Investigating discrepancies in assessments of PFS between study investigators and independent reviewers

Clare Jones, Giles Monnickendam, and Juan Fernando Soto Barrientos. PRMA Consulting Ltd, Fleet, Hampshire, UK

Introduction

Overall survival (OS) is typically the preferred trial endpoint for measuring patient benefit in oncology trials but can require long follow-up. Furthermore, the potential impact of subsequent therapies beyond the experimental treatment phase can make interpretation of differences in OS challenging. Progression-free survival (PFS) requires shorter follow-up, can avoid confounding from subsequent therapy, and is increasingly used as an alternative endpoint to OS. PFS can be assessed by local investigators (INV) or by blinded, independent, centralized review (ICR). The most commonly used criteria for evaluation are subject to measurement error and also usually include subjective factors. There is a risk of systematic bias in assessment with INV (in favor of the experimental treatment), particularly with open-label trials.[1,2] ICR has been recommended as an alternative measure of PFS that is not subject to assessment bias. However, ICR may itself be biased (against the experimental treatment) due to informative censoring.[3]

Past studies of the concordance between INV- and ICR-assessed PFS across clinical studies concluded that, while there is variability between individual INV and ICR results, overall there is strong concordance and no evidence of systematic bias.[4–6] However, more recent individual clinical studies have reported large differences between INV- and ICR-assessed PFS; for example, the SOLO-2 study, which compared olaparib versus placebo as maintenance treatment for platinum-sensitive ovarian cancer, reported an INV-assessed PFS benefit (19.1 months versus 5.5 months, HR = 0.30) substantially below the ICR result (30.2 months versus 5.5 months, HR = 0.25).[7]

We reviewed recent oncology trials to establish the extent of differences between INV- and ICR-assessed PFS.

Objectives

To increase understanding of potential variability in PFS measurement, relevant associations, and possible causal factors to inform appropriate use of PFS in payer and clinician decision-making.

Methods

- We searched ClinicalTrials.gov for "progression free survival" and "cancer", filtering for interventional Phase 2 or 3 studies with results. The studies identified were exported into Excel and records searched for the terms "independent centralized review", "investigator assessed", and similar terms.
- We searched PubMed using the same criteria; full articles were reviewed and studies reporting ICR- and INV-assessed PFS were included.
- For comparative trials with only two treatment groups, we calculated differences in median PFS between experimental and control groups for ICR- and INV-assessed PFS.
- For comparative trials with more than two treatment groups, we calculated the difference in median PFS between each group reported by INV and ICR.
- We calculated the ratio of the ICR and INV hazard ratios (ICR divided by INV) for comparisons where data were available.
- For non-comparative trials, we calculated the absolute difference between ICR- and INV-assessed median PFS.

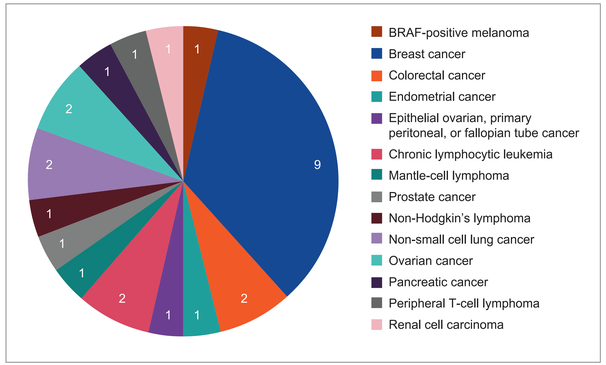
Results

Searches and study characteristics

A total of 365 Phase 2 or 3 interventional studies on ClinicalTrials.gov specified PFS as an endpoint and had reported results. Of these, 48 specified that PFS was assessed by ICR, 45 by INV, and 6 by both. The search of PubMed identified 49 studies, 21 of which reported both INV and ICR PFS values. There was one duplicate between the two databases, so the analysis set comprised 26 studies.

The majority of the study reports were for solid tumors: breast cancer was the most common cancer type (35%, 9), followed by colorectal cancer (8%, 2), CLL (8%, 2), NSCLC (8%, 2), and ovarian cancer (8%, 2). Four hematological malignancies were included across 5 studies. Twenty of the studies (77%) had 2 groups, 1 (4%) had 3 groups, 1 (4%) had 4 groups, and 4 (15%) were non-comparative trials. Nineteen of the studies were Phase 3 (73%), seventeen were open-label (65%), and most were published in the past 5 years (81%, 21).

Figure 1: Cancer type studied in the selected trial reports



Measurement of PFS as an endpoint

In studies of solid tumors, the most commonly used criteria for measuring disease progression were the Response Evaluation Criteria in Solid Tumors (RECIST; various versions were used, e.g., 1.0, 1.0 with modifications, 1.1). It was not always possible to determine whether both INV and ICR used the same RECIST version. In some cases, the criteria were identified for only one assessor, usually the INV, leaving it unclear whether the ICR used the same criteria.

In studies where PFS was the primary endpoint (58%, 15) or co-primary endpoint (31%, 8) it was usually assessed by INV; PFS assessed by INV was the primary endpoint in 11 (42%) of the studies and PFS by ICR in 4 (15%). In only 1 study were INV- and ICR-assessed PFS co-primary endpoints.

Figure 2: Role of INV and ICR endpoints

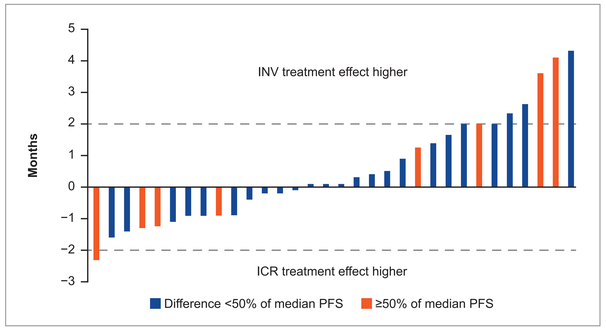
		ICR				
		Primary	Co-primary	Secondary	Other/not specified	**Total**
INV	Primary	–	–	1	10	**11**
	Co-primary	–	1	2	2	**5**
	Secondary	1	2	1	3	**7**
	Other/not specified	3	0	0	0	**3**
	Total	**4**	**3**	**4**	**15**	**26**

Comparison of INV and ICR measurements of PFS

From the 26 studies identified, 32 comparisons were possible. Calculating the PFS gain at the median, the difference between the ICR-assessed gain and the INV-assessed gain ranged from −2.3 months (ICR PFS >INV PFS) to 4.3 months (INV PFS >ICR PFS). In one-quarter of the comparisons the percentage difference in median PFS between the assessors was at least 50%, with the greatest percentage difference being -446% in a study where the ICR-determined median PFS was 1.25 months.

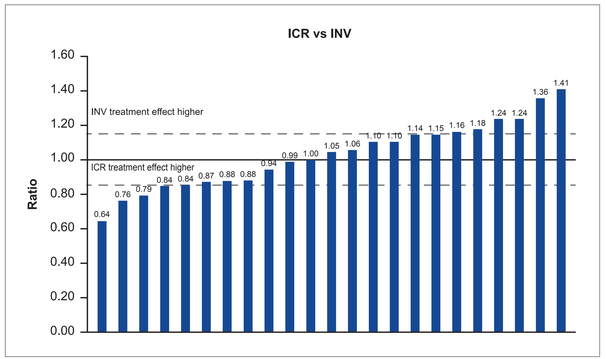
The INV-assessed PFS gain was greater than the ICR-assessed gain in 18 of 32 comparisons (56%).

Figure 3: Difference between INV- and ICR-assessed median PFS



Previous studies by Zhang et *al.*, and Amit et *al.*, used the ratio of hazard ratios to quantify differences and identified a threshold of 15% difference as high variation. In the current study the ratio of INV and ICR hazard ratios vary from 0.64 to 1.41, with 48% (11/23) of trial comparisons resulting in a difference in HRs >15%.

Figure 4: Ratio of hazard ratios (ICR HR/INV HR)



Conclusions

- It is uncommon for studies to report both ICR- and INV-assessed PFS; it is therefore difficult to determine how often results vary and the extent of the variation.
- In most cases, the median difference between INV and ICR assessments are small in absolute magnitude but a substantial proportion of the studies reviewed show large differences (>2 months).
- This work includes more recent trials than previous studies and, based on hazard ratios, found substantial differences in PFS in a greater proportion of comparisons.
- These differences may not lead directly to different conclusions in the registration process, where binary decisions are based on whether measured PFS benefit exceeds a threshold level for statistical significance. However, for other decision makers, where the magnitude of PFS benefit is directly important e.g., in determining the relative value of a new therapy, it is important to understand the variability in assessments of PFS.
- Both INV and ICR measures should be routinely reported in oncology trials, to improve consistency of comparison across trials, indicate where there is uncertainty over the magnitude of treatment effect, and provide data for further investigation of potential sources of variability in PFS measurement.

References

1. US Food and Drug Administration. Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. Washington, DC, 2007, pp. 1–19
2. European Medicines Agency: Methodological consideration for using progression-free survival or disease-free survival in confirmatory trials. December 2012
3. Dodd LE, Korn EL, Freidlin B, et al., Blinded independent central review of progression-free survival in Phase 3 clinical trials: important design element or unnecessary expense? *J Clin Oncol* 2008 26:3791–6
4. Tang PA, Pond GR, Chen EX. Influence of an independent review committee on assessment of response rate and progression-free survival in Phase 3 clinical trials. Ann Oncol 2010 21:19–26
5. Amit O, Mannino F, Stone AM, et al., Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis. *Eur J Cancer* 2011 47:1772–8
6. Zhang JJ, Chen H, He K, et al., Evaluation of blinded independent central review of tumor progression in oncology clinical trials: a meta-analysis. *Drug Inf J* 2013 47:167–74
7. Pujade-Lauraine E, Ledermann JA, Selle F, et al., Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a BRCA1/2 mutation (SOLO2/ENGOT-Ov21): a double-blind, randomised, placebo-controlled, Phase 3 trial. Lancet Oncol 2017