

A Bayesian joint model for zero-inflated integers and left-truncated event times with a time-varying association: applications to senior health care

Xavier Piulachs*¹ | Eleni-Rosalina Andrinopoulou² | Montserrat Guillén¹ | Dimitris Rizopoulos²

¹Department of Econometrics, University of Barcelona, Barcelona, Spain

²Department of Biostatistics, Erasmus MC, Rotterdam, The Netherlands

Correspondence

*Xavier Piulachs, Department of Econometrics, University of Barcelona, Av. Diagonal 690, 08034 Barcelona, Spain.
Email: xavier.piulachs@ub.edu

Summary

Population aging in most industrialized societies has led to a dramatic increase in emergency medical demand among the elderly. In the context of private health care, an optimal allocation of the medical resources for seniors is commonly done by forecasting their life spans. Accounting for each subject's particularities is therefore indispensable, so the available data must be processed at an individual level. We use a large and unique dataset of insured parties aged 65 and older to appropriately relate the emergency care usage with mortality risk. Longitudinal and time-to-event processes are jointly modeled, and their underlying relationship can therefore be assessed. Such an application, however, requires some special features to also be considered. First, longitudinal demand for emergency services exhibits a non-negative integer response with an excess of zeros due to the very nature of the data. These subject-specific responses are handled by a zero-inflated version of the hierarchical negative binomial model. Second, event times must account for the left truncation derived from the fact that policyholders must reach the age of 65 before they may begin to be observed. Consequently, a delayed entry bias arises for those individuals entering the study after this age threshold. Third, and as the main challenge of our analysis, the association parameter between both processes is expected to be age-dependent, with an unspecified association structure. This is well-approximated through a flexible functional specification provided by penalized B-splines. The parameter estimation of the joint model is derived under a Bayesian scheme.

KEYWORDS:

zero inflation, left truncation, morbidity, age-related mortality hazard, shared-parameter joint model, time-varying association, private health care

1 | INTRODUCTION

1.1 | Medical motivation

Nearly all developed countries have experienced a gradual population aging process over the course of the past decades. This has led to a substantial increase in elderly individuals' demand for health care, along with the related financial challenges for health care providers. A reasonable strategy for optimal medical resource allocation consists of monitoring each individual's

service demand from the age of 65 onwards, which in turn reinforces affiliates' loyalty.¹ For this purpose, we need precise data on individuals' health status according to their age, and we also need to assess the role of other demographic risk factors affecting the propensity to seek medical attention, such as gender and socio-economic group. From a health care insurance provider's standpoint, monitoring the use of critical care services becomes crucial; this is the primary source of information on the deterioration of senior affiliates' health, and at the same time it represents the highest costs for the private sector. Since it is common for health care policies to renew annually, the overall risk assessment period can be properly divided into intervals of one calendar year. Then, individually analyzing critical care may provide a useful surrogate indicator or proxy variable when evaluating the subject's instantaneous mortality hazard. When considering these annually-recorded medical responses, the variability in a subject's health status within a specific calendar year is unknown. To deal with this missing information, the time-dependent Cox model assumes that the level of critical care usage over the past calendar year is directly imputed to the following one, along which it remains constant. However, such an unrealistic hypothesis usually leads to severely biased results.² Instead, a more logical smooth evolution over time of the annual emergency demand may be considered by incorporating a suitable model with time-independent random effects. In this manner, longitudinal and mortality hazard processes are jointly modeled by assuming that the random terms are shared by both submodels; see Henderson et al.³ and references therein. For complete overviews in this research field, the reader is referred to Tsiatis and Davidian⁴ and Rizopoulos⁵. On the other hand, as a key feature of our postulated shared-parameter (SP) joint model, we assess the presumed relationship between both outcomes through a time-varying association. This ultimately quantifies the impact of the expected critical care demand (together with other secondary factors) on the age-adjusted mortality rate for insured elderly individuals. In summary, measuring the health characteristics of elderly populations helps identify high-risk individuals⁶, and opens new avenues for health providers to hedge against subject-specific longevity risk beyond the age of 65.

Our research is illustrated by a unique and relatively large dataset from the Spanish health insurance sector, henceforth HI data⁷, which provides a survey of emergency care demand for 5470 elderly policyholders. All individuals are residents of the city of Barcelona and are between 65 and 99.5 years old at the initial observation, with a median age of 74.9. The demand for emergency services is longitudinally recorded in terms of claim counts at the end of each calendar year throughout the entire study window, from January 1, 2006 to February 1, 2014. The monitoring can be considered sufficiently informative, with roughly six measurements per subject on average, along with a mean and median follow-up period of 5.1 and 6.3 years, respectively. In conjunction with critical care demand, gender is presumed to be an important confounding factor in understanding age-related mortality hazard patterns. Out of all subjects, 3415 (62.4%) are female, who historically have longer life expectancies than male. In addition, to identify the potential risk effect of the socio-economic profile, the postal code is used to determine which policyholders reside in neighborhoods with a mean income above the overall city average; this group consists of 911 individuals (16.7%). The primary longitudinal response is the annual number of reported claims for ambulance services, hospitalizations, and non-routine visits; these counts are individually recorded in an aggregated manner by each calendar year within the observation period. Subjects' entry into the study is registered at their first longitudinal response, after which subsequent responses are yearly recorded. Because individual follow-up does not necessarily start at the beginning of a calendar year, the exposure-time effect is considered to correct the first measurement's frequency. Table 1 displays the distribution of the 32269 longitudinal responses collected over the study period and nested within individuals, where zeros represent more than half of overall measurements. This large number of zeros is due to a) the highly infrequent nature of the data itself, and b) the existing interplay in the Spanish health care system between universal coverage, which provides social security services without cost to the individual, and private insurance, which represents an optional extra within the publicly-funded universal system. This latter point entails a group of measurements outside of the private health sector, which systematically supply null values in our data.

TABLE 1 Annual count rates for emergency medical claims across all observations collected over time in the HI data and nested within the 5470 senior policyholders.

	Count rate											Total
	0	1	2	3	4	5	6	7	8	9	≥ 10	
Number	20357	5580	2849	1488	756	487	280	145	95	78	154	32269
Percentage	63.1	17.3	8.8	4.6	2.4	1.5	0.9	0.4	0.3	0.2	0.5	100

The end of individual follow-up can be due either to death or to non-informative right censoring; dead subjects have a share of 9.3%. Since only the trajectories of policyholders who are at least 65 are monitored, all those not living long enough to reach this fixed truncation age cannot be observed. Life spans are therefore affected by left truncation, and the staggered entry of 4365 individuals (79.8%) into the study after 65 must be incorporated as delayed entries in order to avoid overestimating their event times. These are properly considered by adopting the subject's age above 65 as the time scale, in which the difference between the age of entry and 65 is implicitly considered: $t = \text{age.entry} - 65$. Hence, our particular time zero is set at the age of 65, and from then onwards, any time point $t > 0$ is inextricably linked with the corresponding policyholder's age. This duality makes it equivalent to talk either in terms of time or in terms of age; for instance, a policyholder's response at age 75 is recorded at time $t = 10$ years. In our empirical data, the within-subject observations may arise at any time point between $t_{\min} = 0$ and $t_{\max} = 37.5$, since 102.5 years old is the maximum age at which observations are collected for any subject. Interestingly, these delayed entries result in the progressive incorporation of policyholders after the study has already begun. This entails an age-related dynamic fluctuation of the number of subjects at risk, as illustrated in the left panel of Figure 1. As expected, the number of insured women at risk is typically higher than insured men at all ages, even though this gap in the at-risk pattern by gender tends to narrow with increasing age. It is also relevant to compare the average emergency demand patterns between event-free profiles and trajectories for which death is recorded, as displayed in the right panel of Figure 1. Both of these patterns change across the above-65 age range, and there appears to be strong linkage between dead subjects and greater demand rates; a yearly average of 0.80 emergency claims is recorded among subjects who do not die during the study period, whereas this increases to 1.50 for those who experience the event. Further details about the information provided by the HI data can be found in Appendix A (Tables W1-W5 and Figure W1).

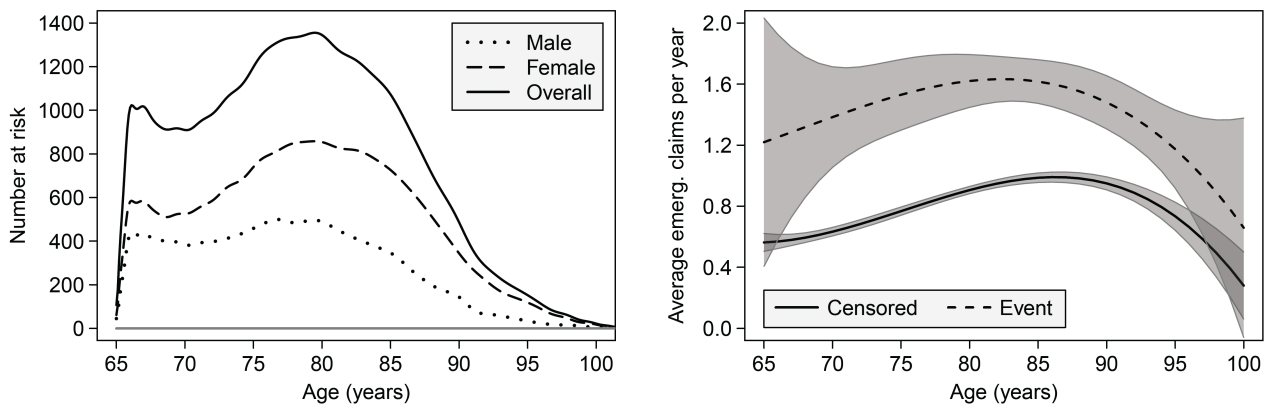


FIGURE 1 Age-specific descriptive plots from the time-to-event analysis of the HI data. Left: Dynamic variation of the number of subjects at risk, overall and by gender. Right: Average trends for annual emergency demand with 95% confidence regions, separately for event-free subjects and those who die during the study.

1.2 | Joint modeling techniques adapted to health care sector

As stated above, we work within the scope of SP joint models for longitudinal and time-to-event data. Recent advances in joint modeling have been made to address different types of non-Gaussian longitudinal responses, including new approaches to better account for a high frequency of zeros with respect to a Poisson process.^{8,9,10} Furthermore, other novel extensions to accommodate left-truncated event times within the scope of SP joint modeling have been also examined.^{11,12} However, none of the previous studies have combined longitudinal counts exhibiting both excess zeros and marked heterogeneity with left-truncated event times. Simultaneously, and as the main challenge of this article, we aim to appropriately measure the relationship between the need for intensive care and the corresponding mortality hazard. The functional shape that interrelates both processes is completely unknown, and so a flexible time-varying association structure is hypothesized. Song and Wang¹³ proposed a joint model in which time-varying regression coefficients are assumed for the hazard process. More recently, Barrett et al.¹⁴ postulated a joint model approach in which the association between survival and random effects can also vary over a discrete set of potential time points. Köhler et al.¹⁵ focused on a flexible Bayesian additive joint model capable of integrating strongly non-linear individual trends and time-varying effects in the association parameter via a penalized B-spline based approach. They summarized their developments by means of the R package `bamlss`.¹⁶ Also within the Bayesian framework, Andrinopoulou

et al.¹⁷ improved the classical SP joint model scheme by expanding the association parameter into penalized B-spline basis functions with quadratic splines of time, and illustrated the benefits in the context of dynamic predictions. Building on their contribution, we propose a joint model to associate zero-inflated discrete responses with left-truncated event times, while also using a time-varying association approximated by penalized cubic B-splines. This modeling framework allows us to incorporate the particular features necessary when analyzing the HI data and provides the tools to properly manage applications beyond the insurance domain.

To appropriately associate the observed emergency claims per year and mortality risk in our empirical data, the main challenges tackled in our analysis are threefold: 1) developing an SP joint model to accommodate correlated count rates with an overabundance of zeros and left-truncated event times, 2) incorporating a flexible time-varying association structure between the longitudinal and time-to-event processes through a penalized B-spline approach with cubic splines, and 3) demonstrating the validity of the proposed model, both empirically and numerically. The article is developed under a Bayesian paradigm by using Markov chain Monte Carlo (MCMC) algorithms. This approach allows for a clearer and more straightforward coding implementation for the referred tasks in comparison to the maximum likelihood approach. In addition, a full posterior inference for any parameter can be obtained, as well as a wide range of diagnostic tools for checking the model fit. These benefits, however, usually go hand-in-hand with a high computational burden derived from using MCMC sampling, especially for sizable datasets.

The remainder of the paper proceeds as follows. Section 2 specifies the Bayesian formulation for jointly tackling the measurements of longitudinal emergency care demand and the time to death. Section 3 describes the estimation procedure under the Bayesian framework. Section 4 conducts an extensive simulation study to assess the performance of the proposed joint model. Section 5 illustrates the application of our model to the motivating dataset. Section 6 describes the model assessment through posterior predictive checking and residual diagnostics. Lastly, Section 7 provides a discussion of the main results and suggests some topics for future research.

2 | THE PROPOSED JOINT MODEL

2.1 | Longitudinal submodel for zero-inflated count rates

Denote $\mathbf{y}_i = \{y_i(t_{ij}), i = 1, \dots, n\}$ as the observed response vector for the i -th subject, recorded at a sequence of time points t_{ij} , $j = 1, \dots, n_i$. For convenience, suppose that every longitudinal count rate can be formally expressed as $y_i(t_{ij}) = v_i(t_{ij}) y_{Ci}(t_{ij})$; here $v_i(t) \sim \text{Bernoulli}\{\pi_{Bi}(t)\}$ is a latent indicator that captures subject's probability $\pi_{Bi}(t)$ to use private health care at a specific time t , while $y_{Ci}(t)$ is the count response derived from an adequately chosen model $f_C(\cdot)$ such that $\mathbb{E}\{y_{Ci}(t) | \mathbf{b}_{iC}\} = \mu_{Ci}(t)$ is its mean response, conditional on the random effects \mathbf{b}_{iC} . Specifically, $v_i(t) = 1$ if the longitudinal measurement corresponds to a senior policyholder who uses private services within a specific time interval. In this case, $y_{Ci}(t)$ comes from $f_C(\cdot)$ and may logically include zeros if the subject's health status is good enough. In contrast, if $v_i(t) = 0$, then $y_i(t) = 0$ systematically. These two zero-generation sources affecting longitudinal count rates are combined into a zero-inflated response as

$$y_i(t) \sim \begin{cases} 0 & \text{if } v_i(t) = 0, \text{ with probability } 1 - \pi_{Bi}(t) \\ f_C\{y_{Ci}(t) | \mathbf{b}_{iC}\} & \text{if } v_i(t) = 1, \text{ with probability } \pi_{Bi}(t) \end{cases} \quad (1)$$

with probability mass function

$$p_y\{y_i(t) | \mathbf{b}_i\} = \{1 - \pi_{Bi}(t)\} I\{v_i(t) = 0\} + \pi_{Bi}(t) f_C\{y_{Ci}(t) | \mathbf{b}_{iC}\} I\{v_i(t) = 1\}, \quad (2)$$

where $I(\cdot)$ is the indicator function and $\mathbf{b}_i = (\mathbf{b}_{iB}^\top, \mathbf{b}_{iC}^\top)^\top$ encompasses the common and unobserved subject-level random effects for the binary and count parts, respectively. The corresponding overall conditional mean and variance expressions are

$$\mathbb{E}\{y_i(t) | \mathbf{b}_i\} = \pi_{Bi}(t) \mu_{Ci}(t), \quad \mathbb{V}\{y_i(t) | \mathbf{b}_i\} = \pi_{Bi}(t) \mu_{Ci}(t) + \{1 - \pi_{Bi}(t)\} \pi_{Bi}(t) \mu_{Ci}^2(t). \quad (3)$$

Apart from the excess of zeros, the discrete measurements $y_{Ci}(t)$ can also be affected by a marked level of patchiness among subject-specific responses, so the conditional variance exceeds the conditional mean. The hierarchical approach to the standard negative binomial (NB) model, commonly termed as NB2 model, arises as the traditional modeling option for $f_C(\cdot)$, since this model provides a second-degree dependence of the variance upon the mean. This quadratic relationship, however, can be relaxed by assuming a NB-Power (NBP) distribution^{18,19}:

$$f_C\{y_{Ci}(t) \mid \mathbf{b}_{iC}\} = \frac{\Gamma\{y_{Ci}(t) + \kappa \mu_{Ci}^{2-P}(t)\}}{\Gamma\{\kappa \mu_{Ci}^{2-P}(t)\} y_{Ci}(t)!} \left\{ \frac{\kappa \mu_{Ci}^{2-P}(t)}{\mu_{Ci}(t) + \kappa \mu_{Ci}^{2-P}(t)} \right\}^{\kappa \mu_{Ci}^{2-P}(t)} \left\{ \frac{\mu_{Ci}(t)}{\mu_{Ci}(t) + \kappa \mu_{Ci}^{2-P}(t)} \right\}^{y_{Ci}(t)}, \quad (4)$$

$$\mathbb{V}\{y_{Ci}(t) \mid \mathbf{b}_{iC}\} = \mu_{Ci}(t) + \mu_{Ci}^P(t)/\kappa.$$

Above, the heterogeneous nature of $y_{Ci}(t)$ is explicitly accommodated through a common shape parameter $\kappa > 0$ for all subjects, whereas the ancillary parameter $P > 0$ allows for a P -th degree mean-variance relationship. The combination of (2) and (4) yields a hierarchical zero-inflated NBP (ZINBP) model, $y_i(t) \sim \text{ZINBP}\{\pi_{Bi}(t), \mu_{Ci}(t), \kappa, P\}$. From this, two common hierarchical models for discrete data with excess zeros can be straightforwardly derived: the zero-inflated NB (ZINB2) longitudinal model is obtained by setting $P = 2$, whereas the zero-inflated Poisson (ZIP) longitudinal model is considered if $\kappa \rightarrow \infty$ for all P .

On the other hand, both $\pi_{Bi}(t)$ and $\mu_{Ci}(t)$ can be respectively related to linear mixed predictors, $\eta_{Bi}(t)$ and $\eta_{Ci}(t)$, through prescribed link functions. Specifically, let

$$\eta_{Bi}(t) = \mathbf{x}_{Bi}^\top(t)\boldsymbol{\beta}_B + \mathbf{z}_{Bi}^\top(t)\mathbf{b}_{iB}, \quad \eta_{Ci}(t) = \mathbf{x}_{Ci}^\top(t)\boldsymbol{\beta}_C + \mathbf{z}_{Ci}^\top(t)\mathbf{b}_{iC}, \quad (5)$$

where $\mathbf{x}_{Bi}^\top(t)$ and $\mathbf{z}_{Bi}^\top(t)$ are the i -th subject's rows of the fixed- and random-effects design matrices in the binary part, respectively, while $\mathbf{x}_{Ci}^\top(t)$ and $\mathbf{z}_{Ci}^\top(t)$ are the corresponding rows of analogous matrices in the count part. These matrices can be defined to summarize the subject-specific evolutions using simple regression patterns, such as linear trends, or to approximate these trajectories through more flexible functional forms, such as splines. Moreover, $\boldsymbol{\beta}_B = (\beta_{B_0}, \beta_{B_1}, \dots, \beta_{B_{p_B}})^\top$ and $\mathbf{b}_{iB} = (b_{iB_0}, b_{iB_1}, \dots, b_{iB_{q_B}})^\top$ denote the $p_B + 1$ fixed and $q_B + 1$ random effects for modeling the binary response, while $\boldsymbol{\beta}_C = (\beta_{C_0}, \beta_{C_1}, \dots, \beta_{C_{p_C}})^\top$ and $\mathbf{b}_{iC} = (b_{iC_0}, b_{iC_1}, \dots, b_{iC_{q_C}})^\top$ include the $p_C + 1$ fixed and $q_C + 1$ random effects for modeling the count rates. This yields the two-part model

$$\text{logit}[\text{Pr}\{v_i(t) = 1 \mid \mathbf{b}_{iB}\}] = \text{logit}\{\pi_{Bi}(t)\} = \eta_{Bi}(t), \quad \log[\mathbb{E}\{y_{Ci}(t) \mid \mathbf{b}_{iC}\}] = \log\{\mu_{Ci}(t)\} = \log\{e_i(t)\} + \eta_{Ci}(t). \quad (6)$$

Here, $e_i(t)$ acts as an offset term which accounts for the existence of an exposure-time effect related to a specific subject interval. Given that count rates are from 1-year periods in our analysis, we have $e_i(t) = 1$ except for the first longitudinal response of those subjects whose trajectory does not start to be observed at the beginning of a calendar year; for these responses, $e_i(t)$ ranges from 0.5 to 1. Moreover, the distribution for the random effects is established to be $\mathbf{b}_i \sim N_q(\mathbf{0}, \mathbf{D})$, $q = q_B + q_C + 2$, with an unspecified structure for the $q \times q$ covariance matrix \mathbf{D} . The latter can be decomposed into the symmetric block matrix

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_B & \mathbf{D}_{BC} \\ \mathbf{D}_{BC} & \mathbf{D}_C \end{bmatrix} \quad (7)$$

where \mathbf{D}_B and \mathbf{D}_C are the $(q_B + 1) \times (q_B + 1)$ and $(q_C + 1) \times (q_C + 1)$ covariance submatrices for \mathbf{b}_{iB} and \mathbf{b}_{iC} , respectively, whereas \mathbf{D}_{BC} is the $(q_B + 1) \times (q_C + 1)$ cross-covariance submatrix that includes the prior correlations between \mathbf{b}_{iB} and \mathbf{b}_{iC} . Specifically, the diagonal terms of \mathbf{D} correspond to the variances of random effects, σ_{kk}^2 , for $k = 1, \dots, q$, whereas the off-diagonal elements denote the corresponding covariances, $\rho_{k\ell} \sigma_k \sigma_\ell$, for $k, \ell = 1, \dots, q$. By combining the overall conditional expectation from (3) with the two equations of (5), we have the globally-expected longitudinal response at time t :

$$\mu_i(t) = e_i(t) \exp\{\eta_{Bi}(t) + \eta_{Ci}(t)\} / [1 + \exp\{\eta_{Bi}(t)\}]. \quad (8)$$

2.2 | Joint modeling with a time-varying association

2.2.1 | Structural features

Let age beyond 65 be the time scale of the study, and let the non-negative random variable T_i^* denote the i -th true event time. Time-to-event data may be subject to a random left truncation process, denoted by the variable $L_i = \text{age.entry}_i - 65 \geq 0$, as well as to random right censoring, denoted by the variable C_i . Suppose further that $\{L_i, T_i^*, C_i\}$ are conditionally independent given the covariate history, and also suppose that L_i and C_i are not related to the subject's death event. For those individuals with $T_i^* > L_i$, we observe both the event time $T_i = \min\{T_i^*, C_i\}$ and the random event indicator $\Delta_i = I(T_i^* \leq C_i)$. In contrast, individuals with $T_i^* \leq L_i$ are not observed since they either do not reach 65 within the study window or pass away

before reaching this age. Our primary research interest lies in assessing the time-dependent relationship between the expected zero-inflated response $\mu_i(t)$, from the longitudinal process, and the hazard for death at time t , from the event history process. These two components are jointly modeled using a shared vector \mathbf{b}_i of time-independent random effects, while also relating both processes with an unknown functional shape whose trajectory varies over the course of time. Our proposed SP joint model with a time-varying association parameter (JMTV-ZINBP) can be expressed as the instantaneous hazard rate

$$h_i \{t \mid \mathcal{M}_i(t), \mathbf{w}_i(t)\} = h_0(t) \exp\{\boldsymbol{\gamma}_w^\top \mathbf{w}_i(t) + \alpha(t) \mu_i(t)\}, \quad t > 0, \quad (9)$$

which is conditioned on: the truncation process, the underlying history $\mathcal{M}_i(t) = \{\mu_i(s) = \pi_{Bi}(s) \mu_{Ci}(s), L_i \leq s \leq t\}$, and a vector $\mathbf{w}_i(t)$ of exogenous, possibly time-varying, covariates, with related regression parameters $\boldsymbol{\gamma}_w$. Additionally, $h_0(\cdot)$ is usually an arbitrary and unspecified function to describe the baseline hazard at the population level, while $\alpha(\cdot)$ denotes some flexible function of time that assesses the effect of the longitudinal outcome on the conditional hazard rate. Thus the quantity $\exp\{\alpha(t)\}$ evaluates the hazard ratio at time point t for a one-unit increase in the value of $\mu_i(t)$, while keeping the rest of variables constant.

A typical major concern when dealing with JM techniques is the specification of the unknown baseline hazard as a function of age. This masks underlying mortality risks in the target population, which may vary gradually over time and ultimately determine the distribution of the event times. Consequently, even though in our analysis the original experience data is aggregated over 1-year calendar periods, survivorship needs to be continuously modeled over the entire individual's lifetime. This guarantees that a subject's conditional hazard rate will reasonably and smoothly rise with increased age. Notice that taking natural logarithms in (9) directly leads to a numerically advantageous linear relationship with time, and so any approximation of the baseline hazard function is usually performed with its logarithmic version, $\log\{h_0(\cdot)\}$. A first alternative for mimicking the baseline hazard consists of approximating its true functional shape through a predefined survival law. The use of a parametric function of age confers two major advantages: the model itself automatically imposes smoothness across the age-related mortality hazard trend, and only a small number of parameters need to be estimated.²⁰ The most basic law to describe mortality trends in human populations is the Gompertz survival model, in which a continuous exponential increase in death risk over time is hypothesized. Although other more general forms are also extensively considered in the literature²¹, for illustrative purposes we set the Gompertz distribution as the default parametric law for describing the baseline hazard in our empirical data. A common parameterization is $h_0(t) = \lambda_0 \exp(\sigma_i t)$, where $\lambda_0 > 0$ and $\sigma_i \geq 0$ represent the initial mortality intensity and the annual proportional growth rate in mortality, respectively. Caution is advisable if λ_0 is close to zero, due to the fact that numerical instabilities are likely to affect this parameter estimation. This can be circumvented by setting $\lambda_0 = \exp(\lambda)$, so that $h_0(t) = \exp(\lambda + \sigma_i t)$. Taking natural logarithms of this equation does not merely linearize it, but also provides an appealing direct relationship between the logarithm of the baseline hazard and the parameters themselves, $\log\{h_0(\cdot)\} = \lambda + \sigma_i t$.

Despite the benefits of using a suitable law to estimate the baseline hazard's functional shape, previous knowledge of risk behavior is required from the researcher when handling complex data. Otherwise, using a predefined law which relies on few parameters could seriously hinder the analysis. An alternative and more flexible option to model this underlying risk is the use of penalized B-spline basis functions, called P-splines²⁴, requiring no precedent for handling this kind of data and allowing as much smoothness as desired. Essentially, the most appealing feature of such a generalized proposal is that the underlying functional shape behind $\log\{h_0(\cdot)\}$, along with that of $\alpha(\cdot)$, are properly approximated through expanding each of these time-dependent functions into P-splines, with respective degrees d_{h_0} and d_α . For these two P-spline regressions, the corresponding B-spline basis can be immediately calculated as differences of truncated power functions, as suggested by Dierckx;²⁵ we also refer to Eilers and Marx²⁶ for computing details and some examples discussed herein. To control smoothness, one can impose discrete penalties on finite differences of the estimated coefficients from adjacent B-splines. Using this strategy, assuming a relatively large number of knots within the time domain is generally recommended, so that potential overfitting problems are circumvented with a penalty that counterbalances the flexibility. This is of primary importance because it solves one of the major questions when approximating an unspecified shape by a smoothing curve. In parallel with the penalty specification, the other cornerstone lies in the adequate selection of the number of knots for each approximation, Q_{h_0} and Q_α , respectively. The great advantage of Bayesian P-splines is precisely the fact that choosing the number of knots is not a critical step, since the penalty corrects possible overfitting concerns. This does not mean, however, that in practice one can indefinitely increase the number of knots placed, since a trade-off between enough smoothness in the fitted curve and an acceptable computational burden is necessary. Eilers and Marx²⁴ suggested including between 15 and 20 knots to avoid thinking about the number of knots needed, especially when dealing with large samples. Along the same lines, Ruppert³³ pointed out that, in the case of functions which are expected to have few oscillations and slow changes, one can barely observe any improvement when using more than 20 knots

(he provides an example with a sample size of $n = 25000$). Taking these previous references into account, the maximum number of knots considered in our generalized proposal is set at 15 for large datasets in which the time-dependent function is expected to have a smooth trajectory without marked oscillations.

To provide a unified and flexible modeling framework, a Bayesian P-spline approach is the default option utilized in our proposal to examine the functional shapes of $\log\{h_0(\cdot)\}$ and $\alpha(\cdot)$ within the joint model. Moreover, for comparison, the parameter estimates and the corresponding time-varying profile for a Gompertz approximation of the baseline hazard function are also included in the results section.

2.2.2 | Prior setting

Regarding the ZINBP longitudinal approach, the univariate prior distributions for the fixed-effects regression coefficients of the binary and count parts are $\beta_{B,\ell_B} \sim N(0, \tau_{\beta_B})$, $\ell_B = 0, \dots, p_B$ and $\beta_{C,\ell_C} \sim N(0, \tau_{\beta_C})$, $\ell_C = 0, \dots, p_C$, respectively, with $\tau_{\beta_B} = 1/\sigma_{\beta_B}^2$ and $\tau_{\beta_C} = 1/\sigma_{\beta_C}^2$. Here prior variances are allowed to be sufficiently large by assuming small precision parameters, $\tau_{\beta_B} = \tau_{\beta_C} = 0.01$, say. Furthermore, the priors assumed for the scale and power parameters are $\kappa \sim U(1, 5)$ and $P \sim U(1, 5)$.

The prior assignment within the survival approach starts with independent and diffuse normal distributions for the components of vector γ_w , namely $N(0, 0.01)$. For the B-spline approximation of the time-dependent functions $\log\{h_0(t)\}$ and $\alpha(t)$, an equally-spaced vector ξ_{h_0} of Q_{h_0} knots is placed on the time domain $[t_{\min}, t_{\max}]$, $t_{\min} = \xi_{h_0,1} < \dots < \xi_{h_0,Q_{h_0}} = t_{\max}$, while a vector ξ_α of Q_α knots is located over $[t_{\min}, t_{\max}]$, $t_{\min} = \xi_{\alpha,1} < \dots < \xi_{\alpha,Q_\alpha} = t_{\max}$, thus dividing the domain for each case into $Q_{h_0} + 1$ and $Q_\alpha + 1$ subintervals. Hence, the baseline hazard function on the log-scale can be approximated through a linear combination of $R_{h_0} = (Q_{h_0} - 1) + d_{h_0}$ B-spline functions, while the true correlation between longitudinal responses and event times can be captured via $R_\alpha = (Q_\alpha - 1) + d_\alpha$ B-splines:

$$\log\{h_0(t)\} = \sum_{r=1}^{R_{h_0}} \gamma_{h_0,r} B_{d_{h_0},r}(t, \xi_{h_0}), \quad \alpha(t) = \sum_{\tilde{r}=1}^{R_\alpha} \gamma_{\alpha,\tilde{r}} B_{d_\alpha,\tilde{r}}(t, \xi_\alpha). \quad (10)$$

Here $\{B_{d_{h_0},r}(t, \xi_{h_0}), r = 1, \dots, R_{h_0}\}$ denotes the set of R_{h_0} B-spline basis functions when approximating $\log\{h_0(t)\}$, while $\gamma_{h_0} = (\gamma_{h_0,1}, \dots, \gamma_{h_0,R_{h_0}})^\top$ is the corresponding vector of unknown regression coefficients (without biological or physical interpretation). Similarly $\{B_{d_\alpha,\tilde{r}}(t, \xi_\alpha), \tilde{r} = 1, \dots, R_\alpha\}$ embraces the R_α B-spline functions to mimic the shape of $\alpha(t)$, while $\gamma_\alpha = (\gamma_{\alpha,1}, \dots, \gamma_{\alpha,R_\alpha})^\top$ reports the regression coefficients. The curve-fitting solutions for $\log\{h_0(t)\}$ and $\alpha(t)$ are positively defined over the $d_{h_0} + 2$ and $d_\alpha + 2$ adjacent knots, respectively, being zero everywhere else. Because of boundary conditions from the B-spline definition, the original support of knots is extended in each approximation by respectively adding d_{h_0} and d_α knots to the left and right of their border-knots, namely $\{\xi_{h_0,1}, \xi_{h_0,Q_{h_0}}\}$ and $\{\xi_{\alpha,1}, \xi_{\alpha,Q_\alpha}\}$. The resulting $Q_{h_0} + 2d_{h_0}$ and $Q_\alpha + 2d_\alpha$ knots will be used to generate two complete basis with B-splines of degrees d_{h_0} and d_α . In this article, we use Bayesian penalized B-splines²⁷ to obtain a parsimonious parameterization of the baseline hazard and, especially, of the time-varying association parameter. Following such an approach, the joint prior distributions for vectors γ_{h_0} and γ_α are conditional on the amount of smoothness introduced by the positive-valued hyperparameters τ_{h_0} and τ_α , which penalize the roughness of $\log\{h_0(t)\}$ and $\alpha(t)$, respectively. Concretely, these overall reference priors are assumed to be

$$\pi(\gamma_{h_0} | \tau_{h_0}) \propto \tau_{h_0}^{\text{rank}(\mathbf{M}_{h_0})/2} \exp\left(-\frac{\tau_{h_0}}{2} \gamma_{h_0}^\top \mathbf{M}_{h_0} \gamma_{h_0}\right), \quad \pi(\gamma_\alpha | \tau_\alpha) \propto \tau_\alpha^{\text{rank}(\mathbf{M}_\alpha)/2} \exp\left(-\frac{\tau_\alpha}{2} \gamma_\alpha^\top \mathbf{M}_\alpha \gamma_\alpha\right), \quad (11)$$

yielding the hierarchical multivariate Gaussian priors $\gamma_{h_0} | \tau_{h_0} \sim N_{R_{h_0}}(\mathbf{0}, \tau_{h_0} \mathbf{M}_{h_0})$ and $\gamma_\alpha | \tau_\alpha \sim N_{R_\alpha}(\mathbf{0}, \tau_\alpha \mathbf{M}_\alpha)$, where the symmetric and positive-definite precision matrices \mathbf{M}_{h_0} and \mathbf{M}_α play in turn the role of appropriate penalty matrices. Validity of such a priors then yields the global penalties $\text{pen}(\gamma_{h_0}) = (\tau_{h_0}/2) \gamma_{h_0}^\top \mathbf{M}_{h_0} \gamma_{h_0}$ and $\text{pen}(\gamma_\alpha) = (\tau_\alpha/2) \gamma_\alpha^\top \mathbf{M}_\alpha \gamma_\alpha$. Here $\mathbf{M}_{h_0} = \Delta_u^\top \Delta_u + \varepsilon \mathbf{I}$ and $\mathbf{M}_\alpha = \Delta_v^\top \Delta_v + \varepsilon \mathbf{I}$, where Δ_u and Δ_v are the difference matrices based upon u -th and v -th order squared finite differences of adjacent B-splines, respectively, while the term $\varepsilon \mathbf{I}$ introduces a small ‘‘ridge penalty’’ to avoid a linearly dependent system. This is done by providing a small enough value for ε , for instance $\varepsilon = 10^{-6}$. We assume here onwards penalties based on second-order differences, yielding $\text{rank}(\mathbf{M}_{h_0}) = R_{h_0} - 2$ and $\text{rank}(\mathbf{M}_\alpha) = R_\alpha - 2$ in (11). The sets of parameters $\{\gamma_{h_0}, \tau_{h_0}\}$ and $\{\gamma_\alpha, \tau_\alpha\}$ are in each case simultaneously estimated by assuming a prior distribution which delivers a large positive support for the unknown smoothing hyperparameters τ_{h_0} and τ_α . For these, the standard prior choice is Gamma(a_τ, b_τ) with expectation a_τ/b_τ and variance a_τ/b_τ^2 , where a_τ and b_τ are chosen to be minimally informative. Traditional options advocate for assigning highly diffuse but proper priors, with small values for $a_\tau = b_\tau$ ^{28,29}, or, alternatively, combining $a_\tau = 1$ with a small value for

b_{τ} .^{27,30,17} Both options are generally proven to supply quite similar results when the sample size is large enough.^{31,32} In our approach, we have considered with $\tau_{h_0} \sim \text{Gamma}(1, 0.005)$ and $\tau_{\alpha} \sim \text{Gamma}(1, 0.005)$.

The prior setting concludes with the specification of the normally distributed random effects, which is completed by assigning a proper hyperprior distribution for the $q(q+1)/2$ distinct entries of the covariance matrix \mathbf{D} . Within the Bayesian framework, the hierarchical covariance matrix of a multivariate normal variable is combined in a natural way with the inverse-Wishart (IW) prior distribution, so $\mathbf{D} \sim \text{IW}_q(\mathbf{S}^{-1}, \nu)$. Here \mathbf{S}^{-1} is a $q \times q$ positive definite scale matrix and ν denotes the degrees of freedom, with $\nu \geq q$ to ensure the properness of the prior. This in turn equates to picking a Wishart prior for the precision covariance matrix, $\mathbf{D}^{-1} \sim \text{W}_q(\mathbf{S}, \nu)$. A classical prior setting is usually established as $\mathbf{S} = \mathbf{I}$ and $\nu = q + 1$.²³

3 | POSTERIOR INFERENCE

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_y^\top, \boldsymbol{\theta}_t^\top, \boldsymbol{\theta}_b^\top)^\top$ denote the unknown complete parameter vector, where $\boldsymbol{\theta}_y = (\boldsymbol{\beta}_B, \boldsymbol{\beta}_C, \kappa, P)^\top$ collects the fixed-effect parameters of the longitudinal submodel, $\boldsymbol{\theta}_t = (\boldsymbol{\gamma}_w, \tau_{h_0}, \boldsymbol{\gamma}_{h_0}, \tau_{\alpha}, \boldsymbol{\gamma}_{\alpha})^\top$ the fixed quantities related to the time-to-event response, and $\boldsymbol{\theta}_b \equiv \mathbf{D}$ the shared variance components. Let $\mathcal{D}_n = \{(\mathbf{y}_i, L_i, T_i, \Delta_i), i = 1, \dots, n\}$ summarize the observed information from the target population. To construct the full conditional likelihood for the joint model, we assume that, given the random effects of the i -th subject, these two processes are independent, as are the subject's n_i longitudinal responses.⁵ Under this conditional independence assumption, the complete conditional joint likelihood can be factorized as

$$\begin{aligned} p(\mathcal{D}_n | \mathbf{b}_i, \boldsymbol{\theta}) &= \prod_{i=1}^n p_y(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\theta}) p_t\{T_i, \Delta_i | T_i^* > L_i, \mathcal{M}_i(t), \boldsymbol{\theta}\} \\ &= \prod_{i=1}^n \prod_{j=1}^{n_i} p_y\{y_i(t_{ij}) | \mathbf{b}_i, \boldsymbol{\theta}\} \frac{p_t\{T_i, \Delta_i | \mathcal{M}_i(t), \boldsymbol{\theta}\}}{\Pr\{T_i^* > L_i | \mathcal{M}_i(t), \boldsymbol{\theta}\}}, \end{aligned} \quad (12)$$

where $p_y(\cdot)$ is the conditional likelihood function of the zero-inflated count rates and $p_t(\cdot)$ is the conditional likelihood for the event times. The resulting posterior distribution of parameters $\pi(\boldsymbol{\theta}, \mathbf{b}_i | \mathcal{D}_n)$ is obtained through the celebrated Bayes' rule, $\pi(\boldsymbol{\theta}, \mathbf{b}_i | \mathcal{D}_n) \propto p(\mathcal{D}_n | \mathbf{b}_i, \boldsymbol{\theta}) \pi(\boldsymbol{\theta}, \mathbf{b}_i)$. Here, the prior distribution for the unknown parameters $(\boldsymbol{\theta}, \mathbf{b}_i)$ can in turn be factorized as

$$\pi(\boldsymbol{\theta}, \mathbf{b}_i) = p_b(\mathbf{b}_i | \boldsymbol{\theta}_b) \pi(\boldsymbol{\theta}), \quad (13)$$

where $p_b(\mathbf{b}_i | \boldsymbol{\theta}_b)$ follows a zero-mean multivariate Gaussian distribution. In addition, prior independence is assumed between the components of $\boldsymbol{\theta}$, so the posterior distribution for the parameters satisfies

$$\begin{aligned} \pi(\boldsymbol{\theta}, \mathbf{b}_i | \mathcal{D}_n) &\propto \prod_{i=1}^n \prod_{j=1}^{n_i} p_y\{y_i(t_{ij}) | \mathbf{b}_i, \boldsymbol{\theta}_y\} \frac{p_t\{T_i, \Delta_i | \mathcal{M}_i(t), \boldsymbol{\theta}_t\}}{\Pr\{T_i^* > L_i | \mathcal{M}_i(t), \boldsymbol{\theta}_t\}} \\ &\times p_b(\mathbf{b}_i | \boldsymbol{\theta}_b) \pi(\boldsymbol{\beta}_B) \pi(\boldsymbol{\beta}_C) \pi(\kappa) \pi(P) \pi(\boldsymbol{\gamma}_w) \pi(\tau_{h_0}) \pi(\boldsymbol{\gamma}_{h_0} | \tau_{h_0}) \pi(\tau_{\alpha}) \pi(\boldsymbol{\gamma}_{\alpha} | \tau_{\alpha}) \pi(\mathbf{D}). \end{aligned} \quad (14)$$

The resulting expression, however, is analytically intractable, so we approximate it using MCMC algorithms. The specific form presented by $p_y(\cdot)$, $p_t(\cdot)$, and $p_b(\cdot)$ is derived in Appendix B.

4 | SIMULATION STUDY

4.1 | General design

A simulation study is conducted to firstly check that the general algorithm proposed to run the JMTV-ZINBP is indeed working, and then to determine the main effects of using more elementary joint model structures to fit data generated from a JMTV-ZINBP. In particular, we begin with a base scenario in which a set of $M = 500$ independent datasets are both simulated and fitted using our complex joint model. Then, three additional scenarios are considered, using in every case a particular simplified version of our proposed joint model to fit the same 500 datasets generated in the first scenario. Each of these three alternative joint models lacks a noteworthy feature of the JMTV-ZINBP, thus enabling us to examine which aspect of the generated data is

not properly characterized. The four distinct scenarios are named according to the joint model assumed therein to fit the common datasets. We therefore have (i) JMTV-ZINBP, (ii) JMTV-NBP, (iii) JMTV-ZIP, and (iv) JM-ZINBP. A replicated dataset reports the longitudinal counts observed over a 7.5-year observation window for $n = 250$ subjects, whose distribution by age is similar to that observed in the HI data. All subjects are 65 or older at the time of their first measurement, and subsequent measurements are then taken annually. To properly account for this age adjustment in the event time process, the time scale is measured in years above the age of 65, so this fixed truncation age is directly linked to our time zero. Particularly, 30% of individuals are assumed to enter the study at $t = 0$, while the remaining percentage enter at any time between $t = 0$ and $t = 30$ years; such a substantial proportion of left-truncated event times typically arises when the observation of a non-selected cohort of senior individuals is conditioned on a fixed age. Aside of truncation, the event times are potentially affected by a non-informative censoring mechanism, expressed by independent and uniformly distributed random variables C_i with a mean of 20 years. This results in about 45% of all subjects being right-censored, so each subject's longitudinal information is gathered through at most $n_i \leq 8$ observations. Additionally, the baseline hazard function for all scenarios is taken to be Gompertz distributed over time.

The parameter setting for the datasets common to all scenarios is established as follows. The binary part only includes fixed and random baseline terms, whereas the count part is described using cubic natural B-splines of time, with a single internal knot φ at the median of time measurements related to each dataset, and considering both a fixed and a random baseline effect. Specifically, the two longitudinal predictors are $\eta_{Bi}(t) \equiv \eta_{Bi} = \beta_{B_0} + b_{iB_0}$, and $\eta_{Ci}(t) = \beta_{C_0} + b_{iC_0} + \beta_{C_1} B_{3,1}(t, \varphi) + \beta_{C_2} B_{3,2}(t, \varphi)$, where the random intercepts of binary and count parts are assumed to be uncorrelated, i.e. $\rho_{BC} = 0$; this way, the specification for the random effects is reduced to $b_{iB_0} \sim N(0, \sigma_{B_0}^2)$ and $b_{iC_0} \sim N(0, \sigma_{C_0}^2)$, with $\sigma_{B_0} = 0.500$ and $\sigma_{C_0} = 0.400$. We let $\beta_{B_0} = 1.525$, $(\beta_{C_0}, \beta_{C_1}, \beta_{C_2}) = (0.405, 0.165, 0.070)$, $\kappa = 1.950$, $P = 1.100$, and the exposure period is set at 1 for all measurements. This yields count values in the range 0-20, with 45% of overall zero counts. The relative risks model in (9) includes a single baseline zero-one indicator group $\mathbf{w}_i(t) \equiv \text{grp}_i \sim \text{Bernoulli}(0.70)$, with $\gamma_w = -0.150$. Likewise, the baseline risk function is fixed as $h_0(t) = \exp(-3.910 + 0.027t)$, while the underlying association curve is assumed to be $\alpha(t) = 2.1610 - 0.0752t + 0.0005t^2$.

4.2 | Computed information

The time-varying association $\alpha(t)$ is approximated through a P-spline basis of degree $d_\alpha = 3$, placing $Q_\alpha = 8$ internal knots between $t_{\min} = 0$ and $t_{\max} = 37.5$ years. The marginal posterior density for a particular component θ from vector $\boldsymbol{\theta}$ is computed within each scenario for the corresponding m -th fitting, $m = 1, \dots, M$. By averaging over the M posterior results of θ , we report: the average of the means, $\bar{\theta}_{\text{av}} = \sum_{m=1}^M \bar{\theta}_m / M$; the average bias, $\text{bias}_{\text{av}} = \bar{\theta}_{\text{av}} - \theta$; the average standard error, $\text{SE}_{\text{av}} = \sum_{m=1}^M \text{SE}_m / M$; the empirical standard error of the mean estimates, $\text{SE}_{\text{emp}} = \left\{ \sum_{m=1}^M (\bar{\theta}_m - \bar{\theta}_{\text{av}}) / (M-1) \right\}^{1/2}$; and the average 95% credible interval, $\text{CI}_{\text{av}} = [p_{2.5, \text{av}}, p_{97.5, \text{av}}]$, where $p_{2.5, \text{av}} = \sum_{m=1}^M p_{2.5, m} / M$ and $p_{97.5, \text{av}} = \sum_{m=1}^M p_{97.5, m} / M$ denote the average 2.5th and 97.5th percentiles of θ , respectively. The estimation procedure is handled using an MCMC algorithm, which sequentially draws random and independent samples from the posterior distribution. We use Gibbs sampling to fit our model by means of the Bayesian software JAGS v. 4.3.0³⁴. For each simulated dataset, we run two parallel chains with 10000 iterations per chain, while discarding the first 9000 iterations per chain as a burn-in period. Thinning is set to keep 500 posterior samples from each chain, thus using a total of 1000 samples for estimation. Convergence is checked via the visual examination of traceplots, as well as through the potential scale reduction factor \hat{R} , defined by Gelman and Rubin³⁵, and the effective sample size, n_{eff} .

4.3 | Simulation results

The average posterior mean estimates for the joint model parameters of each scenario are listed in Tables W6-W9, provided in Section C.1. of Appendix C. Along with the fitting of each dataset under a particular scenario, a new dataset of size $n = 250$ is simulated from the corresponding posterior parameter distribution. Thereby, an average count distribution can be derived for each of the four scenarios from the 500 replicated datasets generated therein, and these frequency distributions can in turn be compared to the theoretical distribution from the initially generated 500 datasets. Such a comparison is presented in Table W10, which appears in Section C.2. of Appendix C.

The average results of scenario (i) suggest that the JMTV-ZINBP algorithm yields little bias in most of the parameter estimates, while the average of the estimated standard errors shows a strong concordance with the empirical standard error. It is also interesting to note the great similarity between the fitted and theoretical frequency distribution of the count rates. In the remaining three scenarios, however, some significant disparities between the average posterior result and the true magnitude may

be consistently observed because of the simplified joint model assumed. In scenario (ii), the the JMTV-NBP does not explicitly account for the excess of zeros, so the additional variability from this source of overdispersion is absorbed by the catch-all parameter κ . In comparison with its theoretical value when generating the data, this term exhibits an average posterior mean that is nearly half, $\kappa = 1.950$ versus $\bar{\kappa}_{\text{av}}(t) = 1.058$. Hence, the longitudinal part becomes highly efficient in accommodating the overall proportion of zeros, even if these are slightly underestimated. Conversely, lower counts are likely to be overfitted. With respect to the shape of the association parameter, we can observe that its corresponding average estimate is fairly close to the true time-varying shape. In scenario (iii), the specific features of the two-component ZIP response are specifically developed to deal with an excess of zeros. The JMTV-ZIP alternative, however, assumes a Poisson model for its count part, thereby neglecting the non-linearities coming from unobserved heterogeneity among clustered observations. This lack of flexibility is well-reflected as a significant reduction of the corresponding average posterior mean of the binary part's intercept, $\beta_{B_0} = 1.525$ versus $\bar{\beta}_{B_0, \text{av}} = 0.903$, yielding a serious overestimation for the excess of zeros. Nonetheless, the fitted frequency of overall zeros remains in line with the theoretical frequency. A further examination of the fitted responses from this model reveals that the relatively long tail displayed by the original data is not captured; there is definitively more latent variability in the data than the model can adjust for. Regarding the estimation of the trajectory $\alpha(t)$, similar results to the previous scenarios are derived, providing the model with a high capability to mimic the true association shape. Scenario (iv) retains the same longitudinal assumptions as the model which generates the data, but the constant association parameter assumed for the JM-ZINBP leads to substantial departures of the longitudinal coefficients from their true corresponding values. This ultimately gives rise to an underprediction of the frequency of zeros, along with an overrepresentation of small and intermediate counts. As a key feature of this scenario, it summarizes the time-varying relationship between longitudinal and event time with a single average posterior mean, $\bar{\alpha}_{\text{av}} = 1.517$; this is actually a particular value within the range covered by the true $\alpha(t)$. For illustrative purposes, Figure 2 displays the 500 time-varying correlation profiles under scenario (i), while Figure W2 in Section C.3 of Appendix C presents the time varying-profiles derived for each of the four scenarios.

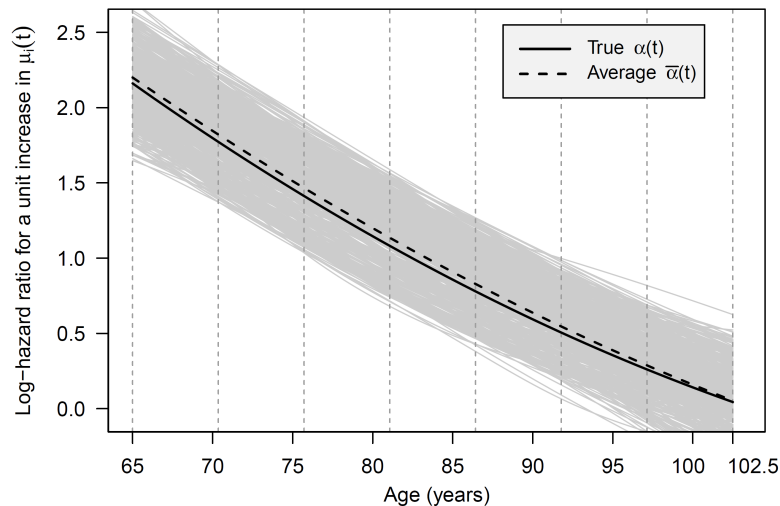


FIGURE 2 Simulation results for the time-varying association under scenario (i). The target shape of $\alpha(t)$ is compared to the average posterior mean $\bar{\alpha}_{\text{av}}(t)$, which is obtained by averaging over the $M = 500$ posterior mean profiles from simulations.

5 | ANALYSIS OF THE EMPIRICAL DATA

5.1 | Longitudinal results

We first consider a Bayesian longitudinal model with a ZINBP response to accommodate the observed emergency claims per year from the HI data. Although this is not a necessary step, it provides a reliable guide about the prior choice for the longitudinal parameters in the subsequent JMTV-ZINBP fit, especially when faced with a relatively large dataset such as ours. For the binary response, we assume that the probability of private medical usage is explicitly dependent on subject-specific behavioral habits, where both fixed and random intercept terms are considered. Such usage is also influenced by two zero-one indicators: the

subject's gender, $\text{sex}_i = I(\text{sex}_i = \text{female})$, and the household income, $\text{inc}_i = I(\text{inc}_i > \text{average})$, which establishes who has a monthly income above the municipal average. By contrast, the count response accounts for non-linear patterns over time by using natural cubic B-splines, with a random term also assumed at a baseline level. The gender and household income factors are also included in this count part. Following (5), we set

$$\begin{aligned}\eta_{Bi}(t) &\equiv \eta_{Bi} = (\beta_{B_0} + b_{iB_0}) + \beta_{B_{\text{sex}}} \text{sex}_i + \beta_{B_{\text{inc}}} \text{inc}_i, \\ \eta_{Ci}(t) &= (\beta_{C_0} + b_{iC_0}) + \beta_{C_1} N_{3,1}(t, \varphi) + \beta_{C_2} N_{3,2}(t, \varphi) + \beta_{C_{\text{sex}}} \text{sex}_i + \beta_{C_{\text{inc}}} \text{inc}_i,\end{aligned}$$

where $\mathbf{b}_i = (b_{iB_0}, b_{iC_0})^\top \sim N_2(\mathbf{0}, \mathbf{D})$. Correlation between the binary and count responses is defined through $\rho_{BC} = \text{Cor}(b_{iB_0}, b_{iC_0})$, which controls how the subject's need for emergency care at time t relates to the claim intensity at other time points. In this regard, if our general model only accounts for random intercepts in both the binary and count parts (i.e. $q = 2$), an easy-to-handle alternative to the standard inverse Wishart prior could be based on the factorization $\pi(b_{iB_0}, b_{iC_0}) = \pi(b_{iB_0})\pi(b_{iC_0} | b_{iB_0})$, where $b_{iB_0} \sim N(0, \tau_{B_0})$ and $b_{iC_0} \sim N(\psi b_{iB_0}, \tau_{\text{cond}})$, with $\psi \in \mathbb{R}$.³⁶ See also Neelon et al.³⁷ for a detailed example of such a strategy in the hierarchical zero-inflated setting. The two referred approaches to account for random effects were previously tested and led to similar results. In this article, however, we only follow the procedure of assigning an IW prior to the random effects in order to provide a general overview of the proposed model.

The Bayesian estimation is conducted via JAGS using 10000 iterations for the adaptation, and then running two parallel chains of 25000 iterations each after a burn-in period of 15000. We kept 1000 posterior samples from each chain. Mean, standard error, and 95% credible interval are sampled for each parameter from the corresponding marginal posterior distribution. The posterior summary statistics for the unknown parameters are provided in Table W11, within Section D.1 of Appendix D.

5.2 | Application of the JMTV-ZINBP

The proposed JMTV-ZINBP for the empirical data is concretized as

$$h_i \{t | \mathcal{M}_i(t), \text{sex}_i\} = h_0(t) \exp\{\gamma_{\text{sex}} \text{sex}_i + \alpha(t) \mu_i(t)\}, \quad t > 0,$$

in which the profiling of age-related functions $\log\{h_0(t)\}$ and $\alpha(t)$ can be approximated using P-splines of the same degree, $d_{h_0} = d_\alpha = d$, while also placing the same quantity $Q_{h_0} = Q_\alpha = Q$ of equally-spaced knots between $t_{\min} = 0$ and $t_{\max} = 37.5$ years; this indicates that the overall observations occur between the ages of 65 and 102.5 years old. We particularly set $d = 3$ and $Q = 15$, which results in computing $R = 17$ regression coefficients within each approximation of the function. Taking advantage of the previous section's results, the univariate normal priors for $\{\beta_{B_0}, \beta_{B_{\text{sex}}}, \beta_{B_{\text{inc}}}, \beta_{C_0}, \beta_{C_1}, \beta_{C_2}, \beta_{C_{\text{sex}}}, \beta_{C_{\text{inc}}}\}$ are now centered on their corresponding longitudinal mean estimates, while the distance between the lower and upper support of the uniform priors for κ and P is reduced. Furthermore, the only survival covariate considered is gender, $\mathbf{w}_i(t) \equiv \text{sex}_i$. The posterior distribution of the parameters, $\pi(\theta, \mathbf{b}_i | \mathcal{D}_n)$, is approximated by running an MCMC algorithm. Similarly to the longitudinal fitting, the number of iterations required in the adaptation phase is unusually high, at 25000. The fitting of the JMTV-ZINBP accounts for two parallel chains with 100000 iterations each after a burn-in of 50000, and a total of 2000 posterior samples are drawn. The JAGS code within the R environment is given in the supplementary files. Sections D.2 and D.3 of Appendix D (Tables W12-W15) provide the main posterior statistics of the unknown parameters for the joint model fit with constant and time-varying associations, respectively, while distinguishing for each case whether the log-baseline hazard is approximated using a Gompertz model or a P-spline approach. Further, the traceplots and the posterior densities for the parameters are shown in Section D.4 (Figures W3-W11), from which a correct mixing and convergence of the chains can be concluded. The results of the two fitted JMTV-ZINBP give rise to the shapes for $h_0(t)$ and $\alpha(t)$ depicted for each case in the four-panel plot shown in Figure 3. Looking at this, it appears that the baseline hazard function exhibits an increasing convex trend for the target population, whether a Gompertz model is assumed or its underlying shape is approximated using P-splines. However, with the former approach, the uncertainty around the mean trend is higher. These functional forms seem reasonable, since mortality risk increases as people get older. The top and bottom right panels show hardly any differences and demonstrate that the annual demand for emergency medical care and the mortality hazard are positively correlated over time, while the posterior statistics obtained for the regression coefficients γ_α provide evidence that this correlation is higher than zero. Most remarkably, these two plots show that the relationship between these outcomes is not stationary from the age of 65 onwards, but instead varies significantly with age. From the age of 65 to the ages around 75, the mortality risk decreases sharply with each one-unit increase

in the annual emergency claims rate. In consistency with this, for relatively younger individuals (within the total range of 65-102.5), seeking emergency services is likely indicative of a more critical health status because the intensity of use itself is lower and because younger retirees are usually more mobile. Additionally, for 75- to 95-year-olds, medical emergency services are relatively more frequently used but not necessarily indicative of a terminal illness. As a result, the time-varying association trend shows barely any noticeable variation in the 75-95 age range. Also recall that the abundance of data for the ages around 85 allows for a reduction of the 95% credible intervals for that age range. Finally, the risk exhibits an increasing trend from the age of 95 onwards, even though emergency service usage typically reduces sharply at very old ages. Here, the survival probability itself is so low that any unusual occurrence leads to an increase in the mortality risk.

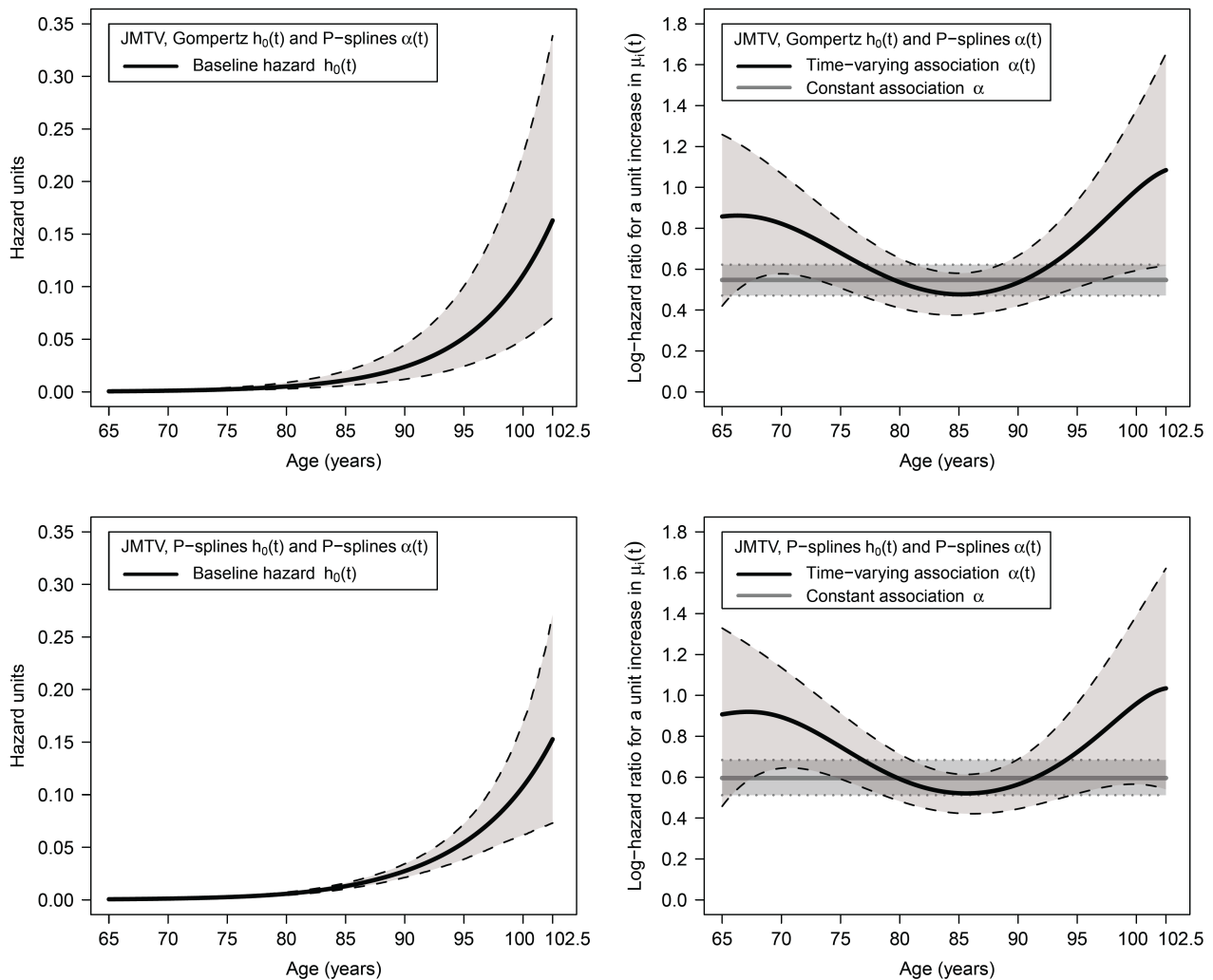


FIGURE 3 Time-varying parameters derived from the two JMTV-ZINBP fitted to the HI data. Top left: Baseline hazard function when assuming a Gompertz survival model for $h_0(t)$. Top right: Comparison between a constant association parameter and a time-varying association when assuming a Gompertz survival model for $h_0(t)$. Bottom left: Baseline hazard function when approximating $h_0(t)$ with P-splines. Bottom right: Comparison between a constant association parameter and a time-varying association when approximating $h_0(t)$ with P-splines. The 95% credible intervals are displayed for all graphs.

The implications of these results are of vital importance within the health care field, giving rise to wider applications that improve resource allocation across different age groups. Indeed, the postulated joint model allows for contouring the age-related hazard ratio $HR(t) = \exp\{\alpha(t)\Delta EC\}$ for a specific change in the annual amount of emergency claims, ΔEC . Then, a first option to concisely capture the age effect on mortality risk relies on transforming $\bar{\alpha}(t)$ into a piecewise-constant trend over a predefined age intervals. These can be chosen in accordance with the three distinct trends for $\bar{\alpha}(t)$ displayed in Figure 3, 65-75, 75-95, and

95-102.5. Constant association values can be separately derived within each of these three age intervals, namely $\bar{\alpha}_k$, $k = 1, 2, 3$, so that each value is obtained by averaging $\bar{\alpha}(t)$ over the corresponding interval. In particular, we obtain the following mean values (with their associated 95% CI): $\bar{\alpha}_1 = 0.873$ (0.607, 1.134), $\bar{\alpha}_2 = 0.593$ (0.471, 0.723), and $\bar{\alpha}_3 = 0.895$ (0.553, 1.284). These quantities provide a simple but useful piecewise expression for the hazard ratio, based on the age range in which the age t of a certain policyholder is located.

A more accurate option for communicating the results is through the use of the shaded contour graph shown in Figure 4, which illustrates the hazard ratio estimations for both a specific age t above 65 and a variation ΔEC , regardless of the number of claims observed in the immediately preceding year. The graph is consistent with the shape of the time-varying association in Figure 3, showing that emergency care usage has a greater impact on survival at the lower and very advanced ages among the senior policyholders. For example, let us consider three policyholders aged 65, 85, and 100, each of them having experienced an increase of one emergency claim. From the graph, the corresponding mortality risk projections lead to a $\overline{HR} = 2.51$ for the individual aged 65, a $\overline{HR} = 1.69$ for the individual aged 85, and a $\overline{HR} = 2.59$ for the centenarian. To gain insight into the impact of emergency care usage on survival at different ages, in Section D.5 of Appendix D, Table W16 presents the estimated hazard ratios that correspond to different increases in emergency claims per year at the given ages between 65 and 102.5 years old.



FIGURE 4 Two-dimensional death hazard map between the ages of 65 and 102.5 years old, showing potential variations of up to eight additional emergency claims per year above the previous level of demand.

6 | BAYESIAN MODEL ASSESSMENT

6.1 | Posterior predictive checking

A basic level of performance assessment consists of judging how well the posited model captures the main features of the observed longitudinal data \mathbf{y}_i , that is, examining the plausibility that such data have really been generated by our model. Indeed, the ultimate goal of our model assessment is to inform about possible misspecifications in the model or, more generally, to identify those aspects in which the model may be improved to provide a better data fitting. A validation procedure should, ideally, consider external data which have not been used for estimating the joint model parameters, but such information is not always available. Nevertheless, the model checking can still be accurately carried out using replicated datasets $\mathbf{y}_{rep,i}$ of the same size and shape as the observed one, each of these replications comprising a new dataset that would theoretically be obtained from our applied JMTV-ZINBP fitting if assuming the estimated $\{\bar{\theta}_y, \bar{\mathbf{b}}_i\}$ from the HI data for $\{\theta_y, \mathbf{b}_i\}$. The replicated datasets are merely simulated from the posterior predictive distribution $p(\mathbf{y}_{rep,i} | \mathbf{y}_i) = \iint p(\mathbf{y}_{rep,i} | \mathbf{b}_i, \theta_y) p(\theta_y, \mathbf{b}_i | \mathbf{y}_i) d\theta_y d\mathbf{b}_i$, so such replications can thereby actually be understood as predictions. Hence, it makes sense to consider that the more indistinguishable the replications are when compared to the original data, the more accurate the fitting is. This approach has sometimes been

criticized by arguing that the data are used twice, both for fitting the model and for checking how this model mimics any relevant aspect of the data; however, within the Bayesian framework, the posterior predictive p -value can be directly interpreted as a conditional probability statement given the data, so in that sense the data would really only be used once.³⁸

The degree of agreement between \mathbf{y}_i and $\mathbf{y}_{\text{rep},i}$ may be expressed either quantitatively or qualitatively. The first option relies on naturally extending some of the classic goodness-of-fit methods to the Bayesian framework, thus informing in probability terms about the potential discrepancies between the HI data and its replications. This is achieved through the construction of specific auxiliary test statistics $T(\cdot)$, which can depend on the data as well as the parameters. The information provided by these diagnostic quantities can be summarized by a 95% credible interval for $T_{\text{rep},i} \equiv T(\mathbf{y}_{\text{rep},i} | \mathbf{b}_i, \boldsymbol{\theta}_y)$, so a narrow-enough 95% credible interval around the observed quantity $T_{\text{obs}} \equiv T(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\theta}_y)$ contributes to providing strong evidence that the hypothetical replications can suitably account for a particular feature of the HI data. Additionally, a Bayesian p -value p_B can be calculated. This, in particular, shows the probability that some of the replicated test statistics exceed the quantity T_{obs} , that is, $p_B = \Pr(T_{\text{rep},i} > T_{\text{obs}} | \mathbf{b}_i, \boldsymbol{\theta}_y) + (1/2) \Pr(T_{\text{rep},i} = T_{\text{obs}} | \mathbf{b}_i, \boldsymbol{\theta}_y)$, for the count rates. Hence, p_B should tend to be close to 0.50 if the fitting of the model is appropriate, due to the uncertainty in $\{\boldsymbol{\theta}_y, \mathbf{b}_i\}$ from T_{obs} . As an example, Table 2 provides the checking results when analyzing how the model approximates two key aspects of our data distribution: the two types of zeros and the extreme values at the right tail. The results provided on the left-hand side of Table 2 summarize the posterior predictive test performed to assess how the replications $\mathbf{y}_{\text{rep},i}$, $i = 1, \dots, 2000$ are able to capture the overall percentage of zeros and the percentage of subjects for which all measurements are zero count rates. In general, the assessment of the number of zeros yields reasonable posterior intervals and Bayesian p -values. On the other hand, as displayed on the right-hand side of Table 2, the predictive checks for the existence of measurements above 20 (the maximum observed value) lead to unexpected near-one p -values. However, although almost all the replications contain some count rate above 20, so this itself is not a problematic result for our model validation since the percentage of such measurements is negligible. Note the agreement between p -values when assessing the percentage of count rates above 20 and those p -values derived from the percentage of subjects with at least one extreme count; it may be concluded that over-20 measurements appear exactly once within each of these subjects. As another numerical check, Table 3 compares the overall distribution of the observed data to the distribution which results from averaging the 2000 replicated datasets from the posterior predictive density.

TABLE 2 Summary of posterior predictive checks for two test statistics regarding the fitting of the zeros in the HI data: (a) percentage of recorded zero values, and (b) subjects without utilization of private health services.

Discrepancy	Assessment for zeros				Discrepancy	Assessment for values > 20			
	T_{obs}	$T_{\text{rep},i}^{2.5}$	$T_{\text{rep},i}^{97.5}$	p_B		T_{obs}	$T_{\text{rep},i}^{2.5}$	$T_{\text{rep},i}^{97.5}$	p_B
Perc. zeros					Perc. > 20				
Overall	63.1	62.6	63.7	0.609	Overall	0	0	0.025	0.985
Male	63.8	62.8	64.7	0.454	Male	0	0	0.043	0.911
Female	62.7	62.1	63.6	0.645	Female	0	0	0.024	0.923
Perc. subjects all zeros					Perc. subjects any > 20				
Overall	21.8	21.2	23.1	0.602	Overall	0	0	0.146	0.985
Male	22.6	21.8	24.9	0.683	Male	0	0	0.243	0.911
Female	21.4	20.1	22.7	0.517	Female	0	0	0.146	0.923

TABLE 3 Comparison between the observed distribution of count rates in the HI data and the distribution that results from averaging the 2000 replicated datasets from the posterior predictive distribution.

	Assessment for count rates											
	0	1	2	3	4	5	6	7	8	9	≥ 10	Total
Perc. observed	63.1	17.3	8.8	4.6	2.4	1.5	0.9	0.4	0.3	0.2	0.5	100
Perc. simulated	63.2	17.1	8.7	4.6	2.5	1.4	0.9	0.5	0.3	0.2	0.5	100

Alternatively, the comparison between the characteristics of the reference data and its replications can be graphically expressed by plotting some of the checks undertaken. In this regard, there is a broad range of checking plots which are proven to be useful^{23,39}. For instance, Figure 5 assesses whether the model is adequate to capture the health care usage by accounting for age and the mortality risk. With this purpose, each subject’s measurement is first allocated to one of three possible age categories, namely [65, 75), [75, 95), and [95, 102.5] years, and is then assigned to one of two possible subgroups distinguishing between alive and deceased subjects. The observed amount of emergency claims is recorded within each of the six possible groups, and observations are compared to the 2000 simulated distributions of annual emergency claims by age and event status. The position of T_{obs} with respect to the histogram generated from the replicated datasets allows us to illustrate how precisely our model fits the annual emergency demand when stratifying by the aforementioned groups. The more centered the observed test statistic is with respect to its corresponding histogram, the better the model captures the distribution of the yearly count rates in that category.

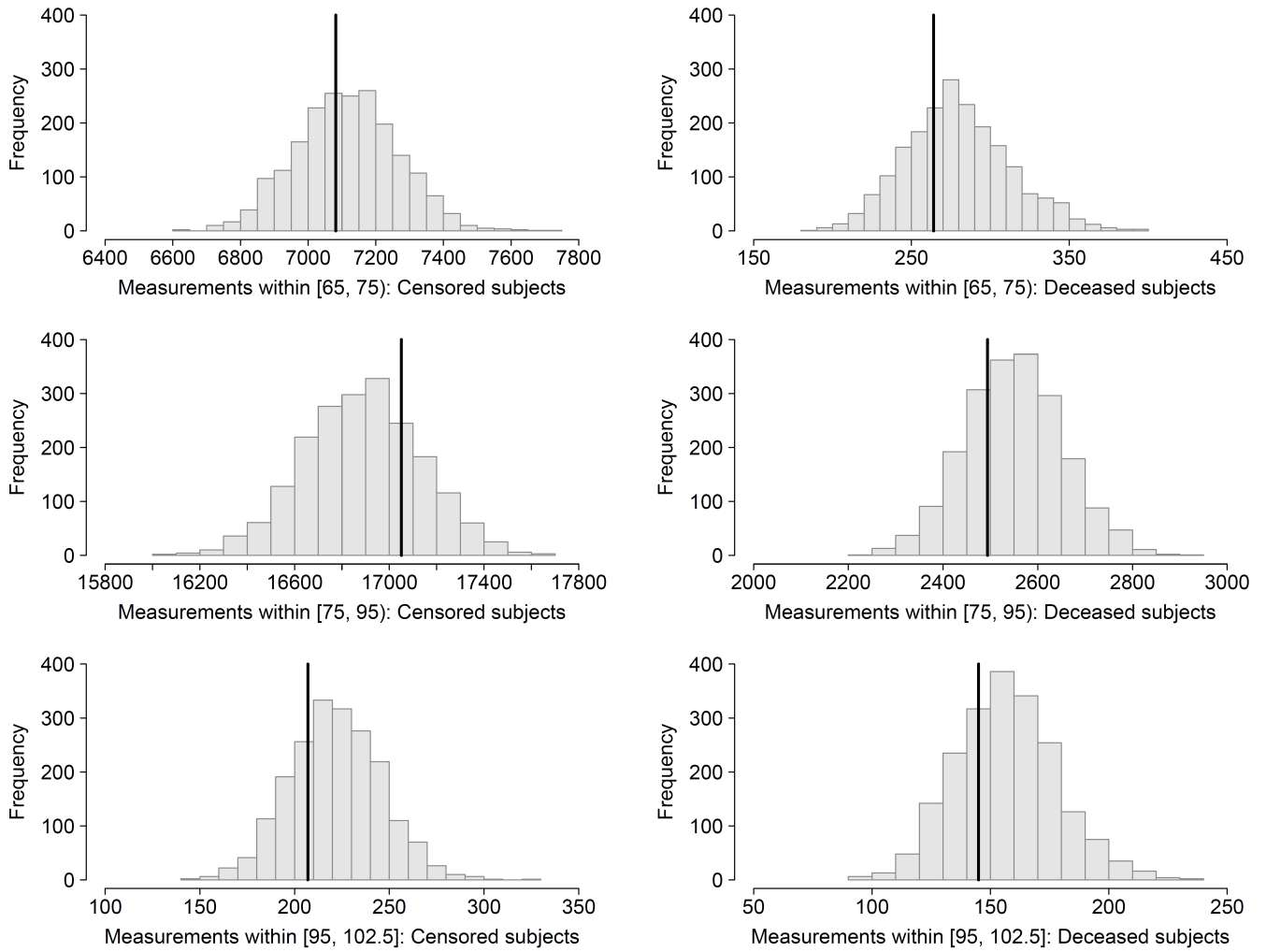


FIGURE 5 Observed amount of emergency claims per year per age group and survival status (vertical line at each plot) compared to 2000 replicated datasets from the posterior predictive distribution.

6.2 | Residual diagnostics

Beyond posterior predictive methodology for model diagnostics, a qualitative analysis of residuals arises as a primary step to validate the model, especially residual plots. These can in fact be considered a classical type of checking plots within the wide group of graphical posterior predictive tools. Since there are two simultaneous processes considered in our model, the analysis of the subject-specific (conditional) residuals must be separately performed for both approaches.⁵

6.2.1 | Residuals of the longitudinal submodel

For the longitudinal part, the analysis of the residuals is focused on a count-response model, so normality and homoscedasticity in the residuals is generally not observed. Instead, the corresponding scatterplot of the residuals versus the fitted values is expected to exhibit a non-homogeneous configuration, with the data being concentrated around as many nearly-parallel curves as there are possible integer responses in the observed data \mathbf{y}_i . An appropriate assessment of longitudinal residuals cannot therefore be directly conducted; a different procedure is necessary to achieve continuous residuals which allow for interpretable results. Going through this idea, Dunn and Smyth⁴⁰ define the so-called randomized quantile residuals, which enable the obtainment of continuous residuals in the case of non-Gaussian responses. To obtain this type of residuals, let us first consider the theoretical cumulative distribution of the current data, $F\{y_i(t) \mid \mathbf{b}_i, \boldsymbol{\theta}_y\}$, which in turn can be viewed as a continuous random variable distributed as $U(0,1)$. The subject-specific quantile residuals are therefore defined by

$$r_{q,i}(t \mid \mathbf{b}_i, \boldsymbol{\theta}_y) = \Phi^{-1}[F\{y_i(t) \mid \mathbf{b}_i, \boldsymbol{\theta}_y\}], \quad (15)$$

where Φ^{-1} refers to the inverse cumulative distribution function of the standard normal model, and $r_{q,i}(t \mid \mathbf{b}_i, \boldsymbol{\theta}_y) \xrightarrow{D} N(0, 1)$ if $\boldsymbol{\theta}_y$ and \mathbf{b}_i are consistently estimated. We next introduce a randomization process to derive continuously distributed residuals; we specifically use the replications $\mathbf{y}_{\text{rep},i}$ from the posterior predictive distribution to estimate the residuals:

$$r_{q,i}(t \mid \bar{\mathbf{b}}_i, \bar{\boldsymbol{\theta}}_y) = \Phi^{-1}[\Pr\{y_{\text{rep},i}(t) < y_i(t) \mid \bar{\mathbf{b}}_i, \bar{\boldsymbol{\theta}}_y\} + (1/2) \Pr\{y_{\text{rep},i}(t) = y_i(t) \mid \bar{\mathbf{b}}_i, \bar{\boldsymbol{\theta}}_y\}].$$

The above simulation-based strategy now allows us to obtain continuously distributed residuals for each response observation, and thus perform the usual residuals assessment based on the Q-Q plot. Figure W12 in Appendix E (left panel) displays the residuals assessment for our ZINBP longitudinal outcomes, where most of the plotted points tend to be concentrated around the 45-degree reference line, that is, the uniform cumulative distribution function. There are, however, two small groups of data for which the corresponding empirical quantiles are clearly larger than the theoretical ones. Specifically, the arrangement of points at the lower-left corner is probably due to the lack of left tail in the distributions of both our original and replicated datasets when comparing them to the Gaussian ones, whereas the group of points at the upper-right corner suggests a heavy right tail in the aforementioned distributions as a consequence of some extreme values.

6.2.2 | Residuals of the relative risk submodel

The residuals diagnostic for the survival part is assessed here through the Cox-Snell residuals⁴¹:

$$r_{cs,i}\{t \mid \mathcal{M}_i(t), \boldsymbol{\theta}_t\} = \mathcal{H}_i\{t \mid \mathcal{M}_i(t), \mathbf{w}_i(t)\} = \int_{L_i}^{T_i} h_i\{s \mid \mathcal{M}_i(t), \mathbf{w}_i(t)\} ds. \quad (16)$$

The basic idea behind this type of residuals comes from the principle that if the assumed model fits the data accurately, then

$$\mathcal{S}_i\{t \mid T_i^* > L_i, \mathcal{M}_i(t), \mathbf{w}_i(t)\} = \Pr\{T_i^* > t \mid T_i^* > L_i, \mathcal{M}_i(t), \mathbf{w}_i(t)\} \sim U(0, 1),$$

and consequently,

$$\mathcal{H}_i\{t \mid \mathcal{M}_i(t), \mathbf{w}_i(t)\} = -\log[\mathcal{S}_i\{t \mid T_i^* > L_i, \mathcal{M}_i(t), \mathbf{w}_i(t)\}] \sim \text{Exp}(1).$$

We can therefore assess the fitting of the survival part by examining if the Cox-Snell residuals approximately follow an exponential distribution of mean one. To properly account for left truncation and right censoring in the residuals' values, we compare the $\mathcal{S}_{\text{exp}}(t) = \exp(-t)$ curve with the Kaplan-Meier (KM) estimates of the survival function associated with the set of residuals $r_{cs,i}\{t \mid \mathcal{M}_i(t), \boldsymbol{\theta}_t\}$. In our case, such residuals are obtained from the posterior mean estimates $\bar{\boldsymbol{\theta}}$:

$$r_{cs,i}\{t \mid \bar{\mathcal{M}}_i(t), \bar{\boldsymbol{\theta}}_t\} = \int_{L_i}^{T_i} \bar{h}_0(s) \exp\{\bar{\gamma}_w^\top \text{sex}_i + \bar{\alpha}(s) \bar{\mu}_i(s)\} ds.$$

We can use the Bayesian computational tools to obtain a large number of replicates for the Cox-Snell residuals, thus including their corresponding KM survival curves in the final comparison. The results are shown on the right panel of Figure W12 in Appendix E, where the survival curve estimates are approximated through $\mathcal{S}_{\text{KM}}(t) \approx a \exp(bt) + c$, with $a, b, c \in \mathbb{R}$.

7 | DISCUSSION

Seeking a better understanding of mortality risk evolution among elderly beneficiaries of private care, we have simultaneously tackled different statistical challenges which may arise within the joint modeling framework. Specifically, our article sets itself apart from previous works by combining different aspects into a single estimation procedure. We have proposed a Bayesian joint model where the relationship between the longitudinal and time-to-event processes is flexibly modeled over time through cubic P-splines, while also accounting for the special features of each process. Assessing the longitudinal response involves tracking data with count rates that are affected by both an abundance of zeros and an inherent heterogeneity among subject-specific measurements. The hierarchical ZINBP regression model has been found to offer an especially optimal level of generality. In the survival process, the usual right censoring is accounted for, along with left truncation. This situation typically arises when individuals are subject to a determined age threshold for study entry, and the subsequent delayed entry bias can be properly addressed by translating each subject's starting point to the corresponding age above the aforementioned threshold. Moreover, an accurate estimation of the baseline hazard function is also proposed using a P-spline based approach. However, based on the researcher's previous knowledge of the data, this assumption may be eventually simplified to use an adequate survival model which has built-in smoothness.

A number of health care databases include survey information that is ideal for assessment via a joint model which combines a time-varying association with the aforementioned features. The proposed JMTV-ZINBP is exemplified using the relationship between the yearly counts for emergency health care usage and the mortality risk for a relatively large sample of policyholders over 65. Although it is broadly accepted that senior policyholders' morbidity increases with age, precisely determining the impact of health deterioration (measured here by an increase in claims, reflecting a rise in emergency care needs) on mortality risk projections at a specific age over 65 is not something that can be done intuitively. In contrast with the stationary hazard ratio for a one-claim increase provided by a traditional joint model approach, our time-varying joint model quantifies the dynamic nature of the true relationship between emergency demand and risk. The association shape obtained from our empirical data shows it to depart from a stationary relationship, exhibiting significant variation across the ages from 65 to 100. Concretely, the emergency care requirement turns out to be much more defining in the mortality risk for subjects under 75 years old, and not so much for ages around 85, where the aging process itself entails the need for greater medical attention. At ages above 95, the relationship between emergency care requirements and mortality risk is again higher than the constant parameter association obtained under a traditional joint model approach. This means that, as in the 65-75 age category, the emergency claims at 95-102.5 are more meaningful in terms of risk, albeit for different reasons. In this latter age group, human biology itself establishes immediate limits on the remaining life expectancy, so requiring emergency care services becomes very critical indeed in this age range. To conclude, our joint model allows for the estimation of the specific age-related hazard ratio for any age above 65. This provides a genuine output in private health care, supplying professionals with an initial measure for determining who is likely to require more emergency attention and thus allowing for the optimal assignment of policies over time.

A numerical study has been performed to test both the consistency of our general model and its advantageous features when compared to simpler joint model alternatives. In addition, a Bayesian model assessment has been performed by comparing the predictive distribution to the observed data, along with using residual plots related to both longitudinal and survival submodels. This makes headway in applying the proposed joint model in a wide range of potential studies focused on assessing the time-changing effect of a zero-inflated discrete outcome on the hazard process. Nonetheless, despite the unified modeling framework provided, further research remains on the agenda. Firstly, it would be very interesting to be able to estimate how many policyholders do not reach the age of 65 in our study due to major health problems. These missing subjects might ultimately lead to a bias in the estimates for both the longitudinal and time-to-event parameters⁴², inasmuch as healthier individuals from our target population are more likely to be observed. Although we expect that the number of policyholders who die before 65 is not very high, we do not currently have this information available, so this issue has not been addressed. Secondly, it would be extremely useful to include multiple longitudinal outcomes in the joint model^{43,44,45}, in order to allow for discerning between the three analyzed reasons for emergency medical attention. Thus, insurers could then identify the types of emergency claims which more decisively contribute to increases in the hazard risk. Delving deeper into this idea, an individual could then be assigned multiple health behaviors over time based on the amount and type of claims experienced, therefore extending survival techniques to model multistate event history data. More concretely, the observation times between sequentially-observed claims could be efficiently modeled using the hazard risk information, through the methodology applied in Richards.²² A third important extension to this work would be rooted on the idea that our data most likely masks a certain number of unknown sub-populations or latent classes⁴⁶; we could then focus on identifying how many individuals belong to each class in order to optimally fit the data.

Implementing the second and third points of improvement would not only entail methodological changes, but would also considerably increase the processing time for the joint model, and these have therefore remained beyond the scope of the present study. In connection with this, several authors have pointed out that the computation time may become the main constraint when applying Bayesian joint models.⁴⁷ For relatively large datasets such as ours, a key point of enhancement would be exploring alternative Bayesian procedures due to the associated computational complexity. For instance, in some fields of investigation, one might be interested in analyzing the main longitudinal response over periods of time shorter than a year, using monthly, weekly, or even daily count rates. From a conceptual point of view, there is no methodological constraint that would prevent the use of these interval lengths. In our case, however, this would entail splitting each of the current longitudinal responses into as many measurements as required for the time frame chosen. In practice, this could lead to unfeasible computational times.

ACKNOWLEDGMENTS

This research has been partially funded by the Spanish Ministry of Economy and Competitiveness, FPI grant BES-2014-070396. Montserrat Guillén thanks ICREA Academia 2018 and grant PID2019-105543GB-C21 from the Spanish Government's plan for scientific and technological research.

Conflict of interest

The authors declare no potential conflict of interests.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

References

1. Guillén M, Nielsen JP, Scheike TH, Pérez-Marín AM. Time-varying effects in the analysis of customer loyalty: A case study in insurance. *Expert Syst Appl*. 2012; 39(3): 3551–3558.
2. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*. 1982; 69(4): 331–342.
3. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time Data. *Biostatistics*. 2000; 1(4): 465–480.
4. Tsiatis AA, Davidian M. Joint Modeling of Longitudinal and Time-to-Event Data: An Overview. *Stat Sinica*. 2004; 14(3): 809–834.
5. Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data with Applications in R*. Chapman & Hall/CRC Biostatistic Series: Boca Raton, Florida (US); 2012.
6. Alemany R, Warner KE. The Lifetime Distribution of Health Care Costs. *Health Serv Res*. 2004; 39(3): 627–642.
7. Piulachs X, Alemany R, Guillén M, Rizopoulos D. Joint models for longitudinal counts and left-truncated time-to-event data with applications to health insurance. *SORT, Stat Oper Res T*. 2017; 41(2): 347–372.
8. Hatfield LA, Boye ME, Hackshaw MD, Carlin BP. Multilevel Bayesian Models for Survival Times and Longitudinal Patient Reported Outcomes with Many Zeros. *J Am Stat Assoc*. 2012; 107(499): 875–885.
9. Zhu H, Luo S, DeSantis SM. Zero-inflated count models for longitudinal measurements with heterogeneous random effects. *Stat Methods Med Res*. 2015; 26(4): 1774–1786.

10. Juarez-Colunga E, Silva G, Dean C. Joint Modeling of Zero-Inflated Panel Count and Severity Outcomes. *Biometrics*. 2017; 73(4): 1413–1423.
11. Rizopoulos D. The R Package JMbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC. *J Stat Softw*. 2016; 72(7): 1–46.
12. Crowther MJ, Andersson TM-L, Lambert PC, Abrams KR, Humphreys, K. Joint modelling of longitudinal and survival data: Incorporating delayed entry and an assessment of model misspecification. *Stat Med*. 2016; 35(7): 1193–1209.
13. Song X, Wang CY. Semiparametric Approaches for Joint Modeling of Longitudinal and Survival Data with Time-Varying Coefficients. *Biometrics*. 2008; 64(2): 557–566.
14. Barrett J, Diggle P, Henderson R, Taylor-Robinson D. Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference. *J R Stat Soc B*. 2015; 77(1): 131–148.
15. Köhler M, Umlauf N, Greven S. Nonlinear association structures in flexible Bayesian additive joint models. *Stat Med*. 2018; 37(30): 4771–4788.
16. Umlauf N, Klein N, Zeileis A. BAMLSS: Bayesian Additive Models for Location, Scale, and Shape (and Beyond). *J Comput Graph Stat*. 2018; 27(3): 612–627.
17. Andrinopoulou E-R, Eilers PHC, Takkenberg JJM, Rizopoulos D. Improved Dynamic Predictions from Joint Models of Longitudinal and Survival Data with Time-Varying Effects Using P-Splines. *Biometrics*. 2018; 74(2): 685–693.
18. Ridout M, Hinde J, Demétrio CGB. A Score Test for Testing a Zero-Inflated Poisson Regression Model Against Zero-Inflated Negative Binomial Alternatives. *Biometrics*. 2001; 57(1): 219–223.
19. Greene WH. Functional forms for the negative binomial model for count data. *Econ Lett*. 2008; 99: 585–590.
20. Richards SJ. Applying survival models to pensioner mortality data. *Brit Actuar J*. 2008; 14(2): 257–326.
21. Manton GK, Stallar E, Vaupel JW. Alternative Models for the Heterogeneity of Mortality Risks Among the Aged. *J Am Stat Assoc*. 1986; 81(395): 635–644.
22. Richards SJ. A handbook of parametric survival models for actuarial use. *Scand Actuar J*. 2012; 4: 233–257.
23. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. Chapman & Hall: Boca Raton, Florida (US); 2004.
24. Eilers PHC, Marx BD. Flexible smoothing with B-Splines and penalties. *Stat Sci*. 1996; 11(2): 89–121.
25. Dierckx P. *Curve and Surface Fitting with Splines*. Oxford University Press: New York (US); 1993.
26. Eilers PHC, Marx BD. Splines, knots, and penalties. *WIREs Comp Stat*. 1996; 2(6): 637–653.
27. Lang S, Brezger A. Bayesian P-Splines. *J Comput Graph Stat*. 2004; 13(1): 183–212.
28. Crainiceanu CM, Ruppert D, Wand MP. Bayesian Analysis for Penalized Spline Regression Using WinBUGS. *J Stat Softw*. 2005; 14(14): 1–24.
29. Yue YR, Simpson D, Lindgren F, Rue H. Bayesian Adaptive Smoothing Splines Using Stochastic Differential Equations. *Bayesian Anal*. 2014; 9(2): 397–424.
30. Komárek A, Brezger A. Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution. *Comput Stat Data Anal*. 2008; 52(7): 3441–3458.
31. Jullion A, Lambert P. Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. *Comput Stat Data Anal*. 2007; 51(5): 2542–2558.
32. Bauer C, Wakefield J, Rue H, Self S, Fenge Z, Wang Y. Bayesian penalized spline models for the analysis of spatio-temporal count data. *Stat Med*. 2016; 35(11): 1848–1865.

33. Ruppert D. Selecting the Number of Knots for Penalized Splines. *J Comput Graph Stat.* 2002; 11(4): 735–757.
34. Plummer M. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. Technische Universität Wien, Vienna (Austria); 2003.
35. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. *Stat Sci.* 1992; 7(4): 457–472.
36. Spiegelhalter DJ. Bayesian graphical modelling: A case study in monitoring health outcomes. *J R Stat Soc C.* 1998; 47(1): 115–133.
37. Neelon BH, O’Malley AJ, Normand S-LT. A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. *Stat Model.* 2010; 10(4): 421–439.
38. Gelman A, Meng X-L, Stern H. Posterior Predictive Assessment of Model Fitness via Realized Discrepancies (with comments). *Stat Sinica.* 2016; 6(4): 733–807.
39. Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A. Visualization in Bayesian workflow. *J R Stat Soc A.* 2019; 182(1): 1–14.
40. Dunn PK, Smyth GK. Randomized Quantile Residuals. *J Comput Graph Stat.* 1996; 5(3): 236–244.
41. Cox DR, Snell EJ. A general definition of residuals. *J R Stat Soc B.* 1968; 30(2): 248–275.
42. Piccorelli AV, Schluchter MD. Jointly modeling the relationship between longitudinal and survival data subject to left truncation with applications to cystic fibrosis. *Stat Med.* 2012; 31(29): 3931–3945.
43. Brown ER, Ibrahim JG, and DeGruttola V. A Flexible B-Spline Model for Multiple Longitudinal Biomarkers and Survival. *Biometrics.* 2005; 61(1): 64–73.
44. Chi Y-Y, Ibrahim JG. Joint Models for Multivariate Longitudinal and Multivariate Survival Data. *Biometrics.* 2006; 62(2): 432–445.
45. Rizopoulos D, Ghosh P. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Stat Med.* 2011; 30(12): 1366–1380.
46. Proust-Lima C, Séne M, Taylor JM, Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: a review. *Stat Methods Med Res.* 2014; 23(1): 74–90.
47. Ibrahim JG, Chen M-H, Sinha D. Bayesian Methods for Joint Modeling of Longitudinal and Survival Data with Applications to Cancer Vaccine Trials. *Stat Sinica.* 2004; 14(3): 863–883.

