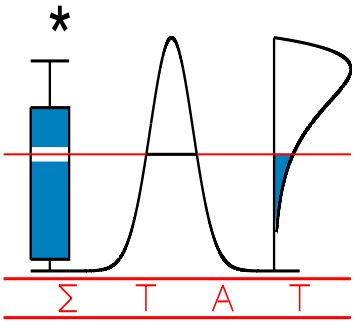
T E C H N I C A L

R E P O R T

0307

EVOLUTION OF RECURRENT ASTHMA EVENT RATE OVER TIME IN FRAILTY MODELS.

L. DUCHATEAU, P. JANSSEN, I. KEZIC and C. FORTPIED



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Evolution of recurrent asthma event rate over time in frailty models

Luc Duchateau †

Department of Physiology, Biochemistry and Biometrics, Faculty of Veterinary Medicine, Ghent University, Merelbeke, Belgium

Paul Janssen

Center for Statistics, Limburgs Universitair Centrum, Diepenbeek, Belgium

Iva Kezic

Catherine Fortpied

UCB S.A., Braine-l'Alleud, Belgium

Summary. To model the time evolution of the event rate in recurrent event data a crucial role is played by the timescale used. Depending on the selected timescale the interpretation of the time evolution will be entirely different, both in parametric and semiparametric frailty models. The gap timescale is most appropriate when studying the recurrent event rate as a function of time since the last event, whereas the calendar timescale keeps track of actual time. We show both timescales in action on data of an asthma prevention trial in young children. The frailty model is further extended to include both timescales simultaneously as this might be most relevant in practice.

Keywords: recurrent events, frailty model, gap time, hazard rate, asthma

1. Introduction

Recurrent event data have mostly been analysed using standard survival techniques with an additional adjustment for the correlation between events originating from the same individual, leading either to marginal models (Prentice *et al.* (1981); Wei *et al.* (1989)) or to frailty models (McGilchrist and Aisbett (1991)). Both marginal models and conditional models have been used in a parametric and semiparametric context (Mahé and Chevret (1999)).

Little attention, however, has been given to the timescale that is used for subsequent events and the interpretation attached to it. Different timescales can be used (Kelly and Lim (2000)). The most often used timescale is the gap time: after an event, the subject starts again at time 0 and the time to the next event corresponds to the number of days it takes to experience the next event. Alternative timescales are the total time and the calendar time. With total time, the time to any event corresponds to the time since randomisation regardless whether other events have been experienced meanwhile. Calendar time keeps track of time since randomisation, as in total time, but the duration of the time at risk for

†*Address for correspondence*: Luc Duchateau, Department of Physiology, Biochemistry and Biometrics, Faculty of Veterinary Medicine, Ghent University, Salisburylaan 133, 9820 Merelbeke, Belgium. E-mail:Luc.Duchateau@rug.ac.be

an event corresponds to the duration of the time at risk in the gap time representation (see Figure 1 and explanation in the next section).
In this paper, we investigate the effect of the use of the different timescales on the frailty model and on its interpretation, based on an asthma prevention trial in young children. Children with high risk for asthma (but not yet experiencing it) enter the study at the age of 6 months and are followed up for 18 months. They are randomised to placebo or drug. Apart from the drug effect, other points of interest are the evolution of the asthma recurrent event rate over time, and how the appearance of an event influences the event rate. These issues are particularly important in this study as children of 6 months differ substantially from children of 2 years and therefore physicians want to know from this study how the asthma event rate changes with age.
These points of interest will be tackled by making use of two different timescales, the calendar time and gap time in the context of frailty models; we also will extend the model to accomodate both processes in one and the same model.

2. Recurrent event data representation

Recurrent event data can be represented in different ways depending on the timescale used. As a starting point, the representations introduced by Kelly and Lim (2000) are used, but they are extended in order to account for a specific issue in asthma data: when a patient has an asthma attack that can last for several days, he/she should not be considered to be at risk for an asthma attack at that time. Kelly and Lim (2000) distinguish between total time, gap time and calendar time. The time at risk for an event in total time representation starts at time 0 when the subject enters the study until the particular event is experienced. It is unclear, however, what to do with the time when the subject was not at risk, either because of censoring or experiencing an asthmatic event. This timespan should be subtracted but then the total time does no longer present the actual time since randomisation. Furthermore, total time representation has little intuitive appeal because a subject is at risk for all his/her events at the start of the study even when it is known that particular events can only happen after previous events have taken place. Therefore, we do not consider total time in the remainder. In case of gap time representation, the time at risk starts at zero, but the length of the time at risk now corresponds to the time since the end of the previous event (or study entry in case of first event) until the time of the particular event. In calendar time formulation, the length of the time at risk period is the same, but the start of the at risk period is not reset at zero but at the actual time since study entry (see Figure 1).

Assume there are in total N subjects. A particular subject has different periods at risk during the total observation time, that are separated from each other by either an asthmatic event that lasts one or more days, or by a period during which the subject was not under observation. If there are r_i at risk periods for patient i, then the complete information on asthma attacks for patient i can be presented by r_i triplets

$$((t_{i11}, t_{i12}, \delta_{i1}), \ldots, (t_{ir_i1}, t_{ir_i2}, \delta_{ir_i}))$$

where for the j^{th} triplet, t_{ij1} is the start of the j^{th} at risk period, t_{ij2} is the end of the j^{th} period and δ_{ij} is the censoring indicator and $t_{i11} = 0$.

The hazard function for the frailty model with calendar time is given by

$$\lambda_i(t) = \begin{cases} \lambda_0(t) U_i \exp(\beta x_i) & \text{for } t_{ij1} \leq t \leq t_{ij2}, j = 1, \ldots, r_i \\ 0 & \text{otherwise} \end{cases}$$

where $\lambda_0(t)$ is the baseline hazard which is assumed to be independent of both the event history and the covariates of the subject and U_i is the frailty term for the i^{th} patient.

The frailty terms $U_1, \ldots, U_N$ are assumed to be independent with common frailty density. As frailty density we will only consider the one parameter gamma density with mean one and variance θ

$$f_{U_i}(u) = \frac{u^{1/\theta - 1} \exp(-u/\theta)}{\theta^{1/\theta} \Gamma(1/\theta)}.$$

The likelihood function corresponding to this hazard is then given by

$$\prod_{i=1}^{N} \prod_{j=1}^{r_i} \left(\lambda_i(t_{ij2})\right)^{\delta_{ij}} \exp\left(-\Lambda_i\left(t_{ij1}, t_{ij2}\right)\right)$$

with cumulative hazard

$$\Lambda_i\left(t_{ij1}, t_{ij2}\right) = \int_{t_{ij1}}^{t_{ij2}} \lambda_i(t) dt.$$

In the case of gap time, part of the information in the triplets is redundant, and we could summarize the information in the triplets alternatively as

$$((t_{i12} - t_{i11}, \delta_{i1}), \ldots, (t_{ir_i2} - t_{ir_i1}, \delta_{ir_i}))$$

In other words, only the length of the time at risk is needed, and not the particular time (relative to study entry time) when the patient is at risk.

The hazard function for the frailty model with gap time is given by

$$\lambda_i(t) = \begin{cases} \lambda_0(t - t_{ij1}) U_i \exp\left(-\beta x_i\right) & \text{for } t_{ij1} \leq t \leq t_{ij2}, j = 1, \ldots, r_i \\ 0 & \text{otherwise} \end{cases}$$

and the corresponding likelihood function is, as before,

$$\prod_{i=1}^{N} \prod_{j=1}^{r_i} \left(\lambda_i(t_{ij2})\right)^{\delta_{ij}} \exp\left(-\Lambda_i\left(t_{ij1}, t_{ij2}\right)\right)$$

but now with different meaning for the hazard $\lambda_i(.)$ and cumulative hazard $\Lambda_i(.,.)$.

3. Parametric frailty models for recurrent event data

The Weibull baseline hazard is taken as an example of the parametric frailty model for recurrent event data; it has a straightforward interpretation. Other parametric hazard rates can be used such as the lognormal. As mentioned in Section 5, the algorithms for these two parametric assumptions are available in R.

The hazard function for calendar time (Model 1: Weibull-calendar) is

$$\lambda_i(t) = \begin{cases} \lambda_c \gamma_c t^{\gamma_c - 1} U_i \exp(\beta x_i) & \text{for } t_{ij1} \leq t \leq t_{ij2}, j = 1, \ldots, r_i \\ 0 & \text{otherwise} \end{cases}$$

with resulting likelihood function

$$\prod_{i=1}^{N} \prod_{j=1}^{r_i} \left[\left(\lambda_c \gamma_c t_{ij2}^{\gamma_c - 1} U_i \exp(\beta x_i) \right)^{\delta_{ij}} \exp\left(-\lambda_c \times \left(t_{ij2}^{\gamma_c} - t_{ij1}^{\gamma_c} \right) U_i \exp(\beta x_i) \right) \right]$$

The hazard function for gap time (Model 2: Weibull-gap) is

$$\lambda_i(t) = \begin{cases} \lambda_g \gamma_g (t - t_{ij1})^{\gamma_g - 1} U_i \exp(\beta x_i) & \text{for } t_{ij1} \leq t \leq t_{ij2}, j = 1, \ldots, r_i \\ 0 & \text{otherwise} \end{cases}$$

with resulting likelihood function

$$\prod_{i=1}^{N} \prod_{j=1}^{r_i} \left[\left(\lambda_g \gamma_g (t_{ij2} - t_{ij1})^{\gamma_g - 1} U_i \exp(\beta x_i) \right)^{\delta_{ij}} \exp\left(-\lambda_g \times (t_{ij2} - t_{ij1})^{\gamma_g} U_i \exp(\beta x_i) \right) \right]$$

In the special case of a constant baseline hazard rate, $\lambda_0(t) = \lambda$, the likelihood functions for the gap time and calendar time are equal, related to the fact that the exponential distribution is memoryless.

All the previous likelihood expressions contain the frailty terms U_i but not the frailty parameter θ. The easiest way, in the parametric case, to deal with the unobserved frailty terms is to integrate out the frailty density, thus obtaining the observable likelihood (Klein (1992); Duchateau *et al.* (2002)). A closed form for the observable likelihood for the one parameter gamma density is given by

$$\prod_{i=1}^{N} \frac{\Gamma(\frac{1}{\theta} + d_i)}{\theta^{1/\theta} \Gamma(\frac{1}{\theta})} \frac{\prod_{j=1}^{r_i} \left(\lambda_c \gamma_c t_{ij2}^{\gamma_c - 1} \exp(\beta x_i) \right)^{\delta_{ij}}}{\left(\frac{1}{\theta} + \sum_{j=1}^{r_i} \lambda_c \times (t_{ij2}^{\gamma_c} - t_{ij1}^{\gamma_c}) \exp(\beta x_i) \right)^{d_i + \frac{1}{\theta}}}$$

for the calendar time with $d_i = \sum_{j=1}^{r_i} \delta_{ij}$, the number of asthmatic events for subject i.

For the gap time the observable likelihood is

$$\prod_{i=1}^{N} \frac{\Gamma(\frac{1}{\theta}+d_i)}{\theta^{1/\theta}\Gamma(\frac{1}{\theta})} \frac{\prod_{j=1}^{r_i}\left(\lambda_g\gamma_g\left(t_{ij2}-t_{ij1}\right)^{\gamma_g-1}\exp(\beta x_i)\right)^{\delta_{ij}}}{\left(\frac{1}{\theta}+\sum_{j=1}^{r_i}\lambda_g\times\left(t_{ij2}-t_{ij1}\right)^{\gamma_g}\exp(\beta x_i)\right)^{d_i+\frac{1}{\theta}}}.$$

The Weibull-calendar and the Weibull-gap model are presented in Figure 2. It is clear that an event does not influence the hazard rate in the Weibull-calendar time model (Figure 2a) whereas the hazard rate changes abruptly after an event in the Weibull-gap time model (Figure 2b).

4. Extension of parametric frailty models

In the parametric models presented in Section 3, we did not take into account that the first event is different in nature from the subsequent events. Subjects that enter the study did not experience an asthmatic event yet. Therefore, Model 2 of Section 2 based on gap time can be extended to include a separate term for the first event. In this extended model, we assume the following hazard rate

$$\lambda_i(t)=\begin{cases}\lambda_f\gamma_f t^{\gamma_f-1}U_i\exp\left(\beta x_i\right) & \text{for } 0\le t\le t_{i12}\\ \lambda_g\gamma_g\left(t-t_{ij1}\right)^{\gamma_g-1}U_i\exp\left(\beta x_i\right) & \text{for } t_{ij1}\le t\le t_{ij2}, j=2,\ldots,r_i\\ 0 & \text{otherwise}\end{cases}$$

with for the special case $\gamma_f = 1$ (Model 3: Weibull-gap-first event exp) a constant hazard rate for the first event and for the more general model (Model 4: Weibull-gap-first event Weibull) a hazard rate based on Weibull distributed event times for the first event.

Parameter estimates for this extended models can be obtained in a similar way as before by maximising the observable likelihood based on this new hazard function. An example of such a model is given in Figure 2c. Before the first event the hazard rate is constant with an abrupt change when the first event occurs.

Up to now, we have considered either gap time or calendar time. The model can be extended further to take into account both time since study entry, as in the calendar time and time since last event, as in the gap time. We will consider the following model (Model 5: Weibull-calendar-gap) based on the Weibull distribution with the hazard function given by

$$\lambda_i(t)=\begin{cases}\lambda_c\gamma_c t^{\gamma_c-1}U_i\exp\left(\beta x_i\right) & \text{for } 0\le t\le t_{i12}\\ \left(\lambda_c\gamma_c t^{\gamma_c-1}+\lambda_g\gamma_g\left(t-t_{ij1}\right)^{\gamma_g-1}\right)U_i\exp\left(\beta x_i\right) & \text{for } t_{ij1}\le t\le t_{ij2}, j=2,\ldots,r_i\\ 0 & \text{otherwise}\end{cases}$$

Parameter estimates for this model can again be obtained by maximising the observable likelihood based on this new hazard function.

The contribution of risk period j from individual i to the cumulative hazard is given by

$$\Lambda_i(t_{ij1}, t_{ij2}) = \begin{cases} \lambda_c t_{i12}^{\gamma_c} U_i \exp(\beta x_i) & \text{for } j = 1 \\ \left(\lambda_c \times \left(t_{ij2}^{\gamma_c} - t_{ij1}^{\gamma_c}\right) + \lambda_g \times (t_{ij2} - t_{ij1})^{\gamma_c}\right) U_i \exp(\beta x_i) & \text{for } j = 2, \ldots, r_i \end{cases}$$

An example of Model 4 is given in Figure 2d. Even before the first event the hazard rate decreases with an abrupt increase in hazard rate upon the occurrence of an event.

5. Semiparametric frailty models for recurrent event data

The baseline hazard function is left unspecified in case of the semiparametric model. The difference between the calendar time and the gap time is now in terms of risk sets.

In calendar time (Model 6: Cox frailty-calendar), the risk set at time 0 consists of all patients being at risk for their first event only as compared to gap time (Model 7: Cox frailty-gap) for which each patient contributes a number of r_i risk periods at time 0.

The partial likelihood function in the case of the calendar time (without ties) is given by

$$\prod_{i=1}^{N} \prod_{j=0}^{r_i} \left[\frac{U_i \exp(\beta x_i)}{\sum_{k=1}^{N} Y_k(t_{ij2})\, U_k \exp(\beta x_k)} \right]^{\delta_{ij}}$$

with

$$Y_k(t_{ij2}) = \begin{cases} 1 & \text{if patient } k \text{ at risk at time } t_{ij2} \\ 0 & \text{otherwise} \end{cases}$$

and the partial likelihood function in the case of the gap time (without ties) is given by

$$\prod_{i=1}^{N} \prod_{j=0}^{r_i} \left[\frac{U_i \exp(\beta x_i)}{\sum_{k=1}^{N} \sum_{l=0}^{r_k} Y_{kl}(t_{ij2})\, U_k \exp(\beta x_k)} \right]^{\delta_{ij}}$$

with

$$Y_{kl}(t_{ij2}) = \begin{cases} 1 & \text{if } (t_{kl2} - t_{kl1}) \geq (t_{ij2} - t_{ij1}) \\ 0 & \text{otherwise} \end{cases}$$

It is shown in Klein (1992) how parameter estimates for these Cox models can be obtained by the EM-algorithm, making use of these partial likelihood expressions in the maximisation step. In order to study the effect of the frailty term in the Cox model, Cox models without the frailty terms are additionally fitted to the data, both for the calendar (Model 8: Cox -calendar) and the gap (Model 9: Cox frailty-calendar) time.

6. Application to recurrent asthma event data

Parameter estimates of the different models introduced in Sections 3, 4 and 5 are shown in Table 1 for the asthma recurrent event data. For Models 1 and 2 (not taking into account the different nature of the first event), the model based on calendar time leads to a larger treatment effect and heterogeneity estimate than the model based on gap time. With respect to the evolution of the recurrent event rate since time of randomisation (calendar time), the baseline hazard rate is close to a constant value ($\lambda_c = 0.2299$) with $\gamma_c = 1.02935$. Thus the recurrent event rate does not seem to increase or decrease as a function of time since randomisation. On the other hand, the baseline hazard rate for gap time is decreasing over time, with $\gamma_g = 0.8286$. Thus the recurrent event rate is larger immediately after an event and is decreasing with time since last event.

When extending the gap time model to include a constant and different hazard rate for the first event (Model 3), the AIC is decreasing substantially and thus this extended model seems to fit the data better. The parameter estimates for the extended model are similar, but the decrease of the recurrent event rate with time since the last event is more pronounced ($\gamma_g = 0.762$) and the constant hazard rate for the first event is substantially smaller ($\lambda_f = 0.217$) than for the subsequent events. When replacing the exponential distribution for the first event time with the Weibull distribution (Model 4), the AIC only decreases slightly. The hazard rate for the first event time increases with time ($\gamma_f = 1.105$). The last parametric model (Model 5) where both timescales are considered jointly leads to a further decrease of the treatment effect. The hazard rate is now decreasing as a function of time since randomisation as well ($\gamma_c = 0.9458$), and the decrease of the recurrent event rate with time since last event is even more pronounced $\gamma_g = 0.687$. This model, however, has a much higher AIC than the two previous models.
Finally, the semiparametric models lead to similar results as their parametric counterparts. In the case of calendar time (Model 6), the treatment effect and heterogeneity estimates are almost the same, whereas in the gap time, both the treatment effect and the heterogeneity are slightly smaller in the semiparametric model (Model 7). Finally, the exclusion of the frailty term does not have a large impact on the estimation of the treatment effect as can be seen in Models 8 and 9. Algorithms have been developed in R (available from http://eduforum.rug.ac.be/ biometrie/software.html) to fit all previous models for the Weibull and the lognormal distribution. The different models are depicted in Figure 2 using as parameter values the estimates obtained from the recurrent asthma event data.

7. Discussion

The use of different timescales leads to quite different interpretation of the data as in the asthma recurrent event data example. The calendar timescale and the gap timescale model two different and important aspects of the data. In the gap timescale, the effect of a recurrent event on the recurrent event rate of the subsequent event can be studied, and this seems to be the most important time evolution of the recurrent event rate in the example, together with the distinct hazard rate for the first event. In the calendar timescale, evolution of the recurrent event rate since randomisation is studied, and there seems to be only a marginal time effect on the recurrent event rate in the example, regardless whether this effect is modelled alone or jointly with the gap timescale. It is remarkable that the parametric and semiparametric models lead to almost the same parameter estimates when

Table 1. Parameter estimates with standard error of different models based on calendar time, gap time or both.

Model	β (s.e.)	θ (s.e.)	λ (s.e.)	γ (s.e.)	Parameter	AIC
1.Weibull -	-0.2998	0.5737	0.2299	1.02935	λ_c/γ_c	3906.8
calendar	(-0.0152)	(0.0055)	(0.00092)	(0.00126)		
2.Weibull -	-0.2544	0.4017	0.3159	0.82862	λ_g/γ_g	3863.0
gap	(0.0121)	(0.0041)	(0.00066)	(0.00059)		
3.Weibull - gap	-0.2513	0.3716	0.3464	0.7620	λ_g/γ_g	3838.4
first event exp	(0.0117)	(0.0040)	(0.00083)	(0.00068)		
			0.2172		λ_f	
			(0.00047)			
4.Weibull - gap	-0.2585	0.3912	0.3439	0.7641	λ_g/γ_g	3837.2
first event Weibull	(0.0121)	(0.0044)	(0.00084)	(0.00068)		
			0.1780	1.105	λ_f/γ_f	
			(0.00076)	(0.00374)		
5.Weibull-gap	-0.2451	0.4212	0.134	0.687	λ_g/γ_g	3885.4
calendar	(0.0126)	(0.0045)	(0.00073)	(0.00203)		
			0.2098	0.9458	λ_c/γ_c	
			(0.00203)	(0.00085)		
6.Cox frailty-	-0.302	0.579				
calendar	(0.123)					
7.Cox frailty-	-0.241	0.396				
gap	(0.11)					
8.Cox -	-0.308					
calendar	(0.071)					
9.Cox-	-0.221					
gap	(0.071)					

using either gap time or calendar time: the difference between gap timescale and calendar timescale is far larger than between parametric and semiparametric models.
If it is expected that the recurrent event rate will not change as a function of the time since randomisation, the analysis can be based solely on the gap timescale and it can thus be studied how the hazard rate evolves after an event has taken place. For instance in asthma trials with adults followed over time, a dramatic change of the recurrent event rate is not expected. In our example, however, which was based on young children, it was expected that the recurrent event rate would evolve over time, so that in such cases we should fit the different models presented in the paper. It turned out, however, that also for these young children, the recurrent event rate was not changing substantially over time but did so as a function of the time since the last event.

Acknowledgements

Paul Janssen acknowledges partial support by the Ministery of the Flemisch Community (Project BIL00/28, International Scientific and Technological Cooperation) and by the Interuniversity Attraction Pole Programme P5/24 - Belgian State - Federal Office for Scientific, Technical and Cultural Affairs.

References

Duchateau, L., Janssen, P., Lindsey, P., Legrand, C., Nguti, R. and Sylvester, R. (2002). The shared frailty model and the power for heterogeneity tests in multicenter trials. *Computational Statistics and Data Analysis*, **40**, 603-620.

Kelly, P.J. and Lim, L. (2000). Survival analysis for recurrent event data: an application to childhood infectious diseases *Statist. Med.*, **19**, 13-33.

Klein, J.P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*, **48**, 795-806.

Mahé, C. and Chevret, S. (1999) Estimation of the treatment effect in a clinical trial when recurrent events define the endpoint. *Statist. Med.*, **18**, 1821-1829.

McGilchrist, C.A. and Aisbett, C.W. (1991) Regression with frailty in survival analysis. *Biometrics*, **47**, 461-466.

Prentice, R.L., Williams, B.J. and Peterson, A.V. (1981) On the regression analysis of multivariate failure time data. *Biometrika*, **68**, 373-379.

Wassell, J.T., Wojciechowski, W.C. and Landen, D.D. (1999) Recurrent injury event-time analysis. *Statist. Med.*, **18**, 3355-3363.

Wei, L.J., Lin, D.Y. and Weissfeld, L. (1989) Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Stat. Assoc.*, **84**, 1065-1073.

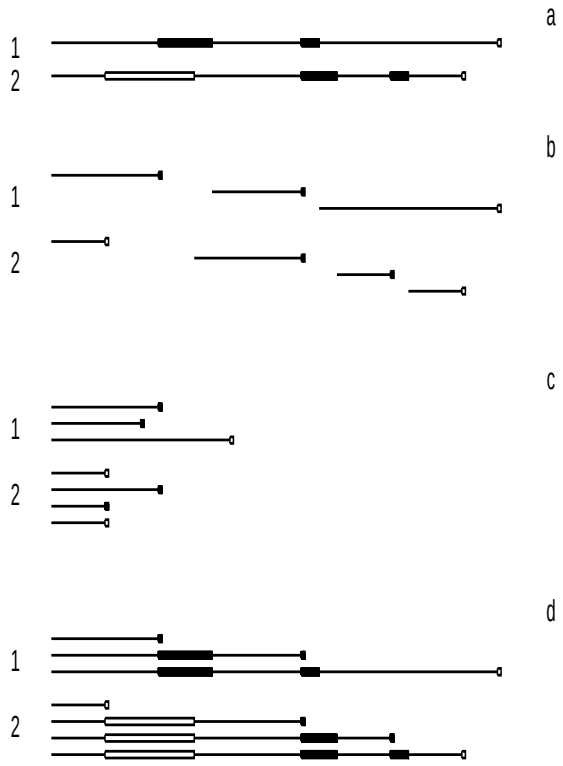


Fig. 1. The event history for two patients (a) where a line denotes time at risk, a filled box corresponds to asthma attack time and an empty box is censored time, together with the calendar time (b), gap time (c) and total time (d) representation.

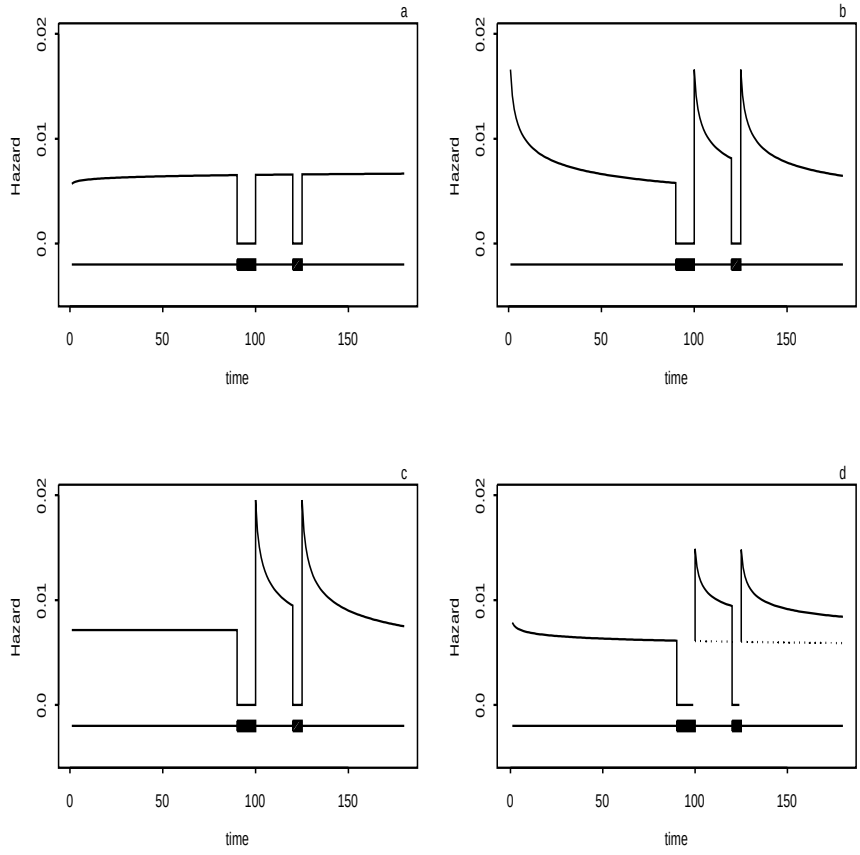


Fig. 2. The hazard as a function of time for a subject depicted at the bottom of the picture according to (a) Model 1: Weibull-calendar, (b) Model 2: Weibull-gap, (c) Model 3: Weibull-gap-first event exp and (d) Model 5: Weibull-gap-calendar