

Statistics Notes

The logrank test

J Martin Bland, Douglas G Altman

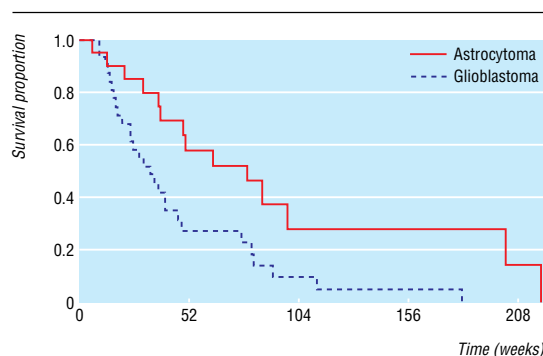
We often wish to compare the survival experience of two (or more) groups of individuals. For example, the table shows survival times of 51 adult patients with recurrent malignant gliomas¹ tabulated by type of tumour and indicating whether the patient had died or was still alive at analysis—that is, their survival time was censored.² As the figure shows, the survival curves differ, but is this sufficient to conclude that in the population patients with anaplastic astrocytoma have worse survival than patients with glioblastoma?

We could compute survival curves³ for each group and compare the proportions surviving at any specific time. The weakness of this approach is that it does not provide a comparison of the total survival experience of the two groups, but rather gives a comparison at some arbitrary time point(s). In the figure the difference in survival is greater at some times than others and eventually becomes zero. We describe here the logrank test, the most popular method of comparing the survival of groups, which takes the whole follow up period into account. It has the considerable advantage that it does not require us to know anything about the shape of the survival curve or the distribution of survival times.

The logrank test is used to test the null hypothesis that there is no difference between the populations in the probability of an event (here a death) at any time point. The analysis is based on the times of events (here deaths). For each such time we calculate the observed number of deaths in each group and the number expected if there were in reality no difference between the groups. The first death was in week 6, when one patient in group 1 died. At the start of this week, there were 51 subjects alive in total, so the risk of death in this week was $1/51$. There were 20 patients in group 1, so, if the null hypothesis were true, the expected number of deaths in group 1 is $20 \times 1/51 = 0.39$. Likewise, in group 2 the expected number of deaths is $31 \times 1/51 = 0.61$. The second event occurred in week 10, when there were two deaths. There were now 19 and 31 patients at risk (alive) in the two groups, one having died in week 6, so the probability of death in week 10 was $2/50$. The expected numbers of deaths were $19 \times 2/50 = 0.76$ and $31 \times 2/50 = 1.24$ respectively.

The same calculations are performed each time an event occurs. If a survival time is censored, that individual is considered to be at risk of dying in the week of the censoring but not in subsequent weeks. This way of handling censored observations is the same as for the Kaplan-Meier survival curve.³

From the calculations for each time of death, the total numbers of expected deaths were 22.48 in group 1 and 19.52 in group 2, and the observed numbers of deaths were 14 and 28. We can now use a χ^2 test of the null hypothesis. The test statistic is the sum of $(O - E)^2/E$ for each group, where O and E are the totals of the observed and expected events. Here $(14 - 22.48)^2 / 22.48 + (28 - 19.52)^2 / 19.52 = 6.88$. The degrees of freedom are the number of groups minus one, i.e.



Survival curves for women with glioma by diagnosis

$2 - 1 = 1$. From a table of the χ^2 distribution we get $P < 0.01$, so that the difference between the groups is statistically significant. There is a different method of calculating the test statistic,⁴ but we prefer this approach as it extends easily to several groups. It is also possible to test for a trend in survival across ordered groups.⁴ Although we have shown how the calculation is made, we strongly recommend the use of statistical software.

The logrank test is based on the same assumptions as the Kaplan Meier survival curve³—namely, that censoring is unrelated to prognosis, the survival probabilities are the same for subjects recruited early and late in the study, and the events happened at the times specified. Deviations from these assumptions matter most if they are satisfied differently in the groups being compared, for example if censoring is more likely in one group than another.

The logrank test is most likely to detect a difference between groups when the risk of an event is consistently greater for one group than another. It is unlikely to detect a difference when survival curves cross, as can happen when comparing a medical with a surgical intervention. When analysing survival data, the survival curves should always be plotted.

Because the logrank test is purely a test of significance it cannot provide an estimate of the size of the difference between the groups or a confidence interval. For these we must make some assumptions about the data. Common methods use the hazard ratio, including the Cox proportional hazards model, which we shall describe in a future *Statistics Note*.

Competing interests: None declared.

1 Rostomily RC, Spence AM, Duong D, McCormick K, Bland M, Berger MS. Multimodality management of recurrent adult malignant gliomas: results of a phase II multiagent chemotherapy study and analysis of cytoreductive surgery. *Neurosurgery* 1994;35:378-8.

2 Altman DG, Bland JM. Time to event (survival) data. *BMJ* 1998;317:468-9.

3 Bland JM, Altman DG. Survival probabilities. The Kaplan-Meier method. *BMJ* 1998;317:1572.

4 Altman DG. *Practical statistics for medical research*. London: Chapman & Hall, 1991: 371-5.

Department of Health Sciences, University of York, York YO10 5DD

J Martin Bland
professor of health statistics

Cancer Research UK/NHS Centre for Statistics in Medicine, Institute of Health Sciences, Oxford OX3 7LF
Douglas G Altman
professor of statistics in medicine

Correspondence to: Professor Bland

BMJ 2004;328:1073

Weeks to death or censoring in 51 adults with recurrent gliomas¹ (A=astrocytoma, G=glioblastoma)

A	G
6	10
13	10
21	12
30	13
31*	14
37	15
38	16
47*	17
49	18
50	20
63	24
79	24
80*	25
82*	28
82*	30
86	33
98	34*
149*	35
202	37
219	40
	40
	40*
	46
	48
	70*
	76
	81
	82
	91
	112
	181

*Censored survival time (still alive at follow up).