PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Overall Distribution of Survival Times for U.K. AIDS Patients

Gillian K. Reeves

Phil. Trans. R. Soc. Lond. B 1989 **325**, 147-151
doi: 10.1098/rstb.1989.0079

Phil. Trans. R. Soc. Lond. B **325**, 147–151 (1989) [147]
Printed in Great Britain

THE OVERALL DISTRIBUTION OF SURVIVAL TIMES FOR U.K. AIDS PATIENTS

By GILLIAN K. REEVES
Department of Mathematics, Imperial College of Science and Technology, London SW7 2BZ, U.K.

Data on the survival times of 997 U.K. AIDS patients are analysed with the aim of deriving a simple form for the overall survival distribution. The exponential and Weibull distributions are modified to accommodate specific features of the data, in particular, the recording of survival times to the nearest month and the occurrence of a significant proportion of cases recorded as having zero time on study. The final model has a probability 0.08 of underlying survival time being zero and, given non-zero survival time, takes the form of an exponential distribution with mean of 14.95 months. The results are in close agreement with those of a study of New York City patients as well as the empirical data.

1. Introduction

The aim of this paper is to find a simple representation for the overall distribution of survival time (time from diagnosis to death) for U.K. AIDS patients. The data consist of information on the first 997 U.K. AIDS cases; that is, all cases reported to the Communicable Disease Surveillance Centre and the Communicable Diseases (Scotland) Unit by the end of September 1987. For an analysis of survival time as a dependent variable based on the same data see Overton *et al.* (1988). The data contain the following features.

A significant proportion of cases have survival times that are effectively zero (diagnosis and death virtually simultaneous). Fewer individuals have survival times that are relatively long. The study of short survival times is further complicated by the fact that recording was to the nearest month.

The first two points mean that models that allow for the possibility of subpopulations at a much greater or much lower risk than the remainder of the population need to be considered; this is a common requirement in survival analysis. The above issues were studied by the maximum likelihood fitting of models based on the commonly used exponential and Weibull distributions.

2. The models

2.1. *Modification of the basic exponential and Weibull distributions*

First, although some of the recorded zeros are accountable via the continuous distribution because of the rounding, it was found necessary to introduce a point concentration at zero into the basic continuous distributions; thus the new density function, $f'(t)$, and survivor function, $S'(t)$, are given by

$$f'(t) = \begin{cases} \theta & (t = 0), \\ (1-\theta) f(t) & (t > 0), \end{cases}$$

$$S'(t) = \begin{cases} (1-\theta) & (t = 0), \\ (1-\theta) S(t) & (t > 0), \end{cases}$$

where $f(t)$ and $S(t)$ denote the basic Weibull or exponential density function and survivor function, respectively and θ represents the probability of zero survival time.

To check whether the simple functional form accounts for long term survivors, we can examine models in which there is a non-zero probability, γ, of 'long-term' survival, i.e. in effect an atom of probability at infinity can be incorporated into the model. In this case, the density and survivor functions, respectively, are,

$$f''(t) = \begin{cases} \theta & (t = 0), \\ (1-\theta-\gamma) f(t) & (t > 0), \\ \gamma & (t = \infty), \end{cases}$$

$$S''(t) = \gamma + (1-\theta-\gamma)\, S(t),$$

where γ now denotes the probability of long term survival and $f(t)$ and $S(t)$ are simple parametric forms as before.

One further adjustment was made to all of the above models. Because all dates were recorded to the nearest month, it is necessary to distinguish between θ, the actual probability of zero survival time, and π_0, a function of θ, which represents the probability of being recorded as having zero survival time. For the present analysis it was assumed that the time of diagnosis was distributed uniformly over a month, in which case π_0 can be written as

$$\pi_0 = \int_0^1 [1-S^*(1-t)]\, \mathrm{d}t = \int_0^1 [1-S^*(u)]\, \mathrm{d}u,$$

where $S^*(t)$ is the survivor function associated with the particular model, so in the case of the single atom model $S^*(t) = S'(t)$ and similarly for the two atom model $S^*(t) = S''(t)$. If the time period involved had been more than a month, for example if dates were given only to the nearest quarter, then a more appropriate distribution for the time of diagnosis, say $g(t)$, might depend in some way on the rate of new cases in which case π_0 would be given by

$$\pi_0 = \int_0^1 g(t)\, [1-S^*(1-t)]\, \mathrm{d}t = \int_0^1 g(1+u)\, [1-S^*(u)]\, \mathrm{d}u.$$

Of course, similar versions of the above expression could be used for the probability of survival time j months $(j = 1, 2, \ldots)$ but were considered unnecessary in this case.

If the exponential distribution or Weibull distribution is adapted to incorporate this discreteness in the data as well as a single atom at zero the resulting log likelihood for a censored sample is

$$\ell = n_0 \log(\pi_0) + \sum_{i \in u} \log [f'(t_i)] + \sum_{i \in c} \log [S'(t_i)],$$

where n_0 denotes the number of patients with time on study recorded as zero, t_i denotes the recorded survival time for the ith individual, and the subgroups C and U refer to the sets of censored and uncensored individuals, respectively. The log likelihood for the two-atom model can be similarly obtained by substituting $f''(t)$ and $S''(t)$ for $f'(t)$ and $S'(t)$, respectively.

3. Application and conclusions

3.1. *Application*

Of the 997 cases discussed in §1, death dates were known for 678 cases; the last known alive date could be specified for 319 including 15 cases that were censored because they were lost to follow-up; all visitors from abroad were excluded from the analysis. The following models (designated (*a*)–(*d*)) discussed in §2, were fitted using NAG subroutines, the integral π_0 was evaluated numerically within each iteration of the maximization procedure.

(*a*) An exponential distribution supplemented by a probability, θ, of zero survival time.

(*b*) A Weibull distribution supplemented by a probability, θ, of zero survival time.

(*c*) An exponential distribution supplemented by a probability, θ, of zero survival time and by a probability, γ, of 'long-term' survival.

(*d*) A Weibull distribution supplemented by a probability, θ, of zero survival time and by a probability, γ, of 'long-term' survival.

3.2. *Conclusions*

The results are summarized in table 1. The estimates of the parameter γ in both models (*c*) and (*d*) were not only non-significant but also negligible in value indicating that the probability of surviving for a relatively long period of time is accounted for by the exponential or Weibull component.

Table 1. Parameter estimates and selected confidence intervals for models (*a*) and (*b*)[a]

model	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\theta}$	$l_{\max}$
(*a*)				
2-parameter exponential	0.067		0.080	−2450.13
(95% c.i.)	[0.061, 0.073]		[0.061, 0.101]	
(*b*)				
3-parameter Weibull	0.067	1.018	0.082	−2450.04

[a] θ, probability at zero; exponential, $S(t) = e^{-\lambda t}$; Weibull, $S(t) = e^{(-\lambda t)^{\kappa}}$.

From the results of the remaining single atom models (*a*) and (*b*) it is clear that the underlying distribution is very close to an exponential distribution as can be seen from the estimated index in the Weibull model, which is almost equal to unity. This is confirmed by the non-significance of the difference in maximized log-likelihoods corresponding to a X^2 statistic of only 0.1804. The parameter θ is of particular interest and in the case of the exponential model has a value of 0.0802; this implies that about 80 of the 110 cases recorded as having zero time on study were genuine zeros, the others being a direct result of the recording procedure.

Because of the possible asymmetry of the log-likelihoods for all of these models standard errors for the estimates were not calculated; instead, confidence intervals were obtained via the profile log-likelihoods (figures 1 and 2); these indicate that $\hat{\lambda}$, in particular, is quite well defined.

Apart from extending the model parametrically, another method for testing the 'goodness of fit' is to compare an empirical version of the survival curve with that obtained under the assumed model. In this case, not only is the difference in maximized log-likelihoods for the exponential and Weibull models non-significant but the fit to the empirical data (figure 3) is

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

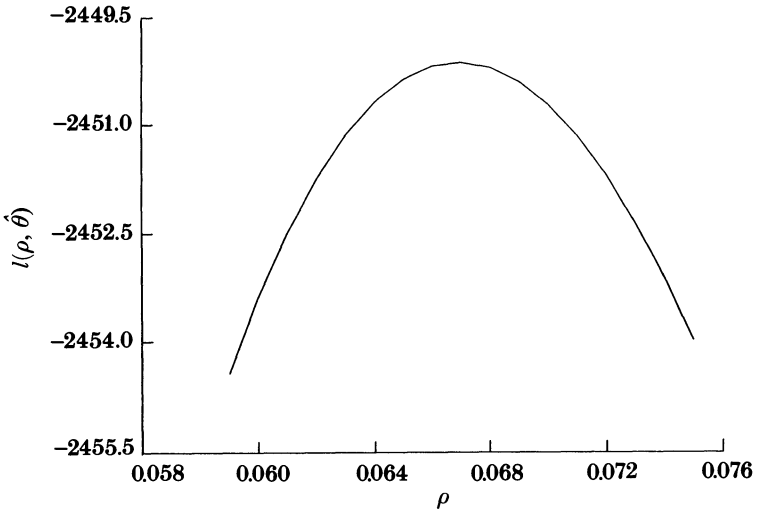

FIGURE 1. Profile log-likelihood for ρ.

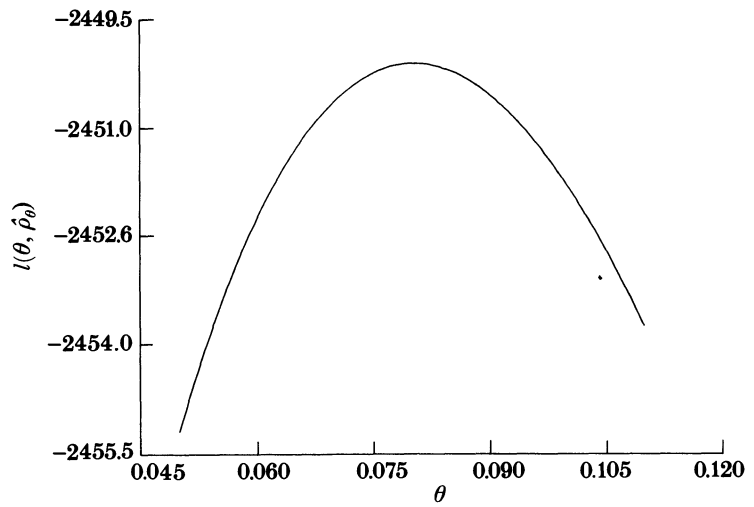

FIGURE 2. Profile log-likelihood for θ.

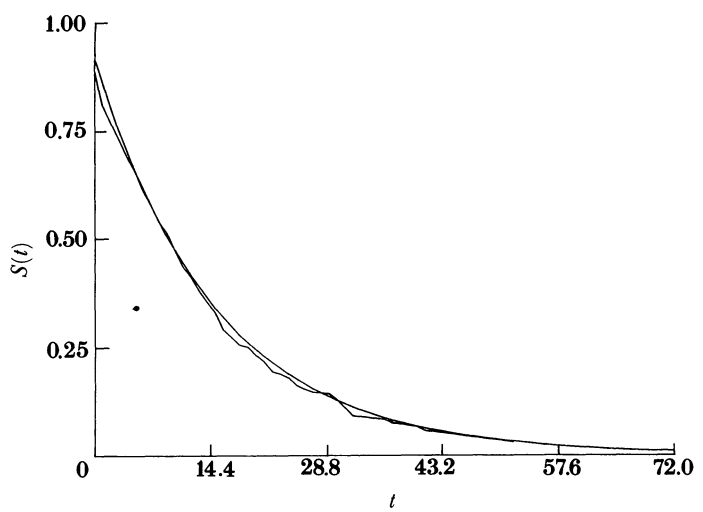

FIGURE 3. Fitted survivor function (smooth curve) and the Kaplan–Meier estimate (jagged curve); t, time/months.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

excellent. Thus it appears that the very simple and convenient exponential model (*a*) is suitable for application. That is, we take a probability 0.080 of underlying survival time time being zero. Conditionally on non-zero survival time we take an exponential distribution with mean 14.950 months. Table 2 summarizes some properties.

TABLE 2. THE FITTED DISTRIBUTION

probability of zero survival time	0.080
probability of surviving at least	
1 month	0.860
3 months	0.753
6 months	0.616
12 months	0.412
24 months	0.185
36 months	0.083
median survival time/months	9.110
mean survival time/months	13.750

These results are in remarkably close agreement with those from a study of some 5833 New York City patients (Rothenberg *et al.* 1987). The number of New York City patients whose time on study was given as zero to the nearest day was 663; this constitutes about 11.37 % of all cases, which is close to the value of 11.03 % for the U.K. data. The median survival time for the New York City data, excluding cases with time on study recorded as zero, was 11.4 months compared with a corresponding value of 10.36 months under the exponential model for the U.K.

The contribution of reporting doctors who supplied the follow-up information and of Elizabeth Overton who prepared the data set is gratefully appreciated.

REFERENCES

Overton, S. E., Juritz, J. M., Gill, O. N., Kennedy, A. N., Wright, A. L. & Marasca, G. 1988 Survival analysis of the first 1000 cases of AIDS reported in the U.K. In *Proceedings of the 4th International Conference on AIDS, Stockholm* (poster no. 4593).

Rothenberg, R., Woelfe, M., Stoneburner, R., Milberg, J., Parker, R. & Truman, B. 1987 Survival with the acquired immunodeficiency syndrome. *New Engl. J. Med.* **317**, 1297–1302.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES