Hazard ratios in cancer clinical trials—a primer

Krastan B. Blagoev,

Physics Division, National Science Foundation, 4201 Wilson Boulevard, Arlington, VA 22230, USA

Julia Wilkerson,

Medical Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA

Tito Fojo

Medical Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA

Abstract

The increase and diversity of clinical trial data has resulted in a greater reliance on statistical analyses to discern value. Assessing differences between two similar survival curves can pose a challenge for those without formal training in statistical interpretation; therefore, there has been an increased reliance on hazard ratios often to the exclusion of more-traditional survival measures. However, because a hazard ratio lacks dimensions it can only inform the reader about the reliability and uniformity of the data. It does not provide practitioners with quantitative values they can use, nor does it provide information they can discuss with patients. Motivated by a non-scientific poll of oncologists in training and those with board certification that suggested only a limited understanding of the derivation of hazard ratios we undertook this presentation of hazard ratios: a measure of treatment efficacy that is increasingly used and often misused.

Introduction

Oncologists rely on the analysis of data from clinical trials to make rational choices. However, discerning the magnitude of a marginal survival benefit can be difficult. In particular, extracting valuable and easy-tointerpret information from survival curves can pose a challenge. Increasingly, investigators report, and often seem to rely on, hazard ratios as the preferred method to assess efficacy. Hazard ratios are often cited in preference to more-traditional efficacy measures such as length of progression-free survival (PFS) and overall survival. Indeed, reports increasingly cite hazard ratios in abstracts (Figure 1), often omitting actual time benefits.[1] Evidence of the prevalence of reporting hazard ratios without the time benefits was obtained from an assessment of articles published in 2011 in seven

journals that report cancer clinical trials (Figure 1b). The figure shows our assessment of studies reporting clinical trials in which a hazard ratio was reported in the abstract; in only 57% of these abstracts were both the hazard ratio and the relevant details of the end points (for example, median overall survival) presented, whereas in 43% of abstracts only a hazard ratio without reference to the absolute values was reported. Because a hazard ratio is a value that has no dimensions, its presentation without the absolute benefit in time provides information of limited value, informing the reader only about the reliability and uniformity of the data. Our strong feeling that reporting a hazard ratio without concurrent presentation of the magni tude of benefit is not appropriate motivated the preparation of this article with the aim to present the use of hazard ratios in a straightforward manner.

Unlike most articles that focus on statistical inference models and are targeted at statisticians, we will discuss the use of hazard ratios in analyzing clinical trial data, how the ratio is obtained from typical survi val plots and caveats to look for in deciding the benefit of a particular therapy. We begin with descriptions of Kaplan–Meier plots and then describe how hazard ratios are calculated. We believe this will allow the reader to understand the value and limitations of hazard ratios.

Kaplan–Meier plots

In many phase III clinical trial reports, data are represented as a Kaplan–Meier plot. To illustrate the advantages of this type of representation, we begin by constructing a simplified Kaplan–Meier plot depicting PFS for two groups receiving different treatments. The first group receives the experimental regimen and the second is treated with a standard therapy. Data for the experimental arm are presented in Table 1; the resulting Kaplan–Meier plot is presented in Figure 2.

In the example in Table 1, the number of patients surviving free of cancer progression is listed at equal assessment intervals of 1 month. At the beginning of the trial there are 100 patients; at the end of each 'assessment interval' it becomes important to know the number of patients who have died or have evidence of disease progression and the number of patients that have been 'censored' for reasons other than cancer progression or death. For each interval, the probability of 'surviving without evidence of progression' (PFS) is the ratio: number of progression-free survivors at the end of the interval/number of progression-free survivors at the beginning of the interval.

For this probability to be accurate, both the numerator and denominator should have those censored during the interval deducted before calculation. An advantage of the Kaplan–Meier approach is that it allows one to include censored patients in probability estimates of survival right up to the evaluation point preceding their censoring.[2] Patients censored in a given interval are then omitted from all subsequent calculations. In clinical trials, variable fractions of patients are censored at each interval depending on the requirements of the trial.

The probability of surviving progression free to the end of an assessment interval is the product of the probabilities of surviving progression free in all preceding assessment intervals multiplied by the probability for the interval of interest. The main assumption is

that the probability of surviving progression free in a given interval is independent of the probabilities of surviving progression free in any other interval. In our example, the percentage of patients who are progression free for any interval is obtained by dividing column 5 (progression free at start of month minus censored during month [n]) by column 6 (progression free at end of the month minus censored during month [n]) and multiplying the resulting fraction by 100. In the first 3 months, since no patients were censored the probabilities of PFS for each individual month were: $90/100 = 0.90 \rightarrow 0.90 \times 100 = 90\%$ (first month); $81/90 = 0.90 \rightarrow 0.90 \times 100 = 90\%$ (second month); $73/81 = 0.90 \rightarrow 0.90 \times 100 = 90\%$ (third month). Overall, for the first 3-month period this would represent a probability of PFS of: $0.90 \times 0.90 \times 0.90 = 0.729$ or $\approx 73\%$.

Alternatively, because no patients were censored, for the first 3 months the probability of PFS could have been calculated by recognizing that 27 patients $(10 + 9 + 8)$ either experienced disease progression or died while $(100 - 27)$ 73 did not, so the probability of PFS could have been more-directly calculated as: $73/100 \times 100 = 73\%$. However, the calculations differ if patients are censored for reasons other than progressive disease or death. To exemplify this we continue with month 4; during this month eight patients were censored and these are then excluded from the calculations. The probability of surviving month 4 then is: $(73 - 8 - 7)/(73 - 8) = 58/65 \times 100 = 89.23\%$.

The denominator $(73 - 8)$ recognizes that eight patients were censored, making 65 the effective number of patients at the beginning of the period, eight less than the original 73. Furthermore, as in previous months, a given number (seven) either died or experienced progression of their disease; the $(73 - 8 - 7)$ numerator subtracts the seven patients that experienced disease progression or died from the effective starting number of 65. While the PFS probability during month 4 is 89.23%, for any patient who began the study, the probability of surviving progression free for 4 months is obtained by multiplying the probability of surviving progression free for the first three months with the probability of surviving for month 4: $0.90 \times 0.90 \times 0.90 \times 0.8923 = 0.6504 \times 100 = 65.04\%$.

Note that in calculations such as this, patients for whom information is not available are censored. In a clinical trial this might be a patient who discontinues treatment in the middle of a cycle because of toxicity without first having an efficacy assessment. However, if the patient completes a period of treatment and is evaluated before the decision to discontinue treatment is made, that patient contributes information for that period and is not censored in assessing efficacy in the just-completed period, but rather is censored going forward beyond the just completed period.

Continuing the above calculations at the end of each defined interval generates the Kaplan–Meier plots used to depict results from clinical trials (Table 2 and Figure 2). With different rates of death, progression, and number of patients censored, Kaplan–Meier plots vary greatly. For example, compare the two panels of Figure 3 that depict Kaplan–Meier plots of PFS in two clinical trials in patients with renal cell carcinoma.[3,4] One might ask to what extent the two curves in each study differ? One measure of value of a therapy option is the median survival time—a value calculated in most studies from a Kaplan– Meier plot. In the examples depicted, the median PFS is the time (x-axis) at which the percentage of

progression-free patients (y-axis) is 50%. If this 50% of progression-free patients extends over a range on the x-axis, then the median PFS is calculated to be the average between the first and last value corresponding to 50%.[5] The curves in Figure 3a have steeper slopes, follow each other closely and demonstrate a modest benefit from sorafenib relative to placebo.[3] By comparison, Figure 3b shows that sunitinib markedly improves the median PFS.[4] These differences can also be quantified using hazard ratios, as discussed later. However, at this point we would like to note that when curves are similar, as in the sorafenib study (Figure 3a), the median survival is easy to measure and is reliable. By contrast, curves such as the ones in the sunitinib study (Figure 3b) can be 'smoothed out' by calculating a hazard ratio, although a median PFS unequivocally reflects the clear benefit.

In addition to PFS and overall survival, Kaplan–Meier plots are often used to assess the viability of cancer biomarkers. However, as has been noted elsewhere,[6] there are a number of limitations associated with using standard analysis of Kaplan–Meier plots to assess markers. For example, a Kaplan–Meier analysis does not answer whether a marker is of value or if a new marker is better than an existing one. If established markers are not considered, then one cannot be certain equivalent separation could be achieved using an established marker or combination of established markers. Even comparisons to established markers demon strating greater or more significant separation of prognostic groups with the new marker are of limited value since they may presume there is only one established marker. Furthermore, a Kaplan–Meier analysis usually assumes that a marker is binary (that is giving a 'yes' or 'no' answer, with no gray area) risking the loss of prognostic value by collapsing the marker information into one of the two groups. In addition, the issues discussed below in using the hazard ratio to establish efficacy also apply to the assessment of markers of efficacy.

Before we move on from Kaplan–Meier plots, we note two important points and offer a note of caution. The first point relates to the issue of patients who are censored, an outcome increasingly reported in oncology trials.[7] Censoring for survival is straightforward because studies are often terminated when a pre-defined number of 'events' (deaths) have occurred, and consequently the date of death will not be known for all participants. In this case, patients alive when data collection is terminated are 'censored' at the study termination date and often appear as 'tick marks' or other symbols in Kaplan–Meier plots. Such tick marks appear throughout the plot reflecting the fact that patients enrolled at different times and hence have been 'alive and on-study' for a range of times. By including them in the analysis even though they have not yet experienced an 'event', in this case death, they contribute a maximum amount of data since they are known to have survived at least this amount of time. Less straightforward is the practice of censoring patients in PFS analyses, since unless a study is reported with only a short time of follow up, the majority of patients have usually been enrolled long enough for progression to have occurred and been 'scored' and assessed within the plot. Censoring in studies with PFS as an end point occurs for a variety of reasons, including treatment discontinuation due to toxic effects before documented progression of disease, in which case the patient cannot be scored as having progressive disease and must be censored. However, because of differences in toxicity profiles, censoring might not occur with equal frequency in both arms of a randomized trial. Censoring also often occurs at a higher rate in the central review analysis than when the data

are analyzed locally.[8,9] The increased rate of censoring by central review can occur, for example, when the independent reviewer's measurements do not agree with the investigator's assessment of progression. We would caution that while censoring allows one to include all patients that participated in the trial until the time of an event or when they are censored, censoring can introduce problems in the interpretation of a trial. Censoring will not alter the survival probabilities in the Kaplan–Meier plot only if that censoring is 'uninformative'. In the case of PFS, for censoring to be 'uninformative' the patients in the censored population must be representative of those in the population that continue the trial; and censoring must not occur to a greater extent in one arm of a randomized trial. If these conditions are not satisfied, censoring can alter the true probabilities for PFS and bias the results. For example, patients censored early in a trial have a higher probability of having continued progression free than patients censored towards the end of the trial, so that discrepancies in censoring could lead one to incorrectly estimate the efficacy of a treatment if the censored population is not representative of the population continuing on the trial.

The second point we would note regarding Kaplan–Meier plots is a variation often used when the number of patients in a clinical trial is small. In the case of a small number of patients, instead of performing calculations at defined intervals as in our example, calculations are made each time an 'event' occurs (disease progression or death) subtracting patient(s) censored during this same time. One then sees 'drops' on the plots at variable intervals, and these are of greater magnitude towards the right of the plot (longer observation times). This increase in 'drop' when an event occurs at a later time point is because over time any 'event' represents a greater fraction of the remaining patients and has a greater impact.

Hazard rate and hazard ratio

To fully understand a hazard ratio we must introduce the concept of hazard rate. The hazard rate quantifies the likelihood a patient will experience a 'hazardous event' or 'hazard'—such as, disease progression or death—during a defined interval of observation as a rate or percentage. Unlike a Kaplan–Meier plot, which focuses on the number of patients continuing to do well at the end of the interval of interest, the hazard rate and hazard ratio focus on the opposite, those patients who have not done well during the interval of interest —estimating the chances that a patient would suffer a hazardous event in a given interval. The hazard rate can easily be obtained from data used to generate Kaplan–Meier plots (Figures 2 and 3). For clarity we use a fictitious clinical trial, (Table 3 and Figure 4) where 5 weeks after the trial began there are 50 patients alive in the control group who have not experienced disease progression (that is, 50 progression-free survivors); and 75 such patients in the experimental group. A week later—at the end of week 6—there are 40 progression-free survivors in the control group and 70 in the experimental arm (Table 3 and Figure 4a). From these data, one can calculate the hazard rate at 6 weeks for patients in the control and treatment groups—that is, the rate or likelihood of the occurrence of a 'hazardous event'. To calculate this value, we need to know the number of patients who experienced a hazardous event (disease progression or death) during week 6. For the control group this number is 10 patients (50 – 40) and for the treatment group it is five patients (75 – 70). To obtain the hazard rate for each group this week we divide the number who experienced a 'hazardous

event' during week 6 by the progression-free survivors at the start of the interval of interest (end of week 5). Therefore, in our hypothetical example, the hazard rates are 0.2 (10/50; 20%) and 0.0666 (5/75; 6.66%) for the control and experimental groups, respectively. We can repeat these calculations for every week (or other time interval of interest) and obtain hazard rates throughout the trial. In our example, the hazard rates increase, because although in each week the same number of patients experience disease progression or die (10 in the control arm and five in experimental arm), this number as a fraction (or rate) of the total number of patients alive without disease progression at the beginning of each week increases. It is important to note that the hazard rate has a unit, in the case of our study we are calculating the hazard rate each week and so the unit is percent per week. Splitting the trial into small intervals in this manner generates a series of hazard rates for each arm of the trial. In the case of PFS, which is being used in our example, the hazard rates are defined as $h_t = (N_{t-\Delta t}^{PFS} - N_t^{PFS})/N_{t-\Delta t}^{PFS}$, where h_t denotes the hazard rate at time t and the interval Δt that follows is sufficiently short that the hazard rate remains approximately constant during the interval. It is important to note that the hazard rate depends on the number of patients (N) that survive without disease progression to time t, but not to time t + Δt (as previously discussed, the latter number is important in a Kaplan–Meier plot). If the number of patients is small, the hazard rate is dominated by statistical variations and the assumption that the hazard rate remains approximately constant during the time interval is unlikely to be valid. This situation often occurs as a clinical trial progresses and the number of patients who have not experienced the hazardous event becomes smaller. Therefore, data gathered towards the end of a trial is often ignored, as discussed below.[1]

Having determined the hazard rates for each group at each time interval we can now determine the hazard ratio for each time interval. As the name implies, the hazard ratio is the ratio of the hazard rates. For our example (Table 3), the hazard rates for week 6 were 0.2 and 0.0666. The hazard ratio—likelihood that a 'hazardous event' will occur in the experimental group relative to the control group—can be readily determined by the ratio of the hazard rates: 0.0666/0.2 = 0.333. In other words, during week 6 a patient on the experimental group was 0.333 (33.3%) as likely to experience disease progression or die as a patient in the control group—conversely 66.7% less likely to encounter a hazardous event.

How does one then calculate the hazard ratio for the entire study and not just for one interval? If the hazard ratio is similar (can be considered to be constant) for all time intervals one can define the hazard ratio for the entire study as $HR = h^{arm\ 1}/h^{arm\ 2}$, where HR denotes the hazard ratio and $h^{arm\ 1}$ and $h^{arm\ 2}$ can be the hazard rates for any time interval. However, in general, uniformity of hazard rates over time is uncommon. In our example, week 6 could have been a good week for the experimental group, or a poor week. Similarly, it frequently occurs in clinical trials that the slope of the graph varies over the study period (Figure 3b). Since the slope of the graph reflects the hazard rate— steeper slopes represent a greater hazard rate—the hazard rate is changing, and this in turn affects the hazard ratio. Specifically, between months 3 and 6 the slope of the sunitinib arm and, in turn, its hazard rate is much less than that of the interferon-α (control) arm (Figure 3b). Consequently, during this time period the hazard ratio favors the sunitinib arm. By contrast, between months 9 and 12 the slope of the sunitinib arm is very steep, which translates into a larger

hazard rate for the sunitinib arm that when divided by the smaller hazard rate of the interferon-α arm, with its flatter slope, results in slopes that translate into a hazard ratio that is not favorable for sunitinib (see also Table 3 and Figure 4). Furthermore, even if the number of patients surviving progression free decreases exponentially such that the hazard rate in each time interval might remain constant, censoring of patients can lead to changes in the hazard rate.

Even though the hazard ratio might change over time, most studies report a single value for the hazard ratio. How is this value obtained? One approach is to determine the hazard ratio for a large number of intervals. With this approach the shorter the intervals, the more accurate the results. Suppose that for the control group the hazard rate is $h^C(t_i)$ and for the experimental group it is $h^E(t_i)$. Dividing the hazard rates for every interval we obtain the hazard ratios, for each interval: $HR(t_i) = h^C(t_i)/h^E(t_i)$, and using the hazard ratios obtained for each interval one can compute the hazard ratio for the entire study by determining an average hazard ratio. Such computations to calculate an average hazard ratio could not be performed if the ratio varied over time, as might occur with the acquisition of an immune response mediated by an agent such as interferon-α or a vaccine. However, if the hazard ratio is approximately constant it is possible to calculate the average and then go on to use further statistical assessment to establish the statistical significance of the hazard ratio. The statistical significance is usually calculated using the log-rank test,[10,11] to disprove the null hypothesis that predicts both treatments have the same survival probabilities when the hazard ratio is significant. A thorough description of the log-rank test is outside the scope for this article; for details we refer the reader to the original manuscripts.[10,11]

Alternatively, one can obtain a hazard ratio using regression models.[12] In these models, one assumes there is a (uniform) mathematical relation between the hazard rate and several independent parameters that characterize the disease being studied (for example, the patient's response to a treatment, biomarkers and side effects of therapy). Using statistical computer software one can then fit the trial data to a model that uses this mathematical relation and obtain one value for the hazard rate, in effect 'smoothing out' the data so that the hazard rate is a constant. The model can be guided (constrained) by details of the trial.[12] From the models one can then establish hazard rates and, in turn, a hazard ratio for the trial and its statistical significance. In many cases, the Cox proportional hazard model[8] is used as the regression model. In this case, a linear model is assumed for the logarithm of the hazard rate—that is, the probability that an event occurs in a small time interval conditioned that no event occurred before the beginning of the time interval.[12] This approach assumes that the hazard ratio does not depend on time and that between the hazard rate and the indepen dent parameters there is a loglinear relationship.[12] For this reason, sometimes the Cox proportional hazard model is said to be semi-parametric. [13] Extensions of the Cox proportional hazard model include assessments of the hazard ratio that depend on time (time dependent).[14] Most statistical packages perform this analysis auto matically.[15–17] In some studies, the two curves in the Kaplan–Meier plot cross, invalidating the Cox proportional hazard model. In this case, other statistical methods might be useful. The advantage of regression-based models is that a number of covariates (confounders) can be captured in a

single model. If the number of confounders is more than three, Kaplan–Meier survival plots are impractical, because one needs to stratify the data in many subsets.[18]

Furthermore, although the ideal hazard ratio would capture the differential benefit throughout the study, in practice the entire time depicted in a Kaplan–Meier plot might not be analyzed in the hazard ratio.[1] As the sunitinib data (Figure 3b) shows, at the extremes and especially towards the end of a clinical trial, as the number of patients who have not yet died or experienced disease progression declines, any event generates a disproportionate effect on the hazard rate and in turn the hazard ratio. Consequently, if the change is not 'statistically significant' the interval should not be used in calculating the overall hazard ratio. While intuitively reasonable, this has the effect of ignoring portions of the Kaplan–Meier curve that are less beneficial for the superior arm of the study, enhancing its hazard ratio.

Finally, returning to the assessment of biomarkers, and addressing the use of hazard ratios in their evaluation, we note that such analyses are encumbered by limitations and do not directly test the value of a marker. For example, the *P* value (which is often used to validate the hazard ratio in these cases) essentially tests whether the hazard ratio is 1, not whether a marker has any predictive value.[6] As discussed regarding the use of Kaplan–Meier plots to assess bio markers, how the biomarker is coded (binary or continuous) affects the results. Similar to survival analyses, the assumption is made that a Cox regression is the best prediction model for the quality of biomarkers, but this might not be the case—precluding conclusions as to whether a new marker can more accurately predict outcomes than standard methods.

Conclusions

In this article, we have discussed the use of hazard ratios in cancer clinical trials. At present, hazard ratios calculated from clinical trial data are offered as evidence of treatment success when comparing new drugs or combinations with established therapies. While a hazard ratio has some value, for the clinician caring for a patient and, more importantly, the patient, it does not convey benefit in terms that are meaningful—how much longer will the patient live or live without experiencing disease progression. This situation occurs because to obtain the hazard ratio one divides the hazard rates and the time variables cancel out and so it is a unitless value. So the units of time on a Kaplan–Meier plot could be hours, days, weeks, months, or years, and the hazard ratio would be the same. Because of this, a hazard ratio has very limited value unless accompanied by a number that quantifies the magnitude of benefit in time. The inclusion of these data should be a requirement of all reports of clinical trials and these values should be included in abstracts. To properly assess the value of a new treatment, clinicians should determine the absolute difference in survival and use this information to make clinical recommendations. Therefore, the clinician would go on to determine what the absolute difference in survival at various time points would be before making a clinical judgment. This is routinely done in clinical practice and so more-detailed information or understanding about hazard ratio methodology would not materially change clinical trial interpretation.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Acknowledgments

References

1. Hernán MA The hazards of hazard ratios. Epidemiology 21, 13–15 (2010). [PubMed: 20010207]

2. Kaplan EL & Meier P Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc 53, 457–481 (1958).

3. Escudier B et al. Sorafenib in advanced clear-cell renal-cell carcinoma. N. Engl. J. Med 356, 125–134 (2007). [PubMed: 17215530]

4. Motzer RJ et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. N. Engl. J. Med 356, 115–124 (2007). [PubMed: 17215529]

5. Armitage P, Berry G & Matthews JNS Statistical Methods in Medical Research, 4th edn, 568–590 (Blackwell Science Ltd, Oxford, 2002).

6. Kattan MW Evaluating a new marker's predictive contribution. Clin. Cancer Res 10, 822–824 (2004). [PubMed: 14871956]

7. Raymond E et al. Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. N. Engl. J. Med 364, 501–513 (2011). [PubMed: 21306237]

8. Yao JC et al. Everolimus for advanced pancreatic neuroendocrine tumors. N. Engl. J. Med 364, 514–523 (2011). [PubMed: 21306238]

9. Baselga J et al. Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. N. Engl. J. Med 10.1056/NEJMoa1109653.

10. Mantel N Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother. Rep 50, 163–170 (1966). [PubMed: 5910392]

11. Peto R & Peto J Asymptotically efficient rank invariant test procedures. J. Roy. Stat. Soc. Ser. A (General) 135, 185–207 (1972).

12. Cox DR Regression models and life tables. J. Roy. Stat. Soc. Ser. B (Methogological) 34, 187–220 (1972).

13. Fan J & Li R Variable selection for Cox's proportional hazards model and frailty model. Ann. Stat 30, 74–99 (2002).

14. Fisher LD & Lin DY Time-dependent covariates in the Cox proportional-hazards regression model. Annu. Rev. Public Health 20, 145–157 (1999). [PubMed: 10352854]

15. SAS/STAT®(SAS Institute Inc Cary, NC, USA).

16. XLSTAT (Addinsoft, New York, NY, USA).

17. MedCalc version 12.0 (MedCalc Software, Mariakerke, Belgium).

18. Miettinen OS Stratification by a multivariate confounder score. Am. J. Epidemiol 104, 609–620 (1976). [PubMed: 998608]
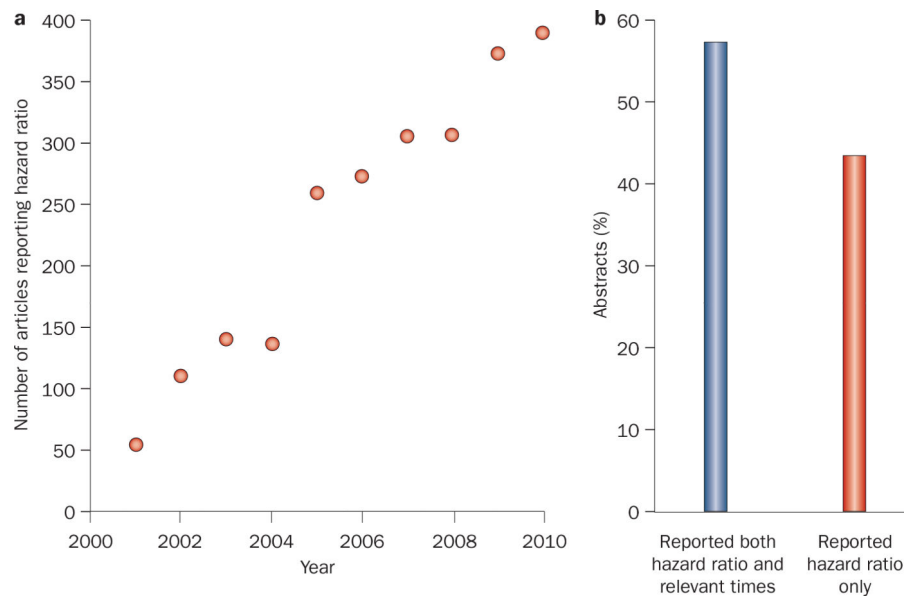
Figure 1 |.

The use of the term hazard ratio in abstracts. **a** | The increase in the use of the term hazard ratio in the abstracts of journal* articles reporting cancer clinical trial data over time. **b** | Abstracts reporting clinical trials* that included a hazard ratio and published in January, February and March of 2011 were examined. Both the hazard ratio and the relevant times were presented in 57% of these abstracts, whereas only a hazard ratio without reference to the absolute values in time could be found in 43% of abstracts. *Journals assessed: *Annals of Oncology*, *Clinical Cancer Research*, *European Journal of Cancer*, *Journal of Clinical Oncology*, *Lancet*, *Lancet Oncology*, *New England Journal of Medicine*.
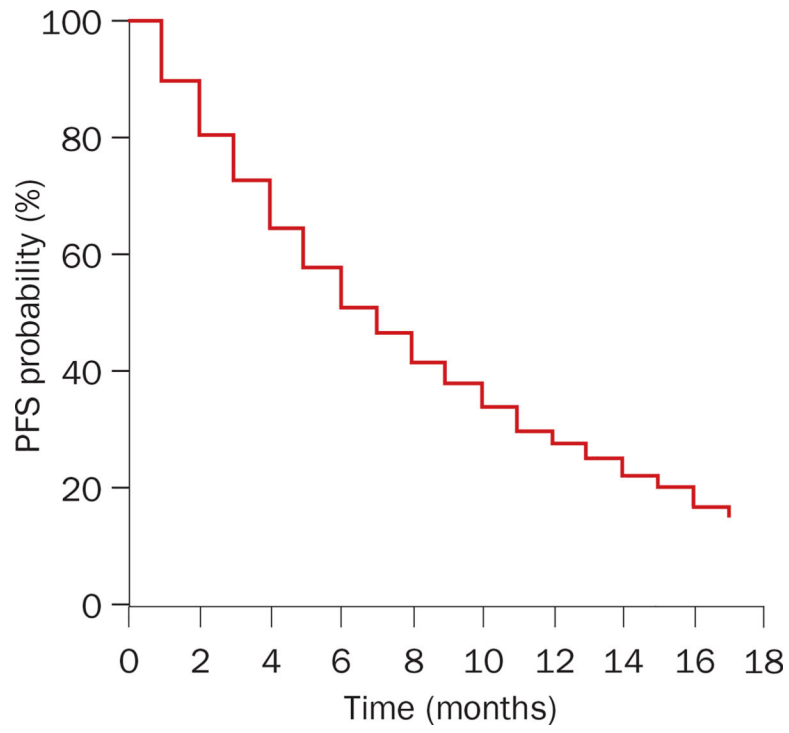
Figure 2 |.
Kaplan–Meier plot of hypothetical clinical trial. The data for this plot are tabulated in Tables 1 and 2.
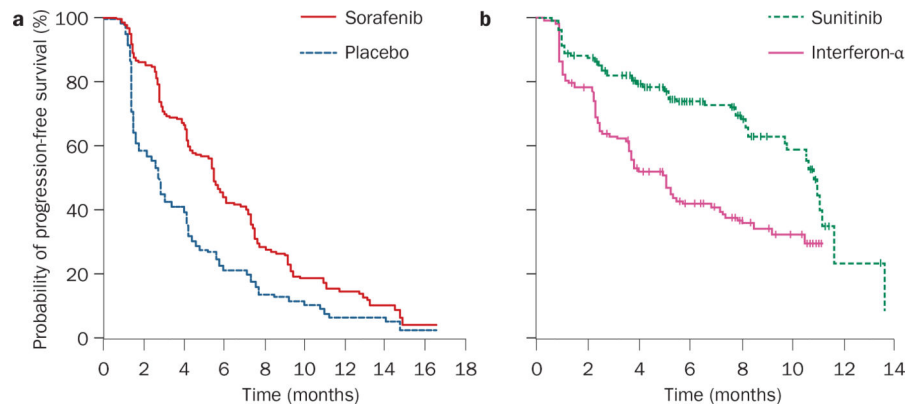
Figure 3 |.

Kaplan–Meier plot of two studies that led to registration of the respective drugs in renal cell carcinoma. **a** | sorafenib and **b** | sunitinib. Permission obtained from Massachusetts Medical Society part a © Escudier, B. *et al. N. Engl. J. Med.* **356**, 125–134 (2007) and part b © Motzer, R. J. *et al. N. Engl. J. Med.* **356**, 115–124 (2007).

Figure 4 |.

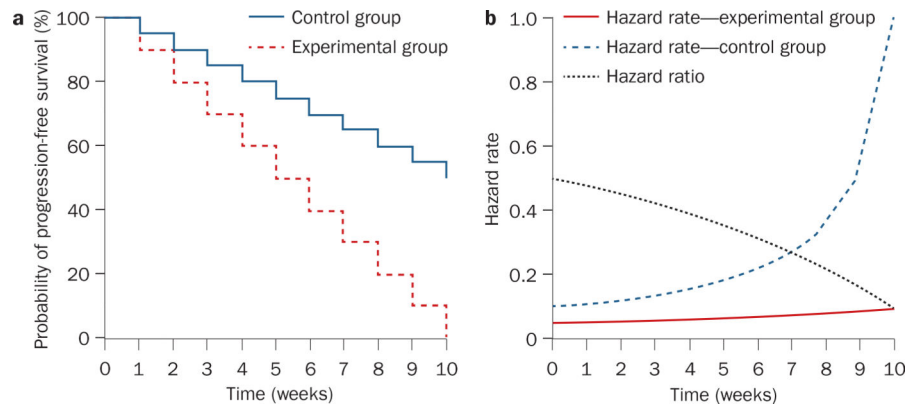Graphical presentation of hypothetical clinical trial data (Table 3) demonstrates how the hazard rate is calculated and how this rate can be used to calculate the hazard ratio.

a | Kaplan–Meier plot for the two arms of the hypothetical study (control and experimental).

b | The hazard rate can vary over time and this variation can disproportionately affect one arm of the trial at different time points.

Table 1 |

Simulated clinical trial data for experimental arm (n = 100)

Time (months)	Alive and progression-free at start of month (*n*)	Died or cancer progressed during month (*n*)	Censored during month (*n*)	Progression free at start of month minus censored during month (*n*)	Progression free at end of the month minus censored during month (*n*)
1	100	10	0	100	90
2	90	9	0	90	81
3	81	8	0	81	73
4	73	7	8	65	58
5	58	6	6	52	46
6	46	5	0	46	41
7	41	4	0	41	37
8	37	4	0	37	33
9	33	3	0	33	30
10	30	3	0	30	27
11	27	3	0	27	24
12	24	2	0	24	22
13	22	2	0	22	20
14	20	2	0	20	18
15	18	2	0	18	16
16	16	2	0	16	14
17	14	2	0	14	12

Table 2 |

Calculation of cumulative survival probability for simulated data

Time (months)	(Progression free at end of month - censored)/(progression free at start of month - censored)	Fraction progression free at end of month multiplied by survival probability at end of previous month (multiplied by 100 to obtain percentage)	Cumulative survival probability (%)
1	$90/100 = 0.9$	$0.9 \times 1.00 \times 100$	90
2	$81/90 = 0.9$	$0.9 \times 0.9 \times 100$	81
3	$73/81 = 0.9012$	$0.9012 \times 0.81 \times 100$	73
4	$58/65 = 0.8923$	$0.8923 \times 0.7299 \times 100$	≈ 65
5	$46/52 = 0.8846$	$0.8846 \times 0.6514 \times 100$	≈ 58
6	$41/46 = 0.8913$	$0.8913 \times 0.5762 \times 100$	≈ 51
7	$37/41 = 0.9024$	$0.9024 \times 0.5135 \times 100$	≈ 46
8	$33/37 = 0.8918$	$0.8918 \times 0.4633 \times 100$	≈ 41
9	$30/33 = 0.9090$	$0.9090 \times 0.4132 \times 100$	≈ 38
10	$27/30 = 0.9$	$0.9 \times 0.3755 \times 100$	≈ 33
11	$24/27 = 0.8888$	$0.8888 \times 0.3380 \times 100$	≈ 30
12	$22/24 = 0.9166$	$0.9166 \times 0.3004 \times 100$	≈ 28
13	$20/22 = 0.9090$	$0.9090 \times 0.2753 \times 100$	≈ 25
14	$18/20 = 0.9$	$0.9 \times 0.2502 \times 100$	≈ 23
15	$16/18 = 0.8888$	$0.8888 \times 0.2252 \times 100$	≈ 20
16	$14/16 = 0.875$	$0.875 \times 0.2001 \times 100$	≈ 18
17	$12/14 = 0.857$	$0.857 \times 0.1750 \times 100$	≈ 15

Table 3 |

Data used to construct Kaplan–Meier plot and hazard rate graph in Figure 4

Week	Patients alive and progression free (*n*)				Hazard rates (fraction per week)		Hazard ratio (hr^E/hr^C)
	Control arm		Experimental arm		Control arm (hr^C)	Experimental arm (hr^E)	
	Start of week	End of week	Start of week	End of week			
1	100	90	100	95	10/100 = 0.1000	5/100 = 0.0500	0.5
2	90	80	95	90	10/90 = 0.1111	5/95 = 0.0526	0.473
3	80	70	90	85	10/80 = 0.1250	5/90 = 0.0555	0.444
4	70	60	85	80	10/70 = 0.1429	5/85 = 0.0588	0.411
5	60	50	80	75	10/60 = 0.1666	5/80 = 0.0625	0.375
6	50	40	75	70	10/50 = 0.2000	5/75 = 0.0666	0.333
7	40	30	70	65	10/40 = 0.2500	5/70 = 0.0714	0.286
8	30	20	65	60	10/30 = 0.3333	5/65 = 0.0769	0.231
9	20	10	60	55	10/20 = 0.5000	5/60 = 0.0833	0.166
10	10	0	55	50	10/10 = 1	5/55 = 0.0909	0.091