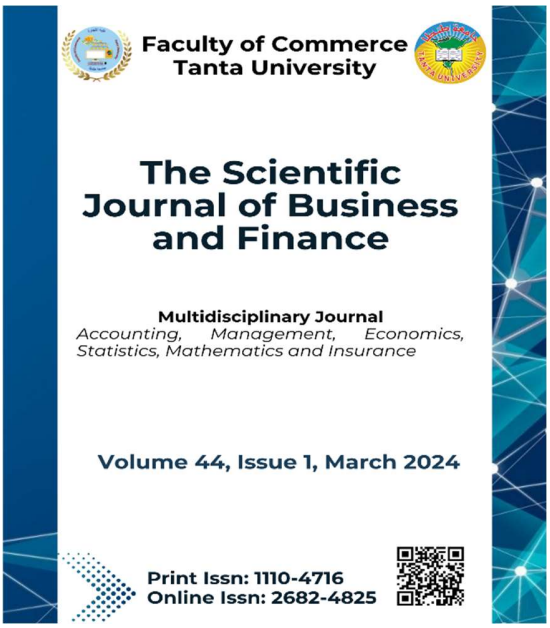

The Scientific Journal of Business and Finance

https://caf.journals.ekb.eg

Parametric Frailty Models for Elapsed Time between Recurrent Events

Khaled F. Bedair [*] [a] **and Dina Abdelhady**[a]

[a]Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Tanta University

Published online: **March 2024 .**

To cite this article: Bedair Khaled F and Abdelhady, Dina. Parametric Frailty Models for Elapsed Time between Recurrent Events, The Scientific Journal of Business and Finance, 44 (1), 319-344.

DOI: https://doi.org/10.21608/caf.2024.351760

*Corresponding author: Khaled.bedair@commerce.tanta.edu.eg

Parametric Frailty Models for Elapsed Time between Recurrent Events

Khaled F. Bedair [*] [a]

[a]Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Tanta University

Dina Abdelhady[a]

[a]Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Tanta University

Article History

Received 11 January 2024, Accepted 28 January 2024, Available online March 2024

Abstract

Analyzing recurring event data is crucial in clinical, epidemiological, and a wide range of other fields, requiring consideration of the interdependence among events within individuals and potential variability in event likelihood across different individuals. This paper introduces a comprehensive suite of models aimed at addressing the gap in parametric frailty models designed for recurrent events. The proposed models encompass five distinct baseline intensities and integrate gamma, inverse Gaussian, and positive stable frailty distributions. Parameter estimation is optimized for maximizing the marginal log-likelihood and accommodates both right-censored and potentially left-truncated data. Model selection is facilitated through the Aikake Information Criterion (AIC) and Bayesian Information Criterion (BIC). Simulation studies assess the computational algorithm and the efficacy of the proposed models. The models' performance is further demonstrated through the analysis of a dataset focusing on recurrent outcomes of phototherapies. Among the proposed models, the one based on a gamma frailty and Weibull baseline intensity stands out with a lower AIC and BIC, establishing it as an enhanced and robust framework for capturing underlying patterns in eczema datasets. The estimates within this model include baseline intensity parameters, incidence relative risk for potential covariates, and the frailty's variance. Beyond capturing the effects of risk factors, the frailty's variance indicates the presence of unobserved heterogeneity in recurrent events not accounted for by the risk factors in the model. The extensive evaluation, involving simulation studies and real-world data analysis, underscores the utility and effectiveness of these parametric frailty models, making a significant contribution to the field of survival analysis.

Keywords: Parametric Frailty Models, Baseline Intensity, Model Selection, Recurrent Events, Frailty Distributions

1. Introduction

The event of interest may occur more than once during a study when dealing with recurrent event data. Hospitalizations, asthma attacks in children, cardiovascular diseases, bleedings, and tumors that recur are a few examples of these kinds of incidents. There is a wealth of literature on examining single-variant recurrent event data. Occasionally, during the study, two or more distinct recurrent event types may occur, and those distinct recurrent event types may be correlated with one another. We call this kind of data as multi-type recurrent event data. It might not be enough to conduct independent analyses for every kind of recurrent event while ignoring the interdependence between event types when dealing with multi-type recurrent event data. Events of interest that repeat themselves within a person or a system are referred to as recurrent events. In a number of disciplines, such as epidemiology, engineering, finance, and medical research, these occurrences are frequent (Van Eck, Berkhof, Nicolson, & Sulon, 1996). Recurrent occurrences in the medical domain may include readmissions to the hospital, recurrences of a particular clinical outcome, or disease relapses. When it comes to fields where repeated occurrences are common, analyzing recurrent events is essential to understanding the underlying processes and making well-informed decisions (Ng, Tawiah, McLachlan, & Gopalan, 2023). For the purpose of examining time-to-event data for a single occurrence, conventional survival analysis methods like the Cox proportional hazards model work well. Nevertheless, these approaches frequently fail to capture the correlation between repeated occurrences within the same subject, making them unsuitable for handling recurrent events. Neglecting this correlation may result in erroneous statistical inference and biased parameter estimations. The realization that individuals or entities within a population may display unobserved heterogeneity or frailty that affects the frequency and timing of recurrent events gives rise to the necessity for frailty models. Frailty can be viewed as a latent variable that captures the unmeasured attributes or circumstances that influence a person's likelihood of experiencing an event (Ding, Kuha, & Murphy, 2017). It represents an individual's susceptibility to the event. Therefore, by adding individual-specific random effects, frailty models offer a more complex and realistic way to model recurrent events.

A subset of frailty models known as parametric frailty models treat the frailty term as a random variable with a known distribution and assume a particular parametric form for the baseline hazard function (Balia & Jones, 2007). The baseline hazard and the frailty effect can be modeled in a more flexible and understandable way thanks to this parametric approach (Tarekegn et al., 2020). When there is significant individual heterogeneity in the study population, it is especially helpful, and capturing this heterogeneity is crucial for a thorough analysis (Borenstein, Hedges, Higgins, & Rothstein, 2017). Frailty models are widely used and have implications for many different fields of study. Frailty models have the potential to enhance our knowledge of the variables that impact the course and recurrence of diseases in medical research. These models can improve the long-term prediction of system failures and maintenance requirements in engineering and reliability analysis. Frailty models are useful in social sciences to study patterns of recurrent events or behaviors in human populations (Bedair, Hong, Li, & Al-Khalidi, 2016). Research in many scientific fields frequently focuses on

studying recurrent events, or situations in which a particular event does not occur once but rather repeatedly over time. Such phenomena are difficult to study analytically because the complexities in the data may be too complex for standard statistical models to fully capture. Frailty models have become essential instruments for tackling these issues and improving analysis accuracy (Cui et al., 2008). By including frailty terms which take into consideration individual-specific factors that might affect the recurrence of events these models provide a sophisticated framework (Tawiah, 2019). Frailty models offer a nuanced understanding of the underlying processes governing recurrent events by taking into account unobservable and time-invariant characteristics specific to each individual (Moguilner et al., 2021).

By using this nuanced approach, researchers can better understand the data and produce a more accurate and thorough depiction of the phenomena they are studying. Research findings are more accurate and insightful when frailty terms are incorporated into recurrent event analyses. These variables function as latent variables, symbolizing the unobservable variation among participants in a research sample. Frailty models provide a more accurate representation of the intrinsic variability in recurrent event data by recognizing and accounting for these latent factors. As a result, because the models take into consideration both known and unknown sources of variation, researchers are able to make well-informed decisions and predictions (Woodman & Mangoni, 2023). When working with repeated occurrences, this thorough understanding improves the validity of statistical inferences, allowing researchers to make more reliable conclusions and more precise predictions. Therefore, frailty models represent an important methodological development that gives researchers a potent tool for deciphering the subtleties of recurring events in a variety of study domains (Hao et al., 2021). Within the framework of survival analysis, fragility refers to unrecognized individual variability that influences an individual's susceptibility to recurrent events. This unobserved frailty component accounts for individual-specific traits that influence variability in the risk of experiencing recurrent events, such as genetic predispositions or other unmeasured factors. This idea is integrated in parametric frailty models by adding a random effect term to the hazard function (Rouast, Adam, & Chiong, 2019). This means that the hazard function is now a function of the individual-specific frailty term and the baseline hazard. Frailty is incorporated into the model to acknowledge that some people may be more or less predisposed to experiencing recurrent events, and to account for the correlation among recurrent events within the same individual (Sarker, 2021).

There is a clear and significant gap in the frailty models that are currently available, particularly when taking into account the analysis of recurring events. Our paper aims to bridge this gap by providing a wide range of frailty models that are specifically made to deal with recurrent events. This closes a significant gap in the literature by giving researchers a flexible and unified toolkit for examining time-to-event data when it comes to recurrent events. Our suggested models are rich because they include a variety of frailty distributions, such as positive stable, inverse Gaussian, and gamma distributions. These distributions are selected to allow for a more nuanced depiction of individual variability in the risk of experiencing

recurrent events because they can capture different aspects of frailty. Moreover, our models are flexible and applicable in a variety of scenarios because they are based on five different baseline hazards. A key component of our methodology is the estimation of model parameters, and we take a conservative approach by maximizing the marginal log-likelihood. This approach takes into account data that has been censored to the right or may have been truncated to the left, recognizing the complexity of real-world situations that arise in survival analysis. Our models' usefulness in real-world applications is increased by resolving these issues and making them more capable of producing precise and trustworthy parameter estimations. In this research, model selection a crucial phase in the analysis process is handled methodically. We assess and contrast the performance of various models using the log-likelihood ratio test, AIC, and BIC. This enables us to determine which model fits the data the best using a combination of model parsimony and statistical rigor. By providing a cohesive and all-inclusive collection of parametric frailty models designed especially for recurrent events, the research addresses a crucial gap in survival analysis. We hope to give researchers a strong toolkit for deciphering the complexities of time-to-event data in the presence of recurrent occurrences through a thorough methodology that includes baseline intensity functions, a variety of frailty distributions, and reliable parameter estimation techniques.

The following are the research study's principal contributions:

- Develop a set of parametric frailty models tailored specifically for the analysis of recurrent events.
- Integration of gamma, inverse Gaussian, and positive stable frailty distributions with five distinct baseline intensity functions.
- Utilizing statistical tools to find the best-fitting parametric frailty model, such as the log-likelihood ratio test, the AIC, and the BIC.
- Developing a cohesive and adaptable modeling framework that can capture the subtleties of recurrent event data.

The following is the arrangement of the remained sections in this article: A summary of relevant studies is given in Section 2. The problem statement for the current system is given in Section 3. In Section 4 of the paper, the methodology of parametric frailty models for recurrent events are described. Section 5 introduces the simulation study. The results of the research and the discussion that followed are presented in Section 6. Section 7 discusses the conclusion of the suggested model and its future application.

2. Relative literature

Talebi-Ghane et al. (2021) proposed a joint frailty model with cure fraction for repeated occurrences and death. The death time could be linked to the underlying recurring process, and there is frequently a correlation between the recurrent events' occurrences. Furthermore, there are some circumstances in which some patients may recover. Within the context of this study, "cured" refers to the possibility that some patients will not die from the disease being studied,

nor experience any recurrent events. In order to analyze the recurrent and terminal events, the research suggested a combined frailty model in with the inclusion of cure fraction. Also estimated the impact of covariates that are on the cure rate and both of the aforementioned events simultaneously. The relationship between the recurrences and lifespan times as well as the dependence between the intercurrences were examined because this model had two independent gamma distributed frailties. Using the maximum likelihood approach for a piecewise value and a parameterized base hazard function, the model's variables were estimated. A study was used to assess the method, and real data on surgically treated breast cancer patients were used to illustrate its features. Furthermore, the interpretability of these intricate models may make it challenging to effectively convey results to a wider audience, which includes legislators and healthcare professionals.

Tawiah et al. (2020) proposed a time-varying and multiple levels frailty mixture cure models for recurring events. A frailty mix cure system has been proposed for these data, assuming that each uncured patient's random subject effect remains constant over successive intervals between recurrent events. Assuming a model based on multivariate variable in time frailty with an AR(1) correlation coefficients structure for each uncured patient, the study utilized two new models in a more general setting that address multiple recurrent event data originating from multi-institutional trials. An effective calculation process via an EM-type technique using REML through the generalized linear mixed methodology to address the challenges in parameter estimation caused by these extremely complex correlation structures. The models' performances are evaluated through the presentation of simulation studies. The findings show a strong positive AR (1) correlation between the frailties over successive gap times, suggesting that a constant frailty might not be practical in all circumstances. Computational efficiency issues may arise in the real-world application of these models, especially when working with sizable datasets or attempting real-time analysis in a clinical trial environment.

Tarekegn et al. (2020) proposed approaches to predictive modeling of frailty conditions in aging individuals. By using various machine learning techniques to create prediction models for frailty conditions, this method aims to close this gap. Using a large administrative health database with 1,095,612 people 65 years of age or older, 58 input variables, and 6 output variables, the study uses resampling to address data imbalance. Machine learning algorithms such as artificial neural networks, GP, SVM, RF, Logistic Regression, and Decision Tree have been compared and found to perform differently depending on the type of frailty outcome. Importantly, SVM performs better in predicting urgent hospitalization than ANN in terms of mortality prediction. The results show that model performance varies across different frailty conditions, suggesting that customized decision-support tools could improve the early identification and forecasting of frail older adults. The disadvantage is that the lack of interpretability may make it more difficult for healthcare professionals to use model results to generate practical insights and may restrict their capacity to customize interventions to the unique requirements of elderly patients.

To facilitate inference and prediction, Boom et al. (2022) presented a novel model that treats the number of instances of recur before termination as a random variable. This method accounts for variables like frailty that may affect both survival and recurrence by introducing a dependency between the two. Defined a combined survival and recurrence distribution, adding more dependency via frailty terms. We utilize an autoregressive model in order to capture temporal dependence. The method allowed for data-driven subject clustering while accounting for population heterogeneity through the use of a non-parametric random effects distribution for frailty terms. The model was tested using data on colorectal cancer, comparing its results to those of previous approaches and drawing conclusions about the frequency of recurrent events. The Gibbs sampler makes posterior inference easier by utilizing reversible slice and jump sampling steps.

Examining biological age quantification and death prediction in aging populations has made the frailty index technique a central focus in Moguilner (2021) research. The method, which is based on the development of health deficits, is useful for capturing differences in health status between people of the same age. Nonetheless, the purpose of this study was to look into how mortality prediction is affected when age is included in the FI. However, the study found that not all FI variables were equally important in predicting mortality, with physical function deficits and self-rated health deficits scoring higher. Notably, chronological age was found to be the most significant characteristic. Therefore, even though the FI is still useful, the study emphasizes how important it is to take chronological age into account when interpreting the prognostic implications of a FI. The limitation is that different demographic or cultural contexts may have different factors influencing mortality prediction, so the study's findings and conclusions might not be applicable everywhere.

Ma et al. (2021) developed a zero-inflated generalized combined frailty framework and a sieve maximum probability approach for zero-inflated recurrent events analysis. The model offers a great deal of flexibility by defining different transformation functions to formulate the impact of variables on both recurrences and the terminal event. Furthermore, the unknown cumulative baseline hazard component is approximated using Bernstein polynomials. The estimating process is quick to compute and simple to implement. We perform extensive simulation studies that show the effectiveness of the method in real-world scenarios. In a clinical trial with cardiovascular outcomes, the research finally applied the method to recognize recurrences of myocardial infarction. The method's robustness and applicability in capturing the complexities of real-world data are its drawbacks, which require additional validation through empirical studies and possibly a variety of clinical datasets.

The examined literature lists a number of prevalent flaws in the suggested models for examining death and recurrent events. Due to the complex nature and interpretability problems with the suggested joint frailty models with cure fraction, one significant limitation is the difficulty of effectively communicating results to a wider audience, including legislators and

healthcare professionals. Furthermore, computational efficiency problems can occur, particularly when working with big datasets or trying to do real-time analysis in clinical trial settings, as demonstrated by the multiple levels and time-varying frailty mixture cure models. While providing a variety of insights, the machine learning approaches to predictive modeling of frailty conditions are not easily interpretable, which could impede healthcare providers' capacity to produce useful insights and tailor interventions for senior citizens. Certain models, like the zero-inflated generalized combined frailty framework, show promise and are useful in describing the intricacies of the real world, but they require further validation using a variety of clinical datasets and empirical research.

3. Problem Statement

The insufficiency of current models in capturing the complex dynamics of recurrent events within the framework of survival analysis is the issue addressed by the literature review. While frailty models have been widely used in survival analysis to address over-dispersion and clustering in data, there is still a notable lack of unified parametric frailty models that are specifically designed for recurrent event analysis. The insufficiency of current frailty models may limit their applicability in situations where recurrent events are common by failing to fully capture the complexities involved (Muscedere et al., 2020). This gap impedes progress in comprehending underlying processes and creating customized interventions for researchers and practitioners looking for reliable and adaptable tools to analyze time-to-event data with recurring events. To improve the analytical toolkit available to researchers in the field, it is imperative that parametric frailty models be developed and validated in order to effectively handle recurrent events.

4. Methods

The research methodology section employs parametric frailty models to investigate and model recurrent events through a methodical approach. The dataset, which was acquired from phototherapy outcomes, includes vital patient data such as demographics, medical history, and time-to-event results. The gamma, inverse Gaussian, and positive stable frailty models three parametric frailty models are the mainstays of the methodology. These models offer a sophisticated extension to traditional survival analysis techniques by taking into account unobservable variability within the population. Model adaptability is enhanced by differential evaluations of frailty distributions, which capture subtleties like skewness, temporal fluctuations, and heavy-tailed patterns. Furthermore, baseline intensity functions are essential to comprehending the underlying risk structures. These include Log-Normal, Gompertz, Weibull, Exponential, and Log logistics distributions. AIC, BIC, and log-likelihood ratio tests are among the criteria used in the model selection process to make sure the best fitting and least expensive model is chosen. With the goal of improving predictive modeling in healthcare settings and advancing our understanding of risk factors, this all-encompassing methodology

attempts to decipher complex survival patterns associated with recurring healthcare events. Figure 1 shows the conceptual diagram of the proposed methodology is given below.

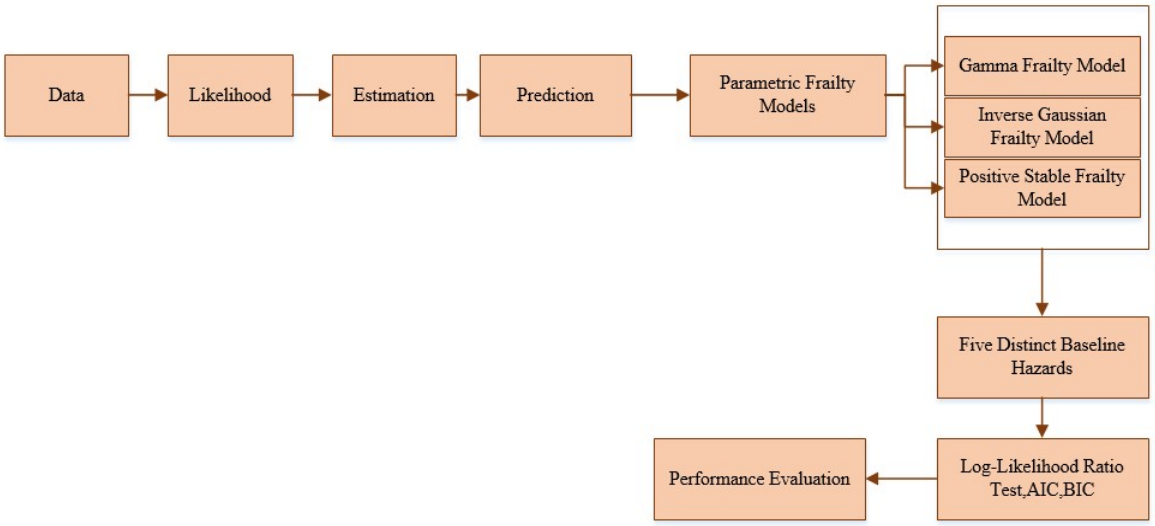

Figure 1: The Block Diagram of the Proposed Methodology

4.1 Data

The pair z_{pq} = = (y_{pq}, δ_{pq}) for event j $\in j_p$= {1,..., n_p } from subject p ∈ I = {1,..., G} is the observation for right-censored survival data, where $y_{p,q}$= min(t_{pq}, c_{pq}) is the minimum between the survival time t_{pq} and the censoring time c_{pq}, and where δ_{pq}= I(t_{pq}≤ c_{pq})) is the event indicator. In the case of gap time, with $y_{pq} = \left(y_{pq2} - y_{ij1}\right)^t$ referring to the start and the end of the interval q. It's possible that covariate data was also gathered; in this instance, z_{pq} = (y_{pq}, δ_{pq} , x_{pq}), where $x_{p,q}$represents the covariate vector for the pq-event . Moreover, truncation time $\tau_p$are collected in the vector τ if left-truncation is also present.

4.2 Frailty Model

The intensity/rate function for a frailty model is

$$\lambda_{pq}\left(t|u_p\right) = u_p\lambda_0(t - y_{pq}\ )\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}(t)_{pq}\ )\ \ for\ y_{pq1} \geq t \leq y_{pq2}\ .\quad (1)$$

Where $\lambda_0(.)$ is the baseline intensity function, $\boldsymbol{X}(t)_{\boldsymbol{pq}}$ is a time-dependent covariate vector associated with the vector of regression parameters $\boldsymbol{\beta}$, and u_p is the frailty term for the p^{th} subject. We assume that the u_p are independently and identically distributed $(i.i.d.)$ from a pre-specified frailty distribution.

4.3 Likelihood

The frailties have been integrated out by averaging the conditional likelihood with respect to the frailty distribution, and in the parametric setting, estimation is based on the marginal likelihood. The marginal log-likelihood of the observed data, z = $\{z_{pq}$;p ∈ I, q ∈ $J_p\}$, can be expressed as follows under the assumptions of non-informative right-censoring and independence between the censoring time and the survival time random variables given the covariate information.

$$l_{marg}(\psi, \beta, \xi;\ z \mid \tau) = \sum_{p=1}^{G} \left\{ \left[\sum_{q=1}^{n_p} \delta_{pq} \left(\log \left(\lambda_0 \left(y_{pq} \right) \right) + x_{pq}^T \beta \right) \right] + \log \left[(-1)^{(d_p)} L^{(d_p)} \left(\sum_{q=1}^{n_p} \Lambda_0 \left(y_{pq} \right) \exp \left(x_{pq}^T \beta \right) \right) \right] - \log \left[\mathrm{L} \left(\sum_{q=1}^{n_p} \Lambda_0 \left(\tau_{pq} \right) \exp \left(x_{pq}^T \beta \right) \right] \right\} \quad (1)$$

with $d_p = \sum_{q=1}^{n_p} \delta_{pq}$ the number of events in the p^{th} subject. λ_0 (y) and Λ_0 $(y_{pq}) = \int_0^{y_{pq2} - y_{pq2}} \lambda_0(s) ds$ are the baseline intensity and baseline cumulative intensity function, respectively. L(q)(·) is the q^{th}derivative of the Laplace transform of the frailty distribution $f(u_p)$ defined as

$$\mathrm{L(s)} = \mathrm{E}\left[\exp(-\mathrm{Us})\right] = \int_0^{\infty} \exp(-u_p s) f(u_p) du_p, \mathrm{s} \geq 0. \quad (2)$$

4.4 Parameters Estimation

Maximizing the marginal log-likelihood in Eqn (1) yields estimates of $\xi = (\psi, \beta, \xi)^t$. The estimated variance-covariance matrix is derived as the inverse of the observed information matrix, evaluated at the parameters estimates $\hat{\boldsymbol{\xi}}$. Standard errors are computed as the square roots of the diagonal elements of the observed information matrix. The detailed formulas for the first- and second-order derivatives of the log-likelihood with respect to the parameters $\boldsymbol{\xi} = (\psi, \beta, \xi)^t$ are provided in Duchateau and Janssen (2008).

4.5 Frailty Prediction

In addition to parameter estimations, it can occasionally be desirable to predict frailties. With z_p and τ_p representing the data and the censoring time for p subject, the frailty term can be predicted using the formula $\widehat{u_p}$ = E $\left(U \mid z_p \tau_p ; \widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\xi}}.\right)$ One can calculate this conditional expectation as

$$\mathrm{E}\ (\mathrm{U} \mid z_p, \tau_p;\ \psi, \beta, \xi) = \frac{L^{(d_p+1)}\left[\sum_{q=1}^{n_p} \Lambda_0(y_{pq}) \exp\left(x_{pq}^T \boldsymbol{\beta}\right)\right]}{L^{(d_p)}\left[\sum_{q=1}^{n_p} \Lambda_0(y_{pq}) exp\ (x_{pq}^T \boldsymbol{\beta})\right]}, \quad (3)$$

which is detailed at [20], along with

$$\mathrm{E}[\mathrm{U}^{\ i} \exp(-\mathrm{Us})] = (-1)^i \mathrm{L}^i\ (\mathrm{s}) \quad (4)$$

4.6 Frailty Distributions

The research introduces a sophisticated method by using parametric frailty models in the context of survival analysis. These models offer a way to account for unobservable variability within a population, which makes them a potent extension of conventional survival analysis techniques (Cui et al., 2008). The fundamental concept is based on adding frailty distributions to the baseline intensity functions, namely the gamma, inverse Gaussian, and positive stable frailty. Frailty can be conceptualized as a random variable that represents individual differences in susceptibilities to the event of interest, capturing unobserved heterogeneity among individuals. Through the incorporation of these frailty distributions, our methodology facilitates a more sophisticated comprehension of survival patterns, accounting for the innate variability found in the population and offering an enhanced examination of the fundamental risk structures.

A positive skewness is introduced by the gamma distribution, which captures situations in which a portion of the population is more susceptible to the event of interest. Conversely, the inverse Gaussian frailty model allows for dynamic modeling of frailty in scenarios where the event's risk fluctuates over time. With its heavy tails, the positive stable frailty distribution is well-suited to containing extreme values and accommodating outliers within the population. The methodology takes into account these various frailty distributions, which not only accepts the complexity of real-world survival data but also offers researchers a sophisticated toolkit to customize their analyses to particular features of the population they are studying (Cui et al., 2008).

In our parametric frailty models, the definition of baseline intensity function is a crucial aspect that serves as the foundation for understanding the fundamental risk of the event of interest. Let $\lambda_0(t; \theta)$ denote the baseline intensity function, where t is the time variable and θ is the vector of parameters associated with the chosen frailty distribution. The subscript 0 emphasizes the baseline intensity functions, acting as a reference before incorporating the impact of frailty.

4.6.1 Gamma Frailty

The gamma frailty model is a parametric method for accounting for unobserved heterogeneity within a population. It introduces a frailty term that follows a gamma distribution (Mazroui et al., 2013). The frailty term is a useful tool when trying to model complex survival data because it acts as a random effect that captures variability not explained by the observed covariates. The gamma distribution was selected for the frailty term due to its versatility in capturing heterogeneity patterns and its mathematical properties. Shape (α) and scale (β) are the two parameters that define the gamma distribution. Within the framework of the gamma frailty model, these parameters are essential in determining how frailty is distributed throughout the population. The distribution's skewness is determined by the shape parameter, which enables it to model both under- and over dispersion. The variability or spread of the frailty values is, however, influenced by the scale parameter. The gamma frailty model offers

an adaptable framework for taking into account varying degrees and kinds of heterogeneity in survival data by varying these parameters.

The gamma frailty model's adaptability makes it a good fit for situations in which the notion that individual risk ratios are constant is dubious. For instance, due to unmeasured factors, patients in medical studies may show different susceptibilities to the same event. Researchers can explicitly model this unobserved heterogeneity using the gamma frailty model, which produces more realistic and accurate survival predictions.

A probability density function-based random variable U ~ Gam*(θ) is known as a gamma frailty term.

$$f(u) = \frac{\theta^{-\frac{1}{\theta}} u^{\frac{1}{\theta}-1} exp\,(-u/\theta)}{\Gamma(1/\theta)}, \theta > 0, \quad (5)$$

where the gamma function is denoted by Γ(•). It is equivalent to a gamma distribution Gam (μ, θ) with identifiability fixed at 1 for the mean μ. Then, its variance is θ. The corresponding Laplace transform is provided in Eqn. (6)

$$L(t) = (1 + \theta t)^{-\frac{1}{\theta}}, \quad t \geq 0, \quad (6)$$

for r ≥ 1,

$$L^{(r)}(t) = (-1)^r (1 + \theta t)^{-r} (\prod_{l=0}^{r-1} 1 + l\theta)\, L(t) \quad (7)$$

As a result, in Eqn. (5), we have

$$log\,\left[(-1)^r L^{(r)}(t)\right] = -(r + \frac{1}{\theta})\, log(1 + \theta t) + \sum_{l=0}^{r-1} log\,(1 + l\theta). \quad (8)$$

In the multivariate case, the gamma distribution, which quantifies the correlation among any two event instances from the same cluster, can be calculated as

$$\tau = \frac{\theta}{\theta + 2} \in (0, 1) \quad (9)$$

4.6.2 Inverse Gaussian Frailty

Among our suggested parametric frailty models, the Inverse Gaussian frailty model is essential. With regard to the frailty distribution, this model presents an original viewpoint by taking the Inverse Gaussian distribution into account. The formulation of the baseline intensity function in this model is complex and aims to capture the interaction between time-dependent risks and the impact of frailty on survival dynamics. A key feature of the baseline intensity function expression is the model's sensitivity to temporal domain variations: the reciprocal of the square root of time (t).

When dealing with situations where the risk of the event of interest is impacted by both time-related factors and unobserved individual-specific characteristics, the Inverse Gaussian frailty model is especially useful.

From a practical standpoint, the Inverse Gaussian frailty model provides a flexible instrument for survival analysis, able to adapt to circumstances in which the continuous intensity function assumption is not relevant. The Inverse Gaussian distribution's flexibility combined with its capacity to capture time-varying frailty improves the model's ability to identify intricate survival patterns. This model can be used by researchers to gain subtle insights into the heterogeneity within populations, which will enhance their knowledge of survival dynamics and make a significant contribution to the larger field of parametric frailty modeling.

The density of the inverse Gaussian frailty distribution IG*(θ) has is gin in Eqn. (10)

$$f(u) = \frac{1}{\sqrt{2\pi\theta}}\, u^{-\frac{3}{2}}\, exp\,(-\frac{(u-1)^2}{2\theta})\,, \theta > 0 \qquad (10)$$

1 and θ, respectively, represent the mean and variance. Regarding the Laplace transform, there is

$$L(t) = exp\left[\frac{1}{\theta}\left(1 - \sqrt{1} + 2\theta t\right)\right], t \geq 0 \qquad (11)$$

Additionally, for r ≥ 1

$$\mathrm{L}^{(\mathrm{r})}(\mathrm{t}) = (-1)^{\mathrm{r}}(2\theta\mathrm{t} + 1)^{\frac{-\mathrm{r}}{2}} \frac{\mathrm{P}_{\mathrm{r}-(\frac{1}{2})}\left[\sqrt{2\theta^{-1}\left(\mathrm{t}+\frac{1}{2\theta}\right)}\right]}{\mathrm{P}_{1/2}\left[\sqrt{2\theta^{-1}\left(\mathrm{t}+\frac{1}{2\theta}\right)}\right]}\mathrm{L}(\mathrm{t}), \qquad (12)$$

where P is the Bessel function,

$$P\gamma(\omega) = \frac{1}{2}\int_0^\infty s^{\gamma-1}\, exp\left[-\frac{\omega}{2}\left(s + \frac{1}{s}\right)\right] ds, \gamma \in R, \omega > 0 \qquad (13)$$

For any distribution for which the moments of U | z_i , τ_i; ψ, β, ξ,, the conditional frailty given the data, are known, the proof of this result outlines a general creative method to obtain the derivatives of the distribution of the Laplace transform

$$\mathrm{P}_{1/2}(\omega) = \sqrt{\frac{\pi}{2\omega}}\exp(-\omega), \qquad (14)$$

we have

$$log\left[(-1)^r L^{(r)}(t)\right] = \frac{-r}{2}\, log(2\theta t + 1) + log\left(P_{r-\left(\frac{1}{2}\right)}(z)\right) - \left[\frac{1}{2}\log\left(\frac{\pi}{2z-z}\right)\right] + \frac{1}{\theta}\left(1 - \sqrt{1+2\theta t}\right), \qquad (15)$$

$$\text{with } z = \sqrt{2\theta^{-1}(t + \frac{1}{2\theta})} \qquad (16)$$

An inverse Gaussian distributed frailty with multivariate data is given in Eqn. (17)

$$\tau = \frac{1}{2} - \frac{1}{\theta} + 2\,\frac{exp(2/\theta)}{\theta^2}\int_{2/\theta}^\infty \frac{exp(-u)}{u}\, du \in (0, 1/2) \qquad (17)$$

4.6.3 Positive Stable Frailty

Among our suggested parametric frailty models, especially designed for recurrent events analysis, the Positive Stable frailty model stands out as a unique and potent component. A strong framework to capture the underlying heterogeneity among subjects is provided by the Positive Stable frailty distribution in the context of recurrent events, where individuals may experience the event of interest more than once (Huang & Liu, 2007). An integral involving the Positive Stable distribution characterizes the baseline intensity function for this model, adding a degree of complexity appropriate for managing recurrent event data. The Positive Stable distribution is a useful tool in scenarios where outliers or high-impact events have a significant impact on the recurrence pattern because of its heavy tails, which enable it to accommodate extreme values.

The study of non-Gaussian and heavy-tailed frailty effects in the context of recurrent events is made possible by the Positive Stable frailty model, which offers a sophisticated insight into the variation in subjects' susceptibility to the event occurring again. The complex relationship between the Positive Stable frailty and the underlying risk structure is captured by the integral representation of the baseline hazard. We enable researchers to identify and interpret intricate patterns of recurrence by integrating this model into our parametric framework, which advances our understanding of the dynamics surrounding repeated events in the study population.

Haugaard presents a family of positive stable distributions with two parameters: an index $\alpha < 1$ and a scale $\delta > 0$. The positively stable frailty distribution PS* (w) with $w = 1 - \alpha$ is obtained by imposing $\delta = \alpha$.

$$f(u) = \frac{-1}{\pi u} \sum_{p=1}^{\infty} \frac{\Gamma(p(1-w)+1)}{p!} (-u^{w-1})^p sin\,(1-w)p\pi, \qquad w\epsilon(0,1) \tag{18}$$

Unlike the probability density function, the corresponding Laplace transform has a very straightforward form.

$$L(t) = exp\,(-t^{1-w}), \quad t \geq 0, \tag{19}$$

$$L^{(r)}(t) = r((1-w)t^{-v})^r / [\sum_{m=0}^{r-1} \Omega_{r,m} t^{-m(1-v)}] L(t), \tag{20}$$

where the polynomials of degree m, denoted $\Omega_{r,m}$,are provided recursively by

$$\Omega_{r,0} = 1,$$

$$\Omega_{r,m} = \Omega_{r-1,m} + \Omega_{r-1,m-1} \left[\frac{r-1}{1-v} - (r-m)\right], m = 1, \ldots, r-2, \tag{21}$$

$$\Omega_{r,r-1} = (1-v)^{1-r} \frac{\Gamma(r-(1-v))}{\Gamma(v)} \tag{22}$$

It follows,

$$log\left[(-1)^r L^{(r)}(t)\right] = r\,[log(1-w) - wlog(t)] + log\,[\sum_{m=0}^{r-1} \Omega_{r,m} t^{-m(1-w)}] - t^{1-w} \tag{23}$$

The Kendall's tau for positive stable distributed frailties with grouped data is

$$\tau = w \in (0,1) \tag{24}$$

4.7 Distributions of Baseline Intensity

Baseline intensity functions are essential for capturing the underlying risk of an event happening at any given time in parametric frailty models for recurrent events. The instantaneous failure rate at time t, in the absence of any additional factors or frailty, is referred to as the "baseline intensity". It stands for the inherent risk that an event will occur at a specific time, independent of personal traits or outside factors. This paper examines five different kinds of baseline intensity functions, each with its own parametric form. These encompass the Log-Normal, Gompertz, Weibull, Exponential, and Log logistics. The Gompertz and Weibull baselines, for instance, add shape and scale parameters to capture various hazard dynamics, whereas the exponential baseline intensity assumes a constant intensity rate over time. The selection of baseline intensity functions takes into account the variety of ways that event risks may appear when frailty is absent. This gives a thorough basis for modeling recurrent events and helps to guide the addition of frailty terms to the suggested parametric models (Munda, Rotolo, & Legrand, 2012).

Table 1: Five Distinct Baseline Intensity Functions

Distribution	$\boldsymbol{\lambda_0(t)}$	$\Lambda_0(t) = \int_0^t \lambda_0(s)ds$	**Parametric Space**
Exponential	λ	λt	$\lambda > 0$
Weibull	$\lambda\, \beta t^{(\beta-1)}$	λt^{β}	$\lambda, \beta > 0$
Gompertz	$\lambda\, [exp\,(\gamma t)]$	$\frac{\lambda}{\gamma}[exp\,(\gamma t) - 1]$	$\lambda, \gamma > 0$
Log-Normal	$\frac{\phi\left(\frac{\log(t)-\mu}{\sigma}\right)}{\sigma t\left[1-\phi\left(\frac{\log(t)-\mu}{\sigma}\right)\right]}$	-log$\left[1-\phi\left(\frac{log(t)-\mu}{\sigma}\right)\right]$	$\mu \epsilon$R, σ>0
Log-Logistics	$\frac{\exp(\alpha)\kappa t^{\kappa-1}}{1+\exp(\alpha)t^{\kappa}}$	$\log[1 + exp(\alpha)t^{\kappa}]$	$\alpha \epsilon$R, $\kappa > 0$

Table 1 shows the parametric distributions for the baseline intensity that are available. The special design of the suggested models for the analysis of recurrent events, which emphasizes an emphasis on comprehending and forecasting the occurrence of repeated events over time, forms the conceptual basis of these models. Within the field of survival analysis, recurring events refer to situations in which subjects encounter a particular event more than once during the course of the study. The models incorporate frailty terms and time-dependent intensity functions to intricately capture the temporal dynamics of event incidences. This design provides a thorough method that goes beyond the traditional study of individual events by enabling a nuanced investigation of the ways in which different factors affect the frequency and timing of recurrent events. The models are especially relevant in healthcare settings, where patients might experience recurrent medical episodes. They are intended to improve the ability to manage healthcare resources and anticipate recurrent event patterns more precisely.

The process of estimating parameters in the models that have been suggested is centred on optimizing the marginal log-likelihood, which represents the total likelihood of observing the provided data in the model that has been assumed. Finding the parameter values that produce the highest likelihood for the observed data depends critically on this optimization. Crucially, the models can handle incomplete data with ease: they can handle data that is potentially left-truncated (observations are available only for events occurring before a specific time) or right-censored (where event times are known only to occur after a certain point). These intricacies are considered during the optimization process, guaranteeing reliable parameter estimation even when there is insufficient data.

Next, the logarithm of the product of these likelihoods for each individual, or the marginal log-likelihood, is maximized. In light of the presumptive model, this maximization looks for parameter values that maximize the likelihood of the observed data.

The models can accommodate data that has been right-censored and may have been left-truncated, demonstrating their flexibility and suitability for real-world scenarios where complete event time information may be missing. This flexibility is especially useful in industries like healthcare, where follow-up times can vary and events happen sometimes. To sum up, the process of estimating the models' parameters, which maximizes the marginal log-likelihood while managing incomplete data, guarantees the models' suitability and resilience in scenarios where event times are observed partially or imprecisely.

4.8 Models Selection Criteria

In order to determine which parametric frailty model best fits the observed data, a number of models are evaluated. The BIC, which penalizes complex models to avoid over fitting, is a crucial factor in model selection, much like the Akaike Information Criterion. However, BIC has a more significant penalty for more parameters, which makes it particularly useful in situations where sample sizes are smaller. The BIC is computed by multiplying the number of parameters by the negative log-likelihood of the model plus half of the log of the sample size. This penalty term discourages the addition of needless parameters that might not have a major impact on the explanatory power of the model, reflecting a stronger preference for parsimonious models. Practically speaking, the BIC is especially helpful when researchers must carefully balance model fit and complexity, especially when there is a lack of data. The BIC formula's penalty term makes sure that models with more parameters are not given preference unless the complexity is justified by the improvement in fit. As a result, models with lower BIC values represent a better balance between simplicity and goodness of fit, which is consistent with the main objective of model selection. The log-likelihood ratio test, AIC, and BIC work together as a complete toolkit during the model selection process. AIC and BIC offer quantitative metrics that direct the model selection process, whereas the log-likelihood ratio test provides a formal statistical comparison of nested models. The model selected is the one that strikes the best balance between accurately capturing the underlying patterns in the data and avoiding needless complexity. A rigorous and well-informed model selection process that takes into account statistical significance as well as the inherent trade-off between model fit and complexity is ensured by this multi-criteria approach.

$$BIC = -2 \cdot log(-likelihood) + pn \cdot ln(sz) \quad (25)$$

- −2·log-likelihood is twice the negative log-likelihood of the model.
- pn is the parameter numbers.
- sz is the sample size.

With smaller sample sizes, the penalty term is especially sensitive to over fitting and provides a stronger penalty for more parameters. It is proportional to the logarithm of the sample size. Lower BIC values indicate models that better balance explaining the data and avoiding needless complexity. The BIC equation, then, quantifies the trade-off between model fit and complexity.

The model complexity and goodness of fit are intended to be balanced by the Akaike Information Criterion (AIC). In order to discourage over fitting, it acts as a tool for model selection by penalizing the addition of excessive parameters. The Eqn. (27) below is used to calculate the AIC:

$$AIC = -2 \cdot log(-likelihood) + 2 \cdot pn \quad (26)$$

The AIC formula combines a penalty term (2·p) proportional to the number of parameters with a term (negative log-likelihood) that assesses how well the model fits the data. Overly complicated models are discouraged by the penalty, which indicates a preference for simpler models that can explain the data adequately. Finding models that balance goodness of fit and model simplicity is the main objective.

Better-fitting models are generally indicated by lower AIC values. In line with the statistical modeling principle of parsimony, the preference for lower AIC values highlights the AIC's function in choosing models that provide a favourable compromise between explaining the observed data and avoiding needless complexity.

5. Simulation Study

In this section, we generate data to assess the practical performance of the developed method in finite samples.

5.1 Simulation Settings

For each subject i we generated the frailties with gamma, Inverse Gaussian, Positive the correspondence distribution parameter to achieve the Kendell taues coefficients. A right-censoring variable ($C_p = 5$) was set at a fixed value in order to have on average a third of censored data. The binary explanatory variables x_{i1}, x_{i2}, were generated from a Bernoulli distribution with $pr = 0.5$. We set $\beta_1 = 1$, and $\beta_2 = -0.5$. Additional information regarding the steps for generating the simulated data can be referenced in Bedair et al. (2016). In the initial two settings, we generated data with a substantial and positive low and moderate dependency between recurrent events (Kendell's tau= 0.10, and 0.5). In the third setting, the frailties were negatively dependent (Kendell's tau=0-.5).

5.2 Simulation Results

Table 2 provides the results for a sample size of $N = 400$, whereas the outcomes for a sample size of $N = 1000$ are summarized in Table 3. The death rate remains around two-thirds, and the average numbers of observed recurrent events per subject are 1.25 and 0.25 in the conducted simulation studies for the respective sample sizes. Approximately 25% and 80% of the subjects did not experience any recurrent events in the two sample sizes, respectively. For each of these sample sizes, we examine low, moderate (positive and negative) frailty association scenarios.

Upon careful scrutiny, it becomes apparent that, in all three settings, parameter estimates for our proposed models (gamma, inverse Gaussian, and Positive stable) are precisely estimated. The empirical biases of the estimates are negligible, with only minor biases (around 5%) observed in all settings. The coverage probabilities closely adhere to the nominal level of 95%. The simulation study shows that the baseline intensity parameters, regression coefficients and variance parameters from the proposed method were well estimated with small MSE values at all settings which implies the performance of the estimates is good with moderate samples.

Table2: Simulation outcomes for frailty distributions (gamma, inverse Gaussian, and positive stable) with a Weibull baseline intensity, considering a total of 400 subjects and 500 events.

	Gamma			Inverse Gaussian			Positive stable		
Parameter	|Bias|	MSE	CP	|Bias|	MSE	CP	|Bias|	MSE	CP
Theta =0.4	0.0102	0.0030	0.9553	0.0301	0.0200	0.9541	0.0314	0.0075	0.9516
Lambda =0.25	0.0063	0.0018	0.9558	0.0280	0.0126	0.9527	0.0310	0.0056	0.9498
beta1=1	0.0262	0.0014	0.9547	0.0278	0.0096	0.9528	0.0284	0.1153	0.9435
beta2=-0.5	0.0128	0.0282	0.9552	0.0323	0.0483	0.9492	0.0308	0.0491	0.9458
Kendall's Tau= 0.10	0.0028	0.0121	0.953	0.0270	0.0207	0.9504	0.0293	0.0303	0.9508
Setting 2									
Theta =0.4	0.0107	0.0132	0.9495	0.0304	0.0224	0.9492	0.0300	0.0096	0.951
Lambda =0.25	0.0071	0.0098	0.9517	0.0264	0.0472	0.9532	0.0232	0.0139	0.9516
beta1=1	0.0281	0.0488	0.9518	0.0261	0.0200	0.9535	0.0265	0.0065	0.9448
beta2=-0.5	0.0159	0.0209	0.9472	0.0239	0.0127	0.9557	0.0242	0.0042	0.946
Kendell's tau= 0.5	0.0153	0.0131	0.9526	0.0245	0.0092	0.9551	0.0271	0.0030	0.9451
Setting 3									
Theta =0.4	0.0117	0.0120	0.948	0.0285	0.0074	0.9494	0.0296	0.0778	0.9499
Lambda =0.25	0.0078	0.0174	0.9485	0.0570	0.0031	0.9565	0.0291	0.0305	0.9494
beta1=1	0.0376	0.0082	0.9511	0.0439	0.0019	0.9531	0.0228	0.0192	0.9565
beta2=-0.5	0.0176	0.0052	0.9485	0.0390	0.0014	0.9556	0.0262	0.0139	0.9531
Kendell's tau= -0.5	0.0226	0.0038	0.9502	0.0342	0.0293	0.9526	0.0239	0.0719	0.9556

Table3: Simulation results for the gamma, inverse Gaussian, and positive stable frailty distributions and Weibull bassline intensity for number of subject n=1000 and number if events=250.

frailty distribution	Gamma			Inverse Gaussian			Positive stable		
Parameter	|Bias|	MSE	CP	|Bias|	MSE	CP	|Bias|	MSE	CP
Setting1									
Theta =0.4	0.0114	0.0121	0.9525	0.0300	0.0304	0.9526	0.0248	0.0309	0.9532
Lambda =0.25	0.0071	0.0075	0.9525	0.0321	0.0193	0.9547	0.0295	0.0227	0.9535
beta1=1	0.0324	0.0056	0.9486	0.0261	0.0140	0.9523	0.0295	0.0507	0.9557
beta2=-0.5	0.0147	0.1185	0.9515	0.0277	0.0744	0.9493	0.0291	0.0207	0.9551
Kendall's Tau= 0.10	0.0029	0.0492	0.9522	0.0286	0.0309	0.9522	0.0265	0.0132	0.948
Setting 2									
Theta =0.4	0.0113	0.0199	0.9495	0.0303	0.0092	0.9492	0.0244	0.0234	0.9515
Lambda =0.25	0.0066	0.0138	0.9474	0.0256	0.0239	0.9541	0.0301	0.0744	0.9520
beta1=1	0.0287	0.1403	0.9534	0.0197	0.0099	0.9601	0.0289	0.0311	0.9456
beta2=-0.5	0.0159	0.0521	0.9519	0.0258	0.0061	0.9537	0.0265	0.0191	0.9535
Kendell's tau= 0.5	0.0144	0.0324	0.9509	0.0260	0.0045	0.9536	0.0291	0.0139	0.9456
Setting 3									
Theta =0.4	0.0111	0.0037	0.9548	0.0281	0.0250	0.9552	0.0271	0.0082	0.9525
Lambda =0.25	0.0069	0.0023	0.9499	0.0276	0.0158	0.9495	0.0271	0.0052	0.9525
beta1=1	0.0252	0.0017	0.9501	0.0339	0.0120	0.9509	0.0307	0.0038	0.9486
beta2=-0.5	-0.0129	0.0352	0.9504	0.0262	0.0604	0.9531	0.0281	0.0972	0.9515
Kendell's tau= -0.75	-0.0248	0.0151	0.9531	0.0260	0.0259	0.9505	0.0274	0.0381	0.9522

6. Real Application and Results

This section reports on our methodical investigation and modelling of gap time between recurrent events in healthcare using parametric frailty models. Utilizing data simulated, with exactly the same characteristics as the real data used in the photherapy dataset that could not be made publicly available. We conducted a retrospective assessment of the response to phototherapy for atopic eczema, focusing on the types of phototherapy, namely NB-UVB (A), UVA1 (C), or PUVA (B). We analyzed 1532 (88%) A, 83 (4%) B, and 129 (6%) Courses administered to 1303 patients. These courses were distributed across four units with A and B used in all treatment units, while C was exclusively available in unit (A).

We categorized outcomes recorded as a favourable outcome and not favorable. The primary outcome measure assessed was the probability of a favorable outcome in relation to the number of treatments per course. Various covariates (refer to Table 1) potentially influencing treatment efficacy were examined. We employed chi-square and t-student tests for comparing proportions of discrete covariates and means of continuous covariates, respectively. To compare "survival curves" we utilized the log-rank test. The final risk prediction model with multi-variables was established through a backward selection process, commencing with the full model encompassing all covariates. The iteration continued until all covariates in the model achieved statistical significance at the 0.05 level. The frailty models, which include gamma, inverse Gaussian, and positive stable, enhanced by various baseline intensity functions, including Log-Normal, Gompertz, Weibull, Exponential, and Log logistics

distributions, offer a comprehensive comprehension of the survival patterns linked to recurrent medical procedures. The selection of the most suitable frailty and models was based on a comparison of the Akaike Information Criterion (AIC) and the Bayesian. By following this model selection procedure, we determine which model is the most cost-effective and appropriate for our dataset. The talk that follows explores the main findings and provides information about how different frailty distributions and baseline intensity functions can be interpreted and adjusted, which enhances our understanding of the complex dynamics that underlie recurring healthcare events.

6.1 Participants

The majority of Participants, comprising 679 (74%), underwent a single course, 175 (17%) had two courses, 46 (5%) had three courses, and only 37 (4%) underwent more than three courses. The primary outcomes are summarized in Table 4. Across A, B, C, and D treatment centers, a total of 763 (59%), 272 (20%), 138 (10%), and 130 (9.9%) courses were conducted, respectively, with corresponding proportions of favorable outcomes at 63.3%, 74.3%, 63%, and 69.2%. Additional features for participants are presented in Table 4.

Table 4. Summary of data on most important variables assessed for possible effects on treatment outcome when treating eczema with the phototherapies.

	Courses frequency	Distribution of courses by the outcomes		P. value
Studied Covariates	1303 (%)	Not good outcome 441 (33%)	Good outcome 862 (66%)	
Treatments per course	(27.3, 0.42) d	(17, 0.654)d	(32.26, 0.46)d	0.001a 0.001b
High	656 (50.00)	90 (13.72)	566(81.28)	
Low	657 (50.00)	351 (53.42)	296(46.58)	
Gender				0.925b 0.444c
Female	689 (53.00)	234 (33.96)	455 (66.04)	
Male	614 (47.00)	207 (33.71)	407 (66.29)	
Course age	(34.12, 0.44) d	(32.63, 0.76)d	(34.88, 0.55)d	0.020b 0.728c
High	653 (50.00)	202(30.93)	451 (69.07)	
Low	620 (50.00)	239 (36.77)	411 (31.54)	
Erythema				0.004b 0.597c
No	1261 (97.00)	435 (34.50)	826 (65.50)	
yes	42 (3.00)	6 (14.29)	36 (85.71)	
Skin Type				0.993b 0.001c
I	875 (67.00)	297 (33.94)	578 (66.33)	
II	401 (30.00)	135 (33.36)	266 (66.44)	
III	27 (03.00)	9 (33.84)	18 (66.67)	
Treatment Centre				0.007b 0.001c
I	763 (59.00)	280 (36.70)	483 (63.30)	
II	272 (20.00)	70 (25.74)	202 (74.26)	
III	138 (11.00)	51 (36.96)	87 (63.04)	

IV	130 (10.00)	40 (30.77)	90 (69.23)	
Treatment Type				0.005b 0001c
A	61(04.00)	22 (36.07)	39 (63.93)	
B	86(07.00)	43 (50.00)	43 (50.00)	
C	1156(89.00)	376 (32.53)	780 (67.47)	
Course duration (mean, SD)	(86.42 ,1.50)d	(65.37, 2.48)d	(97.19, 1.77)d	0.001b 0.001c
High	656 (50.50)	128 (19.00)	528 (80.49)	
Low	647 (49.50)	313 (48.38)	334 (51.62)	
Treatment A doses				0.001b 0.001c
High	576 (44.21)	53 (9.20)	523 (90.8)	
Low	575 (44.13)	320 (55.65)	255 (44.35)	
No A treatment	152 (11.67)	68 (44.74)	84 (55.25)	
Treatment B doses				0.218b 0.237c
High	31 (2.38)	8 (25.81)	23 (74.19)	
Low	30 (2.30)	14 (46.67)	16 (53.33)	
No B treatment	1242 (95.32)	419 (33.74)	823 (66.26)	
Treatment C doses				0.001b 0.634c
High	43 (3.30)	12 (27.91)	31 (72.09)	
Low	43 (3.30)	31 (72.09)	12 (27.91)	
No C	1217 (93.40)	398 (32.70)	819 (67.30)	
Cumulative treatment A				0.001b 0.001c
High ≥ 48m	654 (50.19)	172 (26.30)	482 (73.70)	
Low	642 (49.27)	266 (41.43)	376 (58.57)	
No A treatment	7 (0.54)	3 (42.86)	4 (0.31)	
Cumulative treatment B				0.830a 0.095b
High	185 (14.20)	65 (35.14)	120 (64.86)	
Low	418 (32.08)	137 (32.78)	281 (67.22)	
No B treatment	700 (53.72)	239 (34.14)	461 (65.85)	
Cumulative treatment C				0.001a 0.001b
High	99 (7.6)	35 (35.35)	64 (64.65)	
Low	107 (8.2)	60 (56.07)	47 (43.93)	
No C treatment	1097 (84.2)	346 (31.54)	751 (68.46)	

(*) median, (a) t-test, (b) chi-square test, (c) log-rank test, (m) median, (d) (mean, SD).

6.2 Model Selection Criteria and the Frailty Variance

Table 5 provides a summary of the AIC and BIC values for all 15 candidate models considered in the analysis. Researchers can utilize this table to compare and select the most appropriate distribution, considering the balance between model fit and complexity, as reflected in the AIC and BIC values. In this specific application, the Weibull baseline appears to be a promising candidate. The Weibull baseline with a gamma frailty model exhibited the highest log-likelihood value and the lowest AIC and BIC values, indicating a superior fit to the data compared to other frailty models. In this model, the frailty's variance was estimated to be 1.261 (95% CI 0.85–1.93), implying the existence of unexplained heterogeneity in recurrent events that could not be elucidated by the independent variables in the model.

Table 5: Model Comparison of the Proposed Method

	Gamma		Inverse Gaussian		Positive Stable	
	AIC	BIC	AIC	BIC	AIC	BIC
exponential	681.120	692.889	690.160	702.047	699.373	711.318
Weibull	681.041	690.657	689.213	698.722	699.321	708.877
gompertz	683.261	695.032	691.253	703.140	701.371	713.316
loglogistic	692.036	703.805	698.979	710.866	702.841	714.787
lognormal	685.637	697.407	692.780	704.667	697.479	709.424

6.3 Results of Frailty Models

Table 6 presents the results for applying the final parametric frailty model with gamma frailty and Weibull baseline intensity using the number of treatment courses as a time scale. It is noteworthy that treatment B exhibited a higher probability of a positive outcome compared to other phototherapies, even though it is typically used as a second- or third-line treatment. Among the four phototherapy units, one (II) exhibited better outcomes, possibly attributed to its status as the second-longest established unit, with center I being the longest-established and the sole unit offering treatment C, a treatment reserved for challenging cases.

A recorded painful erythema was associated with more favorable outcomes, although this effect disappeared when considering the number of treatments, suggesting that erythema might serve as a marker for receiving sufficient treatments. Older patients appeared to fare better until accounting for the number of treatments, possibly indicating a lower likelihood of early discontinuation.

Lower cumulative exposures to treatment A, B, and C were linked to better responses, possibly due to continued attempts with treatments for those showing poor initial response or a potential tachyphylaxis effect with reduced efficacy over repeated courses.

In this population with low sun-reactive skin phototypes (I to III), those with phototype II were slightly more likely to respond, and although not statistically significant, those with phototype III may have performed better. While skin phototype is loosely related to minimal erythemal dose, which determines starting doses, it is associated with tolerance development.

Table 6. Relative risks (RR) with 95% confidence intervals (CI) for potential risk factors, adjusted for age and sex variables, from gamma frailty models with a Weibull baseline intensity for the current dose.

Studied Covariates	**P. value**	**95% CI**
Skin Type	0.001	
II vs I		1.33[1.09, 1.62]
III vs I		1.47[0.77, 2.80]
Treatment Centre	0.001	
II vs I		2.03[1.64, 2.51]
III vs I		1.28[0.95, 1.72]
IV vs I		1.11[0.84, 1.47]
Treatment Type	0001	

B vs A		0.19[0.11, 0.33]
C vs A		0.29[0.19, 0.43]
Course duration (mean, SD)	0.001	
Low vs High		11.13[8.91, 13.90]
Treatment A doses	0.001	
Low vs High		4.09[3.23, 5.19]
Cumulative A treatment	0.001	
Low vs High		2.07[1.73, 2.47]
Cumulative C treatment	0.001	
Low vs High		2.47[1.34, 4.54]

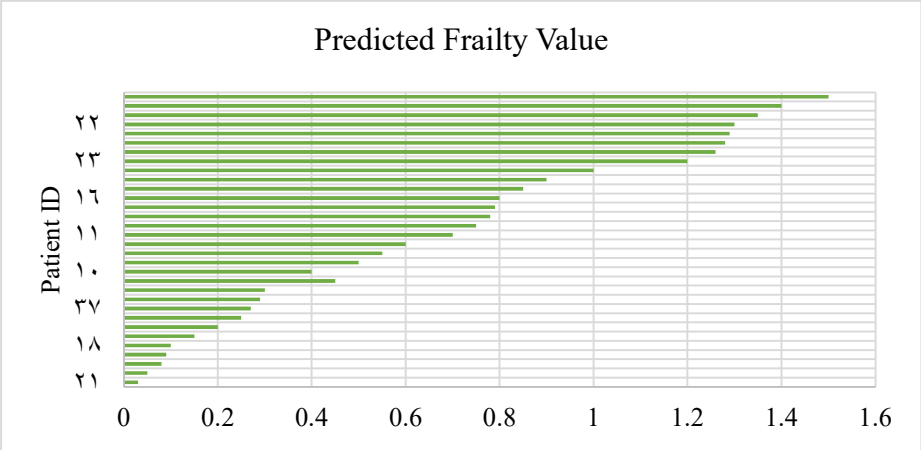

Figure 2: Graphical Illustration of Predicted Frailty Value

Figure 2 illustrates the graphical representation of predicted frailty values for a subset of patients. Each line depicted in the graph may represent a unique participant, with the lines likely reflecting the predicted frailty values derived from a gamma frailty model.

Table 7: Comparison Table of Proposed Model with other Methods

Metrics	**Proposed Parametric Frailty Models**	**Cox Proportional-Hazards Model**	**Kaplan-Meier Estimator**
Model Flexibility	High	Semi-Parametric	Non-Parametric
Handling of Covariates	Yes	Yes	Limited
Assumption about Hazard Shape	Flexible	Proportional	Non-Parametric
Frailty Distributions	Gamma, Inverse Gaussian, Positive Stable	N/A	N/A
Performance Metrics (AIC, BIC, Log-Likelihood)	Available	Available	Not Applicable
Survival Curve Estimation	Yes	No	Yes
Handling of Censored Data	Yes	Yes	Yes
Interpretability	Dependent on chosen frailty distribution	Hazard ratios provided	Visual survival curves

The Cox Proportional-Hazards Model, the Kaplan-Meier Estimator, and the proposed parametric frailty models are just a few of the survival analysis models whose salient features

are highlighted in the Table 3. The suggested parametric frailty models exhibit great adaptability, supporting a range of frailty distributions, and provide performance measures for model assessment, including AIC and BIC.

7. Discussion and Conclusion

We introduced parametric frailty models to investigate relevant risk factors associated with recurring outcome events in a phototherapy study. The proposed models offer flexibility, featuring five distributions for the baseline intensity and three frailty distributions. Parameter estimation is achieved by maximizing the marginal log-likelihood. Emphasizing the importance of performance metrics in model evaluation, particularly AIC and BIC, our study guides researchers in selecting the most suitable model, considering the delicate balance between complexity and goodness of fit. The presented models not only shed light on the versatility of parametric frailty models, showcasing their compatibility with various frailty distributions but also provide performance metrics for a thorough model assessment. Collectively, these elements enhance the analytical framework, aiding in the identification of optimal models and a deeper understanding of survival patterns associated with recurring events in various medical applications.

The models' suitability for use in future projects will be improved by additional investigation and validation across a variety of datasets and disciplines. Including time-varying covariates and integrating machine learning techniques offer viable ways to improve the predictive supremacy of the models. The suggested parametric frailty models can be improved upon and expanded upon thanks to the ongoing development of healthcare data analytics, which will ultimately improve patient care and personalized medicine. In conclusion, this study presents and clarifies the use of parametric frailty models in the context of recurrent event data analysis, specifically in the medical field. The suggested models offer a sophisticated framework for capturing unobservable heterogeneity and improving the understanding of survival patterns. They incorporate various frailty distributions, including gamma, inverse Gaussian, and positive stable. The models' adaptability is enhanced by the assessment of differential frailty distributions, which allows researchers to customize analyses to particular population characteristics. When combined with survival curve estimation, performance metrics like AIC and BIC allow for a thorough evaluation of the effectiveness of the model.

In summary, we implemented a series of parametric frailty models to analyze gap time of recurrent events. The frailty model indicated the presence of unobserved subject-specific risk factors in the study, even after accounting for all the known risk factors in the model. The risk prediction model, focusing on recurrent events, incorporates more information about the occurrence of an event compared to a model based solely on the data from the first event. These risk prediction models can categorize participants into different risk levels, aiding in the targeted application of preventive therapies for recurrent events. Improving the models' relevance for future projects involves conducting additional investigation and validation across a range of datasets and applications, encompassing fields such as medicine, engineering, business, and social sciences, among others.

References

Balia, S., & Jones, A. M. (2007). Unravelling the influence of smoking initiation and cessation on premature mortality using a common latent factor model. HEDG, Department of Economics, University of York.

Bedair, K. F., Hong, Y., Li, J., & Al-Khalidi, H. R. (2016). Multivariate frailty models for multi-type recurrent event data and its application to cancer prevention trial. *Computational Statistics & Data Analysis*, 101, 161-173.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, 8(1), 5-18.

Cui, J., Field, P., Cai, B., & Snyder, J. (2008). Parametric conditional frailty models for recurrent cardiovascular events in the lipid study. *Clinical Trials*, 5(6), 565-574.

Ding, Y. Y., Kuha, J., & Murphy, M. (2017). Multidimensional predictors of physical frailty in older people: identifying how and for whom they exert their effects. *Biogerontology*, 18, 237-252.

Duchateau, L., & Janssen, P. (2008). *The Frailty Model*. Springer.

Hao, L., Kim, J., Kwon, S., & Ha, I. D. (2021). Deep learning-based survival analysis for high-dimensional survival data. *Mathematics*, 9(11), 1244.

Huang, X., & Liu, L. (2007). A joint frailty model for survival and gap times between recurrent events. *Biometrics*, 63(2), 389-397.

Ma, C., Hu, T., & Lin, Z. (2021). Semiparametric analysis of zero-inflated recurrent events with a terminal event. *Statistics in Medicine*, 40(18), 4053-4067.

Mazroui, Y., Mathoulin-Pélissier, S., MacGrogan, G., Brouste, V., & Rondeau, V. (2013). Multivariate frailty models for two types of recurrent events with a dependent terminal event: Application to breast cancer data. *Biometrical Journal*, 55(6), 866-884.

Moguilner, S., Knight, S. P., Davis, J. R., O'Halloran, A. M., Kenny, R. A., & Romero-Ortuno, R. (2021). The importance of age in the prediction of mortality by a frailty index: A machine learning approach in the Irish longitudinal study on ageing. *Geriatrics*, 6(3), 84.

Munda, M., Rotolo, F., & Legrand, C. (2012). parfm: Parametric frailty models in R. *Journal of Statistical Software*, 51, 1-20.

Muscedere, J., et al. (2020). Moving towards common data elements and core outcome measures in frailty research. *The Journal of Frailty & Aging*, 9, 14-22.

Ng, S. K., Tawiah, R., McLachlan, G. J., & Gopalan, V. (2023). Joint frailty modeling of time-to-event data to elicit the evolution pathway of events: a generalized linear mixed model approach. *Biostatistics*, 24(1), 108-123.

Rouast, P. V., Adam, M. T. P., & Chiong, R. (2019). Deep learning for human affect recognition: Insights and new developments. *IEEE Transactions on Affective Computing*, 12(2), 524-543.

Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications, and research directions. *SN Computer Science*, 2(6), 420.

Talebi-Ghane, E., Baghestani, A., Zayeri, F., Rondeau, V., & Akhavan, A. (2021). Joint frailty model for recurrent events and death in the presence of cure fraction: Application to breast cancer data. *Biometrical Journal*, 63(4), 725-744.

Tarekegn, A., Ricceri, F., Costa, G., Ferracin, E., & Giacobini, M. (2020). Predictive modeling for frailty conditions in elderly people: machine learning approaches. *JMIR Medical Informatics*, 8(6), e16678.

Tawiah, R. (2019). Frailty Models for the Analysis of Recurrent Event Data in Studies of Chronic Diseases. Griffith University.

Tawiah, R., McLachlan, G. J., & Ng, S. K. (2020). Mixture cure models with time-varying and multilevel frailties for recurrent event data. *Statistical Methods in Medical Research*, 29(5), 1368-1385.

Van den Boom, W., De Iorio, M., & Tallarita, M. (2022). Bayesian inference on the number of recurrent events: A joint model of recurrence and survival. *Statistical Methods in Medical Research*, 31(1), 139-153.

Van Eck, M., Berkhof, J., Nicolson, N., & Sulon, J. (1996). The effects of perceived stress, traits, mood states, and stressful daily events on salivary cortisol. *Psychosomatic Medicine*, 58(5), 447-458.

Woodman, R. J., & Mangoni, A. A. (2023). A comprehensive review of machine learning algorithms and their application in geriatric medicine: Present and future. *Aging Clinical and Experimental Research*, 35(11), 2363-2397.

Xiao, S., Yan, J., Farajtabar, M., Song, L., Yang, X., & Zha, H. (2017). Joint modeling of event sequence and time series with attentional twin recurrent neural networks. *arXiv preprint arXiv:1703.08524*.