STATISTICAL REVIEW AND EVALUATION

NDA#:	50-740
APPLICANT:	Fujisawa USA, Inc
NAME OF DRUG:	AmBisome® (liposomal amphotericin B)
INDICATION:	Empiric Therapy or Treatment of Systemic Fungal Infections or Prophylaxis in Immunocompromised Hosts and for Visceral Leishmaniasis
DOCUMENTS REVIEWED:	Volumes 1.0001, 6.7, 8.1, 12.1, 13.1, 14.1, 15.1, 15.4, 15.5, 23.1
MEDICAL INPUT:	Jeff Murray, M.D. (HFD-530), Joyce Korvick M.D. (HFD-590), Andrea Meyerhoff, M.D. (HFD-590)

STATISTICAL REVIEW AND EVALUATION

NDA#: 50-740

1. Background

1.1 Overall Objectives

The applicant completed seven trials to compare the safety and efficacy of several doses of ambisome (Ambs) to that of control drugs. One trial, 94-0-002, was a randomized, double blind, multicenter trial comparing four possible doses of ambisome to four possible doses of amphotericin B (Ampb) for the empiric therapy of febrile, neutropenic patients. Two of trials (104-08 and 104-13) were double blind, randomized, placebo controlled trials for prophylaxis in neutropenic subjects with liver transplants or hematological malignancies. The other four trials (104-05, 104-09, 104-10, and 104-14) were open label, randomized trials with a conventional amphotericin B control in the empiric therapy of suspected or confirmed fungal infections. There was also an open label, comparative dose study (104-19) with no comparator drug for the treatment of confirmed Aspergillosis. This study will not be addressed further except with respect to adverse events.

1.2 Summary of Study Designs

1.2.1 Double Blind Empiric Therapy Study

Study 94-0-002 was a randomized, double-blind, active controlled, multi-center trial, with 702 patients at 32 participating sites in the US. Patients were undergoing chemotherapy for cancer, had had a bone marrow transplant, or had had a peripheral blood stem cell transfusion. They were neutropenic (<250 neutrophils/mm^3), were febrile (temperature $>$ $38^{\circ}C$ for 48 hours), and had received $\geq$ 96 hours of empiric antibacterial therapy. Patients were randomized by blocks at each site in a 1:1 ratio to either IV ambisome at 3 mg/kg/day or IV amphotericin B at .6 mg/kg/day. Depending on response, toxicity, and the clinical judgment of the investigator, individual patients on either arm could have their doses reduced by a factor of 1/2 or increased by factors of either 1.5 or 2. Patients were treated for up to 3 days after their absolute neutrophil count (ANC) rose above 250. (The 28 patients who entered the study with neutrophil count $>$ 250 are discussed in section 1.3.1 below.)

1.2.2 Double Blind Prophylaxis Studies

Studies 104-08 and 104-13 were both randomized, double-blind, placebo controlled, multi-center trials. In study 104-08 the sample consisted of all patients undergoing orthotopic liver transplants at two major university hospitals, one each in Sweden and Finland, between Feb 1991 and Apr 1992 who did not have prior fungal infection. Patients were randomized in a 1:1 ratio to either placebo or ambisome at 1 mg/kg/day for 5 days starting on the day of operation. Patients were randomized by blocks at each center. 59 patients were randomized in Sweden and 27 in Finland.

In study 104-13 the sample consisted of patients with hematological malignancies who received either myeloablative therapy in preparation for bone marrow transplants, high dose immunosuppressive treatment for GVHD, or chemotherapy known to consistently cause neutropenia. The patients were selected at five hospitals in Ireland and the UK. Patients were randomized in a 1:1 ratio to either placebo or ambisome at 2 mg/kg 3 times per week until the end of neutropenia, death, or termination of treatment due to adverse event. The average duration of observation was 25 days. Patients were randomized by center.

1.2.3 Open Label Treatment Studies

Studies 104-10, 104-14, 104-05, and 104-09 were open-label (OL), randomized, parallel, multi-center trials. All used amphotericin B as the comparator. The large difference in infusion times for ambisome (45 minutes-1 hour) and amphotericin B (4-6 hours) is a major reason for the open-label design. The studies differed slightly in their inclusion criteria.

Study 104-10 enrolled adult subjects who were immuno-suppressed and belonged to one of two strata. Either they were neutropenic with fever of unknown origin (FUO) which was unresponsive to antibiotics for 96 hours or they had confirmed mycosis (CM). Within each stratum patients were randomized in a 1:1:1 ratio to either amphotericin B at 1 mg/kg/day, ambisome at 1 mg/kg/day, or ambisome at 3 mg/kg/day. Patients randomized to amphotericin were permitted to crossover to ambisome 1 mg/kg/day in the event of non-resolving nephrotoxicity.

Treatment lasted until the first of the following endpoints was reached:

1) Resolution of fever plus other signs, symptoms, or tests indicative of fungal infection. For patients treated prophylactically, return of neutrophil count to greater than $1000/mm^3$ for 3 consecutive days was also required.

2) Occurrence of severe adverse event

3) Patient withdrawal.

The latter two endpoints counted as failure.

Concomitant use of other anti-fungals was not allowed but use prior to study initiation was allowed. Patients were required to have baseline serum creatinine less than twice the upper limit of normal or creatinine clearance greater than 50 Ml/min.

Study 104-14 consisted of patients under 18 years of age with chemotherapy induced neutropenia (neutrophils $<500/mm^3$) who had fever $\geq 38^{\circ}C$ that did not respond to 96 hours of broad spectrum antibiotics (criteria similar to the FUO stratum in trial 10). Patients were required to have baseline serum creatinine < 2 ULN and were not to have either evidence of deep fungal infection nor any systemic anti-fungal treatment within the previous 28 days.

The patients were selected at six hospitals in the UK. Patients were randomized in a 1:1:1 ratio to either amphotericin B at 1 mg/kg/day, ambisome at 1 mg/kg/day, or ambisome at 3 mg/kg/day. Patients were randomized by center.

Study 104-05 was an open label study in which 31 patients were randomized to amphotericin B and 32 to ambisome. This study was stopped early when only 20 patients (out of 63 randomized) were deemed evaluable by one year after the anticipated date at which recruitment would be complete. The unplanned early stopping and the open label selection of only one third of the subjects for inclusion in the analysis make the results from this study statistically uninterpretable and it will not be further discussed except with respect to safety.

Study 104-09 enrolled HIV+ subjects who were over 18 years of age with a primary episode of confirmed cryptococcal meningitis. Only 30 patients were enrolled in this open-label

study. Furthermore, the report submitted by the applicant for this study is a journal article by the investigators. No primary efficacy variable was identified; complete details of the statistical analysis were lacking; and data for FDA re-analysis are unavailable. Given these limitations, this study will not be further discussed with respect to efficacy.

1.3 Patient Accounting and Baseline Characteristics

1.3.1 Double Blind Empiric Therapy Study (Trial 02)

In trial 002, 702 patients were randomized at 32 sites. Of these 702, 15 were excluded before receiving drug for patient refusal or for failure to meet entry criteria during screening. The number of patients per site ranged from 55 to 1, with all but 3 centers having at least 10 patients.

The mean age of patients was 41 with a range of 2 to 80 in the two arms. The subjects were 86% white and 54% male. The patients' primary diagnoses were similar between the two arms. 46% of subjects received baseline systemic antifungal prophylaxis; 5% had elevated serum creatinine at baseline; 15% were anemic at baseline. These figures were nearly the same on each arm. Nine ambisome patients and eight amphotericin patients had fungal infections at baseline. 14 patients on each arm did not have baseline ANC < 250. Among these, 9 ambisome patients and 7 amphotericin patients eventually had ANC below 250.

Table 1.3.1 A shows the reasons for discontinuations for trial 02.

TABLE 1.3.1 A
REASONS FOR DISCONTINUATIONS IN TRIAL 02

	AmBisome	Amphotericin B
Randomized	347	355
Received ≥ 1 Dose	343	344
Completed Treatment	255 (74%)	243 (71%)
Discontinued		
Adverse Event	25 (7%)	25 (7%)
Death	10 (3%)	12 (4%)
Infusion Reaction	8 (2%)	22 (6%)
Lack of Efficacy	13 (4%)	14 (4%)
Other	32 (9%)	28 (8%)

1.3.2 Double Blind Prophylaxis Studies
Trial 08 Liver Transplant Patients

In trial 08, 43 liver transplant patients were randomized to AmBisome and 43 were randomized to placebo. One placebo patient died during surgery and another had no baseline fungal infection data. Four other placebo patients and three ambisome patients received fewer than 5 days of therapy.

The median age of patients was 41 with a range of 1 to 67 in the two arms. The subjects were all white and 51% male. The ambisome arm had more males (63% vs 45%). The patients' primary diagnoses were similar between the two arms.

Trial 13 Neutropenic Hematological Malignancy Patients

In trial 13, 80 bone marrow transplant patients were randomized to AmBisome and 90 were randomized to placebo. Five ambisome patients and two placebo patients never received treatment after randomization and were excluded from all analyses. One of the two placebo subjects, 04012, re-appears in the data set as subject 04016, still on placebo. Of the other six, three did not receive their transplants or begin their chemotherapy as scheduled after randomization; another received ketoconazole as a prophylaxis. Thus, these four violated entry criteria. No information is provided about the other two (one in

each arm). In addition, two patients at site 1 were randomized to ambisome but received placebo and were analyzed in the placebo arm.

The subjects in the efficacy evaluable (EE) subset were randomized to ambisome and placebo, respectively, at the five sites as follows: (18, 21), (16, 18), (20, 20), (12, 19), and (9, 10). Site 4 had 4 subjects randomized to ambisome dropped from the EE set so it was originally (16, 19).

The median age of patients was 39 with a range in the two arms. The subjects were 91% white and 63% male. None of the baseline demographic covariates were significantly different in the two arms. The patients' primary diagnoses were similar between the two arms. AmBisome had significantly more subjects with decreased neutrophils (14% vs 6%) with an exact p-value of .07.

Table 1.3.2 A shows the reasons for discontinuations for trial 13.

TABLE 1.3.2 A
REASONS FOR DISCONTINUATIONS IN TRIAL 13

	AmBisome	Placebo
Randomized	80	90
Evaluable	75	88
Neutrophils Recovered	27 (36%)	30 (34%)
Neutrophils Recovering	1 (1%)	2 (2%)
Suspected Fungal Infec.	25 (33%)	36 (41%)
Severe AE or Death	6 (8%)	3 (3%)
Intercurrent Illness	2 (2%)	0 (0%)
Local Fungal Infec.	10 (13%)	9 (10%)
Toxicity/Reaction	1 (1%)	2 (2%)
Other	3 (4%)	6 (7%)

1.3.3 Open Label Treatment Studies
Trial 10 Adult Empiric Therapy

In trial 10, 134 cases of fever of unknown origin (FUO) and 59 cases of confirmed mycosis (CM) were enrolled. One case of FUO was dropped from both safety and efficacy analyses and 20 cases of CM were dropped from the efficacy analysis but included

in the safety analysis. The investigator for the excluded FUO case did not fill out case report forms (CRF's) for him after learning that the patient had taken itraconazole. Thus, no data was available for this subject. The twenty excluded CM cases were reclassified, after receiving drug, as not having had a confirmed mycosis after all. The FDA medical officer reviewed the CRF's for these subjects and concurred that no useful determinations of response could be made from these subjects.

There were 13 investigators in the study. The numbers of patients per site were 40, 35, 31, 21, 15, 14, 11, 7, 7, 6, 6, 2, and 1. Patient accountability is given in table 1.3.3 A.

TABLE 1.3.3 A
PATIENT ACCOUNTABILITY IN TRIAL 10

	AmpB	AmBisome at 1 mg	AmBisome at 3 mg
Enrolled with FUO	40	47	46
Afebrile	7	12	17
Discontinued	33	35	29
Severe Adverse Event	4	1	0
Death	2	8	5
Crossover to AmBisome	9	NA	NA
Patient/Investigator Choice	18	26	24
Afebrile, non-neutropenic	3	8	11
Afebrile, neutropenic	6	4	3
Treatment Failure	5	8	7
Toxicity	4	4	2
Other	0	2	1
Enrolled with CM	20	21	18
Excluded from Efficacy	4	7	7
Afebrile	1	3	1
Confirmed Mycosis Treated	7	9	4
Discontinued	12	9	13
Severe Adverse Event	0	0	2
Death	4	7	3
Crossover to AmBisome	6	NA	NA
Patient/Investigator Choice	2	2	8
Treatment Success	0	0	2
Clinical Improvement	0	1	4
Treatment Failure	0	0	1
Toxicity	0	0	1
Other	2	1	0

Two errors in treatment assignment occurred. One patient in each of the two ambisome arms actually received the opposite dose from the one to which he was randomized. These patients were analyzed for efficacy according to the dose actually received.

The mean age of patients was 44 in the FUO stratum and 48 in the CM stratum with a range of in the two strata. In the latter stratum the amphotericin subjects averaged 9 years older than the 1 mg/kg ambisome subjects and 12 years older than the 3 mg/kg ambisome subjects. This was the only demographic variable on which treatments arms were statistically

significantly different within either stratum. The subjects were 79% white, 4% black in the FUO stratum and 93% white, 0% black in the CM stratum. The subjects were 63% male subjects in the FUO stratum and 66% male in the CM stratum. The patients were all seriously ill. Primary diagnoses were similar among the three arms.

Trial 14 Pediatric Empiric Therapy

In trial 14, 64 patients were randomized to amphotericin B, 70 were randomized to 1 mg/kg/day AmBisome, and 71 were randomized to 3 mg/kg/day AmBisome. Nine additional patients were randomized but not included in the study. Four were excluded because no CRF's were filed. Three were excluded because the site investigator deliberately violated the protocol by assigning them to 3 mg/kg/day AmBisome without randomization. (This site, site 7, was dropped from further participation after this episode.) One patient never received any study medication. One received three days of amphotericin B but had no CRF data because the site investigator declared him to be non-neutropenic at baseline.

The subjects in the efficacy evaluable subset were randomized to amphotericin B, ambisome 1 mg, or ambisome 3 mg respectively, at the six sites as follows: (14, 15, 14), (4, 4, 5), (16, 20, 20), (7, 7, 8), (9, 9, 9), and (14, 15, 15). Table 1.3.3 B shows the reasons for discontinuations.

TABLE 1.3.3 B
PATIENT ACCOUNTABILITY IN TRIAL 14

	Ampb	Ambs1	Ambs3
Enrolled	64	70	71
Afebrile	31 (48%)	42 (60%)	47 (66%)
Severe AE or Death	6 (9%)	1 (1%)	3 (4%)
Invest/Pat Decision	27 (43%)	27 (39%)	21 (30%)
Afeb & non-neutropnc	13	12	12
Afeb & neutropenic	3	5	3
Trt failure	3	7	1
Toxicity	6	0	0
Other	2	3	5

The median age of patients was 6 years with a range of 0 to 16 in the three arms. The subjects were 82% white and 59% male. None of the baseline demographic covariates were significantly different among the three arms. The patients' primary diagnoses were similar among the three arms. The amphotericin B arm had higher mean neutrophil count ($76/mm^3$ vs 36 and 58), but the standard deviations of neutrophil counts were large enough that this had a p-value > .1 by Kruskal-Wallis or ANOVA.

1.4 Summary of Methods of Assessment

1.4.1 Double Blind Empiric Therapy Study (Trial 02)

During treatment, body temperatures were measured every 4±1 hours but not within 1 hour of infusion, signs and symptoms were recorded daily, fungal blood cultures were taken every other day, neutrophil counts were checked daily, and other blood and urine chemistry specimens were collected three times a week. Body temperature, blood and urine chemistry, and blood fungal cultures were also taken within 48 hours before start of treatment and 7 days after the end of treatment.

The primary efficacy variable was success, defined as all of 1) survival for 1 week post treatment, 2) resolution of fever during the neutropenic period, 3) resolution of confirmed baseline fungal infection, 4) no emergence of fungal infection up to 1 week post treatment, and 5) no premature discontinuation of drug due to toxicity or lack of efficacy. Secondary variables were duration of fever (measured in three different ways) and incidence of emergent fungal infections.

1.4.2 Double Blind Prophylaxis Studies
Trial 08 Liver Transplant Patients

Patients were evaluated daily during the 5 days of prophylaxis, within 48 hours of final infusion, and periodically for the next 30 days. Blood and urine were obtained before, daily during, 48 hours after, and 2 and 4 weeks after, treatment. Fungal infection was evaluated by culture and/or microscopy of suspected sites. Serological tests for Candida and Aspergillus antigens were performed at days 0, and 6, and 2 days post treatment.

The two primary efficacy variables were rates of incidence of proven fungal infections and of use of non-prophylactic systemic fungal therapy within 30 days of transplant. The secondary variable was the 30-day survival rate.

Trial 13 Neutropenic Hematological Malignancy Patients

Patients were evaluated pre-treatment, post-treatment, and weekly during prophylaxis for fungal infection. Hematology, BUN, serum creatinine, Na^+, and K^+ were measured three times per week. Fungal infection were established empirically as persistent fever > 38° C for 48-96 hours of unknown etiology and non-responsive to systemic broad spectrum antibiotics and/or anti-virals. Fungal cultures were also taken from deep sites at unspecified times and used to establish proven fungal infections. The FDA reviewer found that the applicant's computer files showed that cultures were taken at irregular intervals, averaging every 4.6 days for each patient.

The primary efficacy variables were incidence of proven fungal infections, incidence of fever of unknown origin (FUO), and use of non-prophylactic systemic fungal therapy during treatment. Subsequent to completion of the study, incidence of proven infection was selected as the primary efficacy variable.

The secondary variables were the 30-day survival rate and the time to nephrotoxicity. These latter two were used in the safety assessment.

1.4.3 Open Label Treatment Studies

Trial 10 Adult Empiric Therapy

The primary measure of efficacy was defined differently in the FUO and the CM strata. In the FUO stratum, success was all of the three following conditions: 1) three or more consecutive days without fever (temperature $\leq$ 38°C) lasting until study end, 2) no addition of a systemic antifungal during study, and 3) no development of a documented systemic fungal infection while on study.

Two measures of efficacy were defined in the CM stratum. Clinical cure was complete resolution of the patient's signs and symptoms as assessed by the investigator. Mycological cure was

repeated negative cultures from a previously positive site. The applicant treated these as two measures of success and did not produce a single overall measure of success.

Trial 14 Pediatric Empiric Therapy

Patients were evaluated pre-treatment, post-treatment, and during prophylaxis for fever, concomitant use of systemic anti-fungals, and for identification of pathogens. Hematology, BUN, serum creatinine, Na^+, and K^+ were measured twice a week.

The primary efficacy variables were not specified in the protocol but were selected during analysis of the data. Successful response was defined as at least 3 consecutive days without fever at the end of treatment without any of 1) resolution of neutropenia prior to end of fever, 2) use another systemic anti-fungal drug, and 3) development of a confirmed emergent fungal infection.

The secondary variables were the survival time, the time to nephrotoxicity, rate of adverse events, and rate of abnormal laboratory tests. These latter four were used in the safety assessment.

1.5 Summary of Methods of Statistical Analysis

In the double blind empiric therapy trial, success rates were compared by Cochran-Mantel-Haenszel tests, stratifying by center. Clinical equivalence was declared if two-sided 95% stratified confidence intervals lay within the interval -10%, +10%. Kaplan-Meier curves were used to compare survival and time to nephrotoxicity.

In the other four two arm trials (both prophylactic trials 08 and 13 and open label trials 05 and 09), Student's t-tests and Mann-Whitney tests were used to compare continuous baseline covariates; Fisher's exact test and χ^2 tests were used to compare efficacy response rates, adverse event rates and discrete baseline covariates; and log-rank tests and Kaplan-Meier curves were used to compare survival times, times to response, and times to nephrotoxicity.

In the two three arm OL trials, (trials 10 and 14), ANOVA and Kruskal-Wallis tests were used in place of Student's t-tests and Mann-Whitney tests. (The exact ANOVA model is not specified but the FDA reviewer notes that the one way model corresponds to Kruskal-Wallis.) No attempt to establish clinical equivalence was made in these trials.

Analyses in trial 02 were stratified by center and those in trial 10 were stratified by the FUO-CM strata. None of the other NDA volumes discusses any stratified analyses, specifically stratification on randomization site.

2. Summary of Applicant's Results

The results contained in the applicant's written reports were in almost all cases discrepant from the results obtained using the computer data provided by the applicant. In all cases, the written report was more favorable to ambisome than was the computer data. In the following section, either both sets of results will be presented or only the results from the computer data will be presented. Tables in which such discrepancies were found are marked with a §.

2.1 Double Blind Empiric Therapy Study, Trial 02

Table 2.1 A shows the results for the rates for overall success and for each of the five conditions that constituted overall success. The two arms were comparable on all criteria. The 95% confidence interval for the difference in success rates was (-.064, .078), using results stratified by center. The Breslow-Day test found no statistically significant differences in treatment effect among centers (p-value = .30).

TABLE 2.1 A
CLINICAL ENDPOINT OUTCOMES IN TRIAL 02
WITH P-VALUES FOR SIGNIFICANCE

	AmBisome	Amphotericin
Evaluated	343	344
Overall Success	171 (50%)	169 (49%)
Survived 7 days	318 (93%)	308 (90%)
Fever Resolved	199 (58%)	200 (58%)
Base Infection Cured		
Invest. judgment	5/11	6/10
Applicant judgment	7/9	5/8
No New Infection	294 (86%)	297 (86%)
No Severe Toxicity	294 (86%)	280 (81%)

Table 2.1 B shows the incidence rates of presumed and proven new fungal infections. There was a statistically significantly lower rate of confirmed infections in the ambisome arm. Although the applicant did not remark upon it, there was also a statistically signicantly higher rate of presumed infections.

TABLE 2.1 B
INCIDENCE RATES OF NEW FUNGAL INFECTIONS, TRIAL 02

	AmBisome	Amphotericin	CMH P-value
Evaluated	343	344	
Investigator			
Proven Emergent Infect.	16 (4.7%)	32 (9.3%)	.024
Presumed Infections †	28 (8.2%)	10 (2.9%)	.003
Presumed Infections ††	33 (9.6%)	15 (4.4%)	.007
Applicant			
Proven Emergent Infect.	11 (3.2%)	27 (7.8%)	.011
Presumed Infections †	33 (9.6%)	14 (4.1%)	.004
Presumed Infections ††	38 (11.1%)	19 (5.5%)	.009
Blinded Reviewer			
Proven Emergent Infect.	10 (2.9%)	26 (7.6%)	.009
Presumed Infections †	34 (9.9%)	15 (4.4%)	.005
Presumed Infections ††	39 (11.4%)	20 (5.8%)	.010

† Unknown status counted as not infected
†† Unknown status counted as presumed

2.2 Double Blind Prophylaxis Studies

2.2.1 Trial 08 Liver Transplant Patients

Table 2.2 A shows the results for the rates at 30 days for four variables. These are the two primary efficacy variables: incidence of proven fungal infections and use of non-prophylactic systemic fungal therapy within 30 days of transplant, the secondary variable, death, and the combined occurrence of any failure. AmBisome was statistically significantly better than placebo with respect to frequency of proven fungal infections using the applicant's classification. When the FDA clinical reviewer reclassified fungal infections using the CRF's, there were no statistically significant differences in either the proven or the proven + presumed categories. (See the FDA clinical review.)

TABLE 2.2 A §
CLINICAL ENDPOINT OUTCOMES IN TRIAL 08
WITH P-VALUES FOR SIGNIFICANCE

	AmBisome	Placebo	Fisher Exact P-value
Randomized	43	43	
Proven Fungal Infection			
Applicant's Decision	0 (0%)	6 (14%)	< .01
FDA MO's Proven	0 (0%)	3 (7%)	>.2
FDA's Prov/Presumed	1 (2%)	6 (14%)	.11
Non-Proph. Antifungal	6 (14%)	13 (30%)	.069
Death within 30 days	3 (7%)	6 (14%)	>.2
Death within 31 days	4 (9%)	6 (14%)	>.2
Any Failure	9 (21%)	19 (44%)	.04

2.2.2 Trial 13 Neutropenic Hematological Malignancy Patients

Table 2.2 B shows the results for several different measures of success in the two arms of trial 13, together with Fisher exact p-values for a difference between the two arms. These results fail to show a statistically significant difference between ambisome and placebo with respect to fungal infection.

TABLE 2.2 B §
CLINICAL ENDPOINT OUTCOMES IN TRIAL 13
WITH P-VALUES FOR SIGNIFICANCE

	AmBisome	Placebo	Fisher Exact P-value
Enrolled	75	88	
Proven Fungal Infection	1 (1%)	4 (5%)	.082
Non-Proph. Antifungal	22 (29%)	22 (25%)	>.2
Fever Unknown Origin	4 (5%)	11 (13%)	>.2
Death	11 (15%)	12 (14%)	>.2
Any Failure	34 (45%)	39 (44%)	>.2

Table 2.2 C shows five additional endpoints, tested by the applicant. In the opinion of the FDA medical officer, these endpoints provide evidence of drug activity but not of clinical efficacy. Therefore, no p-values for significance are reported but it is noted that the endpoints systematically favor ambisome.

TABLE 2.2 C §
ENDPOINTS REFLECTING ACTIVITY IN TRIAL 13

	AmBisome	Placebo
Enrolled	75	88
Any Fungal Isolate	12 (16%)	30 (34%)
Colonizations	10 (13%)	18 (20%)
Urinary Infection	1 (1%)	7 (8%)
Stool Infection	5 (7%)	7 (8%)

2.3 Open Label Treatment Studies

2.3.1 Trial 10: Fever of Unknown Origin Stratum

Table 2.3 A shows the results in the fever of unknown origin stratum, defining success as being afebrile for last 3 days of treatment without concomitant use of a systemic antifungal or the development of a confirmed fungal infection. The FDA medical officer has re-classified a few of the outcomes and the table gives both the applicant's and the FDA's classification of outcomes.

TABLE 2.3 A
TRIAL 10, FUO STRATUM
NUMBER (PERCENT) OF SUCCESSFUL OUTCOMES

	AmpB	AmBisome 1mg	AmBisome 3mg
Applicant's Classification			
Successes	22 (55%)	21 (45%)	29 (63%)
Failures	18	26	17
FDA MO's Classification			
Successes	20 (51%)	20 (43%)	28 (62%)
Failures	19	26	17

Chi-square and exacts tests for differences in proportion successful both had p-values of .21. A logrank test for differences in the survival times among the three arms had a p-value of .66.

2.3.2 Trial 10: Confirmed Mycoses Stratum

Table 2.3 B shows the results in the confirmed mycoses stratum, measured by clinical and mycological response. (Not evaluable subjects have been counted as failed clinically or persisted mycosis.) In these analyses, only 39 of the originally randomized patients were actually analyzed. The other 20 subjects were not used because they were ultimately not shown to have a confirmed mycosis.

TABLE 2.3 B §
TRIAL 10, CM STRATUM
NUMBER (PERCENT) OF SUCCESSFUL OUTCOMES

	AmB	AmBisome 1mg	AmBisome 3mg
Randomized	16	13	10
Clinical Response			
Cured	7 (44%)	5 (39%)	5 (50%)
Improved	2 (13%)	1 (8%)	1 (10%)
Failed	7	7	4
Mycological Response			
Eradicated	6 (38%)	6 (46%)	5 (50%)
Persisted	10	7	5

Chi-square and exacts tests for differences in proportion clinically cured or proportion mycologically eradicated had

p-values > .5. A logrank test for differences in the survival times among the three arms also had a p-value > .5.

2.3.3 Trial 14: Empiric Therapy in Children

Table 2.3 C shows the success rates in the pediatric empiric therapy trial, with success defined being afebrile for the last three days without fungal infection or use of another systemic antifungal. As was the case with trial 10, the FDA medical officer has re-classified a few of the outcomes and the table gives both the applicant's and the FDA's classification of outcomes.

TABLE 2.3 C §
CLINICAL OUTCOMES IN TRIAL 14
WITH P-VALUES FOR SIGNIFICANCE

	Ampb	Ambs1	Ambs3	χ^2 P-value
Applicant's Classification				
Success	33 (52%)	44 (63%)	45 (63%)	.35
Failure	31	26	26	
Febrile	12	7	9	
< 3d afeb	18	15	16	
Infection	0	2	1	
Use A.Fung	1	2	0	
FDA MO's Classification				
Success	32 (51%)	44 (63%)	45 (63%)	.35
Failure	31	26	26	

Both of the open label empiric therapy trials had several violations of entry criteria and of treatment protocol. These will be discussed below in the FDA statistical reviewer's comments.

2.4 Safety

2.4.1 Double Blind Empiric Therapy Study

Table 2.4 A shows the most common adverse events in the ambisome arm for trial 02. Six events were identified a priori as interesting.: fever, chills/rigor, increased creatinine, increased BUN, anemia, and hypokalemia. The ambisome group experienced statistically significantly lower rates of four of these: chills, increased creatinine, increased BUN, and hypokalemia. The ambisome arm did not experience an observed

incidence rate greater than 5% more than the amphotericin rate for any adverse event.

TABLE 2.4 A
ADVERSE EVENTS IN TRIAL 02

	AmBisome	Amphotericin	Exact P-value
Enrolled	343	344	
Death	25	36	.18
Fungal related death	4	11	.11
Any Serious AE's	62	77	.18
Any AE's	343	343	1.0
Severe Toxicity	198	272	<.0001
Fever	307	313	>.2
Chills	163	261	<.0001
Hypokalemia	147	174	.047
Nausea	136	133	>.2
Vomiting	109	151	.001
Diarrhea	104	94	>.2
Rash	85	84	>.2
Dyspnea	79	100	.08
Hyperglycemia	79	96	.16
Increas Creatinine	77	145	<.0001
Increas Alk. Phosp.	76	66	>.2

2.4.2 Placebo Controlled Prophylaxis Studies

In the liver transplant trial (08), the ambisome arm had 3/43 deaths compared to 6/42 deaths in the placebo arm. AmBisome produced significantly higher serum alkaline phosphatase (mean = 3.3 * ULN vs 1.5 * ULN) with a p-value < .01. No other laboratory parameters were significantly different between the arms.

In the hematological malignancy trial (13), ambisome produced more nephrotoxicity (9/69 vs 6/85, p =.28) and more hyperbilirubinemia (12/75 vs 7/88, p = .14). The nephrotoxicity in the ambisome arm tended to occur late in prophylaxis as shown in the applicant's Kaplan-Meier plot (not reproduced here) for the onset of this event. No other laboratory parameters were significantly different between the arms.

Table 2.4 B shows the most commonly occurring adverse events in this trial.

TABLE 2.4 B
ADVERSE EVENTS IN TRIAL 13

	AmBisome	Placebo	Exact P-value
Enrolled	75	88	
Death	11	12	>.2
Any Serious AE's	9	9	>.2
Any AE's	67	74	>.2
Abdomen Pain	10	9	>.2
Fever	31	40	>.2
Diarrhea	23	31	>.2
Nausea	31	39	>.2
Vomiting	21	25	>.2
Rash	13	14	>.2

2.4.3 Open Label Empiric Therapy Trials

Table 2.4 C shows the most commonly occurring adverse events in the adult empiric therapy trial (10).

TABLE 2.4 C
ADVERSE EVENTS IN TRIAL 10

	Ampb	Ambs1	Ambs3	Exact P-value
Enrolled	60	68	64	
Death	21	21	19	>.2
Any Serious AE's	20	8	9	.006 *
Any AE's	59	45	48	<.0001 *
Abnorm renal func.	9	2	2	.017 *
Toxic Nephropathy	11	4	3	.025 *
BUN increased	5	1	1	.11
Hypokalemia	20	0	5	<.0001 *
Diarrhea	4	10	12	.13
Nausea	3	6	7	>.2
Dyspnea	2	8	4	>.2

The applicant also plotted mean serum creatinine over time and found there to be a noticeably larger increase for the amphotericin B arm than for either ambisome arm. No tests of significance of this effect were performed.

In trial 14 (pediatric empiric therapy), death and other serious adverse events occurred at comparable rates among the three arms: 4 and 1 out of 64 for deaths and other SAE's on amphotericin B, 3 and 1 out of 70 on 1 mg ambisome, and 7 and 2 out of 71 on 3 mg ambisome.

Among less serious adverse events, incidence rates among the three arms were comparable, except that there was a slight increase in diarrhoea or vomiting in the ambisome groups (3.1% vs 5.7% and 8.5%). Among laboratory events, hypokalemia was less common with ambisome (23.4% vs 1.4% and 16.9%, p=.04). Similarly, nephrotoxicity was less prevalent with ambisome (23.7% vs 9.5% and 13.9%, p=.10). The log-rank test for time to nephrotoxicity was significant with a p-value of .03.

In trial 05 (empiric therapy on neutropenic subjects), ambisome produced statistically significantly less nephrotoxicity than did amphotericin B, as measured by increase of greater than 100% in baseline serum creatinine (1/32 vs 10/31). The log rank p-value for time to nephrotoxicity was .003. AmBisome also produced fewer cases of hypokalemia (4/32 vs 21/31) and hyponatremia (3/32 vs 6/31) and more case of elevated SGPT (10/32 vs 4/31). None of these differences were statistically significant (unadjusted for multiple comparisons). There were also insignificantly fewer subjects with adverse events with ambisome (22/32 vs 25/31) and severe adverse events (7/32 vs 10/31).

For trial 09 (Cryptococcal infections), adverse event experience is summarized in table 2.4 D.

TABLE 2.4 D
ADVERSE EVENTS IN TRIAL 09

	Ampb	Ambs	Fisher p-value
Enrolled	14	15	
Death	2	1	>.2
Other Clinical AE's	4	3	>.2
Serum Creatinine > 3*ULN	1	0	>.2
No Clin or Lab AE's	3	8	.14

Nephrotoxicity in this trial was further analyzed by a repeated measures ANOVA on log transformed changes from baseline in serum creatinine. The exact model is not specified. A

statistically significant treatment effect was found, corresponding to amphotericin B producing increases larger by an estimated factor of 1.37 (p=.003). (This analysis was given in a draft of a paper intended for journal publication. No data has been provided to the FDA for review.)

3. Summary of Applicant's Conclusions

3.1 Double Blind Empiric Therapy Trial

In trial 02, the applicant concluded that ambisome and amphotericin B were equivalent with respect to the primary outcome of overall success and that this equivalence was maintained across all subgroups examined. AmBisome was statistically significantly more effective than amphotericin B in preventing emergent systemic fungal infections. AmBisome also had a significantly more favorable safety profile, characterized by reduced nephrotoxicity and reduced incidence of several other adverse events.

3.2 Double Blind Prophylaxis Studies

In the trial with liver transplants (08), the applicant concluded that a routine five-day prophylactic course of 1 mg/kg/day ambisome significantly reduced the incidence of invasive fungal infections in the early post-transplant period. There were few differences in adverse events noted between the ambisome and placebo groups. Changes in laboratory parameters were limited and similar for ambisome and placebo group patients. A relatively greater increase in serum alkaline phosphates values at Day 30 in the ambisome group may represent a drug effect, but the significance of this finding in a liver transplant population remains to be determined.

In the trial with hematological malignancies (13), the applicant concluded that early systemic use of ambisome reduced the incidence of positive deep fungal infections, although the incidence of such infections in the placebo arm was too small to achieve statistical significance. The ambisome arm experienced no statistically significant increases in total adverse events, serious adverse events, in kidney function laboratory abnormalities, in liver function laboratory abnormalities, or in

serum electrolyte laboratory abnormalities relative to the placebo arm.

3.3 Open Label Treatment Studies

In the adult empiric therapy trial (10), the applicant concluded that in the stratum for fever of unknown origin, ambisome was as effective as amphotericin B in inducing an afebrile state without the need for alternative medication or the appearance of a fungal infection. There was a suggestion of better results with 3 mg/kg dose of ambisome.

In the confirmed mycosis stratum, ambisome was comparable to, or better than, amphotericin B.

In both strata, there were significant decreases in nephrotoxicity, hypokalemia, anemia, and the incidence of drug-related adverse events.

In the pediatric empiric therapy trial (14), the applicant concluded that ambisome offered both an improved efficacy and safety profile: a higher percentage of patients became a febrile while experiencing less nephrotoxicity and hypokalemia.

4. Statistical Reviewer's Comments

There is one major issue and a number of minor problems with the applicant's statistical presentation. The most important issue is the strength of the evidence in support of the claim of demonstrated statistically significant superiority with respect to incidence rate of new fungal infections in the blinded empiric therapy trial (trial 02). The statistical reviewer will demonstrate in section 4.1 below that the evidence is insufficiently strong to warrant applicant's claim. The p-value for the treatment effect reported by the applicant has not been adjusted for multiple comparisons; the observed difference is the same size as the effect on presumed infections that is dismissed as chance variation by the applicant; the effect is not confirmed in other trials; and the estimated magnitude of the effect puts it inside the range specified by the FDA medical reviewers as appropriate for equivalence.

A second group of lesser issues concern the strength of the evidence in support of clinical equivalence with respect to the primary efficacy endpoint of empiric success against fever of unknown origin. The statistical reviewer will demonstrate in section 4.2 below that the applicant's claim is robust to adjustments for uncertainty for the true cause of the fever and to cross-overs from amphotericin to ambisome treatment.

A third group of issues concerns the placebo controlled studies. The statistical reviewer will show in section 4.3 below that there is no adequate evidence of efficacy superior to placebo for prophylaxis.

Finally, section 4.4 will supplement the material provided by the applicant with respect to safety.

4.1 Superiority on Secondary Endpoints in Trial 02

In trial 02, the applicant reported p-values of .024, .009, or .007 with respect to incidence rate of proven new fungal infections, defined either by the investigator, the applicant, or the blinded reviewer. This assertion of achieving statistically significant superiority of ambisome to amphotericin B is made without multiple comparison adjustment. This endpoint is one of

four secondary endpoints listed in the original protocol. (The other three secondary endpoints were duration of fever, duration of fever divided by duration of neutropenic period, and duration of fever truncated at the end of the neutropenic period. Before unblinding the data, the applicant proposed a new set of analyses of the secondary endpoints. In this proposal, emergent fungal infections were specified to include only proven infections. In addition, three more secondary endpoints (time to success, duration of survival, and the binary endpoint of fever resolution during neutropenia) were also proposed. Survival (as a binary endpoint) was also examined and must be counted as a secondary endpoint even though the applicant did not list it explicitly.

The applicant gave a clear rule for establishing clinical equivalence with respect to the single primary endpoint in this trial. However, no explicit rules were specified about determining superiority on any secondary endpoint. At a minimum, the applicant tested for superiority on nine endpoints: the primary endpoint, survival, and the seven secondary endpoints explicitly listed in the analysis plan.

Formal adjustment for multiple comparisons requires a clear list of all possible ways of rejecting the null hypothesis. The applicant performed secondary tests on at least nine endpoints. A number of these endpoints are correlated so Bonferroni adjustment for the number of tests performed may be overly conservative. On the other hand, there were, because of the lack of a fully documented analysis plan, an indeterminate number of combination of these endpoints where one might also have found an apparently clinically meaningful and statistically significant difference. It is quite problematic to assign any valid p-value to the most favorable test result appearing from such a procedure.

In the appendix, the FDA reviewer gives the mathematical details for some multiple comparison adjustments that would be appropriate had the applicant explicitly stated that only the nine secondary endpoints listed above would be tested. P-values that adjust for the correlations among these nine endpoints range from .046 to .063 instead of .007 for reviewer designated proven emergent infections and from .16 to .22 instead of .024 for investigator designated proven emergent infections. Although these values are also not strictly correct, the FDA reviewer

suggests taking these more conservative values, rather than the smaller p-values reported by the applicant, as an initial, subjective estimate of the strength of evidence for a treatment difference in the rate of proven emergent fungal infections. Certainly, the support for a real treatment difference on this endpoint is weaker than asserted by the applicant in table 2.1 B above.

A second observation that raises concern about the strength of the conclusion of a treatment difference in the incidence of new infections is that a treatment difference of the same magnitude but opposite direction is seen with respect to presumed new infections. As table 2.1 B shows, both the difference in number of infections between the arms and the unadjusted p-values for that difference are the same for proven and for presumed infections but the incidence rate for amphotericin is lower than that for ambisome with respect to presumed infections.

One possible explanation of the reversed rate among the presumed infections is that the presumed infections correspond to real infections which were reduced in severity by ambisome. If this were the case, then one might consider fungal infections to be an ordinal variable with 3 levels of severity: proven, presumed, and none. One can test for a treatment effect in this ordinal variable with a Wilcoxon rank sum test. The z-statistic (Wilcoxon statistic - its expected value/ standard error) from such a test is .11, with a p-value of .92. This leads to the conclusion that there is not convincing evidence of a treatment effect taking the form of a reduced severity of fungal infection, as measured by certainty of diagnosis.

If one is prepared to ignore the anomalies with respect to presumed infections and take the results adjusted for exactly nine endpoints at face value, the p-values from this one trial are in the range depending on the details of the adjustment for correlations. It is desirable that effects at this level of credibility be supported by results in other trials. The applicant has two other randomized trials comparing ambisome to amphotericin B in neutropenic, febrile patients unresponsive to antibacterials. The incidence rates for emergent fungal infections in each arm of these trials is given in table 4.1 B. One thing that is noticeable in this table is that the proven infection rate, using the applicant's classifications, is

almost twice as a large in trial 02 (which had 89% adults and 11% children) as it was in trials 10 (adults) and 14 (children). The FDA clinical reviewer also reclassified some of the outcomes in trial 10, producing an average incidence rate in this trial that is comparable to that seen in trial 02.

TABLE 4.1 B
EMERGENT FUNGAL INFECTIONS, BY TRIAL

	AmBisome 1 mg	AmBisome 3 mg	Amp_B	Pooled Rate
Trial 02				
Proven		11/343	27/344	5.5%
Proven+Presumed		44/343	41/344	12.3%
Trial 10, FUO Stratum	3/47	1/46	1/40	3.7%
FDA MO reclassified	3/47	3/46	3/40	6.8%
Trail 14	3/70	2/71	1/64	2.9%

Because of their smaller sample sizes, trials 10 and 14 had power of 15% and 20%, respectively, to detect the difference in underlying probabilities of new fungal infection equal to those observed in the two arms of trial 02. Therefore, these trials could not be expected to yield statistically significant results. However, the power of trials of this size to detect differences is not negligible. In fact, if the odds ratio in emergent fungal infection rates were the same as in trial 02, there would be an 83% chance that the point estimate would be positive in trial 10 and a 75% chance of its being positive in trial 14.

One other problem with trials 10 and 14 needs to be discussed. There was an unavoidable design flaw in the adult empiric therapy trial (trial 10) that would lessen the chance of a confirmatory finding. Subjects receiving amphotericin B were permitted to cross-over to ambisome in the event of severe nephrotoxicity. No cross-over in the other direction was allowed. There were 8/40 such cross-overs in the FUO stratum of trial 10. One of these eight patients developed a confirmed fungal infection and another one died. In the pediatric empiric therapy trial (trial 14) the same type of cross-over occurred inadvertently rather than by design. Seven patients were switched by the investigators from amphotericin B to ambisome. One of these seven developed a fungal infection. Counting patients receiving at least some ambisome as being on amphotericin B (the ITT analysis) creates a bias toward the null

hypothesis of no treatment difference. One can assess how much these potential biases effect the conclusions by doing both an ITT analysis and one in which subjects who switched therapies without dying or developing a fungal infection are discarded.

Table 4.1 C contains (unadjusted) 95% confidence intervals for the differences in incidence rates of confirmed infections and for the odds ratios of confirmed infections. (Differences greater than 0 and odds ratios greater than one are favorable to ambisome.)

TABLE 4.1 C

CONFIRMED INFECTION RATES, BY TRIAL

	Amp_B	Ambs_3	Difference 95% Confidence Limits	Odds Ratio Limits
Trial 02	27/344	11/343	4.6% (1.2%, 8.0%)	2.57 (1.25, 5.27)
Trial 14	1/64	2/71	-1.3% (-6.2%, 3.6%)	.55 (.05, 6.19)
Trial 10, FUO	1/40	1/46	0.3% (-6.1%, 6.7%)	1.15 (.07, 19.1)
As reclassified by FDA MO				
Trial 10, FUO	3/40	3/46	1.0% (-9.9%, 11.8%)	1.16 (.22, 6.11)
Discarding Subjects who Switched				
Trial 14	1/58	2/71	-1.1% (-6.2%, 4.0%)	.61 (.05, 6.85)
Trial 10, FUO	3/33	3/46	2.6% (-9.6%, 14.7%)	1.43 (.27, 7.59)

One can see the following features in this table. First, the smaller sample sizes in trials 10 and 14 lead to wider confidence intervals for treatment differences in these two trials. Second, the findings from trials 10 and 14 are fairly robust with respect to changes made by reclassifying the endpoints and discarding subjects who switch therapy. The confidence intervals get wider as the incidence rates increase but all variations of the confidence intervals for a given trial overlap considerably with all variations of the point estimate lying near the common center of the intervals.

Overall, these three trials do not support the conclusion of ambisome superiority. The incidence rate in trial 10 is

essentially the same in both arms; in trial 14, it was lower in the amphotericin arm. The only suggestion of support for the trial 02 result comes from discarding subjects who switched therapy in trial 10. Leaving those subjects out produced an estimated decrease in incidence rate of infections of 2.6% in the ambisome arm, corresponding to an odds ratio of 1.43. Both on the absolute and the odds ratio scales, these differences are about half the size seen in trial 02. Furthermore, the reviewer would be reluctant to recommend that results from an 'on treatment analysis' be used in preference to those from an ITT analysis.

Another way of looking at the data in these three trials is to consider the incidence of adverse outcomes, defined as either death or a confirmed fungal infection. Table 4.1 D gives the incidence rates for such adverse outcomes in the amphotericin and 3 mg ambisome arms of the three trials, together with confidence limits for the differences in rates and odds ratios. Again, the two smaller trials fail to provide support for the claim of superiority for ambisome. The unadjusted confidence limits in trial 02 lie above the null values of difference equal zero and odds ratio equal one, but both trials 10 and 14 have point estimates of the differences less than zero (and odds ratios less than one).

TABLE 4.1 D
ADVERSE OUTCOME RATES, BY TRIAL

	Amp_B	Ambs_3	Difference	Odds Ratio
			95% Confidence Limits	
Trial 02	50/344	32/343	5.2%	1.65
			(.4%, 10%)	(1.03, 2.65)
Trial 10, FUO	11/40	14/46	-2.9%	.87
			(-22%, 16%)	(.34, 2.21)
Trial 14	4/64	8/71	-5.0%	.53
			(-15%, 4.4%)	(.15, 1.83)
Discarding Subjects who Switched				
Trial 10, FUO	11/34	14/46	-1.9%	1.09
			(-19%, 23%)	(.42, 2.84)
Trial 14	4/58	8/71	-4.4%	.58
			(-14%, 5.5%)	(.17, 2.04)

Finally, there is an issue to be raised with respect to a claim of clinical superiority in a study with the announced goal of establishing clinical equivalence. In all of these trials, it was specified in advance that treatment differences that were, with 95% confidence, no more than 10% would be considered as clinically meaningless. The 95% confidence limits in table 4.1 C for the difference in confirmed infection rates in all three trials lie entirely in the interval -10% to +10%.

There is room for debate here as to whether a change in incidence from 8% to 3% is better measured on an additive or a relative scale. This a philosophical rather than a statistical judgment so this review will address this issue simply by presenting the unadjusted confidence intervals for the odds ratio in tables 4.1 C and 4.1 D. One should recall that these intervals should be widened to allow for the fact that they are the best of nine or more intervals that were examined by the applicant.

4.2 Clinical Equivalence Determination

The efficacy analyses in the open-label empiric therapy trials are done by tests for a difference in treatment effects. These trials have active control arms, amphotericin B. It would be sufficient to demonstrate equivalence by showing that confidence intervals for the difference in success rates are narrow intervals close to or containing zero. Conclusion of equivalence based on failure of the test to demonstrate superiority of amphotericin B is unwarranted. The applicant did use the confidence interval method for supporting equivalence in the double blind empiric therapy trial (see section 2.1 above or table 4.2 A below).

FDA medical reviewers had determined prior to the beginning of trial 02 that $\pm$10% would constitute equivalence with respect to the endpoint of clinical success as defined in sections 1.4.1 and 1.4.3 above.

The FDA statistical reviewer has computed confidence intervals for the differences in success rates between amphotericin B and ambisome, according to the various definitions of success used by the applicant. These results are given in

table 4.2 A for the both the double blind and the open label, amphotericin controlled trials (02, 10 and 14). In trials 02 and 14, which were randomized by site, the estimated difference, its standard error, and its confidence limits, are obtained from the Mantel-Haenszel weighted averages of differences by site. In this table, positive values of the difference in rates, DIFFER., indicate better performance for ambisome than for amphotericin.

In table 4.2 A, a single endpoint of success has been defined as follows. For the CM stratum of trial 10 the applicant failed to define a single primary endpoint. Therefore, the FDA reviewer followed the algorithm used in previous NDA's to define success for this stratum as clinical cure plus mycological response of either eradication or unevaluable. For the FUO stratum of trial 10 and for trial 14 the applicant defined a primary endpoint of success as afebrile for the last three days without use of another anti-fungal or a confirmed fungal infection. In addition, in trial 14, data from site 7 has been excluded. Site 7 was guilty of fraud in the assignment of subjects to treatment arms. In addition, the confidence intervals for these three sets of results have been widened to adjust for the presence of two doses of ambisome per trial to compare to the control arm. Results for trial 02 are also included, using the applicant's computer data and the FDA's computation of the difference in rates, weighted by site. The latter were not adjusted for multiple doses since only one ambisome arm was included in this trial.

TABLE 4.2 A
ACTIVE-CONTROL, EMPIRIC THERAPY TRIALS
CONFIDENCE INTERVALS FOR DIFFERENCES IN SUCCESS RATES

TRIAL 02	DIFFER.	LOWER	UPPER
Cure at 3 mg	0%	-7%	8%
TRIAL 10	DIFFER.	LOWER	UPPER
CM STRATUM			
Cure at 1 mg	-5%	-47%	36%
Cure at 3 mg	6%	-39%	51%
FUO STRATUM			
Cure at 1 mg	-8%	-29%	13%
Cure at 3 mg	11%	-10%	32%
TRIAL 14	DIFFER.	LOWER	UPPER
Cure at 1 mg	9%	-6%	23%
Cure at 3 mg	6%	-12%	21%

The results in the four active-controlled, empiric therapy trials are generally supportive of the applicant's claim of clinical equivalence.

This was also demonstrated at the 1 mg/kg dose in trial 14. However, at the 1 mg/kg dose, ambisome was not shown to be clinically equivalent to amphotericin B even up to limits of $\pm$20% in the FUO stratum of trial 10.

In the CM stratum of trial 10, neither dose could be judged to be clinically equivalent to amphotericin B. However, in both strata of trial 10, the results went in the same direction: the 1 mg/kg dose was estimated to be inferior to amphotericin and the 3 mg/kg dose was estimated to be superior.

Two potential problems remain with this conclusion. First, there was concern with the problem of subjects not actually having any fungal infections and thus recovering or failing for reasons unrelated to their anti-fungal treatment. It is unclear whether the pre-specified limits of $\pm$10% for equivalence were intended to already contain an allowance for this effect. Second, there were the matter of subjects switching therapies.

There first issue creates a bias toward apparent equivalence because subjects not actually having fungal infections should have the same success rate in both arms. Any observed difference in the arms is a weighted average of the difference in success rate on fungally infected subjects and the difference in success rate on non-fungally infected subjects, with the latter difference being zero. One can get a tentative estimate of the potential magnitude of this problem from table 4.2 B, which shows that about 30-40% of subjects in trial 02 developed non-fungal infections during treatment. These numbers do not directly measure the number of baseline infections which were non-fungal but they can serve as a starting point for a sensitivity analysis on the effect of the presence of non-fungal infections on confidence limits for the rate of overall success as defined in the protocol.

TABLE 4.2 B
NON-FUNGAL INFECTIONS DURING TREATMENT, TRIAL 02

	AmBisome	Amphotericin
Enrolled	343	344
Bacterial	90 (26%)	75 (22%)
Viral	23 (7%)	21 (6%)
Protozoal	3 (<1%)	1 (<1%)
Other	18 (5%)	15 (4%)

It is straight-forward to adjust the confidence intervals for the difference in success rates to adjust for any postulated proportion of the subjects being truly fungally infected at baseline. If one knew that a fraction w of the subjects in an empiric study were fungally infected and that the other 1-w fraction were not and that the 95% confidence limits for the difference in cure rates between the two arms were q% to p%, then the 95% confidence limits for the difference in cure rates among truly fungally infected patients would be (q/w)% to (p/w)%. This would hold even without knowing which patients were fungally infected.

If one makes such adjustments, one finds that the conclusions of clinical equivalence are fairly robust. Trial 02 would have confidence limits for the difference in treatment effect inside $\pm$20% if only a third of the subjects were fungally infected and inside $\pm$10% if three quarters of the subjects were fungally infected. Both trials 10 and 14 would also have

confidence limits for the difference in treatment effect inside ±20% for the 3 mg dose of ambisome if three quarters of the subjects were fungally infected. Trial 14 would also have confidence limits for the difference in treatment effect inside ±20% for the 1 mg dose if even half the subjects were fungally infected.

Finally, as mentioned in the previous section, subjects switching therapies in trials 10 and 14 created a potential bias in the ITT analysis toward the hypothesis of equivalence. If one considers any amphotericin B subjects who crosses over to be ipso facto failures, the asymmetry of the allowed cross-over biases the design in favor of ambisome superiority. Two analyses that explore the effect of this design bias are possible. Method A is to exclude cross-over subjects from the analysis. Method B is to count any subject who crosses over as on amphotericin B if he is a success and as on ambisome if he is a failure. In neither trial do the switches change the conclusions with respect to clinical equivalence of the arms: the confidence limits for the difference in treatment effect remained inside ± 10% in trial 14 and inside ±20% in trial 10, regardless of method.

4.3 Placebo Controlled Trials

The applicant's results in tables 2.2 A, B, and C above examined a number of endpoints and found some of them to be statistically significant. All of these tests were performed without adjustment for multiple testing.

Some of these issues have already been dealt with in passing in the above tables. In trial 08, the FDA clinical reviewer assessed the fungal serology, cultures, and biopsies reported by the applicant to determine how many proven and possible fungal infections were present. These results have been given in table 2.2 A and changed the unadjusted p-values for proven infections from (using proven only) or .24 (using proven plus possible).

For trial 13, the applicant's reported results were discrepant from those trial 08. In trial 13, the primary endpoint did not even have a point estimate that favored ambisome over placebo, much less show a statistically significant

difference. The applicant presented an argument for several endpoints showing a pattern favorable to ambisome to a statistically significant extent. In the FDA clinical reviewer's judgment, those four of these endpoints which are listed in table 2.2 C indicate drug activity, not clinical efficacy. (See the FDA medical review.) As such, they are only secondary, supportive evidence. The remaining endpoint supporting clinical efficacy was proven fungal infections, in table 2.2 B. This endpoint was one of four components of overall failure. Therefore, the Fisher exact p-value reported in the table should be adjusted from in assessing significance. In consequence, trial 13 shows no evidence of a difference in clinical efficacy between ambisome and placebo.

4.4 Differences in Serum Creatinine and Other Lab Parameters

The applicant's main analysis of the differences in serum creatinine levels (and other important laboratory parameters) focused only on whether laboratory values crossed various prespecified thresholds at any time during treatment. With respect to temporal patterns, the applicant merely gave plots of mean levels over time. The FDA reviewer has supplemented these with three additional items: 1) confidence intervals for differences over time, 2) repeated measures tests for treatment effects, and 3) plots comparing serum creatinine levels between subjects continuing on treatment and those discontinuing treatment. Similar analyses were performed for alkaline phosphatase, SGOT, SGPT, BUN, bilirubin, hemoglobin, and serum potassium.

Figure 4.6 i shows difference between the ambisome arm minus the amphotericin arm in the mean change from baseline over time of serum creatinine in trial 02. The 95% confidence intervals at each time point (unadjusted for multiple looks at the data) are also given. Figures 4.6 ii-iv show comparable graphs for trials 10, 14, and 13. In these figures ii and iii, there are two curves for the differences between each of the two ambisome doses minus amphotericin.

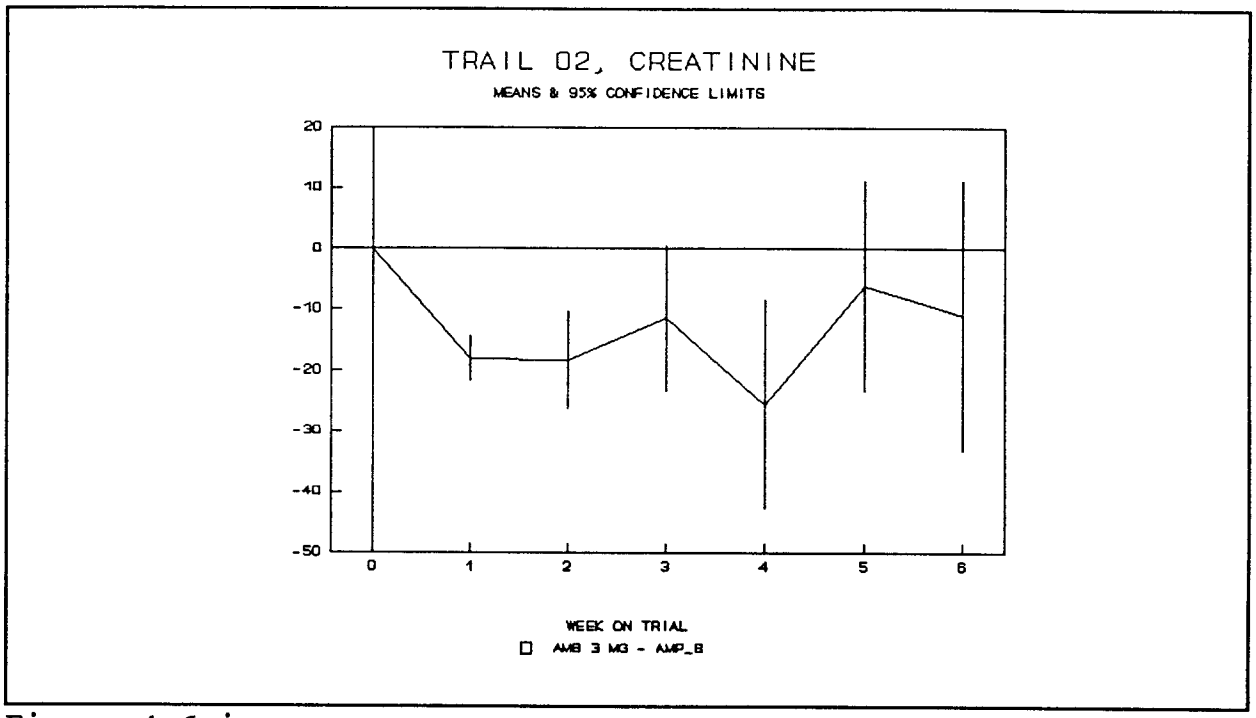

Figure 4.6 i

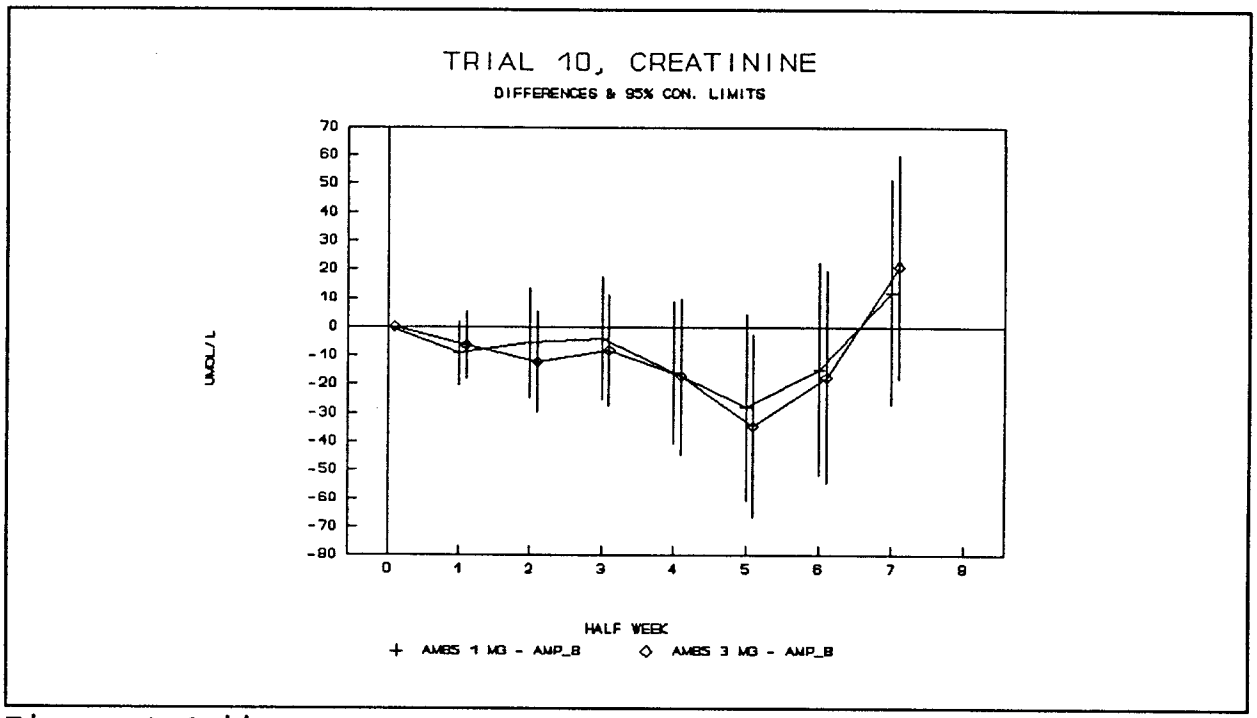

Figure 4.6 ii

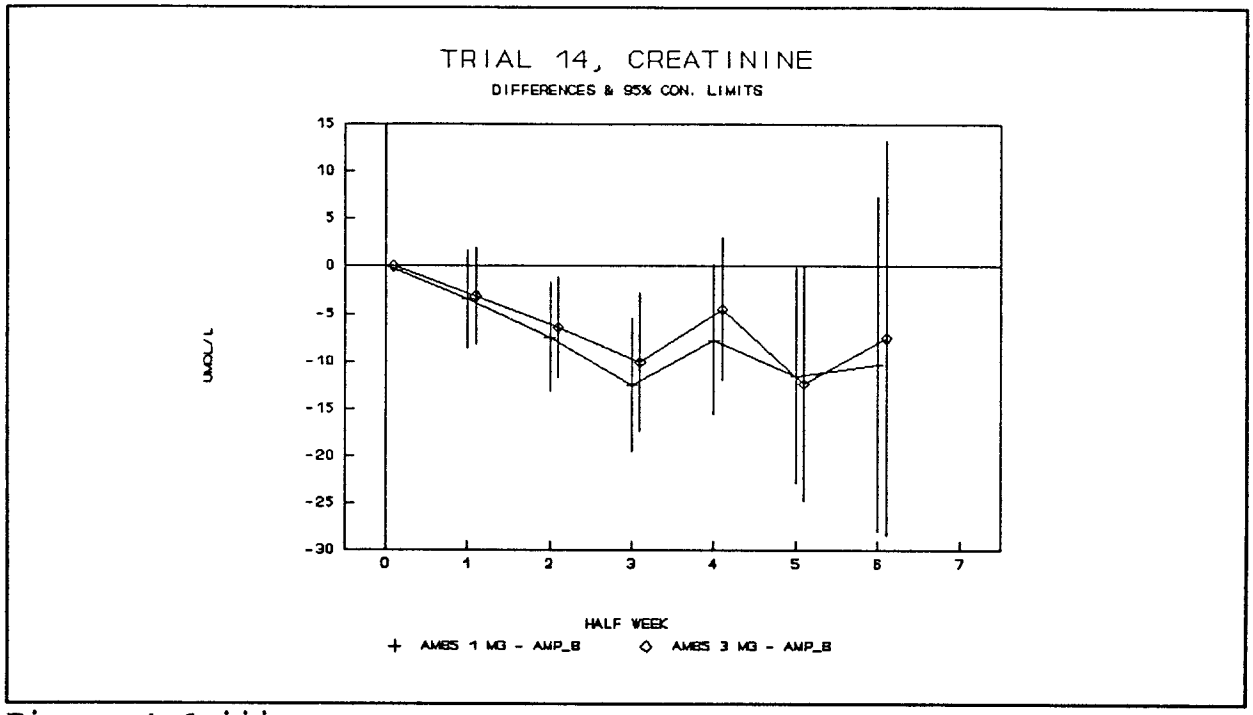

Figure 4.6 iii

One can see that in all these graphs there is a pattern of lower levels of creatinine in the ambisome arms, often approaching statistical significance (using separate tests at each time point). In figure iv, the curve is the difference between ambisome and placebo. Here it should be noted that, although the ambisome tends to have higher levels than placebo, the difference is closer to zero than in the comparison with ambisome and amphotericin. Creatinine levels with ambisome were about lower than with amphotericin in the two trials with adult subjects: trial 02 and both arms of trial 10, while creatinine levels with ambisome were higher than with placebo in trial 13.

A similar graph for trial 08 was also examined but is not included here because the duration of trial 08 was only 5 days and no consequential differences between the ambisome and placebo arms were observed.

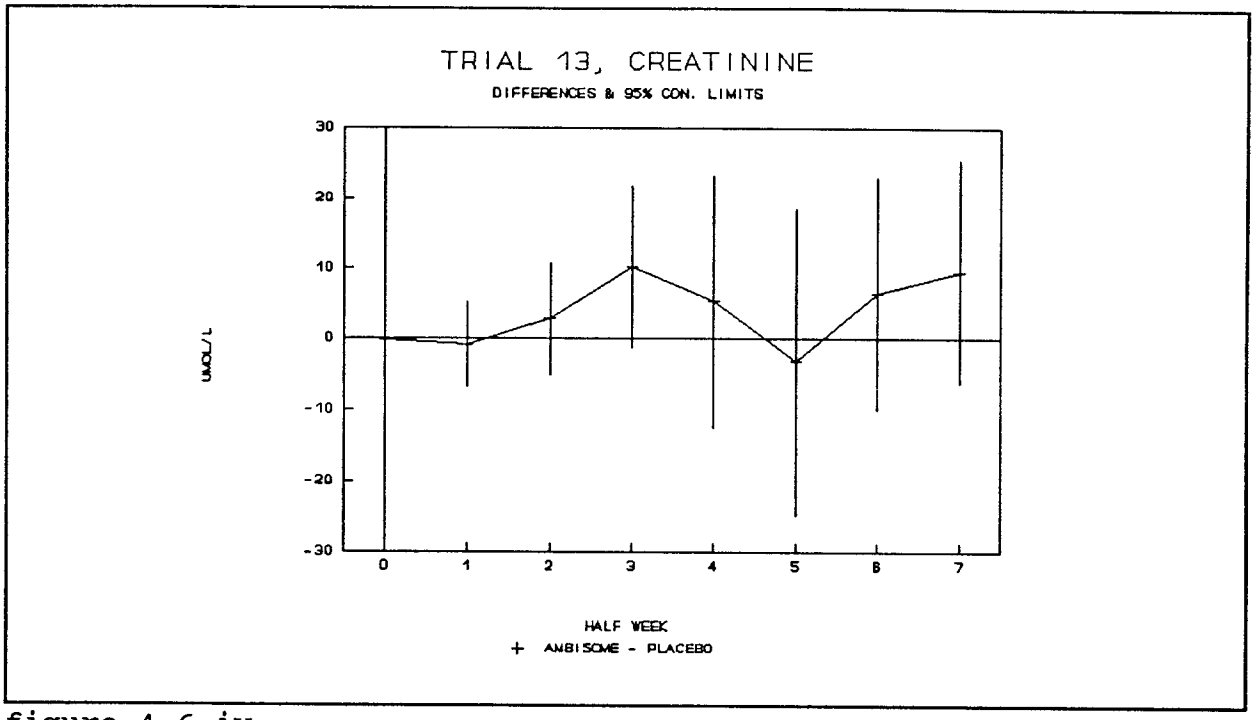

figure 4.6 iv

Formal tests for treatment effects over time were performed using repeated effects multivariate analysis of variance (MANOVA) with a single main effect for treatment. In order to reduce concerns about the sensitivity of results to possible drop-outs, missing laboratory data were replaced by the previous observation carried forward. The results are given in table 4.4 A. The trajectory of serum creatinine was found to be statistically significantly lower for ambisome than for amphotericin in three out of five trial arms tested, with p-values of in the other two instances. The trajectory of serum creatinine in trial 13 was significantly higher for ambisome than for placebo, suggesting that ambisome may reduce but does not eliminate the risk of nephrotoxicity.

There was also some evidence of higher levels of potassium in the ambisome arms than in the amphotericin arms. Significantly lower levels of BUN and sodium were found in one trial, trial 02.

TABLE 4.4 A

MANOVA P-VALUES FOR TREATMENT EFFECTS IN CHANGE OF LAB PARAMETERS FROM BASELINE

AmBisome VS AMPHOTERICIN

Trial	02	10	10	14	14
AmBisome Dose	3 mg	1 mg	3 mg	1 mg	3 mg
Creatinine	.0001 *	.13	.10	.024 *	.0095 *
BUN	.0001 *	>.2	>.2	>.2	>.2
Hemoglobin	>.2	.046 *	>.2	>.2	>.2
Sodium	.0008 *				
Potassium	.14	.003 *	>.2	.18	>.2
Bilirubin	>.2	.19	.13	>.2	.17

No p-values < .15: Alkaline Phosphatase, SGOT, SGPT, Magnesium

AmBisome VS PLACEBO

Trial	13	08
Creatinine	.0245 *	>.2

No p-values < .15: Alkaline Phosphatase, SGOT, SGPT, Potassium, Bilirubin, BUN, Hemoglobin

An additional check for significant treatment differences in lab parameters was performed by t-tests for differences between the ambisome and amphotericin arms on each of the minimum and maximum over time of each lab parameter. The results of these t-tests are summarized in table 4.4 B. The two consistent findings are lower levels of maximum creatinine and higher levels of minimum potassium with ambisome.

TABLE 4.4 B

P-VALUES FOR TREATMENT EFFECTS IN EXTREMES OF LAB PARAMETERS IN AMPHOTERICIN TRIALS

Trial	02	10	10	14	14
AmBisome Dose	3 mg	1 mg	3 mg	1 mg	3 mg
Max Creat	.0001 *	.01 *	.001 *	.006 *	.03 *
Min Potassium	.006 *	.001 *	.005 *	.0001 *	.05 *

Other p-values < .05 scattered with no parameter having p<.05 on two trials

5. Statistical Reviewer's Summary

There is reasonably robust evidence of clinical equivalence of amphotericin B and ambisome in the treatment of febrile neutropenic patients. The lower confidence limits for the difference between ambisome and amphotericin B on the combined endpoint of success were consistently above -10% for all three active controlled trials and remained at least above -20% under sensitivity analyses exploring the effects of including non-fungally infected subjects in the treatment sample and of switching subjects from amphotericin to ambisome. No noticeable interactions with baseline covariates were found in subset analyses performed by the applicant (details of the latter are not included in this review).

The evidence that ambisome is at least as effective as amphotericin is confirmed with respect to the clinically firmer endpoint of emergent fungal infections. This was most evident in the largest study where the rate of proven plus presumed new infections was the same in the amphotericin arm and in the ambisome arm. There is, however, inadequate scientific grounds for a firm conclusion that ambisome is clinically superior to amphotericin with respect to the new infection rate. A higher incidence rate of proven infections on amphotericin B was balanced by a higher incidence rate of presumed infections on ambisome. Given these equal but opposite differences in rates, the observed pattern of proven infections in trial 02 is not adequate to compel belief by itself without a confirmatory trial and the other two trials fail to support a claim of superiority. All three trials give results compatible with differences in the new infection rate of no more than $\pm$10%.

There is reasonable evidence that ambisome is at least as safe as amphotericin B with respect to adverse events. There is, moreover, clear evidence that ambisome is associated with lower levels of nephrotoxicity and hypokalemia.

There is inadequate evidence to support the claim that ambisome is superior to placebo with respect to prophylaxis in neutropenic transplant or chemotherapy patients.

/S/

Thomas Hammerstrom, Ph.D.
Mathematical Statistician

Concur: Dr. Flyer /S/ 12/1/97

cc:
Archival NDA #50-740
HFD-530
HFD-530/Ms. Sage (via Team Links)
HFD-530/Dr. Feigal (via Team Links)
HFD-530/Dr. Freeman (via Team Links)
HFD-530/Dr. Goldberger (via Team Links)
HFD-590/Ms. Frank
HFD-530/Dr. Murray
HFD-590/Dr. Korvick
HFD-590/Dr. Meyerhoff
HFD-725/Dr. Flyer
HFD-725/Dr. Hammerstrom
HFD-725/Dr. Harkins
HFD-725/Ms. Shores

APPENDIX 1
MULTIPLE COMPARISONS WITH CORRELATED DATA

Efron (Biometrika, March 1997, pp. 143-158) gives the following formula for the probability that the largest of J correlated T-statistics exceeds c:

$$P(T_{max} > c) = 1 - \Phi(c) + \phi(c) \Sigma_{j=2J}\{\Phi(cL_j/2) - .5\} / (c/2)$$

where Φ = standard normal cumulative, ϕ = standard normal density, L_j = arccos(Correlation(T_{j-1}, T_j)).

The nine endpoints tested for superiority included four binary endpoints: survival, fever resolution, emergent fungal infection, and overall success. They also included time to success and duration of survival with a correlation of .974. Finally, they included three correlated measures of time to fever resolution. The correlations among the three measures of duration of fever are as follows:

	F	N
R	.446	.197
F		.931

where
R = Relative duration of fever
F = Time to fever resolution and
N = Fever duration while neutropenic

One reasonable overall adjustment for multiple inference with these correlated endpoints is to treat the four binary endpoints as requiring four separate Bonferroni adjustments, to treat the times to success and death as producing one Efron style adjustment, and to treat the three measures of fever duration as requiring a second Efron style adjustment. (Because the tests for treatment effects on the binary endpoints were done using the Fisher exact or Cochran-Mantel-Haenszel rather than Student t-tests, Efron's formula cannot be used to obtain an adjusted p-value for all nine endpoints.) This leads to computations as follows.

Using Efron's formula, the probability, under the null hypothesis of no treatment difference, that the largest of the three t-statistics based on these fever duration endpoints would be significant at two-sided level .024 (.007) is actually .038 (.011) respectively. Similarly, the probability that the larger

of the two t-statistics for time to success and time to death would be significant at two-sided level .024 (.007) is .026, (.008) respectively.

The FDA reviewer used the Bonferroni inequality to combine the four Fisher exact or CMH tests, the maximum of the three t-tests for fever duration variables, and the maximum of the two t-tests for times to success or death. This gives that the probability, under the null hypothesis, that one of the nine p-values is less than .024 (or .007) is $\leq$ 4*.024 +.038 +.026 = .16 (or $\leq$ 4*.007 + .011 + .008 = .046). A more conservative adjustment for multiple inference would be the simple Bonferroni adjustment for nine endpoints. This would replace the unadjusted p-vale of .024 (or .007) by one of .22 (or .063).