

CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER: NDA 21174

STATISTICAL REVIEW(S)

Statistical Review and Evaluation

NDA #: 21-174 **APR 13 2000**

Applicant: Wyeth Ayerst Research

Name of the Drug: Gemtuzumab Zogamicin

Indication: Treatment of Adult Acute Myeloid Leukemia patients at first relapse

Documents Reviewed: Vols: 1.2, 1.102, 1.105, 1.113, 1.117

Medical Reviewer: Peter Bross, M.D.

Statistical Reviewer: Rajeshwari Sridhara, Ph.D.

1. Background and Overview:

Gemtuzumab Zogamicin (G-Z) is an antibody-targeted chemotherapy agent being developed to demonstrate efficacy in Acute Myeloid Leukemia (AML) patients at first relapse. G-Z has 3 components: an antibody directed against CD33 antigen, a derivative of calicheamicin, and a linker connecting the antibody and the calicheamicin derivative. The calicheamicin component is a cytotoxic derivative of the calicheamicin family of antitumor antibiotics. The CD33 antigen is a cell surface protein expressed by the leukemic cell populations of more than 80% of AML patients. G-Z binds to the CD33 antigen on the surface of leukemic cells and other cells expressing CD33, and then the G-Z is internalized. Once inside the cell, evidence suggests that calicheamicin is released from the antibody and is converted to a reactive intermediate that damages DNA, causing cell death.

In order to support labeling for the indication of treatment of patients at first relapse with AML, the sponsor submitted an NDA comprising of one pivotal Phase II trial (protocol 0903B1-201-US) and two supporting Phase II trials (protocol 0903B1-202-EU and protocol 0903B1-203-WW). All the three studies are open label, non-comparative trials of G-Z with patients dosed at a single dose level of 9 mg Protein/m² administered as a 2 hour IV, for up to 3 dose periods. The sponsor's submission included the pooled analysis of the three Phase II trials and individual analysis of the three trials. This review will focus on the efficacy aspect of the study.

2. Description of the Study:

The pivotal study protocol 0903B1-201-US (heretofore referred as study 201), is an open label, multidose, non comparative study with AML patients in first relapse, enrolled at 11 investigational sites in the United States and Canada. The study was started in 1997 and continues to be open. The data cut off for this NDA submission was March 1999 and 59 patients were enrolled by this date.

The supportive study protocol 0903B1-202-EU (heretofore referred as study 202), is an open label, single arm, multidose, multicenter outpatient study with AML patients in first relapse, enrolled in 17 investigational sites in Europe. The study is similar in design to study 201 being conducted in United States. The study was started in April 1998 and continues to be open to date. The data cut off for this NDA submission was April 1999 and 25 patients were enrolled by this date.

The second supportive study protocol 0903B1-203-WW (heretofore referred as study 203), is an open label, single arm, multidose study carried out at 4 and 8 investigational sites in the United States and Europe, respectively, in AML patients in first relapse, who are at least 60 years old. The study is similar in design to study 201. The study was started in December 1997 and continues to be open. The data cut off for this NDA submission was March 1999 and 20 patients were enrolled by this date.

The inclusion and exclusion criteria were similar in all the three studies.

2.1 Efficacy Endpoints

The primary efficacy endpoint was complete remission per protocol, in all the three studies.

A patient's remission status was classified as complete remission (CR), morphologic remission (MR), or no remission (NR). A patient was considered to be in CR if all the following conditions were met: a) leukemic blasts were absent from the peripheral blood; b) the percentage of blasts in the bone marrow was $\leq 5\%$ as measured by morphology studies (aspirate or biopsy) with neither aspirate nor biopsy exceeding 5%; c) peripheral blood reached the following levels: hemoglobin (Hgb) ≥ 9 g/dL, platelets $\geq 100,000/\mu\text{L}$, absolute neutrophil count (ANC) $\geq 1500/\mu\text{L}$, and d) the patient was red-cell and platelet-transfusion independent. Red-cell transfusion independence was defined as 2 weeks without red-cell transfusion; platelet transfusion independence was defined as 1 week without platelet transfusion. An independent consultant reviewed the morphology from the bone marrow samples to confirm the readings of the local laboratory. When the readings of the independent consultant and the local laboratory differed, the judgment of the independent consultant prevailed. The definition of MR was the same as that of CR with the exception that platelet recovery was not required. No Remission Patients were considered to be in NR if they did not meet all of the criteria for CR or MR. These

patients were followed for response and adverse events. Patient response was first evaluated 28 days after the final dose of G-Z. A patient's maximum response was used for the analysis.

Other efficacy endpoints are: rates of MR and OR (overall remission = CR + MR), relapse-free survival, survival, and landmark survival. Total relapse-free survival is defined as the date of first documentation of response to date of documentation of relapse (i.e., relapse date minus date of documentation + 1). Total survival is defined as the date of first administration of the dose of G-Z to date of death (i.e., date of death minus date of first dose + 1). Patients who were still alive on the date of data cut-off, were censored on that date. Landmark survival is defined as the date of first evaluation of remission (28 days after the final dose) to date of death (i.e., date of death minus date of first evaluation + 1). Patients who were still alive on the date of data cut-off, were censored on that date.

2.2 Efficacy Analysis Methods

Since all studies were open-label and without a control treatment group, the statistical analyses for the individual protocols generally consisted of summary statistics and interval estimation. Only the sample size of Protocol 201 was large enough to provide meaningful interval estimates. For each efficacy endpoint the data were summarized and confidence intervals were calculated for each response group.

Time to event data were graphically summarized by plotting Kaplan-Meier survival estimates and/or by scatter plots that differentiated censored and uncensored observations. When the medians were estimable, 95% confidence intervals for the median survival were calculated by using the method of Brookmeyer and Crowley.

The primary analysis was conducted for the intent-to-treat population (i.e., all patients who received at least one dose of G-Z) consisting of the patients who were enrolled on or before the enrollment cut-off date of 31DEC98.

2.3 Study Design and Sample Size

The sample sizes for Protocols 201 and 203 were based on the Simon Two-Stage Design. For Protocol 202 the originally planned sample size (20) was based on the need to obtain additional safety data. To calculate the required sample size the following assumptions were made. The response probability of an ineffective drug (i.e., the uninteresting level) was chosen to be 0.15, the response probability of an effective drug (i.e., the target level) was chosen to be 0.30, level of significance of 0.10, and power of 0.90. Based on these assumptions the sample size of the first stage was 23 and the sample size required for the entire study was 55. According to the Simon Two-Stage Design for this specific design, if 3 or fewer complete remissions (CRs) were observed during the first stage of the study the trial would be halted and the drug would be declared as ineffective. If the total number of CRs observed after completion of the second stage was fewer than 12, further development of the drug would not proceed. However, the study design was modified

during the course of study by protocol amendment, so that the study was not automatically halted even when after the first stage there were 3 or fewer patients with complete remission. By modifying the Simon 2-Stage Design, the type I error of the original Simon Two-Stage Design is inflated. This means that if there were only 2 or 3 CRs and the trial continues, the probability of receiving a "poor" drug, as defined by CR alone, is greater than 0.10.

Furthermore, although the total sample size of 55 patients is adequate to provide good estimates of the primary endpoint, remission rate, and is similar to the sample size requirements of the method of Gehan (1961), the protocols were amended to allow enrollment to continue beyond the originally planned number of patients. As of the date of cut off, study 201 had enrolled 59 patients. The amendment for study 201 did not specify the final total number of patients that will be enrolled into this study. In study 202 where the enrollment was initially fixed at 20 patients to obtain additional safety data, the study has been amended (amendment V) to enroll up to 150 patients. Justification for this increase is not clear. Study 203 has also been amended to not to stop at the first stage of design per Simon's design. However, as in study 201, the total number of patients who will be enrolled into this study is not specified.

2.4 Interim Analysis

For study 201, in accordance with the rules established by the Simon Two-Stage Design, an interim analysis was performed after 23 patients had enrolled and completed Part I of the study. Although the criterion for continuing beyond the first stage was not met (i.e., > 3 CRs), decision was made to continue the study through completion of the second stage. To account for the interim evaluation of the data, confidence intervals for the primary endpoint, remission rates, were adjusted by using the method of Atkinson and Brown.

An interim analysis is also planned for study 203, but not for study 202. However, it should be noted that in fact, an interim analysis has been submitted for study 203 with 20 patients, before reaching the end of first stage with 23 patients, per Simon's design.

2. Summary of Efficacy Results and Comments

2.1 Demographic and Other Baseline Characteristics

Age and duration of first remission have been found to be important prognostic factors for patients with AML in first relapse. These characteristics are summarized in Table 1 (Sponsor's IES Table 14). Other demographic characteristics for the pooled study (study 201 + study 202 + study 203) are presented in Table 2 (Sponsor's IES Table 15).

Table 1: Patient Age and Duration of First Remission

Characteristic	Study 201 (n = 59)	Study 202 (n = 25)	Study 203 (n = 20)	Studies 201/202/203 (n = 104)
Age, years				
Mean (SD)	52.7 (15.8)	57.4 (12.1)	69.6 (6.6)	57.1 (15.0)
Min – Max	22 – 81	30 – 79	60 – 84	22 – 84
Median	53.0	58.0	70.0	60.0
Duration of first remission (before G-Z), months				
Mean (SD)	16.7 (14.7)	21.3 (27.1)	8.9 (7.6)*	16.4 (17.9)
Min – Max	6 – 95	5 – 117	3 – 35	3 – 117
Median	12.6	11	6.7	11.2

*n = 19 because in one patient precise duration of remission could not be calculated.

Table 2: Summary of Demographic and Baseline Characteristics

Characteristic	Studies 201/202/203 (n = 104)
Sex, n (%)	
Women	44 (42)
Men	60 (58)
Ethnic Origin, n (%)	
White	96 (92)
Black	4 (4)
Asian	2 (2)
Other	2 (2)
Height (cm²)	
Mean	170.2 (8.8)*
Min – Max	144.0 – 191.0
Median	170.1
Weight (kg)	
Mean	78.2 (16.5)
Min – Max	42.5 – 143.7
Median	76.9
Body surface area (m²)	
Mean	1.9 (0.2)
Min – Max	1.3 – 2.7
Median	1.9

*n = 102

3.2 Rates of Remission

The primary efficacy end point was complete remission. Table 3 (sponsor's IES Table 17) gives the rates of CR, MR, and OR (CR + MR), in the individual and pooled phase II studies. The rates of complete remission are between 15 – 19% in all the three studies.

Table 3: Number (%) of Patients and 95% CIs by Remission Category

Type of Remission	Study 201* (n = 59)	Study 202 (n = 25)	Study 203 (n = 20)	Studies 201/202/203 (n = 104)
CR				
No (%) of patients	11 (19)	4 (16)	3 (15)	18 (17)
95% CIs	(10, 31)	(5, 36)	(3, 38)	(11, 26)
MR				
No (%) of patients	9 (15)	4 (16)	1 (5)	14 (13)
95% CIs	(7, 27)	(5, 36)	(0, 25)	(8, 22)
OR (CR + MR)				
No (%) of patients	20 (34)	8 (32)	4 (20)	32 (31)
95% CIs	(22, 47)	(15, 54)	(6, 44)	(22, 41)

*Data from 1 patient who had a bone marrow sample taken on day 20, rather than 28, was also included in the assessments.

Furthermore, for the purpose of approval of labeling of G-Z, a comparative table of remission rates of G-Z with conventional combination therapies were presented as detailed in the following Tables 4, 5a, and 5b (Sponsor's IES Tables 34 and 35). The data regarding conventional therapies were collected from a comprehensive search of the medical literature and selection of papers reporting studies in patient populations with AML in first relapse and with prognostic factors, and evaluation of data from several institution/cooperative group databases on AML patients in first relapse. No statistical comparisons were made with the efficacy data, as the data from this study is being compared to historical data.

Table 4: Second CR Rates in Randomized Phase III Trials and in G-Z Phase II Trials

Institution/Group -	Therapy	Second CR Rate % Patients (95% CI)
German Cooperative Group (Kern et al, 1998)	HiDac versus IDAC + mitoxantrone	47 (39, 56)
Southeastern Cancer Group (Vogler et al, 1994)	HiDac + etoposide Versus	45 (30, 61)
	HiDac	40 (26, 56)
Kohseisho Leukemia Group (Ohno et al, 1994)	Cytarabine + mitoxantrone + etoposide + filgrastim Versus	54 (33, 74)
	Cytarabine + mitoxantrone + etoposide	42 (23, 63)
W-AR	G - Z	17 (11, 26) ¹
		31 (22, 41) ²

¹ CR rates; ² OR rates

**Table 5a: Number (%) of Patients with Second Remission
Retrospective Reviews with ≥ 100 Patients and G-Z Phase II Clinical Trials**

Author, Institution	Total n	Duration of first CR				Age			
		< 1 Year		≥ 1 Year		< 60 Years		≥ 60 Years	
		n (%)	95% CI	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI
Rees Medical Research Council	485	251 (13)	8, 18	234 (48)	42, 55	375 (33)	26, 38	110 (19)	2, 28
Keating MD Anderson**	187	105 (19)	12, 28	82 (62)	51, 73	208* (36)	29, 42	35 (14)	5, 30
Thalhammer Univ. of Vienna	168	121 (33)	25, 42	47 (55)	40, 70	NA	NA	NA	NA
Hiddlemann German Coop. Group	136	87 (46)	35, 57	49 (60)	44, 73	104 (54)	44, 64	32 (44)	26, 62
Davis St. Bartholomew's	126	NA (33)	NA	NA (49)	NA	NA (40)	NA	NA (40)	NA
G - Z W - AR	104	56 (14 ¹)	6, 26	47 (21 ¹)	11, 36	50 (18 ¹)	9, 31	54 (17 ¹)	8, 29
		56 (30 ²)	19, 44	47 (32 ²)	19, 47	50 (34 ²)	21, 49	54 (28 ²)	16, 42

* Subgroups by age include relapsed and refractory AML patients; ** MD Anderson cohort age stratified < 65 and ≥ 65; NA = Not available

¹ CR Rates; ² OR Rates.

**Table 5b: Number (%) of Patients with Second Remission
Retrospective Reviews with < 100 Patients and G-Z Phase II Clinical Trials**

Author, Institution	Total n	Duration of first CR				Age			
		< 1 Year		≥ 1 Year		< 60 Years		≥ 60 Years	
		n (%)	95% CI	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI
Angelov Toronto Hospital	51	22 (32)	14, 55	29 (72)	53, 87	35* (69)	51, 83	20* (35)	15, 59
MacCallum St. Bartholomew's	25	NA	NA	NA	NA	19 (63)	38, 84	6 (67)	22, 96
Letendre Mayo Clinic	24	17 (29)	10, 56	7 (57)	18, 90	20 (40)	19, 64	4 (25)	1, 81
Rassam UK Multicentre Group	22	NA (54)	NA	NA (33)	NA	NA	NA	NA	NA
G - Z W - AR	104	56 (14 ¹)	6, 26	47 (21 ¹)	11, 36	50 (18 ¹)	9, 31	54 (17 ¹)	8, 29
		56 (30 ²)	19, 44	47 (32 ²)	19, 47	50 (34 ²)	21, 49	54 (28 ²)	16, 42

* Subgroups by age include relapsed and refractory AML patients; NA = Not available

¹ CR Rates; ² OR Rates.

The above results suggest that G-Z with lesser complete remission rates is inferior to the existing/published literature results. This NDA submission is not based on the primary efficacy parameter complete remission, but based on the overall remission, under the assumption that the CR and MR patients are clinically similar.

The sponsor has tried to establish that the CR and MR are similar, by comparing the time to remission, time to relapse after remission (remission duration), survival time, and landmark survival time. It must be noted that in the pooled data set of 104 patients treated with G-Z, there are only 18 CRs and 14 MRs. These subgroup sizes are too small to perform any statistical comparisons to draw meaningful conclusions. For example, to detect a significant difference between two groups in median time to event (example: relapse or survival) as large as 4 versus 8 months, with a significance level of 0.05 and power of 80%, at least 50 patients are required in each of the two groups. Observed data are not sufficient to conclude that CR and MR patients are similar.

3.3 Time to Remission

The time to remission was between 1 and 3 months for the majority of patients with CRs and MRs. In the 3 pooled studies (201/202/203), the median time to remission was 54 days for patients with CRs and 57.5 days for patients with MRs. Table 6 (sponsor's IES supportive Table 3) provides the median number of days (Kaplan-Meier estimates) and 95% CIs for study 201 and the pooled studies.

Table 6: Time to Remission in Phase II Studies, Median Number of Days and 95% CIs

Type of Remission	Study 201* (n = 59)	Study 202 (n = 25)	Study 203 (n = 20)	Studies 201/202/203 (n = 104)
CR	n = 11	n = 4	n = 3	n = 18
Median, days	56.0	57.0	52.0	54.0
95% CIs	(48, 69)	---	---	(50, 67)
MR	n = 9	n = 4	n = 1	n = 14
Median, days	51.0	64.0	84.0	57.5
95% CIs	(44, 60)	---	---	(51, 80)
OR (CR + MR)	n = 20	n = 8	n = 4	n = 32
Median, days	53.5	60.5	67.0	56.5
95% CIs	(48, 67)	---	---	(50, 67)

3.4 Relapse Free Survival

As of cutoff dates, 11/32 had relapsed and the median total relapse free days (Kaplan-Meier estimate) was 267 days (8.75 months) in the overall remission group. In the CR group, 6/18 patients had relapsed and the median relapse free survival was 216 days (7.1

months, lower 95% confidence bound 203 days). In the MR group, 5/14 patients had relapsed and the median had not been reached by the date of cut off (lower 95% confidence bound 67 days).

3.5 Survival

Among the 104 patients in the 3 phase II studies, 57 patients had died as of the cutoff dates. Total overall survival was evaluated using Kaplan-Meier estimates and the median duration of total overall survival was 226 days (7.5 months). The median survival in the 72 NRs was 120 days (95% CI: 82, 162) and 52/72 were dead at the time of cut off date. In the MR group, 4/14 had died and the median survival had not been reached (lower 95% confidence bound 294 days). In the CR group, 1/18 had died and the median survival had not been reached.

Furthermore, a landmark survival (post-remission survival) analysis in the CR and MR groups was conducted. In this NDA submission, observed median landmark survival of the censored observations (instead of the Kaplan-Meier estimates) are presented as: that the median number of days of landmark survival was greater for MR patients (257.5 days) than for the CR patients (153.0 days) (Sponsor's IES report, item 10, page 71). These are misleading statistics and these median values may not be used in labeling, as these median values do not account for censoring of the observations. In fact, in a subsequent updated communication, dated Dec 21, 1999, sent by the sponsor to the agency, the median landmark survival for CRs (Kaplan-Meier estimate) is reported to be 379 days and the median in the MR's had not been reached as of July 1999.

3.6 Multivariate analysis

Exploratory analysis of potential prognostic factors was performed using pooled data from all 104 patients in studies 201, 202 and 203. A total of 22 variables were examined. A logistic regression analysis was used for predicting response (OR versus NR), and a proportional hazard model was used to evaluate predictors of landmark survival. In both the analyses, first a univariate analysis was performed and then all results significant at the 0.15 level were included in the multivariate model. Because of missing values, these analyses were conducted in a subset of these patients (76/104 patient data for logistic regression and 67/82 patient data for proportional hazards analysis). The results of these analyses are presented in Table 7 (sponsor's IES Table 32). Due to the exploratory nature of this multivariate analysis, the results should be cautiously examined, and furthermore, these results should be confirmed by prospective, randomized trials.

Table 7: Results of Multivariate Exploratory Analysis for Prognostic Factors

Logistic Regression Analysis of OR versus NR

Analysis Variable	Wald Chi-square p-value	Odds Ratio*
CD13	0.005	0.03
MDR Efflux	0.009	0.97
Hemoglobin	0.015	1.71
CD56	0.026	0.08

*For CD13 and CD56, the ratios are for positive baseline values versus negative baseline values. For other baseline values, the odds ratios are per unit increase in that prognostic variable.

Landmark Survival Analysis

Analysis Variable	Wald Chi-square p-value	Risk Ratio for death*
Quantitative CD33 expression	0.002	0.98
Duration of First Remission	0.013	0.92
Peripheral blood blasts	0.029	4.2
CD34	0.035	2.75

*For CD34, the ratio is for positive baseline values versus negative baseline values. For other baseline values, the risk ratios are per unit increase in that prognostic variable.

4. Conclusion

The primary efficacy parameter, complete remission was estimated to be 17% (95% CI: 11, 26) in the three pooled phase II trials of G-Z (Table 4). The CR was estimated to be 21% (95% CI: 11, 36) in the subgroup of patients who had one or more years of duration of first remission and 18% (95% CI: 9, 31) in the subgroup of patients who were less than 60 years old. The duration of second remission in the CR group is 7.1 months (Table 5a). The overall remission defined as the combined complete and morphologic remissions, was estimated to be 31% (95% CI: 22, 41) (Table 4).

With 18 CRs and 14 MRs, and 1 CR and 4 MR deaths, among the total 104 patients treated with G-Z in the three pooled phase II trials, it is not possible to establish that these two subgroups (CR and MR) are similar or otherwise, in order to combine the two response groups as one response group. Therefore, it is not established that G-Z is in par with conventional therapies with respect to remission rates.

The observed median of censored observations (instead of Kaplan-Meier estimates of the median) should not be used in labeling, as this is not a valid statistic.

Because of the open labeled, uncontrolled, non-randomized nature of the phase II trials presented in this NDA, no formal statistical testing or comparisons could be conducted. Therefore, any claims of 'improved efficacy' or 'no significant differences' need to be cautiously examined. The final recommendation should be based on clinical judgement.

5. **Comments/Suggestions following March 2000 ODAC Meeting**

- There is no conclusive evidence that MR (CRp) and CR are similar or otherwise, even with the additional data presented by the sponsor at the ODAC meeting.
- Efficacy data are being compared to historical data not reviewed by FDA. Also, eligibility criteria in these historical studies are not known.
- A post-hoc subgroup analysis of efficacy is presented in the subgroup of ≥ 60 years age.
- Completion of Study 203, which was prospectively planned in the age group ≥ 60 years, is recommended. Final analyses of this study should be presented.
- Comparison of efficacy of G-Z with other existing regimens may be considered using a case-control study (matching by age group and duration of prior response).

References:

Brookmeyer R and Crowley J. "A Confidence Interval for the Median Survival Time" Biometrics. 1982; 38: 29-41.

Gehan, EA. "The Determination of the Number of Patients in a Preliminary and a Follow-up Trial of a New Chemotherapeutic Agent" Journal of Chronic Diseases. 1961; 13: 346-353.

Simon, R. "Optimal Two-Stage Designs for Phase II Clinical Trials". Controlled Clinical Trials. 1989; 10: 1-10.

JS/

Rajeshwari Sridhara, Ph.D.
Statistician

Concur:

Dr. Chen

JS/

4/13/00

Dr. Mahjoob

JS/

Cc:

Archival: NDA 21-174

2 4/13/2000