Title

Sotorasib in KRASG12C mutated lung cancer

Journal

The Lancet, 403(10422)

ISSN

01406736

Authors

Olivier, Timothée
Prasad, Vinay

Publication Date

2024

DOI

10.1016/S0140-6736(23)02035-4

Peer reviewed

European Journal of Cancer 202 (2024) 114022

Contents lists available at ScienceDirect

European Journal of Cancer

journal homepage: www.ejcancer.com

Current perspective

Progression-free survival estimates are shaped by specific censoring rules: Implications for PFS as an endpoint in cancer randomized trials

Vadim Lesan [a], Timothée Olivier [b], Vinay Prasad [c,*]

[a] Hematology and Oncology Department, Saarland University Hospital, Kirrberger Street 100, 66421, Homburg, Germany
[b] Oncology Service, Geneva University Hospital, 4 Gabrielle-Perret-Gentil Street, 1205, Geneva, Switzerland
[c] Department of Epidemiology and Biostatistics, University of California San Francisco, 550 16th St, 2nd Fl, San Francisco, CA 94158, USA

ARTICLE INFO

Keywords:
Mantel cell lymphoma
Censoring
Kaplan-Meier
Progression-free survival
Informative censoring
Ibrutinib
Venetoclax

ABSTRACT

Kaplan-Meier analysis hinges on the assumption that patients who are censored– lost to follow-up, or only recently enrolled on the study– are no different, on average, than patients who are followed. As such, censoring these patients– omitting their future information and taking the average of those who were followed– should not dramatically change the overall estimate. Yet, in a recent clinical trial, two sets of censoring rules– one favored by trialists and one favored by the US Food and Drug Administration– were applied to a progression-free survival (PFS) estimate. In response, the PFS estimate changed dramatically, increasing the median in the experimental arm from 32 to 43 months, while the control arm was essentially unchanged. In this commentary, we explore the reasons why PFS changed so dramatically. We provide a broad overview of censoring in oncology clinical trials, and suggestions to ensure that PFS is a more reliable endpoint.

In cancer medicine, time-to-event outcomes such as progression-free survival (PFS) and overall survival (OS) are frequently used as the primary endpoint of randomized studies. These endpoints are typically analysed with survival methods, such as the Kaplan-Meier (K-M) plot. In this method, some patients are inevitably censored– meaning that beyond some time point, they no longer contribute to the dataset [1,2]. The central assumption of K-M method is that the censored patients are no different, i.e. no more likely to experience the event, than those who are followed.

Censoring can occur because of loss to follow-up, or recent enrolment– i.e. we do not know what happened to the patient beyond some time point– but may also occur due to deviations from the pre-planned protocol. A recent trial in mantle cell lymphoma– the SYMPATICO study– illustrates large variability in PFS estimates between two different censoring strategies [3]. Specific rules for censoring, used by trialists *versus* those favoured by the US Food and Drug administration (FDA), result in a widely different median PFS in the experimental arm alone [3]. Here, we explain how differences in censoring rules can alter estimates of PFS, and draw several broad lessons for investigators regarding progression-free survival.

1. Sympatico study lessons

In the SYMPATICO study, venetoclax was combined with ibrutinib and tested against placebo and ibrutinib among patients with relapsed/refractory mantle cell lymphoma. The primary endpoint was progression-free survival (PFS). In the primary analysis, using rules favoured by the trialists, called 'global censoring rules', patients without progression of the disease or death were censored at the last follow-up visit without the PFS event. A different, and more aggressive censoring method was favoured by the US Food Administration. According to FDA's censoring rules, "patients without progressive disease or death, **with subsequent anticancer therapy, or with two or more missed visits prior** to the PFS event were censored at last follow-up without PFS event **[emphasis ours]**" (Box 1). The FDA's method of censoring includes additional scenarios beyond the global method, and can only result in more patients being censored.

Interestingly, compared to the global censoring rules, the use of US FDA censoring rules result in an additional ten-month gain in median PFS in the experimental arm. Notably, by both sets of rules, control arm median PFS were comparable (22 vs 22 months). Yet, the experimental arm experienced a dramatic improvement (32 *versus* 43 months) in median PFS, when the FDA's additional rules were applied (Table 1). As

* Corresponding author.
E-mail address: vinayak.prasad@ucsf.edu (V. Prasad).

https://doi.org/10.1016/j.ejca.2024.114022
Received 9 February 2024; Received in revised form 9 March 2024; Accepted 12 March 2024
Available online 20 March 2024

such, the US FDA set of censoring rules disproportionately affects the experimental arm, and results in a more favourable estimate. How might this happen?

In order for the FDA's rules to improve the PFS, the additionally censored patients must do worse than those who remain in the data set. We illustrate this in Fig. 1, which simplifies the study and imagines that there are only two patients on the experimental arm. The two patients include one who contributes to the PFS estimate in both censoring scenarios, and one who contributes to the PFS estimate by the global method, but is censored by the FDA method. In other words, the visual simplifies the scenario where one person's data is being treated differently due to different censoring rules.

As shown in Fig. 1, the person whose data is only censored by the FDA is likely experiencing early progression. Omitting this person from the average actually improves the median PFS, as shown in the third bar. This occurs because the only way to improve PFS when you censor more patients is if the censored patients are doing worse than those who remain uncensored.

Astonishingly, these different rules dramatically affect the experimental arm (32 *versus* 43 months) and not the control arm (22 *versus* 22 months). This fact proves that the censoring is *informative*– i.e. the censored patients are not equally likely to experience the event– they are more likely to. How might this exclusively occur on the experimental arm?

Any hypothesis that seeks to explain these results has to do the following: 1) it must postulate some PFS events are only censored by the FDA method and not by the global method; 2) this must apply disproportionately to the experimental arm. We offer two scenarios.

Although the abstract does not report rates of dose reduction and discontinuation, nor rates of hospitalisation while on therapy, we do know that grade $\geq$ 3 adverse events occurred in 84% of patients treated with ibrutinib plus venetoclax and only 76% treated with ibrutinib plus placebo. If the experimental arm (venetoclax plus ibrutinib) is more toxic than the control arm (placebo plus ibrutinib), and if this toxicity results in missed visits or missed scans (because the patient felt unwell or because they were hospitalised) this would disproportionately occur in the experimental arm. Furthermore, the patients who fail to follow-up are likely older, frailer and more likely to progress than those who return. This scenario– toxicity driving extra missed visits in the experimental arm– would result in more censoring, which would improve the PFS estimate if these censored patients were otherwise more likely to progress. We believe this is the most likely explanation.

In the second scenario, toxicity is driving greater rates of subsequent anticancer therapies in the experimental arm prior to progression. This would also result in more censoring; however, it is unclear how these patients would not also be censored, de facto, by 'global rules', as it is standard practice to cease scans in patients when they switch to off-protocol regimens. For this hypothesis to hold, the new therapy should not be censored by 'global censoring rules' but only by 'FDA censoring rules'. We hope the trialist did not allow this to happen.

Ultimately, the large difference in median PFS in the experimental arm and not in the control arm proves that informative censoring occurs with the FDA's rules. But it does not mean that global rules are reliable.

Table 1
PFS per globa and US FDA rules in experimental and placebo arm.

	Venetoclax plus Ibrutinib	Placebo plus Ibrutinib	Difference in PFS, months	HR (95% CI)
PFS per Global censoring, months	31.9	22.1	9.8	0.65 (0.47-0.88)
PFS per US FDA censoring, months	**42.6**	22.1	20.5	0.60 (0.44-0.83)

Global rules likely also have informative censoring as there are surely some patients who don't merely miss two visits. They miss all subsequent visits entirely. These must also occur more often on the experimental arm, and the omission of these patients likely overestimates median PFS. As such all PFS estimates from SYMPATICO trial warrant scepticism.

2. How censoring distorts PFS

A number of prior analyses have examined the correlation between progression-free survival and overall survival. It was shown that PFS is very sensitive to censoring assumptions making it very volatile [2]. For example, the frequency of progression assessment is inversely correlated with PFS [4]. Haslam and colleagues extended these findings and found that surrogate endpoints including PFS have generally a low to moderate correlation with overall survival [5]

A growing body of evidence suggests that PFS may fail to predict OS in part because PFS is more vulnerable to censoring. Tannock and colleagues wrote that for many new approved drugs based on PFS, this rarely translates into improved OS and quality of life [6]. Templeton and colleagues have argued that informative censoring could lead to bias towards a prolonged PFS in the experimental arm [7].

Prasad and Bilal examined the BOLERO-2 trial, which tested the addition of everolimus or placebo to exemestane among hormone receptor positive advanced breast cancer. The trial found a PFS benefit (HR 0.43) but failed to found an OS benefit [8]. In their reanalysis, Prasad and Bilal note high rates of toxicity on the experimental arm, and greater rates of early censoring [2]. If one assumes censored patients are more likely to progress, a reanalysis finds that BOLERO-2 trial may not have a PFS benefit [2]. This points directs to the vulnerability of PFS in face of altered censoring assumptions.

In an empirical analysis led by Rosen, researchers used primary study data from 29 trials to examine early censoring in PFS estimates and found several outlier studies where this occurred disproportionately in one arm or the other [1]. If early censoring is more prevalent in experimental arms, one consideration is that additive toxicity leads to early drop out. If early censoring is prevalent in control arms, a key concern is whether patient disappointment has led to drop out. Notably the QUANTUM-R study had high rates of early discontinuation in the control arm for this reason, and ultimately was denied FDA approval [9].

Box 1
Censoring rules.

- Global rules:
- 1. Patients without progression or death censored at last visit
- FDA rules:
- 1. Patients without progression or death censored at last visit
- 2. Subsequent therapy, but continuing scans
- 3. Missing 2 visits, but then returning for scans

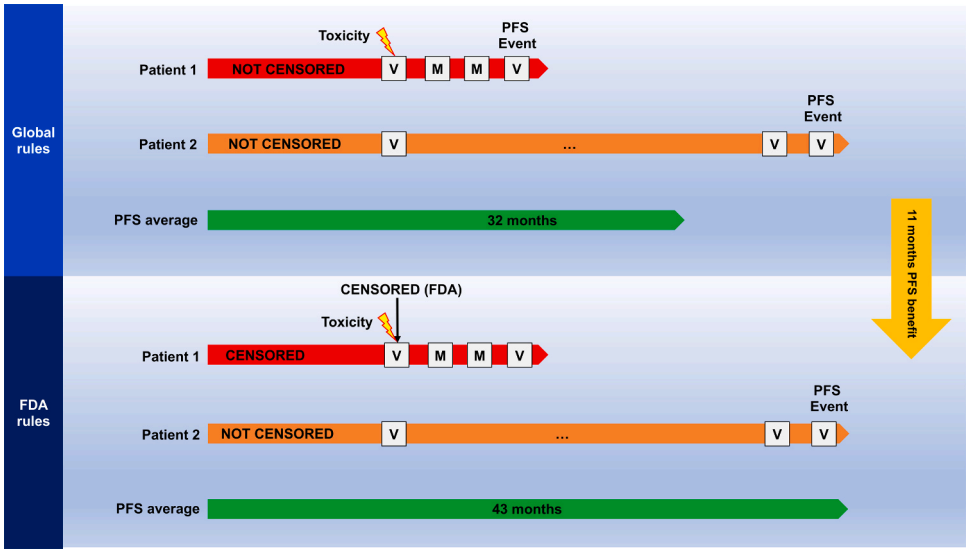

Fig. 1. PFS differences according to censoring rules (global versus FDA). V: visit, M: missed visit.

In both cases, the effect of censoring would be to create a spuriously large benefit– in the former case, by removing the sicker patients from the experimental arm, and in the latter case, removing the patients with the greatest resources to seek out alternatives off-study.

Sometimes, differential censoring can lead to overestimatimation of PFS estimates in the control arm. The CANOVA trial compared venetoclax in combination with dexamethasone (VenDex) to pomalidomide plus dexamethasone (PomDex) in patients with t(11;14)-positive relapsed or refractory multiple myeloma. In this trial, a high proportion of patients in the control arm (PomDex) received subsequent therapy and were censored. This may have occured in patients with suboptimal responses, even without formal progressive disease [10]. Censoring these patients, who were likely destined to progress if they had stayed on similar treatment, may have benefitted the control arm. This could have contributed to the trial's outcome, where no significant improvement in PFS was observed in patients receiving the experimental therapy.

At times, the rate of drop out can be immense. The recent trial VISION has an incredibly high rate of discontinuation on the experimental arm (56%) which was reduced only to 16% with "enhanced trial site education"- a number that still means that the resulting population may be heavily distorted [11,12]. Ultimately, randomization is thwarted with this degree of censoring.

3. We propose several solutions

1. **The number of censored patients in a time interval should be reported in Kaplan-Meier plots.** This practice was adopted by *Lancet*, one of the few medical journals to request reporting of censored patients for each timepoint of the study. Routinely reporting the number of censored patients makes the interpretation of PFS estimates easier. It has also permitted independent analysis of censoring [1]
2. **PFS and OS should be doubted in cases when high rates of drop off after enrollment occur.** The VISION trial and the QUANTUM-R trial illustrate clearly the effect of disproportionate early dropout on both PFS and OS [11,13]. The effect of disproportionate early censoring on PFS or OS can be staggering and should be always accounted for. This bias affects both PFS and OS estimates, as the patients who withdraw may have more or less favourable prognoses.
3. **PFS should be questioned in trials with high rates of discontinuation for toxicity.** High toxicity rates, as in the BOLERO-2 study, should not be ignored, as imbalances in toxicity between the study arms can lead to informative censoring [10]. Reporting the reasons for withdrawal may make the interpretation of PFS estimates more reliable.
4. **PFS should be reported routinely under several different sets of rules or scenarios. This de facto sensitivity analysis will further bolster the credibility of the PFS estimate.** The SYMPATICO trial shows how volatile the PFS can be under different censoring rules [3]. Agencies like US FDA or European Medicine Agency should demand from trialists reporting PFS estimates under different sets of censoring rules before approving drugs based on PFS estimates only. This analysis was performed by the US FDA in the case of sotorasib in the CodeBreak 200 study [14]
5. **Appropriate length and frequency of follow-up.** Assessing PFS (radiologically, biochemically or cytologically) at different timepoints is a surrogate for the true biological PFS. This fact results in data being interval-censored [15]. Interval censoring overestimates the median PFS [5]. In a Phase II study in metastatic colorectal cancer, addition of Bevacizumab to 5-FU resulted in a median PFS of 9.2 months compared to 5.5 months in the 5-FU plus placebo arm [16]. Taking into account the frequency of progression assessment, the true median PFS is between 7.4 and 9.2 months in the Bevacizumab plus 5-FU arm and between 3.7 and 5.5 months in the 5-FU plus placebo arm. Comparing the median PFS as interval data offers a more reliable perspective on the real benefit in the experimental arm.
6. **Generally, and particularly when PFS is not faster, OS should be the primary endpoint of trials.** For instance, in the POLO trial, PFS was used as the endpoint despite the fact that median OS is short in pancreas cancer– it was about 18 months in the POLO trial and the OS results were presented alongside the PFS results [17]. The trialists of POLO could have measured OS instead of PFS without extremely increasing the time needed for the events to happen. Prior work has investigated under what circumstances PFS speeds study results [18]
7. **Primary individual participants data (IPD) should be provided.** Providing patient individual data helps researchers in dissecting the results of a study. Drug approval agencies use patient individual data to ensure that study results are valid. Using individual patient data, US FDA found multiple possible confounders of the PFS estimate in the CodeBreak 200 study [14]. As a result, the FDA did not granted regular approval for sotorasib in KRAS G12C-mutated non-small lung cancer, considering that the PFS in the trial could not be "reliably interpreted" [14]. The proposal of sharing IPD had been championed by the International Committee of Medical Journal Editors, but this was ultimately dropped [19]. We suggest it be reconsidered.

In conclusion, the SYMPATICO study offers an elegant glimpse into the powerful role of censoring and shows the tenuous nature of PFS. Notably, PFS only weakly correlates with OS in trials of mantle cell lymphoma, and is extremely volatile to censoring assumptions [20,21].

PFS should be doubted as a meaningful survival endpoint in relapsed/refractory MCL. Endpoints that inherently include both measures of benefit and toxicity such as overall survival and failure free survival may be preferred in this context.

Funding

This project was funded by Arnold Ventures, LLC through a grant paid to the University of California, San Francisco. The funders had no role in the design and conduct of the study.

CRediT authorship contribution statement

Vinay Prasad: Writing – review & editing, Writing – original draft, Conceptualization. **Timothée Olivier:** Writing – review & editing. **Vadim Lesan:** Writing – review & editing, Writing – original draft.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: ***Vadim Lesan*** received honorari from Abbvie and travel grants from Pfizer, Abbvie, Pierre Fabre, Janssen ***Vinay Prasad*** receives research funding from Arnold Ventures through a grant made to UCSF, and royalties for books and writing from Johns Hopkins Press, MedPage, and the Free Press. He declares consultancy roles with UnitedHealthcare and OptumRX; He hosts the podcasts, *Plenary Session, VPZD, Sensible Medicine*, writes the newsletters, *Sensible Medicine, the Drug Development Letter and VP's Observations and Thoughts*, and runs the YouTube channel *Vinay Prasad MD MPH*, which collectively earn revenue on the platforms: Patreon, YouTube and Substack. ***Timothée Olivier*** declares no conflict of interest.

References

[1] Rosen K, Prasad V, Chen EY. Censored patients in Kaplan-Meier plots of cancer drugs: an empirical analysis of data sharing. Eur J Cancer [Internet] 2020;141:152–61 [cited 2024 Jan 12], ⟨https://pubmed.ncbi.nlm.nih.gov/33160265/⟩.

[2] Prasad V, Bilal U. The role of censoring on progression free survival: oncologist discretion advised. Eur J Cancer [Internet] 2015;51(16):2269–71. Available from: ⟨https://pubmed.ncbi.nlm.nih.gov/26259493/⟩.

[3] Wang M., Jurczak W., Trněný M., Belada D., Wrobel T., Ghosh N., et al. Ibrutinib Combined with Venetoclax in Patients with Relapsed/Refractory Mantle Cell Lymphoma: Primary Analysis Results from the Randomized Phase 3 Sympatico Study. In [cited 2024 Jan 8]. Available from: https://ash.confex.com/ash/2023/webprogram/Paper191921.html.

[4] Haslam A, Gill J, Prasad V. The frequency of assessment of progression in randomized oncology clinical trials. Cancer Rep (Hoboken) [Internet] 2022;5(7) (Available from), ⟨https://pubmed.ncbi.nlm.nih.gov/34821077/⟩.

[5] Haslam A, Hey SP, Gill J, Prasad V. A systematic review of trial-level meta-analyses measuring the strength of association between surrogate end-points and overall survival in oncology. Eur J Cancer 2019;106:196–211.

[6] Tannock IF, Pond GR, Booth CM. Biased evaluation in cancer drug trials—how use of progression-free survival as the primary end point can mislead. JAMA Oncol [Internet] 2022;8(5):679–80. Available from: ⟨https://jamanetwork.com/journals/jamaoncology/fullarticle/2790095⟩.

[7] Templeton AJ, Amir E, Tannock IF. Informative censoring — a neglected cause of bias in oncology trials. Nat Rev Clin Oncol 2020;17:6. ⟨https://www.nature.com/articles/s41571-020-0368-0⟩.

[8] Baselga J, Campone M, Piccart M, Burris HAI, Rugo HS, Sahmoud T, et al. Everolimus in Postmenopausal Hormone-Receptor–Positive Advanced Breast Cancer. https://doi.org/10.1056/NEJMoa1109653 [Internet]. 2012 Feb 9 [cited 2024 Jan 16];366(6):520–9. Available from: https://www.nejm.org/doi/full/10.1056/nejmoa1109653.

[9] FDA Panel Votes Against Quizartinib for AML [Internet]. [cited 2024 Jan 16]. Available from: https://www.onclive.com/view/fda-panel-votes-against-quizartinib-approval-for-aml.

[10] Study Details | A Study Designed to Evaluate the Safety and Efficacy of Venetoclax Plus Dexamethasone (VenDex) Compared With Pomalidomide Plus Dexamethasone (PomDex) in Participants With t(11;14)-Positive Relapsed or Refractory Multiple Myeloma. | ClinicalTrials.gov [Internet]. [cited 2024 Mar 9]. Available from: https://www.clinicaltrials.gov/study/NCT03539744?cond=NCT03539744&rank=1.

[11] Sartor O, de Bono J, Chi KN, Fizazi K, Herrmann K, Rahbar K, et al. Lutetium-177–PSMA-617 for metastatic castration-resistant prostate cancer (Available from: https://www.nejm.org/doi/full/) N Engl J Med [Internet] 2021 Sep 16;385(12):1091–103. https://doi.org/10.1056/nejmoa2107322.

[12] Olivier T, Powell K, Prasad V. Lutetium-177-PSMA-617 in metastatic castration-resistant prostate cancer: limitations of the VISION trial. Eur Urol. 2022 9 sept 2022;S0302-2838(22)02613-6.

[13] Cortes JE, Khaled S, Martinelli G, Perl AE, Ganguly S, Russell N, et al. Quizartinib versus salvage chemotherapy in relapsed or refractory FLT3-ITD acute myeloid leukaemia (QuANTUM-R): a multicentre, randomised, controlled, open-label, phase 3 trial. Lancet Oncol [Internet] 2019 Jul 1;20(7):984–97. Available from: ⟨http://www.thelancet.com/article/S1470204519301500/fulltext⟩.

[14] FDA's ODAC Votes No for Sotorasib in KRAS G12C-Mutated NSCLC [Internet]. [cited 2024 Jan 16]. Available from: ⟨https://www.targetedonc.com/view/fda-s-odac-votes-no-for-sotorasib-in-kras-g12c-mutated-nsclc⟩.

[15] Panageas KS, Ben-Porat L, Dickler MN, Chapman PB, Schrag D. When you look matters: the effect of assessment schedule on progression-free survival (Available from:) JNCI: J Natl Cancer Inst [Internet] 2007;99(6):428–32. https://doi.org/10.1093/jnci/djk091.

[16] Kabbinavar FF, Schulz J, McCleod M, Patel T, Hamm JT, Hecht JR, et al. Addition of bevacizumab to bolus fluorouracil and leucovorin in first-line metastatic colorectal cancer: results of a randomized phase II trial. J Clin Oncol [Internet] 2005;23(16):3697–705 (Available from), ⟨https://pubmed.ncbi.nlm.nih.gov/15738537/⟩.

[17] Golan T, Hammel P, Reni M, Van Cutsem E, Macarulla T, Hall MJ, et al. Maintenance Olaparib for Germline BRCA - mutated metastatic pancreatic cancer (Available from) N Engl J Med [Internet] 2019 Jul 25;381(4):317–27. https://doi.org/10.1056/nejmoa1903387 (Available from), ⟨https://www.nejm.org/doi/full/⟩.

[18] Chen EY, Joshi SK, Tran A, Prasad V. Estimation of study time reduction using surrogate end points rather than overall survival in oncology clinical trials. JAMA Intern Med [Internet] 2019;179(5):642–7. Available from: ⟨https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2729389⟩.

[19] Longo D.L., Drazen J.M. Data Sharing. https://doi.org/101056/NEJMe1516564 [Internet]. 2016 Jan 21 [cited 2024 Jan 18];374(3):276–7. Available from: https://www.nejm.org/doi/full/10.1056/NEJMe1516564.

[20] Korn RL, Crowley JJ. Overview: progression-free survival as an endpoint in clinical trials with solid tumors. Clin Cancer Res [Internet] 2013;19(10):2607 [cited 2024 Jan 9].

[21] Jen MH, Sonksen M, Hess L, Bian F. Relationship between Overall Response Rate (ORR), Progression-Free Survival (PFS), and Overall Survival (OS) in Clinical Trials of Patients with Mantle Cell Lymphoma (MCL). Value Health [Internet] 2023;26(6):S44 (Available from), ⟨http://www.valueinhealthjournal.com/article/S109830152302394X/fulltext⟩.