FDA Briefing Document
Oncologic Drugs Advisory Committee Meeting

July 24, 2012

Evaluation of Radiologic Review

of Progression-free Survival in Non-hematologic Malignancies

Disclaimer*: The attached package contains background information prepared by the Food and Drug Administration (FDA) for the panel members of the advisory committee. The FDA background package often contains assessments and/or conclusions and recommendations written by individual FDA reviewers. Such conclusions and recommendations do not necessarily represent the final position of the individual reviewers, nor do they necessarily represent the final position of the Review Division or Office. We have brought the discussion of evaluation of radiologic reviews in randomized clinical trials using progression-free survival as a primary endpoint in non-hematologic malignancies to this Advisory Committee in order to gain the Committee's insights and opinions, and the background package may not include all issues relevant to the final regulatory recommendation and instead is intended to focus on issues identified by the Agency for discussion by the advisory committee. The FDA will not issue a final determination on the issues at hand until input from the advisory committee process has been considered. The final determination may be affected by issues not discussed at the advisory committee meeting.*

Table of Contents

Tables

Figures

1 Background

This session of Oncologic Drug Advisory Committee (ODAC) will focus on a general discussion of oncologic products seeking marketing approval in the US for the treatment of non-hematologic malignancies based on results from randomized clinical trials with primary endpoint of progression-free survival (PFS), defined as time from randomization to either disease progression or death, whichever occurs first. These discussions will not be specific to any product or disease, and will not discuss whether PFS is the appropriate primary efficacy endpoint. These discussions will in particular consider the merits of an independent audit of investigator assessment of progression in a random sample of patients instead of an independent review of all patients. The expectation is that an independent audit would streamline the conduct of clinical trials, as well as, avoid missing data when no additional protocol-specified progression assessments are mandated. Hematologic malignancies are excluded from this discussion because of other issues (e.g., blood counts, lymph node exams, and other biomarkers) that may influence the assessment of PFS.

2 Current Practice

When PFS is the primary efficacy endpoint of a clinical trial, FDA has generally required independent radiologic review (IRC) of scans under the assumption that local evaluation or investigator assessment (INV) could potentially be biased. Thus, the role of IRC is to mitigate potential evaluation bias by investigators. However, this approach may lead to a greater than 30% disagreement at the patient-level between the investigator and independent reviewer assessments and/or among independent reviewers themselves. Because treatment is generally changed after investigator-determined progression resulting in no further protocol-specified progression assessments, this practice results in missing data and informed censoring for IRC-determined PFS analyses. These disagreements have been attributed to a variety of reasons, including monitoring different target lesions. A common assumption in time-to-event analyses, including the method employed in PFS analyses, is that censoring is not related to the outcome of interest. Censoring is “informative” when the prognosis of the censored patient differs from those on the same treatment arm who are continued to be followed for an event. Informative censoring can lead to significant bias and increased variability in the treatment effect evaluation.

3 Comparison of Investigator *versus* Independent Review

In order to understand the benefits of using IRC and in discussions with FDA, a PhRMA working group analyzed data collected from 27 randomized clinical trials and concluded the existence of a high degree of correlation between INV- and IRC-determined PFS treatment effects as measured by hazard ratios (HR) (Amit et. al., EJC 2011). We have conducted a similar analysis using available data submitted to FDA and observed a high degree of correlation between INV- and IRC-determined PFS treatment effects as measured by HR and also effects on objective response rates (ORR) as measured by odds ratios (Figures 1 and 2). An important finding in our research is that there is no systematic bias that was introduced by the investigator.

Figure 1: PFS - INV vs. IRC Assessment

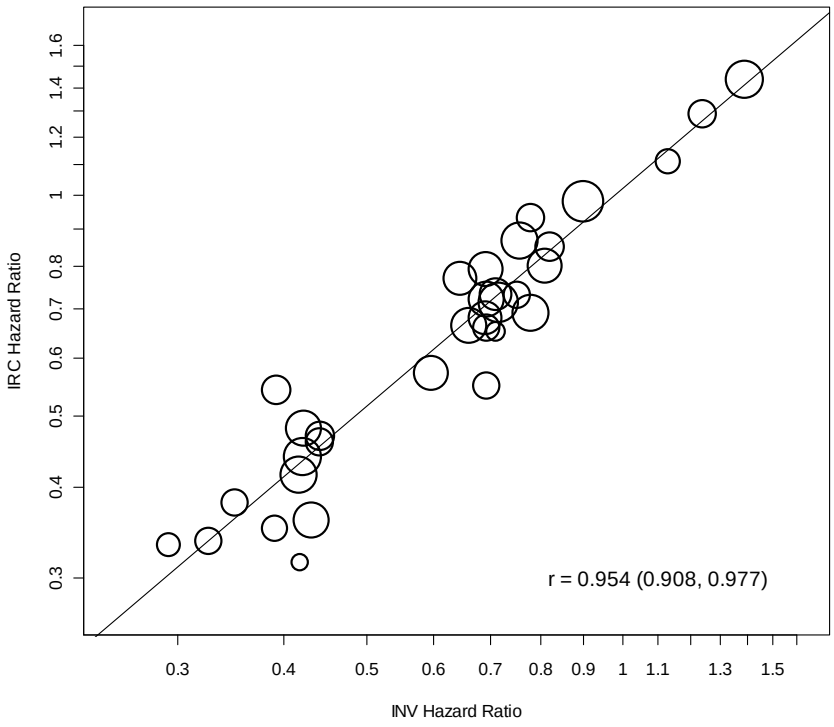
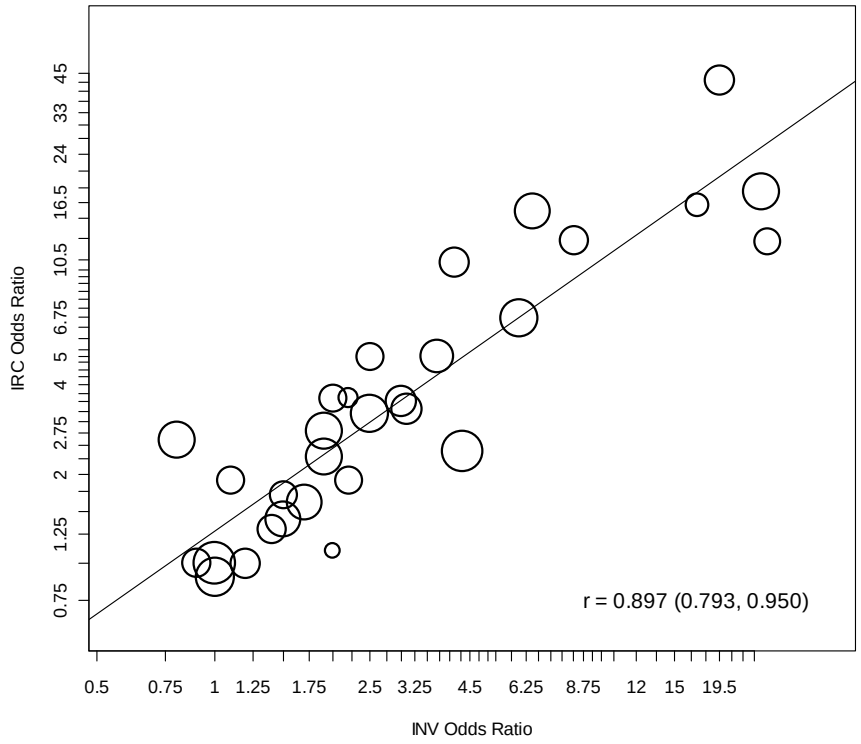


Figure 2: ORR - INV vs. IRC Assessment



The decision on a drug or biologic marketing approval is based on population level statistics such as HRs rather than individual patient level statistics. Given these results from both the PhRMA working group and our own research, we are exploring the efficiency of conducting an independent review audit in a random sample of patients' radiologic scans to ensure the consistency between IRC and INV assessments. This would reduce the cost and burden on the clinical trial investigators, avoid some of the missing data issues, and mitigate informative censoring in analyzing the time-to-event endpoints.

4 Audit Methodology

Currently, two methods that have been proposed for this type of audit (Dodd et.al., Biometrics 2011 and Amit et.al., EJC 2011), neither have been evaluated in prospectively conducted studies. Dodd et. al. (2011) proposes to evaluate the consistency of treatment effect as measured by a HR between the IRC audited assessments and the INV assessments. They also provide a method for audit sample size calculation. Amit et.al. (2011) proposes to evaluate the differential discrepancy rates of INV versus IRC between the treatment and the control arms. Further details of these methods are

provided in the Appendix. With ongoing research efforts, we expect that other audit methods will be proposed.

We have evaluated these two methods by conducting retrospective analyses using available data at FDA from 28 clinical trials in 9 disease settings summarized in Table 1. Results from these analyses will be presented at the ODAC meeting.

Table 1: Summary of Trial Characteristics

Study characteristics	Meta-analysis trials (N = 28)
Tumor type	
MBC	7
RCC	7
MCRC	4
Other[a]	10
Design	
AC / AC-AO / PBO or BSC / SUBST[b]	4 / 7 / 16 / 1
Open-label / Double-blind	13 / 15
Interim / Final analysis	9 / 19
1:1 / 2:1 randomization	20 / 8
1st / subsequent[c] / maintenance line	13 / 13 / 3
Superiority / Non-inferiority	27 / 1
Sample Size	
Median	716.5
Min, Max	171, 1725

[a]Includes trials in non-small cell lung cancer (3), pancreatic neuroendocrine tumors (2), soft tissue sarcoma (2), gastrointestinal stromal tumor (1), ovarian cancer (1), carcinoid tumors (1); [b]AC = active control (e.g. drug A vs. drug B), AC-AO = active control add-on (e.g. A+B vs. A), PBO or BSC = placebo-controlled or best supportive care, SUBST = substitution (e.g. A+C vs. B+C); [c]One trial included both 1st and 2nd line patients and is double counted here

5 Summary

In summary, there is general agreement between IRC and INV with respect to relative treatment effects based on PFS assessments. An inherent measurement error exists in the reading of radiographic scans and disagreements between readers at the patient level are commonly observed. However, regulatory considerations are based on the relative treatment effect at the population level. We also acknowledge that PFS may not be the appropriate primary efficacy endpoint in specific indications and that magnitude of the

treatment effect with the observed risks are important in formulating a benefit to risk evaluation.

If we proceed with implementing a random sample IRC audit method, further considerations need to be addressed, including whether the audit should be conducted only for "positive by INV assessment" trials, and whether more than one independent reviewer is necessary to perform the IRC. Other potential concerns, including uniform training of site radiologists, scan procurement and storage, and quality control will require further discussion. The measurement and reproducibility of the radiologic evaluations will be presented at the ODAC meeting. Selected references are provided below.

Assuming that PFS is an appropriate primary endpoint for a specific indication, and the FDA analysis results, ODAC will be asked to discuss the following issues and provide advice:

1. Irrespective of the audit methodology used, what are the pros and cons of using a random sample audit by the IRC instead of a review of all scans by the IRC?

2. Under what circumstances a review of all scans by the IRC should be considered?

3. How many independent reviewers should be included in the IRC?

4. What threshold value of discrepancy between INV and the audited IRC assessment would lead to discarding the trial results?

5. Discuss whether an audit could be considered for "small" studies, and how to define "small". What factors should be considered in deciding the size of the audit?

6 References

Dodd LE, Korn EL, Freidlin B, et al: "An audit strategy for progression-free survival". *Biometrics* 67:1092-1099, 2011.

Amit O, Mannino F, Stone AM, et al: "Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis". *European Journal of Cancer* 47:1772-1778, 2011.

Hochberg, Y: "A sharper Bonferroni procedure for multiple tests of significance". *Biometrika* 75: 800–802, 1988.

McNitt-Gray MF, Bidaut LM, Armato III SG, et al: "Computed tomography assessment of response to therapy: Tumor volume change measurement, truth, data, and error". *Translational Oncology* 2: 216-222, 2009.

Armato III SG, Meyer CR, McNitt-Gray MF, et al: "The reference image database to evaluate response to therapy in lung cancer (RIDER) project: A resource for the development of change-analysis software". *Clin Pharmacol Ther* 84:448-456, 2008.

Ford R, Schwartz L, Dancey J, et al: "Lessons learned from independent central review". *European Journal of Cancer* 45: 268-274, 2009.

7 Appendix

Dodd et al. (2011) Audit Method

Dodd et al. (2011) have proposed a two-stage auditing strategy as an alternative to a complete-case IRC to detect bias in treatment effect estimators based on local evaluations (INV). The primary goal of the audit is to provide an asymptotically unbiased efficient estimator of the IRC-based hazard ratio (Θ_c), which simultaneously incorporates information from the patient-level INV data on all the cases and the retrospective random sample IRC audit cases. This two-stage testing procedure utilized the Hochberg procedure (Hochberg, 1988) to adjust for multiple comparisons for these two stages. At the first stage, with audit size δ_0, the upper bound of a $1 - \alpha/2$ confidence interval of Θ_c is computed. If the upper bound is below a threshold of clinical significance for PFS improvement, termed the clinical irrelevance factor (CIF), the procedure stops and concludes that the IRC audit has confirmed the INV findings (i.e., consistency of the treatment effect has been shown). Otherwise, the audit proceeds to the second stage, which is a complete-case IRC.

At the second stage, the upper bound of a $1 - \alpha/2$ confidence interval of Θ_c on the complete-case IRC is computed; if that value is below the CIF, the audit procedure stops and consistency of the treatment effect is concluded. If the bound is above this threshold, then the upper bounds of the $1 - \alpha$ confidence intervals of Θ_c for both δ_0 and the full audit are estimated. If both of these bounds fall below the CIF, then consistency of the treatment effect is concluded. Otherwise, inconsistency of the treatment effect is concluded.

This method is limited to superiority clinical trials and a retrospective IRC will only be conducted when the INV hazard ratio indicates a clinically meaningful and statistically significant effect of the experimental treatment.

Amit et al. (2011) Audit Method

Amit et al. (2011) proposed a sample-based IRC procedure based on differential discordance to detect bias in the local evaluation. Differential discordance is evaluated using two measures, the early discrepancy rate (EDR) and late discrepancy rate (LDR), as defined below in conjunction with Table A1.

Table A1: IRC versus INV disease progression assessments

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

Note: a1: number of agreements on timing and occurrence of PD
a2: number of times INV declares PD later than IRC
a3: number of times INV declares PD earlier than IRC
PD: progressive disease

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

EDR quantifies the frequency with which INV declares progression earlier than IRC within each arm as a proportion of the total number of investigator assessed disease progressions. LDR quantifies the frequency that INV declares progression later than IRC as a proportion of total discrepancies within the arm.

A similar distribution of discrepancies between the arms suggests absence of evaluation bias. The differential discordance of EDR and LDR between two arms is defined as the rate on the experimental arm minus the rate on the control arm. A negative differential discordance in EDR and/or a positive differential discordance in LDR are suggestive of a bias in INV favoring the experimental arm.

Based on findings from a meta-analysis of 12 clinical trials and simulations, a sample-based IRC procedure was developed to detect potential evaluation bias in the INV of PD by performing a IRC in a random sample of patients ("audit"). Simulations indicated that a differential discordance of 0.15 predicts a 20-30% relative difference in the HR between INV and IRC evaluations. To detect evaluation bias, the authors proposed a threshold value ranging from 0.075 to 0.1 in a sample-based IRC with a size of 100 to 160 subjects. If the discordance statistics support the absence of any bias in the INV assessment, a complete IRC might not be necessary. If bias cannot be excluded based on the audit, a complete IRC should be implemented.