

TUTORIAL

Finding the right hazard function for time-to-event modeling: A tutorial and Shiny application

Rob C. Van Wijk  | Ulrika S. H. Simonsson 

Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden

Correspondence

Ulrika S. H. Simonsson, Department of Pharmaceutical Biosciences, Uppsala University, Box 591, 75124 Uppsala, Sweden.

Email: ulrika.simonsson@farmbio.uu.se

Funding information

No funding was received for this work.

Abstract

Parametric time-to-event analysis is an important pharmacometric method to predict the probability of an event up until a certain time as a function of covariates and/or drug exposure. Modeling is performed at the level of the hazard function describing the instantaneous rate of an event occurring at that timepoint. We give an overview of the parametric time-to-event analysis starting with graphical exploration by Kaplan–Meier plotting for the event data including censoring and nonparametric hazard estimators such as the kernel-based visual hazard comparison for the underlying hazard. The most common hazard functions including the exponential, Gompertz, Weibull, log-normal, log-logistic, and circadian functions are described in detail. A Shiny application was developed to graphically guide the modeler which of the most common hazard functions presents a similar shape compared to the data in order to guide which hazard functions to test in the parametric time-to-event analysis. For the chosen hazard function(s), the Shiny application can additionally be used to explore corresponding parameter values to inform on suitable initial estimates for parametric modeling as well as on possible covariate or treatment relationships to certain parameters. Moreover, it can be used for the dissemination of results as well as communication, training, and workshops on time-to-event analysis. By guiding the modeler on which functions and what parameter values to test and compare as well as to assist in dissemination, the Shiny application developed here greatly supports the modeler in complicated parametric time-to-event modeling.

INTRODUCTION

Time-to-event analysis is an important tool in the pharmacometric toolbox to analyze event-type data and predict the probability of an event up until at a certain time possibly accounting for covariates and/or treatment metrics. Event-type data are one of the closest representations of

clinical outcomes experienced by patients and therefore relevant and informative for the main objective of drug development. An event is any defined and delimited occurrence or experience at a specific timepoint, for example, a seizure, relapse or exacerbation, or death. Time-to-event analysis not only considers the occurrence of an event but also its timing. Time-to-event analysis is therefore favored

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

over other analysis techniques such as logistic regression because of its incorporation of the event time as outcome. In addition, time-to-event analysis handles censoring of data appropriately.^{1,2} Censoring occurs when the event is not observed because it is outside of the observation time (left- or right-censoring) or between two observations (interval censoring). In clinical trials, this can happen because of end of trial or follow-up, lost to follow-up, or withdrawal from the study. Individuals are commonly right-censored, that is, the event time is larger than the censor time. Left-censoring, when the event time is less than or equal to the censor time, is less common, whereas interval-censoring, when the event time is within a known time interval between two observations, occurs more often. It is important to be aware of competing risks for the event under analysis when considering censoring,^{3,4} but this is outside of the scope of this tutorial.

Because of the special nature of the data, time-to-event analysis or survival analysis relies on two main functions. The two main functions in time-to-event analysis are the survival (S) function, quantifying an individual's probability of reaching a timepoint without an event, and the hazard (h) function, quantifying the instantaneous rate of the event happening at that timepoint. Although the survival function is bounded to [0,1], the hazard function's lower boundary is 0, but it has no upper boundary ($[0, \infty)$). The survival function relates to the hazard function as shown in Equation (1), exponentiating the negative cumulative hazard function.¹

$$S(t) = e^{-\int_0^t h(t) dt} \quad (1)$$

The product of the hazard function and the survival function yields the probability density function of the event times.

$$f(t) = S(t) \cdot h(t) \quad (2)$$

Analysis of event time data can be done empirically, semiparametrically, and parametrically. The simplest approach is empirical and descriptive in nature. The nonparametric Kaplan–Meier estimator is a common example.⁵ The impact of predictive covariates such as demographics or treatment group on the event time as outcome has to be assessed by analysis of the corresponding subset of the data. Semiparametric analyses such as the Cox proportional hazard model⁶ do not define the baseline hazard function and only define the event time in terms of covariates and corresponding estimated coefficient/parameter, from which a hazard ratio can be calculated. Although semiparametric time-to-event analysis does not require assuming a hazard function, other assumptions have to be made on, for example, proportional hazards, proportional odds, or accelerated failure time where the hazards, odds, or failure times

are assumed to be related to each other by a constant factor such as the hazard ratio.⁷ In addition, semiparametric analyses are limited to the observed timepoints in the data and will have output that corresponds to those timepoints, such as the stepwise Kaplan–Meier curve. In contrast, a fully parametric approach defines the underlying baseline hazard function in addition to the covariate coefficient and the baseline hazard function and can be used for simulations in contrast to semiparametric approaches. Therefore, it is a more complete method to characterize the survival and hazard functions. A comprehensive introduction to time-to-event or survival analysis can be found elsewhere¹ as well as a tutorial to its application on pharmacometrics.⁸ To build on that tutorial, the scope of this work is describing the most commonly used hazard functions in parametric time-to-event pharmacometric analyses and supporting model development using those hazard functions.

The hazard function characterizing the outcome distribution is at the core of parametric time-to-event analysis. Especially when analytical solutions are not feasible, for example, with time-varying components, modeling takes place at the hazard function level. Also, coefficients quantifying the effect of covariates and treatment are modeled at the hazard level. It is therefore convenient to graphically explore the observed data and model diagnostics at the hazard level. Recent methods have been proposed to characterize the shape of the hazard profile to assist in time-to-event model development. These nonparametric methods use estimators for the hazard based on histogram or smoothing functions.^{9–12} The hazard-based visual predictive check (VPC) estimates the hazard over time based on the observed data nonparametrically using a binned hazard estimator or a kernel method with either local or global bandwidth.¹¹ The nonparametric hazard estimates reflecting the observed data are subsequently overlaid with (parametric) simulations from the model to diagnose the model fit. The kernel-based visual hazard comparison (kbVHC) in contrast is a simulation-free method that uses the nonparametric kernel estimator of the hazard to calculate a 95% confidence interval of the hazard over time based on the observed data.¹² The parametric model-predicted hazard over time can subsequently be overlaid with the confidence interval reflecting the observed data. Both methods have been proposed as model diagnostics; however, they can also be used to explore the (raw) data graphically as an important first step in pharmacometric model development. Here, we use the kbVHC because its confidence interval reflects the variability in the observed data, but the hazard-based VPC can be used interchangeably.

Theoretically, any user-defined function can be used to describe the hazard, and hazard functions can range from a simple, one-parameter constant to multiparameter,

nonmonotonic, or periodic functions. The choice of an appropriate hazard function can be challenging. Also, more complex hazard functions can be sensitive to the initial estimates for the parameters provided to the estimation algorithm. In addition, as time-to-event analyses are very often collaborations with clinicians or other scientists with less experience in modeling, it would be helpful to visualize hazard functions to make the abstract concept more graspable. The objective of this tutorial is therefore to support time-to-event model building by developing a Shiny application to find the right hazard function(s), including their appropriate initial estimates, to explore parameter sensitivity to inform on, for example, covariate or treatment relationships, and to assist in dissemination.

TIME-TO-EVENT ANALYSIS WORKFLOW

The workflow of a typical time-to-event analysis is shown in Figure 1. Generally, the first step in data analysis is exploration of the data either graphically or numerically. One way to graphically explore the observed individual event times is by showing the follow-up time including occurring event(s) per subject over time (Figure 2a). In addition, the survival probability over time in the population including censoring can be explored by a Kaplan–Meier plot (Figure 2b). A more informative way to graphically explore the observed data regarding the underlying hazard is by using a nonparametric hazard estimation methods based on the histogram or smoothing functions mentioned previously (Figure 2c).^{11,12} Similar to, for example, assessing one- or two-compartmental behavior in pharmacokinetic modeling by log-linear graphics, the nonparametric hazard estimation informs the modeler

of the shape of the hazard over time and which parametric functions to test. Thus, a graphical exploration of the hazard can be taken forward to the choice of the hazard functions to test which shape corresponds to the hazard estimation most. In the case of Figure 2c, the hazard shows an increasing and subsequently decreasing pattern over time, so hazard functions capable of that dynamic, such as log-normal or log-logistic, should be selected to be included in model development. The most common hazard functions are discussed in more detail in the “Hazard Functions for Parametric Time-To-Event Analysis” section. To start model development for the selected hazard functions, the initial estimates for the hazard function parameters need to be chosen, which can be challenging for abstract and less-intuitive hazard functions. Here, a Shiny application was developed that can be used to that aim to explore which parameter values of, for example, the log-normal or log-logistic functions lead to a shape similar to that of the nonparametric hazard estimation (Figure 2c). Similarly, the sensitivity of the hazard over time to certain parameters can be explored. This can inform the modeler on the modeling strategy, which parameter, for example, might be best to start testing covariates or treatment effects on, especially with a nonparametric hazard estimation of the corresponding subset of the data (e.g., male/female, treatment/placebo). This is discussed in more detail in the “Exploration of Hazard Function Parameter Values” section. With the right hazard function(s) and corresponding initial estimates to test, model development using designated nonlinear mixed-effects modeling software can be performed in a well-informed manner. Finally, the results of the analysis will be communicated and disseminated, for which the developed Shiny application can be used as well (“Dissemination Using the Shiny Application” section).

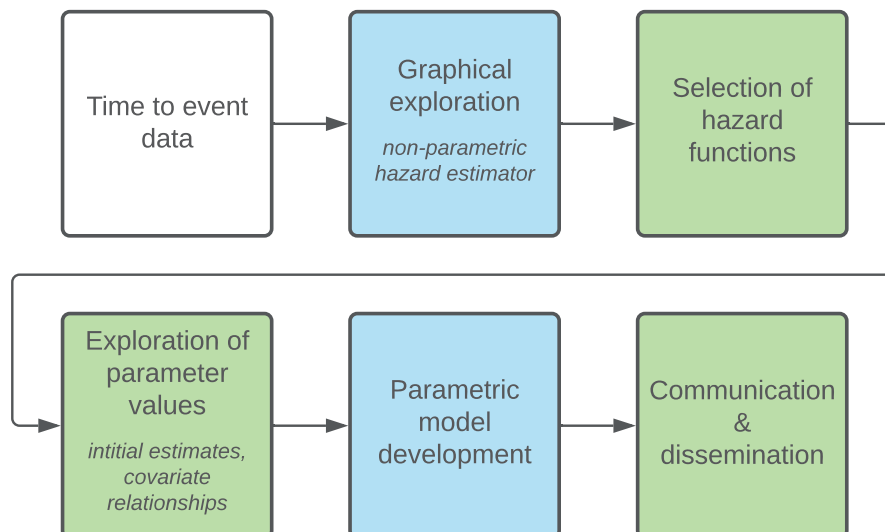


FIGURE 1 Time-to-event analysis workflow. White box represents the input data, green boxes represent actions that can be supported by the proposed Shiny application, and blue boxes represent actions by other algorithms (e.g., nonparametric hazard estimators, nonlinear mixed-effects modeling software)

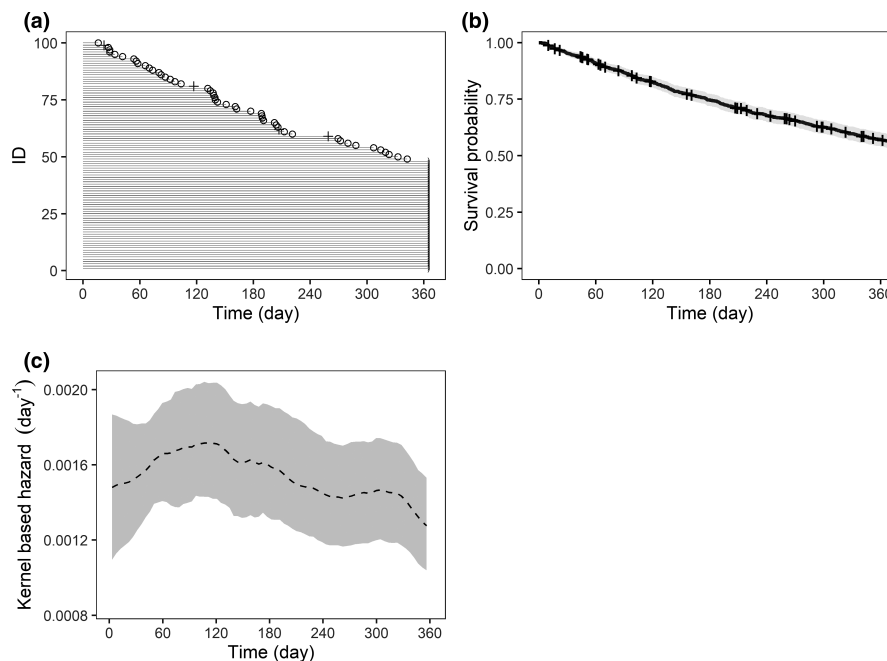


FIGURE 2 Graphical exploration of a typical time-to-event data set. The data set was simulated with a log-logistic hazard function ($\lambda = 0.002$, $\gamma = 1.1$) over 365 days for 1000 individuals (IDs) with 10% random right-censoring. (a) Follow-up time (horizontal line), event (circle symbol), and censoring (plus symbol) for 100 IDs randomly drawn from the simulated data and ordered on event time; (b) Kaplan-Meier plot of the survival probability including censoring (plus symbol) based on the full simulated data set; and (c) kernel-based visual hazard comparison showing the nonparametric kernel-based hazard estimate (dashed line) and its 95% confidence interval (shaded area) based on the full simulated data set

HAZARD FUNCTIONS FOR PARAMETRIC TIME-TO-EVENT ANALYSIS

The hazard functions are described in detail next and shown in Table 1 and Figure 3. The more complicated hazard functions commonly contain a shape parameter (e.g., γ) modifying the curve of the function and a scale parameter (e.g., λ).

Exponential model

The constant hazard function is the simplest hazard function with no change in the hazard over time (Table 1, Figure 3). The survival function is described by an exponential function of minus λ times time, and the probability density function of the events time, which equals the product of the hazard function and the survival function, shows an exponential distribution. Because of this, the constant hazard function is also referred to as the exponential model in time-to-event analysis. The constant hazard function is more restrictive than the semiparametric Cox proportional hazard function because it assumes a constant (baseline) hazard function, whereas the Cox model only assumes the proportionality of hazards when

(undefined) hazard functions do not need to be constant as long as they are proportional. It has, for example, been applied to characterize the time to positivity in a tuberculosis liquid culture assay,¹³ time to acute idiopathic pulmonary fibrosis exacerbations,¹⁴ time to recrudescence of visceral leishmaniasis infection,¹⁵ and overall survival of patients with acute myeloid leukemia.¹⁶

Gompertz model

The Gompertz model is a two-parameter proportional hazard function (Table 1, Figure 3).¹⁷ It is similar to the Cox proportional hazard model but has a defined baseline hazard function with a shape parameter γ following the Gompertz distribution. When $\gamma < 0$, the hazard decreases over time, whereas $\gamma > 0$ characterizes an increasing hazard with time. The Gompertz model reduces to the exponential model for $\gamma = 0$. Its application in pharmacometrics include the characterization of the time to recurrent venous thromboembolism, acute deep vein thrombosis, and pulmonary embolism¹⁸; the time to analgesic remedication after surgery,^{19–21} which can be expected to decrease with time after surgery; and the time to recurrent *Clostridium difficile* infection.²²

TABLE 1 Hazard functions for the most commonly used models in time-to-event analysis

Model	Hazard function
Exponential	$h(t) = \lambda$
Gompertz	$h(t) = \lambda \cdot e^{\gamma \cdot t}$
Weibull	$h(t) = \lambda \cdot \gamma \cdot (\lambda \cdot t)^{\gamma-1}$ Alternatively: $h(t) = \lambda \cdot \gamma \cdot t^{\gamma-1}$
Log-normal	$h(t) = \frac{(\sigma \cdot t \cdot \sqrt{2\pi})^{-1} \cdot e^{-\frac{1}{2} \left(\frac{\log t - \mu}{\sigma}\right)^2}}{1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)}$
Log-logistic	$h(t) = \frac{\lambda \cdot \gamma \cdot (\lambda \cdot t)^{\gamma-1}}{1 + (\lambda \cdot t)^\gamma}$ Alternatively: $h(t) = \frac{\lambda \cdot \gamma \cdot t^{\gamma-1}}{1 + (\lambda \cdot t)^\gamma}$
Generalized gamma	$S(t) = \begin{cases} 1 - \Gamma\left(\lambda^{-2} (e^{-\beta t})^{\frac{\lambda}{\sigma}}; \lambda^{-2}\right) & \text{if } \lambda > 0 \\ \Gamma\left(\lambda^{-2} (e^{-\beta t})^{\frac{\lambda}{\sigma}}; \lambda^{-2}\right) & \text{if } \lambda < 0 \end{cases}$ $f(t) = \frac{ \lambda }{\sigma \Gamma(\lambda^{-2})} \left[\lambda^{-2} (e^{-\beta t})^{\frac{\lambda}{\sigma}} \right]^{\lambda^{-2}} e^{-\lambda^{-2} (e^{-\beta t})^{\frac{\lambda}{\sigma}}}$ $h(t) = \frac{f(t)}{S(t)}$
Circadian	$h(t) = \lambda \cdot \left(1 + \text{amp} \cdot \text{Sin}\left(\frac{2\pi}{\text{period}} \cdot (t + \text{phase})\right) \right)$

Weibull model

The Weibull model is a two-parameter model with flexibility to characterize both increasing and decreasing hazard over time, similar to the Gompertz model, but for the shape parameter >1 or <1, respectively.^{23,24} The Weibull model is a more recent addition to time-to-event analysis compared with the Gompertz model and shows more flexibility in, for example, characterizing an increasing hazard of which the slope decreases with time. If the shape parameter does not vary with covariates, both the accelerated failure time and proportional hazard assumptions hold, which is unique to this model.^{1,2} This makes the Weibull model attractive for intuitive interpretation by nonmodelers regarding the hazard ratio between populations and a treatment or covariate effect on the survival time.

The Weibull model is commonly found in pharmacometric analyses in, for example, time to relapse in multiple sclerosis,²⁵ time to epileptic seizure,²⁶ time to onset of drug toxicity in preclinical experiments with both left- and right-censoring,²⁷ and overall survival in patients with metastatic colorectal cancer,²⁸ and therefore it is also used in more theoretical exercises.^{12,29,30} An alternative parameterization (see Table 1, Figure 3) has been observed in the tutorial by Holford⁸ and the handbook by Kleinbaum and Klein¹ as well as in applications to time

to adverse drug events,³¹ time to anemia,³² and mortality time in patients with amyotrophic lateral sclerosis.³³ The two functions differ by the factor λ and are thus equal at $\lambda = 1$ (and $\lambda = 0$). They collapse for $\lambda_{\text{alternative}} = \lambda^\gamma$. For $\lambda > 1$, parameterization 1 > parameterization 2 for $t > 0$, whereas $0 < \lambda < 1$ results in parameterization 1 < parameterization 2 for $t > 0$. The alternative parameterization 2 seems to be more unstable and sensitive to initial estimates in our experience, running into numerical difficulties with the integration routine. With respect to reproducibility, it is recommended to include the exact formula of the hazard function in the methods in addition to its name.

Log-normal model

The log-normal distribution is an accelerated failure time model, but not a proportional odds or proportional hazard model (Table 1, Figure 3).¹ The hazard increases to a maximum and then decreases over time (non-monotonically), resulting in a log-normal distribution of the event time.³⁴ Although it is a two-parameter model (μ and σ as the mean and standard deviation) similar to the Gompertz and Weibull models, its parameterization is more complex, including the standard normal cumulative distribution function Φ .³⁵ This complex

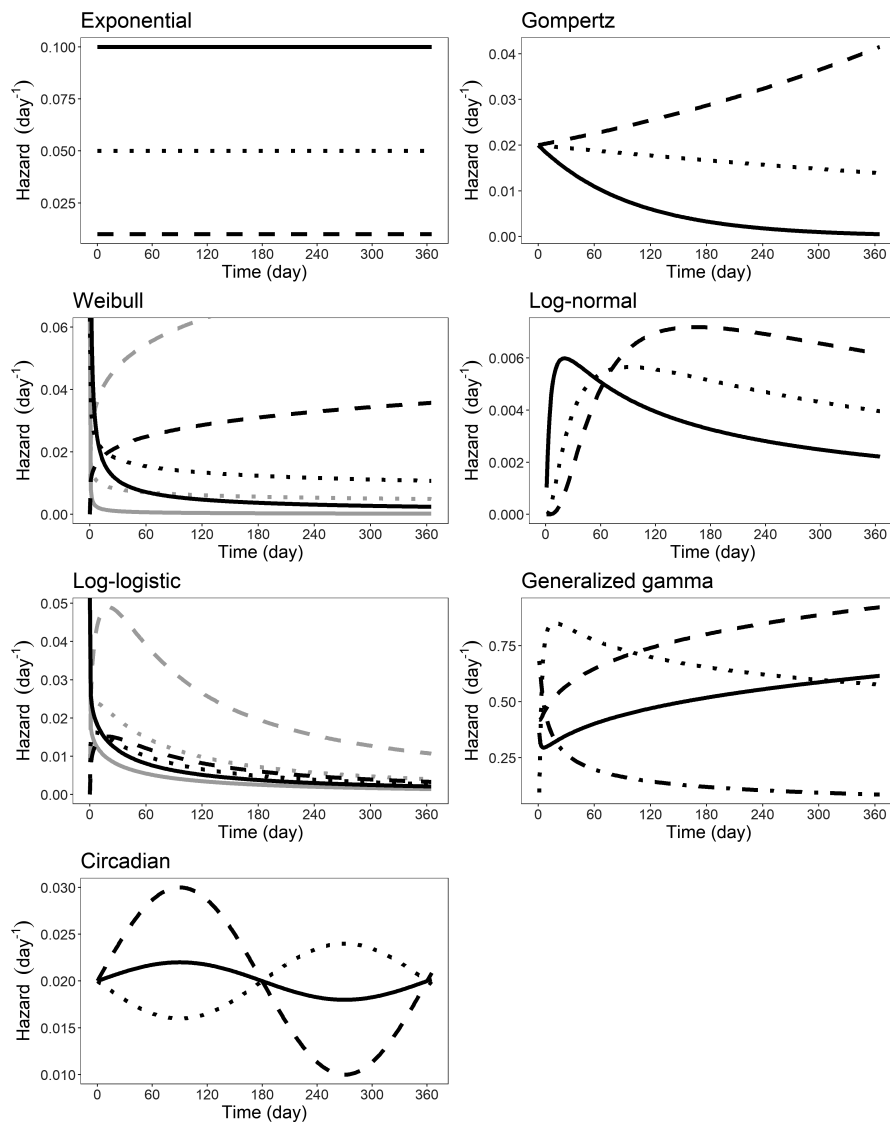


FIGURE 3 Hazard over time for the exponential model (solid line: $\lambda = 0.1$; dotted line: $\lambda = 0.05$; dashed line: $\lambda = 0.01$), the Gompertz model ($\lambda = 0.02$, solid line: $\gamma = -0.01$; dotted line: $\gamma = -0.001$; dashed line: $\gamma = 0.002$), the Weibull model ($\lambda = 0.02$, solid line: $\gamma = 0.4$; dotted line: $\gamma = 0.8$; dashed line: $\gamma = 1.2$; alternative parameterization in gray), the log-normal model ($\mu = 5$, solid line: $\sigma = 1.5$; dotted line: $\sigma = 1$; dashed line: $\sigma = 0.75$), the log-logistic model ($\lambda = 0.02$, solid line: $\gamma = 0.9$; dotted line: $\gamma = 1.1$; dashed line: $\gamma = 1.3$; alternative parameterization in gray), the generalized gamma model (solid line: $\mu = 1.1$, $Q = 1.5$, $\sigma = 1.2$; dotted line: $\mu = 1.2$, $Q = 0.4$, $\sigma = 0.5$; dashed line: $\mu = 0.9$, $Q = 1.2$, $\sigma = 1$; dot-dashed line: $\mu = 0.3$, $Q = 0.5$, $\sigma = 1$; Prentice parameterization), and the circadian model ($\lambda = 0.02$, period = 360, solid line: amplitude = 0.1, phase = 0; dotted line: amplitude = 0.2, phase = 180; dashed line: amplitude = 0.5, phase = 0)

parameterization makes an intuitive choice of initial estimate challenging. The log-normal model has, for example, been applied to time to undetectable levels of hepatitis C virus³⁶ and overall survival and progression-free survival of patients with metastatic gastrointestinal stromal tumors.³⁷

Log-logistic model

The log-logistic distribution allows the hazard to increase for a shape parameter $\gamma > 1$, while it decreases over time for shape parameter $\gamma \leq 1$ (Table 1, Figure 3). If the accelerated failure time assumption holds, the log-logistic distribution is a proportional odds model, that is, the odds ratio is constant over time.¹ The event time approximates a log-logistic distribution.³⁴ Similar to the Weibull model, the log-logistic model has an alternative parameterization.¹ An advantage of the log-logistic model over the log-normal model is that is can

be solved analytically, which decreases run times.³⁸ However, finding initial estimates remains challenging. The log-logistic model has been applied in characterizing the time to pain response in preclinical experiments where it showed better performance compared with the Weibull model in run times and predictive performance³⁹ as well as in modeling dropout in a metastatic colorectal cancer clinical trial²⁸ and progression-free survival in anaplastic lymphoma kinase-positive non-small cell lung cancer.⁴⁰

Generalized gamma model

The generalized gamma model with three parameters (scale σ , shape λ , and location β) allows for even more flexibility than the previous models (Table 1).¹ The flexibility of the generalized gamma model allows its hazard function to take the most common distributions: increasing over time, decreasing over time, decreasing then

increasing over time (U shape or bathtub), and increasing then decreasing over time (inverted U shape).⁴¹ The location parameter accelerates the function with time, whereas the scale and shape parameter determine if the distribution increases, decreases, or has an (inverted) U shape over time as demonstrated by Cox et al.⁴¹ The generalized gamma model can collapse into simpler models, such as the Weibull (for $\lambda = 1$), log-normal (for $\lambda = 0$), or exponential (for $\lambda = \sigma = 1$) model.^{2,41} The generalized gamma model was reparameterized by Prentice to facilitate modeling that is incorporated in statistical software including R.^{42,43} It has been applied to characterize the time to unassisted breathing or death in patients who are critically ill.⁴⁴ It is, however, more commonly used as a tool to find the most appropriate time-to-event model by checking the Weibull or log-normal model assumptions than used as a final model itself² because the generalized gamma model can reduce to the Weibull and log-normal models.^{1,41} One reason is its complexity, including the standardized cumulative gamma distribution function (Γ , mean and variance $\gamma > 0$).

Circadian or periodical model

A more mechanistic characterization of the instantaneous rate of an event over time can be captured by a circadian or more general periodic function, where the hazard changes with the period of the day, month, or year (Table 1, Figure 3).¹² Especially for clinical trials that cover an extended period of time and can be influenced by seasonality, such models can become relevant. The circadian model consists of three additional parameters that cover the amplitude, period, and phase of the periodic relationship in the hazard and can be used as multiplicative factor for any baseline hazard functions (shown in Table 1 for the exponential model).

EXPLORATION OF HAZARD FUNCTION PARAMETER VALUES

Finding the right hazard functions to test is a first step in time-to-event model development (see Figure 1), but model algorithms require initial estimates. The more complex hazard functions are sensitive to the initial estimates the modeler starts with and can easily run into numerical issues. This was even observed for different parameterizations of, for example, the Weibull model. The descriptions of the aforementioned models give a first idea on the constraints of the parameter values given a certain shape (e.g., $\lambda > 1$ for an increasing hazard with time in the Weibull model), but not on the order of magnitude of

the parameters. It is beneficial to modelers to familiarize themselves with the parameter space in relation to the exact shape of the hazard over time of the observed data. In addition, it is important to be aware of the sensitivity of the hazard over time to the different parameters. Exploring the parameter space can inform on which covariate relationships should be tested, especially when a graphical exploration based on a nonparametric hazard estimation was performed on subsets of the data for the covariate(s) of interest. The same holds for testing the treatment effect. If, for example, the peak of the hazard over time in Figure 2c occurs later in the treatment group compared with the placebo group, that effect might be captured best by modeling the treatment at the γ parameter, whereas if the peak decreases, it might be better captured on the λ parameter.

To explore the hazard function parameter values and their influence on the hazard over time, we have developed a Shiny application (available at <https://pqp-uu.shinyapps.io/HazardFunctionsInParametericTTE/> and <https://github.com/rcvanwijk/HazardFunctionsInParametericTTE>; Figure 4 and Supplementary Material) using the Shiny package (Version 1.6.0)⁴⁵ in R (Version 4.0.4) through the RStudio (Version 1.4.1106) interface. The application simulates the six most common hazard functions (exponential, Gompertz, Weibull, log-normal, log-logistic, circadian; Table 1) in a reactive manner, meaning it responds automatically to changes the user makes in the input parameters. The impact of covariates/treatment on the hazard over time can be explored by an additional set of inputs. A covariate/treatment can affect the hazard function proportionally. For each function, a dedicated coefficient input is available in the application. Under the proportional hazard assumption, a hazard ratio can be calculated by exponentiating the coefficient.¹⁸ Alternatively, a covariate/treatment can affect the hazard function's individual parameters for which the application has a relative input per parameter. Both slider and numerical Shiny input widgets are built into the application for function parameters and covariate/treatment effect as well as for the simulation time to give the user full control. For each function, a figure based on the reactive simulation is assigned to a reactive function that is rendered to display in the application and made available in tiff format through a download handler. The Shiny application was code reviewed and tested independently by two researchers.

Figure 4 shows an overview of the Shiny application. Sliders or numerical input for the hazard function parameters (Figure 4a) and the covariate/treatment effect on those parameters (Figure 4b) can be used to simulate the hazard over time profile in the graphical panel (Figure 4c). With the nonparametric hazard estimation based on the observed events, the right functions showing a similar

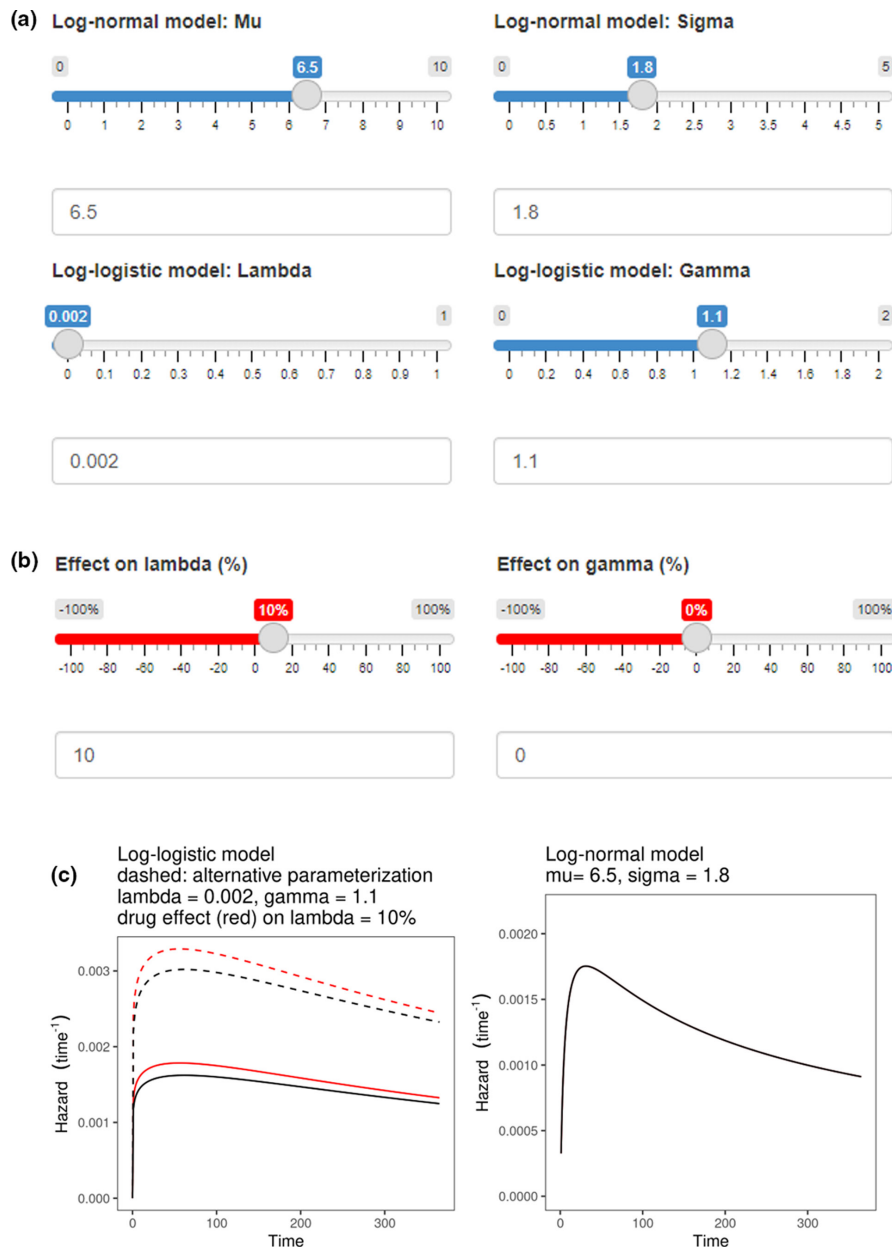


FIGURE 4 Shiny application with (a) slider panel for the hazard function parameters, (b) slider panel for covariate/treatment effect on the hazard function parameters, and (c) the corresponding hazard over time profiles, here illustrated for log-logistic and log-normal functions

shape can be readily selected. Based on Figure 2c, either the log-normal or log-logistic model would be best to describe the data. In addition, the input parameters can be updated to improve the shape toward the observed data. These values can then be used as initial estimates for the nonlinear mixed-effects modeling software. For the log-normal and log-logistic models, the parameter values were adjusted (Figure 4a) to approximate the nonparametric hazard shape (compare Figure 4c with Figure 2c), and these values can now be used as initial estimates into the parametric time-to-event algorithm. Changing the parameters using the sliders will also illustrate the sensitivity of the shape of the function to different parameters. In addition, the covariate/treatment effect sliders can even be used to explore which parameter should be tested for covariate and/or treatment effects. Overall, a better

understanding of the meaning and impact of the parameters will result from the exploration of the hazard functions using the Shiny application developed here, which will improve the following model development steps.

DISSEMINATION USING THE SHINY APPLICATION

Time-to-event analysis in general and the hazard function specifically are abstract concepts that are difficult to intuitively understand by nonexperts. Dissemination of the time-to-event projects can be supported by informative graphics. One example is the Kaplan–Meier VPC, which overlays the time-to-event observations with a prediction interval based on a high number (e.g., 500 or

1000) of simulated time-to-event data sets using the final model.¹⁹ The advantage of the Kaplan–Meier VPC is that it shows the model performance with a graphic commonly used and understood. However, the simulation-based technique is not without limitations, including a longer run time and less sensitivity to more complex hazard functions.¹² In principle it is better to communicate at the level where modeling decisions take place on, for example, treatment and covariate effects. When developing covariate models in pharmacokinetics, the covariate relationship to the pharmacokinetic parameter such as clearance is presented in addition to the concentration-time profile. The same holds for the hazard function, albeit complex. To aid with that complexity, the Shiny application developed here can be used. In presentations or workshops, the different hazard functions can be visualized as well as the meaning and impact of the parameters on their shape. The Shiny application also includes the possibility to export the created figures that can be used in further communication. In addition to specific project dissemination, the Shiny application can be used in educational programs teaching time-to-event analysis.

DISCUSSION AND CONCLUSION

We have developed a Shiny application to find the right hazard function(s) to test in time-to-event modeling, explore the influence of the parameter values on the hazard over time to inform on initial estimates and possible covariate or treatment relationships to test, and support the dissemination of project results.

Time-to-event modeling is an important method in the toolbox of the pharmacometrician. It is an excellent tool to characterize not only the occurrence of a clinical (or preclinical) event but also its time as well as the impact of censored data. That quantitative relationship is essential for interpretation of treatment and covariate effects and translation to unknown phases in drug development, clinical situations, or populations. As time-to-event modeling and quantification of treatment and covariate effects takes place at the hazard level, the hazard functions at the pharmacometrician's disposal must be well understood. Combining recently developed methods to graphically explore the time-to-event data by hazard estimators^{11,12} with knowledge of the theoretical hazard functions is essential for time-to-event model development. The Shiny application developed here assists in that understanding.

The hazard function is an essential element in parametric time-to-event modeling, but other elements are important to address as well. Dropout that is nonrandom between, for example, the treatment and the control group, and therefore informative on the treatment effect, should

be considered.⁴⁶ The Shiny application developed here is limited to visualize informative dropout, which can be detected more readily in, for example, Kaplan–Meier curves displaying dropout as right-censoring. Another element that complicates time-to-event modeling is time-varying elements such as covariates or treatment. Although the Shiny application developed here could visualize periodic hazard functions through, for example, the circadian function, it was not designed to detect time-varying covariate relationships in parametric time-to-event analysis. A combination of tools is at the disposal of the pharmacometrician to elucidate the more complex relationships in drug development to which we add this Shiny application.

In conclusion, we described in detail the most common hazard functions including mathematical and graphical representation and developed a Shiny application to support time-to-event modelers to start model development in a well-informed manner as well as offer graphics to aid in the dissemination of results.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Laurynas Mockeliunas and Lina Keutzer for testing the application and Xiaomei Chen for consultation on the Weibull derivatization. The Shiny application is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

ORCID

Rob C. Van Wijk  <https://orcid.org/0000-0001-7247-1360>
Ulrika S. H. Simonsson  <https://orcid.org/0000-0002-3424-9686>

REFERENCES

1. Kleinbaum DG, Klein M. *Survival analysis, 3rd edition (Statistics for Biology and Health)*. Springer; 2020.
2. Klein JP, Moeschberger ML. *Survival Analysis. Techniques for Censored and Truncated Data (Statistics for Biology and Health)*. Springer; 2003.
3. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*. 2016;133:601-609.
4. Krishnan SM, Friberg LE, Bruno R, Beyer U, Jin JY, Karlsson MO. Multistate model for pharmacometric analyses of overall survival in HER2-negative breast cancer patients treated with docetaxel. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:1255-1266.
5. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53:457-481.
6. Cox DR. Regression models and life-tables. *J R Stat Soc Ser B*. 1972;34:187-220.
7. Guo S, Zeng D. An overview of semiparametric models in survival analysis. *J Stat Plan Inference*. 2014;151-152:1-16.

8. Holford N. A time to event tutorial for pharmacometricians. *CPT Pharmacometrics Syst Pharmacol.* 2013;2:e43.
9. Muller H-G, Wang J-L. Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics.* 1994;50:61-76.
10. Chiang CT, Wang MC, Huang CYU. Kernel estimation of rate function for recurrent event data. *Scand J Stat.* 2005;32:77-91.
11. Huh Y, Hutmacher MM. Application of a hazard-based visual predictive check to evaluate parametric hazard models. *J Pharmacokinet Pharmacodyn.* 2016;43:57-71.
12. Goulooze SC, Väitalo PAJ, Knibbe CAJ, Krekels EHJ. Kernel-based visual Hazard comparison (kbVHC): a simulation-free diagnostic for parametric repeated time-to-event models. *AAPS J.* 2018;20:1-11.
13. Chigutsa E, Patel K, Denti P, et al. A time-to-event pharmacodynamic model describing treatment response in patients with pulmonary tuberculosis using days to positivity in automated liquid mycobacterial culture. *Antimicrob Agents Chemother.* 2013;57:789-795.
14. Tang F, Weber B, Stowasser S, Korell J. Parametric time-to-event model for acute exacerbations in idiopathic pulmonary fibrosis. *CPT Pharmacometrics Syst Pharmacol.* 2020;9:87-95.
15. Dorlo TPC, Kip AE, Younis BM, et al. Visceral leishmaniasis relapse hazard is linked to reduced miltefosine exposure in patients from eastern Africa: a population pharmacokinetic/pharmacodynamic study. *J Antimicrob Chemother.* 2017;72:3131-3140.
16. Lin S, Shaik MN, Chan G, Cortes JE, Ruiz-Garcia A. Population pharmacokinetic/pharmacodynamic evaluation of the relationship between Glasdegib treatment/exposure and overall survival in AML patients. *Blood.* 2018;132:1450.
17. Gompertz B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Phil Trans R Soc London.* 1825;115: 513-583.
18. Nyberg J, Karlsson KE, Jönsson S, et al. Edoxaban exposure-response analysis and clinical utility index assessment in patients with symptomatic deep-vein thrombosis or pulmonary embolism. *CPT Pharmacometrics Syst Pharmacol.* 2016;5:222-232.
19. Juul RV, Rasmussen S, Kreilgaard M, Christrup LL, Simonsson USH, Lund TM. Repeated time-to-event analysis of consecutive analgesic events in postoperative pain. *Anesthesiology.* 2015;123:1411-1419.
20. Elkomy MH, Drover DR, Galinkin JL, Hammer GB, Glotzbach KL. Pharmacodynamic analysis of morphine time-to-Remedication events in infants and young children after congenital heart surgery. *Clin Pharmacokinet.* 2016;55:1217-1226.
21. Goulooze SC, Krekels EH, Saleh MA, et al. Predicting unacceptable pain in cardiac surgery patients receiving morphine maintenance and rescue doses: a model-based pharmacokinetic-pharmacodynamic analysis. *Anesth Analg.* 2021;132:726-734.
22. Yee KL, Kleijn HJ, Zajic S, Dorr MB, Wrishko RE. A time-to-event analysis of the exposure-response relationship for bezlotoxumab concentrations and CDI recurrence. *J Pharmacokinet Pharmacodyn.* 2020;47:121-130.
23. Weibull W. A statistical distribution function of wide applicability. *J Appl Mech.* 1951;103:293-297.
24. Juckett DA, Rosenberg B. Comparison of the Gompertz and Weibull functions as descriptors for human mortality distributions and their intersections. *Mech Ageing Dev.* 1993;69:1-31.
25. Novakovic AM, Thorsted A, Schindler E, Jönsson S, Munafo A, Karlsson MO. Pharmacometric analysis of the relationship between absolute lymphocyte count and expanded disability status scale and relapse rate, efficacy end points. *Multiple Sclerosis Trials J Clin Pharmacol.* 2018;58:1284-1294.
26. Kim MJ, Yum MS, Yeh HR, Ko TS, Lim HS. Pharmacokinetic and pharmacodynamic evaluation of intravenous levetiracetam in children with epilepsy. *J Clin Pharmacol.* 2018;58:1586-1596.
27. Berges A, Cerou M, Sahota T, et al. Time-to-event modeling of left- or right-censored toxicity data in nonclinical drug toxicology. *Toxicol Sci.* 2018;165:50-60.
28. Vera-Yunca D, Parra-Guillen ZP, Girard P, Trocóniz IF, Terranova N. Relevance of primary lesion location, tumour heterogeneity and genetic mutation demonstrated through tumour growth inhibition and overall survival modelling in metastatic colorectal cancer. *Br J Clin Pharmacol.* 2022;88:166-177.
29. Abel UR, Jensen K, Karapanagiotou-Schenkel I, Kieser M. Some issues of sample size calculation for time-to-event endpoints using the freedman and Schoenfeld formulas. *J Biopharm Stat.* 2015;25:1285-1311.
30. Phadnis MA, Sharma P, Thewarapperuma N, Chalise P. Assessing accuracy of Weibull shape parameter estimate from historical studies for subsequent sample size calculation in clinical trials with time-to-event outcome. *Contemp Clin Trials Commun.* 2020;17:100548.
31. Cornelius VR, Sauzet O, Evans SJW. A signal detection method to detect adverse drug reactions using a parametric time-to-event model in simulated cohort data. *Drug Saf.* 2012;35:599-610.
32. Tod M, Farcy-Afif M, Stocco J, et al. Pharmacokinetic/pharmacodynamic and time-to-event models of ribavirin-induced anaemia in chronic hepatitis C. *Clin Pharmacokinet.* 2005;44:417-428.
33. Eijk, R. P. A. van, Eijkemans, M. J. C., Rizopoulos, D., van Den Berg, L. H. & Nikolakopoulos, S. Comparing methods to combine functional loss and mortality in clinical trials for amyotrophic lateral sclerosis. *Clin Epidemiol.* 2018;10:333-341.
34. Lee ET, Go OT. Survival analysis in public health research. *Annu Rev Public Health.* 1997;18:105-134.
35. Nyberg J et al. Simulating large time-to-event trials in NONMEM. *Page.* 2014;23:3166.
36. Saleh MI, Bani Melhim S. A time-to-event analysis describing virologic response in patients with chronic hepatitis C infection. *J Chemother.* 2019;31:274-283.
37. Schindler E, Krishnan SM, Mathijssen R, Ruggiero A, Schiavon G, Friberg LE. Pharmacometric modeling of liver metastases' diameter, volume, and density and their relation to clinical outcome in imatinib-treated patients with gastrointestinal stromal tumors. *CPT Pharmacometrics Syst Pharmacol.* 2017;6:449-457.
38. Claret L, Zheng J, Mercier F, et al. Model-based prediction of progression-free survival in patients with first-line renal cell carcinoma using week 8 tumor size change from baseline. *Cancer Chemother Pharmacol.* 2016;78:605-610.
39. Yassen A, Olofsen E, Dahan A, Danhof M. Pharmacokinetic-pharmacodynamic modeling of the antinociceptive effect of buprenorphine and fentanyl in rats: role of receptor equilibration kinetics. *J Pharmacol Exp Ther.* 2005;313:1136-1149.
40. Gupta N, Wang X, Offman E, et al. Brigatinib dose rationale in anaplastic lymphoma kinase-positive non-small cell lung

- cancer: exposure–response analyses of pivotal ALTA study. *CPT Pharmacometrics Syst Pharmacol*. 2020;9:718-730.
41. Cox C, Chu H, Schneider MF, Muñoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Stat Med*. 2007;26:4352-4374.
 42. Prentice RL. A log gamma model and its maximum likelihood estimation. *Biometrika*. 1974;61:539-544.
 43. Jackson C. Flexsurv: a platform for parametric survival modeling in R. *J Stat Softw*. 2016;70:1-33.
 44. Checkley W, Brower RG, Muñoz A. Inference for mutually exclusive competing events through a mixture of generalized gamma distributions. *Epidemiology*. 2010;21:557-565.
 45. Chang W, Cheng J, Allaire JJ, et al. Shiny: web application framework for R. 2021. <https://cran.r-project.org/package=shiny>
 46. Björnsson MA, Simonsson USH. Modelling of pain intensity and informative dropout in a dental pain model after

naproxinod, naproxen and placebo administration. *Br J Clin Pharmacol*. 2011;71:899-906.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Van Wijk RC, Simonsson USH. Finding the right hazard function for time-to-event modeling: A tutorial and Shiny application. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:991-1001. doi:[10.1002/psp4.12797](https://doi.org/10.1002/psp4.12797)