Review Article

The Devil Lies in the Details: An Analysis of Six Oncology Randomised Controlled Trials

Alfonso Dueñas-Gonzalez[1], Leticia Bornstein-Quevedo[2], Francisco Gutierrez-Delgado[3]

1. Subdireccion de Investigación Básica, Instituto Nacional de Cancerología, Tlalpan, Mexico; 2. Immunohistochemistry and Molecular Biology Laboratory, InmunoQ, Mexico City, Mexico; 3. Cancer Prevention and Research Center (CEPREC), Mexico

Randomized Controlled Trials (RCTs) in oncology are the backbone of advancing the clinical practice of oncology. Their significance is particularly pronounced in high-impact and influential journals such as the NEJM, The Lancet, The Lancet Oncol, JAMA, JAMA Oncol, and JCO. However, previous works have highlighted that RCTs often fail to provide complete reporting of survival endpoints, which can jeopardize the oncologist's interpretation of the study findings.

This is a narrative review in which we provide some basic concepts of survival statistics and apply these in the analysis of six RCTs published in these six journals. These articles were selected solely by scanning the latest issues (printed) on the journals' web pages. The first RCT on a solid tumor to appear in the corresponding journals was selected. No other criteria were applied.

We found that these six trials failed to provide crucial data, such as the number of survival events, patient censoring, and the presentation of both absolute and relative effect sizes on survival, to different extents.

This report study focuses solely on a small subset of RCTs from high-impact journals, which may not represent the reporting practices of oncology RCTs published in other journals or at different times. The most well-informed party should be the patients, who ultimately benefit from improved RCT reporting. We hope this work will inspire oncologists to sharpen their skills in analyzing RCTs and encourage statistician experts to contribute their valuable insights and knowledge on this issue.

Corresponding author: Alfonso Duenas-Gonzalez, alfonso_duenasg@yahoo.com

1. Introduction

Currently, there is a hegemonic narrative about progress in cancer therapy commanded by the "big-six," the NEJM, The Lancet, The Lancet Oncol, JAMA, JAMA Oncol, and JCO, high-impact and prestigious medical journals, whose publications heavily influence the practice of Oncology. The publication in these journals of a randomized controlled trial (RCT) is like a blessing for the tested drug, and the results are greatly amplified by the mass media and pharma industry, which commonly uses "superlatives" to communicate the results of RCT[1][2]. Oncologists and cancer researchers also participate in the narrative by publicizing or sharing the achievements and suppressing unfavorable information on RCT results[3].

On the other hand, it is a fact that despite arguments justifying the use of surrogate markers, specifically progression-free survival (PFS) or even response rate, which is not discussed here, it seems clear that the bar for newer drug approval in oncology has been lowered[4][5][6]. For a proportion of drugs, the benefit of survival has never been confirmed[7][8]. This is not to say that individual cancer patients may not benefit. However, as a society, we must remember that to reduce cancer mortality, we need treatments that, if not a cure, at least increase the overall survival (OS) rates, not only the median survival time. This should serve as a motivation for oncologists and cancer researchers to strive for more effective treatments. Accordingly, a recent study has shown that novel pharmaceuticals increased patient survival by a median of 2.8 months for OS and 3.3 months for PFS[9].

A very well-orchestrated form of feeding the progress narrative is to present the results of the RCT in a statistical manner intended to highlight the good results and neglect the non-favorable results of the trial. This is coupled with the phenomenon termed "spin," which can be defined as the misrepresentation and distortion of research findings that affect clinical decision-making. Spin has been found not only in oncology but in randomized RCTs published in various fields of medicine[10][11][12].

The main aim of this narrative review is to provide some basic concepts of survival statistics and apply these concepts in the analysis of 6 RCTs published in the big-six (NEJM, The Lancet, The Lancet Oncol, JAMA, JAMA Oncol, and JCO). These articles were selected solely by scanning the latest issues from October to August 2024 (when we started to write this work) on the webpage of these journals. The first RCT on a solid tumor appearing was selected in these journal issues. No other criteria were

applied. The analysis of these RCTs was not intended to generalize our findings but to be used to be analyzed under the statistical concepts reviewed. This narrative review might equip clinical oncologists to better analyze RCT results in oncology. It is important to remark that we are peers with no formal training in statistics, and readers are encouraged to get more acquainted with medical statistics.

2. Main elements of the survival analysis

2.1. Survival curves

Because of the nature of oncological diseases, particularly those in advanced stages, that unfortunately cause the death of patients, it is easy to understand why the most valuable endpoint in most RCTs on cancer is survival. The most widely accepted and currently used method for evaluating survival is the Kaplan-Meier method (K-M). In clinical trials, the effect of an intervention is assessed by measuring the number of subjects who survived after that intervention over some time. The time starting from a defined point to the occurrence of a given event, for example, death, is called survival time, and the group data analysis is survival analysis. The K-M outputs a curve that graphically represents the survival rate or function and is the simplest way of computing survival over time, accounting for censored or incomplete observations. Time is plotted on the X-axis in a K-M curve, and the survival rate is plotted on the Y-axis.

The construction of a survival K-M curve needs the time when the patient entered the study and the status: 1 for the event (death) and 0 for alive (alive can be censored because the patient left the study for any reason (left censoring) or remains in the study without having the event (right or administrative censoring). A typical K-M survival curve displays the median survival (when half of patients have died) and the survival estimates at specific time points, i.e., 1-, 2-, 3-, 5-year. The same concepts apply to PFS curves. In this case, the event can be progression or death.

The RCT must report the median follow-up time with 95% CI (minimum-maximum). The information on the number of events (deaths) is critical for interpreting the curve and calculating survival percentages (these numbers are usually placed in the upper right of the curve). It is essential to reinforce that the number of events reported is actual and not actuarial or estimated when a study is reported (either preliminary or final results). Therefore, the survival percentage is marked at the intersection of a vertical line on the median follow-up time with the Y-axis (survival percentage). Any

deviation from this is suggestive of inconsistencies in the survival analysis. This fundamental principle goes beyond the number and time of censored patients because the curve is constructed with only two variables, time and status, and the status cannot be other than 1 for the event and 0 for no event (0 can be either censored because the patient was lost to follow-up for any reason or censored because she/he continues not having the event at the last follow-up), (Recommended articles[13][14][15][16][17]).

A lack of adherence to this principle may lead to higher or lower survival estimations, jeopardizing the study's interpretation. Examples are shown in **Figure 1**. As observed, the percentage of PFS or OS obtained from the total number of patients and events is reflected at the median follow-up time.

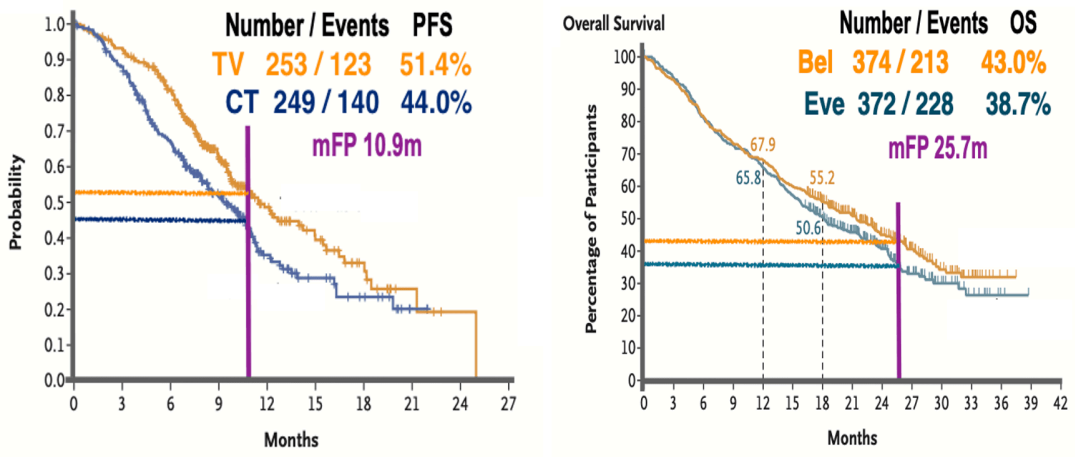

Figure 1. PFS and OS curves from two RCTs published in the NEJM. Left, figure 2 of[18]. Right, Figure 1 of[19].

Figure 2 shows curves from two RCTs of docetaxel versus pembrolizumab in which this principle is observed in the curve on the left but not in the curve on the right.

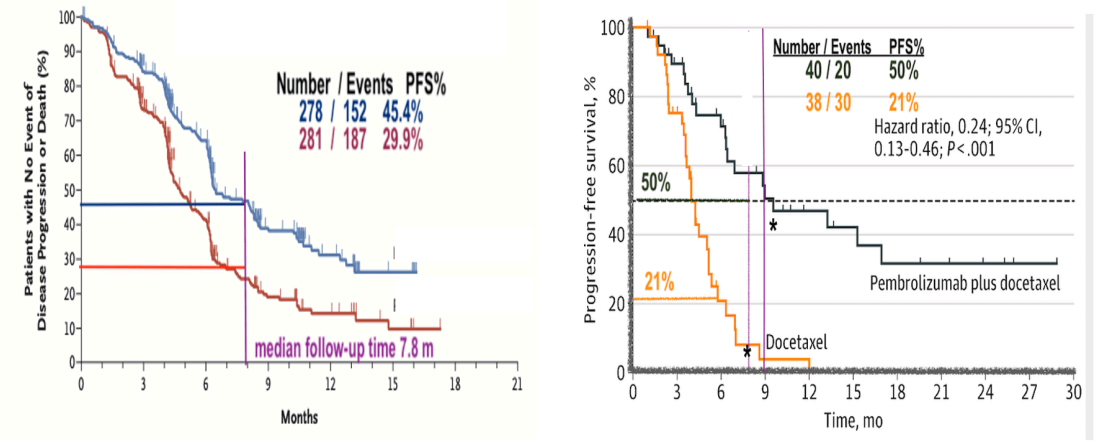

Figure 2. PFS curves from two RCTs in NEJM. (Left, figure 2A from[20] and right, figure 3 from JAMA Oncol[21]. In the left curve, percentages coincide. On the right, the groups slightly differ in median follow-up, 7.9 and 8.9 months. While the percentage coincides with pembrolizumab plus docetaxel (asterisk), it does not in the docetaxel group.

2.2. *Comparison of survival curves*

For RCTs in oncology, whose endpoint is OS or PFS, it is critical to investigate whether the survival differences are statistically different between the control and the experimental arm. This is most often done with the log-rank test[22]. The log-rank test examines the 2 × 2 table consisting of observed versus expected failures in each group whenever a failure occurs and tests significance across all the tables. The test can be extended to more than two groups. While the log-rank test is used to test whether the difference in survival times between two or more groups is statistically different, it does not allow testing the effect of additional independent variables, while the Cox proportional hazard model enables the test of the effect of other independent variables on the survival times of different groups of patients, just like the multiple regression model. Hazard is the dependent variable and can be defined as the probability of having the event at a given time, assuming the patients have survived up to that time. The Hazard Ratio (HR) is the ratio of the hazard occurring at any given time in one group compared with another group at that very time. The log-rank and Cox proportional hazard tests assume the HR is constant over time[23].

Nevertheless, to fully take advantage of the information gathered in RCT published results regarding survival outcomes, the survival analysis should be seen as a "double analysis" presenting i) the results

expressed as Hazard Ratio that informs on the estimated time it takes for a patient to fail (quantitative variable or time to fail) and ii) the survival probabilities of being alive expressed on either relative (RR) or absolute risks (AR) (dichotomous qualitative variable, alive or dead). The presentation of quantitative and qualitative results is needed to have a complete interpretation of the survival outcomes of an RCT.

2.3. Quantitative variable. The Hazard Ratio (HR)

The HR is the outcome of the Cox Proportional Hazard Model, and today, it is increasingly common for the results of survival differences of RCT in cancer to be presented as HR. The HR of the control regimen to the experimental regimen is given by $\exp[\beta]$. When the HR is 1, there are no differences in the time patients in both arms take to suffer the event (death for OS or progression in PFS). An HR <1 indicates the experimental arm is better (patients take longer to have the event). In contrast, an HR >1 indicates that the control arm is better and the experimental arm is worse (experimental arm patients have a shorter time to reach the event).

As the starting point of interpreting HR, the higher the HR decreases, the higher the increase in survival probabilities; however, the magnitude of these survival estimates depends on the survival estimates of the control arm and is not linear. The non-linearity means that at equal HR decreases, the survival probabilities are higher if the control arm survival is low and vice versa. These concepts can be exemplified as follows: At an HR of 0.80 (HR decrease of 20%), when the 1-year OS probability in the control arm is 80%, the corresponding 1-year OS probability in the experimental arm would be 84.5%. On the contrary, under the same HR of 0.80, if the 1-year OS probability in the control arm is 20%, the corresponding 1-year OS probability in the experimental arm would be 29%. When the HR is 0.20 (80% decrease), when the 1-year OS probability in the control arm is 80%, the corresponding 1-year OS probability in the experimental arm would be 96%. On the contrary, at the same HR of 0.20 (80% decrease), if the 1-year OS probability in the control arm is 20%, it would be 72% in the experimental[23][24]. These scenarios are shown in **Figure 3**.

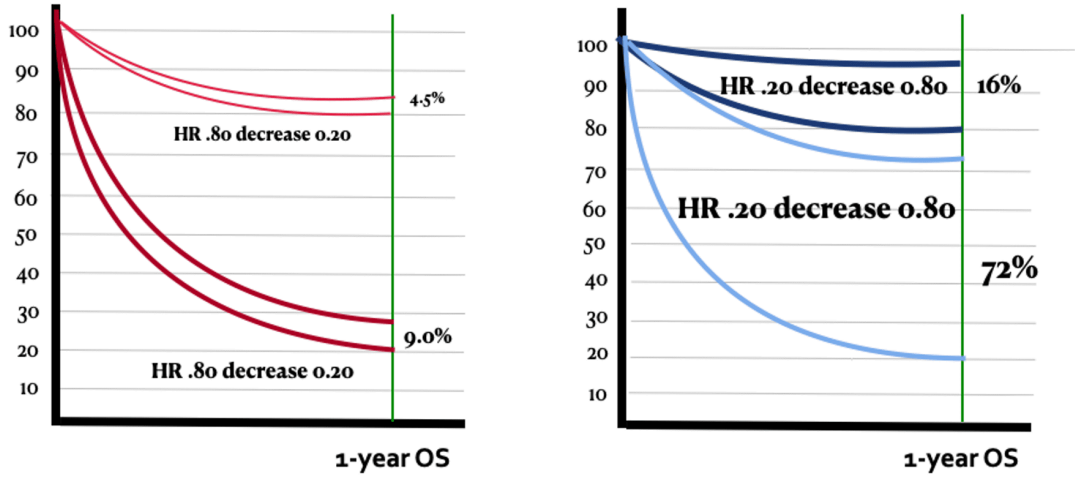

Figure 3. The HR effects on survival depend on OS probabilities of the control arm. The curves are drawn just to illustrate the phenomenon.

It is key to remember that the Cox proportional hazard model is based on the assumption that the ratio of the two hazard functions is approximately constant over time. Such a ratio estimate may capture the relative difference between two survival curves when this assumption is plausible. However, the clinical meaning of such a ratio estimate is difficult, if not impossible, to interpret when the underlying proportional hazards assumption is violated (i.e., the hazard ratio is not constant over time). The most obvious example of a non-proportional hazard is when the survival curves cross. Ideally, the analysis should include a test for the proportionality hazard assumption in each RCT, but this is rarely done. There are several well-known alternatives for quantifying the underlying differences between groups concerning a time-to-event endpoint, but this is a complex issue beyond the scope of this review and suffices to state that the restricted mean survival time (RMST) at a prespecified, fixed time point is a frequently used measure to report the difference between two survival curves when the assumption is not met[25][26][27]. A typical example of a survival curve where the assumption of proportionality was not met is the sotorasib trial[28], as shown in **Figure 4**.

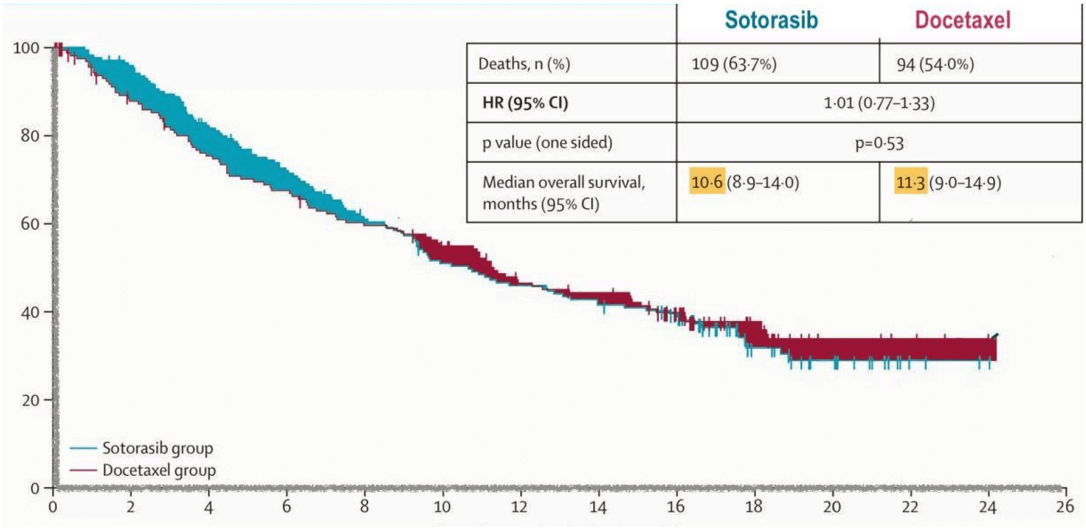

Figure 4. The proportionality hazard assumption is lacking. In the first half, the HRs favor sotorasib, and the opposite favors docetaxel[28]. Curves crossed approximately around month 9, but this fact went unmentioned in the Lancet publication. In this case, with an HR of 1.01, the difference is not statistically significant for OS.

2.4. How the HR is calculated

Several statistical packages can calculate the HR of the Cox proportional hazard model. However, to help understand where the HR values come from, we show in **Figure 5** a simple guide to calculating them with a pencil, a piece of paper, and a pocket calculator.

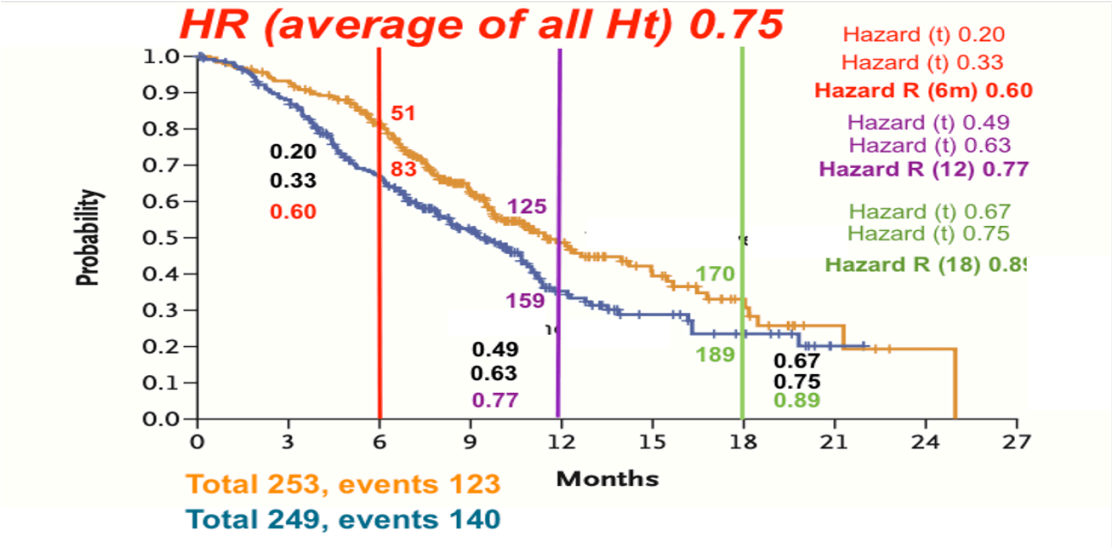

Figure 5. Illustration on an OS curve of how Hazard and Hazard Ratio are calculated at different time points. The averaged individual HRs determine the HR (final) of the study[28].

In this OS figure, yellow is the curve of the experimental arm, and blue is the control arm, which has 253 and 249 total patients, respectively. The HR (final), 0.75, is the mean of many individual HR measurements throughout the study period. Here, we only calculate three individual HR measurements. At 6 months (red vertical line), 12 months (purple vertical line), and 18 months (green vertical line) and averaged them. Because we cannot access the actual data, we visually estimated the number of dead patients from the curves at these three time points. At six months, we estimate 51 and 83 deaths (red numbers), 125 and 159 at 12 months (purple numbers), and 170 and 189 at 18 months (green numbers) for experimental and control arms, respectively. At 6 months, the Hazard for the experimental arm is 0.20 (51÷253= 0.20), and for the control arm, it is 0.33 (83÷249=0.33). To calculate the Hazard Ratio at 6 months, we divide the Hazard for the experimental arm by the Hazard for the control arm (0.20÷0.33=0.60). Therefore, the HR at 6 months is 0.60. The same procedure is repeated for 12 and 18 months, resulting in HR values of 0.77 and 0.89. Finally, we average these individual HR values to obtain the final HR, 0.75. (HR=0.60+0.77+0.83 ÷ 3 = 0.75).

2.5. Statistical significance. P-value and Confidence Interval

P-values mean the probability that an observed difference is due to random chance when the null hypothesis is true. Although a P-value is appropriately considered a statistic interpretable across a

range of values, in contemporary experimental studies, "statistical significance" is now conventionally set at <0.05. The null hypothesis is appropriately rejected if the Type I error probability is <5%. Although a P-value helps determine the reliability with which the null hypothesis can be rejected and the strength of the observed result, it does not provide information regarding its precision[29][30].

The Confidence Interval (CI) is widely used to assess the precision of the P-value result. The CI is calculated around the point estimate of the result to provide a range of values within which the actual value is expected to exist with a given level of confidence. A wide CI suggests an imprecise result and indicates that the results should be interpreted cautiously regardless of statistical significance. Under the conventional acceptance of statistical significance at a P-value of 0.05 or 5%, the CI is frequently calculated at a confidence level of 95%. Generally, if an observed result is statistically significant at a P-value of 0.05, the true value should fall within the 95% CI. Accordingly, the P-value and the 95% CI are taken together to establish if statistical significance exists. For a statistically significant difference to exist in favor of the experimental arm, the HR must be not only lower than 1, but the upper limit of the 95% CI should not reach 1. If it does, then the result cannot be considered statistically significant. It should be noted that the P-value does not necessarily have to be 0.05; it can be lower if contemplated in the statistical plan described in the methods of the RCT. The sample size is calculated based on the assumption of a predetermined HR reduction, type-II error (most commonly 80% to 90%), and type-I error or two-sided p-value and the CI value[31][32].

2.6. One-tailed and two-tailed test

Using a significance level of 0.05 in a two-tailed test, half of the alpha is used to test the statistical significance in one direction, and half is used to test the statistical significance in the other direction. This means that 0.025 is in each tail of the distribution of the test statistic. When using a two-tailed test, regardless of the direction of the relationship you hypothesize, you are testing for the possibility of the relationship in both directions. With the one-tailed test, the possibility of the relationship is in one direction only, completely disregarding the possibility of a relationship in the other direction. According to these simple concepts, choosing a one-tailed test to attain significance is inappropriate. Likewise, choosing a one-tailed test after running a two-tailed test that failed to reject the null hypothesis is inappropriate, no matter how "close" to significant the two-tailed test was. Using inappropriate statistical tests can lead to invalid, non-replicable, and highly questionable results. A

similar situation can occur for the Confidence Interval—a wider CI above 95% results in a higher imprecision of the P-value[29][32]. The RCT study GOG240 testing bevacizumab in advanced cervical cancer is a noticeable example. The trial used a one-sided test with a 98% CI[33], which increased the "acceptance" area. The effect of these two concepts on the probability of attaining statistical significance in an RCT is shown in **Figure 6**.

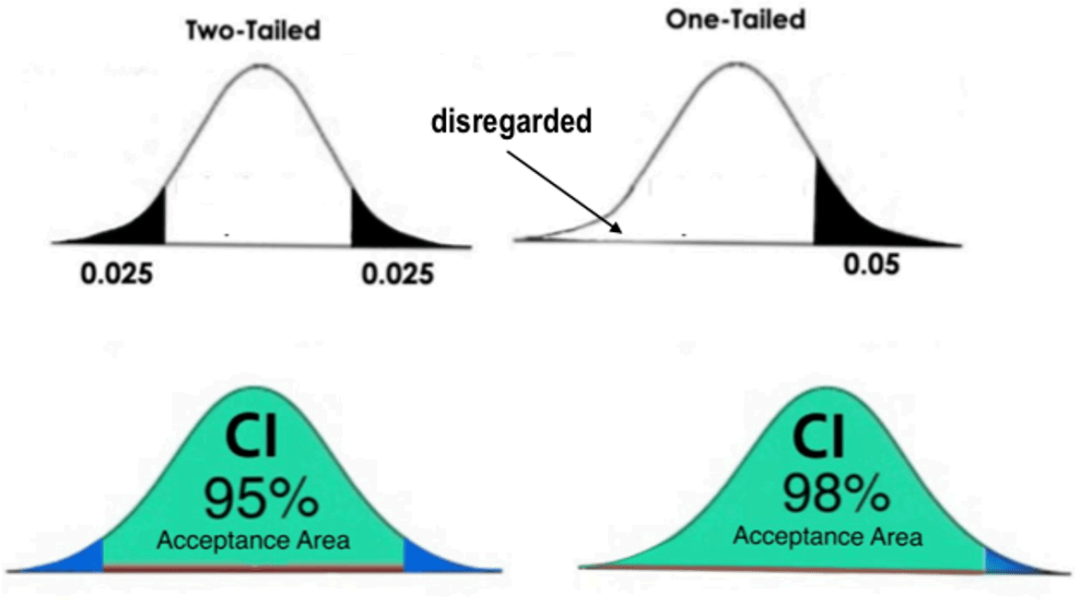

Figure 6. One-sided testing completely disregards any effect in the opposite direction, whereas the wider the CI, the lower the precision of the result.

2.7. Absolute and Relative Risks as dichotomous qualitative variables

Absolute Risk. A measure of the effect of a treatment in an oncology RCT that has OS as an endpoint (the same applies to PFS) is to look at the frequency of bad outcomes (death for OS and progression/death for PFS) of disease in the experimental group being treated compared with those in the control arm and, for instance, **supposing that a well-designed randomized controlled trial in cancer found that 20% of the control group died, compared with only 12% of those receiving experimental treatment**. Without knowing more about the adverse effects, cost, or any other details of the therapy, the treatment reduces the number of deaths in treated patients. Here is where we need to consider the risk of treatment versus no treatment. In healthcare, risk is the probability of a bad outcome in people with the disease.

On these grounds, Absolute Risk (AR) is the most helpful way of presenting research results to help the decision-making process. **In this example, the AR Reduction (ARR) is 8% because 20% minus 12% = 8%.** This means that if 100 patients were treated with the experimental drug, 8 would be prevented from dying. A complementary way of expressing the ARR is the Number Needed-to-Treat (NNT). If 8 patients out of 100 benefit from treatment, the NNT is 13 because 100 ÷ 8 = 12.5 (the NNT is always rounded up). In cases where the experimental treatment leads to a worse outcome, the NNT should be Number Needed-to-Harm (NNH), and the resulting NNH should be rounded down[34]. Absolute Risk and NNT with 95%CI from the RCT discussed here were obtained from the Risk Reduction Calculator http://araw.mede.uic.edu/cgi-bin/nntcalc.pl

2.8. Relative Risk

The relative risk (RR) is essentially the same, but the percentages are divided instead of subtracted, as in Absolute Risks. **Using the same example above, the relative risk reduction (RRR) is 0.60 because (12 percent ÷ 20 percent = 0.60).** When a treatment has an RR higher than 1, the risk of a bad outcome is increased by the treatment; when the RR is less than 1, the risk of a bad outcome is decreased, meaning that the treatment is likely to do well. For example, when the RR is 2.0, the chance of a bad outcome is twice as likely to occur with the treatment as without it, whereas an RR of 0.5 means that the chance of a bad outcome is twice as likely to occur without the intervention. When the RR is exactly 1, the risk is unchanged.

Intuitively, it is more straightforward and more "real" to present the survival results in terms of Absolute risks rather than Relative Risks. Unfortunately, trial sponsors, medical journals, and the media prefer Relative Risks because they magnify the treatment effects[35]. **Figure 7** shows how Relative Risk Reduction can make an enormous difference in Absolute Risk Reduction.

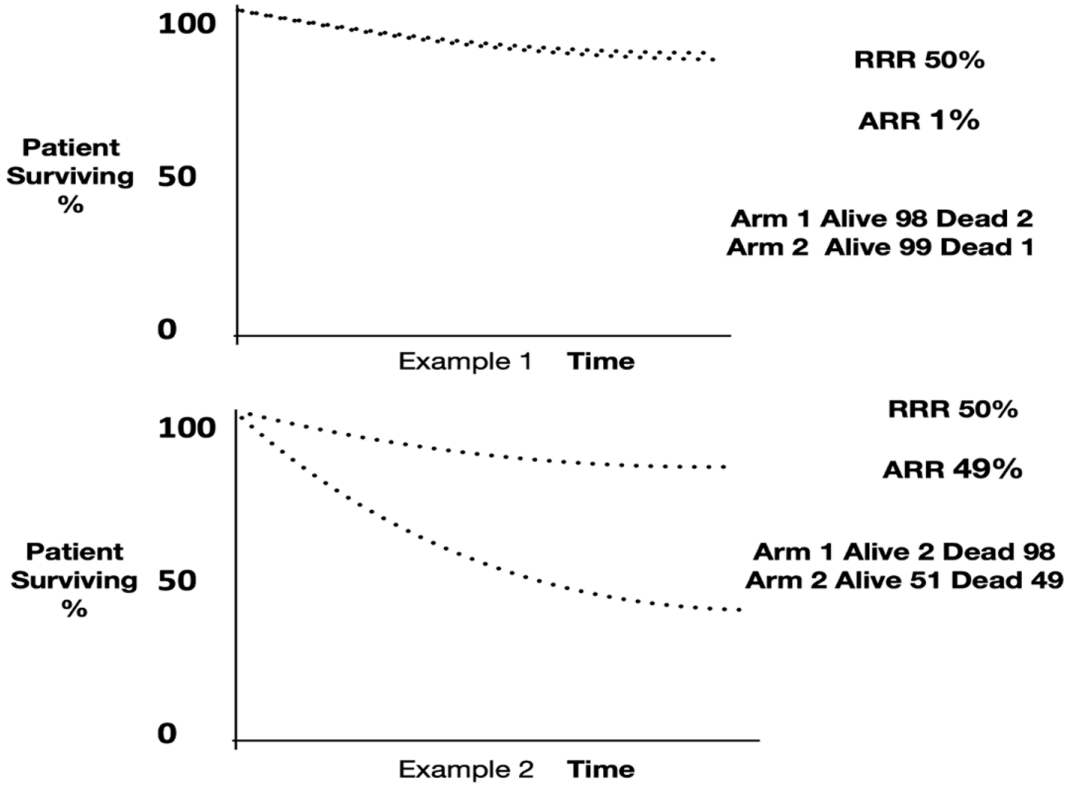

Figure 7. A Relative Risk Reduction of 50% can translate into different Absolute Risk Reductions. Survival curves are imaginary and for illustrative purposes only, as RR and AR are binary data.

2.9. Hazard Ratio (HR) risk reduction must not be presented as Relative Risk Reduction

One additional reason to prefer using absolute risk over relative risk is the wrong but common practice in RCT of expressing the survival benefit based on the HR reduction as if it were a Relative Risk Reduction, which it is not.

Despite the widespread implementation of the Cox Proportional Hazard model, the terminology and interpretation used to describe the estimated HR have become loose and, unfortunately, often incorrect. Although some journals offer guidelines that advise against reporting HRs as relative risks, these guidelines should be more frequently noticed and made mandatory. Due to a lack of understanding, the authors interpret the resultant HR as a relative risk. Such an interpretation is inappropriate and can be misleading. The HR should be described as a relative rate, not a relative risk. While the direction of the HR can be used to explain the direction of the relative risk, the magnitude alone cannot be used to explain the magnitude of the relative risk. Because of that, authors should

refrain from using the HR's magnitude to describe the relative risk's magnitude and be strongly encouraged to ascribe accurate interpretations to the statistics derived from fitted Cox proportional hazards regression models.

In summary, HR should not be interpreted as an RR. If we do not distinguish between HR and RR, the risk reduction (as employed in the publication of clinical trials) implies the durability of the effect in the sense that one is led to believe that for a fraction of the population, the intervention can eliminate the chance of the event occurring (binary or dichotomous result, -death or alive-). This is not the case. The 'risk reduction' based on HR means a reduction in the speed of the event happening (relative rate of the event happening), not the chances of it occurring (a quantitative variable)[24][36].

As such, a typical presentation of results with HR states: Treatment A yields a 30% reduction in the risk of progression or death (for the PFS endpoint) or death (for the OS endpoint) (HR 0.70, 95%CI 0.45-0.90) for progression and death or death, p<0.001, which translated into an increase of 3.5 months of median PFS or OS over treatment B.

The correct statement should be: Treatment A yields a 30% reduction in the **relative rate** of progression and death (for the PFS endpoint) or death (for the OS endpoint) (HR 0.70, 95%CI 0.45-0.90), p<0.001, which translated into an increase of 3.5 months of the median (PFS or OS) over treatment B (as the rate of the event decreases, the survival time increases).

With the above, it is easy to understand why scientific journals, media, and sponsors highlight the results equating the reduction in HR with the reduction of RR because, eventually, they want to convince people of the new treatment's benefits.

2.10. The importance of censoring in analyzing survival results

In a time-to-event analysis, participants are censored when information on the outcome of interest (progression or death event for PFS or death event for OS) is unavailable because the participants are no longer seen in follow-up. Therefore, the K-M method assumes non-informative censoring. In other words, censored patient numbers and clinical characteristics should not differ between the control and experimental arms. When this assumption is not met, the chances of bias increase. Thus, non-informative censoring is "normal or expected to happen." On the contrary, informative censoring may introduce bias[37].

Informative censoring occurs when the reasons are related to the study intervention, potentially introducing post-randomization bias. In quantitative terms, there is a pressing need for consensus on how significant the percentage difference in censoring events between the experimental and the control arms must be to be considered informative censoring. A difference higher than 10% between arms would suggest its existence, emphasizing the need for a clear understanding and agreement on this aspect of censoring.

The pattern of censoring on the time (early or left, late or right) helps interpret the study results. Left censoring, a particularly concerning form of informative censoring, more frequently occurs because of early drug discontinuation, withdrawal of consent, loss to follow-up, or initiating a new anticancer therapy before documenting the event of interest. It typically occurs early during the study, the reason being that early censored patients, in general, are those less fit and more prone to withdraw from the study because of poor tolerance to treatment. Thus, early or left censoring resembles an efficacy analysis as per-treated population instead of intention-to-treat because only the fittest go on follow-up[38][39]. An example of informative censoring is shown in **Figure 8**.

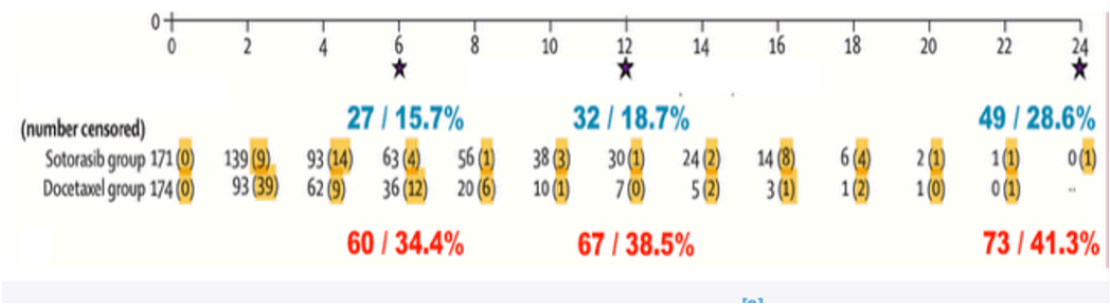

Figure 8. The control arm has a higher percentage of censored patients (41.3% vs. 28.6%), and the imbalance was even more pronounced in the early stages of the study (34.4% vs. 15.7%). Lower part of figure 2B of[28].

In survival analysis, **censored patients cannot be ignored; they have important information to convey, which must be included in the analysis**. By convention, it is assumed that censoring occurs at random. However, when data suggest the presence of informative censoring, sensitivity analyses are required to establish the robustness of the conclusions[40][41][42]. This statistical requirement is recognized by major regulatory agencies such as the European Medicines Agency: "*...sensitivity analysis should show how different assumptions influence the results obtained....*" The National Research Council

"*...sensitivity analyses should be part of the primary reporting of findings from clinical trials*", FDA-ICH E9 addendum[43] "*missing data require particular attention in a sensitivity analysis because the assumptions underlying any method may be hard to justify and impossible to test.*"

Thus, sensitivity analysis is a tool that strengthens the robustness of trial findings. It achieves this by conducting analyses under various plausible assumptions about the methods, models, or data that differ from those used in the pre-specified primary analysis. When the results of the sensitivity analyses align with the primary results, it instills confidence in researchers that the assumptions made for the primary analysis have had minimal impact on the results, thereby reinforcing the trial findings. While several statistical methods exist to perform sensitivity analysis, this is beyond the authors' expertise, and readers are directed to guidance documents. Any RCT should at least mention whether or not a sensitivity analysis was performed. Any result from an RCT with evidence of informative censoring without a sensitivity analysis should be approached with extra caution.

3. Analysis of RCT applying these basic principles

Based on the above statistical concepts, the six RCTs chosen are analyzed in terms of 1) Whether or not the percentage of survival estimated by the total number of patients/number of events at the median follow-time matches with the percentage observed in the curve; 2) the Absolute Risk Reduction and NNT, and 3) Whether the RCT reports on the censored patients (The Absolute Risk Reduction and NNT were calculated using the risk reduction calculator at http://araw.mede.uic.edu/cgi-bin/nntcalc.pl

3.1. NEJM. Amivantanab plus Lazertinib

In this RCT[44], involving patients with previously untreated EGFR-mutated advanced NSCLC, 429 patients were randomized to the Amivantamab-Lazertinib group, 429 to the Osimertinib group, and 216 to the Lazertinib group. The primary endpoint was PFS in the Amivantamab–Lazertinib compared to Osimertinib. The results showed that at a median follow-up time of 22 months, the median PFS was significantly longer in the Amivantamab–Lazertinib group than in the Osimertinib group (23.7 vs. 16.6 months); HR for disease progression or death, 0.70; 95% CI, 0.58 - 0.85; p<0.001). **Figure 9.**

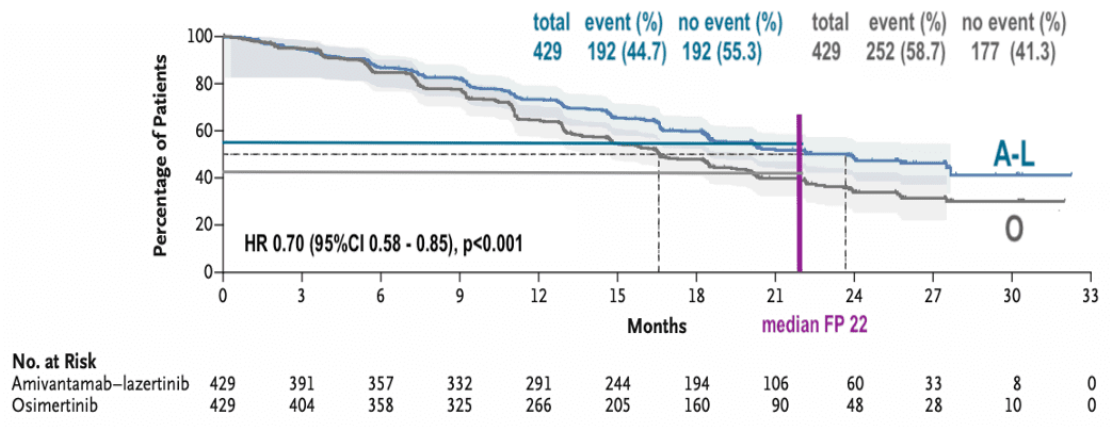

Figure 9. PFS curve of the study. Figure 1A of[44].

1) Survival curves and events

The PFS survival curve and events are not concerns. The PFS percentages in the curve match the progression/death events.

2) Estimation of Absolute Risk Reduction for PFS events

The Absolute Risk Reduction for PFS is as follows (95%CI not shown for simplicity):

Tx	HR	Months	Total	Events	% PFS	ARR	NNT
A+L	**0.70***	23.7	429	192	55.3%	**14***	**8***
O		16.6	429	252	41.3%		

HR, Hazard Ratio, ARR: Absolute Risk Reduction, bold and asterisk, Statistically significant. Event numbers are shown in the forest plot (Figure 1C of 44).

While OS was not the primary endpoint, the preliminary results indicate an HR for death of 0.80 (95% CI, 0.61-1.05), which is not statistically significant. The percentage of survival at the median follow-up time of 22 months is 77.3% vs. 72.3%, which translates into an ARR of 4.66% and an NNT of 26. None of these are statistically significant.

3) Censoring information

No information on the censoring of patients is provided.

This study shows a statistically significant advantage for Amivantamab–Lazertinib compared to Osimertinib regarding HR and AR for PFS. **Unfortunately, the lack of censoring information suggests that caution should be exercised in interpreting the study.**

3.2. The Lancet. Cadonilimab

In this RCT, 45,445 patients with advanced cervical cancer were randomized to cadonilimab (a bispecific antibody targeting PD-1 and CTLA-4) plus platinum-based chemotherapy with or without bevacizumab, or the same regimen plus placebo. The dual primary outcomes were PFS and OS. At a median follow-up time of 18.9 months for PFS, the median PFS was 12.7 months in the cadonilimab group and 8.1 months in the placebo group (HR 0.62, 95% CI 0.49-0.80), p<0.0001). For OS, the median follow-up time was 25.6 months, and the median OS was not reached (27.0 months to not estimable) versus 22.8 months (HR 0.64, 0.48-0.86), p=0.0011). **Figure 10.**

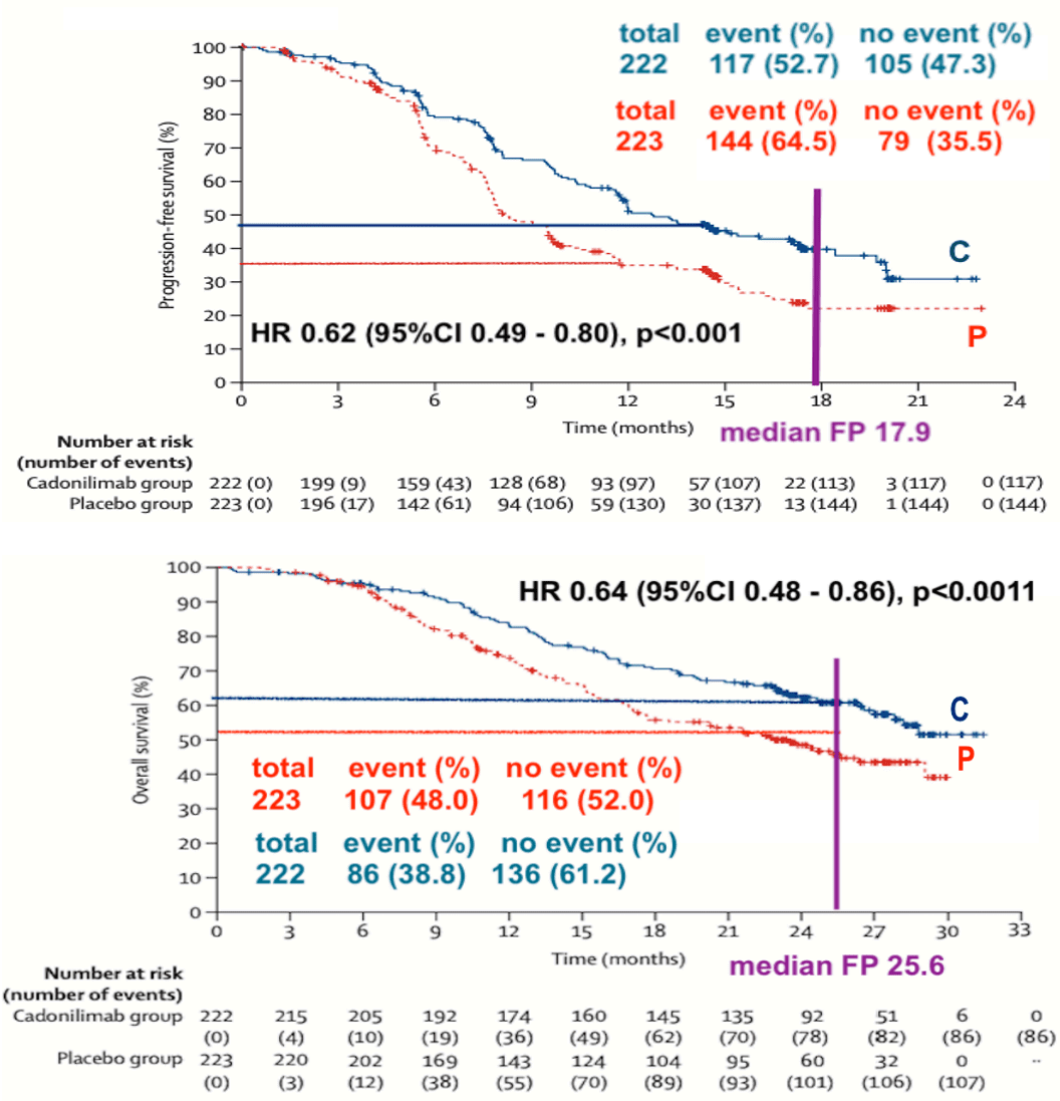

Figure 10. PFS and OS curves of the study. Figures 2A and 3A of[45].

1) Survival curves and events

The PFS percentages in the curve do not match very closely with the percentages from the events, but they did in the OS curve.

2) Estimation of Absolute Risk Reduction for PFS and OS events

The corresponding Absolute Risk Reductions for PFS and OS are as follows (95%CI not shown for simplicity):

Tx	HR	Months	Total	Events	% PFS	ARR	NNT
C	**0.62***	**12.7**	222	117	47.3%	**11.8***	**9***
P		8.1	223	144	35.5%		

Tx	HR	Months	Total	Events	% OS	ARR	NNT
C	**0.64***	**27**	222	86	61.2%	**9.2***	**11***
P		22.8	223	107	52.0%		

HR, Hazard Ratio, ARR: Absolute Risk Reduction, bold and asterisk, Statistically significant. Event numbers are shown in the forest plot (Figure 2B of 45).

3) Censoring information

No information on the censoring of patients is provided.

Overall, this study shows a statistically significant advantage for cadonilimab compared to placebo in terms of HR for PFS and OS. Moreover, the Reduction in Absolute Risks for both survival parameters is statistically significant. **Unfortunately, without information on the censoring of patients and because of the one-sided test design, the results should be taken with caution.**

3.3. The Lancet Oncology. Atezolizumab

This multicenter, double-blind, randomized, placebo-controlled, phase 3 trial studied the comparison of atezolizumab plus chemotherapy (362 patients) versus placebo plus chemotherapy (189 patients) in advanced endometrial cancer[46]. The co-primary endpoints were PFS (in patients with MMR-deficient [dMMR] tumors and the overall population) and OS in the overall population.

At a median follow-up of 28.3 months for the dMMR population, the median PFS was 12.4 months (95%CI 12.4 months–not estimable) in the atezolizumab group and 6.9 months (HR 0.36, 95% CI 0.23-0.57; p=0.0005) in the placebo group. For the overall population, the median follow-up time was 26.7 months, and the PFS was 10.1 months (95% CI 9.5-12.3) versus 8.9 months (8.1-9.6) (HR 0.74, 95% CI 0.61-0·91; p=0.022) respectively. For the OS, the median follow-up time was 28.3 months, and the median OS was 38.7 months (95%CI 30.6 months–not estimable) and 30.2 months (95%CI 25

-37.2), (HR 0.82, 95% CI 0.63-1.07; log-rank p=0.048) respectively. The p-value for the interim analysis of OS did not cross the stopping boundary; therefore, the trial will continue until the required number of events is recorded. **Figure 11.**

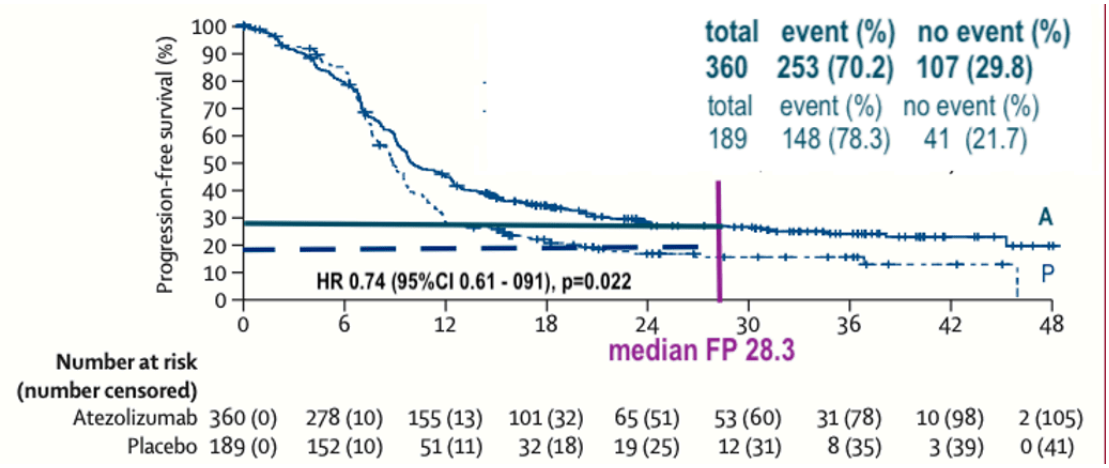

Figure 11. PFS curve from figure 2A of[46].

1) Survival curves and events

The PFS percentages in the curve are close to matching the progression/death percentage of events; hence, there appear to be no concerns regarding survival analysis.

2) Estimation of Absolute Risk Reduction for PFS events

The corresponding Absolute Risk Reductions for PFS and OS are as follows (95%CI not shown for simplicity):

Tx	HR	Months	Total	Events	% PFS	ARR	NNT
A	**0.74***	**10.1**	360	253	29.8%	**8.0***	**13***
P		8.9	189	148	21.7%		

TOTAL POPULATION

Tx	HR	Months	Total	Events	% PFS	ARR	NNT
A	**0.36***	**12.4**	81	37	45.6%	**38.4***	**3***
P		6.9	44	37	16.0%		

dMMR POPULATION

Tx	HR	Months	Total	Events	% OS	ARR	NNT
A	0.82	38.7	360	148	59.9%	5.4	19
P		30.2	189	88	53.5%		

TOTAL POPULATION

HR, Hazard Ratio, ARR: Absolute Risk Reduction, bold and asterisk, Statistically significant. Event numbers are shown in figure 2.

3) Censoring information

Censoring is well-presented, and it indicates it is non-informative.

Overall, this study shows a statistically significant advantage for atezolizumab compared to placebo in terms of HR and ARR for PFS in both the total population and in the dMMR populations. The benefit in terms of PFS is higher in the dMMR population. Regarding OS, the results are preliminary, and so far, atezolizumab is yet to demonstrate it increases survival rates.

3.4. JAMA. Ivonescimab

In this double-blind, placebo-controlled, randomized, phase 3 trial, a total of 322 patients with relapsed advanced or metastatic non-small cell lung cancer with the EGFR variant other than the Thr790Met negative variant were randomized to ivonescimab (161) or placebo (161) plus pemetrexed and carboplatin[47]. The primary endpoint was PFS, and the secondary endpoint was OS. The results are from the first planned interim analysis.

At a median follow-up time of 7.89 months, the median PFS was 7.1 months in the ivonescimab group vs 4.8 months for placebo (difference, 2.3 months; HR, 0.46, 95% CI, 0.34-0.62; p< 0.001). The median OS data were not mature, with 69 deaths (21.4%) having occurred, with 32 and 37 deaths in the ivonescimab and placebo groups, respectively. **Figure 12.**

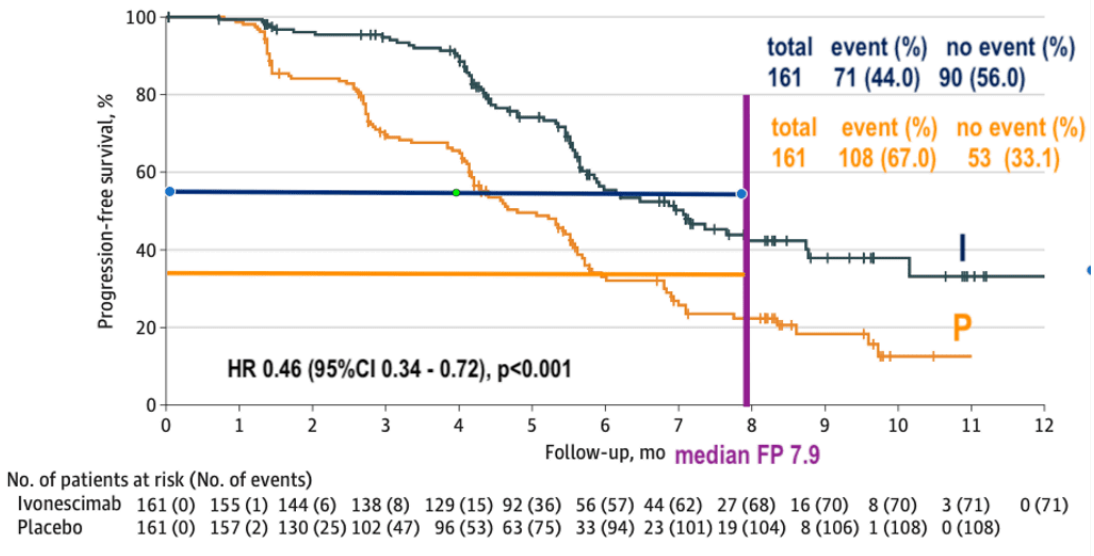

Figure 12. PFS curve from figure 2A of[47].

1) Survival curves and events

The PFS percentages in the curve do not match with the progression/death percentage of events, which may raise some concerns in the survival analysis.

2) Estimation of Absolute Risk Reduction for PFS events

The corresponding Absolute Risk Reduction for PFS is as follows (95%CI not shown for simplicity):

Tx	HR	Months	Total	Events	% PFS	ARR	NNT
I	**0.46***	**7.1**	161	71	56.0%	**22.9***	**5***
P		4.8	161	108	33.1%		

HR, Hazard Ratio, ARR: Absolute Risk Reduction, bold and asterisk, Statistically significant. Event numbers are shown in figure 2.

The OS was a secondary endpoint, and at the date of reporting, there were 32 and 37 deaths for an Absolute Risk Reduction of 3.1% and a NNT of 33, which are not statistically significant.

3) Censoring information

The PFS curve in the printed publication does not show censoring of patients. However, eFigure 1 (K-M plot for investigator-assessed progression-free survival in the intention-to-treat population) shows more censored patients (39 and 29, 25% difference) in the ivonescimab arm.

Overall, this study shows a statistically significant advantage for ivonescimab compared to placebo in terms of HR and Absolute Risk Reduction for PFS. Yet preliminary, there are no differences in death rates. **Unfortunately, the study points to the existence of informative censoring; hence, the results should be taken with caution.**

3.5. JAMA Oncology. Sorafenib plus TACE

In this open-label multicentric randomized phase 3 study, 162 patients with recurrent intermediate-stage HCC after R0 hepatectomy with positive microvascular invasion were randomized to Sorafenib-TACE versus TACE (81 patients in each arm). The primary endpoint was OS. At a median follow-up time of 36.9 and 37.5 months, respectively, the median OS was significantly longer in the SOR-TACE group than in the TACE group (22.2 months vs 15.1 months; HR, 0.55; P <0.001). SOR-TACE also prolonged PFS (16.2 months vs 11.8 months; HR, 0.54; P <.001)[48], **Figure 13**.

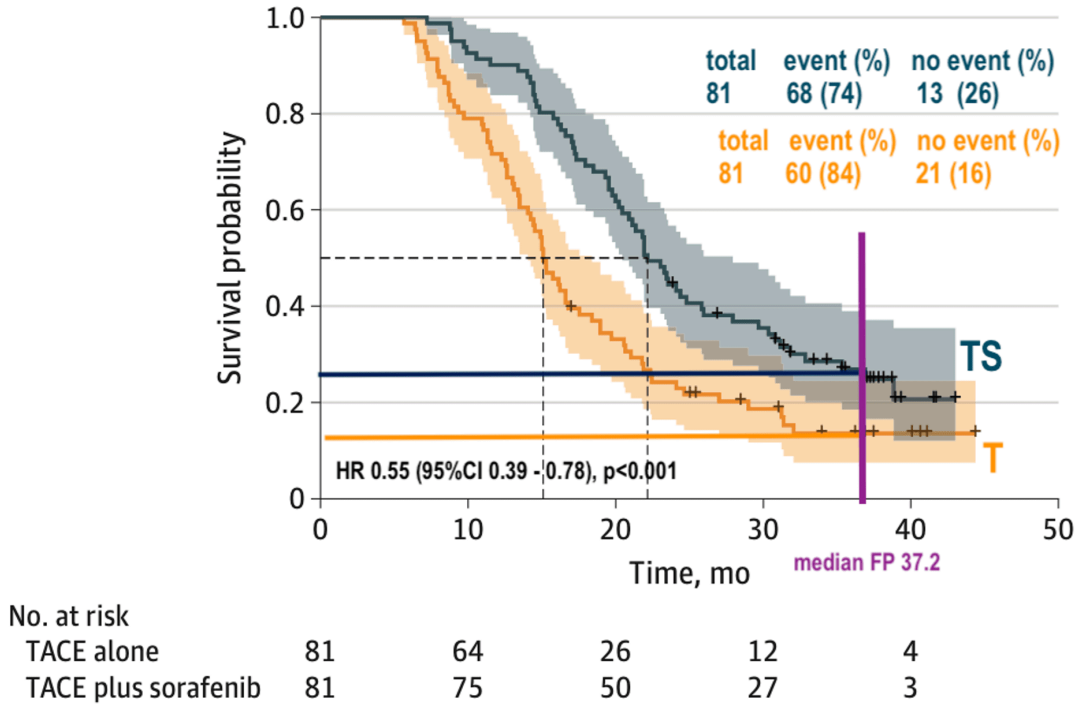

Figure 13. Overall Survival curve is figure 2A from[48].

1) Survival curves and events

The survival percentage estimated from the number of events matches with the curve.

2) Estimation of Absolute Risk Reduction for OS events

The corresponding Absolute Risk Reduction for OS is as follows (95%CI not shown for simplicity):

Tx	HR	Months	Total	Events	% PFS	ARR	NNT
S	**0.55***	**22.2**	81	60	16.0%	10	11
T		15.1	81	68	26.0%		

HR, Hazard Ratio, ARR: Absolute Risk Reduction, bold and asterisk, Statistically significant. Event numbers are shown in the text.

3) Censoring information

No information on the censoring of patients is provided.

Overall, this study shows an increase in the median survival time, which was statistically significant as evaluated by HR; however, neither the Absolute Risk Reduction nor the NNT attained statistical significance. Because the ARR is not statistically significant, it implies that the new therapy may actually increase the risk of death. **Unfortunately, no information on censoring is provided; hence, the results should be taken with caution.**

3.6. Journal Clinical Oncology. Lorlatinib.

This RCT was performed from May 2017 through February 2019, in a total of 296 patients at 104 centers in 23 countries in patients with ALK-positive NSCLC[49]. Patients were randomly assigned to lorlatinib 100 mg once daily (149 patients) or crizotinib 250 mg twice daily (147 patients). The JCO publication[49] informs of a post hoc analysis of updated investigator-assessed efficacy outcomes, safety, and biomarker analyses. At a median follow-up time of 60.2 and 55.1 months for lorlatinib and crizotinib, respectively, the median PFS was not reached (NR) (95%CI, 64.3 - NR) with lorlatinib and 9.1 months (95%CI, 7.4 -10.9) with crizotinib (HR 0.19, 95%CI, 0.13 - 0.27). The 5-year PFS was 60% (95% CI, 51 to 68) and 8% (95% CI, 3 - 14), respectively. To analyze this publication, it is needed to analyze the results of previous publications of this RCT. These are as follows:

N Engl J Med (first publication), **November 2019.**

Data cut-off March 2020. Median follow-up time: 18.3 vs. 14.8 months.

Lancet Respir Med (second publication), **April 2023.**

Data cut-off September 2021. Median follow-up time: 36.7 vs. 29.3 months.

JCO (the third publication described here). **October 2023.**

Data cut-off October 2023. Median follow-up time: 60.2 vs. 51.1 months.

The PFS curve of the JCO[49] publication is the following (**Figure 14**).

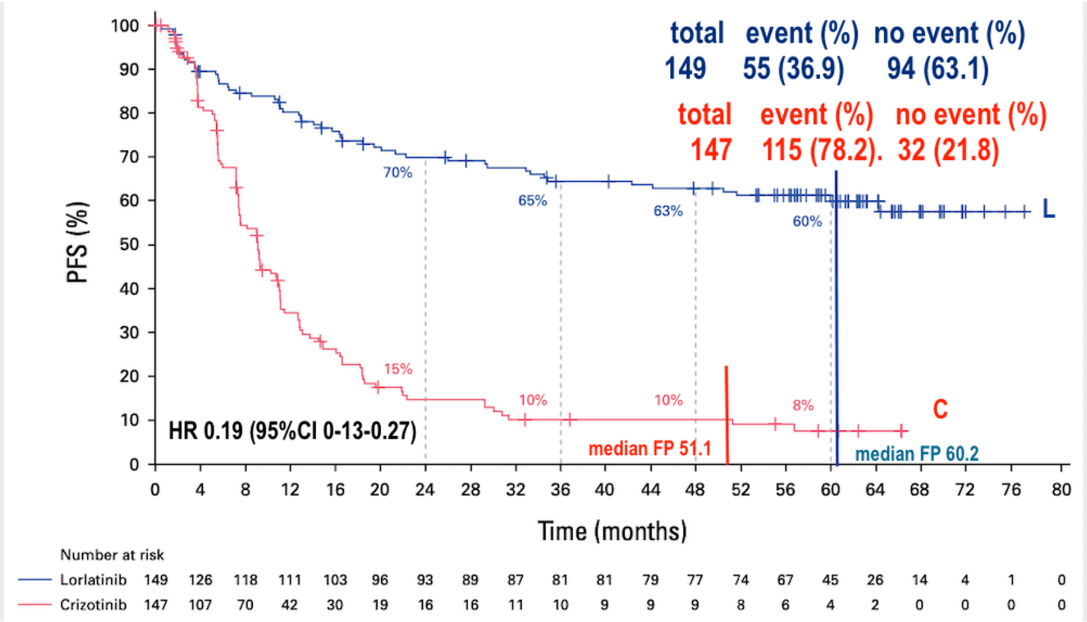

Figure 14. PFS curve from figure 2 of[49].

1) Survival curves and events

The PFS percentage for the crizotinib arm according to the events is 21.8%, while the one shown in the curve is around 10%.

2) Estimation of Absolute Risk Reduction for PFS events

The corresponding Absolute Risk Reduction for PFS is as follows (95%CI not shown for simplicity):

Tx	HR	Months	Total	Events	% PFS	ARR	NNT
L	**0.19***	**NR**	149	55	63.1%	**41.3***	**3***
C		9.1	147	115	21.8%		

HR, Hazard Ratio, ARR: Absolute Risk Reduction, bold and asterisk, Statistically significant. Event numbers are shown in figure 2.

These numbers indicate a high Absolute Risk Reduction for having a PFS of 41.3%, which translates into a NNT of 3. Both are statistically significant.

The results of PFS in terms of HR are very similar to those found in the first two publications. In the NEJM, median PFS was NR vs 9.3 months, HR 0.28 (95%CI, 0.19 to 0.41), P<0.001, whereas in the Lancet Respir Med publication, these were NR vs 9.3 months, HR 0.27, (95%CI, 0.18 to 0.39); P<0.001.

3) Censoring information

As stated above, unfortunately, among these top 6 journals, only the Lancet journals most of the time include censored patients at the bottom of the curve. **Figure 15**, taken from the Lancet Respir Med[50], shows a strong imbalance in the censoring. Overall, 67.1% of the lorlatinib patients were censored (100 out of 149), whereas these were only 55 (almost half) 37.4% for crizotinib. As it can be appreciated, on the right side, as expected, the number of censored patients favored lorlatinib, which most likely results in higher event-free patients in follow-up; however, there is also a high early censoring in crizotinib patients (more than double). **This pattern of censoring clearly shows that it is informative, and in the absence of sensitivity analysis, the results on PFS must be taken with caution.**

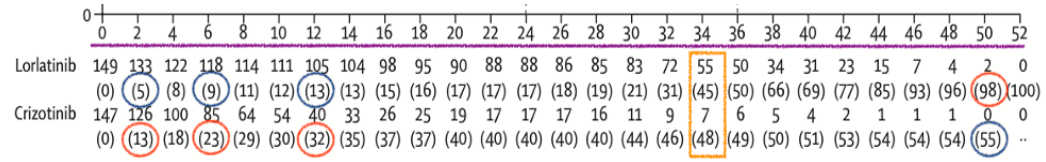

Figure 15. Censoring in the PFS curve. Lower part of figure 2B from[50].

On the other hand, it is at least odd that, as established in the NEJM's first publication on the trial, OS was a secondary endpoint; however, in the last publication of the JCO[49], no word is written about OS. On this basis, **Figure 16** shows the OS curve in the NEJM[51] (first publication).

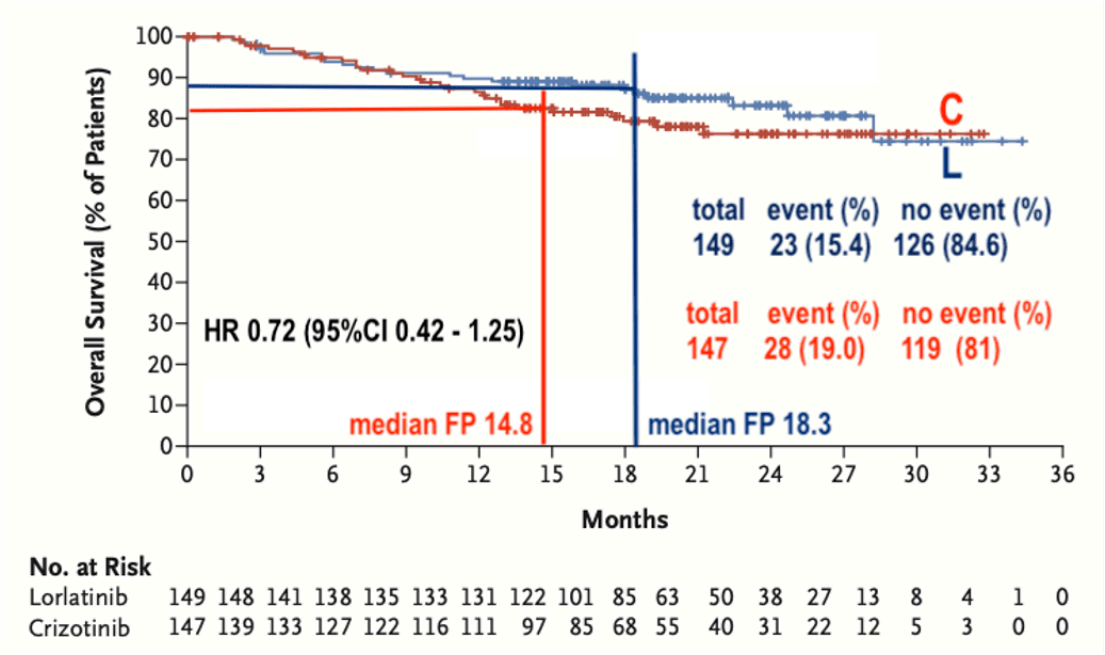

Figure 16. OS curve from 2D of[51].

The median OS has not been reached in both arms, and the HR is 0.72; however, as it can be appreciated, the 95% CI goes beyond 1, therefore this difference is not statistically significant. Likewise, the Absolute Risk Reduction of 3.6% (48.6 – 81 = 3.6) and NNT of 28 (100 ÷ 3.8 = 27.7) are not statistically significant.

Equally odd is that in the second publication in the Lancet Respir Med[50], nothing is mentioned about OS. However, looking at the supplementary data, numerically, in the curve of time to deterioration, there were more events of deterioration in the lorlatinib arm. Neither the HR reduction nor the differences in Absolute Risk were statistically significant. Actually, there is a trend for Absolute Risk increase (5%), and the NNH (Number Needed-to-Harm) is 21%. Both are non-statistically significant (**Figure 17**).

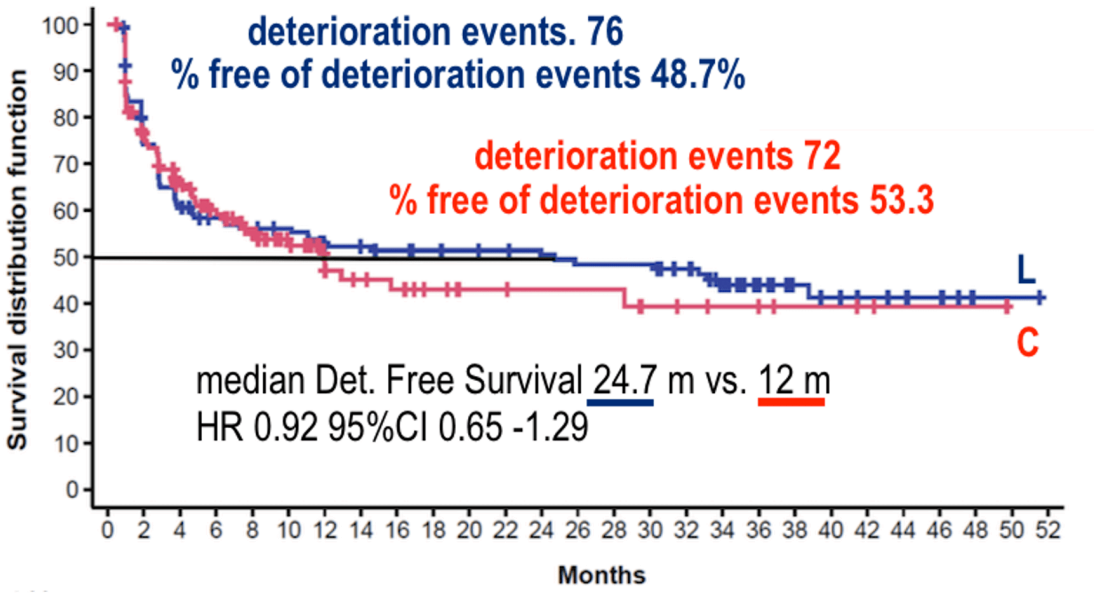

Figure 17. Curve for time to deterioration shown in supplementary appendix S4B from[50].

Finally, we analyzed the total number of deaths registered in the three publications. The information was taken from the corresponding flow diagram figures of these three publications[49][50][51]. These are as follows:

Publication	Deaths Lorlatinib	Deaths Crizotinib
NEJM	23	28
Lancet Respir Med	9	4
J Clin Oncol	12	4
Total	44	36

Based on these data, the Absolute Risk (AR) for OS is as follows:

doi.org/10.32388/RNYMQG

Tx	Total	Events	% OS	AR increase	NNHarm
Lorlatinib	149	44	70.5	5	21
Crizotinib	147	36	75.5		

Overall, this study reports a highly significant effect of lorlatinib over crizotinib in PFS, evaluated by HR and AR as well. However, the clear evidence of informative censoring, the numerically more deterioration events in the lorlatinib arms, and the trend for increased risk of death in lorlatinib suggest that the results should be taken with caution.

4. Comments

The ultimate 'user' of the information provided by the oncology RCT is the cancer patient who has to decide whether to undergo a treatment. This underscores the critical importance of comprehensive and unbiased reporting in RCTs. Such information is not just a guide, but a lifeline for the oncologist and the patient, enabling them to make better treatment decisions. Here, we show in this sample of 6 RCTs from the "big six" that the information on survival outcomes may not be sufficient to assess the efficacy of the investigated treatments. In this sense, we must stress that our analysis of these RCTs is not intended to judge the completeness or adherence of the trial reports to the CONSORT checklist, but is only restricted to three points we consider key for evaluating efficacy on survival: 1) Whether or not the percentage of survival estimated by the total number of patients/number of events at the median follow-time matches with the percentage observed in the curve; 2) the Absolute Risk Reduction and NTT, and 3) Whether the RCT reports on the censored patients.

The amivantanab study at the NEJM[44] failed to provide the censoring of patients. Censoring was also missed in the cadonilimab trial published in The Lancet[45]; moreover, the percentage of PFS does not match that shown in the curve at the median follow-up time. From our analysis, the atezolizumab trial in Lancet Oncol[46] is well presented regarding the PFS curve and data; the trial reports on censoring, which is not informative. The JAMA trial on Ivonescimab shows a mismatch in the PFS percentage between the curve and the number of events. Moreover, while the censoring is not presented at the PFS curve in the printed publication, figure e1 of supplementary data shows 25%

higher censored patients in the experimental arm, which suggests informative censoring[47]. The JAMA Oncol trial[48] of sorafenib plus TACE does not show the censoring. The Lorlatinib trial in JCO[49] shows a mismatch in the PFS percentage between the curve and the number of events for the crizotinib arm. More importantly, the second publication of the trial in The Lancet Respir Med[50] shows highly informative censoring and numerically more deterioration events for lorlatinib despite no statistically significant differences in the HR for time to deterioration. Of note, data from the three trial publications show a 5% absolute increase in the risk of death for lorlatinib, though this difference was not statistically significant.

As usual in the reporting of RCTs in Oncology today, these six publications do not present the results of binary outcomes for PFS and OS in terms of both absolute or relative effect sizes (Absolute and Relative Risks) as recommended by the CONSORT 2010 checklist of information (17b, which establishes that for binary outcomes, the presentation of both absolute and relative effect sizes is recommended. https://legacyfileshare.elsevier.com/promis_misc/CONSORT-2010-Checklist.pdf

Our analysis regarding these three issues is pointed out in **Table 1**.

Trial agent	Journal	Curve match	Absolute effect size	Censoring
Amivantanab Plus Lazertinib	NEJM	Yes	Not shown	Not provided
Cadonilimab	LANCET	No	Not shown	Not provided
Atezolizumab	LANCET ONCOL	Yes	Not shown	Provided
Ivonescimab	JAMA	No	Not shown	Provided*
Sorafenib plus TACE	JAMA ONCOL	Yes	Not shown	Not provided
Lorlatinib	JCO	No	Not shown	Not provided**

Table 1. Summary of the 6 trials.

**In the appendix, informative, not discussed. ** Not provided in JCO, found in a previous trial publication.*

Previous studies have shown that most RCTs in Oncology fail to provide complete reporting of survival endpoints, jeopardizing the oncologist's interpretation of the study findings[52]. In a revision of 125 articles from eight significant journals on RCT in cancer[53], 68% reported insufficient information on the survival analysis, including the lack of data on censoring (42%) and the number of events (28%). In another review of 32 articles reporting survival outcomes in cancer populations, none of the publications reported details relating to the final model validation to analyze survival outcomes. Moreover, in 88% of the studies that reported the use of Cox proportional hazards regression to analyze survival endpoints, most failed to report the validation of the statistical models in terms of the proportionality hazard assumption[54].

While we acknowledge that we cannot generalize from this analysis of 6 RCTs, we must be aware that this problem, previously noted in these publications[52][53][54], could persist to some degree.

In strict adherence to the CONSORT 2010 checklist, RCTs must present the numerical results in absolute numbers, and binary outcomes—progression/death (PFS) and death (OS)—must be presented on both absolute and relative effect sizes. A question in this regard, which is left to expert statisticians, is whether or not the sample size estimated in RCTs to find statistically significant differences in HR is enough to find it in terms of ARR and NNT.

Whether the percentage of survival estimated from the total number of patients and events must match the observed curve issue would be enriched by the input of expert statisticians. Nonetheless, a survival curve that does not provide the number of events leaves the reader with no option to trust that the survival results are correct. On the other hand, the importance of censoring patients cannot be overemphasized. Unfortunately, among these six trials, only one presented the data on censoring in the curve, and another did, but only in the supplementary data appendix. Prasad and Bilal show a remarkable example of the importance of censoring. They show that altering the assumptions for censoring may change the significant conclusions of clinical trials. As such, the number of censored patients at each time interval should be routinely reported in randomized trials to better understand the implications of censoring[55].

As can be observed, this work is not intended to analyze the clinical, pharmacological, or molecular mechanisms of the drugs used in the analyzed RCT. Neither here do we analyze study designs, randomization, control arms, crossover, or any other trial-specific maneuver. It is only intended to point out the three key aspects of how the survival parameters of the trials are presented.

Equally relevant is to keep in mind that the Hazard Ratio cannot be interpreted as a Relative Risk. The Hazard Ratio, a measure of the risk of an event happening in the treatment group compared to the control group, is often misinterpreted as a measure of Relative Risk. However, risk reduction implies the durability of the effect and that for a fraction of the population, the intervention can eliminate the chance of the event occurring (binary or dichotomous result -death or alive-). This is not the case. The 'risk reduction' based on HR reduction means a reduction in the speed of the event happening (relative rate of the event happening), not the chances of it occurring.

In summary, how the results of RCTs in oncology are presented, especially in highly influential journals, must always be stressed because of their tremendous impact on clinical oncology practice. It is possible that "the efficacy," as shown in the publications by authors but above all, by the media, can change depending on how much information is provided and how skilled the oncologist who reads the publication is.

We hope this work encourages oncologists to become more skilled in analyzing RCTs and encourages statistician experts to provide valuable insights and knowledge on this issue.

References

1. ^*Ross JS, Gross CP, Krumholz HM. "Promoting transparency in pharmaceutical industry-sponsored research." Am J Public Health. 2012 Jan; 102(1):72-80. doi:10.2105/AJPH.2011.300187.*
2. ^*Abola MV, Prasad V. "The use of superlatives in cancer research." JAMA Oncol. 2016 Jan; 2(1):139-41. doi:10.1001/jamaoncol.2015.3931.*
3. ^*Elia Ben-Ari. Addressing the Challenges of Cancer Misinformation on Social Media. September 9, 2021. NIH. National Cancer Institute. https://www.cancer.gov/news-events/cancer-currents-blog/2021/cancer-misinformation-social-media*
4. ^*Dilts DM. "Time has come to raise the bar in oncology clinical trials." J Clin Oncol. 2014 Apr 20; 32(12): 1186-7. doi:10.1200/JCO.2013.54.5277.*
5. ^*Schnog JB, Samson MJ, Gans ROB, Duits AJ. "An urgent call to raise the bar in oncology." Br J Cancer. 2021 Nov; 125(11):1477-1485. doi:10.1038/s41416-021-01495-7.*
6. ^*Ellis LM, Bernstein DS, Voest EE, Berlin JD, Sargent D, Cortazar P, Garrett-Mayer E, Herbst RS, Lilenbaum RC, Sima C, Venook AP, Gonen M, Schilsky RL, Meropol NJ, Schnipper LE. "American Society of Clinic*

al Oncology perspective: Raising the bar for clinical trials by defining clinically meaningful outcomes." J Clin Oncol. 2014 Apr 20; 32(12):1277-80. doi:10.1200/JCO.2013.53.8009.

7. ^*Chen EY, Raghunathan V, Prasad V. "An overview of cancer drugs approved by the us food and drug ad ministration based on the surrogate end point of response rate." JAMA Intern Med. 2019 Jul 1; 179(7):91 5-921. doi:10.1001/jamainternmed.2019.0583.*
8. ^*Gyawali B, Hey SP, Kesselheim AS. "Assessment of the clinical benefit of cancer drugs receiving acceler ated approval." JAMA Intern Med. 2019 Jul 1; 179(7):906-913. doi:10.1001/jamainternmed.2019.0462.*
9. ^*Michaeli DT, Michaeli T. "Overall survival, progression-free survival, and tumor response benefit supp orting initial US Food and Drug Administration approval and indication extension of new cancer drugs, 2003-2021." J Clin Oncol. 2022 Dec 10; 40(35):4095-4106. doi:10.1200/JCO.22.00535.*
10. ^*Meursinge Reynders R. "Clinicians: beware of 'spin'." Evid Based Dent. 2024 Mar; 25(1):1-2. doi:10.103 8/s41432-024-00984-x.*
11. ^*Boutron I. "Spin in scientific publications: a frequent detrimental research practice." Ann Emerg Med. 2 020 Mar; 75(3):432-434. doi:10.1016/j.annemergmed.2019.11.002.*
12. ^*Duran M, Boutron I, Hopewell S, Bonnet H, Sidorkiewicz S. "A cross-sectional study assessing visual ab stracts of randomized trials revealed inadequate reporting and high prevalence of spin." J Clin Epidemio l. 2024 Sep 24; 176:111544. doi:10.1016/j.jclinepi.2024.111544.*
13. ^*Rai S, Mishra P, Ghoshal UC. "Survival analysis: A primer for the clinician scientists." Indian J Gastroen terol. 2021 Oct; 40(5):541-549. doi:10.1007/s12664-021-01232-1.*
14. ^*Altman DG. London (UK): Chapman and Hall; 1992. Analysis of Survival times. In: Practical statistics fo r Medical research; pp. 365–93.*
15. ^*Clark TG, Bradburn MJ, Love SB, Altman DG. "Survival analysis part I: basic concepts and first analyse s." Br J Cancer. 2003 Jul 21; 89(2):232-8. doi:10.1038/sj.bjc.6601118.*
16. ^*Goel MK, Khanna P, Kishore J. "Understanding survival analysis: Kaplan-Meier estimate." Int J Ayurve da Res. 2010 Oct; 1(4):274-8. doi:10.4103/0974-7788.76794.*
17. ^*Rich JT, Neely JG, Paniello RC, Voelker CC, Nussenbaum B, Wang EW. "A practical guide to understandi ng Kaplan-Meier curves." Otolaryngol Head Neck Surg. 2010 Sep; 143(3):331-6. doi:10.1016/j.otohns.20 10.05.007.*
18. ^*Vergote I, González-Martín A, Fujiwara K, et al. "Tisotumab vedotin as second- or third-line therapy f or recurrent cervical cancer." N Engl J Med. 2024 Jul 4; 391(1):44-55. doi:10.1056/NEJMoa2313811.*

19. ^Choueiri TK, Powles T, Peltola K, et al. "Belzutifan versus everolimus for advanced renal-cell carcinoma." N Engl J Med. 2024 Aug 22; 391(8):710-721. doi:10.1056/NEJMoa2313906.

20. ^Paz-Ares L, Luft A, Vicente D, Tafreshi A, Gümüş M, Mazières J, Hermes B, Çay Şenler F, Csőszi T, Fülöp A, Rodríguez-Cid J, Wilson J, Sugawara S, Kato T, Lee KH, Cheng Y, Novello S, Halmos B, Li X, Lubinieck i GM, Piperdi B, Kowalski DM; KEYNOTE-407 Investigators. "Pembrolizumab plus chemotherapy for squamous non-small-cell lung cancer." N Engl J Med. 2018 Nov 22; 379(21):2040-2051. doi:10.1056/NEJMoa1810865.

21. ^Arrieta O, Barrón F, Ramírez-Tirado LA, Zatarain-Barrón ZL, Cardona AF, Díaz-García D, Yamamoto Ramos M, Mota-Vega B, Carmona A, Peralta Álvarez MP, Bautista Y, Aldaco F, Gerson R, Rolfo C, Rosell R. "Efficacy and safety of pembrolizumab plus docetaxel vs docetaxel alone in patients with previously treated advanced non-small cell lung cancer: The PROLUNG phase 2 randomized clinical trial." JAMA Oncol. 2020 Jun 1; 6(6):856-864. doi:10.1001/jamaoncol.2020.0409.

22. ^Bland JM, Altman DG. "The logrank test." BMJ. 2004 May 1; 328(7447):1073. doi:10.1136/bmj.328.7447.1073

23. a, bBarraclough H, Simms L, Govindan R. "Biostatistics primer: what a clinician ought to know: hazard ratios." J Thorac Oncol. 2011 Jun; 6(6):978-82. doi:10.1097/JTO.0b013e31821b10ab.

24. a, bSashegyi A, Ferry D. "On the interpretation of the hazard ratio and communication of survival benefit." Oncologist. 2017 Apr; 22(4):484-486. doi:10.1634/theoncologist.2016-0198.

25. ^Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, Schrag D, Takeuchi M, Uyama Y, Zhao L, Skali H, Solomon S, Jacobus S, Hughes M, Packer M, Wei LJ. "Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis." J Clin Oncol. 2014 Aug 1; 32(22):2380-5. doi:10.1200/JCO.2014.55.2208.

26. ^Royston P, Parmar MK. "The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt." Stat Med. 2011 Aug 30; 30(19):2409-21. doi:10.1002/sim.4274

27. ^Huang B, Kuan PF. "Comparison of the restricted mean survival time with the hazard ratio in superiority trials with a time-to-event end point." Pharm Stat. 2018 May; 17(3):202-213. doi:10.1002/pst.1846.

28. a, b, c, dde Langen AJ, Johnson ML, Mazieres J, Dingemans AC, Mountzios G, Pless M, Wolf J, Schuler M, Lena H, Skoulidis F, Yoneshima Y, Kim SW, Linardou H, Novello S, van der Wekken AJ, Chen Y, Peters S, Felip E, Solomon BJ, Ramalingam SS, Dooms C, Lindsay CR, Ferreira CG, Blais N, Obiozor CC, Wang Y, Mehta B, Varrieur T, Ngarmchamnanrith G, Stollenwerk B, Waterhouse D, Paz-Ares L; CodeBreaK 200 Inve

stigators. "Sotorasib versus docetaxel for previously treated non-small-cell lung cancer with KRASG12C mutation: a randomised, open-label, phase 3 trial." Lancet. 2023 Mar 4; 401(10378):733-746. doi:10.1016/S0140-6736(23)00221-0.

29. ^a, ^b*Flechner L, Tseng TY. "Understanding results: P-values, confidence intervals, and number need to treat." Indian J Urol. 2011 Oct; 27(4):532-5. doi:10.4103/0970-1591.91447.*

30. ^*Shreffler J, Huecker MR. Hypothesis Testing, P Values, Confidence Intervals, and Significance. 2023 Mar 13. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan–. PMID: 32491353.*

31. ^*Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. Br Med J (Clin Res Ed). 1986 Mar 15;292(6522):746-50.*

32. ^a, ^b*Dorey FJ. In brief: statistics in brief: Confidence intervals: what is the real result in the target population? Clin Orthop Relat Res. 2010 Nov;468(11):3137-8*

33. ^*Tewari KS, Sill MW, Long HJ 3rd, Penson RT, Huang H, Ramondetta LM, Landrum LM, Oaknin A, Reid TJ, Leitao MM, Michael HE, Monk BJ. Improved survival with bevacizumab in advanced cervical cancer. N Engl J Med. 2014 Feb 20;370(8):734-43. doi:10.1056/NEJMoa1309748. Erratum in: N Engl J Med. 2017 Aug 17;377(7):702. doi:10.1056/NEJMx170002.*

34. ^*Irwig L, Irwig J, Trevena L, et al. Smart Health Choices: Making Sense of Health Advice. London: Hammersmith Press; 2008. Chapter 18, Relative risk, relative and absolute risk reduction, number needed to treat and confidence intervals. Available from: https://www.ncbi.nlm.nih.gov/books/NBK63647/*

35. ^*Dupont WD, Plummer WD Jr. Understanding the relationship between relative and absolute risk. Cancer. 1996 Jun 1;77(11):2193-9. doi:10.1002/(SICI)1097-0142(19960601)77:11<2193::AID-CNCR2>3.0.CO;2-R. PMID: 8635083.*

36. ^*Sutradhar R, Austin PC. Relative rates not relative risks: addressing a widespread misinterpretation of hazard ratios. Ann Epidemiol. 2018 Jan;28(1):54-57. doi:10.1016/j.annepidem.2017.10.014.*

37. ^*Templeton AJ, Amir E, Tannock IF. Informative censoring - a neglected cause of bias in oncology trials. Nat Rev Clin Oncol. 2020 Jun;17(6):327-328. doi:10.1038/s41571-020-0368-0.*

38. ^*Wilson BE, Nadler MB, Desnoyers A, Amir E. Quantifying withdrawal of consent, loss to follow-up, early drug discontinuation, and censoring in oncology trials. J Natl Compr Canc Netw. 2021;19:1433-1440. doi:10.6004/jnccn.2021.7015.*

39. ^*Rosen K, Prasad V, Chen EY. Censored patients in Kaplan-Meier plots of cancer drugs: An empirical analysis of data sharing. Eur J Cancer. 2020;141:152-161. doi:10.1016/j.ejca.2020.09.031.*

40. ^Parpia, S., Morris, T.P., Phillips, M.R. et al. Sensitivity analysis in clinical trials: three criteria for a valid sensitivity analysis. Eye 36, 2073–2074 (2022). doi:10.1038/s41433-022-02108-0

41. ^Morris TP, Kahan BC, White IR. Choosing sensitivity analyses for randomised trials: principles. BMC M ed Res Methodol. 2014;14:1–5.

42. ^Liu Y. Sensitivity analyses for informative censoring in survival data: A trial example. J Biopharm Stat. 2017;27(4):595-610. doi:10.1080/10543406.2016.1167076.

43. ^E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinic al Trials. Accessed December 2024. https://www.fda.gov/regulatory-information/search-fda-guidance -documents/e9r1-statistical-principles-clinical-trials-addendum-estimands-and-sensitivity-analys is-clinical

44. a, b, cCho BC, Lu S, Felip E, Spira AI, Girard N, Lee JS, Lee SH, Ostapenko Y, Danchaivijitr P, Liu B, Alip A, Korbenfeld E, Mourão Dias J, Besse B, Lee KH, Xiong H, How SH, Cheng Y, Chang GC, Yoshioka H, Yang J C, Thomas M, Nguyen D, Ou SI, Mukhedkar S, Prabhash K, D'Arcangelo M, Alatorre-Alexander J, Vázque z Limón JC, Alves S, Stroyakovskiy D, Peregudova M, Şendur MAN, Yazici O, Califano R, Gutiérrez Calder ón V, de Marinis F, Passaro A, Kim SW, Gadgeel SM, Xie J, Sun T, Martinez M, Ennis M, Fennema E, Daks h M, Millington D, Leconte I, Iwasawa R, Lorenzini P, Baig M, Shah S, Bauml JM, Shreeve SM, Sethi S, K noblauch RE, Hayashi H; MARIPOSA Investigators. Amivantamab plus lazertinib in previously untreate d EGFR-mutated advanced NSCLC. N Engl J Med. 2024 Oct 24;391(16):1486-1498. doi:10.1056/NEJMoa 2403614.

45. a, bWu X, Sun Y, Yang H, Wang J, Lou H, Li D, Wang K, Zhang H, Wu T, Li Y, Wang C, Li G, Wang Y, Li D, T ang Y, Pan M, Cai H, Wang W, Yang B, Qian H, Tian Q, Yao D, Cheng Y, Wei B, Li X, Wang T, Hao M, Wan g X, Wang T, Ran J, Zhu H, Zhu L, Liu X, Li Y, Chen L, Li Q, Yan X, Wang F, Cai H, Zhang Y, Liang Z, Liu F, Huang Y, Xia B, Qu P, Zhu G, Chen Y, Song K, Sun M, Chen Z, Zhou Q, Hu L, Abulizi G, Guo H, Liao S, Ye Y, Yan P, Tang Q, Sun G, Liu T, Lu D, Hu M, Wang ZM, Li B, Xia M. Cadonilimab plus platinum-based ch emotherapy with or without bevacizumab as first-line treatment for persistent, recurrent, or metastatic cervical cancer (COMPASSION-16): a randomised, double-blind, placebo-controlled phase 3 trial in Chi na. Lancet. 2024 Oct 26;404(10463):1668-1676. doi:10.1016/S0140-6736(24)02135-4.

46. a, b, cColombo N, Biagioli E, Harano K, Galli F, Hudson E, Antill Y, Choi CH, Rabaglio M, Marmé F, Marth C, Parma G, Fariñas-Madrid L, Nishio S, Allan K, Lee YC, Piovano E, Pardo B, Nakagawa S, McQueen J, Z amagni C, Manso L, Takehara K, Tasca G, Ferrero A, Tognon G, Lissoni AA, Petrella M, Laudani ME, Rull i E, Uggeri S, Barretina Ginesta MP; AtTEnd study group. Atezolizumab and chemotherapy for advanced

or recurrent endometrial cancer (AtTEnd): a randomised, double-blind, placebo-controlled, phase 3 tri al. Lancet Oncol. 2024 Sep;25(9):1135-1146. doi:10.1016/S1470-2045(24)00334-6.

47. a, b, c *HARMONi-A Study Investigators; Fang W, Zhao Y, Luo Y, Yang R, Huang Y, He Z, Zhao H, Li M, Li K, Song Q, Du X, Sun Y, Li W, Xu F, Wang Z, Yang K, Fan Y, Liu B, Zhao H, Hu Y, Jia L, Xu S, Yi T, Lv D, La n H, Li M, Liang W, Wang Y, Yang H, Jia Y, Chen Y, Lu J, Feng J, Liu C, Zhou M, Zhou J, Liu X, Zhou N, He M, Dong X, Chen H, Chen Y, Su H, Li X, Zhang Z, Yang L, Cheng Y, Chen L, Hou X, Zhang Y, Guo J, Wang Z, Lu H, Wu D, Feng W, Li W, Huang J, Wang Y, Song X, Peng J, Liu L, Guo Y, Li W, Lu D, Hu M, Wang ZM, Li B, Xia M, Zhang L. Ivonescimab plus chemotherapy in non-small cell lung cancer with EGFR variant: A randomized clinical trial. JAMA. 2024 Aug 20;332(7):561-570. doi:10.1001/jama.2024.10613.*

48. a, b, c *Fan W, Zhu B, Chen S, Wu Y, Zhao X, Qiao L, Huang Z, Tang R, Chen J, Lau WY, Chen M, Li J, Kuang M, Peng Z. Survival in patients with recurrent intermediate-stage hepatocellular carcinoma: sorafenib Plus TACE vs TACE alone randomized clinical trial. JAMA Oncol. 2024 Aug 1;10(8):1047-1054. doi:10.10 01/jamaoncol.2024.1831.*

49. a, b, c, d, e, f, g *lo Solomon BJ, Liu G, Felip E, Mok TSK, Soo RA, Mazieres J, Shaw AT, de Marinis F, Goto Y, Wu YL, Kim DW, Martini JF, Messina R, Paolini J, Polli A, Thomaidou D, Toffalorio F, Bauer TM. Lorlatin ib versus crizotinib in patients with advanced ALK-positive non-small cell lung cancer: 5-year outcome s from the phase III CROWN Study. J Clin Oncol. 2024 Oct 10;42(29):3400-3409. doi:10.1200/JCO.24.005 81. Epub 2024 May 31.*

50. a, b, c, d, e, f *Solomon BJ, Bauer TM, Mok TSK, Liu G, Mazieres J, de Marinis F, Goto Y, Kim DW, Wu YL, Jas sem J, López FL, Soo RA, Shaw AT, Polli A, Messina R, Iadeluca L, Toffalorio F, Felip E. Efficacy and safet y of first-line lorlatinib versus crizotinib in patients with advanced, ALK-positive non-small-cell lung c ancer: updated analysis of data from the phase 3, randomised, open-label CROWN study. Lancet Respir Med. 2023 Apr;11(4):354-366. doi:10.1016/S2213-2600(22)00437-4.*

51. a, b, c *Shaw AT, Bauer TM, de Marinis F, Felip E, Goto Y, Liu G, Mazieres J, Kim DW, Mok T, Polli A, Thur m H, Calella AM, Peltz G, Solomon BJ; CROWN trial investigators. first-line lorlatinib or crizotinib in adv anced ALK-positive lung cancer. N Engl J Med. 2020 Nov 19;383(21):2018-2029. doi:10.1056/NEJMoa2 027187.*

52. a, b *Abraira V, Muriel A, Emparanza JI, Pijoan JI, Royuela A, Plana MN, Cano A, Urreta I, Zamora J. Repor ting quality of survival analyses in medical journals still needs improvement. A minimal requirements p roposal. J Clin Epidemiol. 2013 Dec;66(12):1340-6.e5. doi:10.1016/j.jclinepi.2013.06.009.*

53. a, b *Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, Kramar A. Survival end point reporting in randomized cancer clinical trials: a review of major journals. J Clin Oncol. 2008 Aug 1;26(22):3721-6. doi:10.1200/JCO.2007.14.1192.*

54. a, b *Batson S, Greenall G, Hudson P. Review of the reporting of survival analyses within randomised cont rolled trials and the implications for meta-analysis. PLoS One. 2016 May 5;11(5):e0154870. doi:10.1371/j ournal.pone.0154870.*

55. ^ *Prasad V, Bilal U. The role of censoring on progression free survival: oncologist discretion advised. Eur J Cancer. 2015 Nov;51(16):2269-71. doi:10.1016/j.ejca.2015.07.005.*

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.