# CENTER FOR DRUG EVALUATION AND RESEARCH

# APPROVAL PACKAGE FOR:

# APPLICATION NUMBER

# 50-794

# Statistical Review(s)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Pharmacoepidemiology and Statistical Science
Office of Biostatistics

# STATISTICAL REVIEW AND EVALUATION

## CLINICAL STUDIES

| | |
|---|---|
| **NDA/Serial Number:** | 50-794 / N000 |
| **Drug Name:** | Vidaza™ |
| **Indication(s):** | Myelodysplastic syndromes / pre-leukemia / smoldering leukemia |
| **Applicant:** | Pharmion Corporation |
| **Date(s):** | Submission Date: December 26, 2003 |
| | Date Review Completed: April 27, 2004 |
| | PDUFA Date: June 26, 2004 |
| **Review Priority:** | Priority |
| **Biometrics Division:** | Division of Biometrics I (HFD-710) |
| **Statistical Reviewer:** | Yong-Cheng Wang, Ph.D. (HFD-710) |
| **Concurring Reviewers:** | Rajeshwari Sridhara, Ph.D. (HFD-710) |
| | Kooros Mahjoob, Ph.D. (HFD-710) |
| **Medical Division:** | Division of Oncology Drug Products (HFD-150) |
| **Clinical Team:** | Edvardas Kaminskas, M.D. & Ann T. Farrell, M.D. (HFD-150) |
| **Project Manager:** | Ms. Amy Baird |

**Keywords:** Retrospective design, Retrospective data collection, Treatment crossover, Response rate, Transfusion rate, Multiple comparison, Fisher's exact test, Wilcoxon test

# Table of Contents

# 1. EXECUTIVE SUMMARY

The study results submitted in this NDA was a study which was conducted by CALGB/NCI in 1994. The results of this study have been reported in literature. The NDA submitted for review is based on a retrospective statistical design and retrospective collection and analyses of data. The primary and secondary endpoints and statistical hypothesis were redefined. The statistical analysis plan was revised from the original protocol. Based on the ICH-E9 guidance: Statistical Principles for Clinical Trials, the study model in this submission is therefore exploratory in nature (Appendix 1).

The study selected for this statistical review is CALGB 9221 which was a Phase 3, open-label, multicenter study designed to compare azacitidine plus supportive care with an observation group receiving best supportive care (BSC) for patients with myelodysplastic syndrome (MDS). Study CALGB 9221 was also designed to allow subjects randomized to the observation group to switch (treatment crossover) to azacitidine treatment after meeting protocol criteria.

The original protocol of the CALGB 9221 study defined the primary objectives as to determine the response rate to azacitidine and the impact of azacitidine on red cell transfusion requirements, platelet counts, ANC, rates of infection and hemorrhage and % BM blasts, in comparison to an untreated observation group. It also defined the response as either Complete Response (CR), Partial Response (PR), or Improvement. However, a retrospective definition of the primary endpoint of overall response (CR+PR) was employed in this NDA submission.

## 1.1 Conclusions and Recommendations

For the overall response (primary endpoint in the revised protocol, complete response and partial response), CALGB 9221 appears to demonstrate a significant benefit of azacitidine compared to observation (1). before crossover to treatment arm in the intern-to-treat (ITT) population ($p < 0.0001$), (2). in ITT population without acute myelogenous leukemia (AML) ($p < 0.0001$), and (3). ITT population without AML or protocol violation ($p = 0.0007$). It also appears to demonstrate a significant benefit of azacitidine compared to observation only group (excluding crossover patients) for ITT population ($p = 0.0033$), ITT population without AML ($p = 0.010$), and ITT population without AML or protocol violation ($p = 0.0276$). However, it failed to demonstrate any significant benefit of azacitidine compared to observation before crossover group as measured by complete response for ITT population without AML and/or protocol violation patients. It also failed to demonstrate any significant benefit of azacitidine compared to observation only group as measured by complete response for ITT population with AML and/or protocol violation patients and ITT population without AML and/or protocol violation patients.

For the change of monthly transfusion requirement, the comparisons of azacitidine versus observation before crossover and azacitidine versus observation only group did not reach the significance level 0.05 under Bonferroni adjustment for both red blood cell transfusion and platelet transfusion. For the change of transfusion requirement per cycle, the comparisons of azacitidine versus patients crossed over to azacitidine group also did not reach the significance level 0.05 under Bonferroni adjustment for red blood cell transfusion and platelet transfusion.

In this statistical reviewer's opinion the data and results of the one, small Phase III study suggest activity of azacitidine in patients with MDS. However, the results are not adequate to support the sponsor's efficacy claim as such since they are based on retrospective statistical design, retrospective data collection and analyses. Furthermore, the strength of statistical significance can not be evaluated based on p-value due to the retrospective nature of the study. Based on the guidance ICH-E9, in this statistical reviewer's opinion this submission can only be considered as exploratory and hypotheses generating analyses. However, azacitidine has been under investigation for over 30 years. The final recommendation should be based on clinical judgment.

### 1.2 Brief Overview of Clinical Studies

This application consists of report of results from registration Study CALGB 9221 conducted during 1994 to 2002 in the treatment of patients with MDS, and supportive data from Study CALGB 8421 and Study CALGB 8921 in the treatment of patients with MDS.

Study CALGB 8421 and Study CALGB 8921 were two Phase II, open-label, uncontrolled, and non-comparative studies, which evaluated azacitidine in subjects with the MDS subtypes of RAEB and RAEB-T, and CMMoL in CALGB 8921 only.

In Study CALGB 8421 up to 20 subjects were initially planned for enrollment; however, based on the results of the first 13 subjects, the enrollment goal was extended to 45 subjects. In this study subjects were to receive azacitidine 75 mg/m2 in Ringer's lactate as a continuous IV infusion for 7 days on a 28-day cycle for a minimum of 4 cycles.

In Study CALGB 8921, up to 50 subjects were initially planned for enrollment over 1.5 years; subjects were to receive 75 mg/m2 azacitidine SC, daily for 7 days on a 28-day cycle for a minimum of 4 cycles. In both of these studies, the azacitidine dose could be adjusted (either increased, decreased, or delayed) at the beginning of any cycle based on predefined hematology and renal laboratory results relating to the well-being of the subject.

CALGB 9221 was a Phase 3, open-label, multicenter controlled study designed to compare azacitidine plus supportive care with an observation group receiving best supportive care in subjects with any of the 5 subtypes of MDS. In this study, up to 174 subjects were initially planned for enrollment over 25 months. Subjects with all 5 subtypes of MDS RA, RARS, RAEB, RAEB-T, and CMMoL were to be balanced between the two treatment arms. Subjects in the azacitidine treatment arm were to receive 75 mg/m2 azacitidine SC, daily for 7 days on a 28-day cycle for a minimum of 4 cycles. Similar to the design of the 2 phase 2 studies, the azacitidine dose could be adjusted (either increased, decreased, or delayed) at the beginning of any cycle based on predefined hematology and renal laboratory results relating to the well-being of the subject.

Additionally, the CALGB 9221 study was designed to allow subjects randomized to the observation group to switch (treatment crossover) to azacitidine treatment after meeting protocol criteria for increases in bone marrow blasts, decreases in hemoglobin or platelets, increased red

blood cell (RBC) or platelet transfusion requirements, or clinical infection with low absolute neutrophil count (ANC) and requiring treatment with intravenous antibiotics.

During the prospective conduct of all 3 CALGB studies, specific response and safety assessments were to be performed at scheduled intervals. Although the CALGB 8421 protocol did not specify criteria for recording adverse events, treatment-related toxicities and complications were to be reported on CALGB Follow-up Forms which were then collected along with reports from source documents as adverse events for the retrospective analysis of this study. After withdrawal from the study, subjects in all 3 CALGB studies were to be followed for relapse or second malignancy and survival.

## 1.3 Statistical Issues and Findings

### 1.3.1 Major Statistical Issues

1. A retrospective definition of primary endpoint and secondary endpoints, statistical hypothesis, and retrospective collection and analyses of data have been included in this NDA submission. Each of these retrospective specifications in this NDA submission may result in a biased efficacy analysis and conclusion. Based on the guidance International Conference on Harmonisation (ICH)-E9: Statistical Principles for Clinical Trials, these retrospective analyses can only be considered as exploratory and hypothesis generating (Appendix 1).
2. The claim of efficacy is based on a subgroup analysis in patients with MDS and excluding all other patients included in this NDA submission. (ICH-E9: "Any conclusion of treatment efficacy (or lack thereof) or safety based solely on exploratory subgroup analyses is unlikely to be accepted".)
3. The pivotal Study CALGB 9221 was designed to allow subjects randomized to the observation group to crossover to azacitidine treatment after meeting protocol criteria. There were a total of 51 subjects (55.4%; 51/92) who crossed over from the observation group to azacitidine treatment group and only 41 subjects left in the observation group until the end of study. The original randomization is no longer valid due to crossover manner and this may result in a biased efficacy analysis. (ICH-E9: "Randomization introduces a deliberate element of chance into the assignment of treatments to subjects in a clinical trial. During subsequent analysis of the trial data, it provides a sound statistical basis for the quantitative evaluation of the evidence relating to treatment effects. It also tends to produce treatment groups in which the distributions of prognostic factors, known and unknown, are similar.")
4. All 3 CALGB studies were designed as an un-blinded study. Blinding is most definitely needed for good scientific reasons, especially when, as in this clinical testing, there are subjective or semi-objective responses. An open-label study usually results in bias based on the fundamental statistical principles when making measurements. (ICH-E9: "Blinding or masking is intended to limit the occurrence of conscious and unconscious bias in the conduct and interpretation of a clinical trial arising from the influence that the knowledge of treatment may have on the recruitment and allocation of subjects, their subsequent care, the attitudes of subjects to the treatments, the assessment of end-points, the handling of withdrawals, the exclusion of data from analysis, and so on. The essential aim is to prevent identification of the treatments until all such opportunities for bias have passed.")

5. There was a high rate of subjects with major protocol violations in both treatment groups. These violations may influence the efficacy analyses of ITT population.
6. In the design of Study CALGB 9221, the number of treatment cycles was varying for each patient. Therefore, it is questionable that the two groups can be compared with respect to response rate when the therapy is different for each patient within the treatment group.
7. Multiple analyses have been conducted and published using the CALGB 9221 study data. No type I error adjustment for the multiple analyses have been made in this application.
8. The NDA submission did not explain how to interpret the results of statistical comparisons of efficacy variables. In fact, the standard statistical comparison can not be employed in this study and p-values are not interpretable based on the retrospective definitions of primary and secondary endpoints, statistical hypotheses, and retrospective collection of data.
9. Quality of life (QoL) data are not interpretable due to missing observations, protocol violations and varying treatment cycles.

### 1.3.2 Statistical Findings

Table 1 presents the major statistical findings for the efficacy analysis of the response rate for the ITT population with/without AML and/or subjects with protocol violation subjects. For each population, two comparisons: azacitidine (N = 99) versus observation before crossover (N =92) and azacitidine (N = 99) versus observation only (N = 41), have been done. At the significance level of 0.05, the comparisons of overall response (CR+PR) reached the significance level for all populations. However, the comparisons measured as complete response did not reach the significance level for all populations except for the comparison of azacitidine versus observation before crossover for ITT population (p = 0.0294).

Table 1: Analysis of Response Rates --- FDA Analyses

| | Azacitidine | Observation before Crossover | Azacitidine vs. Obs bf X-over | Observation without Crossover | Azacitidine vs. Obs w/o X-over |
|---|---|---|---|---|---|
| All ITT Patients | N = 99 | N = 92 | p-value[a] | N = 41 | p-value |
| Overall (CR+PR) n (%) | 16 (16.2) | 0 (0.0) | <0.0001 | 0 (0.0) | 0.0033 |
| Complete (CR) n (%) | 6 (6.1) | 0 (0.0) | 0.0294 | 0 (0.0) | 0.1802 |
| Partial (PR) n (%) | 10 (10.1) | 0 (0.0) | | 0 (0.0) | |
| ITT Patients without AML | N = 89 | N = 83 | p-value | N = 36 | p-value |
| Overall (CR+PR) n (%) | 14 (15.7) | 0 (0.0) | <0.0001 | 0 (0.0) | 0.0100 |
| Complete (CR) n (%) | 5 (5.6) | 0 (0.0) | 0.0596 | 0 (0.0) | 0.3202 |
| Partial (PR) n (%) | 9 (10.1) | 0 (0.0) | | 0 (0.0) | |
| ITT Patients without AML or protocol violation | N = 54 | N = 48 | p-value | N = 22 | p-value |
| Overall (CR+PR) n (%) | 11 (20.4) | 0 (0.0) | 0.0007 | 0 (0.0) | 0.0276 |
| Complete (CR) n (%) | 5 (9.3) | 0 (0.0) | 0.0585 | 0 (0.0) | 0.3133 |
| Partial (PR) n (%) | 6 (11.1) | 0 (0.0) | | 0 (0.0) | |

[a] P-value from Fisher's exact test by comparing the response rates between the azacitidine group and the observation group.
Obs bf X-over=observation before crossover, Obs w/o X-over=observation without crossover, CR=complete response, PR=partial response, CALGB=Cancer and Leukemia Group B

# 2. INTRODUCTION

## 2.1 Overview

### 2.1.1 Background

Myelodysplastic syndromes (MDS), formerly called pre-leukemia or "smoldering" leukemia, consists of a group of diseases characterized by one or more blood cytopenias secondary to bone marrow dysfunction. These syndromes may arise *de novo* (primary MDS) or occur as the result of chemical injury (petrochemicals, benzene, and rubber), treatment with chemotherapy and/or radiation therapy for other diseases (secondary MDS). Secondary MDS generally has a poorer prognosis than does primary MDS. Myelodysplastic syndromes affect all ages, from children to adults, with the highest prevalence in those over 60 years of age. Although the clinical presentation is generally nonspecific, common initial findings of MDS can be attributed to anemia and include fatigue, weakness, pallor, dyspnea, angina pectoris, and cardiac failure. Easy bruising, ecchymosis, epistaxis, gingival bleeding, petechiae, and bacterial infections, particularly respiratory and dermal, are encountered less frequently. Hepatosplenomegaly may also be present in 10% to 40% of patients. Transformation to acute leukemia occurs in up to 40% of patients with MDS. Bone marrow failure eventually leads to death from bleeding and infection in the majority of patients.

Azacitidine is an anti-metabolite, for which the target indication is the treatment of myelodysplastic syndromes, including all 5 subtypes of the French–American–British (FAB) classification:

- Refractory anemia (RA) requiring transfusions, with thrombocytopenia or significant clinical hemorrhage, or with neutropenia and infection requiring treatment with antibiotics;
- RA with ringed sideroblasts (RARS) requiring transfusions, with thrombocytopenia or significant clinical hemorrhage, or with neutropenia and infection requiring treatment with antibiotics;
- RA with excess blasts (RAEB);
- RAEB in transformation (RAEB-T); and
- Chronic myelomonocytic leukemia (CMMoL).

As a general rule, patients with RA and RARS are considered low-risk patients. The median survival ranges between 3 to 6 years and transformation to acute leukemia is rare. Refractory anemia accounts for 20% to 30% of patients while RARS is less common and seen in 2% to 5% of patients with MDS.

Treatment to address the underlying anemias with cytokine therapy (Epo +/- G-CSF or GMCSF) or other supportive care (red blood cell transfusions) constitutes the usual therapy for patients with these subtypes, although bone marrow transplantation is also considered particularly in younger otherwise healthy patients (< 65 years old).

RAEB and RAEB-T carry a worse prognosis both in terms of median survival and risk for transformation to AML. RAEB is associated with a median survival of usually 6 to 9 months and with progression to AML occurring in 40% of patients. Patients with RAEB-T face an even worse prognosis with a median survival < 6 months and progression to AML occurring in 60% diagnosed with this FAB subtype. Approximately one-third of MDS patients are first diagnosed with RAEB while approximately 25% of patients present with RAEB-T.

As the name implies, the hallmark of CMMoL is an increase in the number of monocytes in the blood. Since monocytes are among the most mature white blood cells produced in the bone marrow, CMMoL can be a slowly progressing disease. However, in some cases it can progress rapidly to acute leukemia. Median survival is about 14 to 18 months. CMMoL accounts for approximately 15% to 20% of MDS.

### 2.1.2 History of Drug Development

MDS is a rare and life-threatening disease for which there is an unmet medical need. Currently there is no single agent or combination therapy used as first-line treatment for MDS. The mainstay of therapy is supportive care, which can include the use of red cell or platelet transfusions, treatment of infections, and use of erythropoietin or growth factors when needed. While hematopoietic growth factor therapy (erythropoietin +/- G-CSF) has emerged as effective therapy for the management of anemia (primarily for patients with RA, RARS with serum erythropoietin levels < 500 U/L), treatment for neutropenia with G-CSF has not been shown to reduce infections and has been associated with shorter overall survival in patients with RAEB.

Azacitidine has been under investigation for over 30 years, with most of the clinical development conducted under the sponsorship of the National Cancer Institute (NCI). Since 1971, approximately 70 clinical trials using azacitidine alone or in combination with chemotherapy regimens have been sponsored by the NCI. Approximately 7500 subjects have been treated in clinical trials and NCI-sponsored compassionate programs such as the special exceptions protocol for MDS. In July 2001, the sponsor, Pharmion Corporation, acquired the marketing rights to azacitidine from Pharmacia.

### 2.1.3 Specific Studies Reviewed

The sponsor, Pharmion Corporation, submitted a New Drug Application (NDA 50-794) for accelerated approval of Vidaza™ (azacitidine for injectable suspension) under the Federal Food, Drug, and Cosmetic Act, 21 CFR§314.50 and 21 CFR§314 Subpart H. This application consists of report of results from registration study CALGB 9221 in the treatment of patients with MDS, supportive data from Study CALGB 8421 and CALGB 8921 in treatment of patients with MDS.

Study CALGB 8421 and Study CALGB 8921 were two Phase II, open-label, uncontrolled, and non-comparative studies, which evaluated azacitidine in subjects with the MDS subtypes of RAEB and RAEB-T, and CMMoL in CalGB 8921 only. Since CALGB 8421 and CALGB 8921 were two non-comparative studies and the patient populations were different from in Study

CALGB 9221, the study selected for the full statistical review and evaluation is only the controlled pivotal Study CALGB 9221.

### 2.2 Data Sources

Data used for review is from the electronic submission received on December 26, 2003. The efficacy analysis data were submitted by the sponsor on December 26, 2003. All data sets analyzed are electronic documents and are located in the Electronic Document Room (EDR) of CDER of FDA under the Letter Date "26-DEC-2003". The data sets analyzed in this NDA review are located in the folder of CRT\datasets\9221. The major data sets for the efficacy analyses are "RESPONSE", "TRANS", "VIEW_EOS", and "VIEW_MON" which defined the responses, events, time to relapse after response, and red blood cell transfusions.

## 3. STATISTICAL EVALUATION

### 3.1 Evaluation of Efficacy

#### 3.1.1 Study Design and Endpoints

CALGB 9221 was a Phase 3, open-label, multi-center controlled study designed to compare azacitidine plus supportive care with an observation group receiving best supportive care in subjects with any of the 5 subtypes of MDS conducted during 1994 to 2002. In this study, up to 174 subjects were initially planned for enrollment over 25 months. A total of 191 subjects were enrolled into the study with 99 subjects in azacitidine arm and 92 subjects in the observation arm. Subjects with all 5 subtypes of MDS were to be balanced between the study's two treatment arms. Subjects in the azacitidine treatment arm were to receive 75 $mg/m^2$ azacitidine SC, daily for 7 days on a 28-day cycle for a minimum of 4 cycles. The azacitidine dose could be adjusted (either increased, decreased, or delayed) at the beginning of any cycle based on predefined hematology and renal laboratory results relating to the well-being of the subject.

Additionally, the CALGB 9221 study was designed to allow subjects randomized to the observation group to switch (treatment crossover) to azacitidine treatment after meeting protocol criteria for increases in bone marrow blasts, decreases in hemoglobin or platelets, increased RBC or platelet transfusion requirements, or clinical infection with low ANC and requiring treatment with intravenous antibiotics.

The original protocol of the CALGB 9221 study defined the primary objectives as to determine the response rate to azacitidine and the impact of azacitidine on RBC transfusion requirements, platelet counts, ANC, rates of infection and hemorrhage and % BM blasts, in comparison to an untreated observation group. It also defined the response as either Complete Response (CR), Partial Response (PR), or Improvement.

In this NDA submission of Study CALGB 9221, statistical findings are based on analyses of the retrospectively collected data. In these analyses, overall response (CR + PR) is considered as the only primary endpoint for the CALGB 9221 study. The following variables were classified as secondary endpoints in the revised protocol for CALGB 9221: survival, time to transformation to AML, time to death or transformation to AML, changes in RBC and platelet transfusions, hemoglobin concentrations, ANC, white blood cell (WBC) and platelet counts, rates of infection (as assessed by courses of antibiotic therapy) and hemorrhage, and percent of bone marrow blasts. Responses were assessed based on bone marrow and peripheral blood evaluations performed by the site.

*Reviewer's Comments:*

1) The CALGB 9221 study was designed to allow subjects randomized to the observation group to crossover to azacitidine treatment after meeting protocol criteria for increases in bone marrow blasts, decreases in hemoglobin or platelets, increased RBC or platelet transfusion requirements, or clinical infection with low ANC and requiring treatment with intravenous antibiotics. There were totally 51 subjects (55.4%; 51/92) who crossed over from the observation group to azacitidine treatment group and only 41 subjects left in the observation group until the end of study. The original randomization is no longer valid due to crossover and this may result in biased efficacy analysis.
2) CALGB 9221 was designed as an un-blinded study. Blinding is most definitely needed for good scientific reasons, especially when, as in this clinical testing, there are subjective or semi-objective responses. An open-label study usually results in bias based on the fundamental statistical principles when making measurements. It is not possible to determine if open-label nature of this study resulted in high proportion of crossover.
3) In the design of CALGB 9221, the number of treatment cycles was varying for each patient. Therefore, it is questionable that the two groups can be compared with respect to response rate when the therapy is different for each patient within the treatment group.
4) This NDA submission included retrospective definitions for the primary and secondary endpoints, statistical hypotheses, and retrospective collection and analyses of data. Statistically, any retrospective definition in the design and/or data collection will result in bias. Therefore, the retrospective definition of primary and secondary endpoints, statistical hypotheses and data collection may result in a biased efficacy outcome.
5) The original protocol defined multiple primary endpoints. A multiple comparison adjustment would have been necessary in the efficacy analyses based on the original protocol. However, there is only one primary endpoint in the sponsor's retrospective definition of primary endpoint.
6) There was no adjustment of type I error considered for multiple analyses conducted.

### 3.1.2 Patient Dispositions, Demographic and Baseline Characteristics

There were 191 subjects enrolled in the study with 99 subjects in azacitidine treatment arm and 92 subjects in the observation arm, respectively. There were 51 subjects who crossed over from the observation arm to azacitidine treatment arm and 41 subjects left in the observation arm until to the end of study. Figure 1 shows the patient population disposition.

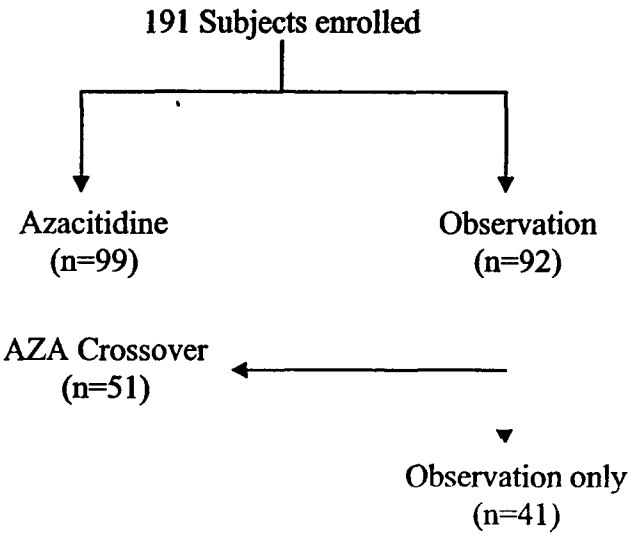



Figure 1. Patient Population Disposition

Table 2 presents the subject demographics. Patients enrolled in CALGB 9221 were representative of the MDS patient population with respect to demographics. Of the 191 subjects enrolled and randomized, 69% (132) were male and 31% (59) were female. The proportion of females was somewhat lower in the azacitidine treatment arm (27%) than in the observation arm (35%), but the proportion in the all azacitidine group (31%; 47/150) was the same as the overall percentage (31%: 59/191) for all subjects. The two randomized treatment groups were similar with respect to race and age: over 90% of the subjects were white and the mean age was about 67 years (range 31 to 92 years). At baseline, the demographics of the azacitidine and observation groups were comparable.

Table 2: Subject Demographic Characteristics[a]

| *Demographic* | *Azacitidine (N=99)* | *Observation (N=92)* | *Crossed over to Azacitidine (N=51)* | *All Azacitidine (N=150)* |
|---|---|---|---|---|
| **Gender (n%)** | | | | |
| Male | 72 (72.7) | 60 (65.2) | 31 (60.8) | 103 (68.7) |
| Female | 27 (27.3) | 32 (34.8) | 20 (39.2) | 47 (31.3) |
| **Race (n%)** | | | | |
| White | 93 (93.9) | 85 (92.4) | 47 (92.2) | 140 (93.3) |
| Black | 1 (1.0) | 1 (1.1) | 1 (2.0) | 2 (1.3) |
| Hispanic | 3 (3.0) | 5 (5.4) | 2 (3.9) | 5 (3.3) |
| Asian/Oriental | 2 (2.0) | 1 (1.1) | 1 (2.0) | 3 (2.0) |
| **Age (n%)** | | | | |
| N | 99 | 91 | 51 | 150 |
| Mean ± SD | 67.3 ± 10.39 | 68.0 ± 10.23 | 67.0 ± 9.33 | 67.2 ± 10.02 |
| Range | 31 – 92 | 35 – 88 | 46 – 82 | 31 – 92 |
| **Age Group (n%)** | | | | |
| Missing | 0 (0.0) | 1 (1.1) | 0 (0.0) | 0 (0.0) |
| ≤ 64 | 36 (36.4) | 33 (35.9) | 17 (33.3) | 53 (35.3) |
| 65 – 74 | 39 (39.4) | 33 (35.9) | 22 (43.1) | 61 (40.7) |
| ≥ 75 | 24 (24.2) | 25 (27.2) | 12 (23.5) | 36 (24.0) |

[a] The sponsor's analyses verified by the statistical reviewer.

Baseline disease characteristics are summarized in Table 3. For this retrospective analysis, baseline diagnosis was adjudicated based on the site, central, or blinded independent review of bone marrow and peripheral blood assessments. Baseline bone marrow slides were reviewed locally for 99% (190/191) of subjects, and centrally for 98% (188/191) of subjects. At the time of retrospective re-collection of data, baseline bone marrow slides for review by the blinded independent reviewer were available for 88 subjects (43% in the azacitidine arm [43/99] and 49% in the observation arm [45/92]). Peripheral blood slides were available for some of these 88 subjects. In cases in which these were not available, the baseline CBC laboratory report was assessed. If the baseline diagnosis by site, central, or blinded review was AML, this diagnosis was assigned to the subject. As shown in Table 3, a single diagnosis of AML at baseline was made by the sites (in a subject randomized to the observation arm). Adjudication by central or blinded review resulted in a baseline diagnosis of AML for 19 subjects, which included the one subject diagnosed with AML by the site. Similar proportions of diagnosis of AML were seen between the azacitidine (10%; 10/99) and observation (10%; 9/92) groups. In the independent blinded review of the 88 subjects with baseline slides for the diagnosis of MDS vs. non-MDS, 83% (73/88) concordance was achieved with the site. Of the baseline slides for the 88 subjects, the central pathology reviewer was able to assess slides for 86 subjects. Concordance between the independent blinded review and the central review was 88% (76/86).

The distribution of MDS subtype based on adjudicated baseline diagnosis was similar for the azacitidine and observation groups. Overall, approximately 40% of subjects had RAEB (azacitidine: 38%; observation: 42%), approximately 20% of each group had RA, and approximately 15% of each group had RAEB-T.

Table 3: Baseline Disease Characteristics[a]

| *Baseline Disease Characteristic* | *Number (%) of Subjects* | | | |
|---|---|---|---|---|
| | Azacitidine (N=99) | Observation (N=92) | Obs Crossed over to Azacitidine (N=51) | All Azacitidine (N=150) |
| Site MDS diagnosis at study entry | | | | |
| RA | 21 (21.2) | 18 (19.6) | 14 (27.5) | 35 (23.3) |
| RARS | 6 (6.1) | 5 (5.4) | 3 (5.9) | 9 (6.0) |
| RAEB | 42 (42.4) | 44 (47.8) | 25 (49.0) | 67 (44.7) |
| RAEB-T | 22 (22.2) | 17 (18.5) | 7 (13.7) | 29 (19.3) |
| CMMoL | 8 (8.1) | 7 (7.6) | 2 (3.9) | 10 (6.7) |
| AML | 0 (0.0) | 1 (1.1) | 0 (0.0) | 0 (0.0) |
| Adjudicated MDS diagnosis at study entry | | | | |
| RA | 21 (21.2) | 18 (19.6) | 14 (27.5) | 35 (23.3) |
| RARS | 6 (6.1) | 5 (5.4) | 3 (5.9) | 9 (6.0) |
| RAEB | 38 (38.4) | 39 (42.4) | 23 (45.1) | 61 (40.7) |
| RAEB-T | 16 (16.2) | 14 (15.2) | 5 (9.8) | 21 (14.0) |
| CMMoL | 8 (8.1) | 7 (7.6) | 2 (3.9) | 10 (6.7) |
| AML | 10 (10.1) | 9 (9.8) | 4 (7.8) | 14 (9.3) |
| Performance Status | | | | |
| 0 Normal | 35 (35.4) | 26 (28.3) | 13 (25.5) | 48 (32.0) |
| 1 Fatigue | 34 (34.3) | 39 (42.4) | 24 (47.1) | 58 (38.7) |
| 2 Impaired | 8 (8.1) | 6 (6.5) | 2 (3.9) | 10 (6.7) |
| 3 Bedrest | 1 (1.0) | 0 (0.0) | 0 (0.0) | 1 (0.7) |
| Unknown/Not Done | 21 (21.2) | 21 (22.8) | 12 (23.5) | 33 (22.0) |
| Transfusion Product used in 3 months before study entry: | | | | |
| Any transfusion product | 70 (70.7) | 59 (64.1) | 36 (70.6) | 106 (70.7) |
| Blood cells, packed human | 66 (66.7) | 55 (59.8) | 34 (66.7) | 100 (66.7) |
| Hetastarch | 0 (0.0) | 1 (1.1) | 1 (2.0) | 1 (0.7) |
| Plasma protein fraction (human) | 1 (1.0) | 0 (0.0) | 0 (0.0) | 1 (0.7) |
| Platelets, human blood | 15 (15.2) | 12 (13.0) | 5 (9.8) | 20 (13.3) |
| Unknown | 2 (2.0) | 2 (2.2) | 1 (2.0) | 3 (2.0) |

[a] The sponsor's analyses verified by the statistical reviewer.

Major protocol violations are summarized in Table 4. The incidence of subjects with at least one major protocol violation during the study was similar between the azacitidine (36%), observation (43%), crossed over to azacitidine (39%), and all azacitidine (37%) groups. The most frequent major protocol violation was the use of hematopoietic growth factors at any time or use of systemic steroids in the month before or during the study, which occurred in about one-third of both randomized groups. For both randomized groups, the majority of the instances of this major protocol violation were due to the use of systemic steroids to treat TEAEs such as bone pain, rash, hives, and infections. The second most common reason for this major protocol violation in both groups resulted from use of systemic steroids for prophylaxis or treatment of transfusion reactions. The remainder of the major protocol violations in this category were due to use of hematopoietic growth factors, and use was similar in both groups.

Failure to meet the eligibility criterion of an established diagnosis of MDS at study entry was observed at similar frequencies in the azacitidine (3%), observation (2%), crossed over to

azacitidine (2%), and all azacitidine (3%) groups. Three subjects in the azacitidine group did not meet the eligibility criterion for the following reasons; Subject 55411 had a prior history of leukemia. Subject 56518 had a site diagnosis of RAEB-T but had myeloblasts of 14.9%, less than the 20% required for this FAB category; Subject 58911 had a site diagnosis of RA but no bone marrow biopsy or aspirate was found despite progress notes indicating the procedure was performed 2 weeks prior to randomization. Two subjects in the observation group did not meet the eligibility criterion for the following reasons: Subject 59106 and Subject 60264 both had a site diagnosis of RAEB-T, but both had > 30% blasts, greater than the cutoff for this FAB category. Subject 60264 later crossed over to azacitidine treatment.

Table 4: Major Protocol Violations[a]

| *Major Protocol Violation* | *Number (%) of Subjects* | | | |
|---|---|---|---|---|
| | Azacitidine (N=99) | Observation (N=92) | Obs Crossed over to Azacitidine (N=51) | All Azacitidine (N=150) |
| Subjects with at least one Major Protocol Violation | 36 (36.4) | 40 (43.5) | 20[d] (39.2) | 56 (37.3) |
| Subject did not receive treatment as randomized, received the wrong treatment, or crossed over from observation to azacitidine without meeting the crossover criteria | 0 (0.0) | 9[e] (9.8) | 9[d, e] (17.7) | 9 (6.0) |
| Subject did not meet the eligibility criterion of an established diagnosis of MDS[b] | 3 (3.0) | 2 (2.2) | 1[d] (2.0) | 4 (2.7) |
| Subject took hematopoietic growth factors at any time, or systemic steroids in the month before or during the study[c] | 34 (34.3) | 32 (34.8) | 12[d] (23.5) | 46 (30.7) |
| Subject received prior cytotoxic therapy for this marrow disease within the past 6 months, or | 0 (0.0) | 2 (2.2) | 0 (0.0) | 0 (0.0) |
| received prior treatment with azacitidine | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
| Subject with documentation stating that they did not receive azacitidine immediately after its reconstitution | | | | |

[a] The sponsor's analyses verified by the statistical reviewer.

[b] Subjects with a prior history of leukemia are included in this category of major protocol violation.

[c] Although use of erythropoietin before study enrollment was permitted in the protocol, use of hematopoietic growth factors at any time prior to study enrollment or during the study was defined retrospectively as a major protocol violation.

[d] Subjects counted in the azacitidine after observation group are also counted in the group of all subjects randomized to observation.

[e] All subjects counted were due to the reason of crossover without meeting crossover criteria.

*Reviewer's Comments:*

There was a high rate of subjects with major protocol violations in each group. Though the sponsor explained that the most frequent major protocol violation was the use of hematopoietic growth factors at any time or use of systemic steroids in the month before or during the study, which occurred in about one-third of both randomized groups, it is likely to influence the efficacy results of intern-to-treat population as the study population has been changed from the protocol defined population.

### 3.1.3 Statistical Methodologies

The methods of data analysis of efficacy endpoints were not described in each of the CALGB protocols. The sponsor developed a statistical analysis plan (SAP) for the studies and finalized before databases were locked, based on the newly created databases from the retrospective collection of data from each of these completed studies. The sponsor proposed their SAP for CALGB 9221 as follows.

Descriptive statistics of the data will include counts, means, standard deviations, minimum and maximum for continuous data, and frequencies and percentages for categorical data.

Since this analysis plan has been written retrospectively with regard to the closing, original analysis, and reporting of the study data, caution must be used in the interpretation of the results. Therefore, for all analyses except the primary assessment of efficacy, the emphasis will be on descriptive summaries and the results of formal statistical hypothesis tests will be considered as exploratory only.

Testing of interactions will be conducted at the $\alpha$=0.10 level. All other testing will be conducted at the two-sided $\alpha$=0.05 level and 95% confidence intervals will be employed, unless otherwise specified. All data analyses will be conducted using SAS Version 8.0 software.

Baseline is defined as the measure taken prior to randomization. Changes from baseline will be calculated as the later value minus the baseline value.

Missing individual data will be treated as missing, and no imputation methods will be used unless otherwise specified. Calculated parameters will be based on non-missing individual data unless otherwise specified for specific parameters. Summaries of categorical data will indicate the number of missing observations.

Time of response data for all unobserved events will be considered right-censored immediately after the last assessment for non-responders, and immediately after the time of loss to follow-up for non-responders lost to follow-up.

Data summaries and analyses will be pooled across sites.

Laboratory measurements which are marked as repeated will be used for data summaries in place of the original routine measurement. All results, whether routine or repeat, will be listed.

The ITT population was to include all subjects randomized in the study and to be used in summaries and analyses of efficacy data.

The original protocol of CALGB 9221 defined response as either CR, PR, or Improvement. However, the primary endpoint in the NDA submission was defined as overall response rate for the analyses of the retrospectively collected data, in which response was defined as CR or PR, with the best response attained during the study used to categorize each subject. Separate

response rates were to be computed for CR, PR, and overall response (CR or PR). Response rates were to be presented for all subjects and were to be presented for all subjects in the ITT population by treatment group, and by MDS subtype and treatment group. In addition, response rates were also to be presented for the specific groups of subjects who crossed over to azacitidine from observation, for subjects who did not have a protocol violation, and for subjects who did not have a baseline diagnosis of AML or an adjudicated baseline diagnosis of AML.

In CALGB 9221, analyses of response rates in the azacitidine group were compared to the observation group. These were to be calculated as the number of subjects in a treatment arm exhibiting each type of response divided by the number randomized to that treatment arm. The comparison of the overall response rate of CR or PR was to be the primary analysis. Response rates from the two groups were to be compared using a multivariable logistic regression model, adjusting for the baseline MDS diagnosis. Because there was a zero response rate in the observation group, the log transformation could not be performed and a simpler Fisher's Exact Test was used to compare treatment groups.

*Reviewer's Comments:*

1) The SAP did not define any analysis method for crossover effect. The SAP also did not define any procedure to adjust the statistical comparisons of efficacy variables. Since there were 55% subjects who crossed over from the observation arm to azacitidine treatment arm, method to adjust the crossover effect is a major statistical issue in this study. This statistical reviewer believes that a sensitivity analysis with/without the crossed over subjects is necessary to evaluate the crossover effect.
2) The SAP did not define any statistical method for the changes in red blood cell (RBC) transfusions, platelet counts, absolute neutrophil count (ANC), rates of infection and hemorrhage, and percent of bone marrow blasts which were defined as the joint primary endpoints in the original protocol.
3) The SAP did not define any method to deal with the protocol violations. Since there was a high rate ($\geq$ 36%) of subjects with major protocol violations in each group, this statistical reviewer believes that a sensitivity analysis with/without the protocol violations is necessary to evaluate the protocol violation effect.
4) The original protocol defined multiple primary endpoints. But the SAP did not define a multiple comparison adjustment to the efficacy analyses.
5) The SAP did not explain how to interpret the results of statistical comparisons of efficacy variables. In fact, the usual statistical comparison can not be employed in this study and p-values are not interpretable based on the retrospective definitions of primary and secondary endpoints, statistical hypotheses, and retrospective collection of data.
6) The SAP did explain that "for all analyses except the primary assessment of efficacy, the emphasis will be on descriptive summaries and the results of formal statistical hypothesis tests will be considered as exploratory only". But the SAP did not explain how the primary assessment of efficacy is not exploratory. Based on major statistical issues stated previously, the whole study would be exploratory in nature.

### 3.1.4 Sponsor's Results

*Overall Response*

The primary endpoint for CALGB 9221 was defined as the overall response rate of CR + PR. The overall response rate and the best response rates of all subjects are shown in Table 5. The best response attained during the study was used to categorize subjects randomized to azacitidine. For observation subjects, the best response during the observation before crossover period, and also best response for subsets who did not crossover and for subjects who did crossover to azacitidine treatment are presented.

Table 5: Analysis of Response Rates of All Subjects[a]

| | **Azacitidine (N=99)** | **Observation before Crossover (N=92)** | | **Observation without Crossover (N=41)** | **Crossed over to Azacitidine (N=51)** | **All Azacitidine (N=150)** |
|---|---|---|---|---|---|---|
| Response | n (%) | n (%) | p-value[b] | n (%) | n (%) | n (%) |
| Overall (CR+PR) | 16 (16.2) | 0 (0.0) | <0.0001 | 0 (0.0) | 6 (11.8) | 22 (14.7) |
| Complete (CR) | 6 (6.1) | 0 (0.0) | 0.0294 | 0 (0.0) | 3 (5.9) | 9 (6.0) |
| Partial (PR) | 10 (10.1) | 0 (0.0) | | 0 (0.0) | 3 (5.9) | 13 (8.7) |
| Improvement, not CR or PR | 33 (33.3) | 18 (19.6) | | 5 (12.2) | 17 (33.3) | 50 (33.3) |
| Stable Disease | 40 (40.4) | 51 (55.4) | | 16 (39.0) | 17 (33.3) | 57 (38.0) |
| Relapse | 0 (0.0) | 0 (0.0) | | 0 (0.0) | 1 (2.0) | 1 (0.7) |
| Disease Progression | 2 (2.0) | 10 (10.9) | | 8 (19.5) | 2 (3.9) | 4 (2.7) |
| Unevaluable | 8 (8.1) | 13 (14.1) | | 12 (29.3) | 8 (15.7) | 16 (10.7) |

[a] The sponsor's analyses verified by the statistical reviewer.

[b] P-value from Fisher's exact test by comparing the response rates between the azacitidine group and the observation before crossover group.

CR=complete response, PR=partial response, CALGB=Cancer and Leukemia Group B

According the sponsor's efficacy conclusion, of the 99 subjects randomized to azacitidine, the overall response rate was 16.2% (16/99); this overall response rate was statistically significantly higher than the response rate for the observation group (0%; p<0.0001). Complete response was achieved by 6.1% (6/99) of azacitidine subjects and this rate was also significantly higher for the azacitidine group compared to observation (p=0.0294). In addition, 10.1% (10/99) of azacitidine subjects achieved PR. The time to onset of response for PR occurred from the 2$^{nd}$ to 19$^{th}$ treatment cycles of the 10 subjects achieving PR (range 51 to 575 days on study). As would be expected given the requirement for complete normalization of the bone marrow, it took more time to see the onset of CR; CR was reported from the 8$^{th}$ to 15$^{th}$ treatment cycle of the 6 azacitidine subjects who achieved CR, corresponding to a range of 197 to 431 days on study.

The majority of subjects who were assessed as CR or PR had either 2 or 3 cell line abnormalities at baseline (81%; 13/16). In addition, all subjects except one (Subject 60118) had elevated bone marrow blasts or were transfusion dependent at baseline. By definition, subjects with a CR or PR achieved the response for all peripheral blood cell lines and bone marrow blasts, which were identified as abnormal at baseline. An additional 6 subjects treated with azacitidine after observation achieved CR or PR resulting in an overall response of 14.7% (22/150) for all

azacitidine. The range of onset of CR or PR in this group of subjects was similar to the ranges seen for subjects randomized to azacitidine (CR: 5th to 11th cycle; PR: 2nd to 21st cycle).

*Duration of Clinical Response*

An assessment of the responders gives some insight into the duration of effect of azacitidine as well as serves as the basis for determining the validity of the response criteria. For the azacitidine responders, the initial positive effect in either bone marrow or peripheral blood was seen as early as Day 29 (Cycle 1) and 75% (12/16) of responders showed an initial positive effect within 4 cycles (~4 months) of treatment. For 3 of the remaining 4 responders, this effect was seen after 5 to 6 months (cycles); one subject showed the initial positive effect as late as Day 477 (Cycle 17). Overall, 94% (15/16) demonstrated evidence of a response by 6 cycles. Of the 9 subjects with elevated bone marrow blasts at baseline, this initial positive response was noted in the bone marrow with a reduction in bone marrow blasts observed in 67% (6/9) of patients. For 2 of the remaining 3 subjects with elevated baseline bone marrow blasts, initial positive effect was seen in peripheral blood before a post-baseline bone marrow aspirate was performed.

The mean and median duration of positive effect was 418.9 days and 284 days, respectively. The mean and median total duration of clinical response of PR or better for the 16 responders was 224.8 days and 165.5 days, respectively. The actual durations of positive effect or response may have been artificially shortened for several subjects due to no further values available to document ongoing results after subjects withdrew from the study. In part, this was due to the protocol specified removal of subjects achieving CR after 3 additional cycles of treatment.

*Other Response Assessments: Improvement*

Subjects who met some but not all the necessary criteria for partial response were assessed as Improvement if, for a minimum of 4 weeks at least one peripheral blood cell line showed $\geq$ 50% restoration in the deficit at baseline for $\geq$ 4 weeks, or if there was a $\geq$ 50% decrease in RBC or platelet transfusion requirement.

Per the sponsor, a best response assessment of Improvement was recorded for 33.3% (33/99) of azacitidine subjects and 12.2% (5/41) of observation without crossover subjects. In the azacitidine group, improvement was observed across all MDS subtypes and adjudicated diagnosis of AML at baseline. Rates of positive change from abnormal baseline values in subjects with a best response of Improvement are presented in Table 6. However, FDA clinical reviewer has a different assessment of improvement which is slightly different than the sponsor's assessment. Please see the clinical review of this application.

Table 6: Subjects with Best Response of Improvements[a]

| Peripheral Blood | Abnormal at Baseline | | Positive Change from Abnormal Baseline | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | ≥50% Increase[b] and Transfusion Free | | <50% Increase[b] but Transfusion Free | | All Positive Change from Baseline | |
| | Azacitidine | Obs w/o crossover | Azacitidine | Obs w/o crossover | Azacitidine | Obs w/o crossover | Azacitidine | Obs w/o crossover |
| Hemoglobin | | | | | | | | |
| n (%) | 32 (97.0) | 5 (100) | 11 (34.4) | 0 (0.0) | 8 (25.0) | 0 (0.0) | 19 (59.4) | 0 (0.0) |
| (n/N) | (32/33) | (5/5) | (11/32) | (0/5) | (8/32) | (0/5) | (19.32) | (0/5) |
| Platelets | | | | | | | | |
| n (%) | 28 (84.8) | 5 (100) | 12 (42.9) | 2 (40.0) | 10 (35.7) | 2 (40) | 22 (78.6) | 3 (80.0) |
| (n/N) | (28/33) | (5/5) | (12/28) | (2/5) | (10/28) | (2/5) | (22/28) | (4/5) |
| WBC[c] | | | | | | | | |
| n (%) | 24 (72.7) | 3 (60.0) | 5 (20.8) | 2[d] (66.7) | 13 (54.2) | 1 (33.3) | 15 (75.0) | 1[d] (33.3) |
| (n/N) | (24/33) | (3/5) | (5/24) | (2/3) | (13/24) | (1/3) | (18/24) | (1/3) |

[a] The sponsor's analyses verified by the statistical reviewer.

[b] Increase for ≥ 4 weeks.

[c] No subjects had WBC transfusions.

[d] For 2 subjects WBC count increased to above the upper limit of normal. These 2 subjects are not included in the 'All Positive Change from baseline' total.

Obs w/o crossover = Observation without crossover

*Impact on Survival and Time to Transformation to AML*

Secondary analyses were performed for time to AML progression (log-rank p-value=0.2211), time to death (log-rank p-value=0.6064), time to disease progression (log-rank p-value=0.0008), and time to relapse after response (log-rank p-value=0.3649) for all ITT patients. In general, it appears that subjects treated with azacitidine, independent of response, had a longer time to death, disease progression, or transformation to AML. Time to AML progression, time to death, time to disease progression, and time to relapse after response for all ITT patients are presented in Figure 2 to Figure 5, respectively.

Figure 2 . Kaplan–Meier Curve of Time to AML Progression

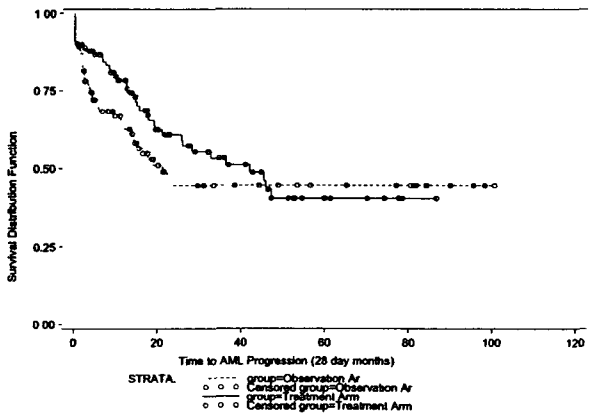
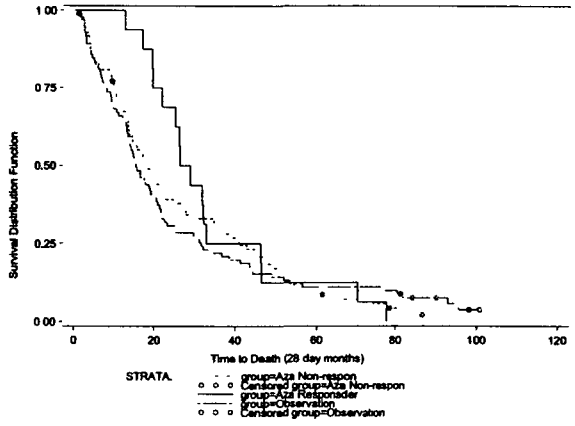
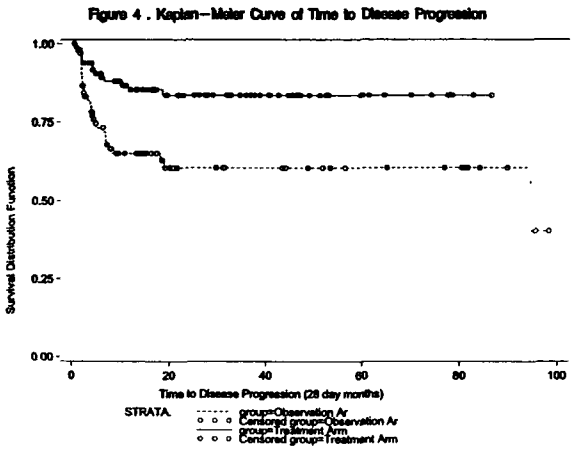
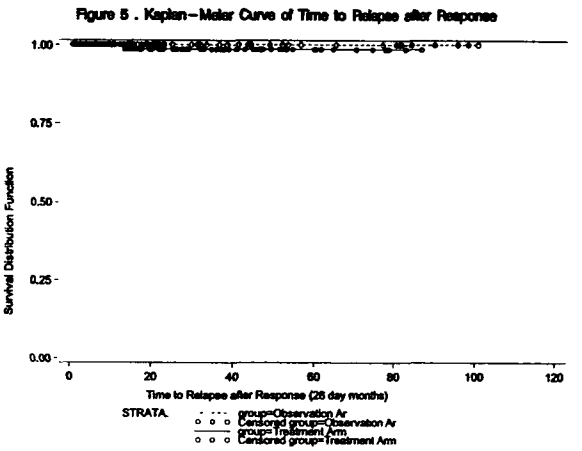



Figure 3 . Kaplan–Meier Curve of Time to Death (AZA Responders vs AZA Non–responders vs OBS)






While not achieving statistical significance, there was a trend for azacitidine treated subjects to have an increase in time to transformation to AML compared to observation subjects. This was most pronounced in the subtypes with the highest risk for transformation. For subjects with RAEB, median time to transformation to AML was 47.2 months for azacitidine group. For RAEB-T, the median time to transformation was 19.2 months for the azacitidine group compared to 4.2 months for observation group.

### 3.1.5 Reviewer's Results based on Exploratory Analyses

For the study CALGB 9221, if the baseline diagnosis by the site, central, or blinded independent review was AML, this diagnosis was assigned to the subject. Since the primary objective of the study CALGB 9221 was to compare the azacitidine group and observation group with respect to efficacy and safety for the 5 subtypes of MDS: RA, RARS, RAEB, RAEB-T, and CMMoL, the efficacy analysis of the primary endpoint should not include subjects with AML.

Since there was a higher rate ($\geq$ 36%) of subjects with major protocol violations in each group, a sensitivity analysis with/without the protocol violations is necessary to evaluate the protocol violation effect.

Finally, there were 55% subjects who crossed over from the observation arm to azacitidine treatment arm. To evaluate the crossover effect, a sensitivity analysis with/without the crossover for the primary endpoint is necessary. Furthermore, the original protocol clearly defined the data analysis as "data obtained after crossover will not be used in the primary analyses".

Table 7 presents the major statistical findings for the efficacy analysis of the response rate for the ITT population with/without AML and/or protocol violation subjects. For each population, two comparisons: azacitidine (N = 99) versus observation before crossover (N =92) and azacitidine (N = 99) versus observation only (N = 41), have been done. At the significance level 0.05, the comparisons of overall response (CR+PR) reached the significance level for all populations. However, the comparisons with efficacy measured as CR did not reach the significance level for

all populations except for the comparison of azacitidine versus observation before crossover for ITT population (p-value=0.0294).

Table 7: Analysis of Response Rates --- FDA Analyses

| | Azacitidine | Observation before Crossover | Azacitidine vs. Obs bf X-over | Observation without Crossover | Azacitidine vs. Obs w/o X-over |
|---|---|---|---|---|---|
| All ITT Patients | N = 99 | N = 92 | p-value[a] | N = 41 | p-value |
| Overall (CR+PR) n (%) | 16 (16.2) | 0 (0.0) | <0.0001 | 0 (0.0) | 0.0033 |
| Complete (CR) n (%) | 6 (6.1) | 0 (0.0) | 0.0294 | 0 (0.0) | 0.1802 |
| Partial (PR) n (%) | 10 (10.1) | 0 (0.0) | | 0 (0.0) | |
| ITT Patients without AML | N = 89 | N = 83 | p-value | N = 36 | p-value |
| Overall (CR+PR) n (%) | 14 (15.7) | 0 (0.0) | <0.0001 | 0 (0.0) | 0.0100 |
| Complete (CR) n (%) | 5 (5.6) | 0 (0.0) | 0.0596 | 0 (0.0) | 0.3202 |
| Partial (PR) n (%) | 9 (10.1) | 0 (0.0) | | 0 (0.0) | |
| ITT Patients without AML or protocol violation | N = 54 | N = 48 | p-value | N = 22 | p-value |
| Overall (CR+PR) n (%) | 11 (20.4) | 0 (0.0) | 0.0007 | 0 (0.0) | 0.0276 |
| Complete (CR) n (%) | 5 (9.3) | 0 (0.0) | 0.0585 | 0 (0.0) | 0.3133 |
| Partial (PR) n (%) | 6 (11.1) | 0 (0.0) | | 0 (0.0) | |

[a] P-value from Fisher's exact test by comparing the response rates between the azacitidine group and the observation group.
Obs bf X-over=observation before crossover, Obs w/o X-over=observation without crossover, CR=complete response, PR=partial response, CALGB=Cancer and Leukemia Group B

The original protocol also defined the changes in RBC transfusions, platelet counts, ANC, rates of infection and hemorrhage, and percent of bone marrow blasts as the joint primary efficacy endpoints. This statistical reviewer performed two transfusion change analyses for two major transfusions: RBC and platelet transfusions required. The first transfusion change analysis is the monthly transfusion change analysis which is defined as the change of transfusion required per month from baseline to post-baseline. The monthly transfusion requirement of baseline is defined as the transfusion requirement in the baseline month (normalized to a 28 days month) before the treatment randomization. The monthly transfusion requirement of post-baseline is defined as the average transfusion requirement after the treatment randomization (normalized each month to a 28 days month). The second transfusion change analysis is the transfusion change per cycle analysis which is defined as the change of transfusion requirement per treatment cycle from baseline to post-baseline. .

Table 8 and 9 present the results of two transfusion change analyses, respectively. At the significance level 0.05 with Bonferroni adjustment, for the monthly transfusion change, the comparisons of azacitidine group versus observation before crossover group and azacitidine group versus observation without crossover group did not reach the significance level. For the transfusion change per cycle, the comparisons of azacitidine group versus crossed over to azacitidine group also did not reach the significance level.

Table 8: Analysis of Transfusions Changed per Month from Baseline --- FDA Analyses

| | Azacitidine (N = 99) | Observation before Crossover (N = 92) | Observation without Crossover (N = 41) |
|---|---|---|---|
| **Number of RBC Transfusion Changed per Month from Baseline** | | | |
| n[a] | 96 | 86 | 38 |
| Mean (SD) | -3.50 (8.9) | -2.44 (7.5) | -2.44 (5.6) |
| Range | | | |
| p-value[b] (compare azacitidine vs. observation before crossover) | | 0.469 | |
| p-value (compare azacitidine vs. observation without crossover) | | | 0.886 |
| **Units of RBC Transfusion Changed per Month from Baseline** | | | |
| n | 96 | 86 | 38 |
| Mean (SD) | -6.05 (13.7) | -4.80 (15.7) | -5.11 (13.9) |
| Range | | | |
| p-value (compare azacitidine vs. observation before crossover) | | 0.626 | |
| p-value (compare azacitidine vs. observation without crossover) | | | 0.729 |
| **Number of Platelet Transfusion Changed per Month from Baseline** | | | |
| n | 96 | 86 | 38 |
| Mean (SD) | -0.62 (7.8) | -0.84 (7.7) | -2.20 (11.4) |
| Range | | | |
| p-value (compare azacitidine vs. observation before crossover) | | 0.047 | |
| p-value (compare azacitidine vs. observation without crossover) | | | 0.075 |
| **Units of Platelet Transfusion Changed per Month from Baseline** | | | |
| n | 96 | 86 | 38 |
| Mean (SD) | -0.58 (11.8) | -5.48 (48.3) | -12.88 (72.3) |
| Range | | | |
| p-value (compare azacitidine vs. observation before crossover) | | 0.322 | |
| p-value (compare azacitidine vs. observation without crossover) | | | 0.214 |

[a] There were 3 and 6 patients in the azacitidine group and the observation group having no transfusions data, respectively.
[b] P-value from Wilcoxon test with normal approximation by comparing the azacitidine group to the observation group.

Table 9: Analysis of Transfusion Changed per Cycle from Baseline --- FDA Analyses

| | Azacitidine (N = 99) | | Crossed Over to Azacitidine (N = 51) |
|---|---|---|---|
| **Number of RBC Transfusion Changed per Cycle from Baseline** | | | |
| n[a] | 96 | | 48 |
| Mean (SD) | -3.40 (9.0) | | -2.06 (8.9) |
| Range | | | |
| p-value[b] (compare azacitidine vs. crossed over to azacitidine) | | 0.046 | |
| **Units of RBC Transfusion Changed per Cycle from Baseline** | | | |
| n | 96 | | 48 |
| Mean (SD) | -5.85 (13.8) | | -3.95 (17.5) |
| Range | | | |
| p-value (compare azacitidine vs. crossed over to azacitidine) | | 0.056 | |
| **Number of Platelet Transfusion Changed per Cycle from Baseline** | | | |
| n | 96 | | 48 |
| Mean (SD) | -0.55 (7.9) | | 0.85 (2.1) |
| Range | | | |
| p-value (compare azacitidine vs. crossed over to azacitidine) | | 0.228 | |
| **Units of Platelet Transfusion Changed per Cycle from Baseline** | | | |
| n | 96 | | 48 |
| Mean (SD) | -0.60 (12.6) | | 2.34 (10.3) |
| Range | | | |
| p-value (compare azacitidine vs. crossed over to azacitidine) | | 0.152 | |

[a] There were 3 and 6 patients in the azacitidine group and the observation group having no transfusions data, respectively.
[b] P-value from Wilcoxon test with normal approximation by comparing the azacitidine group to the observation group.

### 3.1.6 Reviewer's Conclusion and Comments

For the overall response (CR+PR), CALGB 9221 appears to demonstrate a significant benefit of azacitidine compared to observation before crossover group for ITT population ($p < 0.0001$), ITT population without AML ($p < 0.0001$), and ITT population without AML or protocol violation ($p = 0.0007$). It also appears to demonstrate a significant benefit of azacitidine compared to observation only group (excluding crossover patients) for ITT population ($p = 0.0033$), ITT population without AML ($p = 0.010$), and ITT population without AML or protocol violation ($p = 0.0276$). However, it failed to demonstrate any significant benefit of azacitidine compared to observation before crossover group as measured by complete response for ITT population without AML and/or protocol violation patients. It also failed to demonstrate any significant benefit of azacitidine compared to observation only group as measured by complete response for ITT population with/without AML and/or protocol violation patients.

For the change of monthly transfusion requirement, the comparisons of azacitidine versus observation before crossover and azacitidine versus observation only group did not reach the significance level 0.05 under Bonferroni adjustment for red blood cell transfusion and platelet transfusion. For the change of transfusion required per cycle, the comparisons of azacitidine versus crossed over to azacitidine group also did not reach the significance level 0.05 under Bonferroni adjustment for red blood cell transfusion and platelet transfusion.

### 3.2 Evaluation of Safety

No safety evaluation is included in this statistical review. Please refer to the clinical review of this application for the details of safety evaluation.

## 4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

### 4.1 Gender, Race and Age

#### 4.1.1 Gender

The efficacy analysis of the response rates by gender is summarized in Table 10. A significant benefit was demonstrated in the comparisons of azacitidine and observation before crossover groups as measured by overall response (CR+PR) for male ITT population (p-value=0.0005), male ITT population without AML (p-value=0.0019), male ITT population without AML or protocol violation (p-value=0.0068), female ITT population (p-value=0.0386), and female ITT population without AML (p-value=0.0363). A significant benefit was also demonstrated in the comparison of azacitidine and observation without crossover groups as measured by overall response (CR+PR) for male ITT population (p-value=0.0171). All other comparisons were not statistically significant.

Table 10: Analysis of Response Rates by Gender --- FDA Analyses

| | Azacitidine | Observation before Crossover | Azacitidine vs. Obs bf X-over | Observation without Crossover | Azacitidine vs. Obs w/o X-over |
|---|---|---|---|---|---|
| **Male** | | | | | |
| ITT Population | N = 72 | N = 60 | p-value[a] | N = 29 | p-value |
| Overall (CR+PR) n (%) | 12 (16.7) | 0 (0.0) | 0.0005 | 0 (0.0) | 0.0171 |
| Complete (CR) n (%) | 5 (6.9) | 0 (0.0) | 0.0629 | 0 (0.0) | 0.3176 |
| Partial (PR) n (%) | 7 (9.7) | 0 (0.0) | | 0 (0.0) | |
| ITT Population w/o AML | N = 65 | N = 54 | p-value | N = 26 | p-value |
| Overall (CR+PR) n (%) | 10 (15.4) | 0 (0.0) | 0.0019 | 0 (0.0) | 0.0571 |
| Complete (CR) n (%) | 4 (6.2) | 0 (0.0) | 0.1251 | 0 (0.0) | 0.5751 |
| Partial (PR) n (%) | 6 (9.2) | 0 (0.0) | | 0 (0.0) | |
| ITT Population without AML or protocol violation | N = 39 | N = 32 | p-value | N = 16 | p-value |
| Overall (CR+PR) n (%) | 8 (20.5) | 0 (0.0) | 0.0068 | 0 (0.0) | 0.0886 |
| Complete (CR) n (%) | 4 (10..3) | 0 (0.0) | 0.1217 | 0 (0.0) | 0.3105 |
| Partial (PR) n (%) | 4 (10.3) | 0 (0.0) | | 0 (0.0) | |
| **Female** | | | | | |
| ITT Population | N = 27 | N = 32 | p-value | N = 12 | p-value |
| Overall (CR+PR) n (%) | 4 (14.8) | 0 (0.0) | 0.0386 | 0 (0.0) | 0.2916 |
| Complete (CR) n (%) | 1 (3.7) | 0 (0.0) | 0.4576 | 0 (0.0) | 1.0 |
| Partial (PR) n (%) | 3 (11.1) | 0 (0.0) | | 0 (0.0) | |
| ITT Population w/o AML | N = 24 | N = 29 | p-value | N = 10 | p-value |
| Overall (CR+PR) n (%) | 4 (16.7) | 0 (0.0) | 0.0363 | 0 (0.0) | 0.2958 |
| Complete (CR) n (%) | 1 (4.2) | 0 (0.0) | 0.4528 | 0 (0.0) | 1.0 |
| Partial (PR) n (%) | 3 (12.5) | 0 (0.0) | | 0 (0.0) | |
| ITT Population without AML or protocol violation | N = 15 | N = 16 | p-value | N = 6 | p-value |
| Overall (CR+PR) n (%) | 3 (20.0) | 0 (0.0) | 0.1012 | 0 (0.0) | 0.5263 |
| Complete (CR) n (%) | 1 (6.7) | 0 (0.0) | 0.4839 | 0 (0.0) | 1.0 |
| Partial (PR) n (%) | 2 (13.3) | 0 (0.0) | | 0 (0.0) | |

[a] P-value from Fisher's exact test by comparing the response rates between the azacitidine group and the observation group.
Obs bf X-over=observation before crossover, Obs w/o X-over=observation without crossover, CR=complete response, PR=partial response, CALGB=Cancer and Leukemia Group B

### 4.1.2 Race

The efficacy analysis of the response rates by race is summarized in Table 11. The results are consistent with the primary endpoint analysis.

Table 11: Analysis of Response Rates by Race --- FDA Analyses

| | Azacitidine | Observation before Crossover | Azacitidine vs. Obs bf X-over | Observation without Crossover | Azacitidine vs. Obs w/o X-over |
|---|---|---|---|---|---|
| **White** | | | | | |
| ITT Population | N = 93 | N = 85 | p-value[a] | N = 38 | p-value |
| Overall (CR+PR) n (%) | 16 (17.2) | 0 (0.0) | <0.0001 | 0 (0.0) | 0.0059 |
| Complete (CR) n (%) | 6 (6.5) | 0 (0.0) | 0.0296 | 0 (0.0) | 0.1804 |
| Partial (PR) n (%) | 10 (10.8) | 0 (0.0) | | 0 (0.0) | |
| ITT Population w/o AML | N = 83 | N = 77 | p-value | N = 36 | p-value |
| Overall (CR+PR) n (%) | 14 (16.9) | 0 (0.0) | <0.0001 | 0 (0.0) | 0.0098 |
| Complete (CR) n (%) | 5 (6.0) | 0 (0.0) | 0.0595 | 0 (0.0) | 0.3194 |
| Partial (PR) n (%) | 9 (10.8) | 0 (0.0) | | 0 (0.0) | |
| ITT Population without AML or protocol violation | N = 51 | N = 43 | p-value | N = 20 | p-value |
| Overall (CR+PR) n (%) | 11 (21.6) | 0 (0.0) | 0.0008 | 0 (0.0) | 0.0273 |
| Complete (CR) n (%) | 5 (9.8) | 0 (0.0) | 0.0603 | 0 (0.0) | 0.3122 |
| Partial (PR) n (%) | 6 (11.8) | 0 (0.0) | | 0 (0.0) | |
| **Others** | | | | | |
| ITT Population | N = 6 | N = 7 | p-value | N = 3 | p-value |
| Overall (CR+PR) n (%) | 0 (0.0) | 0 (0.0) | NA | 0 (0.0) | NA |
| Complete (CR) n (%) | 0 (0.0) | 0 (0.0) | NA | 0 (0.0) | NA |
| Partial (PR) n (%) | 0 (0.0) | 0 (0.0) | | 0 (0.0) | |
| ITT Population w/o AML | N = 6 | N = 6 | p-value | N = 3 | p-value |
| Overall (CR+PR) n (%) | 0 (0.0) | 0 (0.0) | NA | 0 (0.0) | NA |
| Complete (CR) n (%) | 0 (0.0) | 0 (0.0) | NA | 0 (0.0) | NA |
| Partial (PR) n (%) | 0 (0.0) | 0 (0.0) | | 0 (0.0) | |
| ITT Population without AML or protocol violation | N = 3 | N =5 | p-value | N = 2 | p-value |
| Overall (CR+PR) n (%) | 0 (0.0) | 0 (0.0) | NA | 0 (0.0) | NA |
| Complete (CR) n (%) | 0 (0.0) | 0 (0.0) | NA | 0 (0.0) | NA |
| Partial (PR) n (%) | 0 (0.0) | 0 (0.0) | | 0 (0.0) | |

[a] P-value from Fisher's exact test by comparing the response rates between the azacitidine group and the observation group.
Obs bf X-over=observation before crossover, Obs w/o X-over=observation without crossover, CR=complete response, PR=partial response, CALGB=Cancer and Leukemia Group B

### 4.1.3 Age

The efficacy analysis of the response rates by age is summarized in Table 12. The results support the primary endpoint analysis. A significant benefit was demonstrated in the comparisons of azacitidine and observation before crossover groups as measured by overall response (CR+PR) for 65-74 years old ITT population (p-value=0.0014), 65-74 years old ITT population without AML (p-value=0.0054), 65-74 years old ITT population without AML or protocol violation (p-value=0.0241), and > 74 years old ITT population without AML (p-value=0.0491). A significant benefit was also demonstrated in the comparison of azacitidine and observation before crossover groups as measured by CR for 65-74 years old ITT population (p-value=0.0280). All other comparisons were not statistically significant.

Table 12: Analysis of Response Rates by Age — FDA Analyses

| | Azacitidine | Observation before Crossover | Azacitidine vs. Obs bf X-over | Observation without Crossover | Azacitidine vs. Obs w/o X-over |
|---|---|---|---|---|---|
| **< 65 years Old** | | | | | |
| ITT Population | N = 36 | N = 34 | p-value[a] | N = 17 | p-value |
| Overall (CR+PR) n (%) | 2 (5.6) | 0 (0.0) | 0.4932 | 0 (0.0) | 1.0 |
| Complete (CR) n (%) | 0 (0.0) | 0 (0.0) | NA | 0 (0.0) | NA |
| Partial (PR) n (%) | 2 (5.6) | 0 (0.0) | | 0 (0.0) | |
| ITT Population w/o AML | N = 33 | N = 30 | p-value | N = 14 | p-value |
| Overall (CR+PR) n (%) | 2 (6.1) | 0 (0.0) | 0.4931 | 0 (0.0) | 1.0 |
| Complete (CR) n (%) | 0 (0.0) | 0 (0.0) | NA | 0 (0.0) | NA |
| Partial (PR) n (%) | 2 (6.1) | 0 (0.0) | | 0 (0.0) | |
| ITT Population without AML or protocol violation | N = 22 | N = 14 | p-value | N = 7 | p-value |
| Overall (CR+PR) n (%) | 2 (9.1) | 0 (0.0) | 0.5111 | 0 (0.0) | 1.0 |
| Complete (CR) n (%) | 0 (0.0) | 0 (0.0) | NA | 0 (0.0) | NA |
| Partial (PR) n (%) | 2 (9.1) | 0 (0.0) | | 0 (0.0) | |
| **65 – 74 years Old** | | | | | |
| ITT Population | N = 39 | N = 33 | p-value | N = 11 | p-value |
| Overall (CR+PR) n (%) | 10 (25.6) | 0 (0.0) | 0.0014 | 0 (0.0) | 0.0918 |
| Complete (CR) n (%) | 6 (15.4) | 0 (0.0) | 0.0280 | 0 (0.0) | 0.3168 |
| Partial (PR) n (%) | 4 (10.3) | 0 (0.0) | | 0 (0.0) | |
| ITT Population w/o AML | N = 34 | N = 30 | p-value | N = 10 | p-value |
| Overall (CR+PR) n (%) | 8 (23.5) | 0 (0.0) | 0.0054 | 0 (0.0) | 0.1666 |
| Complete (CR) n (%) | 5 (14.7) | 0 (0.0) | 0.0552 | 0 (0.0) | 0.5730 |
| Partial (PR) n (%) | 3 (8.8) | 0 (0.0) | | 0 (0.0) | |
| ITT Population without AML or protocol violation | N = 21 | N = 17 | p-value | N = 6 | p-value |
| Overall (CR+PR) n (%) | 6 (28.6) | 0 (0.0) | 0.0241 | 0 (0.0) | 0.2843 |
| Complete (CR) n (%) | 5 (23.8) | 0 (0.0) | 0.0529 | 0 (0.0) | 0.5552 |
| Partial (PR) n (%) | 1 (4.8) | 0 (0.0) | | 0 (0.0) | |
| **> 74 years Old** | | | | | |
| ITT Population | N = 24 | N = 25 | p-value | N = 13 | p-value |
| Overall (CR+PR) n (%) | 4 (16.7) | 0 (0.0) | 0.0502 | 0 (0.0) | 0.2756 |
| Complete (CR) n (%) | 0 (0.0) | 0 (0.0) | NA | 0 (0.0) | NA |
| Partial (PR) n (%) | 4 (16.7) | 0 (0.0) | | 0 (0.0) | |
| ITT Population w/o AML | N = 22 | N = 23 | p-value | N = 12 | p-value |
| Overall (CR+PR) n (%) | 4 (18.2) | 0 (0.0) | 0.0491 | 0 (0.0) | 0.2728 |
| Complete (CR) n (%) | 0 (0.0) | 0 (0.0) | NA | 0 (0.0) | NA |
| Partial (PR) n (%) | 4 (18.2) | 0 (0.0) | | 0 (0.0) | |
| ITT Population without AML or protocol violation | N = 11 | N = 17 | p-value | N = 9 | p-value |
| Overall (CR+PR) n (%) | 3 (27.3) | 0 (0.0) | 0.0504 | 0 (0.0) | 0.2184 |
| Complete (CR) n (%) | 0 (0.0) | 0 (0.0) | NA | 0 (0.0) | NA |
| Partial (PR) n (%) | 3 (27.3) | 0 (0.0) | | 0 (0.0) | |

[a] P-value from Fisher's exact test by comparing the response rates between the azacitidine group and the observation group.
Obs bf X-over=observation before crossover, Obs w/o X-over=observation without crossover, CR=complete response, PR=partial response, CALGB=Cancer and Leukemia Group B

### 4.2 Other Special/Subgroup Populations

No other special or subgroup analysis is included in this statistical review.

## 5. SUMMARY AND CONCLUSIONS

### 5.1 Statistical Issues and Collective Evidence

1. A retrospective definition of primary endpoint and secondary endpoints, statistical ypothesis, and retrospective collection and analyses of data have been included in this NDA submission. Each of these retrospective specifications in this NDA submission may result in a biased efficacy analysis and conclusion. Based on the guidance International Conference on Harmonisation (ICH)-E9: Statistical Principles for Clinical Trials, these retrospective analyses can only be considered as exploratory and hypothesis generating (Appendix 1).
2. The claim of efficacy is based on a subgroup analysis in patients with MDS and excluding all other patients included in this NDA submission. (ICH-E9: "Any conclusion of treatment efficacy (or lack thereof) or safety based solely on exploratory subgroup analyses is unlikely to be accepted".)
3. The pivotal Study CALGB 9221 was designed to allow subjects randomized to the observation group to crossover to azacitidine treatment after meeting protocol criteria. There were a total of 51 subjects (55.4%; 51/92) who crossed over from the observation group to azacitidine treatment group and only 41 subjects left in the observation group until the end of study. The original randomization is no longer valid due to crossover manner and this may result in a biased efficacy analysis. (ICH-E9: "Randomization introduces a deliberate element of chance into the assignment of treatments to subjects in a clinical trial. During subsequent analysis of the trial data, it provides a sound statistical basis for the quantitative evaluation of the evidence relating to treatment effects. It also tends to produce treatment groups in which the distributions of prognostic factors, known and unknown, are similar.")
4. All 3 CALGB studies were designed as an un-blinded study. Blinding is most definitely needed for good scientific reasons, especially when, as in this clinical testing, there are subjective or semi-objective responses. An open-label study usually results in bias based on the fundamental statistical principles when making measurements. (ICH-E9: "Blinding or masking is intended to limit the occurrence of conscious and unconscious bias in the conduct and interpretation of a clinical trial arising from the influence that the knowledge of treatment may have on the recruitment and allocation of subjects, their subsequent care, the attitudes of subjects to the treatments, the assessment of end-points, the handling of withdrawals, the exclusion of data from analysis, and so on. The essential aim is to prevent identification of the treatments until all such opportunities for bias have passed.")
5. There was a high rate of subjects with major protocol violations in both treatment groups. These violations may influence the efficacy analyses of ITT population.
6. In the design of Study CALGB 9221, the number of treatment cycles was varying for each patient. Therefore, it is questionable that the two groups can be compared with respect to response rate when the therapy is different for each patient within the treatment group.

7. Multiple analyses have been conducted and published using the CALGB 9221 study data. No type I error adjustment for the multiple analyses have been made in this application.
8. The NDA submission did not explain how to interpret the results of statistical comparisons of efficacy variables. In fact, the standard statistical comparison can not be employed in this study and p-values are not interpretable based on the retrospective definitions of primary and secondary endpoints, statistical hypotheses, and retrospective collection of data.
9. Quality of life (QoL) data are not interpretable due to missing observations, protocol violations and varying treatment cycles.

### 5.2 Conclusions and Recommendations

The NDA submitted for review is based on a retrospective statistical design and retrospective collection and analyses of data. The primary and secondary endpoints and statistical hypothesis were redefined. The statistical analysis plan was revised from the original protocol. Based on the guidance ICH-E9: Statistical Principles for Clinical Trials, the study model in this submission is therefore exploratory in nature (Appendix 1).

For the overall response (primary endpoint in the revised protocol, complete response and partial response), CALGB 9221 appears to demonstrate a significant benefit of azacitidine compared to observation (1). before crossover to treatment arm in the intern-to-treat (ITT) population ($p < 0.0001$), (2). in ITT population without acute myelogenous leukemia (AML) ($p < 0.0001$), and (3). ITT population without AML or protocol violation ($p = 0.0007$). It also appears to demonstrate a significant benefit of azacitidine compared to observation only group (excluding crossover patients) for ITT population ($p = 0.0033$), ITT population without AML ($p = 0.010$), and ITT population without AML or protocol violation ($p = 0.0276$). However, it failed to demonstrate any significant benefit of azacitidine compared to observation before crossover group as measured by complete response for ITT population without AML and/or protocol violation patients. It also failed to demonstrate any significant benefit of azacitidine compared to observation only group as measured by complete response for ITT population with AML and/or protocol violation patients and ITT population without AML and/or protocol violation patients.

For the change of monthly transfusion requirement, the comparisons of azacitidine versus observation before crossover and azacitidine versus observation only group did not reach the significance level 0.05 under Bonferroni adjustment for both red blood cell transfusion and platelet transfusion. For the change of transfusion requirement per cycle, the comparisons of azacitidine versus patients crossed over to azacitidine group also did not reach the significance level 0.05 under Bonferroni adjustment for red blood cell transfusion and platelet transfusion.

In this statistical reviewer's opinion the data and results of the one, small Phase III study suggest activity of azacitidine in patients with MDS. However, the results are not adequate to support the sponsor's efficacy claim as such since they are based on retrospective statistical design, retrospective data collection and analyses. Furthermore, the strength of statistical significance can not be evaluated based on p-value due to the retrospective nature of the study. Based on the guidance ICH-E9, in this statistical reviewer's opinion this submission can only be considered as exploratory and hypotheses generating analyses. However, azacitidine has been under investigation for over 30 years. The final recommendation should be based on clinical judgment.

# APPENDIX 1. STATISTICAL PRINCIPLES AND ICH GUIDELINES

1. All 3 CALGB studies were designed as un-blinded studies. Blinding is most definitely needed for good scientific reasons, especially when, as in this clinical testing, there are subjective or semi-objective responses. The ICH-E9 guidelines, section C of Considerations for Overall Clinical Development, states that: "The most important design techniques for avoiding bias in clinical trials are blinding and randomization, and these should be normal features of most controlled clinical trials intended to be included in a marketing application. ...Blinding or masking is intended to limit the occurrence of conscious and unconscious bias in the conduct and interpretation of a clinical trial arising from the influence that the knowledge of treatment may have on the recruitment and allocation of subjects, their subsequent care, the attitudes of subjects to the treatments, the assessment of end-points, the handling of withdrawals, the exclusion of data from analysis, and so on. The essential aim is to prevent identification of the treatments until all such opportunities for bias have passed." Therefore, an open-label study usually results in bias based on the fundamental statistical principles when making measurements.

2. The NDA submission changed the definition of the primary endpoint. The ICH-E9 guidelines, section B of Considerations for Overall Clinical Development, states that redefinition of the primary variable is unacceptable: "The primary variable should be specified in the protocol, along with the rationale for its selection. Redefinition of the primary variable after unblinding will almost always be unacceptable, since the biases this introduces are difficult to assess. When the clinical effect defined by the primary objective is to be measured in more than one way, the protocol should identify one of the measurements as the primary variable on the basis of clinical relevance, importance, objectivity, and/or other relevant characteristics, whenever such selection is feasible."

3. The efficacy claim in the NDA submission is based on a subgroup analysis. However, any conclusion of treatment efficacy based solely on exploratory subgroup analyses is unlikely to be accepted. The ICH-E9 guidelines, section G of Data Analysis Considerations, states that "In most cases, however, subgroup or interaction analyses are exploratory and should be clearly identified as such; they should explore the uniformity of any treatment effects found overall. In general, such analyses should proceed first through the addition of interaction terms to the statistical model in question, complemented by additional exploratory analysis within relevant subgroups of subjects, or within strata defined by the covariates. When exploratory, these analyses should be interpreted cautiously. Any conclusion of treatment efficacy (or lack thereof) or safety based solely on exploratory subgroup analyses is unlikely to be accepted."

4. The ICH-E9 guidelines, section A of Considerations for Overall Clinical Development, states that "The rationale and design of confirmatory trials nearly always rests on earlier clinical work carried out in a series of exploratory studies. Like all clinical trials, these exploratory studies should have clear and precise objectives. However, in contrast to confirmatory trials, their objectives may not always lead to simple tests of predefined hypotheses. In addition, exploratory trials may sometimes require a more flexible approach to design so that changes

can be made in response to accumulating results. Their analysis may entail data exploration. Tests of hypothesis may be carried out, but the choice of hypothesis may be data dependent. Such trials cannot be the basis of the formal proof of efficacy, although they may contribute to the total body of relevant evidence." Therefore, based on the un-blinded study design, redefinition of primary endpoint, statistical hypothesis, retrospective collection and analyses of data, and subgroup efficacy analysis, the clinical trial results in this NDA submission can only be considered as exploratory in nature.

# APPENDIX 2. STATISTICAL ISSUE OF TREATMENT CROSSOVER

Patients who switch treatment groups (treatment crossover) in randomized clinical trials can cause problems in the statistical design, data analysis and interpretation of the results, because the original randomized manner may be changed to a non-randomized manner. These changes can result in treatment effects of the treatment groups to be more similar. In this appendix, we will discuss this issue and different analysis methods for the survival data with treatment crossover.

Based on our literature review, Robins and Tsiatis (1991) proposed a method in correcting for non-compliance in randomized trials by estimating the parameters of a class of semi-parametric failure time models, the rank preserving structural failure time models, and using a class of rank estimators. These models are the structural or strong version of the "accelerated failure time model with time-dependent covariates" of Cox and Oakes (1984). Robins and Tsiatis (1991) developed asymptotically normal adaptive tests and rank estimators for the parameters of the rank-preserving structural failure time models from data on actual treatment, treatment arm, and survival available in a randomized clinical trial with possibly non-random non-compliance. However, a limitation of the rank estimators is the requirement that there be no censoring prior to the end of follow-up and no other missing data. This assumption greatly limits the application of the method.

Mark and Robins (1993) proposed another method for estimation and analyzing data from the multiple risk factor intervention trial to estimate the causal effect of quitting cigarette smoking. This procedure utilizes the method of Robins and Tsiatis and allows us to take advantage of post-randomization smoking history without requiring untenable assumptions about the comparability of compliers and non-compliers. However, this procedure is also based on the requirement that there be no censoring prior to the end of follow-up and no other missing data.

Law and Kaldor (1996) considered clinical trials with a time to event endpoint in which patients may switch the assigned treatment to the other during the course of follow-up and proposed a model of hazard rates for survival analysis. Since this case is similar to most oncology studies, we will discuss this paper in this appendix and keep the same notations used by Law and Kaldor (1996).

Consider a randomized clinical trial comparing the effects of two treatments A and B with a time to event endpoint, say survival. Let $n_A$ patients be randomized to treatment A, and $n_B$ patients to treatment B. Let the group of patients who were randomized to treatment A and did not switch to treatment B be denoted as group AA, patients who were randomized to treatment A and did switch to treatment B be denoted as group AB, and let groups BB and BA be defined similarly. Let $n_{AA}$, $n_{AB}$, $n_{BB}$, and $n_{BA}$ be denoted number of patients in the group AA, AB, BB, and BA, respectively. Figure 2 shows the study design.

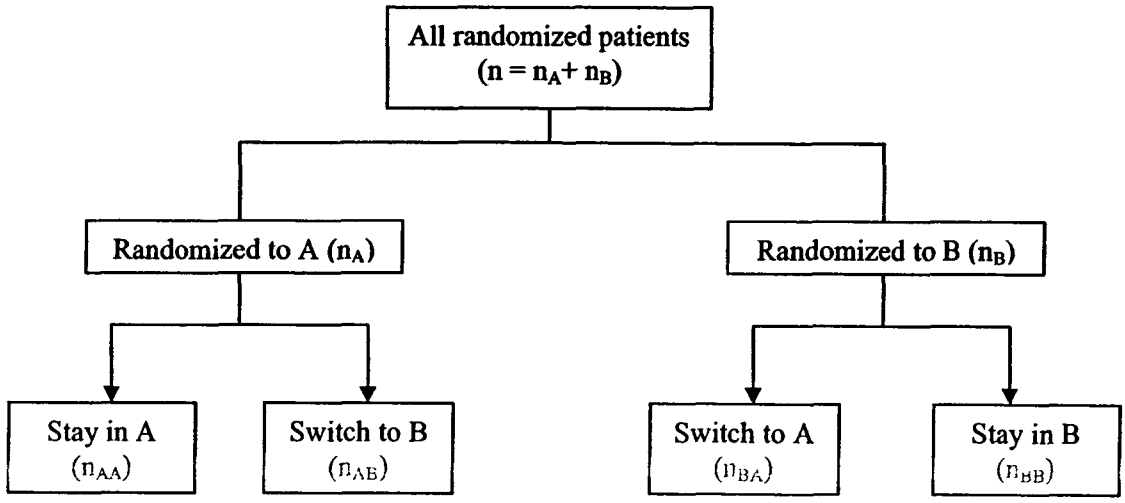



Figure 2. Study Design of Randomized Clinical Trial with Treatment Crossover

Let the hazard ratio of treatment B to treatment A be exp(β), regardless of each patient's underlying hazard. If no patients in either randomized group switch treatments, then the hazard rates would be λ(t) and λ(t)exp(β) in the treatment groups A and B, respectively. The hazard ratio exp(β) can then be estimated using the proportional hazards regression in its simplest two group form. However, some patients did switch treatments in the above study design. Therefore, we need to discuss the hazard rates for the four groups. If patients switched treatments instantaneously, and alter their hazard rate by the factor exp(β), or 1/exp(β), from that time onwards, then the hazard rates for the four groups of patients can be given as:

Group AA: $\lambda_{AA}(t)$ Group AB: $\lambda_{AB}(t)\exp(\beta\delta_i[t])$
Group BB: $\lambda_{BB}(t)\exp(\beta)$ Group BA: $\lambda_{BA}(t)\exp(\beta(1-\delta_i[t]))$

where $\delta_i[t]$ is a time dependent covariate defined to be zero up to the time point of switching treatment in patient *i* and to be one from that time point onwards, and $\lambda_{AA}(t)$, $\lambda_{AB}(t)$, $\lambda_{BB}(t)$, and $\lambda_{BA}(t)$ denote the hazard functions for each subgroup.

This specified model cannot be fitted in a Cox model because the underlying hazard rates in each group are different. Law and Kaldor (1996) proposed a simplified model by assuming that the differing underlying hazard rates in the four groups of patients can be expressed as multiplicative factors acting on a single hazard rate and obtained a simplified model:

Group AA: $\lambda(t)\exp(a_1)$ Group AB: $\lambda(t)\exp(a_2+\beta\delta_i[t])$
Group BB: $\lambda(t)\exp(b_1+\beta)$ Group BA: $\lambda(t)\exp(b_2+\beta(1-\delta_i[t]))$

In Law-Kaldor's model, the hazard for each patient at time *t* is then modeled according to group and current treatment at time *t*. However, White (1997) pointed that the group membership at time *t* may depend on the future: for example, a subject randomized to treatment A who has not

changed treatment by time *t* will be in group AB only if he or she subsequently changes treatment. The use of covariates which depend on the future is dangerous in the proportional hazards regression, and the present example is no exception. Subjects in group AB who have not yet changed treatment are modeled as having hazard $\lambda(t)\exp(a_2)$, yet their true hazard is zero because their membership of group AB means that they cannot die before changing treatment. The effect of this model inadequacy is likely introducing bias and over estimate $\beta$ (the adjusted treatment effect) upwards. This bias may or may not be the same as bias in corresponding model of group BA.

White *et al.* (1999) developed analysis methods based on the semi-parametric estimators for the clinical trials designed to assess the effect of a treatment given repeatedly or continuously during follow-up on the incidence of a disease or other event. However, the semi-parametric estimators are based on an assumed causal accelerated life model.

Rotnitzky *et al.* (2001) proposed methods of conducting sensitivity analysis for the treatment-arm mean difference of an outcome that would have been measured at the end of a randomized follow-up study if, during the course of the study, patients had not initiated a non-randomized therapy or drop out. This method can be only used for a continuous outcome variable.

Based on the above literature review discussion, there is no an appropriate approach for the survival analysis when subjects may change in a non-randomized manner from one randomized treatment to the other during the course of follow-up. Therefore, a sensitivity analysis including and excluding the crossed over subjects could be used for the purpose of exploratory and hypothesis generating analysis.

## APPENDIX 3. REFERENCES


1. Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Landon, Chapman and Hall.
2. Law, M. G. and Kaldor, J. M. (1996). Survival analyses of randomized clinical trials adjusted for patients who switch treatments. *Stat. Medicine* **15**:2069-2076.
3. Mark, S. D. and Robins, J. M. (1993). A method for the analysis of randomized trials with compliance information: an application to the multiple risk factor intervention trial. *Controlled Clinical Trials* **14**:79-97.
4. Robins, J. M. and Finkelstein, D. M. (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* **56**:779-788.
5. Robins, J. M. and Tsiatis, A. A. (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Comm. Stat.* A **20(8)**:2609-2631.
6. Rotnitzky, A., Scharfstein, D., Su., T. L., and Robins, J. (2001). Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics* **57**:103-113.
7. White, I. R. (1997). Letters to the editor. *Stat. Medicine* **16**:2619-2625.
8. White, I. R., Babiker, A. G., Walker, S. and Darbyshire, J. H. (1999). Randomization-based methods for correcting for treatment changes: examples from the Concorde trial. *Stat. Medicine* **18**:2617-2634.

## SIGNATURES/DISTRIBUTION LIST PAGE

| | |
|---|---|
| Primary Statistical Reviewer: | Yong-Cheng Wang, Ph.D. |
| Date Review Completed: | April 30, 2004 |
| Concurring Reviewers: | Rajeshwari Sridhara, Ph.D.<br>Kooros Mahjoob, Ph.D. |
| Statistical Acting Team Leader: | Rajeshwari Sridhara, Ph.D. |
| Biometrics Acting Division Director: | Kooros Mahjoob, Ph.D. |

cc:

HFD-150 / Ms. Amy Baird
HFD-150 / Edvardas Kaminskas, M.D.
HFD-150 / Ann Farrell, M.D.
HFD-150 / Richard Pazdur, M.D.
HFD-710 / Yong-Cheng Wang, Ph.D.
HFD-710 / Rajeshwari Sridhara, Ph.D.
HFD-710 / Kooros Mahjoob, Ph.D.
HFD-700 / Charles Anello, Ph.D.
HFD-700 / Satya D. Dubey, Ph.D.

C:\AAA\NDA\Vidaza (50-794)\REVIEW\Report\Vidaza.doc

**This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.**

/s/

Yong-Cheng Wang
4/27/04 05:18:18 PM
BIOMETRICS

Rajeshwari Sridhara
4/27/04 06:19:23 PM
BIOMETRICS

Kooros Mahjoob
4/27/04 06:26:28 PM
BIOMETRICS

DEPARTMENT OF HEALTH AND HUMAN SERVICES
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH
OFFICE OF BIOSTATISTICS

# Statistical Review and Evaluation

## SECONDARY REVIEW—CLINICAL STUDIES

| | |
|---|---|
| NDA: | 50-794 |
| Name of drug: | Vidaza™ (5-azacytidine) |
| Applicant: | Pharmion Corporation |
| Indication: | Myelodysplastic syndromes/pre-leukemia/smoldering leukemia |
| Document reviewed: | primary review by Yong-Cheng Wang, Ph.D. |
| Project manager: | Ms. Amy Baird |
| Clinical reviewer: | Edvardas Kaminskas, M.D. |
| Dates: | Submission Date: December 26, 2003; PDUFA Date: June 26, 2004. |
| Statistical reviewer: | Rajeshwari Sridhara, Ph.D. |

## 1 INTRODUCTION

This is a secondary review to the primary statistical review by Yong-Cheng Wang, Ph.D. I concur with Dr. Wang's review that there are inherent statistical problems in this application. This review is to clarify some of the concerns raised by the medical team.

## 2 STATISTICAL PROBLEMS

The following are specific statistical problems in evaluating this application:

1. The protocol has been rewritten for this application after the study was completed, analyzed and reported. Specifically, the primary objectives of the original protocol included evaluation of tumor response rate (proportion of patients with response) in the treatment (Vidaza) arm, impact of treatment on red cell transfusion requirements, platelet counts, ANC, rates of infection and hemorrhage and percentage of bone marrow blasts compared to the control (best supportive care or untreated observation) arm. The original protocol did not identify one specific primary endpoint. In the revised protocol the primary endpoint is specified as response rate. In the original protocol response criteria included complete response, partial response and improvement, where as in the revised protocol response criteria included complete response and partial response only.
2. The design of the original study allowed for patients in the untreated control group to cross over to the treatment group. This confounds the treatment effect and there is a potential for bias in this open-label study to switch early from control to treatment group.
3. There were more than a third of patients with major protocol violations in both arms as reported by the applicant.
4. Due to the overall retrospective nature of the study it is not possible to statistically interpret the p-values reported in the analyses.

## 3 IMPACT OF THE ABOVE PROBLEMS

When evaluating response rate, the problems listed above are less critical as detailed below:

1. Response rate was one of the primary objectives of the original study protocol.
2. There were no responses in the best supportive care/untreated observation group.
3. The cross-over of patients from the control arm to treatment arm is likely to favor the control arm as suggested by the sensitivity analyses conducted by Dr. Wang.
4. The subgroup of MDS patients that Dr. Wang refers to is actually per-protocol patients since AML patients were not supposed to be entered per original protocol eligibility criteria.

5. The response rates were consistent in the azacitadine arm as defined in this application: 16.2% in all patients, 15.7% excluding AML patients, and 20.4% excluding all patients with protocol violations.

**This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.**

/s/
---
Rajeshwari Sridhara
5/18/04 04:27:24 PM
BIOMETRICS

Kooros Mahjoob
5/18/04 05:00:24 PM
BIOMETRICS