

# A simple alternative to Kaplan–Meier for survival curves

John P. Costella  
*Peter MacCallum Cancer Centre*

(September 21, 2010)

## Abstract

Survival curves in medical research are almost universally generated by the Kaplan–Meier method, despite numerous warnings over the decades of its shortcomings. In this paper I provide an alternative method that is simple to calculate and understand, and provides a better visual representation of statistical significance.

## 1. Introduction and motivation

Survival curves—depicting the fraction of patients who have not “failed” (however defined) for a given time since an initial event—are ubiquitous in medical research. That there are almost always “censored” patients—those lost to scrutiny (for whatever reason) before failure—means that we cannot compute a sensible survival curve through simple division.

The next simplest approach, first proposed by Böhmer in 1912 [1], and independently rediscovered by Meier and by Kaplan in the 1950s [2, 3], proceeds as follows: At every time  $t_i$  that there is a failure, we determine the number  $n(t_i)$  of patients “at risk” (*i.e.*, still under scrutiny; not lost through censoring or failure). We then estimate that the underlying probability of a patient at risk at  $t_i$  surviving that “danger time”  $t_i$  is

$$\Pr(\text{survive danger time } t_i) \approx \frac{n(t_i) - 1}{n(t_i)}. \quad (1)$$

The logic is that if one of the  $n$  patients at risk failed at that time, then on average it’s probably the same as each of them having a probability of  $1/n$  of failing at that time, *i.e.*, a probability of  $(n-1)/n$  of surviving that “danger time”. If more than one patient fails at the same time  $t_i$  (to within our measurement resolution)—say,  $d_i$  of them fail at  $t_i$  (where  $d_i = 1, 2, \dots$ )—then we just generalize (1) to

$$\Pr(\text{survive danger time } t_i) \approx \frac{n(t_i) - d_i}{n(t_i)}. \quad (2)$$

To determine the probability  $S(t)$  that a given initial patient would survive (if uncensored) to time  $t$ , the Kaplan–Meier method assumes that the probability of surviving each “danger time”  $t_i$  is statistically independent of every other “danger time”, which means that we can just multiply together the probabilities of surviving all “danger times” less than  $t$ :

$$S(t) \approx \prod_{t_i < t} \frac{n(t_i) - d_i}{n(t_i)}. \quad (3)$$

As a simplistic example, imagine we start with 10 patients with a maximum observation time of 7 years; we lose 6 of them to censoring at 1.0, 3.4, 3.7, 5.8, 6.1, and 7.0 years; and the other 4 fail at 1.2, 3.9, 4.1, and 6.3 years. (Any “time” in this paper is the time interval from the “initial event”, however defined, for each individual patient. We use years as the unit of time for definiteness only.) The Kaplan–Meier method assumes that the “danger times” are  $t_1 = 1.2$  years,  $t_2 = 3.9$  years,  $t_3 = 4.1$  years, and  $t_4 = 6.3$  years. The number of patients at risk at each of the “danger times” is  $n(t_1) = 9$  (none have failed but 1 has been censored before 1.2 years),  $n(t_2) = 6$  (we have lost 1 to failure and 3 to censoring before 3.9 years),  $n(t_3) = 5$ , and  $n(t_4) = 2$ . Thus, from (1) or (2), the Kaplan–Meier method estimates that

$$\begin{aligned} \Pr(\text{survive danger time 1.2 years}) &\approx \frac{9-1}{9} = \frac{8}{9}, \\ \Pr(\text{survive danger time 3.9 years}) &\approx \frac{6-1}{6} = \frac{5}{6}, \\ \Pr(\text{survive danger time 4.1 years}) &\approx \frac{5-1}{5} = \frac{4}{5}, \\ \Pr(\text{survive danger time 6.3 years}) &\approx \frac{2-1}{2} = \frac{1}{2}. \end{aligned}$$

Using Eq. (3), the Kaplan–Meier estimate of the survival curve  $S(t)$  is then

$$S(t) \approx \begin{cases} 1 & \text{for } t < 1.2 \text{ years,} \\ \frac{8}{9} \approx 0.889 & \text{for } 1.2 \leq t < 3.9 \text{ years,} \\ \frac{8}{9} \times \frac{5}{6} \approx 0.741 & \text{for } 3.9 \leq t < 4.1 \text{ years,} \\ \frac{8}{9} \times \frac{5}{6} \times \frac{4}{5} \approx 0.593 & \text{for } 4.1 \leq t < 6.3 \text{ years,} \\ \frac{8}{9} \times \frac{5}{6} \times \frac{4}{5} \times \frac{1}{2} \approx 0.296 & \text{for } 6.3 \leq t < 7.0 \text{ years.} \end{cases}$$

This is shown graphically in Fig. 1. By convention, short vertical lines are used to denote the times at which patients were lost to censoring. A more realistic example is shown in Fig. 2, taken from the sample data provided by Bland and Altman [4].

Such Kaplan–Meier curves have attractive properties, which perhaps explains their popularity in medical research for over half a century: they provide a visual depiction of all of the raw data—the failure times (the “steps” down) and the censoring times (the vertical bars)—yet they also provide a mathematical estimate of the underlying probabilistic model. But their apparent simplicity masks severe drawbacks, which has led to rightful criticism of indiscriminating use of the Kaplan–Meier method over the decades [5, 6, 7, 8, 9].

Firstly, the vertical drop at each actual failure draws undue visual attention to those particular “danger times”, with the Kaplan–Meier estimate of  $S(t)$  remaining unchanged until the next failure is encountered. In reality, no practitioner would believe that patients are only at risk at specific times; rather, they are in continuous danger of failure, with the degree of danger perhaps changing with time.

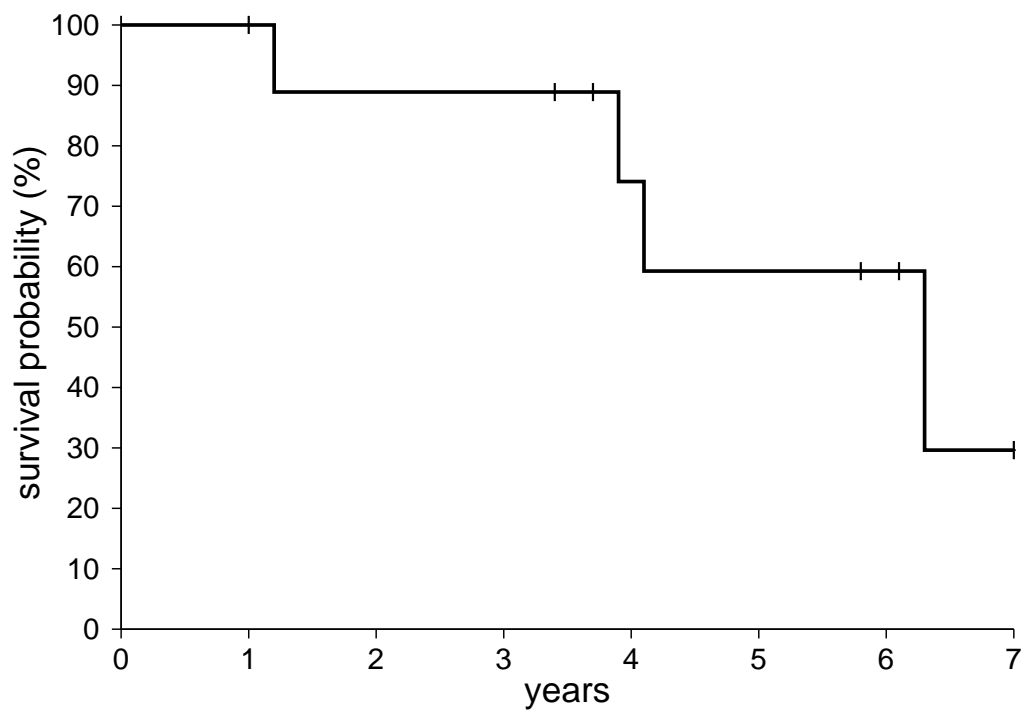


Figure 1: Kaplan–Meier curve for the simple example described in the text.

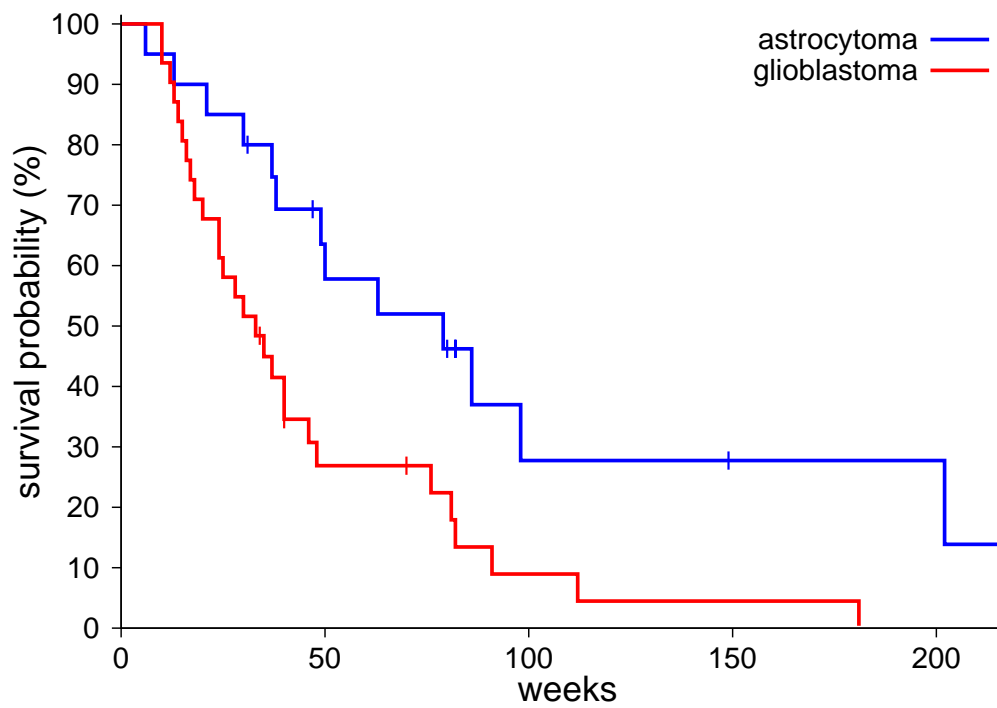


Figure 2: Kaplan–Meier curves for the example of Bland and Altman [4]. (Color-coded.)

Secondly, as time progresses, there are fewer remaining patients at risk. This has two direct effects on the Kaplan–Meier curve: the interval between failures grows, and the effect of each individual failure on the size of the step-down increases. Thus, the visual impact of a single failure is unjustifiedly magnified in both the horizontal and the vertical directions if it occurs at a later time. This is particularly evident in Fig. 2.

Thirdly, if the last remaining patient at risk fails, the Kaplan–Meier estimate of  $S(t)$  drops to zero at that time, whereas the true  $S(t)$  will never reach zero in any physically sensible model. Again, this is evident in both of the curves in Fig. 2.

The fourth drawback of the Kaplan–Meier method is somewhat more subtle. Consider the simple example analyzed above. The Kaplan–Meier estimate of the probability of surviving each “danger time” depends only on the number of patients at risk at that time; for each censored patient, it disregards the time between the last failure and the time of censoring. Consider Fig. 1: the Kaplan–Meier calculations would be unchanged if the patients censored at 3.4 and 3.7 years had instead been censored at 1.3 years, or if those censored at 5.8 and 6.1 years had instead been censored at 4.2 years; in other words, the information contained in these extra years of survival are discarded by the Kaplan–Meier process.

Additions to and modifications of the Kaplan–Meier method have been suggested over the decades, to ameliorate some of these problems, with varying degrees of success.

To draw attention to the second problem—the distorting magnification due to a small number of patients at risk—it has now become standard practice to list the number of patients at risk below the graph at various time points. However, this simply forces the reader to try to use their intellect to override what their own eyes are telling them: if the number of patients at risk becomes small, ignore the largest region of the graph!

To partially repair the mathematical bias of the fourth problem—discarding information between the last failure and the time of censoring—some practitioners and statistical packages now use as denominator the number of patients at risk at the *mid-point* of the time interval between failures, rather than that at the right end. This provides only a leading-order correction, and only if one can assume that the times of censoring are randomly distributed—but such is often not the case in real life. This options now also means that there are at least two subtly different definitions of the Kaplan–Meier estimation method in widespread use.

The second, third, and fourth drawbacks listed above can also be ameliorated somewhat by displaying appropriate uncertainty curves. However, despite the relevant estimate of the variance of the error being provided by Major Greenwood in 1926 [10], it is remarkable that many authors fail to provide uncertainty curves to this day. Fig. 3 shows the curves corresponding to one and two standard errors for the simple example of Fig. 1. Not surprisingly, the amount of information that can be gleaned from just four failures is not great. Fig. 4 does the same for the Bland and Altman example of Fig. 2. Note that Greenwood’s formula for the error variance drops to zero if the Kaplan–Meier estimate of survival drops to zero, which is clearly wrong. However, the formula itself relies on the approximate normality of the distribution, which becomes a bad assumption when the remaining number of patients at risk is less than a couple of dozen, and fails completely when it drops to zero.

Despite this drawback, these uncertainty curves provide a far better picture of the underlying Kaplan–Meier estimate than simply the raw Kaplan–Meier curves. In the case of Fig. 4, it is visually apparent that the two-standard-error envelope of the survival curve for

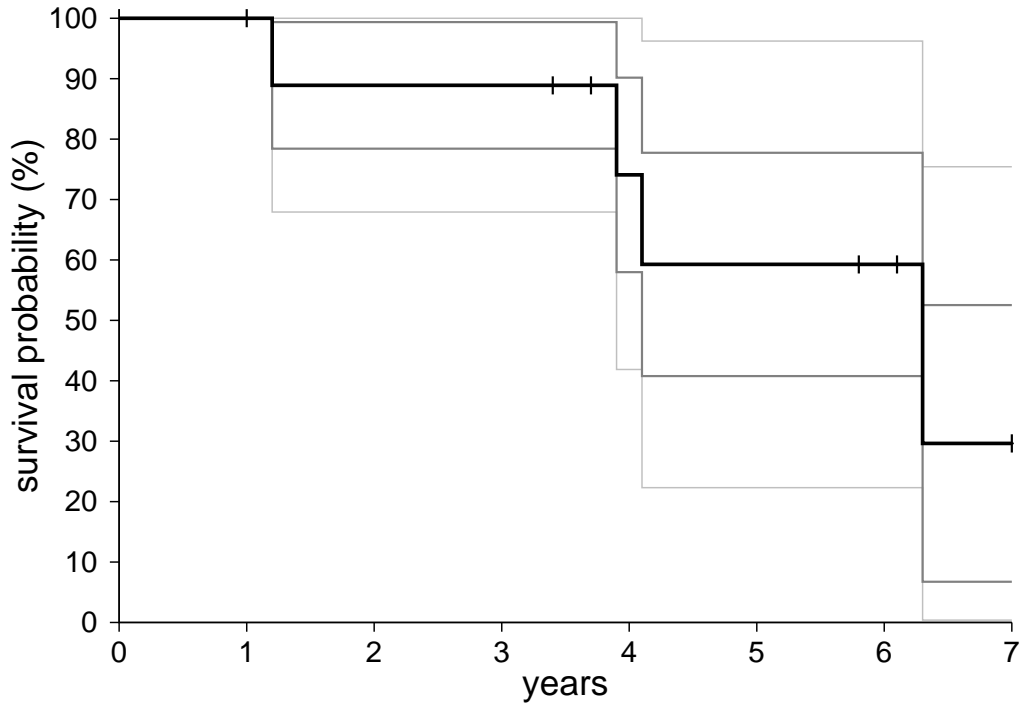


Figure 3: The Kaplan–Meier curve of Fig.1 with uncertainty curves corresponding to one (darker gray) and two (paler gray) standard errors, using Greenwood’s formula.

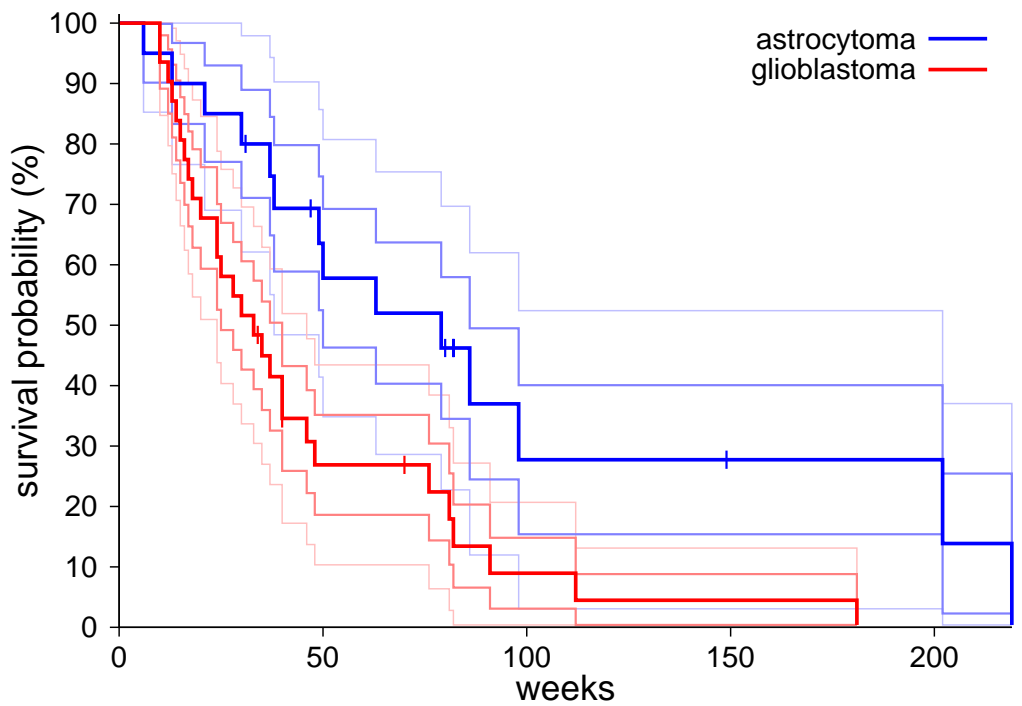


Figure 4: The Kaplan–Meier curves of Fig.2 with uncertainty curves corresponding to one and two standard errors, using Greenwood’s formula. (Color-coded.)

patients with astrocytoma (palest blue) has very little overlap with that of patients with glioblastoma (palest red), at least after 30 weeks. Accordingly, as Bland and Altman show explicitly, the logrank test confirms that the “average” hazard rates for astrocytoma and glioblastoma are statistically significantly different.

Despite these improvements, Figs. 3 and 4 still possess the main drawback of the Kaplan–Meier method: the “stepping down” of the estimate of the survival curve at the actual failure times, and constancy elsewhere. In the next section we will propose an alternative method of presentation and calculation for survival data that minimizes these visual artifacts, while retaining as much connection with the familiar features of the Kaplan–Meier method as possible.

## 2. An alternative method

### 2.1. Patient-time

Arguably the worst defect of a Kaplan–Meier graph is time axis: as the number of patients at risk,  $n(t)$ , decreases, the statistical significance of the remaining patient cohort per unit time likewise decreases; but the time axis of a Kaplan–Meier graph does not take any note of this loss of statistical power: it gives every unit of time the same amount of horizontal space, regardless of whether the number of patients at risk is one or one thousand.

Let us remedy this defect by defining a new time parameter, the *patient-time*, which we shall denote  $\tau$ , and define formally as

$$d\tau \equiv n(t) dt. \tag{4}$$

The patient-time  $\tau$  “ticks more slowly” as the remaining number of patients at risk,  $n(t)$ , decreases:  $\tau$  is a measure of *patient-years*, whereas  $t$  is simply a measure of *years*. Setting  $\tau=0$  at  $t=0$  and integrating Eq. (4), we then formally have

$$\tau(t) = \int_0^t dt' n(t'). \tag{5}$$

Despite this formal definition, computing  $\tau$  in practice for a given set of failure data is straightforward. Consider the simple example described in Sec. 1. We have  $n(t) = 10$  patients at risk from  $t = 0$  years through to  $t = 1.0$  years, so  $\tau$  increases from  $\tau = 0$  patient-years at  $t = 0$  to  $\tau = 0 + 10 \times 1.0 = 10$  patient-years at  $t = 1.0$  years. After we lose one patient to censoring at  $t = 1.0$  years, we have  $n(t) = 9$  patients at risk until another patient fails at  $t = 1.2$  years, so  $\tau = 10 + 9 \times 0.2 = 11.8$  patient-years at  $t = 1.2$  years. Continuing on in this manner, we obtain the mapping of  $t$  to  $\tau$  shown in Fig. 5. The 10 patients in this example provide a total of 42.5 patient-years of data. Note that, unlike the case with the Kaplan–Meier method, no information is thrown away: if any censored patient had been censored at an earlier time, they would have made a smaller contribution to  $\tau$ ; the mapping between  $t$  and  $\tau$  would be altered, and the total amount of patient-time would be less than 42.5 patient-years.

Likewise, for the Bland and Altman example [4], we obtain the mappings between  $t$  and  $\tau$  shown in Fig. 6. Note that relationship between  $t$  and  $\tau$  is specific to each particular set of failure data.

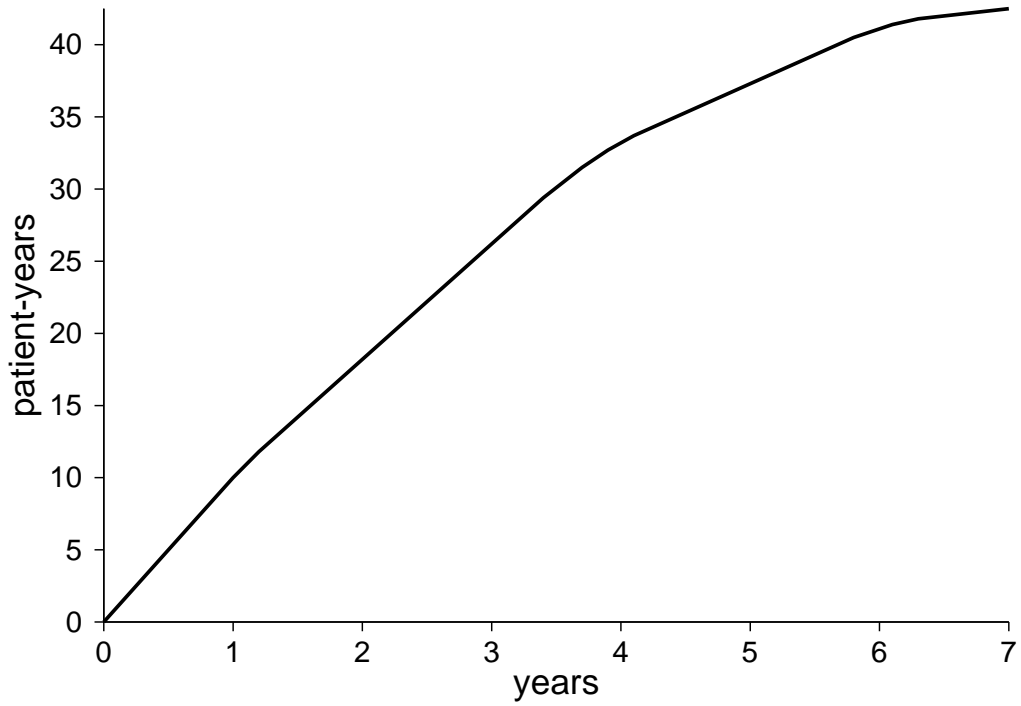


Figure 5: Mapping between time,  $t$ , measured in years, and the patient-time,  $\tau$ , measured in patient-years, for the simple example provided in Sec. 1.

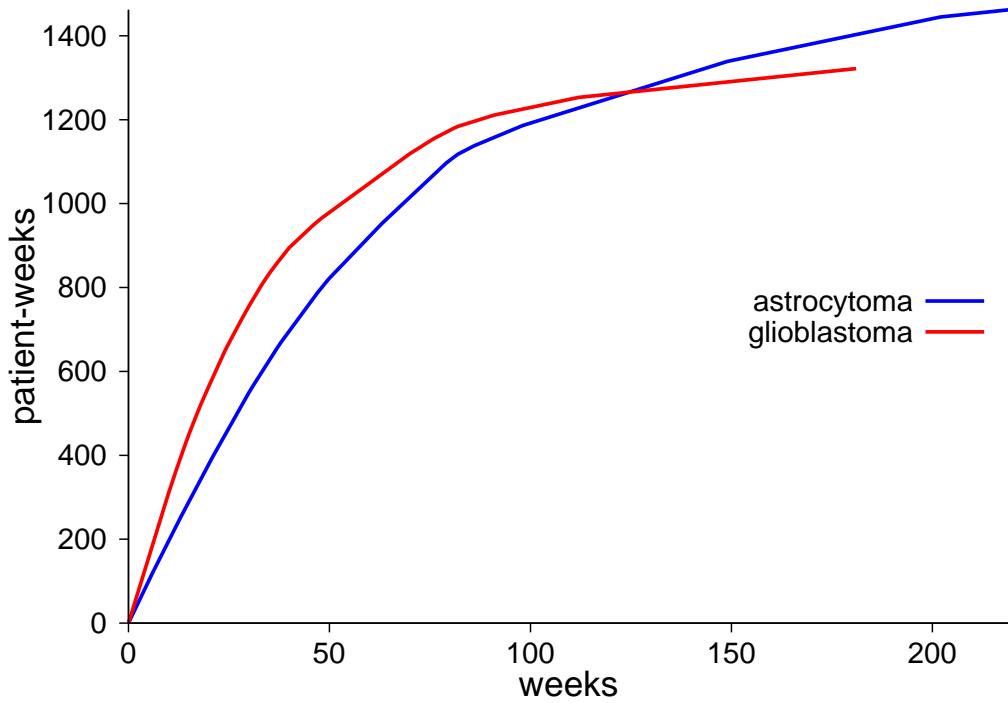


Figure 6: Mapping between  $t$  and  $\tau$  for the example of Bland and Altman discussed in Sec. 1.

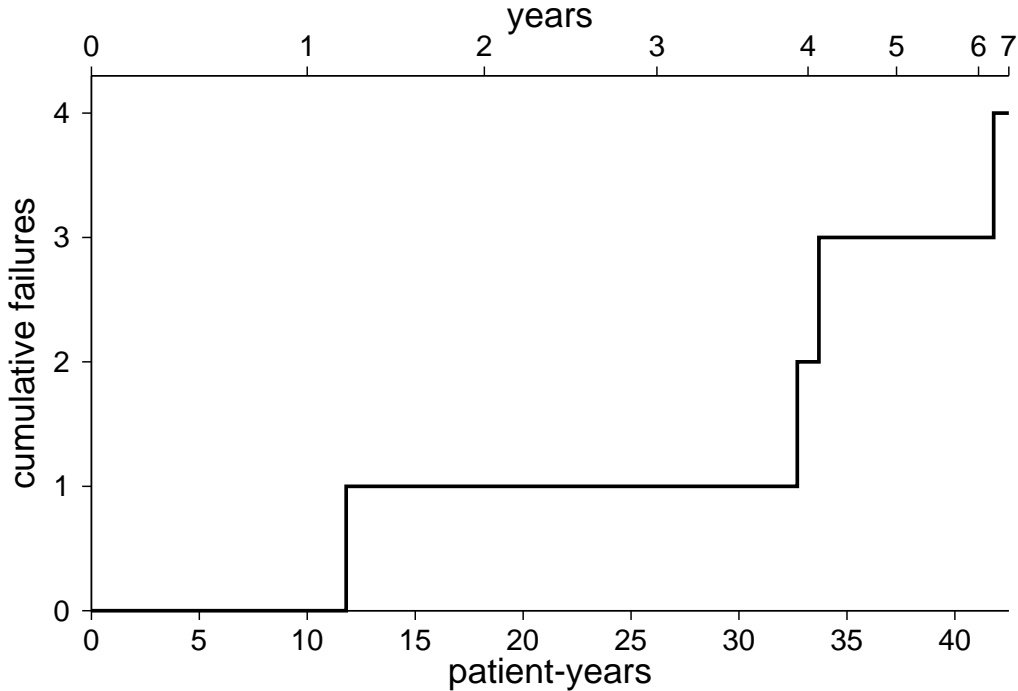


Figure 7: Cumulative failure curve for the simple example.

## 2.2. The cumulative failure graph

Consider, now, a graph of the cumulative number of failures, which we shall denote  $D$ , versus patient-time,  $\tau$ . For the simple example of Sec. 1, such a graph is shown in Fig. 7. Much information can be discerned from this graph alone. The bottom axis tells us that there is just over 40 patient-years of data; each patient-year of data is given equal horizontal distance. The vertical axis tells us that there were 4 failures in total; each failure is given equal vertical distance. For convenience, a time axis is provided along the top of the graph, obviating the need for the separate graph of Fig. 5 to be provided.

Similarly, the cumulative failure graph for the astrocytoma group of the Bland and Altman example is shown in Fig. 8, and of the glioblastoma group in Fig. 9. Unlike the corresponding Kaplan–Meier curves, the steps are quite evenly distributed. Indeed, Fig. 8 is a remarkable approximation to a straight line through the origin; and, apart from the first 300 patient-weeks (10 weeks) containing no failures, Fig. 9 is also remarkably linear.

So what does such an approximately linear portion of a cumulative failure curve signify, physically? Simply that the *hazard rate*,  $\lambda$ , is essentially *constant* during that time: the hazard rate is effectively the underlying slope of the cumulative failure curve, namely, the expected number of failures per unit of patient-time, *i.e.*,

$$\lambda(\tau) \equiv \frac{d\langle D(\tau) \rangle}{d\tau}. \quad (6)$$

Indeed, this is arguably the most intuitive *definition* of the hazard rate possible. To make connection with the standard mathematical formalism, note that the expected number of



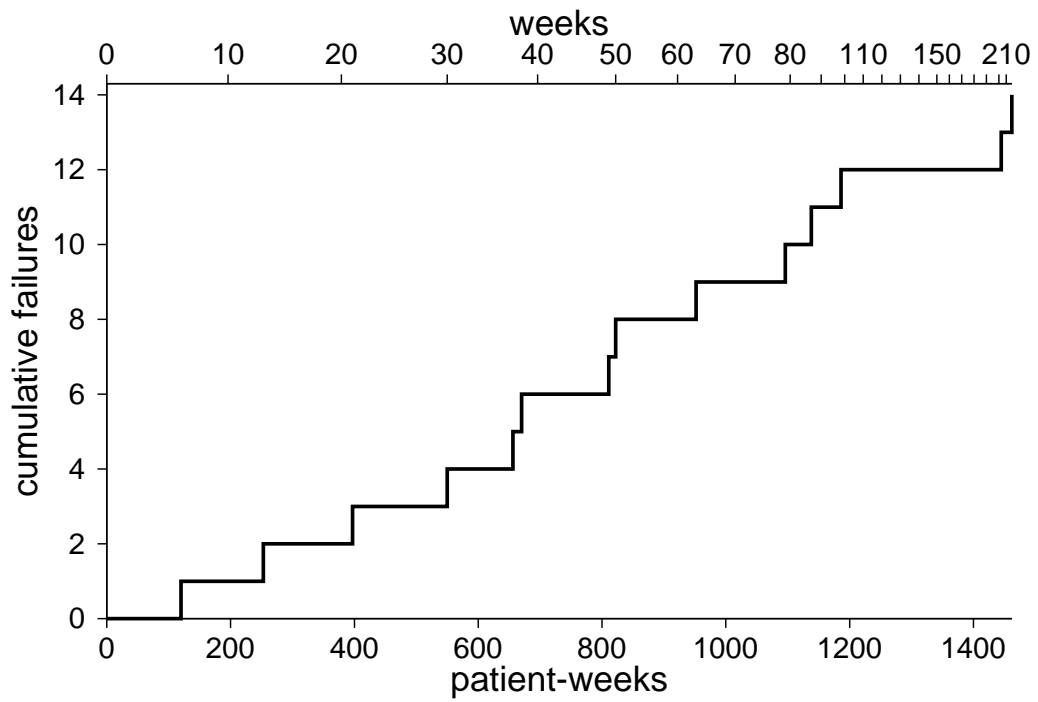


Figure 8: Cumulative failure curve for astrocytoma group of the Bland and Altman example.

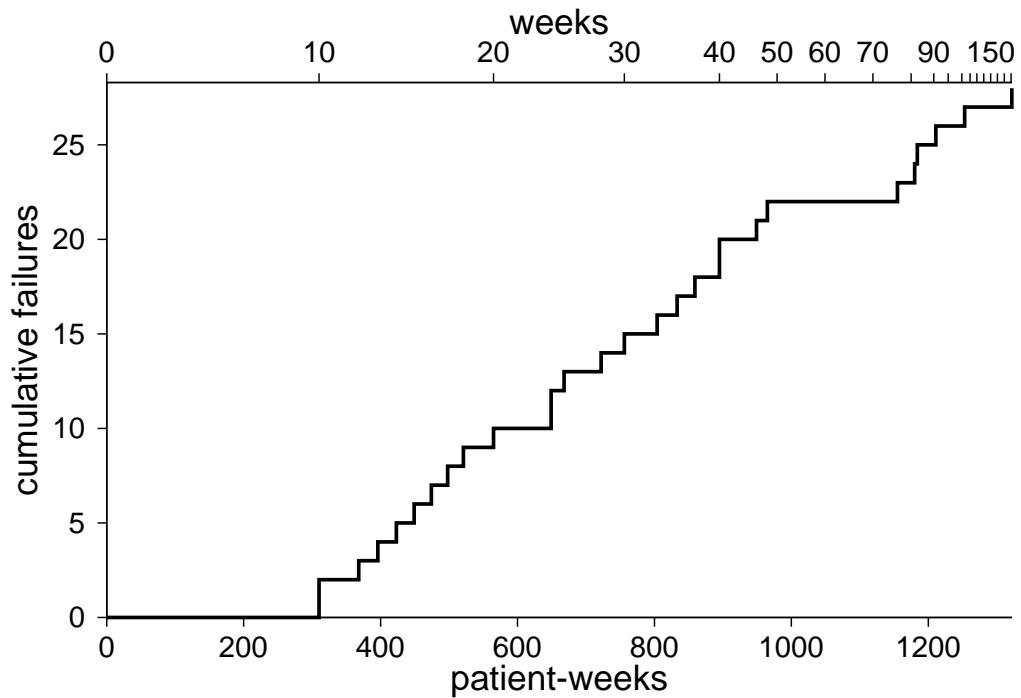


Figure 9: Cumulative failure curve for glioblastoma group of the Bland and Altman example.

cumulative failures to time  $t$  is

$$\langle D(t) \rangle \equiv \int_0^t dt' n(t') \lambda(t'),$$

so that, by variation of the upper limit of integration,

$$\frac{d\langle D(t) \rangle}{dt} = n(t) \lambda(t),$$

and by the chain rule of differentiation,

$$\frac{d\langle D(t) \rangle}{d\tau} \equiv \frac{d\langle D(t) \rangle}{dt} \frac{dt}{d\tau},$$

and from Eq. (5) we have

$$\frac{dt}{d\tau} \equiv \frac{1}{n(t)},$$

which establishes Eq. (6).

### 2.3. Estimating the hazard rate function

We shall now use the cumulative failure graph to estimate the underlying hazard rate function,  $\lambda(\tau)$ .

If we interpreted Eq. (6) naively, we might simply take the  $\tau$ -derivative of  $D(\tau)$ , which would lead to a Dirac delta function at each failure, and zero everywhere else. But this would ignore the fact that  $D(\tau)$  is simply a count of failures, and hence is subject to statistical uncertainty. Now, for a count of  $D$  failures, a reasonable estimate of the standard error of that count is  $\sqrt{D+1}$ . If we display corresponding uncertainty curves on the cumulative failure graph of Figs. 7, we obtain Fig. 10. Clearly, the uncertainties are much greater than the detailed structure implied by the steps corresponding to the four individual failures. It is therefore arguably more sensible to simply state that the data is consistent with a constant hazard rate of  $4/42.5 \approx 0.094$  failures per patient-year, with a standard error of approximately  $\sqrt{5}/42.5 \approx 0.053$  failures/patient-year. This corresponds to the model shown in green in Fig. 11.

Turning to a real-world example, Fig. 12 shows a constant hazard rate model overlaid on the astrocytoma failure data of Fig. 8. Again, it is visually obvious that this simplest model is consistent with the failure data. But how are we to formalize this consistency test? And what should we do if a constant hazard rate model does not fit the failure data?

There are many possible answers to these questions. We shall simply offer what we consider to be the simplest solution, on the understanding that more sophisticated models could be proposed and implemented.

Firstly, we shall deem a constant hazard rate model consistent with the failure data if the line of best fit (the most solid green line in Fig. 11 or Fig. 12) lies completely within the two-standard-error envelope of the failure data (the palest gray curves). This is clearly the case for each of Fig. 11 and Fig. 12.

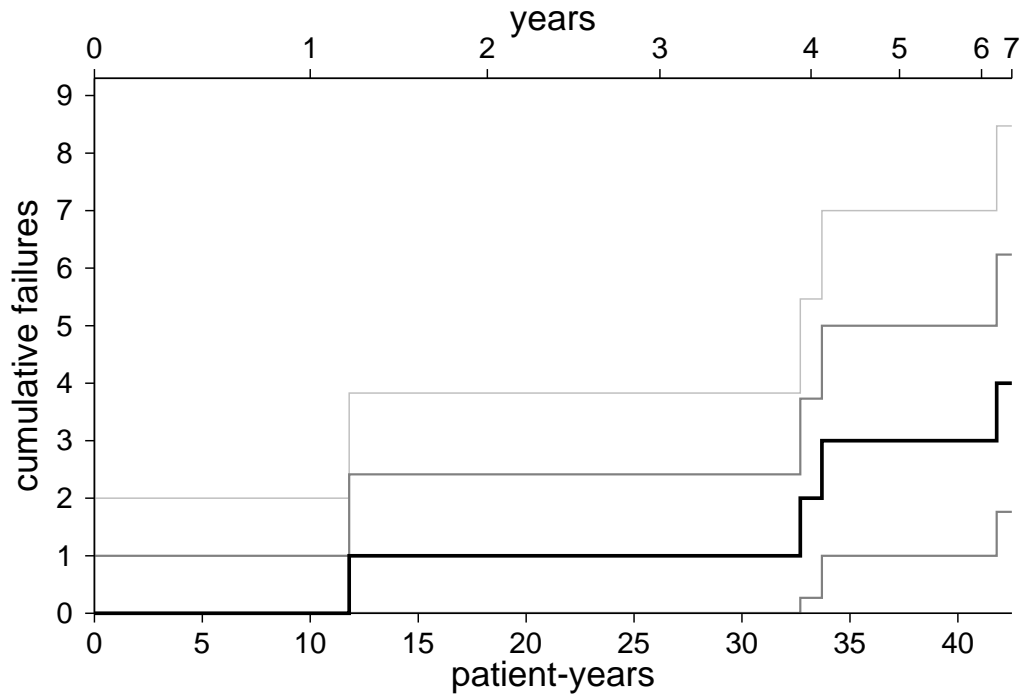


Figure 10: The cumulative failure graph of Fig. 7 for the simple example, with uncertainty curves corresponding to one and two standard errors.

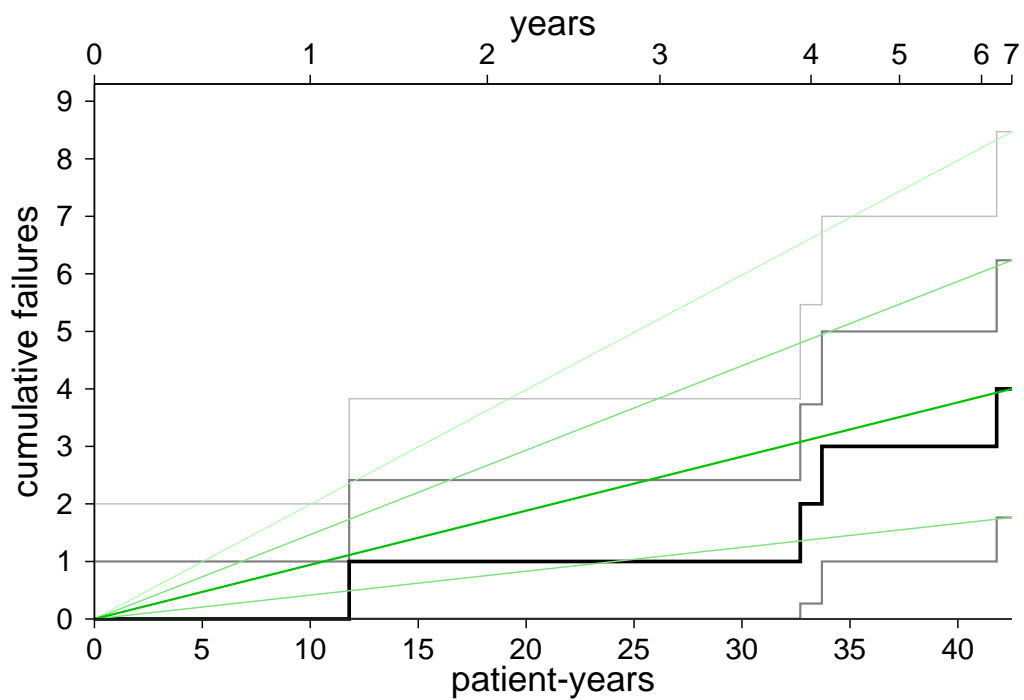


Figure 11: The simplest (constant) hazard rate model, green, overlaid on the cumulative failure graph of Fig. 10, again with uncertainty curves corresponding to one and two standard errors (the lowest of which lies along the horizontal axis).

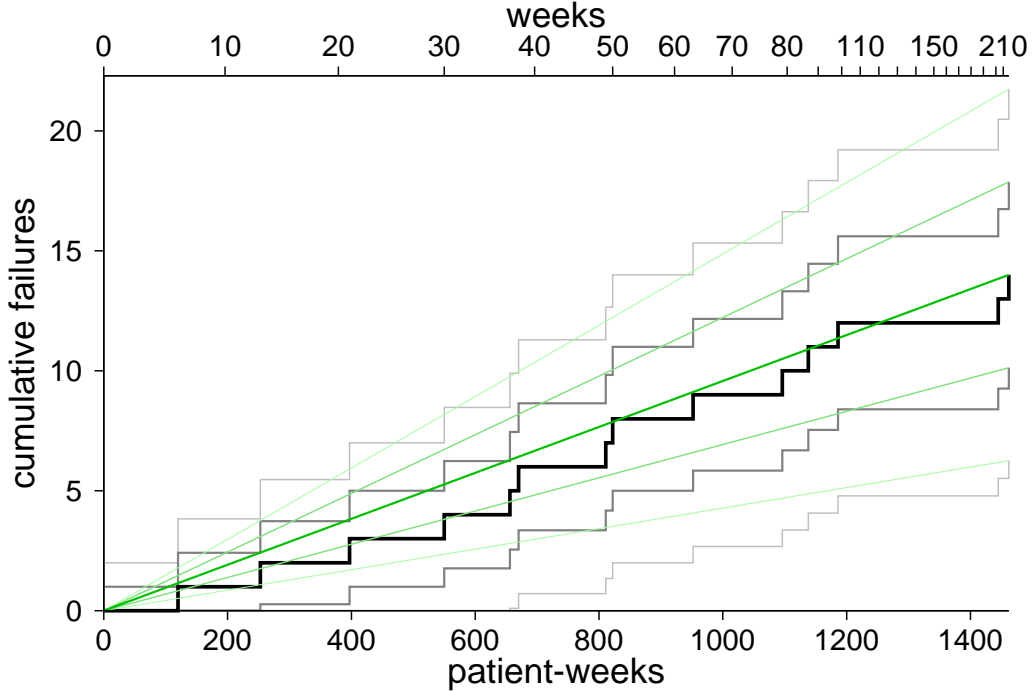


Figure 12: Constant hazard rate model, green, overlaid on the cumulative failure graph of Fig. 8 for the astrocytoma failure data of Bland and Altman.

Secondly, if this test fails, then we proceed as follows. We start at  $\tau = 0$  and  $D = 0$ . We then walk through each of the failures in turn, testing whether a constant hazard rate model from the origin to that  $D(\tau)$  (the “top of the step”), and to the  $D(\tau) - 1$  that applies immediately before the failure (the “bottom of the step”), both fulfil the consistency test described above. If we hit a failure that fails this test, then we return to the previous failure, and find the best-fit constant hazard rate model to that failure, estimating the standard error of the hazard rate from that of the count of failures. We then start from that point anew, walking through successive failures, fitting a (new) constant hazard rate model from the new starting point. Because the new dependent variable is again just a count of failures, it is independent of the previous piece of the model, and so we can apply the same consistency test, where now the count used to estimate the standard error is the cumulative count of failures from the new starting point.

We continue until we have exhausted all of the failure data. The result is a model of the hazard rate that is piecewise constant, *i.e.*, the model of the cumulative hazard is piecewise linear. Because each piece is statistically independent, the standard errors from different pieces add in quadrature in the cumulative hazard, even though the standard error within each piece is linear in  $\tau$ ; *i.e.*, if the pieces are labeled by the index  $k = 1, 2, \dots$ , applying from  $\tau_{k-1}$  to  $\tau_k$  (where  $\tau_0 = 0$ ); if there are  $D_k$  failures in piece  $k$ ; and if  $k(\tau)$  is the value of the index  $k$  corresponding to the piece of the model applicable for the given value of  $\tau$ , then

$$\text{Var}(D(\tau)) = \sum_{\tau_k < \tau} (D_k + 1) + \left( \frac{\tau - \tau_{k(\tau)-1}}{\tau_{k(\tau)} - \tau_{k(\tau)-1}} \right)^2 (D_{k(\tau)} + 1), \quad (7)$$

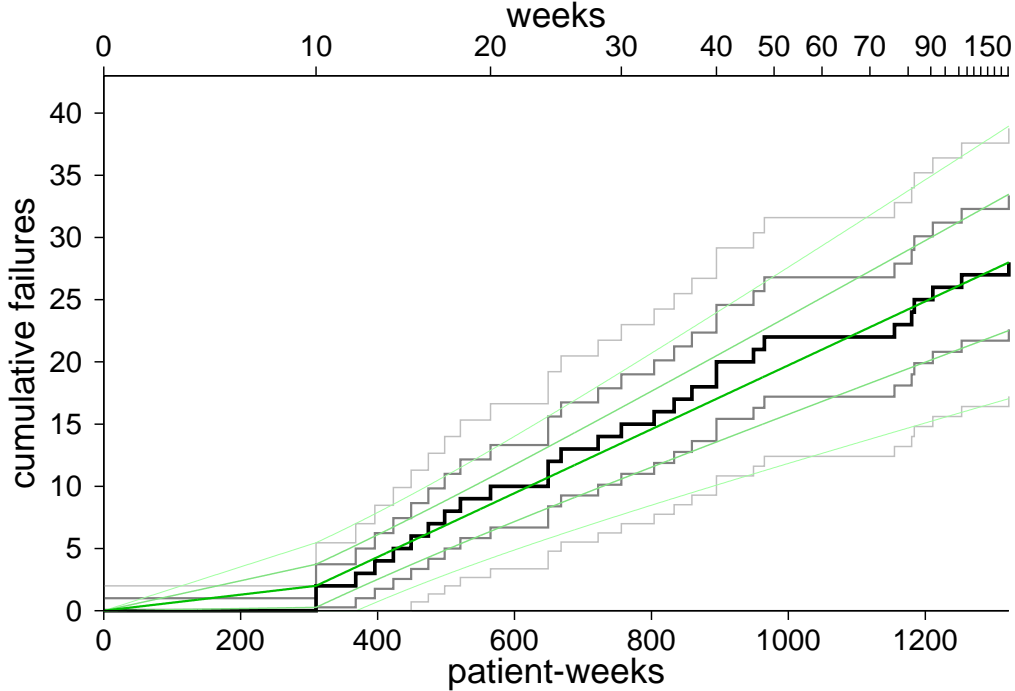


Figure 13: Piecewise constant hazard rate model, green, overlaid on the cumulative failure graph of Fig. 9 for the glioblastoma failure data of Bland and Altman.

where the last term comes from the piece  $k(\tau)$  (which would lead to a standard error linear in  $\tau$  if the sum were not present), and the sum arises from the cumulative contribution to the variance from all previous pieces. In the same notation, the best estimate of the cumulative hazard is

$$E(D(\tau)) = \sum_{\tau_k < \tau} D_k + \left( \frac{\tau - \tau_{k(\tau)-1}}{\tau_{k(\tau)} - \tau_{k(\tau)-1}} \right) D_{k(\tau)}, \quad (8)$$

*i.e.*, the piecewise linear fit.

Applying this method to the glioblastoma data of Bland and Altman shown in Fig. 8., we obtain the model shown in Fig. 13. We can see that the algorithm has picked up what was evident visually, namely, that the lack of failures in the first 300 patient-weeks implies a lower hazard rate than the remaining 1000-odd patient-weeks. Graphically, a constant hazard rate model from the origin to the third failure (at about 360 patient-weeks) would have crossed the corner of the two-standard-error curve at about 300 patient-weeks, and so the algorithm has broken the model into two pieces. Note that the lighter green error lines are slightly curved, not straight, reflecting the sum in quadrature contained in Eq. (7).

Of course, the algorithm only found such a “significant” change in hazard rate between these two pieces because we defined the consistency test in terms of crossing the two-standard-error uncertainty curve, and so we would expect that this algorithm would come to the wrong decision around 5% of the time. But the ramifications of such decisions are rather benign: they simply put extra “kinks” into the model; it still ends up at essentially the same end-point. (The estimate of the standard error is increased slightly for every extra

piece added to the model.) Compared to the Kaplan–Meier method’s stepwise following of every noisy count in the failure data, erroneously adding an extra “kink” every now and again is a relatively small price to pay.

#### 2.4. Returning to real time

Having obtained what is arguably the simplest hazard rate model consistent with the data, we are now in a position to eliminate the patient-time  $\tau$  in favor of the real time  $t$ .

The estimated hazard rate as a function of time,  $\lambda(t)$ , is obtained as a mere change of variable:

$$\lambda(t) \equiv \lambda(t(\tau)), \quad (9)$$

which is thus piecewise constant.

The cumulative hazard as a function of time, defined as

$$\Lambda(t) \equiv \int_0^t dt' \lambda(t'),$$

is therefore piecewise linear. In terms of  $D(\tau)$ , we have, similar to Eq. (8),

$$\mathbb{E}(\Lambda(t)) = \sum_{t_k < t} \left( \frac{t_k - t_{k-1}}{\tau_k - \tau_{k-1}} \right) D_k + \left( \frac{t - t_{k(t)-1}}{\tau_{k(t)} - \tau_{k(t)-1}} \right) D_{k(t)}, \quad (10)$$

where  $t_k \equiv t(\tau_k)$ . Likewise, the uncertainty in this estimate is given by an expression similar to Eq. (7),

$$\text{Var}(\Lambda(t)) = \sum_{t_k < t} \left( \frac{t_k - t_{k-1}}{\tau_k - \tau_{k-1}} \right)^2 (D_k + 1) + \left( \frac{t - t_{k(\tau)-1}}{\tau_{k(\tau)} - \tau_{k(\tau)-1}} \right)^2 (D_{k(\tau)} + 1). \quad (11)$$

Using (10) and (11), the estimated cumulative hazard curves for the Bland and Altman example are shown in Fig. 14. Again, it is clear that the cumulative hazard for glioblastoma becomes statistically significantly higher than that for astrocytoma after about 40 weeks.

Finally, the survival function,  $S(t)$ , is defined as

$$S(t) \equiv e^{-\Lambda(t)}. \quad (12)$$

The estimated survival curves for the Bland and Altman example are shown in Fig. 15. Comparing with the Kaplan–Meier curves of Fig. 4, we can see that the estimates of the survival function and its uncertainty are in broad terms the same; the difference is that Fig. 15 does not have the “staircase” corners arising from the the Kaplan–Meier curve following the statistically insignificant “counting noise” of the failure process.

#### 2.5. Comparing survival curves

Figs. 14 and 15 suggest that the survival rate for astrocytoma is significantly greater than that for glioblastoma after about 40 weeks. We can simplify the visual representation of this

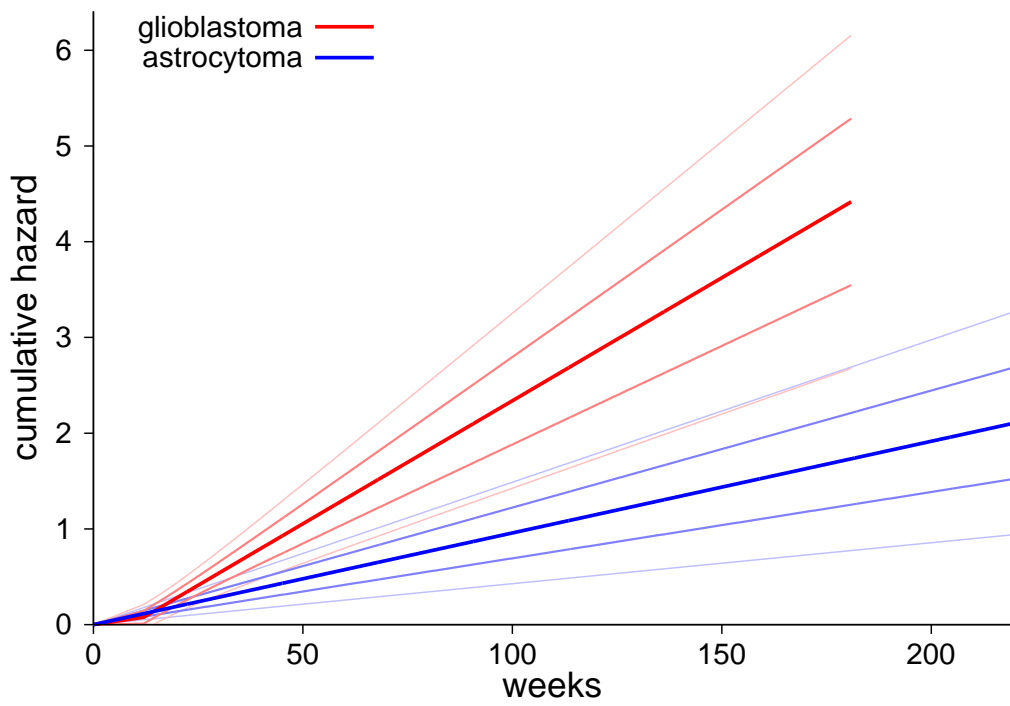


Figure 14: Cumulative hazard estimates for the Bland and Altman example.

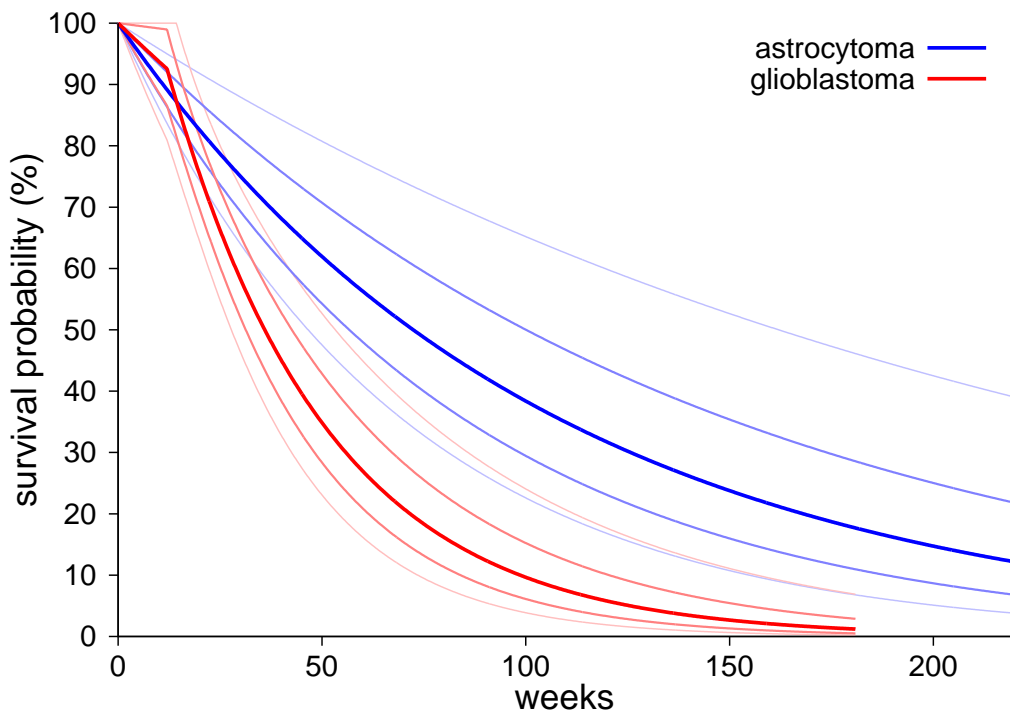


Figure 15: Survival function estimates for the Bland and Altman example.

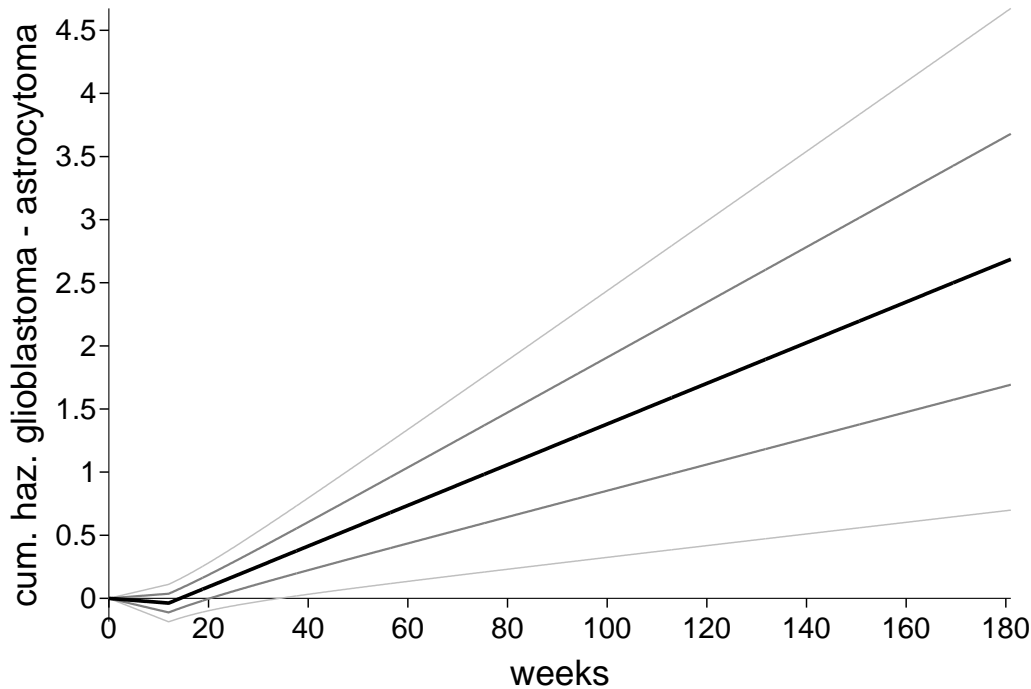


Figure 16: Difference in cumulative hazards for the Bland and Altman example.

statistical test by considering the *difference in cumulative hazard* between the two groups, as a function of time. Since the two sets of failure data are statistically independent, we add their standard errors in quadrature.

Fig. 16 shows the result for the Bland and Altman example. The difference in cumulative hazard becomes significant at the level of two standard errors at around 40 weeks, in agreement with our eyeball estimate of Fig. 14. By 181 weeks (the greatest time available in both data sets), the difference is significant to around 2.7 standard errors, corresponding to  $p < 0.007$ , in broad agreement with the crude logrank estimate of  $p < 0.01$  for the difference in “average” hazard rate obtained by Bland and Altman.

## References

- [1] Böhmer PE. Theorie der unabhängigen Wahrscheinlichkeiten. *Rapports, Mémoires et Procès-verbaux de Septième Congrès International d’Actuaries, Amsterdam 1912*; 2: 327–343.
- [2] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Statist Assn* 1958; 53: 457–481.
- [3] Kaplan EL. This week’s citation classic. *Current Contents* 1983; 24: 14.
- [4] Bland JM, Altman DG. The logrank test. *BMJ* 2004; 328: 1073.
- [5] Miller RG. What Price Kaplan-Meier? *Biometrics* 1983; 39: 1077–81.
- [6] Altman DG, De Stavola BL, Love SB, Stepniowska KA. Review of survival analyses published in cancer journals. *BJC* 1995; 72: 511–8.
- [7] Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *Lancet* 2002; 359: 1686–9.



- [8] Miettinen OS. Survival analysis: up from Kaplan–Meier–Greenwood. *Eur J Epidemiol* 2008; 23: 585–592.
- [9] Carter RE, Huang P. Cautionary note regarding the use of CIs obtained from Kaplan–Meier survival curves. *JCO* 2009; 27: 174–5.
- [10] Greenwood Maj. The natural duration of cancer. *Reports on public health and medical subjects*, No. 33. His Majesty’s Stationery Office, 1926.