lambda_i}(\psi, \lambda_i) = \sum_j \left[\delta_{ij}\ \eta_{ij, \lambda_i}(t_{ij}; \psi, \lambda_i) - H_{ij, \lambda_i}(t_{ij}; \psi, \lambda_i) \right],$$

where $\eta_{ij, \lambda_i}(t; \psi, \lambda_i) = \frac{\partial}{\partial \lambda_i} \log h_{ij}(t; \psi, \lambda_i)$, and $H_{ij, \lambda_i}(t; \psi, \lambda_i) = \frac{\partial}{\partial \lambda_i} H_{ij}(t; \psi, \lambda_i)$, $i = 1, \ldots, n$ and $j = 1, \ldots, k$. Computing the equation (10) with the expressions of $\ell^{ij}_{\psi, \lambda_i}(\cdot)$ given above and with $f_{T_{ij}, \Delta_{ij} / C_i = c_i}(\cdot)$ given in (5), leads to the final explicit equations

$$\begin{aligned} &\sum_{j=1}^{k} \int_0^{c_i} \left[\eta_{ij, \lambda_i}(t; \psi, \lambda_i) - H_{ij, \lambda_i}(t; \psi, \lambda_i) \right]\ p_{ij}(t; \psi_0, \lambda_{i0})\ dt \\ &- \sum_{j=1}^{k} H_{ij, \lambda_i}(c_i; \psi, \lambda_i)\ S_{ij}(c_i; \psi_0, \lambda_{i0}) = 0. \end{aligned} \tag{11}$$

The zero-score expectation parameters ϕ_i are the solutions to these equations after setting $(\psi_0, \lambda_{i0}) = (\hat{\psi}, \phi_i)$.

If data are complete, $\delta_{ij} = 1$ for all subjects and the equations reduce to

$$\sum_{j=1}^{k} \int_0^\infty [\eta_{ij,\lambda_i}(t;\psi,\lambda_i) - H_{ij,\lambda_i}(t;\psi,\lambda_i)] \; p_{ij}(t;\psi_0,\lambda_{i0}) \; dt \; = 0. \tag{12}$$

(ii) *Random censoring.* The expectations involved are taken with respect to the marginal variable $(\mathbf{T}_i, \mathbf{\Delta}_i)$. Since noninformative censoring has been assumed and the c_{ij} are sufficient statistics for the parameter of the censoring distribution, it is convenient to write the expected value as $E_{T,\Delta}(\cdot) = E_C\left[E_{T,\Delta}(\cdot|C)\right]$. We then obtain the zero-score expectations as

$$\begin{aligned} & E_{\mathbf{T}_i,\mathbf{\Delta}_i}\left[\ell^i_{\lambda_i}(\psi,\lambda_i);\; \psi_0,\lambda_{i0}\right]|_{(\psi_0,\lambda_{i0})=(\hat{\psi},\phi_i)} \\ &= \int_0^\infty E_{\mathbf{T}_i,\mathbf{\Delta}_i}\left[\ell^i_{\lambda_i}(\psi,\lambda_i);\; \psi_0,\lambda_{i0}|C_i = c\right]|_{(\psi_0,\lambda_{i0})=(\hat{\psi},\phi_i)} \; g_i(c)\; dc, \;\; i=1,\ldots,n, \end{aligned} \tag{13}$$

where the conditional expectations given $C_i = c_i$ are obtained from (10) and (11). The expected values in (13) are then set equal to zero to find ϕ_i for $i = 1, \ldots, n$.

4 The Weibull model for stratified survival data

We illustrate the inferential procedure based on integrated likelihood for right-censored time-to-event data from a Weibull model. This model is of particular interest also because its logarithmic transformation leads to a parametric accelerated failure time model, frequently used in many areas of application.

Let $\tilde{T}_{ij}$, $i = 1, \ldots, n$ and $j = 1, \ldots, k$, be independent failure times from Weibull distributions with probability density functions of the form

$$p_i(\tilde{t}_{ij}) = \lambda_i \psi(\lambda_i \; \tilde{t}_{ij})^{\psi-1} \exp\{-(\lambda_i \; \tilde{t}_{ij})^{\psi}\},$$

for $\tilde{t}_{ij} \geq 0$, with shape parameter $\psi > 0$ as the parameter of interest and nuisance scale parameters $\lambda_i > 0$ for $i = 1, \ldots, n$. Assume $\tilde{T}_{i1}, \ldots, \tilde{T}_{ik}$ be i.i.d. with common scale parameter λ_i. The hazard and survival functions are $h_i(\tilde{t}_{ij}) = \lambda_i\psi(\lambda_i\tilde{t}_{ij})^{\psi-1}$ and $S_i(\tilde{t}_{ij}) = \exp\{-(\lambda_i\tilde{t}_{ij})^{\psi}\}$, respectively.

Let us define the quantities $\delta_{i\cdot} = \sum_{j=1}^{k} \delta_{ij}$, that is the number of observed events in the ith stratum, $\delta_{\cdot\cdot} = \sum_{i=1}^{n} \delta_{i\cdot}$, which gives the total number of observed events, and $t_{i\cdot,\psi} = \sum_{j=1}^{k} t_{ij}^{\psi}$. Then, the ith contribution to the likelihood function has the form

$$L^i(\psi,\lambda_i) = \psi^{\delta_{i\cdot}} \; \lambda_i^{\psi\delta_{i\cdot}} (\prod_j t_{ij}^{\delta_{ij}(\psi-1)}) \exp\{-\lambda_i^{\psi} t_{i\cdot,\psi}\}. \tag{14}$$

If all the failure times are observed, we have $\delta_{ij} = 1$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, k$, and thus it is sufficient to replace $\delta_{i\cdot} = k$ for all i and $\delta_{\cdot\cdot} = nk$ in equation (14).

From the likelihood, we can obtain the following independent score functions for the nuisance parameters

$$\ell^i_{\lambda_i}(\psi, \lambda_i) = \frac{\delta_{i\cdot}\psi}{\lambda_i} - \psi\lambda_i^{\psi-1} t_{i\cdot,\psi}\,, \qquad i = 1, \ldots, n, \tag{15}$$

where again $\delta_{i\cdot} = k$ when no data are censored. From the corresponding score equations, the resulting constrained estimates are $\hat{\lambda}_{i\psi} = (\delta_{i\cdot}/t_{i\cdot,\psi})^{1/\psi}$. The maximum likelihood estimate $\hat{\psi}$ is obtained as the maximum of the profile log likelihood function

$$\ell_P(\psi) = \delta_{\cdot\cdot}(\log\psi - 1) + (\psi - 1)\sum_i\sum_j \delta_{ij}\log t_{ij} + \sum_i \delta_{i\cdot}[\log\delta_{i\cdot} - \log t_{i\cdot,\psi}]. \tag{16}$$

For computation of the integrated log likelihood for ψ, we may achieve orthogonality of the parameters by constructing the zero-score-expectation parameter $\phi = (\phi_1, \ldots, \phi_n)$. First, define $E\{\cdot; \hat{\psi}, \phi_i\} \equiv E\{\cdot; \psi_0, \lambda_{i0})\} \mid_{(\psi_0,\lambda_{i0})=(\hat{\psi},\phi_i)}$ and let us consider the expectation of (15), for $i = 1, \ldots, n$, which takes the form

$$E\{\ell^i_{\lambda_i}(\psi, \lambda_i); \hat{\psi}, \phi_i\} = k\ \frac{\psi}{\lambda_i} E\{\Delta_{ij}; \hat{\psi}, \phi_i\} - k\ \psi\lambda_i^{\psi-1} E\{T^{\psi}_{ij}; \hat{\psi}, \phi_i\}, \tag{17}$$

since $\ell_{\lambda_i}(\psi, \lambda_i)$ depends on the data only through the statistic $t_{i\cdot,\psi} = \sum_{j=1}^k t^{\psi}_{ij}$, and $T_{i1}, \ldots, T_{ik}$ are i.i.d.

As in Section 3.1, for ease of reading, let us consider the simpler situation where censoring times are equal within each stratum, i.e., $C_{i1} = \ldots = C_{ik} = C_i$. In the following, the computation of the zero-score expectation parameter is presented separately for the type I censoring and random censoring schemes, while a final section is devoted to the specification of the integrated likelihood.

4.1 The Weibull model under type I censoring

For the Weibull model, the zero-score expectation is obtained from equation (17), and the expected values therein are computed as conditional expectations given $C_i = c_i$, with respect to the density function

$$f_{T_{ij}/C_i=c_i}(t_{ij}; \psi, \lambda_i) = p_i(t_{ij}; \psi, \lambda_i)\ I_{[0,c_i)}(t_{ij}) + S_i(c_i; \psi, \lambda_i)\ I_{\{c_i\}}(t_{ij}). \tag{18}$$

This function is equivalent to the density of the observed time $T_{ij} = \min(\tilde{T}_{ij}, c_i)$. Therefore,

$$\begin{aligned} E_{\Delta_{ij}/C_i=c_i}\{\Delta_{ij}; \hat{\psi}, \phi_i, c_i\} &= \Pr(\tilde{T}_{ij} < c_i; \psi_0, \lambda_{i0}) \mid_{(\psi_0,\lambda_{i0})=(\hat{\psi},\phi_i)} \\ &= 1 - S_i(c_i; \hat{\psi}, \phi_i) = 1 - e^{(\phi_i c_i)^{\hat{\psi}}}, \end{aligned} \tag{19}$$

$$\begin{aligned} E_{T_{ij}/C_i=c_i}\{T^{\psi}_{ij}; \hat{\psi}, \phi_i, c_i\} &= \int_0^{\infty} t^{\psi} f_{T_{ij}/C_i=c_i}(t; \hat{\psi}, \phi_i)dt \\ &= \int_0^{c_i} u^{\psi} p_i(u; \hat{\psi}, \phi_i)du\ +\ c_i^{\psi} S_i(c_i; \hat{\psi}, \phi_i) \\ &= \left(\frac{1}{\phi_i}\right)^{\psi}\ \Gamma_I\left((\frac{\psi}{\hat{\psi}} + 1), (\phi_i c_i)^{\hat{\psi}}\right)\ +\ c_i^{\psi}\ e^{-(\phi_i c_i)^{\hat{\psi}}}, \end{aligned} \tag{20}$$

where $\Gamma_I(s,x) = \int_0^x t^{s-1}e^{-t}dt$ is the incomplete gamma function.

The final equations of the zero-score expectation are given as

$$\frac{k\psi}{\lambda_i}\left\{1 - e^{-(\phi_i\ c_i)^{\hat{\psi}}}[1+(\lambda_i c_i)^{\psi}] - (\frac{\lambda_i}{\phi_i})^{\psi}\ a(\phi_i,\psi,\hat{\psi})\ \right\} = 0$$

for $i = 1,\ldots,n$, where $a(\phi_i,\psi,\hat{\psi}) = \Gamma_I\left((\frac{\psi}{\hat{\psi}}+1),(\phi_i\ c_i)^{\hat{\psi}}\right)$. If we allow censoring times to be different within each stratum, the zero-score expectations become

$$\frac{\psi}{\lambda_i}\sum_j\left\{1 - e^{-(\phi_i c_{ij})^{\hat{\psi}}}[1+(\lambda_i\ c_{ij})^{\psi}] - (\frac{\lambda_i}{\phi_i})^{\psi}\ a(\phi_i,\psi,\hat{\psi})\ \right\} = 0.$$

Solving these equations for ϕ_i, for $i = 1,\ldots,n$, does not yield solutions in closed form. However, in order to compute the integrated likelihood, it is sufficient to find the nuisance parameter λ_i as a function of ϕ_i from the above equations, as follows

$$\lambda_i(\phi_i) = \left[\frac{E\{\Delta_{ij};\hat{\psi},\phi_i,c_i\}}{E\{T_{ij}^{\psi};\hat{\psi},\phi_i,c_i\}}\right]^{\frac{1}{\psi}} = \phi_i\left[\frac{1-e^{(\phi_i c_i)^{\hat{\psi}}}}{a(\phi_i,\psi,\hat{\psi})+(\phi_i c_i)^{\psi}\ e^{(\phi_i c_i)^{\hat{\psi}}}}\right]^{\frac{1}{\psi}}. \qquad (21)$$

Formulae for complete data may be obtained as a special case when $c_{ij}\to\infty$; then we have $E\{\Delta_{ij};\hat{\psi},\phi_i\} = 1$ and $E\{T_{ij}^{\psi};\hat{\psi},\phi_i\} = (1/\phi_i^{\psi})\ \Gamma(\psi/\hat{\psi}+1)$. The equations used to find the zero-score expectation parameter simplify to

$$\frac{k\psi}{\lambda_i}\left[1 - (\frac{\lambda_i}{\phi_i})^{\psi}\ \Gamma(\frac{\psi}{\hat{\psi}}+1)\ \right] = 0,$$

which lead to the solutions $\phi_i = \lambda_i\ \Gamma(\psi/\hat{\psi}+1)$, $i = 1,\ldots,n$.

4.2 The Weibull model under random censoring

For simplicity, let us assume that C_{ij} for $i = 1,\ldots,n$ and $j = 1,\ldots,k$, are i.i.d and exponentially distributed, i.e., $C_{ij}\sim \text{Exp}(\nu)$ with density $g(c) = \nu e^{-\nu c}$ and survival $G(c) = e^{-\nu c}$. However, less restrictive assumptions are also possible, e.g., $C_{ij}\sim\text{Exp}(\nu_i)$ for $j = 1,\ldots,k$, with different parameters ν_i across strata.

The zero-score expectations for the Weibull model can be computed again from (17). Note that now the expected values therein are taken with respect to the unconditional variable Δ_{ij} and T_{ij}, respectively, because of random censoring. However, these expected values can be computed similarly to those given in equation (13), considering the conditional expectations given $C_{ij} = c$ and then integrating with respect to c, as follows. First, using the conditional expectation in (19) it can be shown that

$$\begin{aligned} E\{\Delta_{ij};\hat{\psi},\phi_i,\nu\} &= E_{C_{ij}}\left\{E_{\Delta_{ij}}(\Delta_{ij}|C_{ij})\right\} \\ &= 1 - \int_0^{\infty} S_i(c;\hat{\psi},\phi_i)\ g(c;\nu)dc \\ &= 1 - \int_0^{\infty}\nu\ e^{-(\phi_i c)^{\hat{\psi}}-\nu c}dc, \end{aligned}$$

and using the conditional expectation in (20) it can be shown that

$$\begin{aligned} E\{T_{ij}^{\psi};\hat{\psi},\phi_i,\nu\} &= E_{C_{ij}}\left\{E_{T_{ij}}(T_{ij}^{\psi}|C_{ij})\right\} \\ &= \int_0^{\infty} t^{\psi} f_{T_{ij}}(t;\hat{\psi},\phi_i,\nu)\ dt \\ &= \int_0^{\infty} t^{\psi}\left(\hat{\psi}\ t^{\hat{\psi}-1}\phi_i^{\hat{\psi}}+\nu\right)\ e^{-(\phi_i t)^{\hat{\psi}}-\nu t}\ dt, \end{aligned}$$

where $f_{T_{ij}}(t;\psi,\lambda_i,\nu) = p_i(t;\psi,\lambda_i)\ G_i(t;\nu) + g_i(t;\nu)\ S_i(t;\psi,\lambda_i)$. Therefore, it is sufficient to substitute these expected values within equation (17) by considering the maximum likelihood estimate $\hat{\nu}$ as a plug-in estimate for ν. It leads to the final form of the zero-score equations

$$k\ \frac{\psi}{\lambda_i}E\{\Delta_{ij};\hat{\psi},\phi_i,\hat{\nu}\} - k\ \psi\lambda_i^{\psi-1}E\{T_{ij}^{\psi};\hat{\psi},\phi_i,\hat{\nu}\} = 0. \tag{22}$$

Solutions for ϕ are not in closed form. However, in order to compute the integrated likelihood function for ψ, the original nuisance parameter λ_i can be easily written as a function of the zero-score ϕ_i as follows

$$\lambda_i(\phi_i) = \left[\frac{E\{\Delta_{ij};\hat{\psi},\phi_i,\hat{\nu}\}}{E\{T_{ij}^{\psi};\hat{\psi},\phi_i,\hat{\nu}\}}\right]^{1/\psi}. \tag{23}$$

Note that this result is equivalent to what was obtained under type I censoring (equation (21)), except for the way the needed expectations are computed.

4.3 The integrated likelihood

Under both type I and random censoring, the integrated likelihood is constructed as

$$\bar{L}(\psi) = \prod_i \int_{\Phi_i} \tilde{L}^i(\psi,\phi_i)\pi(\phi_i)d\phi_i,$$

where

$$\tilde{L}^i(\psi,\phi_i) = \psi^{\delta_{i\cdot}}\ [\lambda_i(\phi_i)]^{\psi\,\delta_{i\cdot}}\ \exp\left\{-t_{i.,\psi}\,[\lambda_i(\phi_i)]^{\psi}\right\}\prod_j t_{ij}^{\delta_{ij}(\psi-1)}, \qquad i=1,\ldots,n. \tag{24}$$

The factors $\tilde{L}^i(\psi,\phi_i)$ are obtained by reparameterizing the likelihood contributions (14) in (ψ,ϕ_i), i.e., writing λ_i as a function of ϕ_i, which is given in equation (21) for the case of type I censoring, and in equation (23) for the case of random censoring. Integration is then performed numerically.

5 Simulation studies

Monte Carlo simulation studies with 5000 trials were conducted to investigate the performance of integrated likelihood methods for inference on the parameter of interest ψ. Results were also compared with those obtained from the profile likelihood.

We simulated stratified data from the Weibull model (and the related assumptions) presented in Section 4 under no censoring (complete data), type I censoring and random censoring.

To achieve good inferential properties of integrated likelihoods, proper weight functions of the nuisance parameter should not depend on the parameter of interest. To investigate also this issue in simulation studies, we studied an integrated likelihood with constant weights $\pi(\phi_1) = \ldots, \pi(\phi_n) = 1$, denoted as $\bar{\ell}(\psi)$, and an integrated likelihood with improper ψ-dependent weights, denoted as $\bar{\ell}_D(\psi)$. For simplicity, these latter weights are chosen so that the integrated likelihood has a closed form. This is possible when, in order to solve the integral, we set $z = t_{i.,\psi}\,\lambda_i(\phi_i)^\psi$ and $dz/d\phi_i = \frac{\partial}{\partial \phi_i}(t_{i.,\psi}\,\lambda_i(\phi_i)^\psi) = \pi(\phi_i)$.

The simulation studies investigated the coverage probabilities of confidence intervals of level 0.95 based on the signed likelihood ratio statistic. This statistic was computed for the profile log likelihood (R), for the integrated log likelihood with constant weights ($\bar{R}$), and for the integrated log likelihood with ψ-dependent weights ($\bar{R}_D$). The integrated signed likelihood ratio statistic, and the standard signed likelihood ratio statistic, are considered asymptotically standard normal, and the approximation to its distribution is often more accurate for the former.

We considered $n = 5, 20, 100, 250$ for the number of strata (equal to the dimension of the nuisance parameter), and different within-stratum sample sizes, $k = 10, 30, 60$. For censored data, we also assumed different censoring probabilities equal to $0.2, 0.4, 0.6$. Table 1 shows the results from stratified complete data sampled from Weibull distributions with common shape parameter $\psi = 1.5$ and different scale parameter $\lambda_i = 0.2\,i$ for $i = 1, \ldots, n$. The empirical bias of $\bar{\psi}$ is very close to zero for almost all scenarios, while the maximum likelihood estimates for ψ can be severily biased, in particular when the number of observations within strata is low ($k = 10$), and the maximum likelihood estimate bias remains constant when n increases, as known in the literature (Sartori, 2003).

Simulated data under type I censoring were sampled from Weibull distributions with the same parameters as for the complete data. The fixed censoring times c_i were assumed to be equal within each stratum, and were obtained as solutions when setting the survival $G_i(c_i)$ equal to the desired censoring probability. Results from Tables 2 and 3 show a very good performance of integrated likelihood for all different k and n and for all censoring probabilities. The empirical coverage probabilities for $\bar{R}$ are very close to the nominal values, in contrast to the empirical coverages for R, which perform very poorly. In particular, the latter get worse when n increases, for decreasing k, and for lower censoring probabilities, producing in some cases critical inferential results which are substantially wrong (e.g., for prob. cens. $0.2, 0.4$, $n = 250$ and $k = 10$). For the profile likelihood, empirical errors on the tails of the distribution are very asymmetric and the asymmetry worsens when n increases and k decreases. In contrast, the integrated likelihood provides very symmetric empirical errors in all scenarios. The empirical bias for $\bar{\psi}$ is systematically lower than the bias for $\hat{\psi}$ and reaches very low values when k and n increase (e.g., the bias < 0.001 for $n = 100, 250$ and $k = 30, 60$). We note a substantial reduction of the bias for $\bar{\psi}$ with respect to $\hat{\psi}$ for small samples, independently of the value of n. Empirical standard errors are approximately equal. Tables 2 and 3 report also numerical results from the integrated likelihood with improper ψ-dependent weights, as defined above. This counter-example shows very poor performance of the integrated likelihood if the weight function on the nuisance parameter depends on ψ, since the property of orthogonality is violated.

Stratified data under random censoring were simulated again from Weibull distributions with shape parameter $\psi = 1.5$. The scale parameters λ_i for $i = 1, \ldots, n$ were sampled from a Normal distribution with mean 3 and variance 0.5^2. Then, it was assumed that $C_{ij} \sim \text{Exp}(\nu)$; given $(\lambda_1, \ldots, \lambda_n)$ and a certain probability of censoring (P_0), the parameter ν was found as a solution to the equation

$$\frac{1}{n}\sum_{i=1}^{n} \Pr(\tilde{T}_{ij} > C_{ij}) = \frac{1}{n}\sum_{i=1}^{n} E\{\Delta_{ij}; \psi, \lambda_i, \nu\} = P_0.$$

Under random censoring, simulated results based on the integrated likelihood are even better than under type I censoring, as shown in Tables 4 and 5. Empirical coverages for $\bar{R}$ are very near to the nominal levels, whereas empirical coverages for R are substantially wrong for most cases. They perform worse when n increases, k decreases and for lower censoring probabilities, and they fail completely for $n = 100, 250$ and $k = 10, 30$. Numerical results about the estimates for ψ are very similar to those under type I censoring.

Finally, we compare the estimated standard errors for $\hat{\psi}$ and $\bar{\psi}$ given in Tables 3 and 5, respectively. Standard errors obtained for the maximum likelihood estimator are systematically biased downward, especially for lower k, and high censoring probability (0.6), independently of n. This may be the reason why we have the counterintuitive fact that empirical coverages based on profile likelihood improved when the censoring probability increases. Standard errors obtained from integrated likelihood estimation are instead always very accurate.

The robustness of the proposed inferential approach to misspecification of the censoring distribution was investigated with additional simulation studies. Censoring times were generated under a uniform distribution, whereas an incorrect exponential model was assumed. We found that in general the results based on integrated likelihoods are very robust to the assumed misspecification of the censoring distribution (see Tables 6 and 7).

6 The accelerated failure time regression model for real data from HIV-infected patients

In this section we provide an example of how the integrated likelihood method can be applied to regression models for stratified survival data. Let us consider data from a clinical trial comparing two treatments (group 2 vs group 1) for Mycobacterium avium complex, which is a frequent disease in late-stage HIV-infected people. The data were illustrated in Carlin (1999) for fitting a stratified parametric Weibull model, and part of the trial was reported in Cohn et al. (1999). A total of 69 patients coming from 11 different clinical centers were enrolled in the trial, and between them, 5 patients died in the treatment group 1 and 13 in the group 2. The endpoint of interest was time to death or censoring. The interesting aspect of these data is that many centers have enrolled a relatively small number of patients. Moreover, a high proportion of randomly censored patients was observed (74%), no events were observed in 3 centers and few deaths (1 to 4) were observed in each of the remaining centers.

For such data, the censoring mechanism is assumed to be random, the different clinical centers represent the strata and the type of treatment is the covariate of interest, x_{ij}. Consider a regression model with hazard function $h_i(t; x_{ij}) = h_0(t)\ \xi_i\ e^{\beta x_{ij}}$, where $h_0(t)$ is the common baseline hazard, ξ_i for $i = 1, \ldots, n$ are the stratum-specific effects and x_{ij} is the covariate of the jth individual in stratum i. This is a proportional hazards model with stratum-specific baseline hazards $h_{0i}(t) = h_0(t)\xi_i$. Let the survival times to death be independent variables such that $\tilde{T}_{ij} \sim \text{Weibull}(\eta_{ij}, \psi_1)$. Then, in order to have a Weibull regression model, the ith stratum's scale parameter is set to be $\eta_{ij} = e^{-(\alpha_i + \psi_2 x_{ij})}$. If we use the parameterization $\xi_i = e^{-\psi_1 \alpha_i}$ and $\beta = -\psi_1 \psi_2$, the hazard for center i can be written as

$$h_i(t; x_{ij}) = h_0(t)\ \xi_i\ e^{\beta x_{ij}} = \psi_1 t^{\psi_1 - 1}\ e^{-\psi_1(\alpha_i + \psi_2 x_{ij})},$$

with baseline hazard $h_0(t) = \psi_1 t^{\psi_1 - 1}$. Our aim is to make inference for the parameter of interest $\psi = (\psi_1, \psi_2)$, i.e. for the treatment effect β, while treating the vector of stratum-specific effects, $(\alpha_1, \ldots, \alpha_n)$, as nuisance parameter.

For ease of interpretation, the Weibull regression model can also be transformed to an accelerated failure time model, by considering the log-transformation $U_{ij} = \log \tilde{T}_{ij}$. In this case we have the linear model $U_{ij} = \alpha_i + \psi_2 x_{ij} + \sigma\ \epsilon_{ij}$, where ϵ_{ij} has an extreme value distribution and $\sigma = 1/\psi_1$. This model leads to proportional cumulative hazards on the time scale, i.e., $H_1(t) = H_2(t/c)$ for any constant c.

Standard profile likelihood methods applied to the Weibull regression model for the HIV data, lead to the maximum likelihood estimates $\hat{\psi}_1 = 1.149$ (s.e. 0.219) and $\hat{\psi}_2 = -1.012$ (s.e. 0.484). For the invariance property, we obtain that the relative risk is estimated to be $e^{\hat{\beta}} = 3.198$, indicating a higher mortality rate for patients under treatment 2. The likelihood ratio statistic was used for testing the null hypothesis $\psi_2 = 0$ and the treatment effect was found to be significant at the 0.05 level ($W = 4.896$, $p = 0.027$). The likelihood ratio test for $\psi_1 = 1$ provided a nonsignificant result and this suggests that a simpler exponential regression model could be assumed for fitting our HIV data. When we assumed $\psi_1 = 1$, testing for the null effect of ψ_2 gave very similar conclusions. The likelihood ratio-based confidence intervals for ψ_1 and ψ_2 are respectively, $(0.742, 1.669)$, and $(-2.310, -0.111)$.

If inference is based on integrated likelihood, the zero-score expectation for the Weibull regression model can be computed directly from Section 4 , under the assumption of random censoring, by using the parameterization

$$\eta_{ij} = e^{-(\alpha_i + \psi_2 x_{ij})} \tag{25}$$

Now the scale parameters, as well the zero-score expectation, depend also on the index j because of the presence of covariates. Then η_{ij} and ϕ_{ij} play here the same role as, respectively, the parameters λ_i and ϕ_i in Section 4. Because of the invariance property of integrated likelihoods with respect to reparameterizations of the zero-score expectation parameter, the new zero-score expectation parameter, ω_i, is obtained from the relation $\phi_{ij} = e^{-(\omega_i + \hat{\psi}_2 x_{ij})}$. Moreover, the strata have now different sizes and therefore from equations (17) and (22), the zero-score expectations for finding the ϕ_{ij} reduce to

$$E\{\ell^i_{\eta_{ij}}(\psi, \eta_{ij}); \hat{\psi}, \phi_{ij}\} = \frac{\psi_1}{\eta_{ij}} E\{\Delta_{ij}; \hat{\psi}, \phi_{ij}\} - \psi_1 \eta_{ij}^{\psi_1 - 1} E\{T^{\psi}_{ij}; \hat{\psi}, \phi_{ij}\}. \tag{26}$$

The following relation holds

$$E\{\ell^i_{\alpha_i}(\psi, \alpha_i); \hat{\psi}, \omega_i\} = \sum_j (\frac{\partial \eta_{ij}}{\partial \alpha_i}) E\{\ell^i_{\eta_{ij}}(\psi, \eta_{ij}); \hat{\psi}, \phi_{ij}\} = -\sum_j \eta_{ij}\, E\{\ell^i_{\eta_{ij}}(\psi, \eta_{ij}); \hat{\psi}, \phi_{ij}\}, \tag{27}$$

where $E\{\ell^i_{\eta_{ij}}(\psi, \eta_{ij}); \hat{\psi}, \phi_{ij}\}$ is given in equation (26).

Consequently, substituting the new parameterization for η_{ij} and ϕ_{ij} (equation (25)) in the latter expression and setting it equal to zero, yields

$$\alpha_i(\omega_i) = -\frac{1}{\psi_1} \ \log \left[\frac{\sum_j E\{\Delta_{ij}; x_{ij}, \hat{\psi}, \omega_i\}}{\sum_j E\{T^{\psi}_{ij}; x_{ij}, \hat{\psi}, \omega_i\}} \right]. \tag{28}$$

Finally, the integrated likelihood is

$$\bar{L}(\psi_1, \psi_2) = \prod_i \int_{-\infty}^{+\infty} \tilde{L}^i(\psi_1, \psi_2, \omega_i) \pi(\omega_i) d\omega_i,$$

where the integrand has the form

$$\tilde{L}^i(\psi_1, \psi_2, \omega_i) = \psi_1^{\delta_{i\cdot}} A_i \ \exp\{-\psi_1\, \delta_{i\cdot}\, \alpha_i(\omega_i) - \psi_1\, \psi_2\, X_i - B_i\, e^{-\psi_1\, \alpha_i(\omega_i)}\}$$

with

$$A_i = \prod_j t_{ij}^{(\psi_1 - 1)\delta_{i\cdot}}, \qquad B_i = \sum_j t_{ij}^{\psi_1} e^{-\psi_1 \psi_2 x_{ij}}, \qquad X_i = \sum_j x_{ij} \delta_{ij}.$$

The likelihood is computed with $\alpha_i(\omega_i)$ given in (28). The integrand above can be obtained from the integrand $\tilde{L}^i(\psi_1, \psi_2, \phi_{ij})$ by substituting $\eta_{ij}(\phi_{ij}) = e^{-\alpha_i(\omega_i) - \psi_2 x_{ij}}$.

The integrated likelihood approach applied to the data about HIV-infected subjects leads to the estimates $\bar{\psi}_1 = 1.037$ (s.e. 0.207) and $\bar{\psi}_2 = -1.017$ (s.e. 0.536). Using the property of parameterization invariance of integrated likelihoods, the relative risk was estimated to be $e^{\bar{\beta}} = 2.869$. The hazard ratio and the estimated ψ_1 from the integrated likelihood are lower than those from the profile likelihood. Tests and confidence intervals can be computed from integrated likelihood ratio statistics, e.g. for ψ_1, as $\overline{W} = 2[\bar{\ell}(\bar{\psi}_1, \bar{\psi}_2) - \bar{\ell}(\psi_1, \bar{\psi}_{2,\psi_1})]$. The test for the null hypothesis $\psi_2 = 0$ showed that the significance of the treatment effect is borderline at the 0.05 level ($\overline{W} = 3.983$, $p = 0.046$). Thus, unlike the test based on profile likelihood, we conclude here that there is weaker evidence against the null additional effect of treatment 2 with respect to treatment 1. The result of testing the null hypothesis $\psi_1 = 1$ was the same as for the profile likelihood. Confidence intervals based on the integrated likelihood ratio were equal to $(0.649, 1.503)$ for ψ_1, and to $(-2.450, -0.018)$ for ψ_2.

The different results obtained under the two inferential methods are illustrated in Figure 1, where the relative log likelihoods, i.e. $-\frac{1}{2}W$ and $-\frac{1}{2}\overline{W}$, are plotted. Point and interval estimates for the shape parameter ψ_1 under the integrated likelihood have lower values than those under the profile likelihood, however confidence intervals have very similar length. The estimate for ψ_2 is almost the same under the two inferential methods, whereas the confidence interval for ψ_2 computed with the integrated likelihood is slightly wider, as also shown in the right panel of Figure 1. Note that the contribution to the profile likelihood is null for the strata containing only censored observations ($\delta_{i\cdot} = 0$) with possible consequences on the inferential results, while this is not true for the integrated likelihood.

7 Discussion

Standard likelihood inference may be seriously biased when dealing with stratified data (McCullagh and Tibshirani, 1990; Sartori, 2003), especially if the number of strata is large. This problem happens because the profile score function does not provide an unbiased estimating equation, with bias generally of order $O(n)$, therefore increasing with the number of strata n (i.e., number of nuisance parameters). For inferential purposes with stratified data, integrated likelihoods are an appealing alternative since they have a score function with reduced bias and the same asymptotic properties as the modified profile likelihood (De Bin et al., 2013). The latter pseudo likelihood has been studied for censored survival data only to a limited extent (Pierce and Bellio, 2006), and it is not straightforward to compute in practice due to the complexity of the quantities involved. In contrast, integrated likelihoods have not been investigated for censored survival data. The current paper presented this inferential approach for such data and showed the benefits especially for highly stratified survival data.

The simulation studies confirm the superiority of the integrated likelihood over the profile likelihood for stratified right-censored data when a Weibull model and noninformative independent censoring are assumed. The numerical results show that inference based on profile likelihood is very inaccurate and provides serious under-coverages of confidence intervals, which lead to extremely high empirical type I errors. These problems are particularly emphasized when the number of strata (n) increases with respect to the within-stratum size (k). In contrast, the integrated likelihood shows very good performance both in terms of accuracy of the corresponding estimates, and coverage probabilities of confidence intervals in all scenarios, and the results do not seem to be affected by increasing proportions of censoring in the data.

In this work we used the likelihood ratio statistic for constructing confidence intervals. This was motivated by their invariance properties. However, in the literature it has been shown that score, Wald and likelihood ratio statistics are still asymptotically equivalent, even in the extreme scenario provided by stratified data. Moreover, this equivalence is valid independently of the chosen likelihood method (profile, modified profile or integrated likelihood) (Sartori, 2003; De Bin et al., 2013). This fact suggests that also for stratified survival data, the Wald and score statistics are expected to give improved accuracy in inference when they are based on a suitable integrated likelihood.

The proposed integrated likelihood is, as the profile likelihood, parameterization-invariant with respect to interest-preserving reparameterizations. Its computation relies often on numerical integration and therefore careful implementation is needed in order to obtain efficient code. Our implementation in the R framework (R Core Team, 2013) made use of C subroutines to speed up computation in case of many strata. The code is available from the first author upon request.

It should be noted that the integrated likelihood rely on the censoring distribution, which is needed for computing the zero-score expectation parameter, while the profile likelihood is independent of such distribution. Even though we found that the integrated likelihood is fairly robust to misspecification of the censoring distribution, the use of a completely nonparametric censoring distribution in the construction of the integrated likelihood is certainly of interest and will be the subject of future research.

In the motivating example about stratified HIV data, the two treatments are significantly different when using the standard profile likelihood approach, while this difference is not so evident with the proposed inferential method. Moreover, profile likelihoods suffer from ignoring the data information in strata without observed events, as it happens in our example. These conflicting data results were also widely discussed in the literature (Cohn et al., 1999). It is reasonable to think that this problem may be due to lack of accuracy of the inferential approach in presence of stratified data, rather than the results of choosing between unstratified and stratified analysis (Carlin and Hodges, 1999).

It would be of considerable interest to investigate further the integrated likelihood approach for highly stratified survival data, with respect to the within-stratum size, in more general settings such as left-truncated data, informative censoring, and semiparametric models. In light of the results showed in the current paper, we expect that inferential results will benefit greatly from the use of integrated likelihoods also in these settings.

Acknowledgements

The second author acknowledges the financial support of the CARIPARO Foundation Excellence - grant 2011/2012.

References

Barndorff-Nielsen, O. E., and Cox, D. R. (1994). *Inference and Asymptotics.* London: Chapman and Hall.

Bellio, R., and Sartori, N. (2006). Practical use of modified maximum likelihoods for stratified data. *Biometrical Journal* **48**, 876–886.

Berlin, J. A., Kimmel, S. E., Ten Have, T. R., and Sammel, M. D. (1999). An empirical comparison of several clustered data approaches under confounding due to cluster effects in the analysis of complications of coronary angioplasty. *Biometrics* **55,** 470–476.

Carlin, P. B., Hodges, J. S. (1999). Hierarchical proportional hazards regression models for highly stratified data. *Biometrics* **55**, 1162-1170.

Cohn, D. L., Fisher E. J., Peng J. S., Hodges J. S., Chesnut J., Child C. C., Franchino B., Gibert C. L., El-Sadr W., Hafner R., Korvick J., Ropka M., Heifets L., Clotfelter J., Munroe D., Horsburgh C. R., and Terry Beirn Community Programs for Clinical Research on AIDS (1999). A Prospective Randomized Trial of Four Three-Drug Regimens in the Treatment of Disseminated Mycobacterium avium Complex Disease in AIDS Patients: Excess Mortality Associated with High-Dose Clarithromycin. *Clinical Infectious Diseases* **29**, 125-133.

Cox, D. R., and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with Discussion). *Biometrika* **76**, 477–88.

De Bin, R., Sartori, N., and Severini, T. A. (2013). Signed integrated likelihood ratio statistic in models with stratum nuisance parameters. *Submitted.*

McCullagh, P., and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society. Series B (Methodological)* **52**, 325–344.

Pierce, D. A., and Bellio, R. (2006). Effects of the reference set on frequentist inferences. *Biometrika* **93**, 425–438.

R Core Team (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL http://www.R-project.org/.

Ravani, P., Parfrey, P., MacRae, J., James, M., Quinn, R., Malberti, F., Brunori, G., Mandolfo, S., Tonelli, M., Hemmelgarn, B., Manns, B., and Barrett, B. (2010). Modeling survival of arteriovenous accesses for hemodialysis: semiparametric versus parametric methods. *Clinical Journal of the American Society of Nephrology* **5**, 1243-1248.

Sartori, N. (2003). Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika* **90**, 533–549.

Severini, T. A. (2000). *Likelihood methods in statistics.* Oxford University Press, Oxford, UK.

Severini, T. A. (2007). Integrated likelihood functions for non-Bayesian inference. *Biometrika* **94**, 529–542.

Severini, T. A. (2010). Likelihood ratio statistics based on an integrated likelihood. *Biometrika* **97**, 481–496.

Severini, T. A. (2011). Frequency properties on inferences based on an integrated likelihood function. *Biometrika* **21**, 433–447.

Table 1: Complete data: Empirical percentage coverage probabilities of two-sided 95% confidence intervals, and empricial bias of estimates of the parameter of interest ψ, based on the profile (R and $\hat{\psi}$) and integrated ($\bar{R}$ and $\bar{\psi}$) likelihoods. Lower and upper empirical non-coverage probabilities on the left and right tails are given in brackets.

n	k	R	$\bar{R}$	Bias($\hat{\psi}$) (s.e.)	Bias($\bar{\psi}$) (s.e.)
	10	89.8 (0.5, 9.7)	94.9 (2.4, 2.7)	0.130 (0.196)	0.024 (0.184)
5	30	93.8 (1.1, 5.1)	94.8 (2.4, 2.8)	0.039 (0.101)	0.007 (0.099)
	60	94.8 (1.2, 4.0)	95.6 (2.0, 2.4)	0.020 (0.068)	0.004 (0.068)
	10	77.7 (0.1, 22.2)	95.3 (2.3, 2.4)	0.110 (0.095)	0.006 (0.089)
20	30	90.2 (0.4, 9.4)	95.4 (2.3, 2.2)	0.034 (0.049)	0.002 (0.048)
	60	92.4 (0.6, 7.0)	95.0 (2.3, 2.6)	0.017 (0.034)	0.001 (0.034)
	10	25.7 (0.0, 74.3)	95.1 (2.3, 2.6)	0.105 (0.042)	0.001 (0.040)
100	30	68.9 (0.0, 31.1)	94.8 (2.3, 2.9)	0.032 (0.022)	0.001 (0.022)
	60	82.9 (0.1, 17.0)	95.5 (1.9, 2.6)	0.016 (0.015)	0.000 (0.015)
	10	1.5 (0.0, 98.5)	95.3 (2.2, 2.4)	0.105 (0.027)	0.000 (0.025)
250	30	37.1 (0.0, 62.9)	95.7 (2.1, 2.1)	0.032 (0.014)	0.000 (0.014)
	60	63.1 (0.02, 36.8)	94.8 (2.5, 2.7)	0.016 (0.010)	0.000 (0.010)

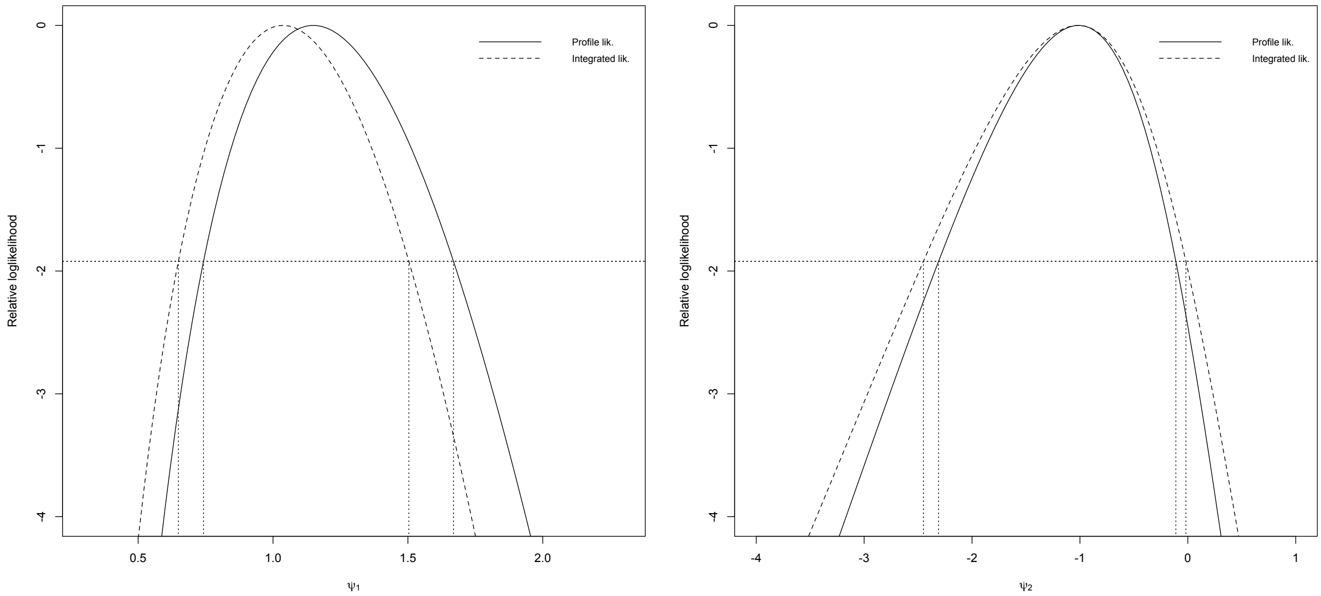

Figure 1: Relative log likelihood for ψ_1 and ψ_2 and corresponding confidence intervals (dotted lines), computed from the profile likelihood (solid lines) and the integrated likelihood (dashed lines).

Table 2: Type I-censored data: Empirical percentage coverage probabilities of two-sided 95% confidence intervals based on the profile (R) and integrated log likelihoods ($\bar{R}$: uniform weights; $\bar{R}_D$: ψ-dependent weights), for different censoring probabilities. Lower and upper empirical non-coverage probabilities are given in brackets.

Prob. cens.	n	k	R	$\bar{R}$	$\bar{R}_D$
0.2	5	10	92.6 (1.2, 6.2)	95.0 (2.4, 2.6)	93.2 (5.9, 0.9)
		30	93.9 (2.4, 3.7)	94.6 (3.4, 2.0)	92.9 (6.0, 1.1)
		60	94.2 (1.8, 4.0)	94.5 (2.5, 3.0)	94.7 (3.4, 2.0)
	20	10	90.1 (0.3, 9.6)	95.1 (2.9, 2.0)	87.6 (11.9, 0.5)
		30	93.2 (1.1, 5.8)	95.3 (2.2, 2.4)	93.0 (6.2, 0.7)
		60	94.6 (1.6, 3.8)	94.5 (2.9, 2.6)	93.5 (5.2, 1.2)
	100	10	69.2 (0.0, 30.8)	93.9 (3.9, 2.2)	57.6 (42.4, 0.0)
		30	87.6 (0.2, 12.3)	95.0 (2.5, 2.5)	82.0 (17.9, 0.1)
		60	91.6 (0.7, 7.7)	94.8 (2.8, 2.5)	88.5 (11.1, 0.4)
	250	10	37.3 (0.0, 62.6)	94.5 (3.9, 1.6)	20.9 (79.1, 0.0)
		30	76.3 (0.0, 23.6)	94.8 (2.9, 2.4)	61.3 (38.7, 0.0)
		60	86.8 (0.2, 13.0)	94.9 (2.5, 2.6)	77.7 (22.1, 0.1)
0.4	5	10	93.7 (1.4, 5.0)	95.3 (2.4, 2.3)	93.8 (5.5, 0.7)
		30	94.5 (1.7, 3.8)	94.8 (2.4, 2.7)	94.5 (4.3, 1.2)
		60	95.4 (1.9, 2.8)	95.5 (2.3, 2.2)	95.3 (3.3, 1.4)
	20	10	91.7 (0.9, 7.4)	94.8 (2.6, 2.6)	87.0 (12.7, 0.3)
		30	93.6 (1.4, 5.0)	94.6 (2.5, 2.9)	92.1 (7.1, 0.8)
		60	94.6 (1.6, 3.7)	94.5 (2.7, 2.8)	93.3 (5.4, 1.2)
	100	10	83.1 (0.2, 16.8)	94.5 (3.8, 1.6)	51.4 (48.6, 0.0)
		30	91.0 (0.7, 8.3)	94.3 (2.8, 2.9)	79.8 (20.1, 0.1)
		60	92.8 (1, 6.2)	94.3 (2.8, 2.9)	87.5 (12.2, 0.3)
	250	10	64.9 (0.1, 35.0)	94.2 (4.4, 1.3)	14.9 (85.1, 0.0)
		30	86.3 (0.3, 13.4)	95.1 (2.6, 2.3)	56.0 (44.0, 0.0)
		60	91.0 (0.4, 8.6)	95.0 (2.4, 2.6)	76.3 (23.7, 0.0)
0.6	5	10	94.2 (1.5, 4.3)	94.8 (2.3, 2.9)	95.6 (3.4, 1.1)
		30	94.5 (2.1, 3.4)	94.7 (2.7, 2.6)	94.7 (3.7, 1.6)
		60	95.0 (2.0, 3.0)	94.9 (2.3, 2.7)	95.1 (3.2, 1.7)
	20	10	93.3 (0.9, 5.8)	95.1 (2.3, 2.6)	92.7 (7, 0.3)
		30	94.6 (1.6, 3.8)	94.5 (2.7, 2.8)	93.8 (5.2, 1.0)
		60	95.0 (1.8, 3.2)	95.2 (2.3, 2.5)	94.6 (4.3, 1.1)
	100	10	89.8 (0.4, 9.8)	94.5 (3.8, 1.7)	74.1 (25.9, 0.0)
		30	93.1 (0.8, 6.1)	95.2 (2.5, 2.4)	86.8 (13.1, 0.2)
		60	93.9 (1.4, 4.7)	94.8 (2.7, 2.5)	90.8 (8.9, 0.3)
	250	10	79.6 (0.2, 20.3)	94.3 (4.4, 1.3)	43.7 (56.3, 0.0)
		30	90.8 (0.7, 8.5)	94.7 (3.1, 2.1)	74.1 (25.9, 0.0)
		60	92.7 (1.0, 6.3)	94.8 (2.9, 2.3)	83.9 (15.9, 0.1)

Table 3: Type I-censored data: Estimates of the parameter of interest ψ obtained from profile and integrated likelihoods, for different censoring probabilities. Bias, simulation-based empirical standard errors (s.e.), and ratios between the average of estimated standard errors (e.s.e.) and s.e. are provided.

Prob.				$\hat{\psi}$			$\bar{\psi}$			$\bar{\psi}_D$	
cens.	n	k	Bias	s.e.	e.s.e./s.e.	Bias	s.e.	e.s.e./s.e.	Bias	s.e.	e.s.e./s.e.
		10	0.095	0.226	0.938	0.020	0.216	0.992	-0.057	0.208	0.990
	5	30	0.016	0.125	0.947	-0.005	0.123	0.969	-0.032	0.121	0.966
		60	0.016	0.085	0.986	0.005	0.084	0.987	-0.008	0.084	0.996
		10	0.074	0.109	0.957	0.001	0.105	1.008	-0.076	0.101	1.010
	20	30	0.023	0.060	0.984	0.002	0.059	0.988	-0.026	0.059	1.003
0.2		60	0.011	0.042	0.997	-0.000	0.042	1.002	-0.014	0.041	1.007
		10	0.069	0.050	0.929	-0.005	0.048	0.980	-0.081	0.046	0.980
	100	30	0.021	0.027	0.984	0.000	0.027	1.007	-0.027	0.026	1.003
		60	0.010	0.019	0.994	-0.000	0.019	0.997	-0.014	0.019	1.004
		10	0.068	0.031	0.943	-0.005	0.030	0.991	-0.081	0.029	0.996
	250	30	0.021	0.017	0.989	-0.001	0.017	1.014	-0.028	0.017	1.007
		60	0.010	0.012	0.994	-0.000	0.012	1.001	-0.014	0.012	1.004
		10	0.092	0.273	0.897	0.026	0.263	0.983	-0.071	0.240	1.020
	5	30	0.028	0.147	0.941	0.010	0.146	0.984	-0.025	0.141	1.011
		60	0.014	0.101	0.963	0.004	0.100	1.005	-0.013	0.099	1.020
		10	0.068	0.134	0.900	0.003	0.129	0.984	-0.092	0.118	1.019
	20	30	0.020	0.074	0.931	0.002	0.073	0.978	-0.033	0.071	0.999
0.4		60	0.011	0.052	0.940	0.002	0.052	0.982	-0.016	0.051	0.995
		10	0.055	0.059	0.912	-0.009	0.057	0.995	-0.104	0.052	1.030
	100	30	0.018	0.033	0.933	0.000	0.033	0.989	-0.035	0.032	1.000
		60	0.009	0.023	0.949	-0.000	0.023	0.991	-0.018	0.022	1.004
		10	0.056	0.037	0.911	-0.009	0.036	0.990	-0.103	0.033	1.029
	250	30	0.017	0.021	0.942	-0.001	0.020	0.998	-0.036	0.020	1.010
		60	0.009	0.014	0.975	0.000	0.014	1.026	-0.018	0.014	1.032
		10	0.116	0.370	0.723	0.052	0.360	0.929	-0.035	0.301	1.012
	5	30	0.038	0.194	0.765	0.021	0.192	0.968	-0.013	0.180	1.001
		60	0.018	0.131	0.796	0.010	0.130	0.995	-0.008	0.126	1.020
		10	0.066	0.167	0.779	0.004	0.162	0.993	-0.076	0.137	1.074
	20	30	0.019	0.093	0.786	0.002	0.092	0.990	-0.031	0.087	1.023
0.6		60	0.009	0.064	0.805	0.001	0.064	1.013	-0.017	0.062	1.029
		10	0.051	0.073	0.791	-0.011	0.071	1.006	-0.088	0.060	1.088
	100	30	0.016	0.041	0.789	-0.001	0.041	0.996	-0.034	0.039	1.028
		60	0.008	0.029	0.791	-0.000	0.029	0.995	-0.018	0.028	1.012
		10	0.051	0.047	0.779	-0.010	0.045	0.989	-0.088	0.038	1.070
	250	30	0.016	0.026	0.800	-0.001	0.026	1.009	-0.034	0.024	1.041
		60	0.008	0.018	0.795	-0.000	0.018	1.001	-0.018	0.018	1.016

Table 4: Randomly-censored data: Empirical percentage coverage probabilities of two-sided 95% confidence intervals based on the profile (R) and integrated ($\bar{R}$) likelihoods, for different censoring probabilities. Lower and upper empirical non-coverage probabilities are given in brackets.

Prob. cens.	n	k	R	$\bar{R}$
0.2	5	10	90.1 (0.7, 9.2)	94.6 (2.5, 2.9)
		30	93.9 (0.8, 5.3)	95.5 (2.0, 2.5)
		60	93.7 (1.6, 4.7)	94.5 (2.7, 2.8)
	20	10	78.1 (0.1, 21.7)	94.9 (2.6, 2.5)
		30	90.0 (0.6, 9.4)	95.1 (2.6, 2.3)
		60	92.2 (0.9, 6.9)	95.1 (2.4, 2.5)
	100	10	27.5 (0.0, 72.5)	95.3 (2.4, 2.4)
		30	71.1 (0.0, 28.9)	95.0 (2.7, 2.3)
		60	83.6 (0.1, 16.3)	94.7 (2.6, 2.7)
	250	10	2.4 (0.0, 97.6)	95.0 (2.8, 2.2)
		30	40.1 (0.0, 59.9)	94.8 (3.0, 2.3)
		60	65.5 (0.0, 34.5)	95.2 (2.6, 2.2)
0.4	5	10	89.7 (0.5, 9.7)	94.9 (2.0, 3.0)
		30	93.1 (1.2, 5.7)	95.1 (2.3, 2.6)
		60	95.1 (1.2, 3.7)	95.6 (2.1, 2.2)
	20	10	79.9 (0.1, 20.0)	94.9 (2.4, 2.7)
		30	89.4 (0.5, 10.0)	95.1 (2.4, 2.6)
		60	92.9 (0.9, 6.3)	94.8 (2.7, 2.5)
	100	10	32.6 (0.0, 67.4)	95.6 (2.2, 2.3)
		30	72.4 (0.0, 27.6)	94.9 (2.5, 2.6)
		60	83.9 (0.1, 16.0)	94.9 (2.6, 2.4)
	250	10	3.6 (0.0, 96.4)	95.1 (2.7, 2.2)
		30	42.5 (0.0, 57.5)	95.2 (2.4, 2.4)
		60	67.6 (0.0, 32.4)	95.0 (2.3, 2.6)
0.6	5	10	89.7 (0.9, 9.5)	94.9 (2.2, 2.9)
		30	93.3 (1.1, 5.6)	95.0 (2.0, 3.0)
		60	94.2 (1.4, 4.4)	94.9 (2.4, 2.8)
	20	10	81.3 (0.3, 18.5)	94.5 (2.4, 3.1)
		30	91.1 (0.8, 8.1)	95.0 (2.7, 2.3)
		60	93.2 (1.0, 5.8)	95.2 (2.6, 2.2)
	100	10	40.2 (0.0, 59.8)	94.6 (2.9, 2.5)
		30	75.6 (0.0, 24.4)	95.1 (2.4, 2.5)
		60	86.0 (0.2, 13.8)	95.6 (1.9, 2.5)
	250	10	7.7 (0.0, 92.3)	94.9 (3.0, 2.1)
		30	48.5 (0.0, 51.5)	94.5 (2.4, 3.0)
		60	71.7 (0.0, 28.3)	95.0 (2.2, 2.8)

Table 5: Randomly-censored data: Estimates of the parameter of interest ψ obtained from profile and integrated likelihoods, for different censoring probabilities. Bias, simulation-based empirical standard errors (s.e.), and ratios between the average of estimated standard errors (e.s.e.) and the s.e. are provided.

Prob.				$\hat{\psi}$			$\bar{\psi}$	
cens	n	k	Bias	s.e.	e.s.e./s.e.	Bias	s.e.	e.s.e./s.e.
		10	0.145	0.223	0.896	0.029	0.205	0.980
	5	30	0.044	0.111	0.959	0.009	0.108	1.000
		60	0.022	0.078	0.955	0.005	0.077	0.984
		10	0.118	0.108	0.907	0.004	0.099	0.992
	20	30	0.035	0.055	0.961	0.001	0.054	1.002
0.2		60	0.018	0.039	0.959	0.001	0.038	0.988
		10	0.114	0.047	0.919	0.001	0.044	1.005
	100	30	0.034	0.025	0.965	-0.000	0.024	1.006
		60	0.017	0.017	0.974	0.000	0.017	1.004
		10	0.112	0.030	0.905	-0.001	0.028	0.990
	250	30	0.034	0.016	0.954	-0.000	0.015	0.995
		60	0.016	0.011	0.975	-0.000	0.011	1.005
		10	0.164	0.269	0.864	0.035	0.241	0.956
	5	30	0.052	0.130	0.950	0.014	0.126	0.981
		60	0.024	0.085	1.004	0.005	0.084	1.021
		10	0.132	0.126	0.901	0.007	0.113	0.993
	20	30	0.042	0.064	0.958	0.004	0.062	0.991
0.4		60	0.019	0.044	0.971	0.001	0.043	0.988
		10	0.124	0.055	0.915	0.001	0.050	1.011
	100	30	0.038	0.028	0.963	0.000	0.027	0.996
		60	0.018	0.019	0.987	0.000	0.019	1.004
		10	0.123	0.035	0.913	-0.001	0.031	1.007
	250	30	0.038	0.018	0.971	0.000	0.017	1.004
		60	0.018	0.012	0.991	0.000	0.012	1.009
		10	0.203	0.339	0.816	0.053	0.294	0.963
	5	30	0.059	0.159	0.910	0.016	0.152	0.978
		60	0.030	0.106	0.942	0.010	0.104	0.992
		10	0.151	0.159	0.838	0.011	0.139	0.982
	20	30	0.044	0.077	0.931	0.003	0.073	1.001
0.6		60	0.021	0.052	0.955	0.001	0.051	1.006
		10	0.139	0.069	0.853	0.001	0.061	0.998
	100	30	0.042	0.034	0.942	0.000	0.032	1.012
		60	0.021	0.023	0.967	0.001	0.023	1.017
		10	0.135	0.043	0.864	-0.002	0.038	1.010
	250	30	0.042	0.022	0.916	0.001	0.021	0.985
		60	0.020	0.015	0.948	0.000	0.014	0.997

Table 6: Randomly-censored data with model misspecification of the censoring distribution: Empirical percentage coverage probabilities of two-sided 95% confidence intervals based on the profile (R) and integrated ($\bar{R}$) likelihoods, for different censoring probabilities. Lower and upper empirical non-coverage probabilities are given in brackets.

Prob. cens.	n	k	R	$\bar{R}$
0.2	5	10	90.2 (0.6, 9.2)	94.6 (2.5, 2.9)
		60	94.3 (1.5, 4.2)	95.4 (2.3, 2.3)
	20	10	80.3 (0.1, 19.6)	95.2 (2.7, 2.1)
		60	92.9 (0.8, 6.3)	95.1 (2.5, 2.4)
	100	10	32.0 (0.0, 68.0)	94.7 (3.0, 2.3)
		60	84.3 (0.2, 15.5)	94.9 (2.8, 2.3)
	250	10	3.4 (0.0, 96.6)	94.7 (3.7, 1.7)
		60	67.4 (0.0, 32.5)	94.7 (2.9, 2.4)
0.4	5	10	90.8 (0.9, 8.3)	95.1 (2.5, 2.4)
		60	94.2 (1.5, 4.3)	94.8 (2.7, 2.5)
	20	10	84.0 (0.2, 15.8)	94.8 (3.1, 2.1)
		60	94.3 (0.9, 4.8)	95.3 (2.7, 1.9)
	100	10	48.6 (0.0, 51.4)	94.4 (4.6, 1.0)
		60	89.0 (0.5, 10.6)	94.3 (4.0, 1.7)
	250	10	12.8 (0.0, 87.2)	91.5 (7.9, 0.6)
		60	78.1 (0.1, 21.8)	94.4 (4.4, 1.2)
0.6	5	10	92.0 (1.4, 6.7)	95.3 (2.6, 2.0)
		60	94.6 (1.9, 3.5)	95.0 (2.6, 2.4)
	20	10	88.4 (0.6, 10.9)	95.1 (3.4, 1.5)
		60	94.1 (1.3, 4.6)	95.3 (2.6, 2.1)
	100	10	68.9 (0.0, 31.0)	92.6 (6.6, 0.8)
		60	91.1 (0.8, 8.2)	94.6 (3.9, 1.5)
	250	10	38.3 (0.0, 61.7)	88.6 (11.0, 0.4)
		60	86.1 (0.3, 13.6)	93.6 (5.2, 1.2)

Table 7: Randomly-censored data with model misspecification of the censoring distribution: Estimates of the parameter of interest ψ obtained from profile and integrated likelihoods, for different censoring probabilities. Bias, simulation-based empirical standard errors (s.e.), and ratios between the average of estimated standard errors (e.s.e.) and the s.e. are provided.

Prob.				$\hat{\psi}$			$\bar{\psi}$	
cens	n	k	Bias	s.e.	e.s.e./s.e.	Bias	s.e.	e.s.e./s.e.
	5	10	0.142	0.224	0.896	0.027	0.206	0.978
		60	0.019	0.077	0.973	0.003	0.076	1.001
	20	10	0.112	0.107	0.915	-0.001	0.099	0.998
0.2		60	0.017	0.038	0.971	-0.000	0.038	0.998
	100	10	0.110	0.049	0.898	-0.002	0.045	0.980
		60	0.016	0.017	0.965	-0.001	0.017	0.993
	250	10	0.109	0.030	0.914	-0.004	0.028	0.998
		60	0.016	0.011	0.948	-0.001	0.011	0.975
	5	10	0.144	0.267	0.903	0.018	0.240	0.971
		60	0.021	0.091	0.983	0.002	0.089	0.994
	20	10	0.111	0.128	0.918	-0.010	0.116	0.984
0.4		60	0.015	0.044	1.004	-0.003	0.044	1.016
	100	10	0.104	0.055	0.931	-0.016	0.050	1.027
		60	0.014	0.020	0.976	-0.004	0.020	0.997
	250	10	0.102	0.035	0.919	-0.018	0.032	1.013
		60	0.015	0.013	0.987	-0.003	0.012	1.007
	5	10	0.148	0.340	0.801	0.016	0.297	0.982
		60	0.020	0.114	0.991	0.001	0.111	1.007
	20	10	0.105	0.159	0.833	-0.018	0.140	1.012
0.6		60	0.016	0.056	0.913	-0.003	0.055	1.010
	100	10	0.094	0.070	0.923	-0.027	0.062	1.012
		60	0.014	0.025	0.907	-0.005	0.025	1.005
	250	10	0.092	0.044	0.847	-0.029	0.039	1.026
		60	0.013	0.016	0.898	-0.005	0.016	0.996

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it
Most of the working papers can also be found at the following url: http://wp.stat.unipd.it