Secondary (Biomarker Qualification) Statistical Review and Evaluation

(An update following the Statistical Review and Evaluation written by the review team and emailed by Dr. Mohammad Huque on December 06, 2010)

Subject: Biomarker qualification (BQ) for detection of Galactomannan (GM) in serum and bronchoalveolar lavage (BAL) fluid by the Platelia Aspergillus enzyme immunoassay (EIA) manufactured by Bio-Rad laboratories & Sanofi Diagnostics

Reference: Primary statistical review and evaluation biomarker qualification GM BQ executive summary and integrated review

Medical Division: Division of Special Pathogen and Transplant Products

Background

The purpose of this review is to facilitate the statistical review and evaluation by primary statistical reviewer, Dr. Cheryl Dixon, in response to the feedback on the GM BQ executive summary and integrated review, discussed at an internal meeting on March 24, 2010 and the updated statistical review and evaluation sent by Dr. Mohammad Huque on December 06, 2010 that has implication for BQ consideration.

Note: the review package for this BQ was literature based. There were no raw data available for data analyses and for providing detailed review evaluation. The recommendations were mainly based on a review summary of 23 articles for serum GM assay and 6 articles for BAL GM assay (the updated statistical review and evaluation has updated this to include 12 articles).

Statistical Issues

Literature has reported that non-differential misclassification *may* introduce bias toward null. In this BQ submission, a positive GM diagnostic assay result will be used to enroll patients into a controlled trial. Assume the clinical outcomes of the misclassified subjects are imputed with the same value as that considered in this BQ submission. Then, individuals without probable invasive aspergillosis (IA) will likely be similar in their treatment outcome, making it more difficult to show a difference and to inflate type I error rate for superiority testing. For a non-inferiority (NI) objective, the potential impact is just the opposite; that is, inclusion of individuals without probable invasive aspergillosis may tend to bias toward falsely showing non-inferiority. Thus, to maximize enrolling patients who have probable invasive aspergillosis, the primary diagnostic measure of critical concern for a NI trial is the percentage of patients with probably invasive aspergillosis among those with a positive GM assay result. This percentage is the positive predictive value (PPV), which is a function of IA prevalence in the intent-to-diagnose patient population.

To properly assess the type I error probability of an NI trial when a diagnostic assay is used to identify eligible patients, there needs to be justification of many assumptions on the diagnostic assay performance characteristics and the NI trial design parameters. These assumptions are interrelated. For a sufficiently powered NI trial, the type I error rate is controlled if the diagnostic assay has 100% PPV. As the percentage of subjects without probable IA being classified as GM positive increases, the bias for concluding NI increases, especially if those non-IA subjects enrolled in the NI trial is assumed to be a success.

In this submission, 23 articles for serum GM assay and 6 articles for BAL GM assay were used to estimate the sensitivity and the specificity. In general, the estimated IA prevalence in these case-control studies is low. The sample sizes for estimating the sensitivity of serum GM assay were small ranging from 2 to 98 with approximately 75% (17) of the studies had less than 30 subjects. The sample sizes for estimating the specificity of serum GM assay ranged from 17 to 751 with about one half of the studies (11 out of 23) had at least 100 subjects. For the BAL GM assay, the sample sizes were not large ranging from 7 to 58 for sensitivity estimates and 10 to 76 for specificity estimates. The statistical review and evaluation written by the statistical review team sent by Dr. Mohammad Huque on December 06, 2010 had updated from 6 articles to 12 articles for BAL GM assay literature.

It is noted that the review comment of the PPV in the restricted patient population will be in the range of 90% assumes the true sensitivity and true specificity of the GM assay of about 90%. When the PPV is in the range of 90%, the type I error rate inflation would be close to two-fold as compared to the conventional one-sided 2.5%, see Figure 1 of Dr. Cheryl Dixon's review copied from memo dated May 30, 2009 by Dennis Wallace to Pete Pappas and Joe Wheat and submitted to biomarker qualification review package.

Our current consideration with a superiority study design requires proper control of the type I error rate at a one-sided 2.5% or a two-sided 5% level. Given the concerns of disease misclassification with a diagnostic assay and the bias toward showing non-inferiority in a non-inferiority trial design, what should be an appropriate level of type I error rate inflation is at issue. If reducing the type I error rate inflation to a potentially tolerable level, such as around an excess of one-third or less of a one-sided 2.5%, then, a PPV in the range of 95% or higher would be needed.

Note that most of the studies reported in the literatures are of relatively small sample sizes and are not from randomized controlled trials. When the uncertainty with the sensitivity and specificity estimates obtained from the literature case-control studies is incorporated, e.g., the first quartile sensitivity (57%) and specificity (93%) seen for the results for "two consecutive tests at a cut off greater than 1.0" (Table 2 in the Appendix – In the updated statistical review and evaluation by Dr. Cheryl Dixon, it is labeled as 'Table 4' and is modified

accordingly as Table 4 in the Appendix in this updated secondary review), it appears that the prevalence of IA would need to be much higher than 50%. For instance, as shown in Table I (see the highlighted cells in page 3), the prevalence of IA may need to be at least 70% or higher in order to maintain a PPV level in the range of 95% or higher. Such prevalence level would help minimize the bias that may increase the risk of falsely concluding non-inferiority to an unacceptable level due to the enrolment of patients with false positive GM assay result.

Summary

It can be shown mathematically that the type I error rate for concluding non-inferiority of a new treatment relative to its active comparator is inflated (higher than the usual 1-sided 2.5%) if the diagnostic assay used to select patient for GM positivity has less than 100% specificity, equivalently, 100% PPV, assuming the constancy of the control effect and the assay sensitivity hold. Given the concerns of disease misclassification with a diagnostic assay and the bias toward showing non-inferiority, what should be an appropriate level of type I error rate inflation will continue to be the major challenge when the diagnostic assay is used to select patients for a non-inferiority study design.

There was an added commentary in the statistical review and evaluation written by Dr. Cheryl Dixon regarding the effect of voriconazole in the voriconazole study. Specifically, “It is believed that there is adequate historical evidence of drug effect and that this drug effect is robust over varying levels of the diagnosis of invasive aspergillosis infection.” Thus, whether future study can be designed to evaluate non-inferiority of a treatment relative to voriconazole is not a concern. The non-inferiority margin needs to be pre-specified for future non-inferiority trials. The main issues of assay sensitivity and constancy assumption in a non-inferiority trial will need to be addressed if voriconazole is the active comparator studied in the patient population identified as positive via GM assay. This includes, e.g., if medical practices and/or background medications change over time from that considered by 2002, patient population comparability due to the use of GM assay versus without the use of GM assay in 2002.

Sue-Jane Wang

Sue-Jane Wang, Ph.D.
Associate Director, Pharmacogenomics and Adaptive Design
Office of Biostatistics, Office of Translational Sciences, CDER

Date: April 22, 2010 (new date: May 02, 2011)

cc: Robert T. O’Neill

Table I. Estimated PPV based on first quartile of sensitivity & specificity

Sensitivity	Specificity	Prevalence						
		30%	40%	50%	60%	70%	80%	90%
83%[1]	75%[1]	59%	69%	77%	83%	89%	93%	97%
65%[2]	86%[2]	67%	76%	82%	87%	92%	95%	98%
65%[3]	75%[3]	53%	63%	72%	80%	86%	91%	96%
57%[4]	93%[4]	78%	84%	89%	92%	95%	97%	99%

[1]two consecutive samples with 0.5 cutoff
[2]single sample with 1.0 cutoff
[3]lowest first quartile from 1 and 2
[4]two consecutive samples with 1.0 cutoff

Appendix

Table 4 Summary of Median and Interquartile Range Values for Serum GM*

Cutoff	# Samples	# Studies	Sensitivity	Specificity	PPV	NPV
0.5	Single	6	97 (65-100)	88 (61-94)	53 (24-68)	99 (94-100)
	Consecutive	6	93 (83-97)	91 (75-98)	71 (34-90)	98 (96-99)
1.0	Single	8	93 (65-99)	90 (86-96)	61 (43-79)	98 (92-100)
	Consecutive	9	88 (**57**-97)	98 (**93**-99)	85 (72-94)	98 (93-99)
1.5	Single	9	69 (38-80)	95 (92-99)	55 (45-93)	96 (90-98)
	Consecutive	9	75 (30-88)	98 (90-99)	63 (50-90)	94 (88-98)

* from Statistical Review and Evaluation Biomarker Qualification by Dr. Cheryl Dixon