Bias Corrected H-likelihood Approach for Joint Models of Longitudinal and Survival Data, With Application to Community Acquired Pneumonia

KARL STESSY BISSELOU, GLEB HAYNATZKI
Department of Biostatistics
University of Nebraska Medical Center
984375 Nebraska Medical Center, Omaha NE 68198
UNITED STATES

Abstract: - Time-to-event coupled with longitudinal trajectories are often of interest in biomedicine, and one popular approach to analysing such data is with a Joint Model (JM). JMs often have intractable marginal likelihoods, and one way to tackle this issue is by using the hierarchical likelihood (HL) estimation approach by Lee and Nelder [12]. The HL approximation sometimes results in biased estimates, and we propose a bias-correction approach (C-HL) that has been used for other models (eg, frailty models). We have applied, for the first time, the C-HL in the context of joint modelling of time-to-event and repeated measures data. Our C-HL method shows efficiency improvement, which comes at a cost of a more expensive computation than the existing HL approach. Additionally, we illustrate our method with a new MIMIC-IV CAP dataset.

Key-Words: H-likelihood; Joint models; Linear mixed effect models; Survival models; Random effects

Received: January 28, 2021. Revised: July 14, 2021. Accepted: August 4, 2021. Published: August 13, 2021.

1 Introduction

In community acquired pneumonia (CAP) studies, C-reactive protein (CRP) blood concentrations are often measured repeatedly over time and used to monitor inflammatory response to bacterial infection in hospitalized adult patients. When a person becomes infected by a bacteria responsible for pneumonia (e.g., Streptococcus pneumoniae), serum levels of CRP begin to increase rapidly stimulated by cytokines, such as interleukin (IL)-6.

The time to a patient's discharge is usually of primary interest; and frailty models are often used to measure cluster specific-risk prediction. Our research interest is to understand longitudinal trajectories of CRP response association with time to discharge or death using the hierarchical likelihood (H-likelihood) estimation approach Ha et al. [5-9, 12]. Three challenges pose in the analysis of such data; the marginal likelihood can have analytically intractable integrals as random effects get larger; the seemingly high censoring rate resulting in bias parameter estimates; and the interrelationship between CRP serum levels and the risk of death. These issues can only be properly accounted for in a joint model when the primary intent is to predict the 30-day mortality in adult CAP patients.

In this paper, we propose a bias correction approach for the H-likelihood estimators of a univariate joint modelling Ha et al., [9-10]. The H-likelihood, originally proposed in by Lee and Nelder, [12] resolves the issue of intractability of integrals in marginal inferences by treating the random effects as parameters. This prevents integrating out random effects over their respective distribution. However, in presence of high censoring rate, the H-likelihood tends to produce substantially bias parameter estimates Jeon et al. [10]. Furthermore, literature of the joint modelling cautions conducting separated analysis of longitudinal measurements and event times. Since these two data are observed directly from a same individual, they are likely correlated. Hence, ignoring this inherent association results in biased regression coefficients [4]. There are several methods capable to correct the all these bias; we use the regression calibration method of Wang et al., [15]. Finally, we conduct a simulation study to assess the performance of the proposed method, and we illustrate it using the Medical Information Mart for Intensive Care (MIMIC)-IV community acquired pneumonia dataset (2008-2019) [2]. This research applies, for the first time, the regression calibration approach to the H-likelihood estimator in the context of joint modelling. Additionally, we illustrate our method with a new MIMIC-IV CAP dataset.

2 Joint models formulation

It is intuitive to use joint modelling methods to understand longitudinal trajectories of CRP response association with the risk of one or multiple events such as onset of an infection, sepsis or death etc. In such study setting, an individual may experience multiple outcomes that may be associated [3, 9]. So-called "shared random intercepts" are widely used to capture the association structure between the longitudinal data and the time-to-event data [9, 14].

2.1 Modelling the longitudinal outcome

Let us denote by Y a random variable, and $\boldsymbol{y}_i(t_{ij})$ its observed value for each subject $(i = 1,\ldots, q)$ measured at time t_{ij}, $(j = 1,\ldots, n_i)$ where n_i is the number of measurement occasions (clusters).
$\boldsymbol{y}_i = \left(y_{i1}, \ldots, y_{in_i}\right)^T$ constitutes the $n_i \times 1$ vector of CRP response for patient i. Let us also consider $\boldsymbol{x}_{1i} = \left(x_{1i1}, \ldots, x_{1in_i}\right)^T$ the vector of covariates for $\boldsymbol{y}_i$ (possibly time-varying). It is common to use linear mixed effects (LME) models to fit the observed value for $\boldsymbol{y}_{ij}$ with respect to some covariates. This can be given by,

$$\begin{cases} \boldsymbol{y}_{ij} = x_{1i}^T\left(t_{ij}\right)\boldsymbol{\beta}_1 + z_{ij}^T(t_{ij})\boldsymbol{b}_i + e_{ij}, \\ \boldsymbol{b}_i \overset{iid}{\sim} N(\boldsymbol{0}, \mathrm{D});\ \boldsymbol{e}_{ij} \overset{iid}{\sim} N(\boldsymbol{0}, \sigma\mathbf{I}_q), \end{cases} \quad (1)$$

where $\boldsymbol{y}_{ij}$ is the immune response of subject i for the j-th measurement occasions, $\boldsymbol{x}_{1ij}$ and $\boldsymbol{z}_{ij}$ are vectors of covariates with mean and random effects parameters $\boldsymbol{\beta}_1 = (\beta_{11},\ldots, \beta_{1p_1})^T$ and $\boldsymbol{b}_i$, respectively. D denotes the between subjects unknown variance covariance matrix, σ is the variance parameter for the within-subject measurements, $\boldsymbol{e}_i = \left(e_{i1}, \ldots, e_{in_i}\right)^T$ is the vector of the random errors, $\mathbf{I}_q$ is the q-th dimension identity matrix, and iid refers to independent and identically distributed. Other covariance structures such as compound symmetric (CS), autoregressive (AR1) or unstructured (UN) can also be used. The best fit for a model's variance-covariance structure can be decided by estimating the Akaike information criterion (AIC) with the lowest value relative to models with other variance structures.

2.2 Modelling the time-to-event

The response profiles of a biomarker, namely C-reactive protein in adult patients with community acquired pneumonia, can be associated with the risk of death, since they are likely to be observed for a same individual. Though, the intrinsic nature of this association be may be complex in practice, we assume that the time to death depends on essential characteristic of the CPR trajectories.
Denote by T_i the random variable of time to death and $\boldsymbol{x}_2 = \left(x_{21}, \ldots, x_{2q}\right)^T$ the $q \times 1$ vector of covariates (possibly time-varying) for T_i. Denote also by $\boldsymbol{C}_i$, the non-informative right censoring time with respect to T_i, $t_i^* = \min(T_i, \boldsymbol{C}_i)$ and $\delta_i = I(T_i \leq \boldsymbol{C}_i)$, where $I(\cdot)$ is an event indicator function. In practice, censoring time are non-negligible in 30-day mortality for CAP studies. We fit the Cox model [3] for the observed the survival times $\{(t_i^*, \delta_i), i = 1,\ldots, q\}$, and can be given by,

$$\lambda_i(t|\boldsymbol{b}_i) = \lambda_0(t)\exp(x_{2i}^T\boldsymbol{\beta}_2 + z_{ij}\boldsymbol{b}_i^T\gamma), \quad (2)$$

where, $\lambda_0(t)$ is an unspecified baseline hazard function, $\boldsymbol{\beta}_2$ is a $p_2 \times 1$ vector of fixed effects, and γ is a real value parameter of association that links positively (negatively) the CPR trajectories to the risk of death. Thus γ has a critical implication for our joint model.

2.3 Joint modelling

The key assumptions below are essential for our joint model:
(i). $\boldsymbol{y}_i$ and $\boldsymbol{T}_i$ given $\boldsymbol{b}_i$ are conditionally independent
(ii). $\boldsymbol{T}_i$ and $\boldsymbol{C}_i$ given $\boldsymbol{b}_i$ are conditionally independent
(iii). $\boldsymbol{b}_i$ and $\boldsymbol{e}_{ij}$ are independent.
Thus, the basic joint model can be given by,

$$\begin{cases} \boldsymbol{y}_{ij} = x_{1ij}^T\left(t_{ij}\right)\boldsymbol{\beta}_1 + z_{ij}^T\left(t_{ij}\right)\boldsymbol{b}_i + e_{ij}, \\ \lambda_i(t|\boldsymbol{b}_i) = \lambda_0(t)\exp(x_{2i}^T\boldsymbol{\beta}_2 + z_{ij}\boldsymbol{b}_i^T\gamma) \\ \boldsymbol{b}_i \overset{iid}{\sim} N(\boldsymbol{0}, \mathrm{D});\ \boldsymbol{e}_{ij} \overset{iid}{\sim} N(\boldsymbol{0}, \sigma\mathbf{I}_q). \end{cases} \quad (3)$$

Now, let $\theta = (\boldsymbol{\beta}_1{}^T, \boldsymbol{\beta}_2{}^T, \boldsymbol{b}_i^T)^T$ be the collection of all fixed effects parameter in our joint model, and denote by $\xi = (\sigma, \gamma, \mathrm{D})$ the collection of all dispersion parameters. For all observed random variables $\boldsymbol{y}_i$, and (t_i^*, δ_i), the marginal joint modelling inference is given by,

$$L(\theta, \xi) = \prod_{i=1}^{q} \int f(y_i|\boldsymbol{\beta}_1, \boldsymbol{b}_i, \sigma) \times f(t_i^*, \delta_i|\Lambda_0, \boldsymbol{\beta}_2, \boldsymbol{b}_i, \boldsymbol{\gamma}) \times f(\boldsymbol{b}_i|\mathrm{D})d\boldsymbol{b}_i$$

where, $f(y_i|\boldsymbol{\beta}_1, \boldsymbol{b}_i, \sigma)$ is the conditional density in (1), $f(t_i^*, \delta_i|\Lambda_0, \boldsymbol{\beta}_2, \boldsymbol{b}_i, \gamma)$ is the survival function in (2) and $f(\boldsymbol{b}_i|\mathrm{D})$ is the density function for the random intercepts given D. The dimension of $\boldsymbol{b}_i$ can

be large and integrating it out can be computationally expensive or analytically intractable. The Fisher likelihood does not provide inference on $\boldsymbol{b}_i$, and evaluating $L(\boldsymbol{\theta}, \xi)$ using Bayesian MCMC yields a slow convergence.

2.4 Corrected H-likelihood approach

For simplicity, suppose $\eta_{1ij} = x_{1ij}^T\boldsymbol{\beta}_1 + \boldsymbol{b}_i$, and $\eta_{2i} = x_{2i}^T\boldsymbol{\beta}_2 + \gamma\boldsymbol{b}_i$. Then, the log h-likelihood is given by,

$$h = h(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \Lambda_0, \sigma, D|\boldsymbol{b}_i) = \sum_{ij}\ell_{1ij} + \sum_i \ell_{2i} + \sum_i \ell_{3i} \quad (4)$$

and,

$$\ell_{1ij} = \log f(y_i|\boldsymbol{\beta_1}, \boldsymbol{b}_i, \sigma) = -\frac{1}{2}\log(2\pi\sigma) - \frac{1}{2\sigma}\left(y_{ij} - \eta_{1ij}\right)^2,$$

$$\ell_{2i} = \log f(t_i^*, \delta_i|\boldsymbol{\beta_2}, \boldsymbol{b}_i, \Lambda_0) = \delta_i\left[\log\left(\lambda_0(t_i^*)\right) + \eta_{2i}\right] - \Lambda_0(t_i^*)\exp(\eta_{2i}),$$

$$\ell_{3i} = \log f(\boldsymbol{b}_i|\mathrm{D}) = \frac{1}{2}\log(2\pi D) - \frac{1}{2D}b_i^2,$$

are the conditional log-likelihoods for $\boldsymbol{y}_{ij}$, (t_i^*, δ_i) and $\boldsymbol{b_i}$, respectively.

Thus, $\exp\{-\Lambda_0(t_i^*)\exp(\eta_{2i})\} = S_i\left(t_i^*|\boldsymbol{b}_i, \boldsymbol{x}_{2i}^T\right)$ is the estimated survival function.

The maximum h-likelihood estimate can be obtained using iterative approximation method based on the score function (U), Hessian matrix (H) and Newton Raphson method, and can be given by,

$$\boldsymbol{U}(\theta) = \begin{pmatrix}\frac{\partial h}{\partial \boldsymbol{\beta}} \\ \frac{\partial h}{\partial \boldsymbol{b}}\end{pmatrix} = \begin{pmatrix}\boldsymbol{X}^T(\boldsymbol{Y} - \boldsymbol{\mu}) \\ \boldsymbol{Z}^T(\boldsymbol{Y} - \boldsymbol{\mu})\boldsymbol{\gamma} + \boldsymbol{\nabla}^{\boldsymbol{1}}_{\boldsymbol{\ell_b}}\end{pmatrix}, \quad (5)$$

$$\boldsymbol{H}(\theta) = \begin{pmatrix}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{Z} + \boldsymbol{\nabla}^{\boldsymbol{2}}_{\boldsymbol{\ell_b}}\end{pmatrix}, \quad (6)$$

where,

$$\boldsymbol{X} = \begin{pmatrix}\boldsymbol{X}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_2\end{pmatrix}, \boldsymbol{Z} = \begin{pmatrix}\boldsymbol{Z}_1 \\ \boldsymbol{Z}_2\end{pmatrix}, \boldsymbol{Y} = \begin{pmatrix}\boldsymbol{Y}_1 \\ \boldsymbol{Y}_2\end{pmatrix},$$

$$\boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2\end{pmatrix}, \boldsymbol{\gamma} = \begin{pmatrix}1 \\ \gamma\end{pmatrix}^T, \boldsymbol{W} = \begin{pmatrix}\boldsymbol{W}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \gamma^2\boldsymbol{W}_2\end{pmatrix},$$

$\boldsymbol{\nabla}^{\boldsymbol{1}}_{\boldsymbol{\ell_b}} = \partial\ell_{\boldsymbol{b}}/\partial\boldsymbol{b}$, and $\boldsymbol{\nabla}^{\boldsymbol{2}}_{\boldsymbol{\ell_b}} = -\partial^2\ell_{\boldsymbol{b}}/\partial\boldsymbol{b}^2$.

Following Ha et al., [5-6, 9], and Breslow [1], the non-parametric maximum H-likelihood estimator of the cumulative baseline hazard function can be given by,

$$\widehat{\Lambda}_{\boldsymbol{0}}(\beta_2, b_i) = \sum_{k:X_{(k)}\le t}\lambda_{0k} = \sum_{k:X_{(k)}\le t}\frac{d_{(k)}}{\sum_{i\in R(X_{(k)})}\exp(\eta_{2i})}$$

where, $X_{(k)}$ is the kth smallest distinct failure time of the $X_{(ij)}$'s, and $R\left(X_{(k)}\right) = \{ij: X_{ij} \ge X_{(k)}, i = 1,\ldots, q, j = 1,\ldots, n\}$ is the risk set at time $X_{(k)}$ and $d_{(k)}$ is the number of failures at $X_{(k)}$.

To estimate the mean parameter θ, we use the log-H profiled likelihood h_p by substituting the estimated baseline hazard $\Lambda_{\boldsymbol{0}}(\cdot) = \widehat{\Lambda}_{\boldsymbol{0}}(\cdot)$ in (4). After some algebra, we obtain the logH-profile likelihood

$$h_p \propto \sum_{ij}\ell_{1ij} + \sum_k d_{(k)}\log\frac{d_{(k)}}{\sum_{i\in R(X_{(k)})}\exp(\eta_{2i})} + \sum_i \delta_i\eta_{2i} - \sum_k d_{(k)} + \sum_i \ell_{3i}.$$

Then, the adjusted h-likelihood (h_A) is used to obtain estimates of the dispersion parameters ξ by solving the score equations $\partial h_A/\partial\xi = 0$ and using Newton Raphson approximation, respectively [12, 14]. The adjusted h-likelihood,

$$h_A = h|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}, \boldsymbol{u}=\widehat{\boldsymbol{u}}} - \frac{1}{2}\log\{\det(\boldsymbol{H})\}|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}, \boldsymbol{u}=\widehat{\boldsymbol{u}}} + \frac{(2(p+d)+1)}{2}\log 2\pi$$

where d is dispersion parameters, and p is the number of fixed effects.

2.5 Derivation of the score equations

Given $\theta = (\boldsymbol{\beta}_1{}^T, \boldsymbol{\beta}_2{}^T, \boldsymbol{b}_i^T)^T$ and $\xi = (\sigma, \gamma, \mathrm{D})$ the maximum h-likelihood estimators for θ are obtained by solving $\partial h_p/\partial\boldsymbol{\theta} = \boldsymbol{0}$. In other words,

$$\frac{\partial h_p}{\partial\boldsymbol{\beta_1}} = \sum_i\frac{\partial\ell_{1ij}}{\partial\boldsymbol{\beta_1}} = \frac{1}{\sigma}\sum_i x_{1ij}^T\left(y_{1ij} - \eta_{1ij}\right) = \frac{1}{\sigma}\boldsymbol{X}_1^T(\boldsymbol{y} - \boldsymbol{\eta}_1), \quad (7)$$

$$\frac{\partial h_p}{\partial\boldsymbol{\beta_2}} = \sum_i\frac{\partial\ell_{2i}}{\partial\boldsymbol{\beta_2}} = \sum_i x_{2i}^T\left(\delta_i - \eta_{2i}\right) = \boldsymbol{X}_2^T(\boldsymbol{\delta} - \boldsymbol{\mu_2}), \quad (8)$$

$$\frac{\partial h_p}{\partial\boldsymbol{b_i}} = \sum_i\frac{\partial\ell_{3i}}{\partial\boldsymbol{b}_i} = \frac{1}{\sigma}\boldsymbol{Z}_1^T(\boldsymbol{y} - \boldsymbol{\eta_1}) + \gamma\boldsymbol{Z}_2(\boldsymbol{\delta} - \boldsymbol{\mu_2}) - \boldsymbol{b}_i\mathrm{D}^{-1}, \quad (9)$$

where $\boldsymbol{\mu_2} = \exp(\log\Lambda_0(t^*) + \boldsymbol{\eta}_2)$. Z_1 is $n\times q$ cluster indicator matrix whose elements z_{ijk} are $\partial\eta_{1ij}/\partial b_i$, $(n = \sum_i n_i)$ and $Z_2 = \boldsymbol{I}_q = \boldsymbol{I}_{q\times q}$ is an identity matrix.

The maximum h-likelihood estimators $\widehat{\xi}$ of ξ can be derived using iterative methods and the diagonal of the Hessian matrix $\boldsymbol{H}$ and is given by,

$$\boldsymbol{H} = \begin{pmatrix}X_1^TW_1X_1 & \boldsymbol{0} & X_1^TW_1Z_1 \\ \boldsymbol{0} & X_2^TW_2X_2 & X_2^T(\gamma W_2)Z_2 \\ Z_1^TW_1X_1 & Z_2^T(\gamma W_2)X_2 & \boldsymbol{Z^TWZ} + \boldsymbol{\nabla}^{\boldsymbol{2}}_{\boldsymbol{\ell_b}}\end{pmatrix}$$

where, $W_1 = -\frac{\partial^2 h_p}{\partial\boldsymbol{\eta}_1\partial\boldsymbol{\eta}_1^T} = \sigma^{-1}I_q$,

$W_2 = -\frac{\partial^2 h_p}{\partial\boldsymbol{\eta}_2\partial\boldsymbol{\eta}_2^T}$, $\boldsymbol{\nabla}^{\boldsymbol{2}}_{\boldsymbol{\ell_b}} = -\frac{\partial^2 h_p}{\partial\boldsymbol{b}\partial\boldsymbol{b}^T} = D^{-1}I_q$,

$$\boldsymbol{W} = \begin{pmatrix} W_1 & \mathbf{0} \\ \mathbf{0} & \gamma^2 W_2 \end{pmatrix},$$

$$\boldsymbol{\nabla^1_{\ell_b}} = -D^{-1}\boldsymbol{b}, \boldsymbol{\gamma} = \begin{pmatrix} 1 \\ \gamma \end{pmatrix}^T \text{ and } \boldsymbol{\eta_1} = \boldsymbol{\mu_1}.$$

We then can update the score equation (5) and the Hessian matrix (6), where

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X_1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X_2} \end{pmatrix}, \boldsymbol{Z} = \begin{pmatrix} \boldsymbol{Z_1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z_2} \end{pmatrix}, \boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y_1} \\ \boldsymbol{Y_2} \end{pmatrix},$$

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix}, \boldsymbol{\gamma} = \begin{pmatrix} 1 \\ \gamma \end{pmatrix}^T, \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \end{pmatrix}.$$

We then take the partial derivative of the adjusted profile likelihood, h_A with respect to ξ,

$$\frac{\partial h_A}{\partial \xi} = \frac{\partial h_p}{\partial \xi}\Big|_{\widehat{\boldsymbol{\theta}}} - \frac{1}{2}\text{trace}\left(\boldsymbol{H}^{-1}\frac{\partial \boldsymbol{H}}{\partial \xi}\right)\Big|_{\widehat{\boldsymbol{\theta}}} - \frac{\partial}{\partial \xi}\left(\frac{1}{2}\log[\det(\boldsymbol{H})]\right)\Big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}, \boldsymbol{b}=\widehat{\boldsymbol{b}}, \lambda_0=\widehat{\lambda}_0} = \mathbf{0} \quad (10)$$

This leads to the partial restricted maximum likelihood estimate

$\hat{\sigma} = (\boldsymbol{y} - \widehat{\boldsymbol{\mu}}_1)^T(\boldsymbol{y} - \widehat{\boldsymbol{\mu}}_1)/(n - \omega_0)$, and

$\widehat{D} = \widehat{\boldsymbol{b}}^T\widehat{\boldsymbol{b}}/(q - \omega_1)$, where

$$\omega_0 = -\sigma \times \text{trace}\left(\boldsymbol{H}^{-1}\frac{\partial \boldsymbol{H}}{\partial \sigma}\right)\Big|_{\widehat{\boldsymbol{\theta}}},$$

$$\omega_1 = -D \times \text{trace}\left(\boldsymbol{H}^{-1}\frac{\partial \boldsymbol{H}}{\partial D}\right)\Big|_{\widehat{\boldsymbol{\theta}}}.$$

The estimate of γ is obtained through Newton-Raphson method. For the partial derivative of h_A with respect to γ,

$$\frac{\partial h_A}{\partial \gamma} = \frac{\partial h_p}{\partial \gamma}\Big|_{\widehat{\boldsymbol{\theta}}} - \frac{1}{2}\text{trace}\left(\boldsymbol{H}^{-1}\frac{\partial \boldsymbol{H}}{\partial \gamma}\right)\Big|_{\widehat{\boldsymbol{\theta}}} \quad ;$$

the first term of $\partial h_A/\partial\gamma$

$$\frac{\partial h_p}{\partial \gamma}\Big|_{\widehat{\boldsymbol{\theta}}} = (\boldsymbol{Zb})^T\boldsymbol{\delta} - \boldsymbol{d_{(k)}}(\boldsymbol{Zb})^T = (\boldsymbol{Zb})^T\left(\boldsymbol{\delta} - \boldsymbol{d_{(k)}}\right)\Big|_{\widehat{\boldsymbol{\theta}}},$$

and its the second term is,

$$\frac{\partial \boldsymbol{H}}{\partial \gamma} = \frac{\partial}{\partial \gamma}\begin{bmatrix} X_1^T W_1 X_1 & \mathbf{0} & X_1^T W_1 Z_1 \\ \mathbf{0} & X_2^T W_2 X_2 & X_2^T(\gamma W_2) Z_2 \\ Z_1^T W_1 X_1 & Z_2^T(\gamma W_2) X_2 & \boldsymbol{Z^T(W_1 + \gamma^2 W_2)Z + D^{-1}} \end{bmatrix}\Bigg|_{\widehat{\boldsymbol{\theta}}}$$

Since, $W_2 = -\frac{\partial^2 h_p}{\partial \boldsymbol{\eta}_2 \partial \boldsymbol{\eta}_2^T} = \boldsymbol{X_2^T \mu_2} X_2$, then,

$$\frac{\partial W_2}{\partial \gamma} = \boldsymbol{X_2^T \Lambda_0} \exp(\eta_2)\, X_2 \boldsymbol{Zb} = \boldsymbol{X_2^T \mu_2} X_2 \boldsymbol{Zb}$$

Now, we take the partial derivative of $\partial h_A/\partial\gamma$ with respect to γ to obtain the Hessian matrix $\boldsymbol{H_A}$

$$\boldsymbol{H_A} = \frac{\partial^2 h_A}{\partial \gamma^2} = \frac{\partial^2 h}{\partial \gamma^2}\Big|_{\hat{\boldsymbol{\tau}}} - \frac{1}{2}\frac{\partial}{\partial \gamma}\left(\text{trace}\left(\boldsymbol{H}^{-1}\frac{\partial \boldsymbol{H}}{\partial \gamma}\right)\right)\Big|_{\widehat{\boldsymbol{\theta}}},$$

$$\frac{\partial^2 h}{\partial \gamma^2}\Big|_{\hat{\boldsymbol{\tau}}} = (\boldsymbol{Zb})^T \boldsymbol{d_{(k)}} \boldsymbol{Zb}\Big|_{\widehat{\boldsymbol{\theta}}}.$$

Finally, estimates of ξ, and γ are obtained through iterative approximation of the score function and Newton-Raphson equation (11).

$$\widehat{\boldsymbol{\gamma}}^{(k+1)} = \widehat{\boldsymbol{\gamma}}^{(k)} + \left(\boldsymbol{H}^{-1}\boldsymbol{S}(\theta)\right)\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(k)}} \quad (11)$$

where, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{b})$, and $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{b}}\right)$.

2.6 Fitting procedure with bias correction

Following Ha et al., [9], the h-likelihood method can be extended and calibrated. Let $\boldsymbol{b} = \left(\boldsymbol{b}_1^T, \ldots, \boldsymbol{b}_q^T\right)^T$ and starting with initial values$\left(\xi^{(0)}, \theta^{(0)}, \hat{\lambda}_0^{(0)}\right)$, we iterate in j-steps as follows:

(S1) Set initial values $(\boldsymbol{\beta}_0, \boldsymbol{b}_0, \lambda_0, \xi_0)$.

(S2) Evaluate quantities $\boldsymbol{S}(\boldsymbol{\theta})$ and $\boldsymbol{H}(\boldsymbol{\theta})$ using equation (5) and (6), respectively.

(S3) Estimate $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}^{(j)}$, and $\boldsymbol{b} = \widehat{\boldsymbol{b}}^{(j)}$ using the Newton Raphson formula.

(S4) Update $\boldsymbol{H}^{(j)}$ using $\widehat{\boldsymbol{\beta}}^{(j)}$and $\widehat{\boldsymbol{b}}^{(j)}$ and estimate $\boldsymbol{H}_{22}^{(j)}$ the lower right corner using $\boldsymbol{H}^{-1(j)}$ matrix.

(S5) Estimate $\hat{\sigma}^{(j)}$ using results from (10), by replacing $\boldsymbol{b}$ with $E\left[\boldsymbol{b}\middle|\widehat{\boldsymbol{b}}^{(j)}\right] = \boldsymbol{\zeta}^{(j)}\widehat{\boldsymbol{b}}^{(j)} = \widehat{\boldsymbol{b}}^{c(j)}$, where, $\boldsymbol{\zeta} = \left(\zeta_1, \ldots, \zeta_q\right)$, $\widehat{\boldsymbol{b}}^{c(j)}$ is the corrected random effect at j^{th} iteration,

$\zeta_i^{(j)} = \frac{\hat{\sigma}^{(j-1)}}{\left(\hat{\sigma}^{(j-1)} + \boldsymbol{\tau}_i^{(j)}\right)}$, $\hat{\sigma}^{(j-1)}$ is the sample variance of $\hat{b}_1^{\ (j-1)}, \ldots, \hat{b}_q^{\ (j-1)}$, $\boldsymbol{\tau}_i^{(j)} = \boldsymbol{H}_{ii}^{(j)}$, $\boldsymbol{H}_{ii}$ is the i^{th} diagonal element of the lower right corner matrix of 2×2 $\boldsymbol{H}^{-1(j)}$ block matrix $\left((\boldsymbol{H}^{-1})_{22}^{(j)}\right)$ evaluated at the j^{th} iteration. By asymptotic properties of $\widehat{\boldsymbol{b}}$, $(\boldsymbol{H}^{-1})_{22} = (\boldsymbol{H}_{22})^{-1}$.

(S6) Update h by replacing $\boldsymbol{b}$ with $\widehat{\boldsymbol{b}}^{c(j)}$, and $\exp(\boldsymbol{b})$ with $E\left[\exp(\boldsymbol{b})\middle|\widehat{\boldsymbol{b}}^{(j)}\right] = \exp\left(\widehat{\boldsymbol{\zeta}}^{(j)}\widehat{\boldsymbol{b}}^{(j)} + \left(\hat{\sigma}^{(j)}\left(1 - \hat{\boldsymbol{\zeta}}^{(j)}\right)\right)/2\right)$.

(S7) Repeat steps S2 to S6 until the convergence criteria is met, which is defined as

$\max\left\{\left|\widehat{\boldsymbol{\beta}}^{(j+1)} - \widehat{\boldsymbol{\beta}}^{(j)}\right|, \left|\hat{\sigma}^{(j+1)} - \hat{\sigma}^{(j)}\right|\right\} < \varepsilon$,

where ε is a predetermined tolerance limit.

(S8) Given $\left(\boldsymbol{\theta}^{(j)}, \xi^{(j)}\right)$ we obtain estimates of the unspecified baseline hazard $\Lambda_0^{\ (j)}$,

$$\Lambda_0^{\ (j)}(t) = \sum_{i=1}^{q}\left(\frac{\delta_i I(t_i^* = t)}{\sum_{i=1}^{q}\exp\left(x_{2i}^T\boldsymbol{\beta}_2^{(j+1)} + \left(\boldsymbol{b}^{(j+1)}\right)^T\boldsymbol{\gamma}^{(j+1)}\right) I(t_i^* \geq t)}\right),$$

where $I(\cdot)$ is an indicator function.

3 Simulation studies

In this section, we conducted simulation studies, based on 500 replicates to evaluate the performance of the proposed bias corrected H-likelihood (C-HL) univariate joint model with shared random intercepts. We compared it with the H-likelihood (HL) approach [5, 6, 9], and the Gauss-Hermite quadrature (GHQ) approximation method. The *R* packages "**frailtyHL**" [7] and "**JM**" [13] were used to fit the HL and the GHQ (15 quadrature nodes), respectively. The initial values of the parameters used for the simulation below were the estimates obtained from fitting the joint model to the CAP dataset.

From the joint model in (3), we generated random subject intercepts $b_i \sim N(0, \mathrm{D} = 0.70)$ for i =1,…, 50. Next, we obtained y_{ij} for $n_i = 8$ (cluster) repeated measurements using the linear mixed model:

$y_{ij}|b_i \sim N(\beta_{10} + \beta_{11}\mathrm{Age} + \beta_{12}\mathrm{Time} + b_i, \sigma = 0.35)$, where $\beta_{10} = 1.50$, $\beta_{11} = 0.15$, $\beta_{12} = 0.15$. Age was a binary random variable generated from the Bernoulli distribution with probability 0.5 (where: 1 = 65 and older, 0 = less than 65), and Time = 0, 5, 10, 15, 20, 25, 30 and 35 (days).

Next, we used the model in (2) to generate event times for (t_i^*, δ_i),

$$\lambda_i(t|\boldsymbol{b}_i) = \lambda_0(t)\exp(\beta_{21}\mathrm{Age} + \gamma b_i),$$

where $\lambda_0(t)$ =1, $\beta_{21} = 0.60$ and $\gamma = 0.20$. Finally, censoring rates were generated from the exponential distribution resulting in three different cases of how heavy the data are censored: 32%, 58% and 86%; and the maximum follow-up time as equal to 35.

Results are summarized in Table 1. We calculated mean, standard deviation (SD) and Mean squared error (MSE) for 500 replicates to obtain $\widehat{\boldsymbol{\theta}}$ and $\hat{\xi}$. Results suggest that the original HL estimators $\widehat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\xi}}$ of $\boldsymbol{\theta} = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{21})^T$ and $\boldsymbol{\xi} = (\sigma, \gamma, \mathrm{D})$, respectively, are underestimated and this bias increases with censoring. However, the C-HL reduces this bias significantly as the censoring rate increases. When the number of quadrature nodes become larger, C-HL and GHQ estimates are closer (results not shown). However, the computation cost increases with the model complexity, the number of quadrature points and the censoring rate [13]. Our studies showed that for data with high censoring rate, C-HL is preferred to the HL, particularly for γ and β_{21} inferences. But C-HL becomes computationally more intensive than HL since the sample variance for the random intercepts are being recalculated; and this trend is maintained as the sample size increases.

4 Data analysis of CAP data

4.1 Data analysis results

In this section, we analyze the Medical Information Mart for Intensive Care (MIMIC)-IV publicly available community acquired pneumonia (CAP) dataset described in section 1, based on our proposed method.

Table 1: Simulation results for 500 replicates of data (number of subjects =50, number of measurement occasions =8). GHQ, Gauss-Hermite quadrature, HL, h-likelihood and C-HL corrected h-likelihood.

Cens. rate (%)	Par.	True	GHQ			HL			C-HL		
			Mean	SD	MSE	Mean	SD	MSE	Mean	SD	MSE
			LME for y								
32	β_{10}	1.50	1.482	0.118	0.018	1.479	0.120	0.020	1.481	0.122	0.021
	β_{11}	0.15	0.148	0.052	0.028	0.139	0.074	0.038	0.145	0.083	0.023
	β_{12}	0.15	0.145	0.034	0.009	0.144	0.019	0.039	0.146	0.091	0.012
	σ	0.35	0.349	0.099	0.010	0.352	0.101	0.023	0.350	0.127	0.015
	D	0.70	0.696	0.012	0.012	0.686	0.010	0.012	0.691	0.015	0.011
			Frailty Model for *T*								
	β_{21}	0.60	0.601	0.205	0.041	0.593	0.176	0.041	0.597	0.271	0.040
	γ	0.20	0.198	0.028	0.022	0.187	0.031	0.089	0.194	0.039	0.032
			LME for y								
58	β_{10}	1.50	1.495	0.147	0.044	1.399	0.159	0.040	1.502	0.182	0.041
	β_{11}	0.15	0.146	0.074	0.048	0.141	0.055	0.063	0.144	0.100	0.067
	β_{12}	0.15	0.139	0.045	0.011	0.143	0.039	0.032	0.149	0.097	0.009
	σ	0.35	0.333	0.118	0.041	0.372	0.091	0.043	0.336	0.130	0.045
	D	0.70	0.688	0.021	0.007	0.691	0.017	0.011	0.693	0.029	0.008
			Frailty Model for *T*								
	β_{21}	0.60	0.603	0.037	0.021	0.584	0.185	0.027	0.596	0.067	0.017
	γ	0.20	0.191	0.019	0.013	0.179	0.030	0.084	0.189	0.045	0.013
			LME for y								
86	β_{10}	1.50	1.501	0.205	0.065	1.414	0.220	0.123	1.498	0.227	0.067
	β_{11}	0.15	0.151	0.099	0.072	0.132	0.102	0.181	0.148	0.105	0.079
	β_{12}	0.15	0.148	0.058	0.014	0.144	0.061	0.019	0.146	0.095	0.022
	σ	0.35	0.347	0.100	0.006	0.312	0.103	0.020	0.351	0.139	0.007
	D	0.70	0.688	0.012	0.000	0.746	0.001	0.002	0.691	0.041	0.000
			Frailty Model for *T*								
	β_{21}	0.60	0.597	0.242	0.001	0.581	0.208	0.003	0.600	0.337	0.001
	γ	0.20	0.201	0.041	0.000	0.151	0.041	0.002	0.199	0.054	0.000

Par. = parameter, SD = standard deviation, MSE = mean squared error. Cens. =censoring

The dataset includes patients diagnosed with CAP (ICD-9 486) and admitted in ICU at the Beth Israel Deaconess Medical Center from 2008 to 2019. Our goal was to assess whether the longitudinal trajectories of CRP biomarker are associated with the risk of death in adult patients during a 30-day hospital stay. Each patient's electronic record was accessed and recorded based on the first admission time (baseline). This analysis may be limited because it is intended primarily for illustration of our method. Since CRP test is a cheap and readily available, it can be ordered daily during in-hospital duration. Thus CRP measurements fit the characteristic of a longitudinal immune response of patients with primary clinical outcome being the time to hospital discharge or death (Figure 1 and 2). The key feature of this CAP dataset is the high censoring rate due to patients' lost-to-follow up.

4.2 Data models

We used the *R* package "**nlme**" to fit linear mixed effect models and we obtained the Akaike Information Criterion (AIC) values to select the best longitudinal model for the data (Table 2). The model with the lowest AIC in boldface was selected for the

joint modelling analysis. The random intercept term was shared by all models formalized by equation (3). Finally, we used the Cox model with gender and age group as covariates to fit the event time model. We also conducted the analysis using the *R* package "**JM**" via GHQ maximum likelihood method with a piecewise constant baseline hazard. This was intended for a simple comparison since "**JM**" is a commonly used package for joint modelling analysis.

Table 2. AIC values of linear mixed models for $\log_{10}$ CRP

Models	Random Effect	AIC value
Time, gender, age	**Intercept**	**9067**
Time, gender	Intercept	9571
Time, age	Intercept	9074
Time	Intercept	9092

4.3 Data analysis results

Table 3 and 4 summarize all results and estimations based on the C-HL. Of 3469 medical records retrieved, 53% were females and 47% males. Age ranged from 18 to 89 with a median of 67 years. Patients were predominantly white (78%) and the median hospital stay was 7 days, and CRP levels measured as high as 397 mg/dL of blood serum. In total, 91.3% of the patients were censored.

Additionally, we focused on the inference on γ in the survival model (2), since it characterizes the association between the shared random intercepts in the longitudinal model (1) and the risk of death. We can see that both methods yield a positive association between CRP trajectories and the risk of death. However, this association is more pronounced in the C-HL method. Taken together, this suggests that the C-reactive protein is a non-specific biological marker and yet can somewhat predict the risk of death. Finally, consistent with CAP literature, we found that patients older than 65 years are at higher risk of death.

5 Conclusion

This paper proposed the regression calibration method of Wang et al., [15] to correct for the biased estimates in H-likelihood inference of the univariate joint model of longitudinal data and survival data in presence of high censoring. Our method shows efficiency improvement, which comes at a cost of a more expensive computation than the existing h-likelihood approach. Last but not least, our C-HL method yields the best results for joint modelling of longitudinal and survival data in the presence of very high censoring.

Table 3. Summary of demographic characteristics

Demographic variable	N	Mean (min, max) n (%)
Age	3469	67 (18, 89)
Age group	3469	
65+		1644 (47%)
<65		1825 (53%)
Gender	3469	
Male		1644 (47%)
Female		1825 (53%)
Ethnicity	3469	
White		2702 (78%)
Black		490 (14%)
Hispanic		165 (5%)
Asian		112 (3%)
Hospital Stay	3469	7 (1, 30)
CRP mg/dL	3469	98 (0.9, 397)
Log_{10}**CRP**	3469	1.73 (SD=0.76)
Deaths	3469	301 (8.68%)

min = minimum; max = maximum; n = number in category.

Table 4. CAP data analysis of the joint model under the C-HL and GHQ

Parameter	JM (GHQ)			JM (C-HL)		
	Est.	SE	*P*	Est.	SE	*P*
Longitudinal model						
Intercept	1.48	0.02	<0.001	1.61	0.22	<0.001
Time	0.13	0.003	<0.001	0.03	0.00	<0.001
Age, >65	0.16	0.02	<0.001	0.13	0.02	<0.001
Male	0.066	0.02	0.006	0.08	0.02	0.00
Event time model						
Age, >65	0.63	0.14	<0.001	0.62	0.14	0.00
Male	0.10	0.12	0.43	0.07	0.13	0.60
D	0.67	-	-	0.34	-	-
σ	0.40	-	-	0.17	-	-
γ	0.05	0.05	0.39	0.21	-	-

D = var(b_i) in the frailty model for death; Est. = point estimate; SE = standard error. γ = association parameter in frailty model; JM(C-HL), joint model under corrected h-likelihood; *P* = p-value; JM(GHQ), joint model under Gauss-Hermite quadrature.

Figure 1. All subject-specific trajectories of CRP measurements. The red line is marginal change.

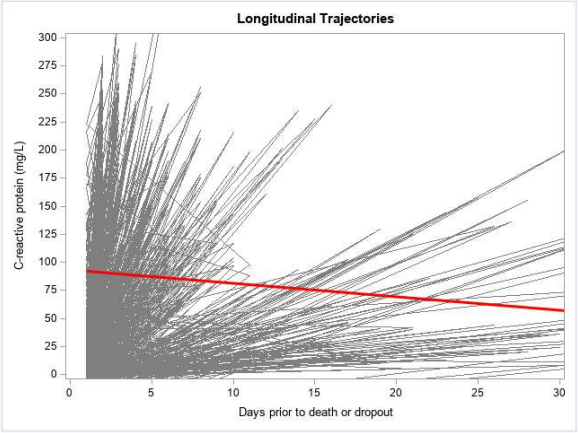
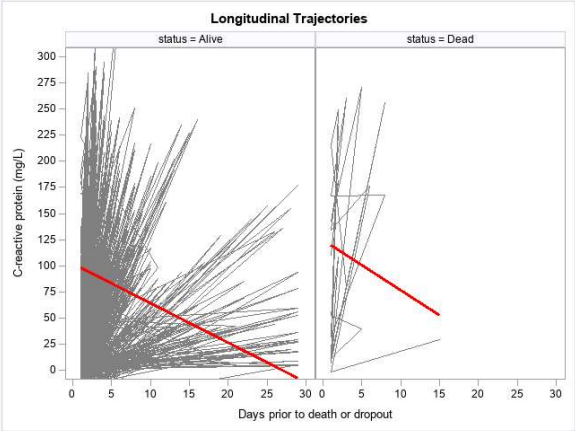


Figure 2. Subject-specific trajectories of CRP measurement, separately by survival status. The red line is the marginal change.



Conflict of interest

The authors declare that there is no conflict of interests regarding the publication of this paper

Data availability

Due to user agreement, the authors are unable to share the data used to support the findings of this study. However, the CAP dataset is available at https://physionet.org/content/mimiciv/1.0

References:

[1] Breslow, N. E., Discussion of Professor Cox's paper, *Journal of the Royal Statistical Society B*, Vol.34, 1972, pp.216-217.

[2] Coelho, L., Póvoa, P., Almeida, E. *et al.* Usefulness of C-reactive protein in monitoring the severe community-acquired pneumonia clinical course, *Critical Care*, Vol. 11, 2007.

[3] Cox, D. R., Regression models and life tables, *Journal of Royal Statistical Society*, Vol.34, 1972, p.187–220.

[4] Guo, X., Carlin, B., Separated and joint modeling of longitudinal and event time data using standard computing packages, *American Statistician*, Vol.58, 2004, p.16–24.

[5] Ha, I.D., Lee, Y., Song, J.K., Hierarchical likelihood approach for frailty models. *Biometrika*, Vol. 88, 2001, pp.233-243.

[6] Ha, I.D., and Lee, Y., Estimating frailty models via Poisson Hierarchical generalized linear models, *Journal of Computational and Graphical Statistics*, Vol. 12, 2003, pp.663–681.

[7] Ha, I.D., Noh, M., Lee, Y., frailtyHL: A package for Fitting Frailty Models with H-likelihood., *The R Journal*, Vol. 4, 2001.

[8] Ha, I.D., Jeong, J., Lee, Y., *Statistical Modelling of Survival Data with Random Effects: H-likelihood approach*, Springer, 2017.

[9] Ha, I.D., Noh, M., Lee, Y., H-likelihood approach for joint modeling of longitudinal outcomes and time-to-event data. *Biometrical journal. Biometrische Zeitschrift*, Vol. 59, 2017.

[10] Jeon, J., Hsu, L, Bias correction in the hierarchical likelihood approach to the analysis of multivariate survival data, *Biostatistics*, Vol.13, 2012, pp.384–397.

[11] Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, A., Mark, R. MIMIC-IV, *PhysioNet*, 2020. https//doi.org/10.13026/a3wn-hq05

[12] Lee Y. and Nelder J.A, Hierarchical generalized linear models (with discussion). *Journal Royal Statistical Society B*, Vol. 58, 1996, pp.619-678.

[13] Rizopoulos, D. JM: An R package for the joint modelling of Longitudinal and Time-to-event data, *Journal of Statistical software*, Vol.35, 2010.

[14] Rizopoulos, D. *Joint models for longitudinal and time-to-event data with applications in R*, Chapman and Hall, 2012.

[15] Wang, C., Hsu, L., Feng Z.D., Prentice, L.R., Regression calibration in failure time regression. *Biometrics* Vol.53, 1997, pp.131-145