EUROPEAN JOURNAL OF CANCER 47 (2011) 1772–1778

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Blinded independent central review of progression in cancer clinical trials: Results from a meta-analysis ☆

O. Amit [a,*], *F. Mannino* [a], *A.M. Stone* [b], *W. Bushnell* [a], *J. Denne* [c], *J. Helterbrand* [d], *H.U. Burger* [e]

[a] *GSK, Collegeville, PA, USA*
[b] *AstraZeneca, Macclesfield, UK*
[c] *Eli-Lilly, Indianapolis, IN, USA*
[d] *Genentech, South San Francisco, CA, USA*
[e] *Hoffman LaRoche, Basel, Switzerland*

ARTICLE INFO

Article history:
Received 29 October 2010
Accepted 16 February 2011
Available online 21 March 2011

Keywords:
Progression-free survival
Independent review
Meta-analysis
Discordance

ABSTRACT

Purpose: Progression free survival (PFS) is increasingly used as a primary end-point in oncology clinical trials. This paper provides observations and recommendations on the use of a blinded independent central review (BICR) for progression.

Patients and methods: The findings and recommendations are based on extensive simulations and a meta-analysis based on 27 previously conducted randomised phase III trials with BICR performed by the Pharmaceutical Research and Manufacturers Association (PhRMA) sponsored PFS Independent Review Working Group.

Results: Results of the meta-analysis demonstrate a strong correlation between LE and BICR estimates of treatment effect ($R = 0.947$). Further, differences between treatment groups in discordance rates predict the differences between LE and BICR estimates of treatment effect supporting the use of a sample-based BICR on a subgroup of patients.

Conclusion: The meta-analysis demonstrates that local evaluation (LE) provides a reliable estimate of the treatment effect with little evidence for systematic evaluation bias. Therefore, when a trial is blinded or a large effect on PFS is observed a BICR may not be warranted. When a BICR is warranted, a sample-based BICR may provide a reduction in operational complexity without compromising the credibility of trial results. While for large trials that are not adequately blinded a sample-based BICR may be recommended. A full BICR should be considered in the case of smaller trials or in situations in which there is a particular need to increase the confidence in the LE results.

1. Introduction

Progression-free survival (PFS) is increasingly becoming a key end-point for the approval of new therapeutics in oncology. PFS is typically defined as the time from randomisation or treatment initiation until the first of disease progression or death where progression is assessed by some objective criteria. The use of PFS as a primary end-point requires detailed

☆ Data presented at DIA/FDA/NCI/PhRMA Progression Free Survival Oncology Workshop, October 8–9th 2009, Washington, DC.

* *Corresponding author*: Address: GlaxoSmithKline 1250 S. Collegeville Road, Collegeville, PA 19426, USA. Tel.: +1 610 917 6893; fax: +1 610 917 4701.
E-mail address: ohad.amit@gsk.com (O. Amit).

0959-8049/$ - see front matter
doi:10.1016/j.ejca.2011.02.013

and careful methodologic consideration.[1] The focus of this paper is on methods to identify and measure the impact of evaluation bias.

Recently, many phase III trials with PFS as a primary end-point have employed blinded independent central review (BICR) of radiologic images. This typically involves a review of all scans for all patients to independently evaluate both the presence and timing of progression.

It has been argued that central review provides a mechanism for reducing bias in the evaluation of progression.[2] A recently published literature review to assess differences between BICR and local evaluation (LE) for both tumour response rate (RR) and PFS[3] concluded that there is significant variability between the LE and BICR results and, therefore. that all phase III trials with a PFS end-point should be subjected to a BICR.[3] With respect to PFS, the review focused on differences in the medians and, therefore, did not directly examine concordance of treatment effect, as measured by the hazard ratio (HR) for BICR and LE.[3] Based on a review of BICR for PFS, by other authors, which covered results from seven randomised clinical trials there appeared to be no additional information added by performing a BICR and it was concluded that the implementation of BICR for trials with a PFS end-point should not be universally recommended.[4] Further, BICR introduces new potential biases in estimating treatment effects due to handling of patients who progress per the LE but not by BICR.[4] BICR should, therefore, be viewed as an 'audit' tool to confirm the results of the LE.[4]

The goals of the PHRMA PFS independent review working group were to (1) quantify the relationship between the estimated treatment effect based on LE versus BICR and (2) to understand the impact of discordance between the LE and BICR on the estimation of treatment effects and subsequently to make recommendations on the use of BICR in phase III trials. The findings are summarised in this paper.

When considering evaluation of PFS it is critical to differentiate between measurement error and bias. We define bias as any systematic process that can lead to an expected estimate of the treatment effect that does not reflect the true effect. In this paper we focus on evaluation bias which derives from the subjective aspects of PFS evaluation.[1] As an example, in an open-label trial progression could be systematically declared earlier in one treatment arm if in the opinion of the investigator, that arm is perceived to be inferior to the other arm(s) in the trial.

From the trial design and importantly from a drug development perspective the impact of bias is far more troublesome than the impact of measurement error. When systematic bias is present, depending on the direction of the bias, the chance of a false positive finding (type I error) from a trial can be inflated and the potential for overestimating the treatment benefit introduced. The impact of evaluation bias on the estimates of treatment effect is discussed more fully in subsequent sections of this paper. The companion paper addresses the separate issue of measurement error and demonstrates that it can attenuate the observed treatment effect, therefore, decreasing the power of the trial.[5]

When PFS is subjected to BICR, there may be discordance with the LE. Discordance can be defined in several ways but fundamentally it represents a disagreement between LE and BICR with respect to either the progression status of the patient or the timing at which progression occurs. A key controversy is whether discordance represents systematic bias/unreliability in the estimate of the treatment effect, or whether such discordance is simply a reflection of the variability inherent in the process.

In this paper we show that in previously conducted trials observed discordance is largely a reflection of variability in the process and not systematic bias. We further develop the concept of discordance by evaluating the utility of traditional measures of discordance for identifying evaluation bias versus some new proposed measures. We evaluate the impact of discordance on the key goal of any clinical trial: the reliable estimation of treatment effect. We introduce the concept of differential discordance, defined as the difference between treatment arms in discordance rates, as the key framework for evaluating bias in the LE. Results are provided from a meta-analysis of 27 trials which employed BICR and LE of PFS and a more detailed evaluation of discordance from 12 of those trials. Finally we propose a framework for performing a sample-based BICR as an audit to confirm the reliability of LE.

2. Patients and methods

The meta-analysis was performed using summary information based on results of 27 randomised phase III trials in solid tumours that employed prospectively or retrospectively defined independent review of scans for progression. The trials identified in a recently published literature search were used and expanded upon to capture additional data available in the public domain.[3] The expanded search included a review of key FDA regulatory meetings, ASCO presentations and key oncology publications. The correlation of LE and BICR HRs was evaluated using a Pearson inter-class correlation coefficient. A zero intercept linear regression model was also fit after testing for the significance of the intercept term. To quantify the magnitude of agreement in the estimated treatment effect the HR ratio (HRR) was calculated as described in a recently published meta-analysis of cardiovascular outcomes.[6] The HRR was defined as the ratio of the BICR HR to LE HR. The overall HRR and associated 95% confidence interval across the 27 trials were estimated using a random effects model weighted by the study size with log HR as the dependent variable and trial as a random effect.

Additional detailed data on discordance rates were provided directly from the sponsor for 12 of the 27 trials. Based on their sensitivity for detecting evaluation bias, two measures, the early discrepancy rate and late discrepancy rate as defined below, were chosen and differential discordance rates were calculated. The agreement between BICR and LE within a treatment arm is represented in a tabular form below (Table 1.)

The early discrepancy rate (EDR) is defined as:

$$\mathrm{EDR} = \frac{b + a3}{a + b}$$

The EDR represents the positive predictive value of investigator assessment and quantifies the frequency with which the LE declares progression early relative to BICR within each

Table 1 – BICR versus LE disease progression assessments.

	BICR	
	PD	No PD
Investigator PD	$a = a1 + a2 + a3$	b
No PD	c	d

Note: In practice a LE PD occurring later than a BICR PD ($a3$) would be observed rarely.
$a1$: number of agreements on timing and occurrence of PD.
$a2$: number of times LE declares PD later than BICR.
$a3$: number of times LE declares PD earlier than BICR.

arm as a proportion of the total number of investigator assessed PD's.

The late discrepancy rate (LDR) is defined as

$$\text{LDR} = \frac{c + a2}{b + c + a2 + a3}$$

The LDR quantifies the frequency that LE declares progression later than BICR as a proportion of the total number of discrepancies within the arm. If the distribution of discrepancies is similar between the arms then this suggests the absence of evaluation bias favoring a particular arm.

The EDR and LDR can be calculated for each treatment arm and the differential discordance around each measure can be defined as the rate on the experimental arm minus the rate on the control arm. A negative differential discordance for the EDR and/or positive differential discordance for the LDR are suggestive of a bias in the LE favoring the experimental arm.

The methodology for a sample-based BICR was developed based on empirical methods using simulation and results from the meta-analysis. Details of the simulation parameters and simulation results are provided in Appendix A.

3. Results

A summary of the characteristics for the trials included in the meta-analyses is presented in Table 2. Additional details on each study are provided in Appendix B, Table B.1.[8–33]

The degree of association between LE and BICR estimates of the treatment effect is high ($R = 0.947$ (95% Confidence interval, CI: 0.88, 0.97), Fig. 1); as demonstrated in the figure, in the majority of cases the LE HR and the BICR HR are highly concordant. The estimated HR ratio between BICR and LE is 1.03 (95% CI: 0.98–1.08) indicating on average a 3% difference between the two evaluations. Seventeen (63%) of the trials had an HR ratio between 0.9 and 1.1 (<10% difference between BICR and LE), while only 2 studies (7%) had an HR ratio indicating >15% difference between BICR and LE.

The findings of the meta-analysis were supported by simulation. The simulation results (Appendix A, Tables A.1 and A.2) demonstrate that when no evaluation bias is present, HR estimates are nearly identical even with a marked level of discordance present within each arm. The simulation results in Appendix A further demonstrate that a marked level of differential discordance, resulting from substantial evaluation bias, must be present in order to significantly bias the estimation of the treatment effect.

Table 2 – Meta analysis study characteristics.

	Full meta-analysis ($N = 27$)	Meta-analysis subgroup ($N = 12$)[b]
Tumour type		
Colorectal	6	5
Breast	9	3
Renal	6	1
Other[a]	6	3
Study size		
Median	579	488.5
Min, Max	200, 1286	200, 1286
Study characteristics		
Blinded	12	3
RECIST criteria	21	7

[a] Includes trials in non-small cell lung cancer, prostate cancer, melanoma, mesothelioma, Glioblastoma and ovarian cancer.
[b] Trial data contributed from GlaxoSmithKline, Eli Lilly, Hoffman La Roche, AstraZeneca.

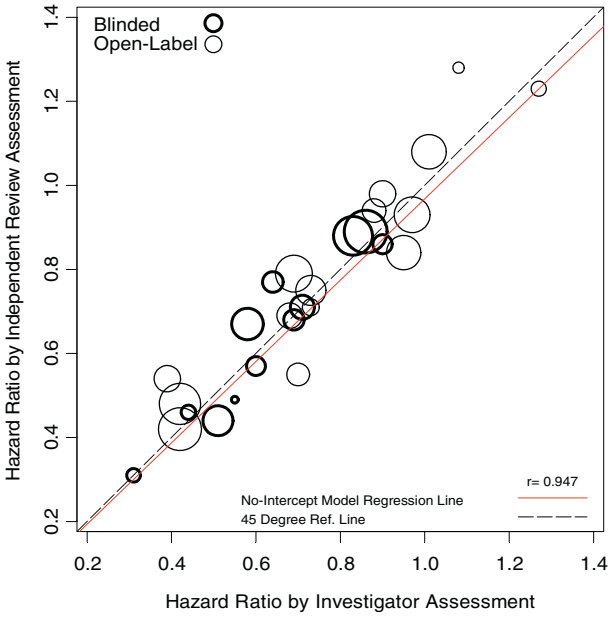

Fig. 1 – BICR versus LE HR by blinding status of trial.

3.1. *Discordance as a measure of evaluation bias in a sample based BICR*

The typical metrics for quantifying discordance are confounded by efficacy as they divide the number of patients who have a disagreement between the BICR and LE by the number of patients randomised. Even in the absence of evaluation bias, the less efficacious arm will have a greater opportunity to have disagreements by virtue of having more events (Appendix A, Tables A.1 and A.2). Further, the degree of differential discordance also depends on the maturity of the data (Appendix A, Fig. A.1).

The LDR and EDR by contrast, only show differential discordance in the presence of evaluation bias independent of data maturity (Appendix A, Tables A.1 and A.2 and Fig. A.1).

This finding is further supported by the meta-analysis (Fig. 2) where the difference in LDR was associated with the extent of disagreement in the HR as measured by the HRR ($R = 0.582$, 95% CI: 0.015, 0.086). A similar correlation was observed for differential discordance as measured by the EDR and the HRR (Fig. 3, $R = -0.601$, 95% CI: –0.867, 0.013). The association observed in Figs. 2 and 3 for the EDR and LDR has been confirmed through simulation (Appendix A, Fig. A.2). Across the 12 trials in the analysis, disagreement on timing or occurrence of progression was observed on average in 51% and

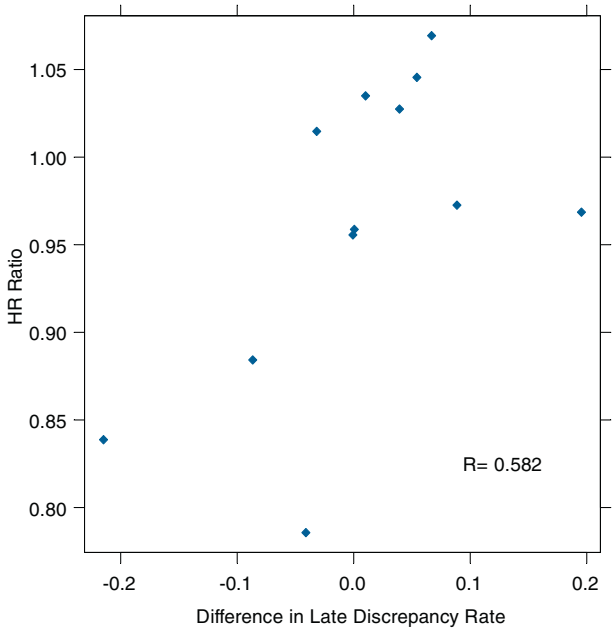

Fig. 2 – Difference in LDR versus ratio of hazard ratios (HR ratio).

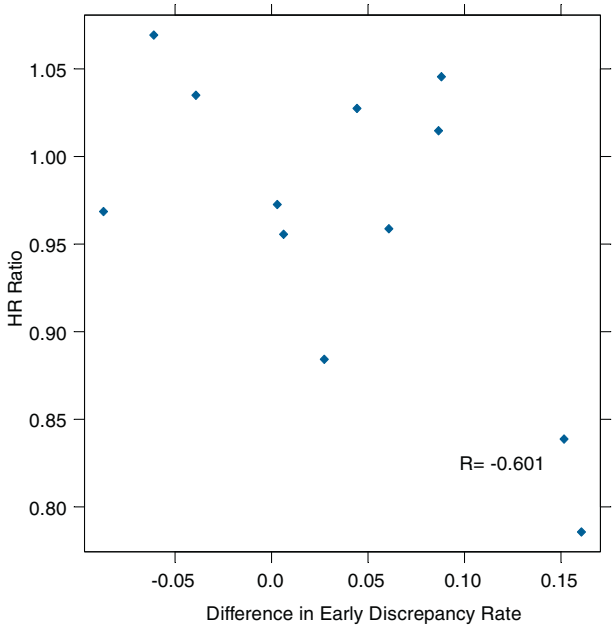

Fig. 3 – Difference in EDR versus ratio of hazard ratios (HR ratio).

Table 3 – Summary of discordance measures.

Discordance measure	Control ($N = 12$)	Experimental ($N = 12$)
Proportion disagreement on censoring status		
Mean (SD)	0.36 (0.10)	0.37 (0.10)
Median	0.37	0.37
Min, Max	0.17, 0.54	0.14, 0.48
Proportion disagreement on timing or occurrence of PD		
Mean (SD)	0.51 (0.13)	0.52 (0.13)
Median	0.51	0.55
Min, Max	0.27, 0.71	0.23, 0.67
EDR		
Mean (SD)	0.41 (0.13)	0.45 (0.12)
Median	0.43	0.49
Min, Max	0.18, 0.63	0.13, 0.56
LDR		
Mean (SD)	0.33 (0.17)	0.33 (0.13)
Median	0.30	0.28
Min, Max	0.15, 0.77	0.21, 0.56

52% of subjects and BICR disagreed with LE progression on average in 38% and 41% of subjects in the control and experimental arms respectively (Table 3). Further, discordance rates were similar between the arms across all measures with an average differential discordance close to 0. Both the LDR and EDR have utility as tools to screen for the presence of evaluation bias in the LE.

Based on the findings from the meta-analysis and statistical simulations, we developed a procedure for a sample-based BICR. The goal of the procedure is to detect potential evaluation bias in the LE of PD by performing a BICR in a random sample of patients. This is achieved through an analysis of differential discordance. The sample-based BICR is designed to detect evaluation bias based on a large underlying differential discordance rate of 0.15 or greater using threshold values ranging from 0.075 to 0.1. Differential discordance rates of 0.15 predict for relative differences in the HR between the LE and BICR evaluations of 20–30% (Appendix A, Fig. A.2). To detect this level of differential discordance with adequate sensitivity and specificity sample-based BICRs of between 100 and 160 subjects in total are recommended. A sample-based BICR of this size provides sensitivity of 80% to detect a difference in EDR and LDR rates when the true difference between arms is 0.15 or greater and a similar specificity (Table 4). With the LDR the number of subjects needed may vary depending on the number of discrepancies in timing or occurrence of progression that are observed. Based on a review of discordance data from the 12 trials on average such discrepancies occurred in 50% of subjects with the rates exceeding 40% in all but one of the trials.

4. Discussion and conclusion

PFS is an end-point that is increasingly used for the registration of anti-cancer therapies. The increased use of this end-point in phase III settings has resulted in the use of BICR as

Table 4 – Sensitivity and specificity for the EDR and LDR

Discordance measure	Threshold value[c]	Discrepancy rate control/experimental	Sensitivity[d] (%)	Specificity[e] (%)
Early discrepancy rate ($N = 80$)[a]	$\Delta \leqslant -0.100$	0.40/0.20	81	87
		0.45/0.25	83	85
		0.50/0.30	83	84
	$\Delta \leqslant -0.075$	0.35/0.20	78	80
		0.40/0.25	77	78
		0.45/0.30	76	77
Late discrepancy rate ($N = 160$)[b]	$\Delta \geqslant 0.100$	0.40/0.60	82	82
		0.45/0.65	82	82
		0.50/0.70	82	81
	$\Delta \geqslant 0.075$	0.40/0.55	75	75
		0.45/0.60	75	75
		0.50/0.65	75	75

Sensitivity and specificity calculated assuming binomial distributions for the discordance rates in each arm.

[a] Assumes 80 PD events are sampled. Specificity is calculated assuming experimental discrepancy rate in the control arm.

[b] One hundred and sixty subjects sampled to attain 80 subjects with discordance in timing or occurrence of PD. Specificity is calculated assuming control discrepancy rate in the experimental arm.

[c] Threshold value represents the differential discordance value at which evidence for evaluation bias is concluded.

[d] *Sensitivity*: Proportion of time threshold is exceeded assuming differential discordance in column.

[e] *Specifity*: Proportion of time threshold is not exceeded assuming equal discrepancy rates in the two arms.

a common feature of trial design and analysis. The use of BICR has, not unexpectedly, resulted in disagreements for individual patients between BICR and LE with respect to both timing and occurrence of progression. In general discordance between the LE and BICR can be a function of measurement error, which is known to be present in the process, or systematic bias in the LE.

Given the measurable level of disagreement observed between two blinded reviewers[7] and our own results, there is now strong evidence to believe that discordance is largely reflective of the variability inherent in the process and not systematic bias. These disagreements have been magnified by the fact that follow-up disease assessments are typically not available once progression is declared per LE and, therefore, the progression time for the BICR is often not known. This feature of the design leads to informative censoring in the BICR analysis of PFS. If evaluation bias is present, neither the LE nor BICR evaluation will provide an unbiased estimate of the treatment effect.[4] This result has been confirmed through simulations (Appendix A, Tables A.1 and A.2). The LE would be directly impacted by evaluation bias while the BICR evaluation would be impacted by different rates of informative censoring between the arms.

The meta-analysis presented, which included both blinded and unblinded trials in several solid tumours, demonstrated that HR estimates based on the LE and BICR are highly concordant. This finding exists in the presence of a marked degree of individual patient discordance in the timing or occurrence of progression. This suggests that in most cases the LE provides a reliable estimate of treatment effect. The results covered a broad range of outcomes and tumours to allow for meaningful conclusions regarding the reliability of the LE. While we did not examine differences between prospective and retrospective BICRs both were represented in the meta-analysis. These potential differences merit further investigation. Our proposed sample-based BICR is intended to be prospectively defined as part of the study design.

Two metrics have been proposed that are sensitive measures of evaluation bias. It has been shown both through clinical trial results and statistical simulation that a large magnitude of differential discordance can be associated with divergent estimates of treatment effect between the LE and BICR. While some guidance is provided regarding a threshold value of differential discordance, the ultimate decision as to whether to proceed to a full BICR should be based on a weight of evidence including careful examination of the point estimates and confidence intervals for differential discordance across both the EDR and LDR. Based on the simulation results and the sensitivity and the specificity of the proposed procedure, we believe that observed differential discordance values of 0.075 or greater are suggestive of the possibility of evaluation bias.

The use of BICR increases trial complexity and cost and is an added burden to both investigators and sponsors. Based on the observations from our meta-analysis and simulations, the routine use of BICR in all randomised trials evaluating PFS as an end-point is not recommended. In cases where BICR is utilised we believe the intent should not be to re-estimate the treatment effect. Therefore, we have proposed the use of a sample-based BICR. There are several situations in which a sample-based BICR might be useful. In non-blinded trials for example, a sample-based BICR in lieu of a 100% BICR would be desirable, particularly where the anticipated effect size on PFS is unknown. Alternatively, in situations where a large effect on PFS is observed or anticipated or where the trial is blinded, neither a sample-based or full BICR may be necessary. However, due to the toxicity profiles of anti-cancer agents true blinding may be difficult to achieve and, therefore, careful thought should be given to whether or not to perform a BICR. There are also situations where a full BICR may be

warranted. In smaller trials ($N < 300$ patients) a sample-based BICR may not be feasible or offer limited savings in terms of cost and complexity. There may also be situations in which there is a strong desire to increase the confidence in the LE (e.g. in tumour types where progression is more difficult to define or measure). In such cases a full BICR may be considered.

There are several logistical aspects of a sample-based BICR that need to be considered. All scans should be centrally managed, to allow maximum flexibility regarding management of the sample-based BICR, and to prevent significant delays and logistical hurdles associated with having to retrospectively retrieve the scans from the investigational centres. A typical approach would be to take a random sample of all subjects for the sample-based BICR including those who have progressed by LE and those who have not. An alternative approach which reduces cost and complexity is to sample and centrally review only the LE PD events, however, this reduces the number of discordance measures that can be evaluated and may introduce bias if investigators know which scans will be subjected to a BICR. As to the timing of the sample-based BICR, a simple approach is to prospectively identify a random sample of patients once enrolment is complete, and perform the central review of scans after the clinical cutoff has been achieved but prior to breaking the randomisation code. A second approach, if one is sampling events only, is to perform the central review during the conduct of trial, randomly selecting subjects who progressed by LE. An additional approach for consideration is to draw a random sample while the study is ongoing from the first set of patients enrolled in the trial. Results could be shared with an IDMC who could in turn make recommendations regarding the need for further central review. Lastly, one may choose to look at the PFS results of the LE before deciding on whether to proceed with a sample-based BICR. This may be a useful approach in a setting where large, robust effects on PFS are observed.

In summary, we feel that the recommendations presented, advocating a careful assessment of the need for a BICR and the possibility of a sample-based approach, will lead to more efficient, cost-effective randomised trials in oncology with a PFS end-point. These gains in efficiency and cost, in our view, will not compromise the interpretability of the results or the integrity of the trial as a whole, and as such will ultimately allow new therapies to patients more rapidly and efficiently.

Conflict of interest statement

The authors disclose the following: O. Amit, F. Mannino and W. Bushnell own stock in Glaxosmithkline; J. Denne owns stock in Eli Lilly; J. Helterbrand and H.U. Burger own stock in Hoffman LaRoche; A.M. Stone owns stock in AstraZeneca.

Acknowledgements

The authors would like to acknowledge the contribution of the following members of the PhRMA PFS Independent Review Working Group: Paul Bycott: Pfizer, Steve Dahlberg: Amgen, Andrew Strahs: Wyeth, Sandra Chica: Novartis, Grant Williams, Vicki Goodman: GSK, Sam Givens: Roche, Nicole Blackman: GSK, Bill Capra: Genentech, Bee Chen: Novartis.

The authors would additionally like to acknowledge the following for their assistance: Michelle Casey: GSK, Michael Durante: GSK, Soumi Lahiri: GSK, Dan Sargent: Mayo Clinic, Karen Fitch: GSK, Lauren McCann: GSK.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejca.2011.02.013.

REFERENCES

1. Dancey JE, Dodd LE, Ford R, et al. Recommendations for the assessment of progression in randomised cancer treatment trials. *Eur J Cancer* 2009;**45**:281–9.
2. Shamsi K, Patt RH. Onsite image evaluations and independent image blinded reads: close cousins or distant relatives? *J Clin Oncol* 2009;**27**:2103–4.
3. Tang PA, Pond GR, Chen EX. Influence of an independent review committee on assessment of response rate and progression-free survival in phase III clinical trials. *Ann Oncol* 2010;**21**(1):19–26.
4. Dodd LE, Korn EL, Freidlin B, et al. Blinded independent central review of progression-free-survival in phase III clinical trials: important design element or unnecessary expense? *J Clin Oncol* 2008;**26**:3791–6.
5. Stone AM, Bushnell W, Denne J, et al. Research outcomes and recommendations for the assessment of progression in cancer clinical trials from a PhRMA working group. *Eur J Cancer* 2011;**47**(12):1763–71.
6. Pogue J, Walter SD, Yusuf S. Evaluating the benefit of event adjudication of cardiovascular outcomes in large simple RCTs. *Clin. Trials* 2009;**6**:239–51.
7. Kemsley K, Ghiorgiu D, Schmitt N, et al. Use of progression-free survival in advanced melanoma: comparison of central and site review. *J Clin Oncol* 2007;**27**:18s [Suppl.; abstr. 1011].
8. Sternberg CN, Szczylik C, Lee E, et al. A randomized, double-blind phase III study of pazopanib in treatment-naïve and cytokine-pretreated patients with advanced renal cell carcinoma (RCC). *J Clin Oncol* 2009;**27**:15s [Suppl.; abstr. 5021].
9. Escudier BJ, Bellmunt J, Negrier S, et al. Final results of the phase III, randomized, double-blind AVOREN trial of first-line bevacizumab (BEV) + interferon-α2a (IFN) in metastatic renal cell carcinoma (mRCC). *J Clin Oncol* 2009;**27**:15s [Suppl.; abstr. 5020].
10. Motzer RJ, Hutson TE, Tomczak P, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med* 2007;**356**:115–24.
11. Escudier B, Eisen T, Stafler W, et al. Sorafenib in advanced clear-cell renal-cell carcinoma. *N Engl J Med* 2007;**356**:125–34.
12. FDA: FDA Medical Review: NDA 22-334 Afinitor (everolimus), <http://www.fda.gov/ohrms/dockets/ac/07/briefing/2007-4332b1-01-FDA.pdf>.
13. Ravaud A, Hawkins R, Gardner JP, et al. Lapatinib versus hormone therapy in patients with advanced renal cell carcinoma: a randomized phase III clinical trial. *J Clin Oncol* doi:10.1200/JCO.2007.14.5029.
14. Van Cutsem E, Peeters M, Siena S, et al. Open-label phase III trial of panitumumab plus best supportive care alone in patients with chemotherapy – refractory metastatic colorectal cancer. *J Clin Oncol* 2007;**25**:1658–64.

15. Hecht JR, Mitchell E, Chidiac T, et al. A randomized phase IIIb trial of chemotherapy, bevacizumab, and panitumumab compared with chemotherapy and bevacizumab alone for metastatic colorectal cancer. *J Clin Oncol* doi:10.1200/JCO.2008.19.8135.
16. Hoff PM, Ansari R, Batist G, et al. Comparison of oral capecitabine versus intravenous fluorouracil plus leucovorin as first-line treatment in 605 patients with metastatic colorectal cancer: results of a randomized phase III study. *J Clin Oncol* 2001;**19**(8):2282–92.
17. Van Cutsem E, Twelves C, Cassidy J, et al. Oral capecitibine compared with intravenous fluorouracil plus leucovorin in patients with metastatic colorectal cancer: results of a large phase III study. *J Clin Oncol* 2001;**19**(21):4097–106.
18. Rothenberg ML, Cox JV, Butts C, et al. Capecitabine plus oxaliplatin (XELOX) versus 5-fluorouracil/folinic acid plus oxaliplatin (FOLFOX-4) as second-line therapy in metastatic colorectal cancer: a randomized phase III non-inferiority study. *Ann Oncol* 2008;**19**(10):1720–6.
19. Kabbinavar FF, Schulz J, McCleod M, et al. Addition of bevacizumab to bolus 5-FU/leucovorin in first-line metastatic colorectal cancer: results of a randomized phase II trial. *J Clin Oncol* 2005;**23**(16):3697–705.
20. Albain KS, Nag S, Calderillo-Ruiz G, et al. Global phase III study of gemacitabine plus paclitaxel (GT) vs. paclitaxel (T) as frontline therapy for metastatic breast cancer (MBC): first report of overall survival. *J Clin Oncol* 2004;**22**(14s, Suppl. 510).
21. Miller KD, Chap LI, Holmes FA, et al. Randomized phase III trial of capecitibine compared with bevacizumab plus capecitibine in patients with previously treated metastatic breast cancer. *J Clin Oncol* doi:10.1200/JCO.2005.05.098.
22. Geyer CE, Forster J, Lindquist D, et al. Lapatinib plus capecitibine for HER2-positive advanced breast cancer. *N Engl J Med* 2006;**355**:2733.
23. Di Leo A, Gomez H, Aziz Z, et al. Lapatinib (L) with paclitaxel as first-line treatment for patients with metastatic breast cancer: a phase III randomized, double-blind study of 580 patients. *J Clin Oncol* 2007;**27**(18s) [Suppl.; abstr 1011].
24. Blackwell KL, Burstein HJ, Storniolo AM, et al. Randomized study of lapatinib alone or in combination with trastuzumab in women with ErbB2-positive, trastuzumab-refractory metastatic breast cancer. *J Clin Oncol* 2010;**28**(7):1124–30.
25. Johnston S, Pippen J, Pivot X, et al. Lapatinib combined with letrozole versus letrozole and placebo as first-line therapy for postmenopausal hormone receptor- positive metastatic breast cancer. *J Clin Oncol* doi:10.1200/JCO2009.23.3734.
26. FDA: FDA Briefing Document Oncology Drug Advisory Committee Meeting December 5, 2007: BLA STN 125085/91.018 Avastin (bevacizumab), <http://www.fda.gov/ohrms/dockets/ac/07/briefing/2007-4332b1-01-FDA.pdf>.
27. Fine HA, Puduvalli VK, Chamberlain MC, et al. Enzastaurin (ENZ) versus lomustine (CCNU) in the treatment of recurrent, intracranial glioblastoma multiforme (GBM): a phase III study. *J Clin Oncol* 2008:26 [Suppl.; abstr 2005].
28. Vogelzang NJ, Rusthoven JJ, Symanowski J, et al. Phase III study of pemetrexed in combination with cisplatin versus cisplatin alone in patients with malignant pleural mesothelioma. *J Clin Oncol* 2003;**21**(14):2636–44.
29. Dummer R, Robert C, Chapman B, et al. AZD6244 (ARRY – 142886) vs temozolomide (TMZ) in patients (pts) with advanced melanoma: an open-label, randomized, multi-center, phase II study. *J Clin Oncol* 2008:26 [Suppl.; abstr 9033].
30. FDA: FDA Briefing Document Oncology Drug Advisory Committee Meeting July 15, 2009: NDA 22-447 Yondelis (Trabectedin). <http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/OncologicDrugsAdvisoryCommittee/UCM171149.pdf>.
31. Sternberg CN, Petrylak D, Witjes F, et al. Satraplatin (S) demonstrates significant clinical benefits for the treatment of patients with HRPC: results of a randomized phase III trial. *J Clin Oncol* 2007:25 [Suppl.; abstr 5019].
32. FDA: FDA Briefing Document Oncology Drug Advisory Committee Meeting July 15, 2009: NDA 22-447 Tarceva (Erlotinib). <http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/OncologicDrugsAdvisoryCommittee/ucm193916.htm>.
33. Robert NJ, Dieras V, Glaspy J, et al. RIBBON-1: randomized, double-blind, placebo-controlled, phase III trial of chemotherapy with or without bevacizumab (B) for first-line treatment of HER2-negative locally recurrent or metastatic breast cancer (MBC). *J Clin Oncol* 2009;**27**(15s) [Suppl; abstr 1005].