CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

021164Orig1s000

STATISTICAL REVIEW(S)

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	21-164
Drug Name:	Gepirone ER
Indication(s):	Major Depression Disorder
Applicant:	Fabre-Kramer Pharmaceuticals, Inc.
Date(s):	May 1, 2007
Review Priority:	Priority
Biometrics Division:	Biometrics I (HFD-710)
Statistical Reviewer:	Fanhui Kong
Concurring Reviewers:	Peiling Yang, Kooros Mahjoob, Hsien Ming J. Hung
Medical Division:	Division of Psychiatry Products
Clinical Team:	Earl Hearst, Ni Aye Khin, Mitchell Mathis, Thomas Laughren
Project Manager:	Renmeet Grewal

Keywords: MDD, ANOVA, ANCOVA, LOCF, Meta-Analysis, MMRM, Sexual Dysfunction, Suicidality

Table of Contents

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

In this submission, two pivotal short-term gepirone ER studies, FKGBE007 and FKGBE008 were conducted in the United States. In these studies, the primary objective was to evaluate the therapeutic efficacy of gepirone ER tablets in comparison with placebo at the end of an 8-week treatment period in subjects with MDD. The sponsor also conducted *post hoc* meta-analyses on ten previously conducted short-term studies along with two current studies FKGBE007 and FKGBE008 and reevaluated the relapse prevention Study 28709. In addition, the sponsor also intended to seek a claim for lack of sexual dysfunction for gepirone ER in labeling.

In Study FKGBE007, the effectiveness of gepirone ER in the treatment of adult patients with MDD is supported by the primary efficacy analysis using LOCF, and the analyses using OC and MMRM. Further *post hoc* subgroup analyses suggest that the treatment effect appeared to be mainly driven by Caucasians and female patients. In addition to Study FKGBE007, there is one more positive Study 134001 among a total of 10 previously conducted short-term gepirone ER studies.

The fact that **only two out of a total of 12 short-term efficacy studies are positive in support of the effectiveness of gepirone ER during a period of 12 years raises concerns on the reproducibility of the treatment effect observed**. Statistical procedures using meta-analysis or mixed-effects model on different combinations of the remaining 10 studies (i.e., after excluding the two positive studies) do not seem to provide further evidence supporting the effectiveness of gepirone ER over placebo. Furthermore, in three out of the five so called "lack of assay sensitivity" studies, active control significantly outperformed gepirone ER (Table 3.10). Active control significantly outperformed placebo in 2 out of these 5 studies (Table 3.10). In the positive study FKGBE007, the treatment effect seems to be driven by Caucasians and female patients only. The reevaluation of the relapse prevention Study 28709 does not provide valid evidence supporting the efficacy of the treatment. The collective evidence seems to provide only a weak support for the effectiveness of gepirone ER in the treatment of MDD among adults.

When comparing patient sexual functioning, although gepirone ER seems to be superior to its active comparators (fluoxetine or paroxetine) in some actively controlled studies, there does not seem to be enough consistent evidence in supporting the claim that gepirone ER did not reduce the quality of patient's sexual functioning (b) (4)

1.2 Brief Overview of Clinical Studies

Gepirone ER (Org 33062 ER) is a 5-HT_{1A} agonist that has been under development as an antidepressant, both in IR and ER formulations. An NDA for the ER formulation was originally submitted on September 30, 1999, but was refused to be filed by FDA. It was resubmitted on May 18, 2001, however, a non-approvable (NA) letter was issued on March 15, 2002, citing inadequate evidence of efficacy (the agency considered one ER study [134001] and one IR study [03A7A-003] positive, but required one additional positive ER study). The NDA was resubmitted on December 23, 2003 with data from a randomized withdrawal study. However, the agency considered this randomized withdrawal trial to be problematic, and issued a second NA letter on June 23, 2004, citing a need for a robustly positive short-term trial and a

positive randomized withdrawal trial. The sponsor now proposes to submit the results of two additional short-term studies (007 & 008) in major depression disorder (MDD) in support of an NDA for MDD.

In this submission, two pivotal studies FKGBE007 and FKGBE008 were submitted for the evaluation of the efficacy of gepirone ER in the treatment of MDD. Study FKGBE007 was conducted between October 8, 2003 and August 21, 2004 in the United States. It was a Phase III, multicenter, randomized, double-blind, placebo-controlled, flexible dose study in which 248 (124 in gepirone ER; 124 in placebo) moderately to severely depressed outpatients received gepirone ER or placebo once each morning with food for 8 weeks (56 days). The primary analysis was on the change in HAMD-17 total score based on the ITT set and on the LOCF approach. Of the 248 subjects who received either gepirone ER or placebo, 199 (102 in gepirone ER and 97 in placebo) completed the study. Sixty five percent (65%) of the patients were Caucasian and 23% were Africa-Americans. Thirty one percent (31%) were male and 69% were female. All the patients were between 18 and 64 years of age (inclusive).

Study FKGBE008 was conducted between October 20, 2003 and August 23, 2004 in the United States with the exact design as Studies FKGBE007. Of the 206 subjects randomized and received at least one dose of study medication (102 in gepirone ER and 104 in placebo), 159 (77 in gepirone ER and 82 in placebo) completed the study. Eighty one percent (81.4%) of the patients were Caucasian and 8.5% were Africa-Americans. Thirty six percent (35.7%) were male and 64.3% were female. All the patients were between 18 and 64 years of age (inclusive).

In order to provide further evidence for the efficacy of gepirone ER in the treatment of MDD, the sponsor conducted meta-analyses on the previously conducted studies along with two current studies FKGBE007 and FKGBE008. Of the 13 relevant adequate and well controlled studies identified above, the meta-analyses included all 12 short-term studies but excluded the relapse prevention Study 28709 because its design was incompatible with the short-term studies. A total of three meta-analyses were conducted. One was on the five relevant short-term so called “supportive studies” without the two positive studies (FKGBE007 and 134001) and one on these “supportive studies” together with the two positive studies. An additional meta-analysis was conducted on all 12 short-term studies.

1.3 Statistical Issues and Findings

In this submission, the sponsor conducted 2 pivotal short-term gepirone ER studies, FKGBE007 and FKGBE008. Only FKGBE007 is positive and FKGBE008 is not. In these studies, the primary efficacy measure was the change from baseline to the end of study of the HAMD-17 total score. The treatment efficacy was analyzed using ANCOVA with LOCF data. The sponsor also conducted *post hoc* meta-analyses on the previously conducted studies along with two current studies FKGBE007 and FKGBE008. Among the three meta-analyses performed, the sponsor suggested to adopt the positive one which included two positive studies and five so called “supportive” studies.

In Study FKGBE007, the effectiveness of gepirone ER in the treatment of adult patients with MDD is supported by the primary efficacy analysis using LOCF, and the sensitivity analyses using OC and MMRM. Further *post hoc* subgroup analyses suggest that the treatment effect was mainly driven by Caucasians and female patients only.

Only two (Studies FKGBE007 and 134001) out of a total of 12 short-term efficacy studies are positive in support of the effectiveness of gepirone ER in adults with MDD. Statistical procedures using meta-analysis or mixed-effects model on different combinations of the remaining 10 studies (i.e., after excluding the two positive studies) did not provide further information supporting the efficacy of gepirone

ER. Active control significantly outperformed gepirone ER in three out of the five so called “lack of assay sensitivity” studies. Active control significantly outperformed placebo in 2 out of these 5 studies. In the positive study FKGBE007, the treatment effect seems to have been driven by Caucasians and females only. The reevaluation of the relapse prevention Study 28709 does not provide valid evidence supporting the efficacy of the treatment. The collective evidence seems to provide only a weak support for the effectiveness of gepirone ER in the treatment of MDD among adults.

1.4 Safety Issues and Findings

Although the relative risk of developing sexual dysfunction-related AE in subjects treated with gepirone ER was not significantly different from placebo, the non-inferiority of gepirone ER cannot be determined due to the lack of non-inferiority margin. And it cannot be determined if the advantage of gepirone ER in not increasing sexual dysfunction related AE was consistently over the studies because no information for specific study was provided.

In summary, although gepirone ER seems to be superior to its active comparators (fluoxetine or paroxetine) in some active-controlled studies, there does not seem to be enough evidence to support the claim that gepirone ER did not reduce the quality of patient’s sexual functioning in these studies according to the criteria set by the Agency.

2. INTRODUCTION

2.1 Overview

This submission contains two new pivotal positive efficacy studies FKGBE007 and FKGBE008 for the evaluation of the therapeutic efficacy of gepirone ER tablets in comparison with placebo at the end of an 8-week treatment period in subjects with MDD (diagnosed according to DSM-IV criteria). The secondary objectives were to describe the safety profile of 8 weeks of treatment with gepirone ER in comparison with placebo in subjects with major depression and, if sufficient data were available, to evaluate the therapeutic efficacy of gepirone ER in subjects with atypical depression.

In Study FKGBE007, a total of 248 subjects were randomized and received at least one dose of study medication (124 in the gepirone ER group and 124 in the placebo group); 238 subjects were analyzed for efficacy in the intent-to treat (ITT) population (116 in the gepirone ER group and 122 in the placebo group). Male or female subjects 18 to 64 years of age who met DSM-IV criteria for moderate to severe MDD, had a HAMD-17 score of ≥20 at screening and baseline, and had significant daily dysphoria for 4 weeks prior to screening.

All subjects participated in a placebo washout screening period during which subjects received 1 placebo tablet/day for 4 to 7 days. Subjects were titrated to study medication as follows: 1 gepirone ER 20 mg tablet or 1 placebo tablet on Days 1-3 and 2 gepirone ER 20 mg tablets or 2 placebo tablets on Days 4-7. Thereafter, the doses on the preferred dosing schedule were 40 to 60 mg gepirone ER or placebo qd (2-3 tablets) on Days 8-14 and 40 to 80 mg gepirone ER or placebo qd (2-4 tablets) on Days 15-56.

The primary efficacy measure was the change from baseline in HAMD-17 total scores at Week 8. The secondary measures included various HAMD scores, MADRS scores, various CGI scores; numbers of HAMD, MADRS, and CGI responders; number of HAMD-17 remitters; and number of subjects who discontinued due to lack of efficacy.

In Study FKGBE008, a total of 206 subjects were randomized and received at least one dose of study medication (102 in the gepirone ER group and 104 in the placebo group); 199 subjects were analyzed for efficacy in the intent-to treat (ITT) population (99 in the gepirone ER group and 100 in the placebo group). Since this study does not provide evidence supporting the efficacy of gepirone ER, the efficacy results will not be reviewed in the following.

2.2 Data Sources

The Clinical Study Reports and SAS transport data sets for the studies were provided in electronic form in \\CDSESUB1\N21164\N_000\2007-05-01.

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Baseline Demographic Characteristics

The patient baseline demographic characteristics are summarized in Table 3.1 for Study FKGBE007. Demographic characteristics were similar for both treatment groups in the ITT population. The mean (SD) age of all subjects was 38.1 (11.20) years and ranged from 18 to 64 years. The majority (nearly 70%) of subjects in both treatment groups were female, and the distributions of race were also comparable for the two treatment groups, with the majority of subjects being Caucasian (64.7% overall) or Black (23.1% overall). Among the 27 subjects listed as other race, 25 were Hispanic, one was Indian/Hispanic and one was Asian/Hispanic.

Table 3.1 Demographic Characteristics for Study FKGBE007 at Baseline

Characteristic	Treatment Group		Overall
	Org 33062 ER (n=116)	Placebo (n=122)	(N=238)
Age (yrs)			
Mean (SD)	38.2 (11.43)	38.0 (11.03)	38.1 (11.20)
Median	39	38	38.5
Minimum, Maximum	18, 58	18, 64	18, 64
Gender			
Male	37 (31.9)	37 (30.3)	74 (31.1)
Female	79 (68.1)	85 (69.7)	164 (68.9)
Race			
Caucasian	73 (62.9)	81 (66.4)	154 (64.7)
Black	28 (24.1)	27 (22.1)	55 (23.1)
Asian	--	2 (1.6)	2 (0.8)
Other	15 (12.9)	12 (9.8)	27 (11.3)

Source: Table 12 of sponsor's Clinical Study Report of Study FKGBE007.

3.1.2 Baseline Disease Characteristics

The two treatment groups also had similar disease characteristics at baseline. Overall, episodes of depression upon entering the study had lasted >12 months for 39.9% of the subjects, between 1 and 6 months for 36.6% of the subjects, and between 7 and 12 months for 23.5% of the subjects. Subjects had an overall mean (SD) age of 27.7 (11.92) years when they experienced their first episode of depression. Most subjects (58.8% overall) were suffering from recurrent depression with full recovery when they entered the study; 22.3% of subjects were suffering from their first episode of depression at baseline. The majority of subjects (95.4% overall) did not have a comorbid anxiety disorder; however, the incidence of comorbid anxiety disorder was higher for the gepirone ER (Org 33062) group compared to the placebo group (6.9% vs. 2.5%, respectively).

Table 3.2 Patient Baseline Illness Characteristics in Study FKGBE007

Characteristic	Treatment Group		Overall
	Org 33062 ER (n=116)	Placebo (n=122)	(N=238)
Duration of present episode			
1 to 6 months	43 (37.1)	44 (36.1)	87 (36.6)
7 to 12 months	26 (22.4)	30 (24.6)	56 (23.5)
>12 months	47 (40.5)	48 (39.3)	95 (39.9)
Age at first episode of depression (yrs)			
Mean (SD)	27.8 (12.28)	27.7 (11.61)	27.7 (11.92)
Median	25	25	25
Minimum. Maximum	7, 56	7, 61	7, 61
Course of illness			
First episode	27 (23.3)	26 (21.3)	53 (22.3)
Chronic (full criteria for MDD met)	12 (10.3)	8 (6.6)	20 (8.4)
Recurrent with partial recovery	10 (8.6)	15 (12.3)	25 (10.5)
Recurrent with full recovery	67 (57.8)	73 (59.8)	140 (58.8)
Comorbid anxiety disorder			
No	108 (93.1)	119 (97.5)	227 (95.4)
Yes	8 (6.9)	3 (2.5)	11 (4.6)
Social phobia (social anxiety disorder)	7 (6.0)	3 (2.5)	10 (4.2)
Posttraumatic stress disorder	1 (0.9)	0	1 (0.4)

Source: Table 13 of sponsor's Clinical Study Report of Study FKGBE007.

3.1.3 Patient Discontinuation

A total of 248 subjects were randomized and treated in this study, 124 subjects each in the gepirone ER and placebo groups. Two hundred and thirty-eight subjects (116 in gepirone vs. 122 in placebo) were in the ITT population, 220 subjects (103 in gepirone vs. 117 in placebo) were in per-protocol population, and 199 subjects (97 in gepirone vs. 102 in placebo) completed the study and 49 discontinued. The overall discontinuation rate was slightly higher for the gepirone ER group; [3.2% vs. 2.4%, respectively, for lack of efficacy; 4% vs. 2.4% for AE or SAE; and 14.5% vs. 12.9% for other reasons (lost to follow-up, withdrawal of consent, noncompliance with treatment, and other unspecified reasons)]. In addition to the discontinuations for AEs or SAEs noted on the end of trial page, 3 additional subjects were included because the AE form indicated they were discontinued for AEs.

Table 3.3 Subjects Discontinuation by Primary Reasons for Withdrawal in Study FKGBE007

	Number (%, where applicable) of Subjects		
Disposition	Org 33062 ER	Placebo	Overall
Total number of subjects randomized	124	124	248
Total number of subjects treated	124	124	248
Total number of subjects completed	97	102	199
Total number of subjects who discontinued	27	22	49
Lack of efficacy	4 (3.2)	3 (2.4)	7 (2.8)
AE or SAE	5 (4.0)	3 (2.4)	8 (3.2)
Non compliance with treatment	1 (0.8)	0 (0.0)	1 (0.4))
Withdrew consent	5 (4.0)	2 (1.6)	7 (2.8)
Lost to Follow-up	11 (8.9)	12 (9.7)	23 (9.3)
Other	1 (0.8)	2 (1.6)	3 (1.2)

Source: Table 9 in the Clinical Study Report of Study FKGBE007.

3.1.4 Statistical Issues and Results

According to the protocol and SAP, the primary efficacy analyses were to be performed in the ITT population. Supportive efficacy analyses were to be performed on the PP population if the difference between the numbers of subjects in these two populations was more than 15%. Statistical analyses were also to be performed on OC and LOCF populations. No adjustments for multiple comparisons would be made.

The null hypothesis for primary analysis was that there was no difference between treatment groups in the change from baseline to endpoint of the HAMD-17 total score. The estimates of treatment effects and the corresponding 95% confidence intervals would be based upon the additive two-way ANOVA with factors treatment and center. The interaction between treatment and center would be tested. In case of a significant interaction ($p < 0.10$), the kind of interaction (i.e., the differences between the centers with respect to the treatment effects) would be discussed and further explored to evaluate whether it's still justified to present an overall estimate of the treatment effect.

The protocol for this study was approved by the sponsor and was finalized on July 17, 2003. According to the sponsor, the statistical analysis plan (SAP) was internally approved on September 29, 2004. The SAP was not submitted to FDA for review until 21 March 2006, upon the request of the agency.

Using the data sets provided by the sponsor, this reviewer derived the efficacy results which are almost the same as they derived. The ANCOVA analysis using the baseline HAMD-17 as the covariate gave similar significance results as shown in Table 3.4.

Given the high percentages of patient dropout as indicated in Table 3.3, the reliability and interpretability of the efficacy results using LOCF data set could be an issue. In general, LOCF procedure is reliable only when the mean outcome measure is stable over the whole study period. This does not seem to be the case as the mean HAMD-17 total score decreased 9 points on average from a mean baseline score of 24 to the time they left the study. Alternatively, the MMRM method may give more reliable efficacy results if the

patient dropouts were non-informative, with dropouts only depending on the observed outcome values, not on the unobserved values. This seems to be a reasonable assumption in this study.

Table 3.4: Treatment Effects on the Change from Baseline of Primary Efficacy Measures of HAMD-17 at the Endpoint in Study FKGBE007 --- ITT Population

	Placebo	Org 33062 ER
	(N=122)	(N=116)
Baseline Mean	23.9 (2.69)	24.2 (2.93)
Median change from baseline	-6.0	-10.0
ANOVA Analysis (LOCF)		
LS Mean change from baseline (SE) [a]	-7.8 (0.74)	-10.1 (0.76)
Difference between LS Means and C.I.[a]	-2.4 (-4.4, -0.3)	
P-value [a]	0.023	
ANOVA with Interaction (LOCF)		
LS Mean change from baseline (SE) [b]	-7.6 (0.75)	-10.3 (0.77)
Difference between LS Means and C.I.[b]	-2.7 (-4.8, -0.6)	
P-value [b]	0.012	
ANCOVA Analysis (LOCF)		
LS Mean change from baseline (SE) [c]	-7.8 (0.73)	-10.2 (0.75)
Difference between LS Means and C.I.[c]	-2.4 (-4.5, -0.4)	
P-value [c]	0.018	
MMRM Analysis		
LS Mean change from baseline (SE)[d]	-8.0 (0.77)	-10.4 (0.78)
Difference between LS Means and C.I.[d]	-2.3 (-4.5, -0.2)	
P-value [d]	0.033	
OC Analysis		
N	106	105
LS Mean change from baseline (SE) [e]	-8.1 (0.83)	-10.5 (0.83)
Difference between LS Means and C.I.[e]	-2.4 (-4.6, -0.2)	
P-value [e]	0.037	

a: Test for no difference between treatments at the endpoint using ANOVA model with treatment and center as factors.
b: Test for no difference between treatments at the endpoint using ANOVA model with treatment, center and treatment center interaction as factors.
c: Test for no difference between treatments at the endpoint using ANCOVA model with treatment and center as factors and the baseline total HAMD-17 score as covariate.
d: Test for no difference between treatments at the endpoint using MMRM model with treatment, center, visit and the interaction between treatment and visit as factors and baseline total HAMD-17 score as covariate. The MMRM model uses unstructured variance-covariance structure.
e: Test for no difference between treatments at the endpoint using ANCOVA model with treatment and center as factors and baseline total HAMD-17 score as covariate.
Note: Negative change in score indicates improvement.

Source: Reviewer.

Table 3.5: Summary of Mean HAMD-17 Total Score for Study FKGBE007 – LOCF ITT Population (Baseline - Weeks 8)

	Placebo (N=122)	Org 33062 ER (N=116)
Baseline		
Baseline Mean (SD)	23.9 (2.69)	24.2 (2.93)
Median change from baseline	-6.0	-10.0
Week 2		
N	120	112
LS mean change from baseline (SE) [b]	-4.2 (0.41)	-4.9 (0.43)
LS mean treatment effect and 95% CI	-0.7 (-1.9, -0.5)	
P-value [a]	0.27	
Week 3		
N	122	116
LS mean change from baseline (SE) [b]	-5.5 (0.51)	-6.7 (0.52)
LS mean treatment effect and 95% CI	-1.2 (-2.7, 0.2)	
P-value [a]	0.08	
Week 4		
N	122	116
LS mean change from baseline (SE) [b]	-6.4 (0.56)	-8.7 (-0.57)
LS mean treatment effect and 95% CI	-2.3 (-3.9, -0.8)	
P-value [a]	0.004	
Week 6		
N	122	116
LS mean change from baseline (SE) [b]	-7.4 (0.62)	-9.9 (0.64)
LS mean treatment effect and 95% CI	-2.5 (-4.2, -0.7)	
P-value [a]	0.006	
Week 8 / ET		
N	122	116
LS mean change from baseline (SE) [b]	-8.0 (0.73)	-10.2 (0.75)
LS mean treatment effect and 95% CI	-2.3 (-4.3, -0.2)	
P-value [a]	0.032	

a: p-value for treatment based on the reduced model without interaction (treatment and center in the model)
b: Negative differences in LS Means indicate positive effect of the active treatment over placebo.

Source: Table 15 in the Clinical Study Report of Study FKGBE007.

3.1.5 Other Short Term Efficacy Studies

The sponsor proposed to conduct some *post hoc* meta-analyses to combine all the short-term efficacy studies. In a meeting with the sponsor on May 31, 2006, the Agency indicated that our efficacy review would focus on individual study results rather than on ISE. But we were nevertheless in agreement with their plan for a new comprehensive summary of efficacy.

There were a total of 12 short term gepirone ER efficacy studies (134001, 134002, CN105-078, CN105-083, CN105-052, CN105-053, 134004, 134006, 134017, 134023, FK-GBE-007, FK-GBE-008). All 12 studies employed multicenter, randomized, double blind, parallel-group designs. In all 12 studies, the efficacy of gepirone ER was compared to that of placebo in adult outpatients who met either DSM-III-R criteria or DSM-IV criteria for MDD. In addition, in Studies 134004 and 134006, subjects were also required to have MDD with atypical features. All 12 studies had a short-term, double-blind treatment period during which subjects were treated for 6 to 8 weeks; three of the 12 studies had a long-term, double-blind extension period during which subjects were treated for an additional 20 to 44 weeks. In all the studies, a 3- to 14-day baseline observation period proceeded the short-term, double-blind treatment period; during this observation period, either placebo or no medication was administered. The meta-analyses were performed using the 12 studies of gepirone ER in the treatment of MDD that were considered relevant to the determination of gepirone ER efficacy, see Table 3.6 and 3.7.

Table 3.6 All the Controlled Gepirone ER Studies Included in the Meta-Analyses

Study Number	Sponsor Classification	Early Termination Y/N	Date submitted to NDA 21-164
134001	Positive		December 2003 Amendment
134002	Supportive		December 2003 Amendment
FKGBE007	Positive		Current amendment
FKGBE008	Supportive		Current Amendment
CN105-078	Supportive	Y	December 2003 Amendment
CN105-083	Supportive	Y	December 2003 Amendment
CN105-052	Failed	Y	December 2003 Amendment
CN105-053	Failed	Y	December 2003 Amendment
134004	Failed		December 2003 Amendment
134006	Failed		Current Amendment
134017	Failed		Current Amendment
134023	Negative		Current Amendment

Source: Table 3 in sponsor's Integrated Summary of Efficacy

Table 3.7 Designs and Characteristics of All the Controlled Short-term Gepirone ER Efficacy Studies

Study Number & No. of Centers	Study Design & Primary Endpoints	Starting Date Country & Completion Status	Treatment Group & Doses	Sample Size per Group	Gender (M/F) & Race (B/M/O)	Duration of Short-Term Studies
ORG 134001 5 centers	R, DB, PC, MC, PG HAMD-17	6/1999 USA Completed	Gepirone ER 20-80 mg/day Placebo	102 106	82M/126F 19/152/37	8 weeks
ORG 134002						

5 centers	R, DB, PC, MC, PG HAMD-17	6/1999 USA Completed	Gepirone ER 20-80 mg/day Placebo	110 108	83M/135F 19/191/8	8 weeks
FKGBE007 9 centers	R, DB, PC, MC, PG HAMD-17	10/8/2003 USA Completed	Gepirone ER 20-80 mg/day Placebo	116 122	74M/164F 55/154/29	8 weeks
FKGBE008 8 centers	R, DB, PC, MC, PG HAMD-17	10/20/2003 USA Completed	Gepirone ER 20-80 mg/day Placebo	99 100	71M/128F 17/162/20	8 weeks
CN105-078 2 centers	R, DB, PC, MC, PG 6-wk DB titration & 20-wk DB extension HAMD-17	12/18/1991 USA Terminated	Gepirone ER 20-100 mg/day Gepirone ER 10-50 mg/day Placebo	45 50 49	66M/77F 2/125/17	6 weeks
CN105-083 2 centers	R, DB, PC, MC, PG 8-wk DB titration & 44-wk DB extension HAMD-17	12/27/1991 USA Terminated	Gepirone ER 20-100 mg/day Gepirone ER 10-50 mg/day Placebo	39 37 41	51M/65F 7/88/22	6 weeks
CN105-052 2 centers	R, DB, PC, MC, PG 8-wk DB titration & 42-wk DB extension HAMD-17 & CGI	6/10/1991 USA Terminated	Gepirone ER 20-60 mg/day Fluoxetine 20 mg/day, Placebo	36 36 38	36M/72F 0/102/8	8 weeks
CN105-053 2 centers	R, DB, PC, MC, PG 8-wk DB Titration & 44-wk DB extension HAMD-17 & CGI	4/15/1991, USA Terminated	Gepirone ER 10-60 mg/day Imipramine 50-200 mg/day Placebo	58 54 56	73M/95F 1/154/13	8 weeks
ORG 134004 10 centers	R, DB, PC, MC, PG MDD with	6/2000 USA Completed	Gepirone ER 20-80 mg/day Fluoxetine 20-	135 138 136	140M/269F 33/339/37	8 weeks

	atypical features HAMD-25		40 mg/day Placebo			
ORG 134006 13 centers	R, DB, PC, MC, PG MDD with atypical features HAMD-25	12/2000 USA CANADA Completed	Gepirone ER 20-80 mg/day Paroxetine 10-40 mg/day Placebo	147 148 142	106M/331F 33/371/33	8 weeks
ORG 134017 10 centers	R, DB, PC, MC, PG MADRS	10/2002 USA Completed	Gepirone ER 20-80 mg/day Fluoxetine 20-40 mg/day Placebo	165 166 164	180M/315F 77/372/46	8 weeks
ORG 134023 12 centers	R, DB, PC, MC, PG HAMD-17	5/29/2003 USA Completed	Gepirone ER 20-80 mg/day Placebo	127 128	81M/173F 31/198/25	8 weeks

Abbreviations: B/W/O = black/white/other; DB = double-blind; ER = extended release; F = female; IR = immediate release; M = male; MC = multicenter; MD = multiple dose; No. = Number; PC = placebo-controlled; PG = parallel group; R = randomized; wk = week.

Source: Table 4 from Summary-2007 by the sponsor

The two positive studies (Studies 134001 and FK-GBE-007) gave statistically significant results in favor of gepirone ER (p=0.013 and 0.018, respectively) in reducing the HAMD-17 total score. The treatment differences were not statistically significant in the five so called "supportive studies" ($p \geq 0.195$), neither were in the five so called "lacking assay sensitivity" studies (134004, 134006, 134017, CN105-052 and CN105-053, $p \geq 0.167$) when analyzed on a per study basis.

In the five so called "lacking assay sensitivity" studies, the sponsor evaluated the efficacy of gepirone ER and the active comparator controls. By keeping only the subjects in the two groups for comparison in the statistical model and leaving out the subjects in the group which was not in the comparison, the sponsor's efficacy results suggested that in all of these studies, neither gepirone ER nor the active comparator statistically significantly improved the primary measure over placebo at the nominal significance level of 0.05.

In three of the five what the sponsor called "lacking assay sensitivity" studies, the primary endpoint was not HAMD-17. The primary endpoint was MADRS in Study 134017 and was HAMD-25 in Studies 134004 and 134006. Since the measurements of HAMD-25 was not provided in the data sets submitted to FDA, we couldn't confirm the sponsor's assertion of "lacking the assay sensitivity" through a statistical comparison between active control on the HAMD-25 total score in these two studies. Instead, we made the comparisons through the reduction of the HAMD-17 total score using ANOVA model with only the subjects in the two comparison groups as done by the sponsor. Similar comparison was made on MADRS total score for Study 134017. The results are provided in Table 3.8. In study 134004, active control Fluoxetine numerically improved placebo by 1.0 (p=0.23) and statistically significantly improved gepirone ER by 1.9 with a p-value of 0.02. In study 134006, Paroxetine statistically significantly improved both placebo and gepirone ER, by 1.73 (p=0.03) and 1.8 (p=0.014), respectively. In study 105-

053, active control Fluoxetine numerically improved gepirone ER by 1.4 (p=0.36) and almost statistically significantly improved placebo by 3.2 with a p-value of 0.051.

Table 3.8 Analysis of Change in Primary Endpoint in 5 "Lack Assay Sensitivity Studies" -- Gepirone ER (LOCF for ITT Population)

Study Number	Gepirone ER vs. Active Control	Active Control vs. Placebo
CN105-052		
LS Means (SE) difference in HAMD-17[a]	-0.11 (2.04)	-0.67 (2.00)
p-values	P=0.96[b]	P=0.74[c]
CN105-053		
LS Means (SE) difference in HAMD-17 [a]	1.4 (1.53)	-3.21 (1.62)
p-values	P=0.36[b]	0.051[c]
ORG 134004		
LS Means (SE) difference in HAMD-17 [a]	1.9 (0.79)	-1.0 (0.80)
p-values	P=0.02[b]	P=0.23[c]
ORG 134017		
LS Means (SE) difference in MADRS [a]	1.7 (1.11)	-1.1 (1.12)
p-values	P=0.13[b]	P=0.33[c]
ORG 134006		
LS Means (SE) difference in HAMD-17[a]	1.8 (0.74)	-1.73 (0.79)
p-values	P=0.014[b]	P=0.029[c]

a: Least squares means obtained using ANOVA model with terms for treatment and center, with non-comparison group deleted from the analyses.
b: p-value obtained using ANOVA model with terms for treatment and center, with placebo group deleted from the analyses.
c: p-value obtained using ANOVA model with terms for treatment and center, with active control group deleted from the analyses.
Source: Reviewer.

Alternatively as explorative analyses, the reviewer reanalyzed the efficacy data sets on the HAMD-17 total score by keeping all the subjects in the statistical model using the ANCOVA model with the baseline measure as covariate, center and treatment indicator as factors. The results are reported in Table 3.10. According to these results, perhaps only Study CN105-052 may be properly called "lack of assay sensitivity" at the nominal significance level of 0.05 based on the reduction of the HAMD-17 total score in which the active comparator arm failed to show superiority over placebo. In Studies CN105-053 and 134006, the active control statistically significantly improved placebo. In all of the five studies, the active control showed more improvement than placebo, numerically. The improvement of the active control over gepirone ER in Studies 134004, 134006 and 134017 appeared to be statistically significant. In these three studies, the placebo appeared to be numerically better than gepirone ER, even though they were not statistically significant. In these studies and Study CN105-053, the numerical improvement of the active control over gepirone ER was above 1.2.

Table 3.9 Analysis of Change in HAMD-17 from Baseline to End of the Short-Term Double-Blind Treatment Period in Controlled Studies of Gepirone ER

	Number of subjects (N)		(Adjusted) mean change		Treatment difference and 95% CI		
Study (1)	Gepirone ER	Placebo	Gepirone ER	Placebo	(Gepirone ER - Placebo)	SE	p-value
Pivotal Studies							
ORG 134001	101	101	-9.04	-6.57	-2.47 (-4.41, -0.53)	0.98	0.013
FKGBE007	116	122	-10.24	-7.79	-2.45 (-4.47, -0.43)	1.02	0.018
Supportive Studies							
ORG 134023 (2)	123	123	-7.93	-8.05	0.13 (-1.79, 2.04)	0.97	0.898
FKGBE008	96	99	-9.86	-8.48	-1.38 (-3.48, 0.71)	1.06	0.195
ORG 134002	102	103	-9.95	-9.24	-0.71 (-2.44, 1.02)	0.88	0.417
CN105-078	88	47	-7.42	-6.42	-1.00 (-3.17, 1.16)	1.10	0.362
CN105-083	73	39	-9.46	-8.97	-0.49 (-3.52, 2.53)	1.53	0.747
Studies Lacking AS							
ORG 134017	159	159	-10.36	-10.99	0.63 (-0.88, 2.13)	0.76	0.412
ORG 134004	124	130	-5.63	-6.66	1.03 (-0.55, 2.61)	0.80	0.199
CN105-052	35	37	-10.94	-10.28	-0.66 (-4.86, 3.55)	2.11	0.757
ORG 134006	140	143	-6.89	-7.13	0.24 (-1.18, 1.66)	0.72	0.742
CN105-053	56	56	-10.20	-8.15	-2.05 (-4.96, 0.87)	1.47	0.167
Meta-analyses (3)							
Supportive/Pivotal Studies Combined	699	634			-1.22 (-1.99, -0.45)	0.39	0.002
Treatment by study interaction							0.470
Supportive Studies Combined	482	411			-0.68 (-1.60, 0.24)	0.47	0.149
Treatment by study interaction							0.874
All 12 Studies Combined	1213	1159			-0.48 (-1.03, 0.08)	0.28	0.093
Treatment by study interaction							0.108

(1) Individual study statistics obtained using ANCOVA model with terms for treatment and center and baseline value (as a covariate), with active control group deleted from the analyses.
(2) Considered a negative study, but included with supportive studies for the purpose of meta-analysis.
(3) Combined estimates of the gepirone-placebo difference obtained as weighted averages of the gepirone-placebo differences with reciprocals of the squares of the standard errors of the by-study differences used as the weights. The standard errors of the overall estimates are the reciprocals of the square roots of the sums of the weights.

Source: Table 24 in sponsor's ISE

Table 3.10 Analysis of Change in HAMD-17 from Baseline to End of All the Short-Term Double-Blind Treatment Period -- Gepirone ER (LOCF for ITT Population)[a]

Study Number	Placebo	Gepirone ER	Active Control	Gepirone ER vs. Placebo	Gepirone ER vs. Active Control	Active Control vs. Placebo
ORG 134001						
N	101	101	NA		NA	NA
LS Means (SE)	-6.57	-9.04		-2.47 (0.98)		
p-values				p=0.013		
FKGBE007						
N	122	116	NA		NA	NA
LS Means (SE)	-7.79	-10.24		-2.45 (1.02)		
p-values				P=0.018		
ORG 134002						
N	103	102	NA		NA	NA
LS Means (SE)	-9.24	-9.95		-0.71 (0.88)		
p-values				P=0.42		
FKGBE008						
N	99	96	NA		NA	NA
LS Means (SE)	-8.48	-9.86		-1.38 (1.06)		
p-values				P=0.20		
CN105-078			NA		NA	NA
N	47	88		-1.0 (1.10)		
LS Means (SE)	-6.42	-7.42		P=0.36		
p-values						
CN105-083						
N	39	73	NA	-0.49 (1.53)	NA	NA
LS Means (SE)	-8.97	-9.46		P=0.75		
p-values						
ORG 134023						
N	123	123	NA		NA	NA
LS Means (SE)	-8.05	-7.93		0.13(0.97)		
p-values				P=0.90		
CN105-052			(Fluoxetine)			
N	37	35	36			
LS Means (SE)	-10.29	-10.98	-10.95	-0.69 (2.05)	0.02 (2.06)	-0.67 (2.03)
p-values				P=0.74	P=0.99	P=0.74
CN105-053			(Imipramine)			
N	56	56	54			
LS Means (SE)	-8.16	-10.16	-11.35	-2.0 (1.51)	1.2 (1.53)	-3.19 (1.52)
p-values				P=0.19	P=0.44	0.038
ORG 134004			(Fluoxetine)			
N	130	124	134			
LS Means (SE)	-6.79	-5.75	-7.46	1.04 (0.78)	1.71 (0.77)	-0.68 (0.76)
p-values				P=0.18	P=0.027	P=0.38
ORG 134017			(Fluoxetine)			
N	159	159	159			
LS Means (SE)	-11.02	-10.37	-11.92	0.65 (0.76)	1.54 (0.76)	-0.90 (0.76)
p-values				P=0.39	P=0.042	P=0.24

ORG 134006 [b]			(Paroxetine)			
N	143	144	136			
LS Means (SE)	-7.31	-7.09	-8.94	0.22 (0.72)	1.85 (0.73)	-1.63 (0.73)
p-values				P=0.76	P=0.012	P=0.026

a: Individual study statistics obtained using ANCOVA model with terms for treatment and center and baseline value (as a covariate), with active control group included in the analyses.
b: Similar results are obtained for Study ORG 134006 when centers 1 and 12 are pooled together.

Source: Reviewer

3.1.6 Further Support for Efficacy through Meta-Analysis

Three meta-analyses were conducted by the sponsor. These include the analysis to combine all 12 studies, the analysis to combine two positive studies (so called "pivotal studies" in Table 3.9) and five so called "supportive studies", and the analysis to combine only five so called "supportive studies", see Table 3.9. A treatment difference (gepirone ER vs. placebo) was considered by the sponsor to be statistically significant if the corresponding p-value was below 0.05. The first meta-analysis gave a p-value of 0.093, the second gave an estimated treatment effect of -1.22 in favor of gepirone ER and p-value of 0.002. The third gave a p-value of 0.149.

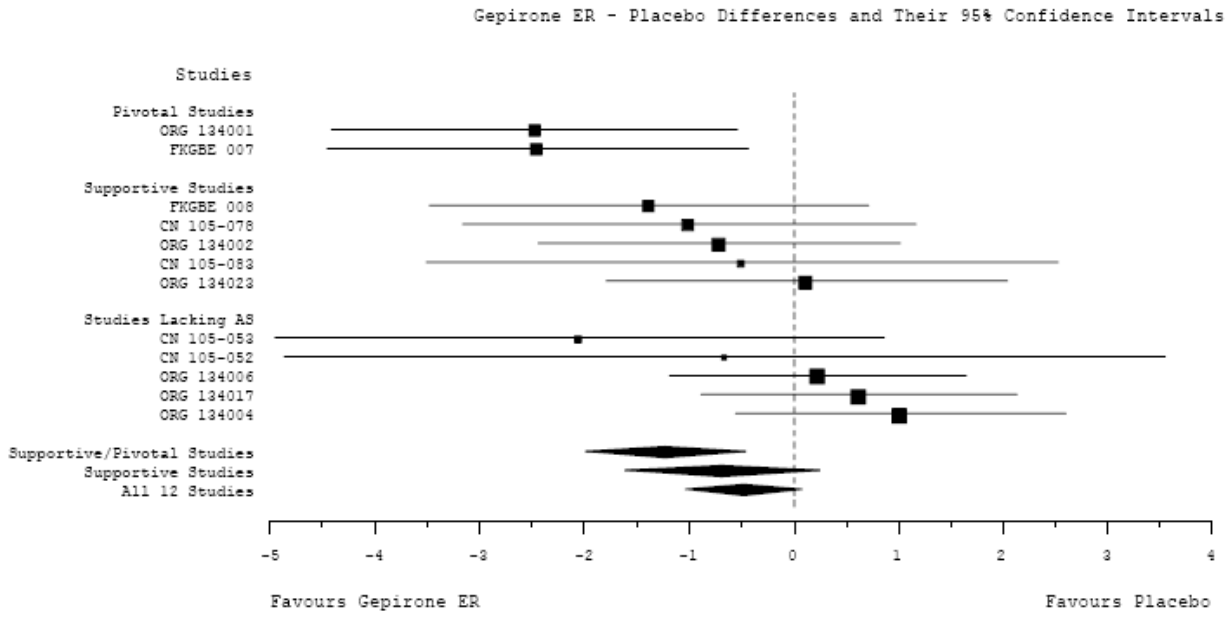
Using the data provided by the sponsor, this reviewer confirmed their efficacy results. At the same time, since the data for all the subjects are available in all studies, in particular the HAMD-17 total score at the last endpoint was available for all the subjects (with LOCF for the dropouts), the reviewer used the mixed effects model with PROC MIXED in SAS to explore the same combinations of the studies. By introducing a random effect to catch the differences among different studies, this approach makes use of all the subjects available and may give as reliable efficacy results as the sponsor's meta-analysis. This approach gives the following results. The combination of the two positive and five so called "supportive studies" gives an estimated treatment effect of -1.21 in favor of gepirone ER on the adjusted change in HAMD-17 total score (p=0.002). The combination of the five so called "supportive studies" gives a p-value of 0.20 for the treatment efficacy, see Table 3.11. The combination of all 12 studies gives a p-value of 0.05. A forest plot together with the meta-analysis results are given in Figure 3.1.

Table 3.11 Efficacy of the Different Combinations of Twelve Short-Term Efficacy Studies for Gepirone ER

Combination of Studies	Number of Subjects		Treatment Effect of Org 33062 ER			
	Placebo	Org 33062 ER	Meta-analysis		Mixed effects model	
			Efficacy (SE)	p-value	Efficacy (SE)	p-value
Five "supportive studies"	411	482	-0.68 (0.47)	0.15	-0.64 (0.48)	0.20
"Supportive/Pivotal studies"	634	699	-1.22 (0.39)	0.002	-1.23 (0.40)	0.002
All 12studies	1159	1213	-0.48 (0.28)	0.09	-0.57 (0.29)	0.05

Source: Reviewer.

Figure 3.1 Forest Plot for the Change in HAMD-17 from Baseline to End of Short-term Double Blind Treatment Period (LOCF)



Source: Figure 4 in sponsor's ISE

Among the above meta-analysis results, the sponsor suggested the use of the result of the combination of positive and so called "supportive studies" as the support of the effectiveness of gepirone ER in the treatment of MDD. This analysis gives a significant result in favor of gepirone ER over placebo with a p-value of 0.002. The reasons the sponsor removed the five so called "lack assay sensitivity" studies were: the active controls did not show statistically significantly superiority over placebo in these studies, none of the secondary endpoints was statistically significant, some studies (CN105-052, CN105-053) were terminated before reaching the planned sample size, and in some studies the HAMD-17 total score was not the primary endpoint (134004, 134006, and 134017).

These reasons don't seem to be convincing enough. As was pointed out above, the active control in all of the five studies outperformed placebo numerically. The superiority of the active control over gepirone ER in Studies 134004, 134006 and 134017 were statistically significant. In these three studies, the placebo seemed to outperform gepirone ER numerically even though they were not statistically significant. Secondly, although the HAMD-17 total score was not the designed primary endpoint in studies 134004, 134006, and 134017, it was collected and can be used for the efficacy analysis.

To examine if there is any additional evidence supporting the efficacy of gepirone ER in the remaining 10 studies (i.e., after excluding the two positive studies) in the treatment of MDD, this reviewer explored different ways to combine the ten non-positive studies using both the sponsor's meta-analyses approach and the mixed effects model approach. These included the combination of all of the ten negative short-term efficacy studies, all the negative short-term efficacy studies except Study CN105-052 which lacked assay sensitivity according to the results in Table 3.10. The results are depicted in Table 3.12.

Table 3.12 Efficacy of the Combinations of Ten Short-Term Efficacy Studies of Gepirone ER (Excluding the Two Positive Studies)

Combination of Studies	Number of Subjects		Efficacy of Org 33062 ER			
	Placebo	Org 33062 ER	Meta-analysis		Mixed effects model	
			Efficacy (SE)	p-value	Efficacy (SE)	p-value
Five "supportive studies"	411	482	-0.68 (0.47)	0.15	-0.64 (0.48)	0.18
Ten negative studies	936	996	-0.094 (0.31)	0.62	-0.14 (0.32)	0.67
Ten negative studies minus CN105-052	899	961	-0.081 (0.31)	0.60	-0.14 (0.32)	0.67

Source: Reviewer.

None of these analyses seems to give additional support for the effectiveness of gepirone ER over placebo in the treatment of MDD. So it seems that in a total of 12 short-term efficacy studies, only studies FKGBE007 and 134001 provide support for the efficacy of gepirone ER in the treatment of MDD. In addition, given that these are *post hoc* analyses, the multiplicity issue could be a concern if one only compares the p-values with the regular nominal significance level of 0.05. That means the adjusted p-value could be even larger than what they are observed here.

3.1.7 Study 28709

Study 28709 was submitted on December 23, 2003 as the response of the first NA letter of the same NDA. After reviewing the study we did not agree with the sponsor in the *post hoc* redefinition and removal of the 5 relapsers from their analysis on gepirone ER arm who had relapse based on discontinuation due to lack of efficacy as determined by Investigator. After adding these patients back, Gepirone ER did not statistically significantly reduce the rate of relapse over placebo in ITT population. The Agency came to the conclusion that this was not a positive study, see the NA letter on June 23, 2004.

Subsequent to the data analysis and reporting, sponsor reexamined the results of this study and found that 40 subjects randomized to double-blind treatment in the continuation period violated the randomization criteria at the end of the open-label period, by not achieving HAMD-17 total scores of 8 or less by Week 12. They removed these 40 patients and analyzed the subsequent PP population. PP population had 104 gepirone ER subjects and 106 placebo subjects, of which 25 (24.0%) gepirone ER subjects and 41 (38.7%) placebo subjects relapsed during the continuation phase. Without adjusting for pooled centers or country, gepirone ER statistically significantly reduced the relapse rate of depression compared to placebo (p=0.023). The adjusted analyses gave non-significant results in favor of gepirone ER (p=0.083 and p=0.059, respectively).

In Section 7.1.2 of the Protocol, however, the sponsor stated that: "All protocol violations will be determined by medical, clinical and biometrics personnel **prior to breaking the blind and will be done at least during 'blind review**'". According to the protocol, the *post hoc* reclassification of the patients to be protocol violators and removing them from the PP population is not justified for efficacy analyses therefore does not provide valid evidence supporting the efficacy of gepirone ER in reducing the relapse rate in this study.

3.1.8 Overall Statistical Evidence

In this submission, the primary efficacy results using LOCF data sets in Study FKGBE007 support the effectiveness of gepirone ER in the treatment of MDD in adult patients. The significance is also supported by the results using OC data and MMRM procedure. On the other hand, Study FKGBE008 does not seem to provide support for the efficacy of gepirone ER.

The fact that **only two out of a total of 12 short-term efficacy studies are positive in support of the effectiveness of gepirone ER during a period of 12 years raises concerns on the reproducibility of the treatment effect observed**. Procedures using meta-analysis or mixed-effects model methods on different combinations of the remaining 10 studies (i.e., after excluding the two positive studies) do not seem to provide further evidence supporting the effectiveness of gepirone ER over placebo. Furthermore, in three out of the five so called “lack of assay sensitivity” studies, active control significantly outperformed gepirone ER. Active control significantly outperformed the control in 2 out of 5 studies. In the positive study FKGBE007, the treatment effect seems to have been driven by Caucasians only. The collective evidence seems to provide only a weak support for the effectiveness of gepirone ER in the treatment of MDD among adults.

3.2 Evaluation of Safety

The review of safety is focused on the review of sexual dysfunction and patient suicidality. The studies that contain the sexual dysfunction questionnaires are 134001, 134002, 134004, 134006, 134017, 134501, 134502, 134503 and 134506. Sexual functioning was measured by scales such as DISF score, CSFQ score and was diagnosed using the criteria for sexual disorder by DSM-IV criteria in 8 clinical trials. The sponsor intends to seek a claim for lack of sexual dysfunction for gepirone ER in labeling.

3.2.1 Statistical Analyses proposed for Sexual Dysfunction Study

(b) (4)

5 Page(s) have been Withheld in Full as b4 (CCI/TS) immediately following this page

(b) (4)

In summary, although gepirone ER seems to be superior to its active comparators (fluoxetine or paroxetine) in some active controlled studies, such superiority has not been able to be reproduced therefore doesn't substantiate the claim that gepirone ER did not reduce the quality of patient's sexual functioning in these studies.

3.2.3 Suicidality Analyses

The ISS summarizes competed suicides and suicide attempts made by subjects participating in the clinical studies, and presents the analysis results of possibly suicide-related adverse events (PSRAEs) and the results of the change from baseline in the suicide questions from HAMD suicide item, and the MADRS item 10 at the 8-week time point.

In addition to 3 completed suicides, there were 13 suicide attempts (7 in gepirone ER, 2 in gepirone IR, 1 in imipramine, 2 in paroxetine and 1 in placebo). These were considered as being unrelated to the study drugs.

Possibly suicide-related AEs (PSRAEs) are summarized for gepirone ER and IR Phase II/III placebo-controlled double-blind randomized studies in depression with at least 20 subjects per treatment arm. Summaries are presented for all AEs (regardless of causality), by severity, relationship to study drug, and duration. Related events were those judged by the investigator as possibly, probably, or definitely related to study drug.

Two groupings based on the Columbia rating of the PSRAE, as assigned by the blinded reviewer, were also identified. A subject was considered to have suicidal behavior or ideation if the event was assigned a Columbia rating of 1-4, and a suicidal behavior event if it was assigned a rating of 1-3. Columbia ratings of 5-9 were deemed to be non-events. Risk ratios and 95% CIs are also presented. Relative risk was calculated in relation to subjects who received placebo and in relation to those receiving another active control. Risk ratios greater than 1.0 indicate a higher relative risk of an AE in subjects treated with the comparator versus placebo or with gepirone versus active control. Confidence intervals that exclude 1.0 suggest a statistically significant difference between the given treatment group versus placebo (or gepirone, in the case of the gepirone/active control data).

In the Phase II/III studies included in the PSRAE analysis, 2143 subjects received gepirone ER, 1527 subjects received gepirone IR, 2450 received placebo, 457 received fluoxetine, 142 received paroxetine, and 220 received imipramine.

Based on the grouping of Columbia rating categories, although examination of the risk ratios suggested that treatment with gepirone (ER, IR, or ER+IR combined group) did not statistically significantly increase the risk of suicidal behavior or ideation (categories 1-4) relative to placebo or to another active control, it seems to have increased the suicidal behavior (categories 1-3) and suicide attempt for gepirone ER compared to placebo. In fact, there were 5 patients with suicide behavior (who had a total of 8 suicide behaviors) in a total of 2143 gepirone ER patients compared none in a total of 2450 placebo patients. A Fisher's exact test gives a statistically significant p-value of 0.022, indicating a possible increase of

suicide behavior among gepirone ER patients. In the mean time, there were 4 patients with suicide attempt among a total of 2143 gepirone ER patients compared to none in a total of 2450 placebo patients. A Fisher's exact test gives a statistically significant p-value of 0.048, indicating a possible increase of suicide attempt among gepirone ER patients.

In the controlled Phase II/III studies included in the PSRAE analysis, 1723 subjects received gepirone ER, 1393 subjects received gepirone IR, 2292 received placebo, 457 received fluoxetine, 142 received paroxetine, and 220 received imipramine.

Based on the grouping of Columbia rating categories, it does not seem to have increased the suicidal behavior and suicide attempt for gepirone ER compared to placebo. In fact, there were 3 patients with suicide behavior (who had a total of 3 suicide behaviors) and 2 suicide attempts in a total of 1723 gepirone ER patients compared none in a total of 2292 placebo patients. They do not seem to give statistically significant results. The Fisher's exact test gives p-values of 0.08 for the differences of suicide behavior between gepirone ER and placebo and a p-value of 0.12 for suicide attempts.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

Using LOCF data, subgroup analyses were performed for the pivotal Study FKGBE007 on age, gender and race (Caucasian versus non-Caucasian). All subgroup analyses were considered exploratory. The treatment-by-subgroup interaction was tested using an ANCOVA model including the terms for baseline, treatment, center, subgroup, and the treatment-by-subgroup interaction. The treatment-by-subgroup interaction was tested to find out whether treatment effects in the primary efficacy measure were similar for each subgroup.

In the study, neither sex nor the interaction between sex and treatment group was statistically significant at the nominal significance level of 0.05 in the ANCOVA analysis. When sex and the interaction between sex and treatment group were adjusted, the treatment effects becomes -2.2 and its significance levels becomes 0.054, indicating possible difference of treatment efficacy between male and female subjects. Table 4.1 suggests that the improvement on the primary endpoint was mainly driven by female rather than male patients.

The original race has four groups: Caucasian (154), Black (55), Asian (2) and Others (27). Due to the low frequency of non-Caucasian groups, we combine them together in the subgroup analysis to form two groups: Caucasian (154) and non-Caucasian (84). In the ANCOVA analysis, race is not statistically significant at the nominal significance level of 0.05. But the interaction between race and treatment gives a p-value of 0.06, below the nominal significance level of 0.10 for interactions. When race and the interaction between race and treatment group are adjusted, the treatment effects becomes -1.8 and its significance level becomes 0.11. Furthermore, when the ANCOVA analysis is performed on non-Caucasians, the treatment effect is 0.06 and the corresponding p-value is 0.97. While the same analysis is performed on Caucasians, the treatment effect is -3.9 and the corresponding p-value is 0.002, which is highly statistically significant at the nominal significance level of 0.05. This suggests that the treatment effect was mainly driven by Caucasians. These results are given in Table 4.1.

Age is a continuous variable and it gives a p-value of 0.55 in the ANCOVA model which is not statistically significant at the nominal significance level of 0.05. It was not separated into different subgroups for further statistical analysis.

Table 4.1 Treatment Effect by Sex and Age Groups on the effect size in Study FK-GBE-007 (LOCF Analysis)

	Placebo	Org 33062 ER	Difference (p-value*)
Sex Effect			
Male	N=37	N=37	
Change from Baseline of HAMD-17 Total Score (SE)	-7.3 (1.22)	-7.7 (1.26)	-0.4 (0.32)
Female	N=85	N=79	
Change from Baseline of HAMD-17 Total Score (SE)	-8.2 (0.87)	-11.0 (0.96)	-2.8 (0.04)
Race Effect			
Caucasian	N=81	N=73	
Change from Baseline of HAMD-17 Total Score (SE)	-7.4 (0.84)	-11.3 (0.98)	-3.9 (0.002)
Non-Caucasian	N=41	N=43	
Change from Baseline of HAMD-17 Total Score (SE)	-8.9 (1.36)	-8.6 (1.21)	0.3 (0.97)

*: For each subgroup, the nominal p-value is derived using the ANCOVA model with baseline HAMD-17 as covariate, center and treatment as factors.

Source: FDA analysis.

4.2 Other Special/Subgroup Populations

Not available.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

Two pivotal short-term gepirone ER studies, FKGBE007 and FKGBE008 were submitted. FKGBE007 is positive and FKGBE007 is not. In these studies, the primary efficacy measure was the change from baseline to the end of study in the HAMD-17 total score. The treatment efficacy was analyzed using ANCOVA with LOCF data. The sponsor also conducted *post hoc* meta-analyses on the previously conducted studies along with two current studies FKGBE007 and FKGBE008 and reevaluated the relapse prevention Study 28709. Among the three meta-analyses performed, the sponsor suggested to adopt the positive one which combineed two positive studies and five so called “supportive” studies.

In Study FKGBE007, the effectiveness of gepirone ER in the treatment of adult patients with MDD is supported by the primary efficacy analysis using LOCF, and the analyses using OC and MMRM. Further *post hoc* subgroup analyses suggest that the treatment effect was mainly driven by Caucasians, and female patients.

The fact that **only two out of a total of 12 short-term efficacy studies are positive in support of the effectiveness of gepirone ER during a period of 12 years raises concerns on the reproducibility of the treatment effect observed**. Procedures using meta-analysis or mixed-effects model methods on different combinations of the remaining 10 studies (i.e., after excluding the two positive studies) do not seem to provide further evidence supporting the effectiveness of gepirone ER over placebo. Furthermore, in three out of the five so called “lack of assay sensitivity” studies, active control significantly outperformed gepirone ER. Active control significantly outperformed the control in 2 out of these 5

studies. In the positive study FKGBE007, the treatment effect seems to have been driven by Caucasians and females only. The collective evidence seems to provide only a weak support for the effectiveness of gepirone ER in the treatment of MDD among adults.

In addition, the sponsor also intends to seek a claim for lack of sexual dysfunction for gepirone ER in labeling.

When comparing patient sexual functioning, although gepirone ER seems to be superior to its active comparators (fluoxetine or paroxetine) in some active controlled studies, there does not seem to be enough consistent evidence in supporting the claim that gepirone ER did not reduce the quality of patient's sexual functioning (b) (4)

5.2 Conclusions and Recommendations

In this submission, two pivotal short-term gepirone ER studies, FKGBE007 and FKGBE008 were conducted in the United States. In these studies, the primary objective was to evaluate the therapeutic efficacy of gepirone ER tablets in comparison with placebo at the end of an 8-week treatment period in subjects with MDD. To have an integrated summary of efficacy data, the sponsor also conducted *post hoc* meta-analyses on the previously conducted efficacy studies along with two current studies FKGBE007 and FKGBE008 and reevaluated the relapse prevention Study 28709. In addition, the sponsor also intended to seek a claim for lack of sexual dysfunction for gepirone ER in labeling.

In Study FKGBE007, the effectiveness of gepirone ER in the treatment of adult patients with MDD is supported by the primary efficacy analysis using LOCF, and the analyses using OC and MMRM. Further *post hoc* subgroup analyses suggest that the treatment effect was mainly driven by Caucasians, and female patients. In addition to Study FKGBE007 there is one more positive Study 134001among a total of 10 previously conducted short-term gepirone ER studies.

The fact that **only two out of a total of 12 short-term efficacy studies are positive in support of the effectiveness of gepirone ER during a period of 12 years raises concerns on the reproducibility of the treatment effect observed**. Procedures using meta-analysis or mixed-effects model methods on different combinations of the remaining 10 studies (i.e., after excluding the two positive studies) do not seem to provide further evidence supporting the effectiveness of gepirone ER over placebo. Furthermore, in three out of the five so called "lack of assay sensitivity" studies, active control significantly outperformed gepirone ER. Active control significantly outperformed the control in 2 out of 5 studies. In the positive study FKGBE007, the treatment effect seems to have been driven by Caucasians and female patients only. The reevaluation of the relapse prevention Study 28709 does not provide valid evidence supporting the efficacy of the treatment. The collective evidence seems to provide only a weak support for the effectiveness of gepirone ER in the treatment of MDD among adults.

When comparing patient sexual functioning, although gepirone ER seems to be superior to its active comparators (fluoxetine or paroxetine) in some active controlled studies, there does not seem to be enough consistent evidence in supporting the claim that gepirone ER did not reduce the quality of patient's sexual functioning (b) (4)

C:\Data\My Documents\NDA Review\NDA 2007\21164

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

/s/

Fanhui Kong
10/9/2007 05:19:26 PM
BIOMETRICS

Peiling Yang
10/9/2007 05:25:22 PM
BIOMETRICS

Kooros Mahjoob
10/9/2007 06:13:26 PM
BIOMETRICS

James Hung
10/10/2007 08:18:34 AM
BIOMETRICS

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	**21-164/N000**
Drug Name:	**Gepirone HCl (Org 33062) Extended Release 40, 60, and 80 mg Tablets**
Indication(s):	**Antidepressant**
Applicant:	**Organon, Inc.**
Date(s):	**12/23/03**
Review Priority:	**Priority**
Biometrics Division:	**Division of Biometrics 1, HFD-710**
Statistical Reviewer:	**Roswitha Kelly, MS**
Concurring Reviewers:	**Kun Jin, Ph.D., Team Leader**
	Jim Hung, Ph.D., Acting Deputy Director
Medical Division:	**Division of Neuropharmacological Drug Products, HFD-120**
Clinical Team:	**Earl Hearst, M.D.**
Project Manager:	**Paul David**

Keywords: Relapse, Cochran-Mantel-Haenszel Test, Log-Rank Test.

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

1. EXECUTIVE SUMMARY[1]

1.1 Conclusions and Recommendations

In this reviewer's opinion, Study 28709-2003, a relapse trial, which the sponsor had identified as fulfilling the outstanding need for one more well-controlled positive trial in Gepirone ER, did not reach its goal. The sponsor's primary statistical analysis did not use all ITT patients. When using all ITT patients, the comparison of relapse rates between Gepirone-treated patients and placebo-treated patients did not reach statistical significance. This finding held for the Cochran-Mantel-Haenszel test when adjusting for Center or for Country.

Five patients may have relapsed but were not treated as such in the sponsor's primary analysis. These patients all received Gepirone and their reclassification reduces treatment differences to statistical non-significance.

One important secondary endpoint is time to first relapse. The log-rank test did not reach statistical significance for the original data, when stratified by center or by country, nor when the five patients in question were considered to have relapsed.

In discussion with the Medical Officer, Dr. E. Hearst (HFD-120), it was decided that no subgroup analyses were necessary.

The sponsor had identified trial 28709-2003 as the only pivotal study. Therefore, none of the other studies were statistically evaluated by agreement with the Medical Division.

This review does not address safety.

The findings were discussed with the Medical Division (HFD-120).

1.2 Brief Overview of the Clinical Studies

In the previous submission (05/18/01), the sponsor had failed to show unequivocally that the ER formulation of Gepirone HCl was statistically superior to placebo (cf. Statistical Review and Evaluation of Gepirone, March 2002). In a Not Approvable Action Letter and subsequent communications the sponsor was told that one more successful, robustly positive, placebo-controlled trial in Gepirone ER in the MDD population would satisfy the concerns regarding Gepirone's efficacy. The 12/23/03 submission is intended to address all issues stated in the Not Approvable Action Letter.

[1] The reviewer would like to acknowledge and express her appreciation for the help in data manipulation and SAS coding received from Dr. Ohid Siddiqui.

The sponsor identified Study 28709-2003 from the 12/23/03 submission as the pivotal trial satisfying the outstanding efficacy requirements. This is the only study being addressed in this statistical review and evaluation.

1.3 Statistical Issues and Findings

1.3.1 Sponsor's Results and Conclusions

The primary analysis of the primary endpoint was the comparison of relapse rates of depression during the continuation phase by the Cochran-Mantel-Haenszel test adjusting for centers. For subjects treated with Gepirone the relapse rate was 23.0% at endpoint compared to 34.7% for placebo-treated patients. This difference was statistically significant (p=0.024).

A supportive analysis of time to first relapse did not significantly distinguish between the two treatments (p=0.065).

Additional five Gepirone patients may have relapsed. The Cochran-Mantel-Haenszel test was no longer statistically significant when these patients were defined as relapses (p=0.101).

1.3.2 Reviewer's Results and Conclusions

The sponsor's statistical methodology was not acceptable without modification. The reviewer's statistical approach did not show statistically significant differences between the Gepirone treated and placebo treated patients in Study 28709-2003 with respect to proportion of relapse ($p>0.080$) or time to first relapse ($p\geq0.089$). When reclassifying five Gepirone patients identified by the sponsor as having relapsed, the treatment difference for each statistical test is even smaller and less significant.

The main concerns with the trial and the sponsor's primary statistical analysis are:

- The Cochran-Mantel-Haenszel (CMH) test as performed by the sponsor, i.e. without appropriate pooling of centers, excluded centers that had either only one treatment arm or that had no relapses. Consequently, 32 ITT patients were not part of the sponsor's primary analysis. These patients needed to be grouped into one center to become part of the analysis.

- The sponsor did not specify how small centers should be pooled. Therefore, the reviewer used the approach exercised in HFD-120 and pooled centers with less than 5 patients. Grouping patients from centers with either only one treatment arm or no relapses or having at most 4 patients into one fictitious Center and performing the CMH test uses the results of all ITT patients. This treatment comparison of relapse rates adjusted for Center is not statistically significant (CMH, p=0.0971).

- Grouping the original centers into countries also avoids the loss of any information. The treatment comparison of relapse rates adjusted for Country is also not statistically significant (CMH, p=0.0805).

- The use of the CMH test in the presence of censoring may be inappropriate. Censored patients are implicitly classified as successes. The log-rank test for time to first relapse uses the information of all ITT patients and treats censored patients as such. The treatment comparison of the log-rank test did not reach statistical significance (p=0.0891). Stratifying by center or by country further reduced the treatment difference in time to first relapse.

As noted by the sponsor, five patients, who all received Gepirone, may have relapsed (based on information recorded under the investigator's discontinuation variable 'Reasons not mentioned above, please specify _____'), but were not classified as such in the primary analysis. Classifying these patients as relapsed renders the treatment comparisons completely non-significant (CMH adjusted for Center, p=0.3302; CMH adjusted for Country, p=0.3145; log-rank, $p \geq 0.2782$).

1.3.3 Extent of Evidence in Support of Efficacy Claim

Study 28709-2003 is the only study reviewed here because the sponsor had identified it as the only trial that meets the outstanding efficacy requirements. In the reviewer's opinion, it did not show statistical superiority of Gepirone ER over placebo in relapse rates when appropriate statistical methods were applied. In addition, the log-rank test for time to first relapse also did not reach statistical significance, whether stratified by center or country or not. Furthermore, additional five Gepirone patients appear to have relapsed and their reclassification further reduced any treatment differences.

1.3.4 Statistical Issues Which May Impair the Efficacy Conclusion

The sponsor's use of the Cochran-Mantel-Haenszel statistic adjusting for each individual center was not appropriate, because it excluded 32 ITT patients. These patients came from centers, which had either only one treatment arm or where no patients had relapsed and such centers do not contribute to the CMH statistic. Furthermore, no provision was made to pool small centers in general or to discuss the CMH test in the presence of censoring. The CMH statistic treats censored patients as successes.

Using all ITT patients in the analysis resulted in non-significant findings.

2. INTRODUCTION

2.1 Overview

The original NDA submission for Gepirone Extended Release (ER) was accepted May 18, 2001. After review, the information presented was found inadequate and a 'Not Approvable' Action Letter was issued 03/15/02. Subsequent communications between the sponsor and the medical division stated that one 'robustly positive', adequate and well-controlled trial in ORG 33062 ER could make the NDA approvable. Study 134004 was identified by the sponsor to satisfy this requirement, but subsequently failed to show a statistically significant difference between Gepirone ER and placebo. Organon requested that the long-term relapse study (Protocol 28709) be used in lieu of the necessary short-term well-controlled trial. The Agency noted that this study would be accepted but could not guarantee that it would be sufficient to support product registration given the preponderance of negative trials.

This statistical review focuses on the efficacy results of the Phase III study #28709-2003. This is the long-term relapse trial in outpatients with major depressive disorder (MDD) treated with Gepirone ER tablets and is identified by the sponsor as positive and as satisfying the outstanding requirement for efficacy approval. Its data and reports are part of the 12/23/03 submission.

2.2 Data Sources

The data sets were submitted as xpt files according to the Guideline for electronic submissions. The primary efficacy parameter for study 28709-2003 was relapse during the double-blind period, defined as a HAMD-17 score of 16 or greater or as a notation by the investigator that relapse criteria were fulfilled. The HAMD-17 results resided in the sponsor's HAMDEPR1-4.xpt files and the investigator's decision was the variable DCRELAPX (discontinued due to relapse) in the EOT.xpt file. However, it was not transparent which HAMD-17 score was the last one for each patient and the reviewer had difficulty reproducing the sponsor's relapse rates. She therefore asked for a new dataset, which contained for each patient the final HAMD-17 result, the investigator's decision, an efficacy endpoint (relapse Y/N) created by the sponsor for the ITT population and an explanation how these variables were derived from the original datasets. Upon receipt of the new data set and the description of the variable derivations from the originally submitted data files, this reviewer could independently create the ITT patient data set for the double blind period. The reviewer's analysis of time to first relapse uses a different time variable as seen in the sponsor's analysis. The sponsor's 'Time_R' variable was not found in the data sets. The reviewer used 'Daydrend', i.e. the day treatment was stopped.

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Overview of the Clinical Program and Studies Reviewed

Gepirone HCl (Org 33062) is a novel azapirone derivative that has both antidepressant and antianxiety activity. It was originally developed as an immediate release formulation by Bristol-Meyer-Squibb. Organon Inc. acquired the drug and developed the extended release formulation. The original 05/18/01 submission for Gepirone ER contained the results of 24 trials (18 randomized, placebo-controlled trials with or without active control; 6 uncontrolled trials) in both Gepirone IR and Gepirone ER (8 controlled ER trials; 1 uncontrolled ER trial). The sponsor received a Not Approvable letter March 15, 2002.

The 12/23/03 submission is in response to the Not Approvable Action letter and subsequent communications between the sponsor and the Agency. It contains the results of all old and new trials conducted with Gepirone IR and ER. Most of these (24) were addressed in the previous review. It appears that seven studies were conducted or completed after the initial submission and another 11 studies are ongoing. Of these the sponsor first identified Study 134004-2003 as showing robustly positive findings against placebo. However, this study was not successful and the sponsor presented Study 28709-2003 as the only pivotal trial, which will satisfy the outstanding efficacy deficiencies.

3.1.2 Description of Pivotal Study # 28709-2003

3.1.2.1 Trial Design and Patient Population

Study 28709-2003 is a relapse trial conducted exclusively in Europe, which lasted up to one year. Patients with a qualifying diagnosis of Major Depressive Disorder (MDD) were treated with open-label Gepirone ER 40-80mg/day for up to 12 weeks. Patients whose HAMD-17 score fell below 9 at week 8 or at week 12 were classified as responders and randomized equally to Gepirone ER (at the same dose) or to placebo for a double-blind continuation phase that lasted for up to the remainder of one year (Table 1).

Table 2 shows the time line of this relapse study. There were three phases to the study. The single-blind placebo washout period served to withdraw subjects from unacceptable drugs and could range from 3-14 days. During the open-label phase all subjects were to receive Gepirone ER according to a preset dosing schedule. The purpose of the OL phase was to elicit a clinical response to the patients' Major Depressive Disorder. When the response criteria for remission (HAMD-17 Total $\leq$ 8) was met at either week 8 or week 12, the subject could be randomized and enter the continuation phase. At entrance to this double blind treatment phase, the subjects were randomized equally to continuation treatment with Gepirone or to placebo.

Recruitment was to go on until 200 subjects belonged to the ITT group, i.e. until 200 subjects had taken at least one dose of double-blind trial medication and had had at least one efficacy assessment within the continuation phase at which the primary parameter could be determined. Relapse was to be evaluated at every visit during the continuation phase.

Entrance criteria required that subjects were outpatients (but could be hospitalized if needed) who presented a primary diagnosis of recurrent MDD (DSM-IV 296.3). Age was limited from 18 - 70 years and a screening and baseline HAMD-17 total score of $\geq$ 20 were necessary for entry. Females had to be postmenopausal for at least one year, surgically sterile or non-pregnant using acceptable methods of birth control. Written informed consent was also required. Exclusion criteria were spelled out for not entering the trial and for not being accepted into the continuation phase.

Table 1: Design Characteristics for Study 28709-2003

Study	Design	Patient Population	Treatment Dose (mg/day)	No. of Patients Randomized	Trial Duration Week	Titration Schedule	Titration Period
28709-2003 (Organon)	R, DB, MC, PG, PC	Adult OP with MDD Met: DSM-IV 296 3	Gepirone **ER** 40- 80 Placebo	126 124	52-wks	Starting: 1 40-mg tablet in the morning Increase:1 20-mg tablet each week up to Max of 80 mg/day	OL

R: Randomized; DB: double-blind; SC/MC: Single/Multi-center; PC: Placebo Controlled; PG: Parallel-group; OP: Outpatient; MDD: Major Depressive Disorder; DMS- IV: Diagnostic and Statistical Manual of Mental Disorders, 4th ed.; ER = Extended release; OL: Open Label.

Table 2: Time Line for Study 28709-2003

	Screen		Open-label Phase					Double-blind Continuation Phase										FU
Assessment Week→	S	BL	1	2	4	8	12	16	20	24	28	32	36	40	44	48	52	54

S = Screening; BL = Baseline; FU = Follow up

This study was conducted exclusively in Europe, namely in Germany, France, Poland, Finland and Turkey. The sponsor reported 9 centers in France, giving a total of 29 centers. The data file contained only 8 centers in France. The sample sizes per center ranged from one to 27 patients.

3.1.2.2 Efficacy Parameters

The primary endpoint was relapse defined as either an HAMD-17 score of 16 or greater or a decision by the investigator that relapse criteria were met. The primary analysis was

a Cochran-Mantel-Haenszel center-adjusted treatment comparison of the proportions of relapse in the ITT population at study end. There were several secondary endpoints and analyses.

3.1.2.3 Demographics

Table 3 is a reproduction of the sponsor's Table 15 giving the demographic distribution of the patients in Study 28709-2003. This reviewer performed no sub-group analyses for demographic factors as suggested by the reviewing medical officer, Dr. Earl Hearst.

Table 3: All Subjects Treated Demographics of Study 28709-2003

	Open-label phase (AST-OL)	Double-blind continuation phase (AST)		
	Org 33062 (N=420)	Org 33062 (N=126)	Placebo (N=124)	Overall (N=250)
Gender				
Male	130 (31.0%)	33 (26.2%)	48 (38.7%)	81 (32.4%)
Female	290 (69.0%)	93 (73.8%)	76 (61.3%)	169 (67.6%)
Race				
Caucasian	415 (98.8%)	124 (98.4%)	124 (100.0%)	248 (99.2%)
Black	2 (0.5%)	1 (0.8%)	0 (0.0%)	1 (0.4%)
Other	3 (0.7%)	1 (0.8%)	0 (0.0%)	1 (0.4%)
Age (years)				
N	419	125	124	249
Mean (SD)	44.3 (10.5)	43.7 (10.9)	45.8 (9.6)	44.7 (10.3)
Median (range)	45 (17-72)	45 (17-70)	46 (19-67)	45 (17-70)
Body height (cm)				
N	420	126	124	250
Mean (SD)	166.8 (9.0)	166.9 (8.2)	167.6 (9.6)	167.3 (8.9)
Median (range)	165 (140-199)	166 (150-192)	166 (145-194)	166 (145-194)
Body weight (kg)				
N	420	126	124	250
Mean (SD)	71.45 (15.19)	70.27 (14.77)	72.96 (15.70)	71.60 (15.27)
Median (range)	70.0 (38.0-133.0)	68.0 (45.0-117.0)	73.0 (38.0-133.0)	70.0 (38.0-133.0)
Body Mass Index (kg/m^2)				
N	420	126	124	250
Mean (SD)	25.59 (4.61)	25.14 (4.48)	25.88 (4.84)	25.51 (4.67)
Median (range)	24.9 (15.6-40.6)	24.6 (17.3-39.5)	25.0 (16.2-39.3)	24.8 (16.2-39.5)

3.1.3 Sponsor's Analysis, Results and Conclusions

3.1.3.1 Statistical Methodologies

The primary efficacy parameter was defined as the number of subjects with a relapse. Relapse was obtained during the double blind continuation period when a subject either reached an HAMD-17 total score of 16 or greater or the subject was discontinued due to lack of efficacy as indicated by the investigator checking "Relapse Criteria Fulfilled" on the EOT CRF. The primary time-point for treatment comparisons was the endpoint assessment of the continuation phase based on the ITT population. The primary statistical analysis compared the proportions with relapse in the ITT populations of placebo- and Gepirone-treated subjects using LOCF. The statistical method for comparison was a Cochran-Mantel-Haenszel test adjusting for centers.

3.1.3.2 Analysis and Findings

A total of 435 patients were screened, of whom 428 selected to participate in the open label (OL) phase. Of the 420 patients receiving open label Gepirone ER, 250 (59.52%) were in remission at the end of the OL phase. They were randomized 1:1 to Gepirone at the final titrated dose (n=126) or to placebo (n=124). The sponsor reported 55 Gepirone and 54 placebo patients discontinuing prematurely. The reasons are given in Table 4, which is a reproduction of the sponsor's Table 11.

Table 4: Reasons for Patient Discontinuation

Reason for discontinuation (according to the EOT CRF)	Open-label phase (AST-OL)		Double-blind continuation phase (AST)					
	Org 33062 (N=420)		Org 33062 (N=126)		Placebo (N=124)		Overall (N=250)	
	n	%	n	%	n	%	n	%
Adverse event / Serious adverse event	46	11.0	4	3.2	5	4.0	9	3.6
Not willing or cannot cooperate for reasons not related to the study	26	6.2	11	8.7	7	5.6	18	7.2
Insufficient therapeutic effects which makes continuation of the trial conditions unjustifiable	26	6.2	--[a]	--	--	--	--	--
Relapse criteria fulfilled	--	--	26	20.6	35	28.2	61	24.4
Reason not mentioned above	19	4.5	14	11.1	7	5.6	21	8.4
Total	117	27.9	55	43.7	54	43.5	109	43.6

[a] not applicable;

The primary efficacy endpoint analysis was the pre-specified comparison of relapse rates at the end of the double blind phase using the ITT population. The statistic was the Cochran-Mantel-Haenszel test with a two-sided p-value. Relapse was defined as an

HAMD-17 score of $\geq$ 16 or the investigator's decision that relapse criteria were fulfilled. Several secondary efficacy parameters and analyses were also specified.

The end-of trial relapse rates in the ITT population using LOCF were 29/126 (23.0%) for Gepirone and 43/124 (34.7%) for placebo. The Cochran-Mantel-Haenszel test of the relapse rates adjusted for center produced a p-value of 0.024, which was considered statistically significant. This analysis had been specified in the protocol and no deviations from the protocol seem to be of concern.

There were five patients, all receiving Gepirone ER, who may be considered having relapsed based on information supplied by the investigator in the item "Discontinued for Reasons not mentioned above, please specify_____" . Considering these five patients as relapsed, rendered the p-value for the CMH test non-significant.

A secondary objective was to compare the time to relapse during the continuation phase between subjects receiving Gepirone at the final titrated dose and those receiving placebo. Time to first relapse was evaluated by a Kaplan-Meier approach. Three subjects in the placebo group continued the trial after having had a relapse. Because they did not have a relapse at endpoint, they were not counted as such in the analysis of relapse rates at endpoint. However, in the survival analysis their time to first relapse was included. The log-rank test for comparing time to relapse of the two treatment groups did not reach statistical significance (p=0.065). The lack of a significant difference in the time to first relapse analysis is attributed to nine patients on Gepirone who relapsed early during the continuation phase. Such worsening was considered an artifact because the majority of patients did not show such a pattern

3.1.3.3 Conclusion

The sponsor concluded that the results of this trial demonstrated that Gepirone ER is effective in preventing relapse, as defined in this study, in subjects who achieved remission during 8-12 weeks of open-label Gepirone ER therapy. Findings across time points and most secondary parameters indicate better maintenance of effect for patients continuing with Gepirone than for patients who were randomized to placebo. The lack of a significant difference in time to relapse was attributed to an artifact of nine Gepirone patients relapsing early in the continuation phase. The sponsor stated that none of the parameters showed a benefit for placebo over Gepirone at any time point in any analysis.

In the study report (28709-2003.pdf) the sponsor does not address whether this trial meets the outstanding efficacy deficiency. The ISE of the 12/23/03 submission seems to address only studies of the original submission. However, the sponsor had proposed to use Study 28709-2003 in lieu of a short-term well-controlled trial as noted in the 07/14/03 Meeting Minutes.

3.1.4 Reviewer's Analysis, Results and Conclusions

3.1.4.1 Statistical Methodologies

This reviewer considers the sponsor's primary analysis as not appropriate because, as specified, it did not use all ITT patients. A feature of the CMH statistic is that strata, which have two or more of the four cells unpopulated, do not contribute to the test statistic. Therefore, the findings from centers, which either enrolled patients to only one treatment arm or where no relapses occurred, did not contribute to the statistic and consequently not to the p-value. If low enrollment per center can be anticipated, this feature of the CMH statistic should have been recognized. As seen in Table 5, there were 10 centers with a total of 32 ITT patients that were not included in the sponsor's CMH test statistic. None of the 12 placebo patients had relapsed but 5 of the 20 (25%) Gepirone ER patients had relapsed. Furthermore, the sponsor did not specify any method for pooling centers with low enrollment. In the absence of a pre-specified method for pooling centers, HFD-120's practice is to combine centers with less than five patients. Combining centers with at least two empty cells with centers, which have four or less patients, into one fictitious center or grouping centers into their respective countries, will use all ITT patients in the CMH test.

Table 5: Centers that did not Contribute to CMH Statistic

Center	Number of Patients (Relapse Rate)		Reason
	Gepirone	Placebo	
D_101		1 (0/1)	Only one treatment arm
PL_068	1 (1/1)		Only one treatment arm
SF_016	1 (1/1)		Only one treatment arm
TR_006	1 (0/1)		Only one treatment arm
F_258	2 (1/2)		Only one treatment arm
TR_005	2 (1/2)		Only one treatment arm
TR_007	2 (1/2)		Only one treatment arm
PL_067	4 (0/4)	4 (0/4)	No relapses
SF_044	3 (0/3)	2 (0/2)	No relapses
TR_003	4 (0/4)	5 (0/5)	No relapses

A further concern with using a CMH test is that it implicitly treats censored patients as successes. The log-rank test or similar methods use all ITT patients and allow for censoring. The reviewer used an unstratified log-rank test as well as Score tests stratified by center and by country in the analysis of time to first relapse.

3.1.4.2 Analysis and Findings

The data for the primary endpoint analysis needed to be derived from the files submitted by the sponsor. The reviewer independently reproduced the end of trial dataset based on the sponsor's specification.

A feature of the CMH statistic is that under certain circumstances a stratum/center does not contribute to the test statistic. Therefore, centers with only one treatment arm or with no patients relapsing need to be grouped. Small centers are also usually grouped. Combining these centers into one Center, the CMH test has a p-value of 0.0971 (Table 6). Or similarly, grouping the original centers into their respective countries, the comparison of relapse rates adjusted for Country is also not statistically significant (p=0.0805).

As noted by the sponsor and the Medical Officer, five patients may actually have relapsed according to notation in the EOT CRF Item "Reasons not mentioned above, please specify _____". These five patients all received Gepirone ER and when considering them as relapsed the p-value for any of the CMH tests is no longer statistically significant (p=0.3302 for CMH adjusted for grouped centers and p=0.3145 for CHM adjusted for Country).

Table 7 gives the chi-square test results per country, with and without the five patients classified as relapsed. When the five patients are treated as censored, only Germany, which represents 16% of the total sample size, showed a nominally significant difference in relapse rates between the two treatments. In Turkey, the relapse rates favored placebo. When the five patients are considered having relapsed, the treatment effect in Germany is weakened and the opposite finding in Turkey is strengthened.

Another drawback of the CMH analysis is that only relapses are counted and censored patients are implicitly treated as successes. The time to relapse analysis uses all ITT patients and treats censored patients as such. Table 8 gives the results when comparing time to relapse between the two treatment groups. This reviewer used a different time variable than the sponsor and the p-values for the log-rank tests are somewhat different, but both are not statistically significant. When stratifying by center or by country, there is a further diminished difference between the two treatments. Figure 1 shows the time to relapse when the five patients in questions were censored. Figure 2 shows the same plots but classifying the five patients as having relapsed. In each case, there is no separation of the survival curves for the first six months on study.

A time to relapse analysis per Country (Table 9) shows an apparent treatment effect in only one of the five countries. However, when the five patients in question are treated as having relapsed, there are no longer any significant treatment differences in favor of Gepirone. On the contrary, the finding in Turkey favoring placebo appears to approach statistical significance.

Figure 3 and Table 8 show that there is no difference in time on study, whether due to censoring or due to relapse, between the two treatments.

Table 6: Summary of Relapse Rates

	Gepirone Relapse Rate	Placebo Relapse Rate	CMH p-value
Sponsor's Analysis	29/126 (23.0%)	43/124 (34.7%)	0.0244
Reviewer's Analyses:			
Grouping Small Centers	29/126 (23.0%)	43/124 (34.7%)	0.0971
Grouping Centers into Countries	29/126 (23.0%)	43/124 (34.7%)	0.0805
Grouping Small Centers and Reclassifying 5 Patients as Relapses	34/126 (27.0%)	43/124 (34.7%)	0.3302
Grouping by Country and Reclassifying 5 Patients as Relapses	34/126 (27.0%)	43/124 (34.7%)	0.3145

Table 7: Relapse Rates per Country

Country	Gepirone Relapse Rate	Placebo Relapse Rate	Chi-square p-value	Chi-square with Additional 5 Patients
Finland	5/24 (6/24)	9/29	0.4018	0.6274
France	17/45 (17/45)	15/37	0.7986	0.7986
Germany	1/12 (2/12)	14/28	0.0126	0.0486
Poland	2/27 (4/27)	5/21	0.1102	0.4283
Turkey	4/18 (5/18)	0/9	0.1255 *	0.0798*

*In favor of placebo

Table 8: Time to Relapse

	Gepirone (mean time)	Placebo (mean time)	Log-Rank P-Value	Stratified by Center*	Stratified by Country*
Time to Relapse	325	323	0.0891	0.2223	0.2047
Time to Relapse with Additional Five Patients Relapsed	318	322	0.2782	0.5319	0.4972
Time on Study	296	295	0.3263	N/a	N/a

*Based on Score test

Table 9: Relapse Rates and Time to Relapse per Country

Country	Treatment	Relapse Rates (Percent)	Log-Rank Test	Relapse Rates (Percent) with Additional Five Relapses	Log-Rank Test with Additional Five Relapses
Finland	Placebo	9/29 (31.0)	0.3819	9/29 (31.0)	0.5583
	Gepirone ER	5/24 (20.8)		6/24 (25.0)	
France	Placebo	15/37 (40.5)	0.8792	15/37 (40.5)	0.8792
	Gepirone ER	17/45 (37.8)		17/45 (37.8)	
Germany	Placebo	14/28 (50.0)	0.0488	14/28 (50.0)	0.1272
	Gepirone ER	1/12 (8.3)		2/12 (16.7)	
Poland	Placebo	5/21 (23.8)	0.0958	5/21 (23.8)	0.3606
	Gepirone ER	2/27 (7.4)		4/27 (14.8)	
Turkey	Placebo	0/9 (0.0)	*0.1120**	0/9 (0.0)	*0.0806**
	Gepirone ER	4/18 (22.2)		5/18 (27.8)	

* In favor of placebo.

Figure 1: Kaplan Meier Time to Relapse by Treatment Group

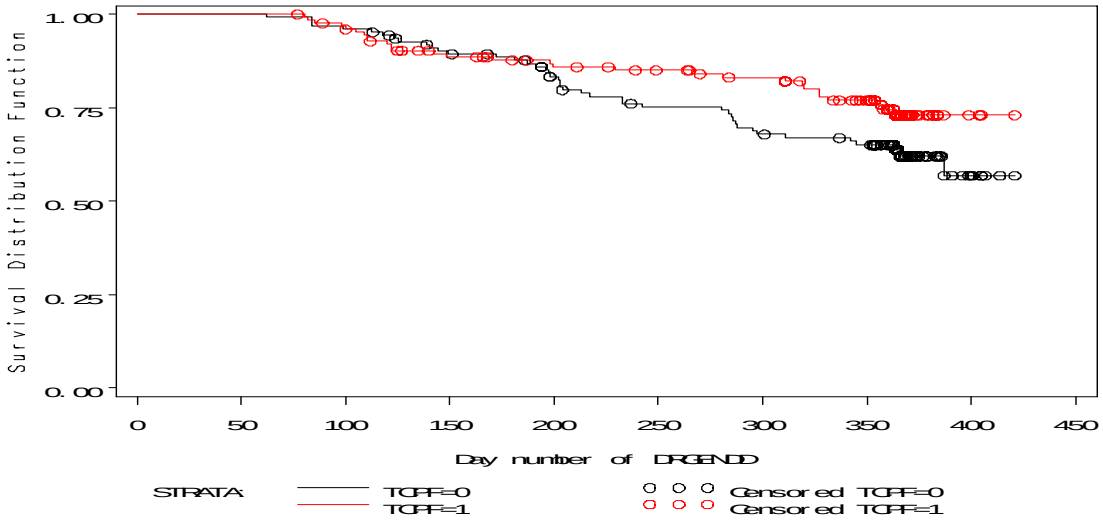


Figure 2: Kaplan Meier Time to Relapse Including Five Additional Patients

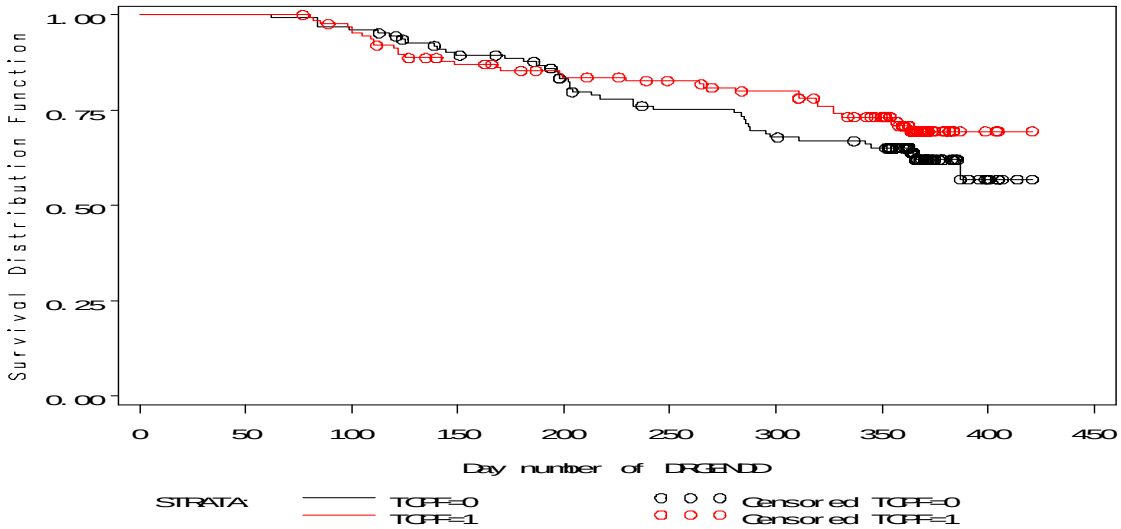
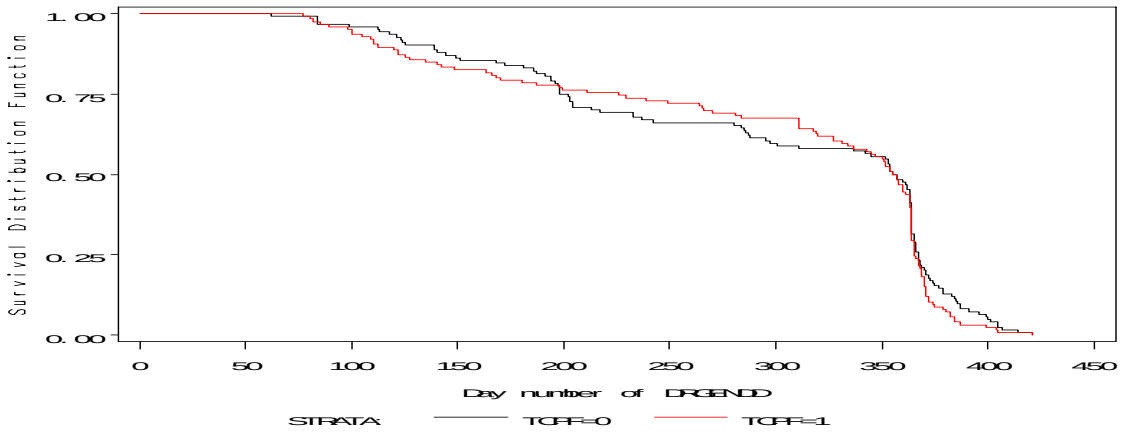


Figure 3: Kaplan Meier Time on Study by Treatment Group



3.1.4.3 Conclusion

This was a relapse trial in 250 MDD patients who had responded to open label Gepirone ER. The primary endpoint was relapse defined as an HAMD-17 total score of 16 or greater or a decision by the investigator that relapse criteria were met. The sponsor's primary analysis was a Cochran-Mantel-Haenszel center-adjusted treatment comparison of the proportions of relapse in the ITT population at study end. However, the sponsor's analysis did not use the results of 32 ITT patients. When using all ITT patients, the Cochran-Mantel-Haenszel test did not show relapse rates of Gepirone patients to be significantly lower than relapse rates of placebo patients when adjusting for centers or for countries. Relapse rates across center or across countries were not consistent. One country even showed numeric superiority of placebo over Gepirone. Furthermore, the CMH test implicitly treats censored patients as successes. The log-rank test for time to first relapse uses all ITT patients and treats censored patients as such. It did not reach statistical significance. The Score test, which is similar to a log-rank test, stratified by center or by country also did not approach statistical significance. The log-rank tests for time to first relapse per country mimicked the findings based on the relapse rates.

There are five patients who should have been classified as relapse but were not in the sponsor's primary analysis due to an apparent logistic mix-up. All five patients received Gepirone and reclassifying them as having relapsed further reduces any treatment differences.

3.2 Evaluation of Safety

This review did not address any safety issues.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

In discussion with the reviewing medical officer, Earl Hearst, M.D., HFD-120, it was decided that no subgroup analyses were required.

The analysis of relapse rates per Country can be considered a subgroup analysis but is discussed in the general body of the review (2.5.2.2)

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

As there is only one study, the statistical issues were discussed in the body of the review. In brief, they are the issues with the sponsor's primary analysis excluding 32 ITT patients, not dealing with censoring, and the five additional Gepirone patients who have relapsed.

The evaluation of the collective evidence of the 12/23/03 submission is based on one trial, Study 28709-2003, which the sponsor identified as meeting the outstanding requirement of a robustly-positive placebo-controlled trial. In this reviewer's opinion the results did not achieve this goal. The analysis of all ITT patients did not result in statistical significance in favor of Gepirone. An investigation of the treatment effect across countries found only one of the five countries with an apparent statistical superiority of Gepirone over placebo in relapse rates. One country favored placebo numerically. Time to first relapse also did not reach statistical superiority of Gepirone over placebo, whether stratified for center or country or not stratified. Time to first relapse per country mimicked the results of the relapse rates.

Five patients, all receiving Gepirone, were identified by the sponsor as likely relapses but were not coded as such for the primary analysis due to logistic reasons. Considering these patients as relapsed renders any statistical analysis non-significant.

There were additional six trials in the 12/23/03 submission. None of them were reviewed here because the sponsor did not identify them as demonstrating statistical superiority of Gepirone over placebo. It appears that there are additional 11 ongoing trials.

5.2 Conclusions and Recommendations

In the reviewer's opinion, the pivotal study submitted by the sponsor as satisfying the outstanding efficacy requirements identified in the Not Approvable Action letter of 05/12/02 did not reach its goal. The sponsor's statistically significant result based on the Cochran-Mantel-Haenszel test can not be accepted as it excludes 32 ITT patients. The proper application of the CMH test to the primary endpoint using all ITT patients did not reach statistical significance. Furthermore, the log-rank test of time to first relapse also did not distinguish significantly between Gepirone and placebo. When using a stratified (by center or by country) time to first relapse methodology, the treatment difference was further reduced. If the five patients identified by the sponsor as potentially having relapsed are included in the statistical analyses, none reach statistical significance.

None of the other studies submitted by the sponsor in the 12/23/03 submission were reviewed because the sponsor did not identify any as potentially meeting the goal.

File Directory: *C:\Data\MyDocuments\Gepirone_NDA21164_7.doc*

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

/s/

Roswitha Kelly
5/4/04 04:04:12 PM
BIOMETRICS

Kun Jin
5/6/04 08:45:11 AM
BIOMETRICS

James Hung
5/6/04 09:25:17 AM
BIOMETRICS

STATISTICAL REVIEW AND EVALUATION

Medical Division: Neuropharmacological Drug Products (HFD-120)
Biometrics Division: Division of Biometrics I (HFD-710)

NDA NO.:	21-164
SERIAL NO.:	Original NDA Resubmission
DATE OF RECEIVED BY THE CENTER:	05/18/01
DRUG NAME:	Gepirone HCl (Org 33062), Extended Release
DOSAGE:	ER Tablets: 20, 40, 60 and 80 mg
INDICATION:	Antidepressant
SPONSOR:	Organon Inc.
DOCUMENTS REVIEWED:	Electronic Submissions Received 05/18/01 and 01/15/02
NAME OF PROJECT MANAGER:	Paul David (HFD-120)
NAMES OF STATISTICAL REVIEWERS:	Roswitha Kelly, M.S. (HFD-710) Kooros Mahjoob, Ph.D. (HFD-710)
NAME OF CLINICAL REVIEWER:	Earl Hearst, M.D. (HFD-120)

Roswitha Kelly, M.S.
Mathematical Statistician

Kooros Mahjoob, Ph.D.
Mathematical Statistician

Concur:

Kun Jin, Ph.D.
Statistical Team Leader

George Chi, Ph.D.
Director, Division of Biometrics I

Distribution:
NDA 21-164 (Gepirone HCL ER)
HFD-120/Paul David
HFD-120/Earl Hearst, M.D.
HFD-120/Thomas Laughren, M.D.
HFD-710/Kooros Mahjoob, Ph.D.
HFD-710/Roswitha Kelly, M.S.
HFD-710/Kun Jin, Ph.D.
HFD-710/George Chi, Ph.D.
HFD-700/Charles Anello, Sc. D.

File Directory: *C:\Data\MyDocuments\Gepirone_NDA21164.doc*

TABLE OF CONTENTS

List of Tables

List of Figures

1. EXECUTIVE SUMMARY OF STATISTICAL FINDINGS[1]

1.1 Conclusions and Recommendations

Organon, Inc., filed this submission (NDA 21-164) in support of the safety and efficacy of Gepirone Extended Release (**ER**) in the treatment of outpatients with major depressive disorder (**MDD**). The sponsor submitted 18 studies as adequate and well-controlled, and designated four specific studies that reached statistical significance based on the sponsor's analyses, as providing 'proof of efficacy'. Of these four studies, one study (Study 134001) is a Phase III trial conducted by Organon with the ER dosage form. The other three were Phase II studies conducted by Bristol Myers Squibb (BMS) with the older Immediate Release (IR) formulation (Studies 03A7A-003, 03A7C-001-B and 03A7A-002). The primary efficacy parameters specified by the sponsor were 'change from baseline in the HAMD-17 Total score' for Studies 134001, 03A7A-003 and 03A7C-001-B. In addition, 'Percent Responders' based on CGI was the co-primary efficacy variable in Studies 03A7A-003 and 03A7C-001-B. Study 03A7A-002 was called a 'relapse' study by the sponsor with the 'time to relapse' endpoint defined post-hoc, even though the study was not designed as a relapse trial in the protocol. The primary analyses were performed at study end for the intent-to-treat population using last-observation-carried-forward. The reviewers were able to reproduce the sponsor's results. However, as will be elaborated on in this review, the reviewers' own analyses and evaluations have led to substantially different conclusions:

Study 134001 is an adequate and well-controlled Phase III trial with the ER dosage form conducted by Organon. It showed statistical significance in favor of Gepirone for the primary efficacy variable (HAMD-17) at the pre-specified time point. It is the only study (out of eight ER trials) that reached statistical significance with the to-be-marketed dosage form, and its efficacy has, therefore, not been replicated. Furthermore, the patients in this study were more aggressively titrated than those in the parallel study of 134002. This aspect of efficacy has also not been replicated.

Study 0A7A-003 is a Phase II trial conducted by (BMS) on the IR dosage form. It showed statistically significant superiority of Gepirone over placebo in both efficacy measures. However, this was a single-center trial, which started with small sample sizes (30 per treatment arm). High dropout rates (40%) further reduced the study size and the results may not be representative of the MDD patient population of interest. Interim analyses specified in the protocol, if carried out, may have introduced operational bias and the integrity of the findings may be in doubt. The finding of this small single-center trial is unreliable.

Study 03A7C-001-B is also an IR trial conducted by BMS. Based on the sponsor's analyses, it showed statistically significant Treatment and Treatment-by-Center effects. The reviewers found that the overall treatment effect was driven by a very small center, the Cole Center, which had an unusually low placebo response (see **Table 20** and **Figure 5**). The analyses without this outlier showed no difference between Gepirone and placebo (p=0.84).

[1] The reviewers would like to acknowledge and express their appreciation of the help received by Dr. Ohid Siddiqui. Due to the extensive analyses necessary, Dr. Siddiqui had volunteered his expertise in SAS programming and management of the data sets.

Study 03A7A-002 was the third IR trial identified by the sponsor. It was identified and analyzed by the sponsor as a relapse study. It is apparent from the protocol that this study was not designed to be a relapse study. There were no definitions of relapse or corresponding analyses specified in the BMS protocol. The original efficacy parameter was HAMD. The analysis based on HAMD-17 did not reach statistical significance.

Small sample sizes, single-center trials, high dropout rates, improper analyses, and planned interim analyses variously affect the strength of evidence provided by the three IR studies designated by the sponsor as supportive of the ER dosage form. In the reviewers' opinion, this support was not substantiated.

Further concerns with submission relate to the fact that 14 other adequate and well-controlled trials with either dosage form had failed. Several of these studies are very similar in design and conduct to the four identified in support of the claim. In the reviewer's opinion they should be included in an evaluation of the extent of the evidence.

The sponsor’s summary results for the four identified studies (Studies 134001, 03A7A-003, 03A7C-001-B, and 03A7A-002) can be found in **Table 1**. The reviewers' summary findings for these studies can be found in **Table 2**. The reader is referred to Sections 2.5 to 2.9 as well as Appendices 1 and 3 for the details.

Conclusion: The reviewers performed extensive analyses. The results were discussed within the Division of Biometrics I and with the Medical Division (HFD-120). The reviewers found that from the four studies identified by the sponsor, the single ER trial (out of eight) showed statistical significance in favor of Gepirone. However, three other comparable ER trials showed no efficacy and the efficacy of the ER dosage form has not been replicated. After the proper analyses were applied to the three IR studies, which were meant to provide support to the ER product, only one reached statistical significance. The validity of the findings, however, is questionable, because the study was small in size, single-center, suffered from high dropout rates, and may not be representative of the MDD patient population of interest. Furthermore, the study may have been compromised due to unblinding through interim analyses as specified in the protocol. Two additional IR trials, which were similar to the ones identified by the sponsor, did not show Gepirone as efficacious.

1.2 Overview of the Clinical Program and Studies Reviewed

Gepirone Hydrochloride (**Org 33062**) is a novel azapirone derivative that has both antidepressant and antianxiety activity. Organon Inc., the sponsor of Gepirone, wishes to market the extended release (**ER**) formulation of Gepirone for the antidepressant indication. In support of its application, Organon submitted the results of 18 randomized, placebo-controlled, with or without active control, 6-8 week studies, and 6 uncontrolled studies to the NDA (see **Table 3**). The clinical development program focused on the 18 controlled studies categorized by Organon as "adequate and well-controlled".

From the 18 studies, four were designated as providing "proof of efficacy". These are Studies 134001, 03A7A-003, 03AC-001B and 03A7A-002. Only Study 134001 was conducted using the ER dosage form by Organon in 1999; the other three studies were conducted by Bristol-Myers Squibb in 1987 and 1988 with the IR formulation. These studies are highlighted in Table 3. This review will focus on these 4 studies. However, the sponsor's results from the other 14 studies will be considered in reaching an overall conclusion.

1.3 Principal Findings

1.3.1 Sponsor's Results and Conclusions

The sponsor's methodological approach appeared prespecified and followed. The reviewers identified several technical issues in these studies which will be addressed in their analyses and discussed in detail in **Section 2.7 Statistical and Technical Issues**. The sponsor's results are presented here (**Table 1)**.

Table 1: Summary of Sponsor's Principle Findings

(Entries are extracted from Tables, 9, 11, 13 and 15)

	Study 134001		Study 03A7A-003		Study 03A7C-001-B			Study 03A7A-002	
Treatment	Placebo	Gepirone ER 20-80	Placebo	Gepirone IR 10-90	Placebo	Gepirone IR 5-45	Gepirone IR 10-90	Placebo	Gepirone IR 20-90
HAMD-17 LS Means ♣ P-Value vs. Placebo	6.8	9.4 0.018	2.7	6.7 0.009	6.4	10.1 0.015	11.3 0.001	9.0	13.0 0.070
CGI % Responders♦ P-Value vs. Placebo	36	44 0.251	23.3	57 0.009	39	59 0.018	55 0.05	41.2	56 0.279
Percent of Relapse♠ P-Value vs. Placebo	--	--	--	--	--	--	--	60.6	30.6 0.035

♣: Results are based on ANOVA, Using LOCF at Last Visit, with Treatment, Center, and Treatment-by-Center Interaction in the model.

♦: Results are based on CMH test.

♠: Results are based on Log-Rank test.

The principal analyses were ANOVA on 'change from baseline in HAMD-17 Total score' and Cochran-Mantel-Haenszel (CMH) on 'percent CGI responders'. In Study 03A7A-002, the 5th

relapse definition was the primary efficacy variable, which was evaluated by the log-rank test. All analyses were performed on the ITT populations and LOCF was used where appropriate.

The sponsor's ANOVA contained Treatment, Center, and Treatment-by-Center Interaction in the model.

Based on the sponsor's analyses, three of the studies reached statistical significance with respect to HAMD-17 at the pre- specified endpoint. The fourth study reached statistical significance with respect to the relapse parameter. All results favored Gepirone. The sponsor concluded that the results provided evidence, from some studies strong evidence, that Gepirone is an efficacious antidepressant in subjects with major depressive disorder and is safe and well tolerated.

1.3.2 Reviewers' Results and Conclusions

The sponsor's ANOVA analyses with the Treatment-by-Center interaction term in the model were prespecified in the protocols and are reasonable. The ICH E9 guidance proposes a somewhat different approach on modeling multi-center trials. The reviewers adopted the ICH recommendations and performed additional analyses to examine the robustness of the treatment effect. The reviewers' ANCOVA model contained Baseline, Treatment, and Center, (no interaction) according to ICH E9. LOCF to the last visit for the primary efficacy variable of 'change from baseline in HAMD-17 Total Scores' was used in the ANCOVA as well as for Fisher's Exact test on 'CGI Responders'.

A summary of the reviewers' results is presented in **Table 2**.

Table 2: Summary of Reviewers' Principle Findings

(Entries are extracted from Tables 17-22)

	Study 134001		Study 03A7A-003		Study 03A7C-001-B			Study 03A7A-002	
Treatment	Placebo	Gepirone ER 20-80	Placebo	Gepirone IR 10-90	Placebo	Gepirone R 5-45	Gepirone IR 10-90	Placebo	Gepirone IR 20-90
HAMD-17 LS Means ♣†	6.8	9.0	2.7	6.6	7.4	10.1	10.1	9.8	13.7
P-Value vs. Placebo		0.0186		0.0073		0.0657	0.0706		0.0840
CGI % Responders♦†	35.6	43.6	23.3	56.7	39.1	58.6	55.2	41.2	55.6
P-Value vs. Placebo*		0.314, [.251]		0.017[.009]		0.056 [0.044]	0.170[.122]		0.224[.232]
Results for Study 03A7C-001-B without the Cole Center									
HAMD-17 LS Means ♣†	--	--	--	--	9.0	11.0	10.1	--	--
P-Value vs. Placebo						0.2774	0.8406		
CGI % Responders♦†	--	--	--	--	45.0	63.9	56.1	--	--
P-Value vs. Placebo*						0.074	0.460		

♣: Results are based ANCOVA using LOCF at Last Visit with Baseline, Treatment and Center (no Interaction) in the model.
♦: Results are based on Fisher Exact test; ∗: The value inside [ ] is the P-value based on CMH test.
†: P-Values are adjusted for the multiple testing of Gepirone doses vs. placebo where appropriate.

The differences in P-values between Table 1 and Table 2 are not so much due to the different methodological approaches between the sponsor (ANOVA) and the reviewers (ANCOVA, ICH E9), but due to performing the proper analyses for Studies 03A7C-001-B and 03A7A-002. Briefly, the significant effect observed by the sponsor for Study 03A7C-001-B was driven by a single small center with an usually low placebo response (**Table 20** and **Figure 5**) and the

primary efficacy parameter for Study 03A7A-002 was HAMD, not relapse. In addition, the reviewers repeat that the significant findings of Study 0A7A-003 may not be reliable. These issues are addressed in detail in Sections 2.7 and 2.7 as well as in Appendices 1 and 3.

1.3.3 Extent of the Evidence in Support of the Efficacy Claim

Based on the four studies identified by the sponsor, the extent of evidence in support of the efficacy claim is limited to the ER Study 134001. Although the analysis results for IR Study 03A7A-003 was in favor of Gepirone, the reviewers consider the findings unreliable. This was a small-scale, single-center Phase II trial with 30 patients and a high percentage of dropouts. In addition, if the planned interim analyses were carried out, the blind may have been broken and an operational bias may have been introduced. The other two studies (03A7C-001-B and 03A7A-002) did not reach statistical significance, either because the sponsor had not fully addressed the outlier results of a center or had not used the per-protocol efficacy variable. In addition, there were 14 (7 ER and 7 IR) adequate and well-controlled studies included in the submission which all resulted in lack of efficacy. The sponsor gave justifications for excluding them from the overall evaluation of evidence as well as why each study did not reach statistical significance. In the reviewers' opinion, three Phase III trials using the ER formulation (Studies 134002, CN105-078 and CN105-083) were very similar in study design, size, dropout rates, etc., to the four primary studies and the sponsor's reasons for exclusion (incomplete sample size and high dropout rates) may have been insufficient. Study 134002 was identical in design and conduct to Study 134001, but did not approach statistical significance (**Table 41**, P-value=0.446). The efficacy results of Study CN105-078 were very weak and Study CN105-083 was outright negative showing no benefit of Gepirone over placebo. Of the additional 7 IR trials, two were found to be adequate in design. One of these studies was numerically in favor of Gepirone, but the other one was in favor of placebo. However, very high dropout rates in these studies would question the representativeness of any findings (for details, please see Section **3.3 Appendix 3: Evidence from Supportive Studies**). Therefore, the collective evidence is based on one positive ER trial and one positive but small single-center IR trial with questionable reliability. Two to seven other adequate and well-controlled studies can be considered, the results of which range from numerically favoring Gepirone to numerically favoring placebo.

1.3.4 Statistical Issues Which May Impair the Efficacy Conclusion

Statistical issues, which impact on the conclusions, are listed below. Some of these have already been mentioned in the Extent of Evidence section, because they affect both considerations.

- Only one study (Study 134001) using the ER formulation reached statistical significance and, therefore, the experience with the to-be-marketed product has not been replicated successfully.

- Study 03A7A-003 was a small single-center IR study that was positive for Geprione. However, due to high dropout rates and informative censoring and the potential operational

bias, the results may not be reliable. It, therefore, does not provide the support of the ER dosage form sought for by the sponsor.

- The sponsor's significant findings in Study 03A7C-001-B were driven by one, the smallest, center with unusually low placebo response (**Table 20** and **Figure 5**). The results excluding this center were not statistically significant (P-value=0.84).

- Study 03A7A-002 was analyzed by the sponsor as a relapse trial. However, this study was not defined as such in the protocol nor were there any relapse definitions specified in the protocol. The efficacy parameters used by the sponsor were developed post hoc after the blind had been broken.

- In all studies, high rates of discontinuation due to side effects, lack of efficacy, and lost to follow-up may have biased results for or against efficacy and may have impaired the generalizability of the results.

- Fourteen adequate and well-controlled trials did not show statistical significance. In the reviewers' opinion, several of these studies should enter into the overall evaluation of evidence.

- Three of the four designated efficacy studies (the BMS IR trials) had planned interim analyses in their protocols. The sponsor appears to have no record as to whether these interim looks were carried out. Therefore, there is no confidence that no operational bias was introduced if the interim analyses were carried out.

The reviewers will address each of these points in the following sections.

2 STATISTICAL REVIEW AND EVALUATION OF EVIDENCE

2.1 Introduction and Background

Gepirone Hydrochloride (**Org 33062**) is a novel azapirone derivative that has both antidepressive and anti-anxiety activity. It was originally developed by Bristol-Myers Squibb (BMS) and was first studied in an immediate-release (**IR**) formulation indicated primarily for anxiety disorders and major depression, but also for obsessive-compulsive disorders or in substance abuse.

After the safety assessment of the IR formulation, an extended-release (**ER**) formulation was developed to reduce the incidence of adverse experiences such as dizziness, nausea, and insomnia. The clinical trials with the ER formulation were primarily aimed at the antidepressive indication.

In 1992, BMS discontinued all trials with Gepirone ER in the major depressive disorder (**MDD**) indication. In 1993, Fabre-Kramer Pharmaceuticals, Inc., acquired the rights to Gepirone ER and completed a series of Phase I trials in special populations. In May 1998, Organon Inc. made an agreement with Fabre-Kramer to further develop and market Gepirone.

2.2 Overview of the Clinical Program and Studies Reviewed

The efficacy of Gepirone was assessed in 24 trials: 18 controlled and 6 uncontrolled clinical trials evaluating the efficacy/safety of Gepirone as an antidepressant agent. Of the 18 controlled trials, 8 were conducted using the ER dosage form and 10 with the IR formulation. The clinical development program focused on randomized, double-blind, placebo-controlled, 6-8 week Phase II and Phase III studies, with or without an active control. The 18 controlled studies randomized a total of 2263 patients: 1168 to Gepirone; 350 to Active Controls, and 745 to Placebo. Most studies lasted about one year. All but two uncontrolled trials were US studies. According to the sponsor, these studies are sufficiently diverse to include most population subgroups in need of antidepressant therapy.

Table 3 lists the 18 controlled and 6 uncontrolled studies by dosage form.

The product's development plan started in 1986 by BMS. By 1989 two of six adequate and well-controlled trials had shown statistical significance in support of the efficacy of Gepirone IR in the treatment of major depression and one had shown supportive evidence. In 1990 and 1991 an additional 10 adequate and well-controlled trials were conducted, none reaching statistical significance. Of these 10, six had been with the ER formulation. In 1999, Organon conducted two trials using the ER formulation, one demonstrated statistical significance, the other did not.

Table 3: Clinical Trial Program of Gepirone

Trial Number	Dosage Form	Sponsor	Control	Sample Size			Indication
				Gepirone	Placebo	Active Contrl	
Adequate and Well Controlled Trials for ER Dosage Form (8)							
134001	**ER**	**Organon**	**Plac**	**102**	**106**	**--**	**MDD not further specified**
134002	ER	Organon	Plac	107	104	--	**MDD** not further specified
CN105-078	ER	BMS	Plac	176	47	--	**MDD** not further specified
CN105-083	ER	BMS	Plac	146	39	--	**MDD** not further specified
CN105-052	ER	BMS	Plac	35	37	36	**MDD** not further specified
CN105-053	ER	BMS	Plac +	15	16	15	**MDD** not further specified
CN105-064	ER	BMS	Plac +	46	21	19	**MDD** not further specified
CN105-057	ER	BMS	Plac	196	49	--	**MDD** not further specified
Uncontrolled ER Trial (1)							
CN105-055	ER	BMS	--	--	--	--	--
Adequate and Well Controlled Trials for IR Dosage Form (10)							
CN105-043	IR	BMS	Plac	59	60	--	**MDD** or dysthymia
CN105-037	IR	BMS	Plac +	126	63	65	**MDD** not further specified
CN105-022	IR	BMS	Plac +	67	69	67	**MDD** or bipolar disorder
CN105-029	IR	BMS	Plac +	18	18	19	**MDD** or bipolar disorder
CN105-028	IR	BMS	Plac +	68	68	71	**MDD** or bipolar disorder
03A7A-003	**IR**	**BMS**	**Plac**	**30**	**30**	**--**	**MDD with an atypical profile**
03A7C-001A-2496	IR	BMS	Plac	154	41	--	**MDD** not further specified
03A7C-001B	**IR**	**BMS**	**Plac**	**141**	**70**	**--**	**MDD not further specified**
03A7C-001A-2486	IR	BMS	Plac	166	38	--	**MDD** not further specified
03A7A-002	**IR**	**BMS**	**Plac**	**36**	**34**	**--**	**MDD not further specified**
Uncontrolled IR Trials (5)							
CN105-039	IR	BMS	--	--	--	--	--
CN105-050	IR	BMS	--	--	--	--	--
CN105-019	IR	BMS	--	--	--	--	--
03A7A-001	IR	BMS	--	--	--	--	--
030L1-0004	IR	BMS	--	--	--	--	--

Note: Table is Reproduced from the Sponsor's Submission.

Trials listed by Dosage Form (ER and IR) and Design (adequate and well controlled or uncontrolled) and start date, from most recent to least recent.

+ includes active control arm.

BMS = Bristol-Myers Squibb; ER = Extended release; IR = Immediate release.

Table 4 provides a time line of the drug development program.

The selection of patients was based on the diagnostic criteria in the Diagnostic and Statistical Manual of Mental Disorders 3rd revised or 4th edition (DSM-III-R/IV) or based on the Research Diagnostic Criteria (RDC).

From the 18 "adequate and well-controlled" studies, the sponsor designated 4 (Studies 134001, 03A7A-003, 03AC-001B, and 03A7A-002) as providing "proof of efficacy". Of these, only Organon's Study 134001 was conducted with the to-be-marketed ER dosage form; the other 3

were IR formulation studies and had been conducted by Bristol-Myers Squibb. To facilitate the discussions, we refer to these four as "**4-designated-studies**". They are highlighted in Table 3. Our review focuses on the results of the 4-designated-studies. However, the sponsor's results from the other 14 adequate and well-controlled studies will be incorporated in reaching the final conclusion.

Section 2.4 will provide a summary description of the design, efficacy parameters, and the demographic characteristics of the 4-designated-studies.

Table 4: Drug Development Time Line

	1986	1987	1988	1989	1990	1991	1999	Total
# uncontrolled	2				1	3		6
# controlled		4	1	1	4	6	2	18
Subset:								
# with Efficacy		2	1				1	4
# Lack of Efficacy		2		1	4	6	1	14

2.3 Data Analyzed and Sources

The data from the primary efficacy parameter(s) were analyzed. HAMD-17 Total Scores were evaluated in all four studies and Percent Responders based on CGI in Studies 03A7A-003, 03AC-001B, and 03A7A-002. In addition, Time to Relapse was analyzed for completeness sake for Study 03A7A-002. Physician's Questionnaire (PQ) was not analyzed by Organon though it had been specified in the original protocol of Study 03A7A-002 as a primary efficacy variable. These efficacy parameters were either pre-specified in the protocols or identified by the sponsor as primary. The data were available as part of the electronic submission in rectangular format suitable for use with SAS software. The submission also provided the SAS programs used by the sponsor in the efficacy assessment of the data.

The data as submitted with the May 18, 2001, electronic submission did not permit the full reproduction of the sponsor's analyses. Therefore, the sponsor was requested to resubmit the data with the appropriate flags for identifying each patient with respect to the population (e.g. ITT) he belonged to, etc. Upon receipt of the new data set in January 2002, the reviewers were able to reproduce the sponsor's results and felt confident about the accuracy of the data submitted.

2.4 Description of the Four Studies Identified by the Sponsor

2.4.1 Trial Design and Patient Population

Table 5 summarizes the study characteristics of the 4-designated-studies reviewed. A more detailed description of each trial follows. All four studies were randomized, double-blind, parallel, placebo-controlled, single- or multi-center trials in outpatients diagnosed with major depressive disorder. Some studies included subjects with minor or moderate depression or with an atypical profile. The diagnosis was confirmed according to DSM-III/R or -IV or RDC criteria.

Sample sizes ranged between 30 and 106 subjects per treatment arm. Except for the low dose arm in one study (5-45 mg/day), most doses were titrated to 10-90 mg/day depending on tolerance and symptom control. The double-blind phase of the studies lasted 6 - 8 weeks.

Table 5: Design Characteristics of the 4-Designated-Studies

Study	Design	Patient Population	Treatment Dose (mg/day)	No. of Patients Randomized	Trial Duration Week	Titration Schedule	Titration Period
134001 (Organon)	R, DB, MC, PG, PC	Adult OP with MD to SD Met: DSM-IV	Gepirone **ER** 20- 80 Placebo QD	102 106	8-wk	Starting: 1 20-mg tablet in the Morning Increase:1 20-mg tablet each week up to Max of 4 tablets daily (QD)	8-wk DB
03A7A-003 (BMS)	R, DB, SC, PG, PC	Adult OP with MaD, MiD, InD, Met RDC With Atypical Profile	Gepirone **IR** 10-90 Placebo BID	30 30	8-wk	Starting: 1 10-mg capsule at HS Increase:1 10-mg capsule each 2-4 days up to Max of 9 capsules daily (BID)	3-wk DB
03A7C-001B (BMS)	R, DB, MC, PG, PC	Adult OP with MD to SD Met: RDC	Gepirone **IR** 10- 90 Gepirone **IR** 5-45 Placebo	70 71 70	8-wk extended 44-wk	Starting: 1 5-mg capsule at HS Increase:1 10-mg capsule each 2-4 days up to Max of 9 capsules daily (BID)	3-wk DB
03A7A-002 (BMS)	OL: 6-wk Then DB: 6-wk R, DB, MC, PG, PC	Adult OP with MD to SD Met: RDC	OL: Gepirone **IR** 10- 90 DB: Gepirone **IR** 10- 90 Placebo	134 36 34	6-wk OL Responders Ran. to 6-wk DB	Starting: 1 10-mg capsule with meal Increase: Not more than 1 10-mg capsule every other day up to Max of 9 capsules daily (BID or TID) during open label period.	6- wk OL

R: Randomized; DB: double-blind; SC/MC: Single/Multi-center; PC: Placebo Controlled; PG: Parallel-group; OP??: Outpatient; MD: Moderate Depression; SD: Sever Depression; MiD: Minor Depression; MaD: Major Depression; InD: Intermediate Depression; DMS-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th ed.; ER = Extended release; IR = Immediate release; RDC = Research Diagnostic Criteria studies; OL: Open Label;
QD: Once a day; BID = Twice a day; TID: Three times a day; HS = Just before sleep.

Study 134001

Study 134001 was a 5-center, double-blind parallel study where a total of 208 outpatients were randomized to Gepirone ER (20-80 mg) once per day in the morning (n=102) or to placebo (n=106).

Prospective subjects were those suffering from major depressive disorder (**MDD**). They were required meeting all of the following criteria in order to be included in the trial:

- Age between 18 and 70 years of age.
- Meet diagnostic criteria for moderate to severe major depressive disorder according to the DSM-IV criteria.
- Have a total score of 20 or greater on the HAMD-17 at both screening and baseline assessments.

- Have significant daily dysphoria for the past four weeks.
- Provide written informed consent.

Subjects suffering from atypical MDD were included in the study but were not evaluated rigorously as a distinct subgroup.
The study lasted eight weeks after a washout week. Doses were titrated from 20mg to 80 mg, depending on acceptability of the medication and therapeutic response. Subjects who completed the 8-week placebo-controlled phase were offered to continue for another 44 weeks in an open-label extension.

Study 03A7A-003

Study 03A7A-003 is a single-center, double-blind trial of Gepirone IR (10-120 mg/day) and placebo. A total of 60 subjects were randomized to Gepirone (n=30) or placebo (n=30).

Prospective subjects were outpatients who met RDC criteria for major, minor, or intermittent depression with a specified atypical profile. They were required meeting all of the following criteria to be included in the trial:

- Be at least 18 years of age.
- Male or female outpatients. Females of child-bearing potential must not be pregnant and must be using an acceptable form of birth control. Subjects must have a minimum score of 10 on the first 21 items of the HAMD-25 at baseline (amended from $\geq$ 18), must have maintained mood reactivity while depressed, and showed one or more of the following: increased appetite or weight gain while depressed, oversleeping, or spending more time in bed while depressed, severe fatigue, creating a sensation of leaden paralysis, or extreme heaviness of arms or legs when depressed, or hypersensitivity to rejection as a trait throughout adulthood.

The study lasted 8 weeks after a 7-14 day baseline period.

Study 03A7C-001-B

This was a 3-center study consisting of a short-term and a long-term Phase. The short-term, double-blind phase is the subject of evaluation in this review. It lasted 8 weeks. Overall, the trial was designed as a randomized, double-blind study with three parallel treatment arms to investigate the efficacy and safety of Gepirone in outpatients meeting the RDC for MDD. Those subjects responding to treatment during the short-term phase were eligible for additional 44 weeks of double-blind treatment (long-term phase).

Eligible subjects were outpatients satisfying RDC for a major depressive disorder (MDD) who met the following criteria:

- Male or female outpatients who were $\geq$18 years of age.
- Female subjects who were postmenopausal, surgically sterile, or using an adequate method of birth control. These subjects were given a pregnancy test before entering the study. If the

subject became pregnant during the trial, all treatment was to be stopped and the subject followed.

- Subjects who met RDC criteria for MDD of at least four weeks duration of the following subtypes:
  - Single episode
  - Recurrent episodes
- Subjects who had ≥20 on the HAMD 25 at the end of baseline (amended from a score of ≥20 on the first 17 items).
- Subjects who gave written informed consent prior to entering the study.

A total of 360 eligible subjects were to be recruited but due to premature termination by BMS only 211 (58.6%) subjects were randomized into the short-term phase. The low dose Gepirone IR (5-60 mg/day) had 71 subjects, the high dose Gepirone IR (10-120 mg/day) had 70, and the placebo arm contained 70.

Study 03A7A-002

This is a 5-center trial to compare Gepirone IR (10-90 mg/day) with placebo. After the conclusion of a six-week open-dose titration phase in subjects with MDD, responders to Gepirone IR were recruited to participate in a six-week, double-blind, placebo substitution, randomized withdrawal phase. Of the 134 subjects who entered the open-label phase, 70 were considered responders at the end of that phase. Of these, 36 were randomized to Gepirone and the remainder to placebo. **The sponsor called it a relapse trial giving six post hoc definitions of relapse as primary efficacy parameters after the blind had been broken. The original protocol defined only of a comparison of HAMD scores between treated and untreated subjects.**

For the double blind phase, the study included outpatients with MDD according to RDC who met the following criteria:

- Between 18 and 65 years of age.
- Baseline total HAMD ≥ 22 on the first 17 items, later amended to require a ≥20 on all 25 items of the HAMD (Amendment No. 4).
- Completion of the open-label phase with a reduction on the HAMD-17 Total Score of ≥ 12 or by ≥50% and a Clinical Global Impression (CGI) score of at least "moderately improved."

2.4.2 Efficacy Parameters

Table 6 presents the primary and secondary efficacy parameters of the 4-designated-studies as identified in the protocols or the sponsor's study reports.

2.4.3 Demographics

Table 7 summarizes the demographic characteristics of the subjects of the 4-designated-studies. Average age and height appear fairly consistent across studies. However, it seems unusual to

observe the highest average weight in the study with a high proportion of females. The observed large standard deviations for weight may indicate that this anomaly is due to a few individuals. The subject populations were mostly white, between 1/2 and 2/3 female, with a mean age of around 42 years.

Table 6: Efficacy Parameters in the 4-Designated-Studies

Study	Primary	Secondary	Primary	Secondary or Tertiary									Other
	HAMD-17 Total Score Change from Baseline	HAMD-17 % Responder	GCI % Responder	GCI-S Change from Baseline	CGI-I Score	HAMD 1	HAMD 21, 25/28	HAMD Factor 1	HAMD [a] Factors V, VI	HAMA Change from Baseline	MADRS Total	SCL [b] Factors Change from Baseline	
134001	X*	X	X	X	X	X	X	X			X		X
03A7A-003	X*	X	X*	X	X	X	X	X	X	X	X	X	X
03A7C-001B	X*	X	X*		X	X	X	X	X	X	X	X	X
03A7A-002	X	X	X	X	X	X	X	X	X	X		X	X[c] *

CGI = Clinical Global Impression; CGI- I = Clinical Global Impression- Improvement; CGI- S = Clinical Global Impression- Severity; HAMA = Hamilton Anxiety Scale; HAMD = Hamilton Depression Scale; MADRS = Montgomery and Åsberg Depression Rating Scale; SCL = Symptom Checklist a Includes Factor V (retardation) and Factor VI (sleep disturbance) b SCL- 87 includes 7 factors (somatization, obsessive- compulsive, interpersonal sensitivity, depression, anxiety, anger/ hostility, phobic Anxiety and 4 indices (total score, general symptom index, positive symptom index, positive symptom distress index); SCL- 90 includes 2 additional factors: paranoid ideation and psychoticism
c Includes the proportion of subjects with relapse
* Indicates primary efficacy parameter as identified by the sponsor

Table 7: Demographics of the 4-Designated-Studies

Treatment	N	Age (yr.)	Height (in.)	Weight (lb.)	Gender % F/ M	Race %				
						Bl	His	Ori	Wh	Other
Study No. 134001										
Gepirone ER 20-80 mg/day	101 (102)	39.5 ± 11.3	66.6 ± 3.6	183 ± 42.7	66/34	8	NA	1	72	19
Placebo	103 (106)	40.6 ± 11.7	66.8 ± 4.3	182 ± 53.0	54/46	11	NA	3	75	12
Study No. 03A7A-003										
Gepirone IR 10-90 mg/day	30	41.4 ± 10.0	67.8 ± 6.0	162 ± 28.8	33/67	3	0	0	97	0
Placebo	30	37.4 ± 8.8	69.2 ± 4.6	168 ± 40.1	37/63	3	0	0	93	3
Study No. 03A7C-001B										
Total Gepirone IR	137	41.7 ± 9.9	67.7 ± 3.9	167 ± 39.4	55/45	7	0	0	93	0
Gepirone IR 5-45 mg/day	70	43.0 ± 10.1	67.3 ± 3.9	159 ± 35.1	54/46	3	0	0	97	0
Gepirone IR 10-90 mg/day	67	40.3 ± 9.6	68.1 ± 3.7	175 ± 42.3	55/45	10	0	0	90	0
Placebo	69	44.0 ± 11.3	66.5 ± 3.6	169 ± 38.7	67/33	4	0	0	96	0
Study No. 03A7A-002[a]										
Gepirone IR 20-90 mg/day	36	42.7 ± 9.1	67.5 ± 3.9	169 ± 38.1	47/53	8	0	0	92	0
Placebo	34	42.7 ± 10.1	67.5 ± 3.5	174 ± 37.6	41/59	0	0	3	97	0

ER = Extended release; IR = Immediate release.

For race, Bl = Black; His = Hispanic; Ori = Oriental; Wh = White; NA = Not applicable: recorded as "Other" on CRF for Protocol 134001.

2.5 Statistical Evaluation of Evidence of Efficacy

2.5.1 Sponsor's Analysis, Results and Conclusions

2.5.1.1 Statistical Methodologies

Except for Study 03A7A-002, the primary efficacy parameter was change from baseline in the (re-calculated) HAMD-17 total score. The primary statistical analysis compared the ITT populations of the placebo and Gepirone treated subjects using LOCF to the defined study endpoint. The statistical analysis used an ANOVA model with treatment, center, and their interaction as terms. In SAS representation the model was:

ΔHAMD-17 = Treatment Center Treatment*Center

Where '**ΔHAMD-17'** represents change from baseline in HAMD-17 Total Score.

Interaction was tested at $\alpha = 0.10$ and if significant, findings were discussed whether they still could present an overall estimate of the treatment effect. A two-sided P=0.05 was used for testing treatment effect. The HAMD-17 total score was re-calculated if 2 ('4' in one study) or less items were missing, using the following adjustment.

$$\text{Adjusted HAMD-17 Total} = 17 \times \left[\frac{\text{Total for non-missing items}}{\text{Number of non-missing items}}\right]$$

If more items were missing, the total score was not calculated.

Assumptions of ANOVA were investigated by statistical tests and a non-parametric model (Wilcoxon test adjusting for center) was used if necessary.

Two studies (03A7A-003 and 03A7C-001-B) had a co-primary efficacy parameter, namely proportion of GCI responders as defined by subjects having a 1 or 2 on the CGI global improvement score. This parameter was analyzed using Cochran-Mantel-Haenszel pair-wise comparisons. Again, LOCF was used and the main comparison was at study end. A finding with a two-sided $P \leq 0.05$ was considered statistically significant.

In the relapse study (Study 03A7A-002), the primary endpoint was time to first relapse and was analyzed by the log-rank test with fixed right censoring. Six (6) definitions of relapse, see below, were developed. Definition 5 was considered to be the most accurate measure of relapse and hence, was considered the primary efficacy parameter.

1. Return to $\geq$ 75% of the (pre-treatment) baseline HAMD-17 total score
2. CGI Improvement score of 'No Change' or 'Worse than (pre-treatment) baseline' ($\geq$4).
3. Return to $\geq$ 75% of the (pre-treatment) baseline HAMD-17 total score, or discontinuation.
4. CGI Improvement score of 'No Change' or 'Worse than (pre-treatment) baseline' ($\geq$4), or discontinuation.
5. Return to $\geq$ 75% of the (pre treatment) baseline HAMD-17 total score, or discontinuation due to lack of efficacy.

6. CGI Improvement score of 'No Change' or 'Worse than (pre-treatment) baseline' ($\geq$4), or discontinuation due to lack of efficacy.

For the simplicity we shall use hereafter:

- ΔHAMD-17 to represent change from baseline in HAMD-17 Total Score
- CGI-PR to represent CGI percent responders

2.5.1.2 Analysis and Findings

This section presents the patient dispositions and the sponsor's main efficacy results of the 4-designated-studies. For the design, efficacy parameters, and demographic make-up of these studies the reader is referred to Sections 2.4.1-2.4.3. The statistical methodology utilized by the sponsor can be found in Section 2.5.1.1.

The results presented consist of: (1) Patient disposition to show patients dropouts and (2) sponsor's results with respect to the primary efficacy parameters:

Study 134001	ΔHAMD-17
Study 03A7A-003	ΔHAMD-17 and CGI-PR
Study 03A7C-001-B	ΔHAMD-17 and CGI-PR
Study 03A7A-002	Percent Relapse

Study 134001

This is a 5-center, double-blind, parallel group, 8-week, Phase III study using the extended release (ER) formulation of Gepirone conducted by Organon. A total of 208 subjects were randomized to treatment (n =102 to Gepirone 20-80 mg and n=106 to placebo).

ANOVA using LOCF on change from baseline in HAMD-17 in the ITT population at study end (week 8 = visit 6) was the primary analysis.

Patient Disposition

Table 8 shows that of the total of 208 subjects, 53 (25.5%) discontinued the trial and that the discontinuation was very similar among the Gepirone and placebo treated subjects (28 (27.5%) and 25 (23.6%), respectively). A higher proportion of subjects in the Gepirone group (10 [9.8%]) discontinued due to AEs/SAEs than did in the placebo group (3 [2.8%]). Four subjects in each treatment group (3.9% in Gepirone and 3.8% in placebo) discontinued due to lack of efficacy. Fourteen (13.7%) subjects in the Gepirone and 18 (17.0%) subjects in the placebo group withdrew from the trial for reasons not mentioned above; most of them were lost to follow-up.

Efficacy Results

Table 9 presents the efficacy results with respect to ΔHAMD-17 (primary) and CGI-PR (secondary).

Based on these findings, the sponsor concluded that there was a statistically significant difference in favor of Gepirone over placebo with respect to ΔHAMD-17 at week 1, (P=0.052), week 3 (P=0.013), week 5 (P=0.051), and at visit 6/Endpoint of treatment (P=0.018). Supportive evidence was observed at all other time points and in secondary efficacy parameters, which also reached statistical significance at various time points.

Table 8: Patient Disposition (Study 134001)

APPEARS THIS WAY ON ORIGINAL

	Treatment Group		Total
	Gepirone	Placebo	
Number	103	106	209
Number Actually	102 (100%)	106 (100%)	208 (100%)
Number of Patients			
Week 1	4 (3.9%)	4 (3.8%)	8
Week 2	6 (5.9%)	3 (2.8%)	9
Week 3	4 (3.9%)	4 (3.8%)	8
Week 4	4 (3.9%)	3 (2.8%)	7
Week 5	0 (0.0%)	5 (4.7%)	5
Week 6	3 (2.9%)	3 (2.8%)	6
Week 7	4 (3.9%)	1 (0.9%)	5
Week 8	3 (2.9%)	2 (1.9%)	5
Total Discontinued	28 (27.5%)	25 (23.6%)	53 (25.5%)
Total Completed	74 (72.5%)	81 (76.4%)	155

[a] Subjects (b) (6) were randomized to Gepirone but were treated with Placebo.
Subjects (b) (6) were randomized to Placebo but were treated with Gepirone. [b] All-Subjects-Treated Group
Data in this table were taken from Appendix F8.1.1-1.

Table 9: Sponsor's HAMD-17 and CGI Responders Results (Study 134001)

APPEARS THIS WAY ON ORIGINAL

Endpoint	Treatment	Statistic	Week of the trial					
			Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6/ET*
HAMD-17 (Total Score)	Gepirone N=101	N	98	100	101	101	101	101
		LS Mean	3.3	5.74	7.86	8.23	8.44	9.04
		SE	0.47	0.58	0.65	0.67	0.75	0.78
	Placebo N=103	N	99	101	101	101	101	101
		LS Mean	2.17	4.4	5.86	6.78	6.63	6.75
		SE	0.46	0.57	0.64	0.66	0.74	0.77
	P-Value		0.052	0.059	0.013	0.078	0.051	0.018
CGI (% Responders)	Gepirone N=101	N	98	100	101	101	101	101
		%	10	24	34	41	39	44
		SE	10.2	24	33.7	40.6	38.6	43.6
	Placebo N=103	N	99	101	101	101	101	101
		%	6	14	24	35	37	36
		SE	6.1	13.9	23.8	34.7	36.6	35.6
	P-Value		0.288	0.067	0.121	0.385	0.772	0.251

* Visits 1-4 correspond to Weeks 1-4; Visit 5 took place at Week 6 and Visit 6 represents Week 8/end of study.

With respect to CGI-PR specifically (this was not a co-primary efficacy parameter), there was no statistically significant difference between the Gepirone and placebo treatment groups at any visit. However, a marginally ($0.05 < P \leq 0.10$) significant difference in favor of Gepirone, was observed in CGI-PR at Visit 2.

Overall, the sponsor concluded that the results provide strong evidence that Gepirone at a dose range of 40-80 mg/day is an efficacious antidepressant in subjects with major depressive disorder, and is safe and well tolerated.

Study 03A7A-003

This is a single-center, double-blind study, classified as a Phase II trial, of Gepirone IR 10-120 mg/day. A total of 60 subjects were randomized to Gepirone (n=30) or placebo (n=30). The study lasted 8 weeks after a 7-14 day baseline period.

ΔHAMD-17 and CGI-PR were the primary efficacy variables. ANOVA using the LOCF on ΔHAMD-17 and Cochran-Mantel-Haenszel on CGI responders were the primary efficacy analyses.

Patient Disposition

Table 10 shows the patient disposition. Of the 60 subjects total, 24 (40.0%) discontinued the trial. Of these, 13 (43.3%) and 11 (36.7%) had received Gepirone and placebo, respectively. The most frequent reason for discontinuation was AEs in the Gepirone group (8 (26.7%) subjects) and lack of efficacy in the placebo group (6 (20.0%) subjects). More Gepirone (26.7%) than placebo subjects (13.3%) discontinued the study because of an AE. On the other hand, more placebo (20.0%) than Gepirone (3.3%) subjects discontinued due to lack of efficacy.

Efficacy Results

Table 11 presents the efficacy results with respect to ΔHAMD-17 and CGI-PR. ΔHAMD-17 was statistically significantly superior to placebo at week 2 (p=0.036), week 5 (p=0.004), week 7 (p=0.038), and week 8/endpoint (p=0.009). With respect to CGI-PR, statistical significance in favor of Gepirone over placebo was achieved at weeks 5 (p= 0.010) and 8/Endpoint (p = 0.009).

The sponsor concluded that overall there were no safety concerns in this study. The efficacy findings provide definite evidence of efficacy of Gepirone in the treatment of subjects with major, minor, or intermittent depression who manifested atypical symptoms.

Table 10: Patient Disposition (Study 03A7A-003)

APPEARS THIS WAY ON ORIGINAL

	Treatment Group		Total
	Gepirone 10-90	Placebo	
Randomized	30	30	60
All-Subjects-Treated	30 (100%)	30 (100%)	60 (100%)
Population			
Intent-to-Treat Population	30 (100%)	30 (100%)	60 (100%)
Evaluable Population[a]	29 (96.7%)	29 (96.7%)	58 (96.7%)
Discontinued Treatment by:			
Week 1	1 (3.3%)	0 (0.0%)	1 (1.7%)
Week 2	5 (16.7%)	1 (3.3%)	6 (10.0%)
Week 3	2 (6.7%)	1 (3.3%)	3 (5.0%)
Week 4	0 (0.0%)	3 (10.0%)	3 (5.0%)
Week 5	1 (3.3%)	0 (0.0%)	1 (1.7%)
Week 6	1 (3.3%)	0 (0.0%)	1 (1.7%)
Week 7	0 (0.0%)	3 (10.0%)	3 (5.0%)
Week 8	3 (10.0%)	3 (10.0%)	6 (10.0%)
Total Discontinued[b]	13 (43.3%)	11 (36.7%)	24 (40.0%)
Completed Study	17 (56.7%)	19 (63.3%)	36 (60.0%)

a Subjects from the ITT population who had a minimum of two weeks of documented exposure to study medication. b One placebo subject (b) (6)) discontinued the study after > 59 days of treatment and is included in the total discontinued group and in discontinuations at week 8.
Note: Data for this table were derived from Appendix F 6.1-1 and the Supplement to report table in Appendix F.

Table 11: Sponsor's HAMD-17 and CGI Responders Results (Study 03A7A-003)

Parameter	Treatment	Weeks of Study							
		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
HAMD-17 Change from Baseline	Placebo N =30 **Δ**	1.3	2.1	2.2	2.8	2.7	3.5	3.2	2.7
	Gep 10-90 N=30 **Δ**	2.2	4.7	4.3	4.8	6.5	5.1	6.2	6.7
	P-value =	0.457	0.036	0.115	0.146	0.004	0.278	0.038	0.009
CGI Percent Responders	Placebo N=30 **%**	9	24	23	27	27	37	30	23
	Gep 10-90 N=30 **%**	12	27	37	47	60	47	53	57
	P-value =	0.749	0.825	0.264	0.111	0.010	0.436	0.069	0.009

Study 03A7C-001-B

This was a 3-center, double-blind, randomized trial with three parallel treatment arms to study the efficacy and safety of Gepirone IR. The study consisted of a short-term and a long-term phase. A total of 211 subjects were randomized into the short-term phase, namely 71 subjects to low-dose Gepirone (5-60 mg/day), 70 subjects to high-dose Gepirone (10-120 mg/day), and 70 subjects to placebo.

The primary efficacy analyses compared ΔHAMD-17 and CGI-PR of placebo with each treatment arm in the ITT populations. The ΔHAMD-17 variable was analyzed via ANOVA using LOCF at week 8. Statistical significance was established without adjusting for multiple

comparisons (combined doses), as well as by using Dunnett's multiple comparison procedure (individual dose-placebo comparisons). CGI-PR was analyzed using the Cochran-Mantel-Haenszel (**CMH**) test.

Patient Disposition

Table 12 gives the patient disposition. Of the total of 211 randomized subjects, 105 (49.8%) completed the short-term phase, namely 41 (57.7%) subjects in the Gepirone 5-45 mg/day group, 32 (45.7%) subjects in the Gepirone 10-90 mg/day group, and 32 (45.7%) subjects in placebo group. A total of 106 (50.2%) subjects discontinued the short-term phase, including 30 (42.3%) from the Gepirone 5-45 mg/day arm, 38 (54.3%) from the Gepirone 10-90 mg/day arm, and 38 (54.3%) from the placebo arm. The most frequent reason for subjects discontinuing treatment were AEs, namely 23.9% from the low-dose Gepirone group, 40.0% from the high-dose Gepirone group, and 11.4% from the placebo group. There was a 30.0% dropout due to lack of efficacy in placebo group as compared to 7.1% and 11.3% of the high- and low-dose Gepirone groups, respectively.

Efficacy Results

Table 13 summarizes the efficacy results with respect to ΔHAMD-17 and CGI-PR.

With respect to ΔHAMD-17 Table 13 shows:

- When both Gepirone arms were combined (not the pre-specified primary analysis), the placebo-Gepirone comparison reached statistical significance at Week 4 (P=0.006), Week 6 (P=0.002), and Week 8/endpoint (P=0.001).

- Using adjusted p-values for the high-dose Gepirone - placebo comparison, there were statistically significant differences at weeks 4, 6, and 8/endpoint (P=0.006, 0.002 and 0.001, respectively).

- Using adjusted p-values for the low-dose Gepirone - placebo comparison, there were significant differences at Week 6 (P = 0.020) and Week 8/endpoint (P=0.015).

With respect to CGI-PR Table 13 shows:

- When both Gepirone doses were combined, there were statistically significant (adjusted) differences from placebo at Week 6 (P=0.032) and Week 8/endpoint (P=0.013).

- The comparison of low-dose Gepirone with placebo showed statistical significance (adjusted) at Week 6 (P=0.042) and Week 8/endpoint (P=0.018).

- The comparison of high-dose Gepirone with placebo showed statistical significance (adjusted) at Week 8/endpoint (P=0.050).

Table 12: Patient Disposition (Study 03A7C-001-B)

	Treatment Group			Total
	Gepirone 5-45 N (%)	Gepirone 10-90 N (%)	Placebo N (%)	
Randomized Subjects	71	70	70	211
Subjects in Treated Population	71 (100%)	70 (100%)	70 (100%)	211 (100%)
Subjects in ITT Population	70 (98.6%)	67 (95.7%)	69 (98.6%)	206 (97.6%)
Subjects in Evaluable Population	66 (93.0%)	57 (81.4%)	65 (92.9%)	188 (89.1%)
Subjects Discontinued Treatment by Week				
Week 1	5 (7.0%)	9 (12.9%)	4 (5.7%)	18 (8.5%)
Week 2	2 (2.8%)	8 (11.4%)	4 (5.7%)	14 (6.6%)
Week 3	3 (4.2%)	3 (4.3%)	4 (5.7%)	10 (4.7%)
Week 4	7 (9.9%)	6 (8.6%)	7 (10.0%)	20 (9.5%)
Week 5	6 (8.5%)	8 (11.4%)	11 (15.7%)	25 (11.9%)
Week 6	7 (9.9%)	3 (4.3%)	7 (10.0%)	17 (8.1%)
Week 7	0 (0.0%)	1 (1.4%)	1 (1.4%)	2 (1.0%)
Total Number of Subjects Discontinued	30 (42.3%)	38 (54.3%)	38 (54.3%)	106 (50.2%)
Total Number of Subjects Completed	41 (57.8%)	32 (45.7%)	32 (45.7%)	105 (49.8%)

a Last visit for evaluable data for the subject.
Note: Data for this table were derived from Appendices F 7.1.1-1, 7.1.1.1A, 7.1.1-3, 7.1.1.3A, 7.1.1-4, 7.1.1.4A, 7.1.1-6, and 7.1.1.6A.

Table 13: Sponsor's HAMD-17 and CGI Responders Results **(**Study 03A7C-001-B)

Parameter	**Treatment**	**Weeks of Study**					
		Week 1	Week 2	Week 3	Week 4	Week 6	Week 8
HAMD-17 Change from Baseline	Placebo N=69 **Δ =**	3.3	6.0	7.1	6.3	6.6	6.4
	Gep 5-45 n=70 **Δ =**	4.1	6.9	8.2	9.2	10.2	10.0
	P-value =	0.515	0.509	0.490	0.053	0.020	0.015
	Gep 10-90 n=67 **Δ =**	5.0	8.0	9.3	10.4	11.3	11.3
	P-value =	0.152	0.143	0.131	0.006	0.002	0.001
	Both Gep n=137 **Δ =**	4.7	7.5	8.8	9.9	10.8	10.7
	P-value =	0.190	0.207	0.194	0.006	0.002	0.001
CGI Percent Responders	Placebo n=69 **% =**	11	33	36	35	38	39
	Gep 5-45 n=70 **% =**	18	35	39	49	54	59
	P-value =	0.248	0.859	0.772	0.091	0.042	0.018
	Gep 10-90 n=67 **%=**	5	31	42	43	51	55
	P-value =	0.178	0.825	0.461	0.285	0.096	0.05
	Both Gep n=137 **% =**	12	33	40	46	53	57
	P-value =	0.904	0.988	0.547	0.110	0.032	0.013

P-Values for the comparisons of Gepirone 10 to 90mg/day and Gepirone 5-45mg/day vs. Placebo are adjusted for multiple comparisons.

Overall, the sponsor concluded that at both low and high dose ranges, Gepirone was effective in the treatment of subjects with MDD based on both primary and secondary outcome measures.

Study 03A7A-002

This was a 5-center study in subjects with MDD, comparing Gepirone IR (10-90 mg/day) with placebo. After the conclusion of a six-week open-dose titration phase, 70 subjects of the original 134 (52.2) met the criteria for responders (per protocol definition) and were randomized to the double-blind controlled phase (36 to Gepirone and 34 to placebo). The sponsor used six definitions of relapse to assess efficacy.

Patient Disposition

Table 14 shows the patient disposition. Of the total of 70 randomized subjects, 39 (55.7%) completed the double-blind phase: 21 (58.3%) subjects in the Gepirone 20-90 mg/day group and 18 (52.9%) subjects in the placebo group. A total of 31 (44.3%) subjects discontinued the double-blind phase, namely 15 (41.7%) of Gepirone group and 16 (47.1%) of placebo group. The most frequent reason for treatment discontinuation was lack of efficacy: 9 (25.0%) Gepirone subjects and 10 (29.4%) placebo subjects.

Table 14: Patient Disposition During the Double-Blind Phase (Study 03A7A-002)

	Treatment Group		
	Gepirone 20-90 mg/day N (%)	Placebo N (%)	Total N (%)
All-Subjects-Treated Population	36 (100%)	34 (100%)	70 (100%)
Intent-to-Treat Population	36 (100%)	34 (100%)	70 (100%)
Evaluable Population[a]	33 (91.7%)	25 (73.5%)	58 (82.9%)
Discontinued Treatment by:			
Week 7	6 (16.7%)	3 (8.8%)	9 (12.9%)
Week 8	4 (11.1%)	6 (17.7%)	10 (14.3%)
Week 9	2 (5.6%)	4 (11.8%)	6 (8.6%)
Week 10	1 (2.8%)	2 (5.9%)	3 (4.3%)
Week 11	1 (2.8%)	1 (2.9%)	2 (2.9%)
Week 12	1 (2.8%)	0 (0.0%)	1 (1.4%)
Total Discontinued	15 (41.7%)	16 (47.1%)	31 (44.3%)
Completed Double-Blind Phase	21 (58.3%)	18 (52.9%)	39 (55.7%)

[a]Subjects from the ITT population who had a minimum of two weeks of documented exposure to study medication.
Note: Data for this table were derived from Appendix F 6.1-1 and Supplement to report table in Appendix F.

Efficacy Results

Among the following 6 definitions of relapse, Definition 5 was considered by the sponsor to be the most accurate measure of assessing relapse and was, therefore, used as the primary efficacy parameter.

1. Return to $\geq$ 75% of the (pre-treatment baseline HAMD-17 total score
2. CGI Improvement score of 'No Change' or 'Worse than (pre-treatment) baseline' ($\geq$4).
3. Return to $\geq$ 75% of the (pre-treatment baseline HAMD-17 total score, or discontinuation.
4. CGI Improvement score of 'No Change' or 'Worse than (pre-treatment) baseline' ($\geq$4), or discontinuation.
5. Return to $\geq$ 75% of the (pre-treatment) baseline HAMD-17 total score, or discontinuation due to lack of efficacy.

6. CGI Improvement score of 'No Change' or 'Worse than (pre-treatment) baseline' (≥4), or discontinuation due to lack of efficacy.

The log-rank test was the primary method of analysis.

Table 15 presents the efficacy results with respect to percent of subjects relapsing. Definitions 1, 2, 3, and 5 resulted in statistically significant differences in time to relapse between the Gepirone and placebo groups by Week 12 (P-values: 0.030, 0.023, 0.049 and 0.035, respectively).

Table 15: Sponsor's Relapse Results (Study 03A7A-002)

Relapse Definition	Treatment	Weeks of Study					
		Week 7	Week 8	Week 9	Week 10	Week 11	Week 12
Definition 1	Placebo (N=33) **% =**	15	36	45	55	55	55
	Gep 20-90 (N=35) **% =**	20	23	23	26	26	26
	Log-Rank P-Value =						0.030
Definition 2	Placebo (N=33) **% =**	27	48	55	61	61	61
	Gep 20-90 (N=35) **% =**	26	26	26	31	31	31
	Log-Rank P-Value =						0.023
Definition 3	Placebo (N=33) **% =**	18	45	61	70	73	73
	Gep 20-90 (N=35) **% =**	23	31	40	43	43	46
	Log-Rank P-Value =						0.049
Definition 4	Placebo (N=33) **% =**	30	58	67	73	73	73
	Gep 20-90 (N=36) **% =**	29	37	46	51	51	54
	Log-Rank P-Value =						0.124
Primary Definition 5	**Placebo (N=33) % =**	**15**	**36**	**48**	**58**	**61**	**61**
	Gep 20-90 (N=36) % =	**20**	**26**	**29**	**31**	**31**	**31**
	Log-Rank P-Value =						**0.035**
Definition 6	Placebo (N=33) **% =**	27	48	55	61	61	61
	Gep 20-90 (N=36) **% =**	26	31	34	40	40	40
	Log-Rank P-Value =						0.114

The reviewers added here the efficacy results based on ΔHAMD-17 and CGI-PR **(Table 16)** because they will later discuss that HAMD was the original primary efficacy variable, and CGI-PR results are given for comparative purposes to the other studies.

The results show:

- There was no statistically significant difference between Gepirone 20-90 mg/day and placebo with respect to ΔHAMD-17.

- There was no statistically significant difference between Gepirone 20-90 mg/day and placebo with respect to CGI-PR.

Overall, the sponsor concluded, that the trial provided evidence of the efficacy of Gepirone in the prevention of relapse in this population of depressed outpatients.

Table 16: Sponsor's HAMD-17 and CGI Responders Results (Study 03A7A-002)

Parameter	Treatment	Weeks of Study					
		Week 7	Week 8	Week 9	Week 10	Week 11	Week 12
HAMD-17 Change from Baseline	Placebo N= 34 **Δ =**	11.7	9.6	9.5	9.6	9.4	9.0
	Gep 20-90 N=36 **Δ =**	13.0	12.8	12.9	13.1	12.8	13.0
	P-value =	0.532	0.103	0.111	0.107	0.125	0.070
CGI Percent Responders	Placebo N=33 **% =**	45	36	52	48	45	42
	Gep 20-90 N=36 **% =**	62	58	53	53	50	56
	P-value =	0.183	0.070	0.917	0.724	0.708	0.279

2.5.1.3 Conclusion

The following are the sponsor's conclusions for the 4-Designated-Studies from the study reports and the ISE:

- Studies 134001 showed Gepirone to be highly effective in the treatment of subjects with MDD based on both the primary and secondary outcome measures.

- Study 03A7A-003 showed that Gepirone was highly effective in subjects with major, minor, or intermittent depression and a specified atypical profile.

- Study 03A7C-001-B reached statistical significance in both the low and high dose treatment differences from placebo and was effective in the treatment of subjects with MDD based on both primary and secondary outcome measures. Overall, the results of this study show that Gepirone is an effective and well-tolerated antidepressant.

- Study 03A7A-002 reached statistical significance at the end of the double-blind phase based on the primary definition of relapse. The sponsor concluded, that the positive findings in four of six relapse indices and change from baseline on the HAMD 25 strongly support the efficacy of Gepirone in the prevention of relapse in this population of depressed outpatients. Overall results provide evidence that Gepirone is effective in reducing the acute symptoms of MDD, in the maintenance of symptomatic relief, and in the prevention of relapse.

2.5.2 Reviewers' Analysis, Findings, and Conclusions

2.5.2.1 Statistical Methodologies

The sponsor's ANOVA was per protocol and contained the **Treatment-by-Center interaction in the model**. This methodological approach is acceptable in principle and the reviewers verified the correctness of the sponsor's results based on this method. In addition, the following statistical methods were applied by the reviewers:

- Analysis of Covariance (**ANCOVA**) on the change from baseline in HAMD-17 Total Score (HAMD-17 Total Score at baseline is the covariate). The approach follows the ICH E9 suggestions where the **Treatment-by-Center interaction not included** in the model. The model in SAS representation was:

  Change from Baseline in HAMD-17 = Baseline Treatment Center

  The analyses at each time point used:

  - Last Observation Carried Forward (**LOCF**)
  - Observed Cases (**OC**)

- Robustness analyses using a mixed-effects model with repeated measures on change from baseline in HAMD-17 Total Scores. SAS PROC MIXED was used in these analyses. The results are not presented separately because they were similar to the OC analyses.

- Additional analyses to investigate the significant Treatment-by-Center interaction in Study 03A7C-001-B

- Fisher Exact Test at each time point on CGI responders, using

  - Last Observation Carried Forward (**LOCF**)
  - Observed Cases (**OC**)

- HAMD-17 change from baseline plots

The reviewers decided on the above analyses to investigate the robustness of the findings. The reviewers followed the ICH E9 guidance document, which suggests that the interaction term is added to the model only after the treatment effect has been found significant. If the treatment effect was not significant in the model without the interaction term, one does not go any further. If the treatment effect is significant and the added interaction term is found also to be significant, the consistency of the treatment effect across centers is investigated. The interaction term is tested for statistical significance at α**= 0.10**.

For the Study 03A7A-002, the log-rank test results are presented on the **relapse** data in Section 2.10, because they were not pre-specified in the protocol. HAMD-17 total scores are analyzed to

follow the original protocol and CGI responders results are given for comparison's sake with the other studies.

As the studies suffer from large numbers of dropouts, imputation of the missing values becomes a crucial issue. LOCF assumes non-informative censoring which does not hold for these studies. Mixed-effects models with repeated measures also assume non-informative censoring, but there is no need for the imputation of the missing values in the estimation of the treatment effects. The results of the mixed-effects models are not presented in the tables, but the reviewers will comment on which of the classical approaches they followed closest. For a more detailed discussion on the applicability and limitations of the statistical methods, please refer to Section **2.7 Statistical and Technical Issues**. For patient disposition please consult Tables 8, 10, 12, and 14 under ***2.5.1.2***, the sponsor's ***Analysis and Findings*****.**

2.5.2.2 Analysis and Findings

Study 134001

LOCF and Observed Case analysis results are summarized in **Table 17**.

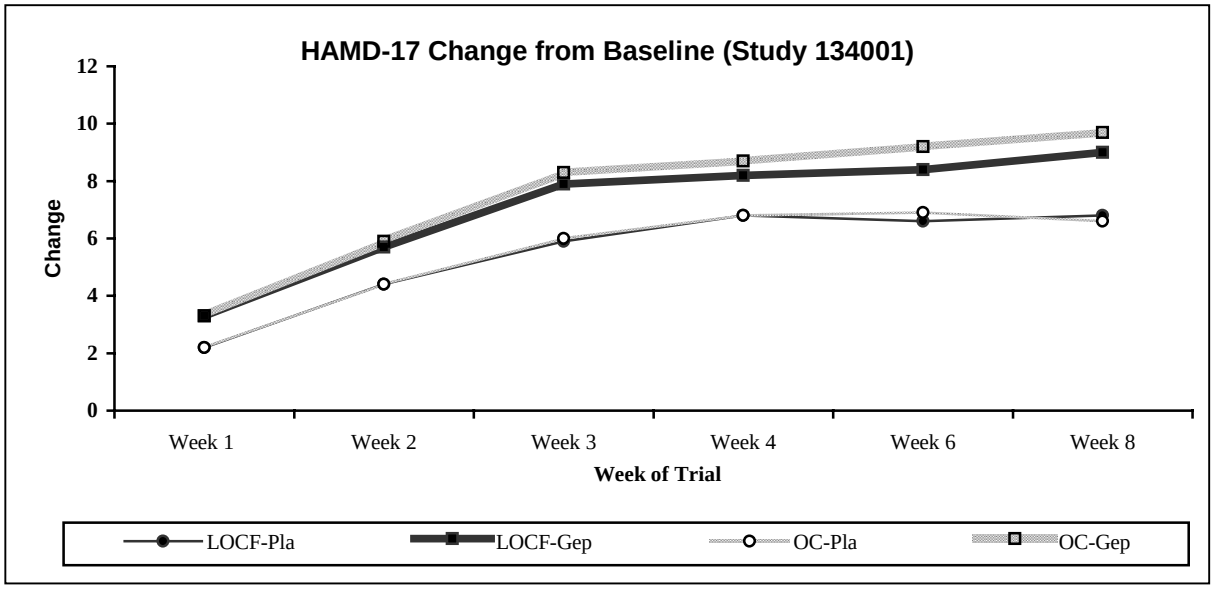
For change from baseline on the HAMD-17 Total score, ANCOVA with LOCF or using OC present a consistent picture of statistically significant superiority of the Gepirone treated group over the placebo treated group. The findings from LOCF ANCOVA above are very close to those from the sponsor's ANOVA analyses. The ICH E9 approach was a mute issue since there was a significant treatment effect but not a significant Treatment-by-Center interaction effect. The mixed-effects models' approach appeared to follow the results of the OC analysis. **Figure 1** below visualizes the higher average scores of the Gepirone treated subjects as compared to the placebo treated group. A placebo response is also apparent. Comparisons on CGI percent responders did not distinguish between the two treatment groups. This measure was not specified as a primary efficacy variable. Therefore, this study showed statistical significance on the primary efficacy parameter.

Table 17: Reviewers' HAMD-17 and CGI Responders Results (Study 134001)

Parameter	Treatment	Weeks of Study					
		Week 1	Week 2	Week 3	Week 4	Week 6	Week 8
Primary Analysis: Last Observation Carried Forward (**LOCF**)							
HAMD-17 Change from Baseline	Placebo **Δ =**	2.2	4.4	5.9	6.8	6.6	6.8
	Gep 10-90 **Δ =**	3.3	5.7	7.9	8.2	8.4	9.0
	P-value =	0.0531	0.0597	0.0128	0.079	0.0519	0.0186
CGI Percent Responders	Placebo **% =**	6.1	13.9	23.8	34.7	36.6	35.6
	Gep 10-90 **% =**	10.2	24.0	33.7	40.6	38.6	43.6
	P-value =	0.310 *[0.288]*	0.074 *[0.067]*	0.161 *[0.121]*	0.468 *[0.385]*	0.885 *[0.772]*	0.314 *[0.251]*
Secondary Analysis: Observed Cases (**OC**)							
HAMD-17 Change from Baseline	Placebo **Δ =**	2.2	4.4	6.0	6.8	6.9	6.6
	Gep 10-90 **Δ =**	3.3	5.9	8.3	8.7	9.2	9.8
	P-value =	0.053	0.0431	0.0071	0.0252	0.0344	0.0080
CGI Percent Responders	Placebo **% =**	6.1	14.3	25.3	35.6	38.6	37.0
	Gep 10-90 **% =**	10.2	25.6	35.6	45.9	43.0	50.7
	P-value =	0.310 *[0.228]*	0.064 *[0.055]*	0.148 *[0.134]*	0.170 *[0.166]*	0.632 *[0.563]*	0.130 *[0.098]*

For CGI responder analysis, the first P-Value is based on the Fisher Exact test; the P-value in [ ] is from the CMH test.

Figure 1: HAMD-17 Change from Baseline by Week (Study 134001)



Study 03A7A-003

LOCF and Observed Case analysis results are summarized in **Table 18**. The various methods of analysis present a consistent picture for change in HAMD-17 Total Scores. Statistical

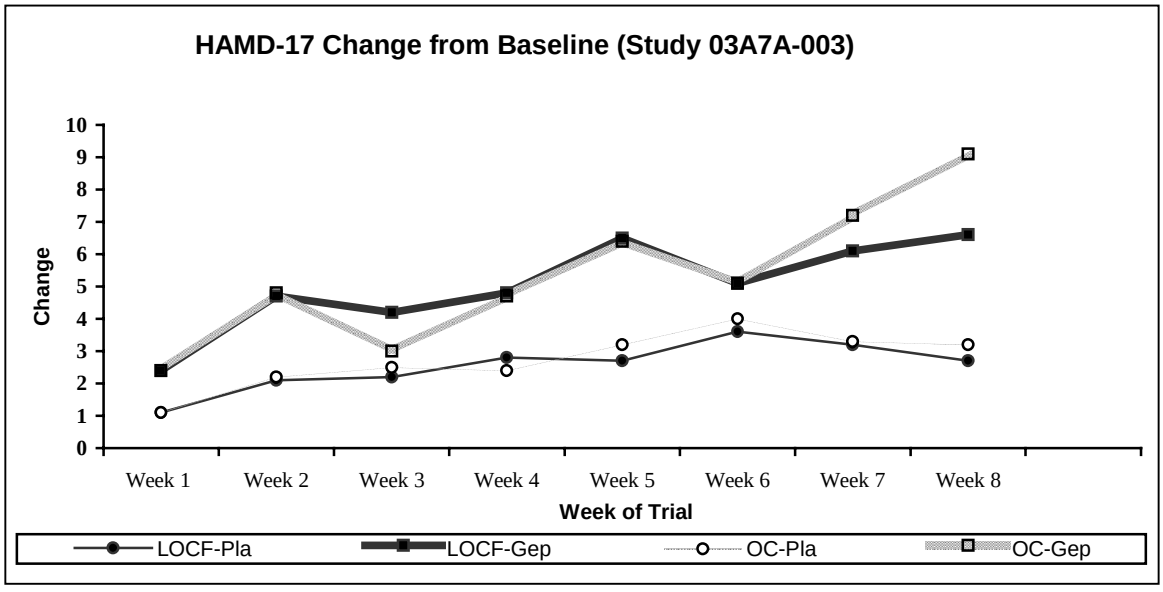
significance was achieved at weeks 2, 5, 7, and 8. These findings corroborate the sponsor's results based on the ANOVA analysis. The mixed-effects models approach seems to follow the classical analyses. As this was a single center trial, there is no possible issue due to interaction. **Figure 2** shows that the numeric superiority of Gepirone over placebo is increasing over time, and that statistical significance at the end of the study may be due to a continued increase from baseline HAMD-17 scores for the treated group compared to an apparent decline in the placebo response. This study used Percent Responders as co-primary efficacy measure. There, the superiority of Gepirone was weaker in terms of level of significance and frequency of reaching statistical significance. Overall, at the end of the study, the primary time point, Gepirone was statistically significantly superior to placebo in all primary analyses.

Table 18: Reviewers' HAMD-17 and CGI Responders Results (Study 03A7A-003)

Parameter	Treatment	Weeks of Study							
		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Primary Analysis: Last Observation Cary Forward (**LOCF**)									
HAMD-17 Change from Baseline	Placebo **Δ=**	1.1	2.1	2.2	2.8	2.7	3.6	3.2	2.7
	Gep 10-90 **Δ=**	2.4	4.7	4.2	4.8	6.5	5.1	6.1	6.6
	P-value =	0.2665	0.0208	0.0943	0.1151	0.0018	0.2696	0.0292	0.0073
CGI Percent Responders	Placebo **% =**	9.1	24.1	23.3	26.7	26.7	36.7	30.0	23.3
	Gep 10-90 **% =**	12.0	26.7	36.7	46.7	60.0	46.7	53.3	56.7
	P-value =	1.00 *[.749]*	1.00 *[.825]*	0.399 *[.264]*	0.180 *[.111]*	0.018 *[.010]*	0.601 *[.436]*	0.115 *[.069]*	0.017 *[.009]*
Secondary Analysis: Observed Case Analysis (**OC**)									
HAMD-17 Change from Baseline	Placebo **Δ=**	1.1	2.2	2.5	2.4	3.2	4.0	3.3	3.2
	Gep 10-90 **Δ=**	2.4	4.8	3.0	4.7	6.4	5.1	7.2	9.1
	P-value =	0.2665	0.0223	0.6791	0.2025	0.0268	0.5642	0.0293	0.0049
CGI Percent Responders	Placebo **% =**	9.1	25.0	26.9	24.0	31.8	38.1	33.3	22.2
	Gep 10-90 **% =**	12.0	27.6	33.3	60.0	63.2	47.1	70.6	75.5
	P-value =	1.00 *[.749]*	1.00 *[.826]*	0.752 *[.636]*	0.042 *[.025]*	0.063 *[.047]*	0.743 *[583]*	0.049 *[.024]*	0.005 *[.002]*

For CGI responder analysis, the first P-Value is based on the Fisher Exact test; the P-value in [ ] is from the CMH test.

Figure 2: HAMD-17 Change from Baseline by Week (Study 03A7A-003)



Study 03A7C-001-B

The sponsor's ANOVA approach with Treatment-by-Center in the model is acceptable for this study as it was for the others. However, the sponsor should have investigated the diverse findings from the three centers further. The reviewers' conclusions do not depend on the fact that they used a different starting model (ICH E9).

The approach suggested in the ICH E9 Guidance would result in a different conclusion for this study, because the model without the interaction term resulted in non-significant treatment effects for the individual drug-placebo comparisons. As **Table 19** shows, only the combined treatment arms reach statistical significance for HAMD scores. However, this comparison was not pre-specified in the protocol. In the supportive analysis of using OC, statistical significance was not achieved at study end. **Figures 3 and 4** show that the average Gepirone responses are not much above the observed placebo responses. The mixed-effects models' approach resulted in the high dose versus placebo comparison reaching statistical significance, but not the low dose comparison. Again, these approaches are presented to explore the robustness of the findings based on the primary analysis and are not consistent. For CGI percent responders, besides the combined treatment, the low-dose Gepirone - placebo comparison reached statistical significance at study end based on one of the statistical methods.

Because the reviewers' methodological approach lead to different conclusions regarding the treatment effect, further investigation of the results is necessary. The sponsor's pre-specified analysis showed a significant Treatment-by-Center interaction. Therefore, the treatment effect needs to be explored for each center. From **Table 20** and **Figure 5** it can be seen that the one small center (Cole) had an unusually low placebo response, which carried the overall significant treatment effect. Since the placebo response from the Cole center appears to be an outlier, the treatment effect was investigated based on **the remaining two centers**. **Table 21** shows that the previous significance of any Gepirone-placebo comparison completely disappeared for either HAMD scores or for CGI percent responders. These results are based on a model with Baseline, Treatment, and Center, since there was no significant interaction between the two remaining centers. These issues are discussed further in Section **2.7 Statistical and Technical Issues**. As will be discussed later, this study as well as the other two BMS studies had interim analyses plans specified in the protocol. If these interim analyses were carried out (the sponsor has no record of whether they were), operational bias could have been introduced which may make the results further suspect and unreliable.

In summary, any treatment effects observed by the sponsor for Study 03A7C-001-B were due to a single small center. Since this center had an unusually low placebo response, one cannot consider its findings as representative. Upon exclusion of this center from the analysis, the data do not distinguish between either Gepirone treatment arm and placebo.

Table 19: Reviewers' HAMD and CGI Responders Results (Study 03A7C-001-B)

Parameter	Treatment	Weeks of Study					
		Week 1	Week 2	Week 3	Week 4	Week 6	Week 8
Primary Analysis: Last Observation Cary Forward (**LOCF**)							
HAMD-17 Change from Baseline	Placebo **Δ** =	3.5	6.3	7.8	7.1	7.7	7.4
	Gep 5-45 **Δ** =	4.7	7.1	8.1	9.3	10.1	10.1
	P-value =	0.4338	0.9416	1.0000	0.1554	0.1314	0.0706
	Gep 10-90 **Δ** =	4.0	6.9	8.4	9.1	10.1	10.1
	P-value =	1.0000	1.0000	1.0000	0.2218	0.1316	0.0796
	Both Gep **Δ** =	4.4	7.0	8.3	9.2	10.1	10.1
	P-value =	0.3216	0.4895	0.6848	0.0516	0.0331	0.0160
CGI Percent Responders	Placebo **%** =	10.9	33.3	36.2	34.8	37.7	39.1
	Gep 5-45 **%** =	18.5	34.8	38.6	48.6	54.3	58.6
	P-value =	0.642 *[0.460]*	1.00 *[1.000]*	1.000 *[1.000]*	0.244 *[0.202]*	0.248 *[0.200]*	0.056 *[0.044]*
	Gep 10-90 **%** =	4.8	31.3	41.8	43.3	50.8	55.2
	P-value =	0.648 *[0.414]*	1.000 *[1.000]*	1.000*[1.000]*	0.760 *[0.622]*	0.334 *[0.252]*	0.170 *[0.122]*
	Both Gep % =	11.8	33.1	40.2	46.0	52.6	56.9
	P-value =	1.00 *[0.859]*	1.00 *[0.972]*	0.651 *[0.587]*	0.137 *[0.125]*	0.055 *[0.044]*	0.018 *[0.016]*
Secondary Analysis: Observed Cases (**OC**)							
HAMD-17 Change from Baseline	Placebo **Δ** =	3.5	6.4	9.0	8.0	9.0	10.4
	Gep 5-45 **Δ** =	4.7	7.4	8.5	9.6	11.3	12.1
	P-value =	0.4338	0.8584	1.0000	0.4628	0.1984	0.4606
	Gep 10-90 **Δ** =	4.0	7.1	10.6	10.2	11.8	13.5
	P-value =	1.0000	1.0000	0.5420	0.2358	0.1124	0.0856
	Both Gep **Δ** =	4.4	7.3	9.5	9.9	11.5	12.7
	P-value =	0.3216	0.4262	0.7011	0.1138	0.0392	0.0762
CGI Percent Responders	Placebo **%** =	10.9	37.9	40.7	42.1	48.0	59.4
	Gep 5-45 **%** =	18.5	37.1	41.5	52.5	60.0	72.5
	P-value =	0.642 *[0.460]*	1.00 *[1.000]*	1.00 *[1.000]*	0.552 *[0.534]*	0.632 *[0.462]*	0.632 *[0.488]*
	Gep 10-90 **%** =	4.8	37.0	57.1	56.3	71.1	77.4
	P-value =	0.648 *[0.414]*	1.00 *[1.000]*	0.244 *[0.180]*	0.346 *[0.302]*	0.072 *[0.046]*	0.354 *[0.254]*
	Both Gep % =	11.8	37.1	49.0	54.1	65.3	74.7
	P-value =	1.00 *[0.859]*	1.00 *[0.912]*	0.329 *[0.308]*	0.191 *[0.142]*	0.052 *[0.045]*	0.163 *[0.120]*

Shaded areas show statistical significance at α=0.05 after Bonferroni adjustment.

Due to multiple testing of Gep 5-45 and Gep 10-90 with the placebo, a Bonferroni adjustment was applied for these two Gep doses by multiplying the resulting P-values from SAS by the factor 2. No adjustment was necessary for the comparison of combined Geps with placebo (α=0.05).

For the CGI percent responders, the first p-value is from Fisher's Exact test, the second [inside] from Cochran-Mantel-Haenszel test.

Figure 3: HAMD-17 Change from Baseline for Low Dose Gepirone by Week (Study 03A7C-001-B)

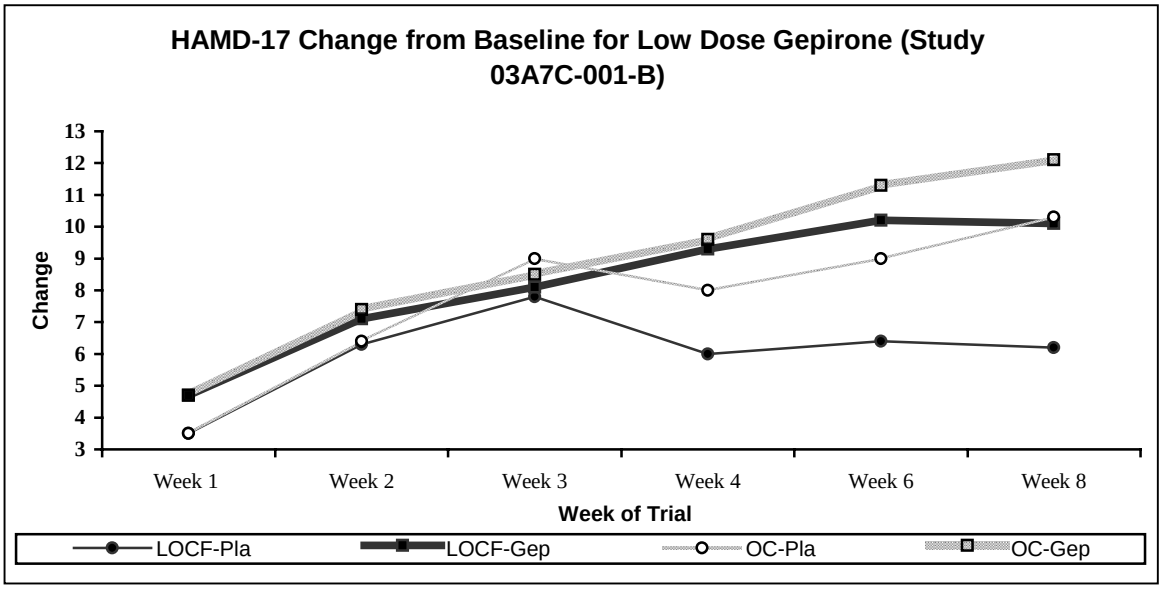


Figure 4: HAMD-17 Change from Baseline for High Dose Gepirone by the Week (Study 03A7C-001-B)

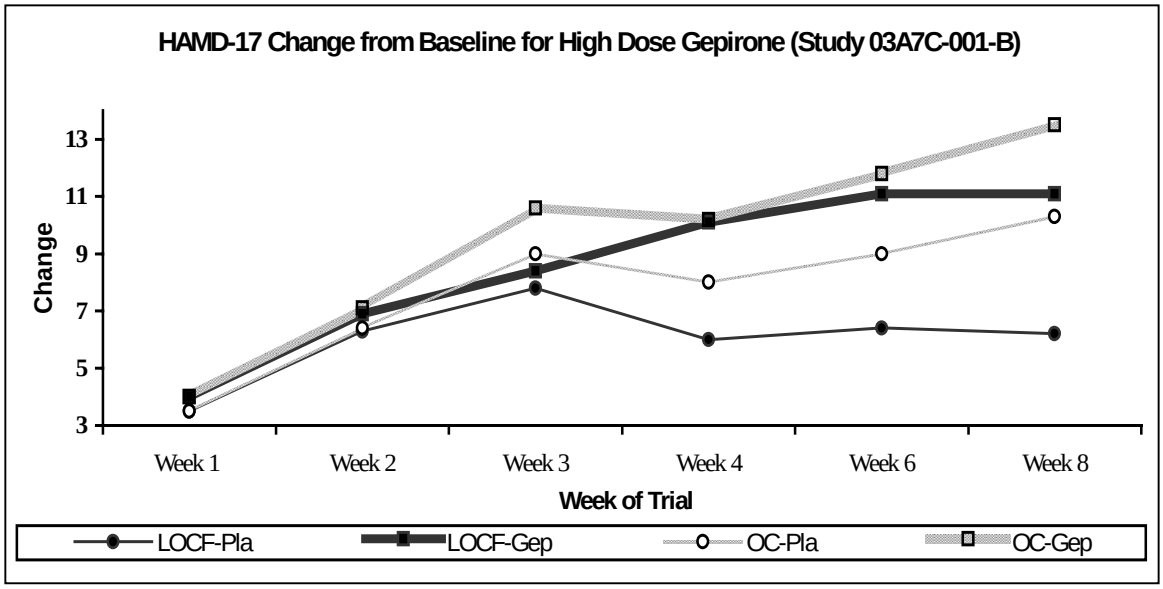


Table 20: ΔHAMD-17 Mean Response by Center (Study 03A7C-001-B)

Investigator	Treatment	N	Week of Trial	
			6	8
Carman	Placebo	30	9.5	8.5
	Gep 5-45	30	11.5	10.5
	Gep 10-90	29	10.8	10.1
Cole	Placebo	9	1.1	1.1
	Gep 5-45	9	9.0	8.2
	Gep 10-90	10	13.4	13.6
Haggerty	Placebo	30	9.2	9.5
	Gep 5-45	31	10.0	11.3
	Gep 10-90	28	9.8	10.2

Figure 5: HAMD-17 Change from Baseline at Week 8 by Center (Study 03A7C-001-B)

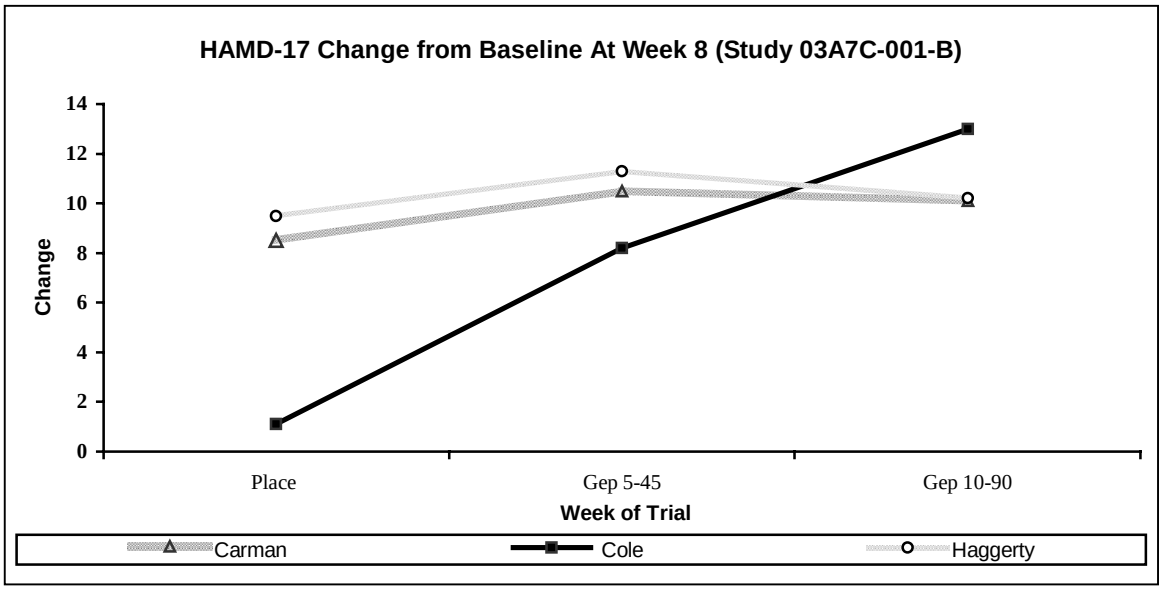


Table 21: Reviewers' Results without the Cole Center (Study 03A7C-001-B)

Parameter	Treatment	Weeks of Study					
		Week 1	Week 2	Week 3	Week 4	Week 6	Week 8
Primary Analysis: Last Observation Cary Forward (**LOCF**)							
HAMD-17 Change from Baseline	Placebo **Δ** =	4.4	7.6	9.3	8.8	9.3	9.0
	Gep 5-45 **Δ** =	5.7	8.1	8.9	10.4	10.9	11.0
	P-value =	0.3770	1.0000	1.000	0.4740	0.5334	0.2774
	Gep 10-90 **Δ** =	3.8	7.6	8.7	9.4	10.2	10.1
	P-value =	1.0000	1.0000	1.0000	1.0000	1.0000	0.8406
	Both Gep **Δ** =	4.8	7.7	8.8	9.9	10.5	10.6
	P-value =	1.0000	1.0000	1.0000	0.6856	0.6222	0.3620
CGI Percent Responders	Placebo **%** =	12.3	38.3	41.7	40.0	43.3	45.0
	Gep 5-45 **%** =	21.1	38.3	44.3	54.1	60.7	63.9
	P-value =	0.632 *[0.486]*	1.00 *[1.00]*	1.000 *[1.000]*	0.292 *[0.244]*	0.140 *[0.118]*	0.090 *[0.074]*
	Gep 10-90 **%** =	3.9	33.3	45.6	45.6	56.1	56.1
	P-value =	0.325 *[0.198]*	1.000 *[1.000]*	1.000 *[1.000]*	1.000 *[1.000]*	0.394 *[0.340]*	0.538 *[0.460]*
	Both Gep % =	12.8	35.9	44.9	50.0	58.5	60.2
	P-value =	1.00 *[0.983]*	0.745 *[0.751]*	0.750 *[0.681]*	0.266 *[0.207]*	0.059 *[0.056]*	0.058 *[0.055]*

Shaded areas show statistical significance at α=0.05 after Bonferroni adjustment.
Due to multiple testing of Gep 5-45 and Gep 10-90 with the placebo, a Bonferroni adjustment was applied for these two Gep doses by multiplying the resulting P-values from SAS by the factor 2. No adjustment was necessary for the comparison of combined Geps with placebo (α=0.05).
For the CGI percent responders, the first p-value is from Fisher's Exact test, the second [inside] from Cochran-Mantel-Haenszel test.

Study 03A7A-002

This study was described by the sponsor as a relapse study with six definitions of relapse specified in the protocol. But in fact, there were no definitions of relapse or corresponding analyses specified in the protocol (for details see Section **3.1 Appendix 1: Issues with Primary Efficacy Parameter for Study 03A7A-002)**. The protocol mentioned a comparison of HAMD scores to the baseline as well as an analysis of the data from the Physician's Questionnaire. The reviewers found that the sponsor had used BMS's study reports, where the first four definitions of relapse had been developed. Organon added two more, apparently fully aware that even the first four definitions had been developed after the blind had been broken. Therefore, the reviewers do not accept any results of the relapse analyses as appropriate

The protocol did not specifically define which HAMD (e.g. -17, -25) measures would be used. The reviewers present the findings from the HAMD-17 measures, which will give comparability to the other studies. The Physicians Questionnaire data were not used by Organon, nor by these reviewers. Baseline in this study is week 6 when the responders of the open label phase were randomized to placebo or Gepirone. **Table 22** shows that change from this baseline in the HAMD-17 Total scores for the Gepirone-treated subjects were not statistically superior to the change observed among the placebo treated subjects. OC and mixed-effects model approaches

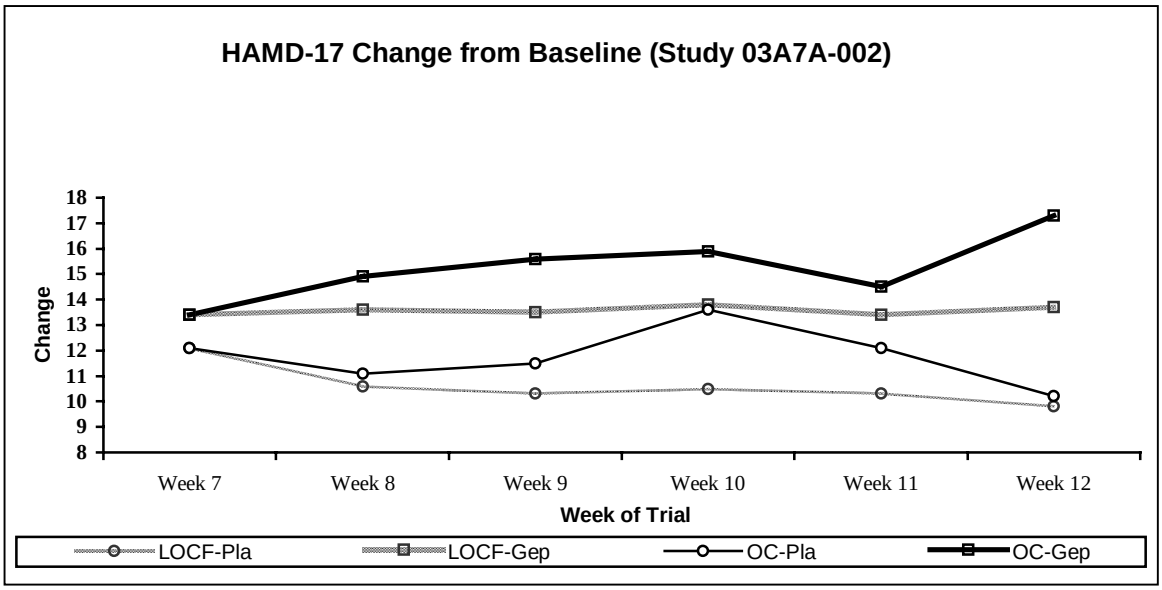
produced similar results and showed a single significant difference at week 8, but not at the end of the double-blind peirod. As **Figure 6** shows, there appears to be a slight decline over the six-week period in the HAMD-17 response among the placebo subjects, whereas the Gepirone subjects seem to be able to maintain their response. However, the averages of the two groups do not differ greatly. Analyzing Percent Responders on the CGI scale (not a pre-specified efficacy parameter) resulted in two statistically significant findings when using OC only. Overall, the analysis of the efficacy parameter specified in the protocol shows only numeric superiority of responders remaining on Gepirone versus responders who were randomized to placebo.

Table 22: Reviewers' HAMD-17 and CGI Responders Results (Study 03A7A-002)

Parameter	Treatment	Weeks of Study					
		Week 7	Week 8	Week 9	Week 10	Week 11	Week 12
Primary Analysis: Last Observation Cary Forward (**LOCF**)							
HAMD-17 Change from Baseline	Placebo **Δ=**	12.1	10.6	10.3	10.5	10.3	9.8
	Gep 20-90 **Δ=**	13.4	13.6	13.5	13.8	13.4	13.7
	P-value=	0.5314	0.0909	0.1204	0.1275	0.1436	0.0840
CGI Percent Responders	Placebo **% =**	45.2	35.3	50.0	47.1	44.1	41.2
	Gep 20-90 **% =**	61.8	58.3	52.8	52.8	50.0	55.6
	P-value =	0.218 *[.183]*	0.061 *[.055]*	1.00 *[.818]*	0.811 *[.635]*	0.641 *[.625]*	0.244 *[.232]*
Secondary Analysis: Observed Case Analysis (**OC**)							
HAMD-17 Change from Baseline	Placebo **Δ =**	12.1	11.1	11.5	13.6	12.1	10.2
	Gep 20-90 **Δ =**	13.4	14.9	15.6	15.9	14.5	17.3
	P-value =	0.5314	0.0290	0.0694	0.3702	0.4605	0.1150
CGI Percent Responders	Placebo **% =**	45.2	37.9	50.0	50.0	33.3	33.3
	Gep 20-90 **% =**	61.8	66.7	70.4	76.2	72.7	100.0
	P-value =	0.218 *[.183]*	0.038 *[.028]*	0.161*[.141]*	0.108 *[.093]*	0.162 *[.126]*	0.006 *[.003]*

For the CGI percent responders, the first p-value is from Fisher's Exact test and the second [inside] is from Cochran-Mantel-Haenszel test.

Figure 6: HAMD-17 Change from Baseline by the Week (Study 03A7A-002)



2.5.2.3 Conclusion

The statistical methods (ANOVA, ANCOVA, use of LOCF or OC, etc.) used either by the sponsor or by the reviewers are standard approaches. However, due to the high dropout rates in these studies, results may be biased for or against efficacy. The analyses, which were pre-specified by the sponsor, were acceptable by the reviewers in principle. The reviewers' robustness analyses supported the significant findings of two of the studies (134001 and 03A7A-003). However, the findings of Study 03A7A-003 may not be reliable due to several crucial issues mentioned again below, all of which weaken the evidence of supporting the ER Study 134001. With respect to the multi-arm study (03A7C-001-B), where the Treatment-by-Center interaction was significant, the reviewers disagree with the sponsor and conclude that the treatment effect is driven by a single center with very unusual placebo response, and on balance, Gepirone was not statistically superior to placebo. The reviewers strongly disagree with the sponsor analyzing Study 03A7A-002 as a relapse study. Relapse or corresponding analyses were **not specified** in the protocol, and were defined after the blind had been broken. The data for one of the two originally pre-specified efficacy measures were submitted (HAMD). Organon had decided not to analyze results from the Physician's Questionnaire. Based on the HAMD results, Gepirone did not reach a statistically significant difference from placebo. Of the four studies promoted by the sponsor as 'proof of efficacy', the reviewers found statistically significant superiority of Gepirone over placebo in one study using the ER dosage form and in one small single-center BMS trial using the IR formulation. This small center suffered from high dropout rates and the results may not be representative of the MDD patient population. Further, the study may have been compromised, if unblinding occurred with the interim analyses specified in the protocol. As noted earlier, 14 additional adequate and well-controlled trials were part of the submission. The reviewers will discuss their views in Section 2.12 as to which of these trials should enter into an overall evaluation of evidence.

From the four studies identified by the sponsor, the single unreplicated study with the to-be-marketed ER dosage form reached statistical significance. The single significant IR study lends very weak support due to the shortcomings just mentioned.

2.6 Findings in Special/Subgroup Populations

The reviewers also performed subgroup analyses on the 4-designated-studies with respect to gender and age. Age was categorized into two classes, ages < 40 years being labeled "Young" and ages $\geq$ 40 years being labeled "Old". No analysis was performed on race because the subjects were predominately white (>72%).

For each subgroup of each study the primary efficacy variable, namely change from baseline in HAMD-17 Total Score (**ΔHAMD-17),** was analyzed as had been done for the whole studies. The findings are summarized in **Tables 23** below.

A more detailed description of the analyses follows:

For each subgroup of Female, Male, Young, and Old, separately, the ΔHAMD-17 data was analyzed using ANCOVA with the LOCF to the last visit (endpoint). The ANCOVA model included baseline, treatment, and center. In SAS representation the model is:

ΔHAMD-17 = BHAMD-17 Treatment Center.

In addition, for the subgroups of each treatment, the raw means of ΔHAMD-17 for the last visit were calculated and are presented in Tables 23.

For Study 03A7C-001-B, the term Treatment-by-Center in the ANCOVA model was significant for each subgroup. Therefore, **Table 24** is provided to present the results when the interaction term was included in the model, namely:

ΔHAMD-17 = BHAMD-17 Treatment Center Treatment*Center.

The reviewers emphasize that subgroup analyses are associated with lack of power, large number of dropouts, presence of interaction, etc., and therefore, the emphasis is on descriptive analyses and the P-values presented in the tables should be interpreted with the caution. The comparison of the raw means and the LS-Means, which resulted from the ANCOVA, may provide some insight on the effects of using LOCF in the presence of high dropout rates.

Table 23 shows that from the 20 subgroup analyses, nine reached statistical significance, namely the Females from the low-dose Gepirone portion of Study 03A7C-001-B, the Males in Study 134001, in Study 03A7A-003, and from the high-dose Gepirone portion of Study 03A7C-001-B, the Young in Study 03A7A-003 as well as in both the low- and high-dose portions of Study 03A7C-001-B, and the Old in the high-dose portion of Study 03A7C-001-B as well as in the relapse Study 03A7A-002. Therefore, each study achieved statistical significance in at least one subgroup, yet no subgroup seems to be consistently sensitive to Gepirone treatment. Raw means based on observed cases at the end of the study and tend to be higher than the means based on the ANCOVA model. The reason for this difference may be due to the model selected and/or due to LOCF.

In **Table 24** the subgroups were re-analyzed for Study 03A7C-001-B leaving the significant treatment-by-center interaction in the model (sponsor's method). With this approach, all subgroups but the Old of the low-dose portion of the study experienced statistically significant superiority of Gepirone treatment over placebo. Again, the raw means based on the small groups of subjects completing the study are substantially higher than the LS means based on the model.
It is noted that the overall results were driven by a single small center (Cole) which appears to be the case in the subgroup analyses as well.

Table 23: Reviewers' Results by Gender and Age Groups

Study	Treatment	Sex		Age Class		Interaction P-Value
		Female	**Male**	**Young** Age< 40 Years	**Old** Age≥ 40 Years	
134001	**Placebo**					T*I →0.7402 T*S→0.5958 T*A→0.8921
	Raw Mean Δ	*6.3, n=37*	*8.6, n=36*	*8.1, n=29*	*7.0, n=44*	
	LS-Mean Δ	5.9, n=55	8.5, n=46	7.2, n=49	6.4, n=52	
	Gepirone 20-80					
	Raw Mean Δ	*9.5, n=49*	*12.9, n=22*	*11.9, n=38*	*9.1, n=33*	
	LS-Mean Δ	8.2, n=67	11.9, n=34	9.7, n=53	8.2, n=45	
	P-Value	0.0681	0.0400	0.0657	0.2344	
03A7A-003	**Placebo**					Single Center T*S→0.4924 T*A→0.9818
	Raw Mean Δ	*1.4, n=7*	*4.5, n=11*	*3.3, n=12*	*3.5, n=6*	
	LS-Mean Δ	2.4, n=11	2.9, n=19	2.4, n=18	3.4, n=12	
	Gepirone 10-90					
	Raw Mean Δ	*8.5, n=6*	*9.2, n=10*	*7.1, n=9*	*11.3, n=7*	
	LS-Mean Δ	6.9, n=10	6.6, n=20	6.6, n=16	6.6, n=14	
	P-Value	0.0680	0.0502	0.0035	0.1562	
03A7C-001-B **without T*I interaction in the model**	**Placebo**					
	Raw Mean Δ	*10.7, n=22*	*10.5, n=11*	*9.1, n=12*	*11.5, n=21*	
	LS-Mean Δ	6.8, n=46	6.7, n=23	5.9, n=27	7.3, n=42	
	Gepirone 5-45				*11.3, n= 22*	
	Raw Mean Δ	*13.7, n= 22*	*10.1, n=18*	*13.0, n=18*	7.9, n=41	
	LS-Mean Δ	9.0, n=38	8.2, 32	9.9, n=29		
	P-Value	0.0008	0.1024	0.0002	0.5680	
	Gepirone 10-90					
	Raw Mean Δ	*15.8, n= 13*	*13.8, n=15*	*14.0, n=15*	*15.2, n=16*	
	LS-Mean Δ	8.0, n=37	8.8, n=30	7.6, n=33	9.2, n=34	
	P-Value	0.1178	0.0112	0.0396	0.0030	
03A7A-002	**Placebo**					T*I➜ 0.5396 T*S➜ 0.6629 T*A➜0.2797
	Raw Mean Δ	*-3.0, n=1*	*13.6, n=5*	*9.0, n= 4*	*14.5, n= 2*	
	LS-Mean Δ	11.3, n=14	9.1, n=20	11.2, n=17	8.0, n=17	
	Gepirone 20-90					
	Raw Mean Δ	16.8, n=8	16.0, n=3	16.0, n= 4	16.9, n=7	
	LS-Mean Δ	15.7, n=17	13.4, n=19	10.8, n=15	14.4, n=21	
	P-Value	0.2432	0.1394	0.9115	0.0484	

A=Age, I=Investigator=Center, S=Sex, T=Treatment,

Table 24: Reviewers' Additional Results by Gender and Age Group (Study 03A7C-001-B)

Study	Treatment	Sex		Age Class		Interactions P-Values n=§
		Female	**Male**	**Young** Age<40 years	**Old** Age> 40 years	
03A7C-001-B with T*I Interaction In the Model	Placebo ***Raw Mean§*** *Δ* LS-Mean Δ	*10.7, n=22* 5.9, n= 46	*10.5, n=11* 5.7, n=23	*9.1, n=12* 4.7, n=27	*11.5, n=21* 6.6, n=42	T*I➜ 0.0588 T*S➜ 0.4924 T*A➜0.9818
	Gep 5-45 ***Raw Mean§*** *Δ* LS-Mean Δ	*13.7, n=22* 9.6, n=38	*10.1, n=18* 8.1, n=32	*13.0, 18* 10.1, n=29	*11.3, n= 22* 7.6, n=41	
	P-Value	0.0002	0.0054	0.0002	0.2758	
	Gep 10-90 ***Raw Mean§*** *Δ* LS-Mean Δ	*15.8, n= 13* 9.1, n=37	*13.8, n=15* 9.3, n=30	*14.0, 15* 8.9, n=33	*15.2, 16* 9.9, n=34	
	P-Value	0.0002	0.0002	0.0002	0.0002	

§: Raw Means and associated sample sizes are based on the observed cases (OC)

2.7 Statistical and Technical Issues

There are nine main statistical issues in this submission which impact on the efficacy conclusion. These issues are elaborated upon below.

1. In the protocols of the BMS studies 03A7A-003 and 03A7C-001-B provision for two interim analyses were made, including which alpha levels to use at each time, as well as at the end. Organon acknowledged in the study reports that these analyses were planned but stated that they have no record that they were actually carried out. The execution of interim analyses is complicated and the preservation of the blind is of utmost importance, as bias can easily be introduced once the blind is broken. Organon apparently cannot vouch for the integrity of the studies and any findings from the BMS studies have to be interpreted with caution. A single interim look was discussed in the protocol of Study 03A7A-002, but the alpha levels for the interim and final analyses were not specified. Again, Organon states that they have no record of whether the interim look was actually carried out by BMS, and validity of the results of this study are also uncertain due to this issue alone. In addition to the concern about the overall validity of the results, the alpha level for claim of statistical significance is no longer 0.05. However, it is the reviewers' impression that this is a minor concern given the observed results.

2. This concern about the validity of the results is heightened by the fact that Organon had reported the definitions of relapse in Study 03A7A-002 as specified in the protocol, which they were not. This study had not been designed as a relapse study by BMS and the pre-specified efficacy parameter did not show statistical significance. A detailed discussion on this issue can be found in Section **3.1 Appendix 1: Issues with Primary Efficacy Parameter for Study 03A7A-002**.

3. Only one of the trials using the to-be-marked Extended Release (ER) formulation of Gepirone was submitted as part of the 'proof of efficacy'. There are additional seven controlled trials with the ER formulation submitted, none of which reached statistical significance on the primary efficacy parameter(s). Therefore, the experience with the ER product has not been replicated.

4. There were three IR trials, which the sponsor had designated to lend support to the ER dosage form. However, only one reached statistical significance. This study was small in size, had high dropout rates, and may have been impaired by operational bias. It does not provide the strength of evidence and level of confidence required to support the efficacy of the ER dosage form.

5. In general, there were additional 14 adequate and well-controlled trials submitted to the NDA, none of which showed a significant treatment effect to support the results of the 4-designated-studies. A detailed discussion of this issue is presented in Section 3.3 **Appendix 3: Evidence from Supporting Studies**. The weight of evidence would be further weakened if any of these studies qualify for consideration in an overall assessment of efficacy.

6. The 4-designated-studies suffered from high rates of discontinuation due to side effects, lack of efficacy, and lost to follow-up. High rates of dropouts can bias statistical findings towards or against significance and can impair the generalizability of the results to the patient population at large. **Table 25** provides the total dropout rates for each study as well as due to lack of efficacy and adverse events. In addition, it is noted that seven of the 14 controlled but not significant studies used high dropout rates as reason for not showing statistical significance. It therefore appears, that the administration of Gepirone may be associated with this limitation.

Table 25: Dropout Pattern in the 4-designated-studies

Number and percent of patients	Study 134001		Study 03A7A-003		Study 03A7C-001-B			Study 03A7A-002	
	Placebo	Gep ER 20-80	Placebo	Gep IR 10-90	Placebo	Gep IR 5-45	Gep IR 10-90	Placebo	Gep IR 10-90
Randomized	106	102	30	30	70	71	70	34	36
Completed	81	74	19	17	32	41	32	18	21
Dropped	25	28	11	13	38	30	38	16	15
Total Dropout %	23.6	27.5	36.7	**43.3**	**54.3**	**42.3**	**54.3**	**47.1**	**41.7**
% Due to Lack of Efficacy	3.8	3.9	20.0	3.3	30.0	11.3	7.1	29.4	25.0
% Due to Adverse Events	2.8	9.6	13.3	26.7	11.4	23.9	40.0	2.9	11.1

As Table 25 shows, the high dropout rates in the placebo arms are mainly due to lack of efficacy, whereas in the Gepirone arms they are due to adverse events. As a result, in addition to the problem of high censoring rates, the censoring patterns in these trials are informative and raise the following concerns.

The reliability of ANOVA treatment estimation with LOCF in the presence of high dropout rates (up to 54% dropouts) needs to be carefully examined. A mixed-effects model with repeated measures estimates the treatment effect differently than does LOCF, but it also

assumes random censoring. The high dropout rates and the informative nature of the censoring cannot be overcome by any currently available statistical methodology, and therefore, the study results may not be generalizable to the MDD patient population.

7. The sponsor's ANOVA approach usually included the Treatment-by-Center interaction term in the model as pre-specified in their protocols. Inclusion of the Treatment-by-Center interaction affords equal weight to each center in the estimation of the treatment effect. The ICH E9 Guideline proposes to estimate the treatment effect first without the interaction in the model. If there is no statistically significant treatment effect at this point, the process should be terminated with a conclusion of no efficacy. In case of a statistically significant treatment effect, the ANOVA model with the Treatment-by-Center interaction should be evaluated. If the interaction is found to be statistically significant, the treatment effect at each center should be explored. The ICH E9 is for consideration and was used by the reviewers as an alternate approach. The sponsor's methodology is classical and was prespecified, and therefore, acceptable in principle.

8. Though the sponsor's analysis of the data for Study 03A7C-001-B was per-protocol and was acceptable, the reviewer cannot accept the conclusion that this study gave support to the efficacy claim. When exploring the treatment effect across centers, it becomes apparent that the overall efficacy is driven by only one of the three centers, which in fact is the smallest center. As was seen in **Table 20,** the placebo response in the Cole center is unusually low (ΔHAMD-17 mean = 1.1 as compared to 8.5 or 9.5 for the other two centers). As a result, the overall apparent efficacy was driven by the results from the Cole center and is mostly due to the low placebo response. **Figure 5** displayed this issue clearly for Week 8. As the Cole center had an unusually low placebo response, the analysis was repeated for the Carman and Haggerty centers without the Cole center. As was seen in **Table 21** there is no statistically significant treatment effect of Gepirone over placebo at any time point for either LOCF or OC analyses. The model used for change from baseline in HAMD-17 Total score contained Baseline, Treatment and Center, because the Treatment-by-Center interaction was non-significant for the remaining centers. It is noted, that the problem with the Treatment-by-Center interaction which affects this study is in addition to the issues resulting from the high dropout rates and the LOCF problem. **Table 26** below is a reproduction of the sponsor's Table 8 from the Study Report. It is an extreme example of small sample sizes and high dropout rates. Seven of the nine subjects (77.8%) randomized to placebo dropped out. No statistical methodology can adequately deal with this problem and any conclusions based on these data cannot represent the patient population at large.

9. The issue as to which is the proper efficacy variable for Study 03A7A-002 will be discussed in detail in Section 3.1 Appendix 1: **Primary Efficacy Parameter for Study 03A7A-002**.

Table 26: Dropout Rates per Center (Study 03A7C-001-B)

Reason for Discontinuation	Org 33062 mg/day		Placebo	Total
	5 to 45	10 to 90		
All Sites Combined	(N = 71)	(N = 70)	(N = 70)	(N = 211)
Adverse Event	17 (23.9%)[b]	28 (40.0%)[c]	8 (11.4%)[d]	53 (25.1%)
Death	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Lack of Efficacy	8 (11.3%)	5 (7.1%)	21 (30.0%)	34 (16.1%)
Lost to Follow-Up	1 (1.4%)	2 (2.9%)	6 (8.6%)	9 (4.3%)
Other[a]	4 (5.6%)	3 (4.3%)	3 (4.3%)	10 (4.7%)
Total Discontinued	30 (42.3%)	38 (54.3%)	38 (54.3%)	106 (50.2%)
Site 2593 (Carman)	(N = 30)	(N = 30)	(N = 30)	(N = 90)
Adverse Event	11 (36.7%)	15 (50.0%)	5 (16.7%)	31 (34.4%)
Death	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Lack of Efficacy	4 (13.3%)	1 (3.3%)	9 (30.0%)	14 (15.6%)
Lost to Follow-Up	1 (3.3%)	1 (3.3%)	4 (13.3%)	6 (6.7%)
Other[a]	1 (3.3%)	0 (0.0%)	1 (3.3%)	2 (2.2%)
Total Discontinued	17 (56.7%)	17 (56.7%)	19 (63.3%)	53 (58.9%)
Site 2483 (Cole)	(N = 10)	(N = 10)	(N = 9)	(N = 29)
Adverse Event	3 (30.0%)	3 (30.0%)	1 (11.1%)	7 (24.1%)
Death	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Lack of Efficacy	0 (0.0%)	2 (20.0%)	6 (66.7%)	8 (27.6%)
Lost to Follow-Up	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Other[a]	1 (10.0%)	0 (0.0%)	0 (0.0%)	1 (3.4%)
Total Discontinued	4 (40.0%)	5 (50.0%)	7 (77.8%)	16 (55.2%)
Site 2485 (Haggerty)	(N = 31)	(N = 30)	(N = 31)	(N = 92)
Adverse Event	3 (9.7%)	10 (33.3%)	2 (6.5%)	15 (16.3%)
Death	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Lack of Efficacy	4 (12.9%)	2 (6.7%)	6 (19.4%)	12 (13.0%)
Lost to Follow-Up	0 (0.0%)	1 (3.3%)	2 (6.5%)	3 (3.3%)
Other[a]	2 (6.5%)	3 (10.0%)	2 (6.5%)	7 (7.6%)
Total Discontinued	9 (29.0%)	16 (53.3%)	12 (38.7%)	37 (40.2%)

[a]Other includes at least one subject with one of the following reasons for discontinuation: patient withdrew consent, patient unreliability, data handling discontinuation, or other known cause.
[b]This number does not include Subject (b) (6) and Subject (b) (6) who discontinued due to AE, because the subjects' final dispositions were listed as "Lack of Efficacy" on th End-of-Study CRF.
[c]The correct number of discontinuations due to AE is 27. This number includes Subject (b) (6) who is also counted (correctly) as a discontinuation from the long-term phase of the study on Table 9.
[d]This number does not include Subject (b) (6) who discontinued due to AE, because the subject's final disposition was listed as "Lack of Efficacy" on th End-of-Study CRF.
Note: Data for this table were derived from Appendices F 6.5-1, 6.5-5, and the Supplement to report tables in Appendix F.

2.8 Statistical Evaluation of Collective Evidence

The four designated efficacy studies suffered from substantial dropout rates (**Table 25**). The studies most affected by dropout rates were 03A7C-001-B and 03A7A-002. Using the LOCF approach in the presence of high dropout rates may bias the results for or against efficacy. Other statistical methods also assume non-informative censoring which did not hold for these trials. Therefore, the results of these studies need to be interpreted with caution. In addition, two studies (03A7A-003 and 03A7A-002) started out with small sample sizes to begin with.

Of the 14 studies selected by the sponsor as not contributing to the evaluation of evidence, the reviewers considered five as having possibly insufficient justification to be excluded. Three of these were Phase III trials conducted with the ER dosage form, and two were conducted with the

IR dosage form. All were well planned and very similar in design, sample sizes, and dropout rates to the 4-Designated-Studies. A detailed discussion of these studies is given in Section **2.12 Appendix of Evidence from Supportive Studies**. All studies were found at best showing numeric superiority of Gepirone, some studies were outright negative (placebo showing a better response than Gepirone). Therefore, if any of these studies enter into the overall evaluation, the evidence of efficacy of Gepirone would be further weakened.

2.9 Conclusions and Recommendations

The sponsor identified four studies as providing evidence of the efficacy of Gepirone ER in the treatment of outpatients with MDD. The four studies were placebo-controlled Phase II (three studies) and Phase III (one study) trials, only one of which used the ER formulation. According to the sponsor, these studies showed statistically significant superiority of Gepirone over placebo at study end on the primary efficacy variable(s).

In the statistical evaluation of this submission, the reviewers evaluated the quality of the four selected studies, the quality of the other 14 adequate and well-controlled trials, the sponsor's methods of analyses, put forth their own methodological approaches, and addressed technical difficulties of the submission.

In the evaluation of the four selected studies, the reviewers found two studies, one ER and one IR, showing statistical superiority of Gepirone over placebo. However, the IR study was judged to be unreliable and not to provide the level of support needed for the single ER study, because of its small sample size, the high dropout rates, and the potential introduction of operational bias. The other two studies did not reach statistical significance. One study had three investigators, one of which showed a significant Gepirone effect but had an extremely low placebo response. The two investigators showed no difference between Gepirone and placebo. The fourth study had been labeled a relapse trial by the sponsor. However, according to the protocol, it was to compare HAMD scores similarly to the other trials and the definitions of relapse and their analyses were made post hoc after the blind had been broken. In this comparison, Gepirone did not reach statistical significance. Therefore, the reviewers conclude, that only one of the four studies identified by the sponsor as 'proof of efficacy' demonstrated the efficacy of Gepirone ER. In addition, one of the three IR studies, though statistically significant, does not have sufficient strength of evidence due to the concern about its validity.

The sponsor had selected four of a total of 18 adequate and well-controlled trials in support of their claim of efficacy. The other 14 studies did not reach statistical significance. The reviewers investigated the validity of the exclusion of these 14 studies. They considered three ER, Phase III, studies to be almost identical in design, sample size, and dropout rates as the four studies put forth in support of the efficacy claim, except that the results did not reach statistical significance (one study was outright negative). Of the IR studies, two trials from a split protocol seemed to qualify for consideration. One trial showed numeric superiority of placebo over Gepirone, the other study showed numeric superiority of Gepirone over placebo, but not statistical significance when adjusted for multiplicity. If the two studies would be combined, as was apparently planned in the original protocol, Gepirone would show very weak numeric superiority over placebo

(though there may be a significant Treatment-by-Center interaction). However, very high dropout rates made the results of these two studies questionable in terms of their representativeness. Therefore, in the reviewers' opinion, at least three ER studies and possibly two IR trials could be considered for inclusion in an overall evaluation of efficacy. The additional trials range from numerically favoring Gepirone to showing clearly negative results.

All studies suffered from fairly large numbers of dropouts. As many of the subjects dropped out due to adverse events or lack of efficacy, the censoring was informative and therefore results from the statistical analyses may be biased (for or against efficacy). Statistical methodology cannot offset this problem, but different approaches may reduce the effects. Both the sponsor's use of LOCF in the ANOVA analyses and the reviewers' mixed-effects models for repeated measures assume random non-informative censoring, which does not hold for these studies. The mixed-effects model uses more information in the estimation of the treatment effect. The results of the mixed-effects models were not presented because the reviewers found that both approaches gave consistent results in the absence of a significant Treatment-by-Center interaction and the overall conclusions did not depend on the methodology used in those cases. However, it is unknown whether the results were biased due to the substantial informative censoring.

The sponsor had prespecified the primary efficacy variable(s), the primary endpoint (end of study), the primary population (ITT), the method of imputation of missing values (LOCF), and the primary method of analysis (ANOVA, usually with interaction). The ICH E9 guidance document suggests that the first analysis should be one without the interaction term in the model, and only if the treatment effect is significant, should the interaction and the consistency of the treatment effect be investigated. The reviewers used the ICH E9 recommendations. For Study 03A7C-001-B, the overall Treatment effect depended on which approach was used. However, more importantly, this significant Treatment effect was solely due to the unusual findings of one small center. Once this center was excluded from the analysis, the treatment effect was clearly non-significant, as was the interaction term.

In their subgroup analyses, the reviewers did not address race because the subjects were predominately white. As the sponsor had not performed any of these analyses, the reviewers chose two major age groups: below 40, and 40 and above. In the gender and age-group analyses of each study, no consistent pattern of benefit for a particular subgroup emerged.

In the reviewers' opinion, only one study (134001) clearly supported the efficacy claim of Gepirone ER. One additional IR study (03A7A-003) reached statistical significance, but the results are considered unreliable due to the issues noted above. The other two of the four studies identified by the sponsor, did not reach statistical significance. If additional studies of either dosage form are allowed to enter into the overall evaluation, the evidence of efficacy is further reduced, since none of the other 14 adequate and well-controlled trials reached statistical significance.

3 APPENDICES

3.1 Appendix 1: Issues with Primary Efficacy Parameter for Study 03A7A-002

The BMS protocol of Study 03A7A-002 calls it a multi-center comparison of Org 33062 (Gepirone) with placebo in MDD. The protocol states under the Synopsis, that the efficacy and safety of Gepirone will be assessed and that there will be two phases during the study. Under Objectives, the protocol states: "The objective of this study is to compare the safety and efficacy of gepirone as compared with placebo in the treatment of depression. The randomized withdrawal design will reduce the exposure of patients to placebo while still enabling a double-blind comparison. It is intended that data from all centers be pooled." When reviewing the protocol, the reviewers found no definitions of relapse, nor any mention of testing differences in relapse rates between Gepirone and placebo, nor any references to appropriate statistical analyses. However, Organon's study reports (individual and ISE) refer to six definitions of relapse **according to the protocol**. After inquiring with the sponsor, they also said that there is no mention of relapse in the protocol. Organon said, that they used BMS's study report as basis, where four definitions of relapse and the analysis were discussed. The original BMS study reports are not part of the electronic submission and the reviewers requested and received pages from the original BMS report and relevant pages of Organon's Clinical Report 03A7A-002 (faxed to the reviewers by the sponsor 2/22/02). On p. 38 of the BMS study report of 10/13/93, BMS discusses four relapse criteria and the survival analyses they had used. There is no mention that these measures were not part of the protocol. On page 56 of Organon's Clinical Report, under '5.12 Changes to Planned Analysis', Organon states the original protocol statistical plan by BMS and then proceeds with "In addition, the following have been added, clarified or changed from the original statistical plan written by BMS, **after the data had been unblinded**." Item 11. of the changes reads (emphasis added): "The **protocol** provided for four **definitions of relapse**. Two additional definitions have been included in the analyses in order to have a more appropriate presentation of subject discontinuations. These definitions are described in Section 5.8.1.1." It was inappropriate of Organon to call the relapse measures primary efficacy variables and to claim that they were specified in the protocol. In addition, defining efficacy measures after the data had been unblinded are open to bias and cannot be accepted in a claim of efficacy. The variables specified in the protocol were HAMD (not further specified) and Physician's Questionnaire. The HAMD results showed no statistically significant difference between Gepirone and placebo (P-value = 0.070 or 0.084 for sponsor's and reviewers' analyses, respectively). The Physician's Questionnaire data were not analyzed by the sponsor and neither by the reviewers.

In case reasons can be found which find the results of the relapse data acceptable, the reviewers performed these analyses as well. **Table 27** gives the total number of subjects who relapsed on each of the six definitions of relapse, as well as the number and percent of subjects not relapsing (i.e. 'censored' in the analysis). The percent of placebo subjects not relapsing during the six weeks ranged from 27.3 to 45.5 percent and from 45.7 to 74.3 percent for the Gepirone group. The measures of quartiles show that at least one half of the placebo patients relapsed, whereas

among the Gepirone subjects, only one criterion (# 4) had more than half relapsing. Q1 - Q3 represent the first, second, and third quartiles of subjects relapsing. For example, for the 5th criterion, at least 25% of the placebo subjects had relapsed by week 8 and at least 50% had relapsed by week 10. Since overall 60.6% (20/33) of the placebo subjects relapsed, the 75% quartile was not reached. Similarly, at least 25 % of the Gepirone subjects had relapsed by week 8; however, since the total number of relapses was 11 (31.4%), the 50% threshold (Q2) was never reached. For neither group was relapse as high as 75 % (Q3). The log-rank test for these data reached statistical significance for four of the six criteria of relapse, including the fifth definition, defined as the primary efficacy parameter by Organon. It is noted, though there were adverse reactions called 'lack of efficacy', the reviewers found that all of them had been classified as relapses.

Table 27: Reviewers' Results of Relapse Data (Study 03A7A-002)

Relapse Criteria	Number of Censored and Uncensored						Central Tendency (**Weeks**)						Log-Rank P-Value
	Placebo			Gepirone			Placebo			Gepirone			
	Failed	Cens	% Cens	Failed	Cens	% Cens	Q1*	Q2	Q3	Q1	Q2	Q3	
1	18	15	45.5	9	26	74.3	8.0	10.0	.	10.0	.	.	0.0296
2	20	13	39.4	11	24	68.6	7.0	9.0	.	7.0	.	.	0.0234
3	24	9	27.3	16	19	54.3	8.0	9.0	.	8.0	.	.	0.0492
4	24	9	27.3	19	16	45.7	7.0	8.0	.	7.0	10.0	.	0.1235
5	20	13	39.4	11	24	68.6	8.0	10.0	.	8.0	.	.	0.0348
6	20	13	39.4	14	21	60.0	7.0	9.9	.	7.0	.	.	0.1141

*Q1, Q2, and Q3 are the first, second (median), and third quartiles

3.2 Appendix 2: Technical Discussions of Statistical Issues

No further technical discussion needed.

3.3 Appendix 3: Evidence from Supporting Studies

Since the sponsor plans to market the ER formulation, the primary evidence should come from trials using this dosage form. There are eight controlled trials in the submission using the ER formulation. Only one study (Study 134001) showed statistical superiority of Gepirone over placebo for the primary efficacy parameter(s) at the prespecified endpoint. The sponsor excluded the remaining seven studies from the evaluation of efficacy for various reasons. **Table 28** lists these seven studies with the time points when the primary efficacy parameter(s) reached statistical significance as well as the main reasons given by the sponsor for not reaching statistical significance.

The sponsor claimed that Gepirone dosing had been inadequate in several (both ER and IR) studies. If dose ranges up to 45 mg/day are considered insufficient to demonstrate efficacy, Studies CN105-057 and CN105-064 had inadequate dosing (though the sponsor does not claim this limitation for study -057). It is noted however, that in both studies the average reduction from baseline in HAMD-17 total scores are lower for the Gepirone treatment arms than for the placebo arms.

Table 28: Summary of Seven Controlled ER Trials

	Week(s) of Significant Difference from Placebo on Primary Efficacy Parameter(s)				Justifications for not Reaching Significance			
Study/ Formulation	Gepirone Low	Gepirone High	Combined Gepirone	Active Control	Gepirone Doses Inadequate	High Placebo Response *	Early Termination (Total Sample Size)	High/Disparate Drop-out Rates
134002 ER		NS				38.5%*		X
CN105-078 ER			NS			28.4%	X (N=146)	
CN105-083 ER			NS			37.1%	X (N=117)	X
CN105-057 ER	Wk 6					38.3%*		X
CN105-053 ER	NS			Wks 2-8		41.7%*	X (N=170)	X
CN105-052 ER		NS		NS		37.2%*	X (N=111)	
CN105-064 ER	NS	NS		NS	X	40.8%*	X (N=93)	X

*Identified by sponsor as high placebo response

The sponsor had identified five of the ER studies as showing a high placebo response. The response was calculated as the average reduction from baseline at study end expressed as a percentage of the average baseline value. In the absence of an active control it is subjective to decide when lack of a treatment effect is due to a high placebo response or when a treatment effect is due to a low placebo response. Studies CN105-052, -053, and -064 had active control arms. For these studies the placebo response ranged between 37 - 42 percent, whereas the active controls had responses in the 44 - 48 percent range (**Table 29**). Though for these studies the case can be made that the placebo response is high, the numeric value of the response cannot be applied to other studies with no active control arm.

Table 29: Response Rates for Active Controlled ER Trials

	CN105-052	CN105-053	CN105-064
Fluoxetine 20-80 mg	43.7%	---	---
Imipramine 50-200 mg	---	47.5%	46.2%
Placebo	41.7%	37.2%	40.8%

Based on the reasoning given above, there seems sufficient justification to exclude studies CN105-052, -053, -057, and -064 from an overall efficacy evaluation.

The reviewers evaluated the design and conduct of Studies 134002, CN105-078, and CN105-083 to decide whether they should be included in an overall evaluation of evidence. They relied on the sponsor's results in this determination.

As the sponsor noted, Studies 134001 and 134002 are both Phase III ER trials conducted by Organon, **identical** in design and conduct. The reader is referred back to the discussion of Study 134001 for details on the study design. The sponsor detailed the similarities and differences in results of these two studies and concluded that both studies provided 'compelling' evidence of efficacy. However, differences in discontinuations, dosing, and placebo response resulted in one study reaching statistical significance whereas the other one failed. The reviewers address each issue:

Table 30: Dropout Pattern for Studies 134001 and 134002

	Study 134001		Study 134002	
Reason for Dropout	Gepirone (N=103)	Placebo (N=106)	Gepirone (N=110)	Placebo (N=108)
AE/SAE	10 9.8%	3 2.8%	11 10%	8 2.8%
Lack of Efficacy	4 3.9%	4 3.8%	3 2.7%	3 2.8%
Other	14 13.7%	18 17.0%	21 19.1%	20 18.5%
Total	28 27.5%	25 23.6%	35 31.8%	31 28.4%

As can be seen from **Table 30**, the dropout rates were somewhat higher in Study 134002. However, they showed the same pattern and were (see Tables 10, 12, and 14) well within what was observed for the other three studies used in support of efficacy.

The dosing in the two studies was identical. However, the sponsor claimed that subjects in Study 134001 were more aggressively titrated and therefore received a higher benefit of Gepirone. If aggressive titration is a necessity to efficacy, this aspect has not been replicated.

The placebo response (final HAMD-17 score expressed as percent HAMD-17 baseline) in Study 134002 was 38.5% as compared to 29.8% for Study 134001. As noted above, in the absence of an active control arm, it is not clear how to judge this magnitude. The other three studies used by the sponsor in support of efficacy had placebo responses ranging from 19.7 - 36.6%. It is the reviewers' opinion that there are no compelling reasons to exclude Study 134002 from the overall evaluation of evidence. As can be seen from **Table 31**, Study 134001 showed clear statistical superiority of Gepirone over placebo especially at the primary endpoint, whereas Study 134002 shows no evidence of an effect.

Table 31: Efficacy Results of Studies 134001 and 134002

Parameter	Treatment	Weeks of Study					
		Week 1	Week 2	Week 3	Week 4	Week 6	Week 8
Primary Analysis: Last Observation Carried Forward (**LOCF**)							
Study 134001 HAMD-17 Change from Baseline	Placebo **Δ =**	2.2	4.4	5.9	6.8	6.6	6.8
	Gep 10-90 **Δ =**	3.3	5.7	7.9	8.2	8.4	9.0
	P-value =	0.0531	0.0597	0.0128	0.079	0.0519	0.0186
Study 134002 HAMD-17 Change from Baseline	Placebo **Δ =**	3.7	6.1	8.0	8.7	9.7	9.3
	Gep 10-90 **Δ =**	4.3	6.7	8.7	9.5	9.9	10.0
	P-value =	0.235	0.375	0.370	0.322	0.841	0.446

Study CN105-078 was a Phase III trial conducted at two sites using a low dose (10-50 mg/day) and high dose (20-100 mg/day) range of Gepirone ER and placebo in outpatients with non-psychotic MDD. The six-week double-blind, randomized, parallel group design was followed by a 20-week long extension for subjects who responded during the short-term phase. Patient characteristics with respect to age, race, and diagnostic criteria, and primary efficacy parameters (HAMD-17 change from baseline and CGI responders) and analyses (combined Gepirone versus placebo) appeared similar to the 4-designated-studies.

Of the 180 planned patients, 146 (81.1 %) were enrolled as described below in **Table 32**. It is noted that Study 03A7C-001-B had also been terminated early by BMS with only 58.6% of the planned sample size.

Table 32: Sample Sizes per Treatment Arm (Study CN105-078)

Data set	Org 33062 ER 10-50 mg/day			Org 33062 ER 20-100 mg/day			Org 33062 ER Total			Placebo			Total
Site →	0001	0002	comb	0001	0002	comb	0001	0002	comb	0001	0002	comb	
Randomized	30	20	50	30	17	47	60	37	97	30	19	49	146
All-Subjects-Treated	30	20	50	29	16	45	59	36	95	30	19	49	144
Intent-to-Treat Population	28	20	48	26	14	40	54	34	88	29	18	47	135
Evaluable Population	26	16	42	20	12	32	46	28	74	28	16	44	118

In the long-term phase, 63 subjects were enrolled: 23 in the Org 33062 ER 10-50 mg group, 19 in the Org 33062 ER 20-100 mg group, and 21 in the placebo group. Data derived from Table 10.

Between 31 (placebo) and 51 (high dose Gepirone) percent of the subjects discontinued during the six-week double-blind phase (**Table 33**). More treated subjects discontinued due to adverse

events than placebo subjects, whereas only one of the treated subjects discontinued due to lack of efficacy versus 10 % of the placebo subjects discontinuing for this reason.

Table 33: Dropout Pattern (Study CN105-078)

Reason for Discontinuation	Org 33062 ER 10-50 mg/day (N=50)	Org 33062 ER 20-100 mg/day (N=45)	Org 33062 ER Total (N=95)	Placebo (N=49)	Total (N=144)
Adverse Event	5 (10.0%)	14 (31.1%)	19 (20.0%)	4 (8.2%)	23 (16.0%)
Death	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Lack of Efficacy	1 (2.0%)	0 (0.0%)	1 (1.1%)	5 (10.2%)	6 (4.2%)
Lost to Follow-up	1 (2.0%)	1 (2.2%)	2 (2.1%)	0 (0.0%)	2 (1.4%)
Discontinued by BMS[a]	4 (8.0%)	0 (0.0%)	4 (4.2%)	1 (2.0%)	5 (3.5%)
Patient Unreliability	2 (4.0%)	3 (6.7%)	5 (5.3%)	1 (2.0%)	6 (4.2%)
Other[b]	4 (8.0%)	5 (11.1%)	9 (9.5%)	4 (8.2%)	13 (9.0%)
Total Discontinued	17 (34.0%)	23 (51.1%)	40 (42.1%)	15 (30.6%)	55 (38.2%)

[a] Subjects discontinued when study was terminated by BMS.

[b] Other includes at least one subject with one of the following reasons for discontinuation: subject withdrew consent, randomized in error, data-handling, or other known cause.

Note: Data for this table were derived from Appendices F 6.5-1 and 6.5-1A and the Supplement to report table in Appendix F.

From the efficacy tables below, it is apparent that Gepirone (combined arms) afforded little superiority over placebo based on change from baseline in the HAMD-17 total score (**Table 34**). Similarly, it showed numeric, but generally not statistically significantly more CGI responders on Gepirone than on placebo (**Table 35**).

Table 34: Sponsor's Results for HAMD17 (Study CN105-078)

Combined sites	Baseline	LS Mean Change from Baseline for HAMD 17 Total Score: LOCF				
Treatment group	N (Mean)	Week 1	Week 2	Week 3	Week 4	Week 6/Endpoint[a]
Org 33062 ER 10-50 mg/day	48 (22.7)	-2.3	-4.9	-6.4	-7.3	-7.5
Org 33062 ER 20-100 mg/day	40 (21.9)	-2.8	-4.8	-5.9	-7.4	-7.5
Org 33062 ER Total	88 (22.3)	-2.5	-4.8	-6.1	-7.2	-7.4
Placebo	47 (22.9)	-3.1	-5.1	-6.5	-6.5	-6.5
Site 0001 (Cohn)	Baseline	LS Mean Change from Baseline for HAMD 17 Total Score: LOCF				
Treatment group	N (Mean)	Week 1	Week 2	Week 3	Week 4	Week 6/Endpoint[a]
Org 33062 ER 10-50 mg/day	28 (22.5)	-1.9	-6.1	-8.4	-9.4	-9.0
Org 33062 ER 20-100 mg/day	26 (21.8)	-2.5	-5.0	-5.0	-5.7	-6.3
Org 33062 ER Total	54 (22.1)	-2.2	-5.6	-6.8	-7.6	-7.7
Placebo	29 (23.1)	-3.2	-5.2	-7.0	-7.8	-7.5
Site 0002 (Ferguson)	Baseline	LS Mean Change from Baseline for HAMD 17 Total Score: LOCF				
Treatment group	N (Mean)	Week 1	Week 2	Week 3	Week 4	Week 6/Endpoint[a]
Org 33062 ER 10-50 mg/day	20 (23.0)	-2.7	-3.7	-4.3	-5.3	-6.0
Org 33062 ER 20-100 mg/day	14 (22.2)	-3.1	-4.6	-6.9	-9.0 [t]	-8.8
Org 33062 ER Total	34 (22.7)	-2.9	-4.1	-5.4	-6.8	-7.1
Placebo	18 (22.7)	-2.9	-4.9	-5.9	-5.3	-5.6

[a] Last visit for evaluable data for the subject.

Note: Data for this table were derived from Appendices 7.1.1-1, 7.1.1-1A, 7.1.1-3, 7.1.1-3A, 7.1.1-4, 7.1.1-4A, 7.1.1-6, and 7.1.1-6A.

[t] $0.05 < p \leq 0.1$ (unadjusted for multiple comparisons)

Table 35: Sponsor's Results for CGI Responders (Study CN105-078)

Combined sites	N (%) of Subjects "Much Improved" or "Very Much Improved": LOCF				
Treatment Group	Week 1	Week 2	Week 3	Week 4	Week 6/Endpoint[a]
Org 33062 ER 10-50 mg/day(N=48)	0 (0%)	11 (23%)	21 (44%)	25 (52%)	24 (50%)
Org 33062 ER 20-100 mg/day(N=40)	4 (11%)	11 (28%)	13 (33%)	17 (43%)	22 (55%)
Org 33062 ER Total (N=88)	4 (5%)	22 (25%)	34 (39%)	42 (48%)	46 (52%)
Placebo(N=47)	1 (2%)	9 (19%)	15 (32%)	17 (36%)	18 (38%)
Site 0001 (Cohn)	N (%) of Subjects "Much Improved" or "Very Much Improved": LOCF				
Treatment Group	Week 1	Week 2	Week 3	Week 4	Week 6/Endpoint[a]
Org 33062 ER 10-50 mg/day(N=28)	0 (0%)	8 (29%)	17 (61%) *	18 (64%)	16 57(%)
Org 33062 ER 20-100 mg/day(N=26)	3 (12%)	7 (27%)	7 (27%)	9 (35%)	13 (50%)
Org 33062 ER Total (N=54)	3 (6%)	15 (28%)	24 (44%)	27 (50%)	29 (54%)
Placebo(N=29)	1 (4%)	5 (17%)	10 (34%)	14 (48%)	13 (45%)
Site 0002 (Ferguson)	N (%) of Subjects "Much Improved" or "Very Much Improved": LOCF				
Treatment Group	Week 1	Week 2	Week 3	Week 4	Week 6/Endpoint[a]
Org 33062 ER 10-50 mg/day(N=20)	0 (0%)	3 (15%)	4 (20%)	7 (35%)	8 (40%)
Org 33062 ER 20-100 mg/day(N=14)	1 (8%)	4 (29%)	6 (43%)	8 (57%) *	9 (64%) *
Org 33062 ER Total (N=34)	1 (3%)	7 (21%)	10 (29%)	15 (44%) *	17 (50%)
Placebo (N=18)	0 (0%)	4 (22%)	5 (28%)	3 (17%)	5 (28%)

[a] Last visit for evaluable data for the subject.
Note: Data for this table were derived from Appendices F 7.1.2-1, 7.1.2-1A, 7.1.2-2, and 7.1.2-2A.
*p-value ≤ 0.05.

It is the reviewers' impression that this study is not very different from the 4-designated-studies with respect to dropout rates or sample size. It appears justified that this study should be included in an overall evaluation of efficacy and as such would have to be counted as giving very weak support.

The design of Study CN105-083 appeared to be identical to Study CN105-078 and is therefore not repeated. It was also a Phase III trial. Of the planned 180 patients, 121 (67.2) were randomized to two study centers. **Tables 36-38** give the samples sizes per treatment arm and the rates of discontinuation.

Table 36: Sample Size per Treatment Arm (Study CN105-083)

Data set	Org 33062 ER 10-50 mg/day			Org 33062 ER 20-100 mg/day			Org 33062 ER Total			Placebo			Total
Site →	0001	0002	**comb**	0001	0002	**comb**	0001	0002	**comb**	0001	0002	**comb**	
Randomized	22	18	**40**	23	17	**40**	45	35	**80**	23	18	**41**	**121**
All-Subjects-Treated	20	17	**37**	22	17	**39**	42	34	**76**	23	18	**41**	**117**
Intent-to-Treat Population	19	17	**36**	21	16	**37**	40	33	**73**	23	16	**39**	**112**
Evaluable Population	18	15	**33**	19	13	**32**	37	28	**65**	19	14	**33**	**98**

In the long-term phase, 44 subjects were enrolled: 14 in the Org 33062 ER 10-50 mg/day group, 15 in the Org 33062 ER 20-100 mg/day group, and 15 in the placebo group.

Table 37: Patient Disposition (Study CN105-083)

Treatment Group	Org 33062 ER			Placebo	Total
	10-50 mg/day	20-100 mg/day	Total		
Randomized Subjects	40	40	80	41	121
All-Subjects-Treated Population	37 (100%)	39 (100%)	76 (100%)	41 (100%)	117 (100%)
Intent-to-Treat Population	36 (97.3%)	37 (94.9%)	73 (96.1%)	39 (95.12%)	112 (95.7%)
Evaluable Population[a]	33 (89.2%)	32 (82.1%)	65 (85.5%)	33 (80.49%)	98 (83.8%)
Discontinued Treatment by:					
Week 1	3 (8.1%)	6 (15.4%)	9 (11.8%)	6 (14.6%)	15 (12.8%)
Week 2	1 (2.7%)	4 (10.3%)	5 (6.6%)	2 (4.9%)	7 (6.0%)
Week 3	5 (13.5%)	2 (5.1%)	7 (9.2%)	5 (12.2%)	12 (10.3%)
Week 4	3 (8.1%)	3 (7.7%)	6 (7.9%)	1 (2.4%)	7 (6.0%)
Week 6	2 (5.4%)	0 (0.0%)	2 (2.6%)	0 (0.0%)	2 (1.7%)
Total Discontinued	14 (37.8%)	15 (38.5%)	29 (38.2%)	14 (34.1%)	43 (36.8%)
Completed Short-term phase	23 (62.2%)	24 (61.5%)	47 (61.8%)	27 (65.9%)	74 (63.2%)

[a]Subjects from the ITT population who had a minimum of two weeks of documented exposure to study medication.
Note: Data for this table were derived from Appendices F 6.1-1, 6.1-5, 6.1-1 A, 6.1.-5A and in the Supplement to report table in Appendix F.

Table 38: Dropout Pattern (Study CN105-083)

Reason for Discontinued	Org 33062 ER		Placebo (N=41)	Total (N=117)
	10-50 mg/day (N=37)	20-100 mg/day (N=39)		
Adverse Event	5 (13.5%)	5 (12.8%)	5 (12.2%)	15 (12.8%)
Death	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Lack of Efficacy	1 (2.7%)	2 (5.1%)	2 (4.9%)	5 (4.3%)
Withdrew Consent	2 (5.4%)	2 (5.1%)	2 (4.9%)	6 (5.1%)
Lost to Follow-up	1 (2.7%)	1 (2.6%)	0 (0.0%)	2 (1.7%)
Discontinued by BMS[a]	2 (5.4%)	3 (7.7%)	5 (12.2%)	10 (8.5%)
Other[b]	3 (8.1%)	2 (5.1%)	0 (0.0%)	5 (4.3%)
Total Discontinued	14 (37.8%)	15 (38.5%)	14 (34.1%)	43 (36.87%)

[a] Subjects discontinued when study was terminated by BMS.
[b] Reason includes improvement, factors related to study design, subject unreliability, randomized in error, data handling discontinuation, and other known cause.
Note: Data for this table were derived from the Supplement to report table in Appendix F; Discrepancies in numbers of discontinuations due to AEs between supplement and Appendix F8.1.1-1/8.1.2-2 come from the fact that two sources were used: SAE forms or End-of-Trial CRFs.

These tables show that the number of subjects discontinuing and the reasons for such are very consistent across the study arms. The rates (34-39%) are comparable to those observed in the 4-designated-studies.

Based on change from baseline in HAMD-17 Total Scores and percent CGI responders, there were no statistically significant differences between either Gepirone arm and placebo at any time-point (**Tables 39 and 40**).

Table 39: Sponsor's Results for HAMD-17 (Study CN105-083)

Treatment Group	Baseline N (Mean)	LS Mean Change from Baseline for HAMD 17 Total Score: LOCF				
		Week 1	Week 2	Week 3	Week 4	Week 6/ Endpoint
Org 33062 ER 10-50 mg/day	36 (24.8)	-4.0	-7.5	-8.6	-10.3	-9.8
Org 33062 ER 20-100 mg/day	37 (23.1)	-4.6	-7.6	-8.4	-9.0	-9.2
Org 33062 ER Total	73 (23.9)	-4.3	-7.5	-8.5	-9.6	-9.4
Placebo	39 (24.0)	-4.3	-7.3	-9.0	-8.7	-8.9

[a] Last visit for evaluable data for the subject.
Note: Data for this table were derived from Appendices F 7.1.1-1, 7.1.1-3, 7.1.1-1A, and 7.1.1-3A.

Table 40: Sponsor's Results for CGI Responders (Study CN105-083)

Treatment Group	N (%) of Subjects "Much Improved" or "Very Much Improved": LOCF				
	Week 1	Week 2	Week 3	Week 4	Week 6/Endpoint[a]
Org 33062 ER 10-50 mg/day (N=36)	2 (6%)	11 (31%)	15 (42%)	18 (50%)	18 (50%)
Org 33062 ER 20-100 mg/day (N=37)	5 (14%)	12 (32%)	14 (38%)	17 (46%)	20 (54%)
Org 33062 ER Total (N=73)	7 (10%)	23 (32%)	29 (40%)	35 (48%)	38 (52%)
Placebo (N=39)	6 (16%)	11 (28%)	15 (38%)	13 (33%)	17 (44%)

[a] Last visit for evaluable data for the subject.
Note: Data for this table were derived from Appendices F 7.1.2-1 and 7.1.2-1A.

The sponsor claimed that though there was no quantitative Treatment-by-Center interaction ($p>0.10$), there was a qualitative one based on the different placebo response at each center. Percent placebo responders were 38 at site Cohn and 48 at site Fieve. At site Cohn, Gepirone treated subjects had a numeric better response than did placebo subjects, whereas at site Fieve the reverse was true. However, it is the reviewers' opinion that this study was similar to the 4-designated-studies, but that it is negative in support of the efficacy of Gepirone in the treatment of MDD in the population studied.

Table 41 gives the mean results of the summary efficacy parameter(s) at study end as defined in the four ER studies which these reviewers consider should enter into the overall evaluation of efficacy. In the reviewers' opinion the four studies selected below are of sufficient quality in design and execution, that they provide evidence of efficacy of Gepirone in the treatment of subjects with MDD. However, the overall evidence is not strong as only one study of the four reaches statistical significance.

Table 41: Efficacy Summary of Four ER Studies

	Treatment	N	HAMD-17 Change	% CGI Responders
134001	Gepirone 20-80mg/day	101	9.0	--
	Placebo	103	6.8	--
	P-value		0.0186	
134002	Gepirone 20-80 mg/day	107	9.96	--
	Placebo	104	9.29	--
	P-value		0.446	
CN105-078	Pooled Gepirone	88	7.4	52%
	Placebo	47	6.5	38%
	P-value		0.451	0.123
CN105-083	Pooled Gepirone	73	9.4	52%
	Placebo	39	8.9	44%
	P-Value		0.743	0.422

The IR formulation had been the earlier formulation and was used in 10 of the 18 well-controlled studies. Three studies with the IR formulation (O3A7A-002, 03A7C-001-B, and 03A7A-002) were reviewed as part of the sponsor's support of efficacy. The other seven studies are listed in **Table 42** with the sponsor's reasons why they should be excluded from the overall evaluation of evidence.

Table 42: Summary of Seven Controlled IR Trials

	Week(s) of Significant Difference from Placebo on Primary Efficacy Parameter(s)				**Justifications for not Reaching Significance**			
Study/ Formulation	Gepirone Low	Gepirone High	Combined Gepirone	Active Control	Gepirone Doses Inadequate	High Placebo Response *	Early Termination (Total Sample Size)	High/Disparate Drop-out Rates
CN105-043 IR	NS				X	41.4%*		
03A7C-001A-2496 IR	NS	NS				39.7%*		X
03A7C-001A-2486 IR	Wk 4	Wks 1, 3, 4, 6, 8				19.7%		X
CN105-037 IR	NS	NS		NS	X	44.3%*		
CN105-022 IR	Wk 2			Wks 2-8	X	32.2%		
CN105-029 IR	NS			Wk 3	X	36.1%	X (N=57)	
CN105-028 IR	NS			Wks 1, 2, 3	X	45.8%*		

* Identified by sponsor as high placebo response

Again, if dose ranges of Gepirone up to 45 mg/day are considered inadequate, Studies CN105-43, -037, -022, -029, and -028 qualify. Studies 03A7C-001A-2496 and -2486, however, had two treatment arms, with the high dose having a range of 10-90 mg/day. It is noted that the sponsor separated a single protocol (03A7C-001A) into these two single center studies. Sample sizes ranged from 38-42 per treatment arm. Entrance criteria and study design were consistent with the other efficacy trials and are not discussed here. **Table 43** below is a partial copy of the sponsor's Table 13 from the ISE.

Table 43: Dropout Rates for Studies 03A7C-001A-2496 and -2486

Study	Treatment	Rando-mized N	Adverse Events		Lack of Efficacy		Termination by BMS		Other		Total Dropouts	
			n	%	n	%	N	%	n	%	n	%
03A7C-001A-2496	Total Org IR	84	32	38.1	14	16.7	0	0	10	11.9	56	66.7
	Org IR 5-45	42	13	31.0	8	19.1	0	0	5	11.9	26	61.9
	Org IR 10-90	42	19	45.2	6	14.3	0	0	5	11.9	30	71.4
	Placebo	42	3	7.1	11	26.2	0	0	4	9.5	18	42.9
03A7C-001A-2486	Total Org IR	87	13	14.9	9	10.3	0	0	36	41.4	58	66.7
	Org IR 5-45	44	5	11.4	4	9.1	0	0	17	38.6	26	59.1
	Org IR 10-90	43	8	18.6	5	11.6	0	0	19	44.2	32	74.4
	Placebo	43	4	9.3	18	37.2	0	0	11	25.6	31	72.1

Excerpt from sponsor's Table 13 in ISE

With such high dropout rates, any efficacy findings from these two studies would be questionable in their applicability to the subject population entering the study, and even more so to the subjects with MDD at large. Further, study - 2496 showed superiority of placebo over each Gepirone treatment arm. Were the findings combined as was planned originally, Gepirone HAMD-17 results would be barely numerically higher than placebo's. These studies do not provide any support towards the efficacy of Gepirone (**Table 44**).

Table 44: Efficacy Summary of Two IR Studies

Study	N	Treatment	HAMD-17	CGI Percent Responders
03A7C-001A-2496	40	Placebo	9.8	46%
	34	Low Dose	7.6 P-value=0.335	36% P-value=0.346
	36	High Dose	6.7 P-value=0.142	29% P-value=0.134
03A7C-001A-2486	36	Placebo	4.9	39%
	38	Low Dose	7.7 P-value=0.150	55% P-value=0.174
	39	High Dose	8.3 P-value=0.064	59% P-value=0.092

P-values are based on Dunnett's multiple comparison test

Overall, none of the additional seven IR trials can be used to support the efficacy claim of IR Gepirone. From the three IR trials submitted by the sponsor as primary support, only one small single-center study shows statistically significant superiority of Gepirone over placebo. Therefore, the collective evidence of efficacy of the Gepirone IR formulation is very weak.

3.4 Appendix 4: Bibliography and References

No published papers were used by the reviewers.

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

/s/

Roswitha Kelly
3/1/02 06:29:58 PM
BIOMETRICS

Kooros Mahjoob
3/2/02 05:58:04 PM
BIOMETRICS

Kun Jin
3/4/02 10:19:41 AM
BIOMETRICS

George Chi
3/4/02 10:29:19 AM
BIOMETRICS

Statistical Review and Evaluation

Review of Mouse Carcinogenicity Data

NDA#: **21-164**

Applicant: **Organon Inc.**

Drug Name: **Gepirone HCl (Org 33062)**

Indication: **Depression**

Documents Received: **Electronic Submission of 05/18/01; Electronic Data Set Submitted 01/09/02**

Pharmacologist: **Linda Fossom, Ph.D. (HFD-120)**

Statistical Reviewer: **Roswitha Kelly, MS (HFD-710)**

Project Manager: **Paul David (HFD-120)**

Cc: Orig. NDA 21-164
HFD-120/Division Files
HFD-120/Mr. David
HFD-120/Dr. Fossom
HFD-120/Dr. Rosloff
HFD-700/Dr. Anello
HFD-710/Dr. Chi
HFD-710/Dr. Jin
HFD-710/Ms. Kelly
This review consists of 9 pages of text and 3 pages of appendix. Gepirone_mouse.doc. 01/15/02

TABLE OF CONTENTS

TABLES

FIGURES

APPENDIX

1.0 Background

The electronic data sets for both the mouse and the rat carcinogenicity studies of the 05/18/01 submission contained errors and inconsistencies and the sponsor was requested to re-submit. The revised rat data were submitted 12/10/01 and have been reviewed (01/14/02). The revised mouse data were received 12/28/01 but were found inadequate (only 85 records submitted). Another mouse data set was submitted on 01/09/02 and is analyzed in this review. This reviewer had to make some modifications to the submitted data as the sponsor did not follow the Guidance for Industry, Providing Regulatory Submissions in Electronic Format, 1999, sufficiently well (e.g. re-group the four causes of death submitted by the sponsor into the three specified in the Guidance document).

2.0 The Mouse Study

2.1 Introduction

Five hundred Charles River CD-1 mice were exposed for two years to gepirone in doses that after repeated adjustments reached 0, 5, 25, and 250 (females) or 350 (males) mg/kg/day. If the doses are averaged over exposure time, they are 0, 5, 25, 224.28 for the females and 0, 5, 24.33, 246.63 mg/kg/day for the males. The control groups consisted of 100 animals per gender; the treated groups were 50 animals per sex. One male animal of the low dose group escaped and is lost to the investigation. All animals had complete histopathologic evaluation. In addition, the examining pathologist classified all tumors of animals dying before terminal sacrifice as fatal or incidental. Tumors observed in terminally sacrificed animals were classified as incidental.

2.2 Sponsor's Findings

Statistical analyses using the methods of Peto and Pike were performed for individual tumor types with at least 2 tumor-bearing animals in the high dose, or with at least 4 tumor-bearing animals in the combined intermediate and high dose groups, on selected combined tumors, on sets of all animals having tumors, and on all animals having malignant tumors. The sponsor formed weekly time intervals for determining observed and expected number of tumors (both for fatal and incidental tumors) and combined these to an overall Chi-square test derived by Tarone. Doses were weighed 0, 1, 2, 3 for control, low, medium, and high dose, respectively.

Mortality analyses showed no increase with dose for either male (p=0.97) or female (p=0.99) mice. Both high dose groups lived longer than the control animals. Using the minimum number of tumor-bearing animals as mentioned above, the following tumors had sufficient numbers of occurrences for statistical analysis: adrenal pheochromocytoma (females), hepatocellular adenomas and carcinomas (males and females), pulmonary adenomas and adenocarcinomas (males and females), pituitary adenomas (females), ocular accessory adenomas (males), uterine endometrial stromal sarcomas, leiomyomas, and leiomyosarcomas (females). In addition, various combinations of tumors were formed. Statistical analysis of the tumor data showed no increases of any individual tumor type or for the overall tumor- or malignancy rates.

2.3 Reviewer's Findings

As noted above, doses were adjusted upward at weeks 7, 12, 19 for the females and at weeks 7, 12, 19, and 79 for the males. Averaging the doses received by the animals over the study resulted in levels of 0, 5, 25, 224 mg/kg/day for the females and 0, 5, 24, 247 mg/kg/day for the males. This reviewer used these levels as scale factors in the trend test analyses.

As can be seen from Tables 1 a/b - 2 a/b and Figures 1 a/b, survival for both the male and female mice was better among the treated than the control animals. This differential reached statistical significance for the males with $p<0.02$ and for the females with $p<0.05$. The sponsor's p-values of >0.97 are consistent with this reviewer's findings, as their test was based on a one-sided trend increasing with dose, whereas this reviewer's test was two-sided.

Table 1a: Number of Animals Dying during Time Interval

Species: Mouse, Sex: Male

	Treatment Group				Total
	CTRL	LOW	MED	HIGH	
	N	N	N	N	N
Week					
0-52	6	2	1	4	13
53-78	13	12	7	1	33
79-91	23	11	8	4	46
92-103	8	4	8	8	28
104-105	50	20	26	33	129
Total	100	49	50	50	249

Table 1b: Dose-Mortality Trend Tests

This test is run using Trend and Homogeneity Analyses of Proportions and Life Table Data

Version 2.1, by Donald G. Thomas, National Cancer Institute

Species: Mouse, Sex: Male

Method	Time-Adjusted Trend Test	Statistic	P Value
Cox	Dose-Mortality Trend	5.40	0.0202
	Depart from Trend	2.50	0.2859
	Homogeneity	7.90	0.0481
Kruskal-Wallis	Dose-Mortality Trend	5.58	0.0182
	Depart from Trend	2.87	0.2381
	Homogeneity	8.45	0.0376

Table 2a: Number of Animals Dying during Time Interval

Species: Mouse, Sex: Female

	Treatment Group				Total
	CTRL	LOW	MED	HIGH	
	N	N	N	N	N
Week					
0-52	10	2	3	1	16
53-78	18	7	12	9	46
79-91	16	6	8	5	35
92-103	18	10	7	6	41
104-105	38	25	20	29	112
Total	100	50	50	50	250

Table 2b: Dose-Mortality Trend Tests

This test is run using Trend and Homogeneity Analyses of Proportions and Life Table Data

Version 2.1, by Donald G. Thomas, National Cancer Institute

Species: Mouse, Sex: Female

Method	Time-Adjusted Trend Test	Statistic	P Value
Cox	Dose-Mortality Trend	3.97	0.0462
	Depart from Trend	2.58	0.2750
	Homogeneity	6.56	0.0875
Kruskal-Wallis	Dose-Mortality Trend	3.65	0.0561
	Depart from Trend	2.91	0.2330
	Homogeneity	6.56	0.0873

Figure 1a: Survival Curves Male Mouse

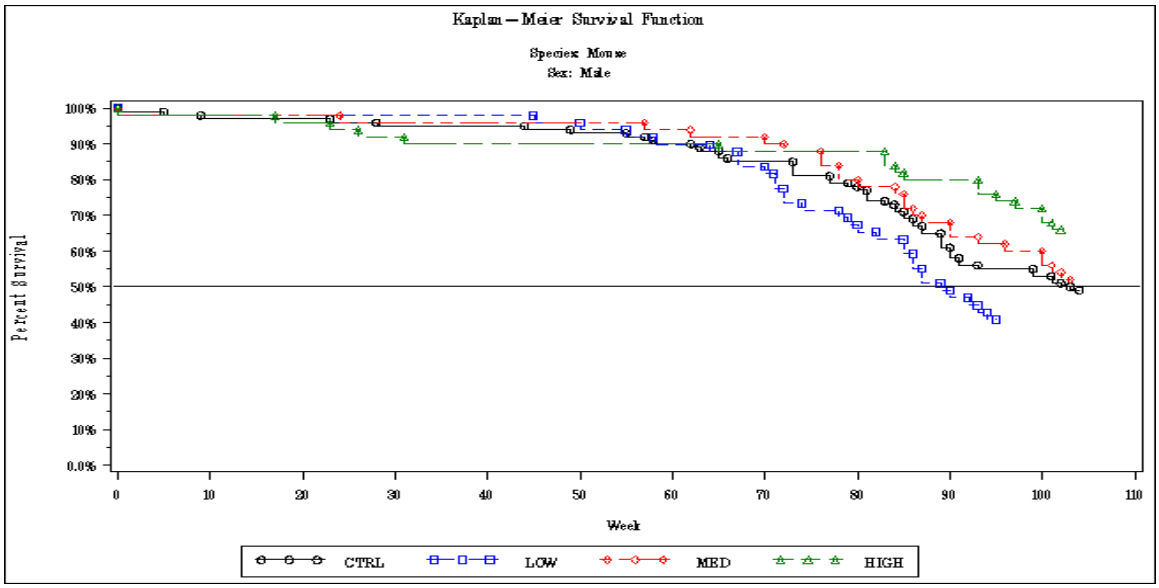
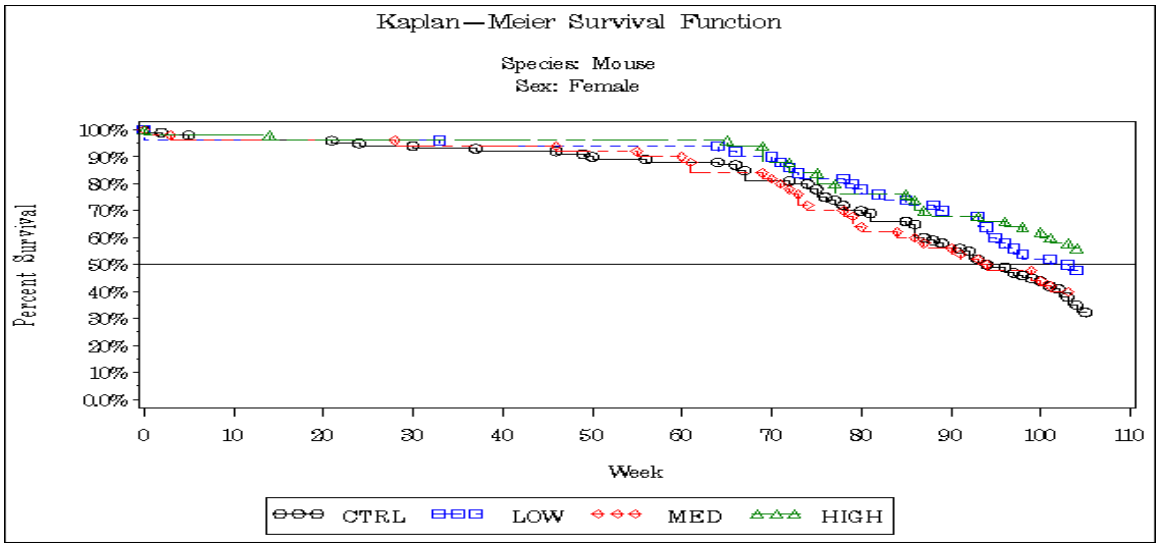


Figure 1b: Survival Curves Female Mouse



The sponsor analyzed tumor findings only if a minimum number (c.f. above) of tumor-bearing animals were observed, whereas all tumor findings were analyzed by this reviewer. Mortality adjusted exact permutation trend tests with increasing dose were computed for tumors observed in the incidental or fatal context. When a tumor presented itself in both contexts within a given time interval, the normal approximation is used and the 'asymptotic' p-value is more appropriate unless the number of tumors is small. In the latter case the true p-value is expected to be bounded by the exact and asymptotic calculations. As there were two two-year carcinogenicity studies in this submission, the

appropriate α-levels for trend tests are 0.025 and 0.005 for rare and common tumors, respectively. Dose groups were weighed by the averaged doses, reflecting most closely the actual exposure of the animals. The sponsor's ordinal (0, 1, 2, 3) weighing should have minimal effect on p-values from exact tests but may affect p-values from asymptotic tests to a greater extend.

All tumor findings are presented in the Appendix. Spot-checking of incidence numbers between the sponsor and this reviewer produced no discrepancies. Neither for the male nor for the female mice were there any statistically significant increases in tumors with dose.

3.0 Validity of the Mouse Study

Survival was good for all groups of either gender, particularly for the high dose animals, which had the best survival (66% for males and 58% for females at week 104). Therefore, it is concluded that a sufficient number of animals of either gender lived long enough. As survival was better in the high dose groups than in the controls, this finding does not contribute to assessing whether the high dose presented a sufficient tumor challenge. The sponsor's Table 2 and their Figure 1 (reproduced below as Figures 2a/b) showed little difference between mean body weights of the male control and high dose animals. For the first six weeks there was no difference at all in mean body weights. With the first adjustment of the high dose there was a slight reduction in mean body weights of the HD males, which remained around 3% for most of the study. The greatest differential (5-6%) was seen roughly between weeks 30 and 65, after which time the difference shrunk again to less than 3% for the remainder of the study. This modest reduction in mean body weights of the high dose male mice may not be sufficiently large to suggest that the high dose was close to the MTD. It is noted again, that the high dose was repeatedly increased from the original 50 mg/kg/day to a final 350 mg/kg/day, which resulted in an average dose of 247 mg/kg/day. The difference in mean body weights between the female control and high dose animals was clearer, starting with about 4% early in the study and reaching differentials as high as 12%. Therefore, based on mean body weight data, the high dose (averaged to 224 mg/kg/day) appears to have been close to the MTD for the female mice.

Figures 2a/b (Sponsor's Figure 1): Group Mean Body Weights, Male/Female Mouse

Figure 1. Mean Body Weight
Male

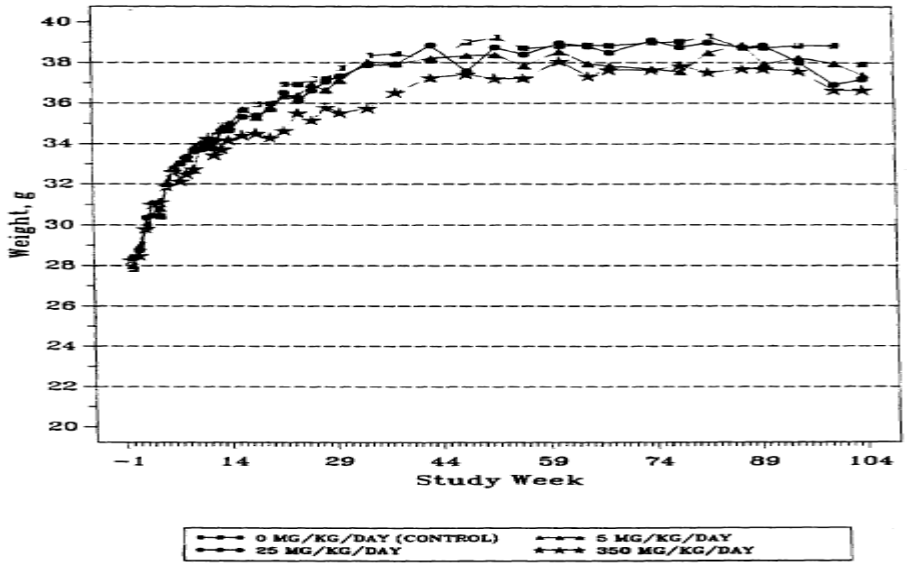
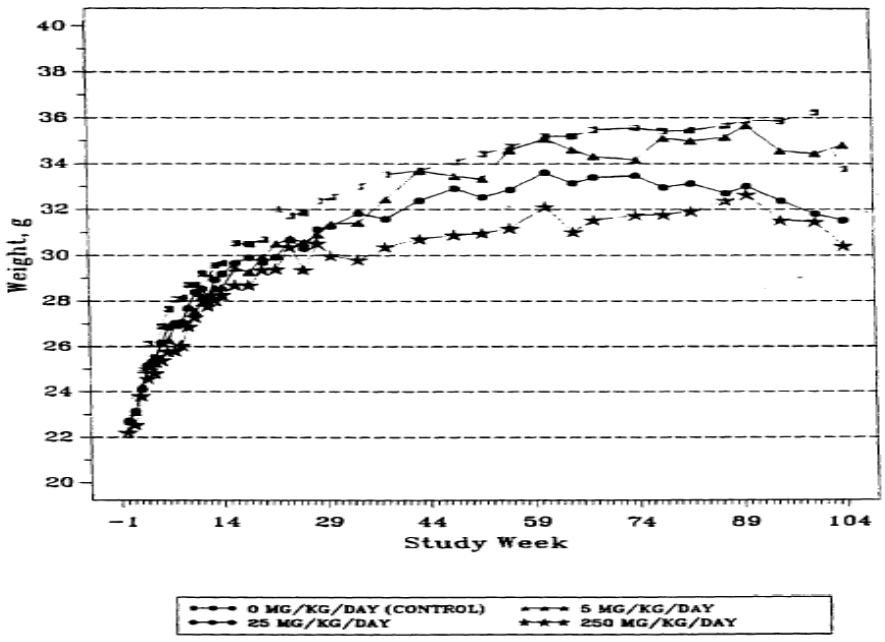


Figure 1. Cont. Mean Body Weight
Female



4.0 Summary

In this two-year study of Charles River CD-1 mice, gepirone HCl was available in the diet at levels of 0, 5, 25, and 250 (females) or 350 (males) mg/kg/day after several upward adjustments. When averaged over the adjustment periods, doses were 0, 5, 25, 224 mg/kg/day for the females and 0, 5, 24, 247 mg/kg/day for the males. Survival was good, especially among the high-dose animals of either sex. This reviewer agrees with the sponsor's conclusion that no statistically significant increases in tumors were observed in either gender. As there were no statistically significant tumor findings in either gender, the validity of both study sections was investigated. Based on the reduction (4-12%) of mean body weights in the high dose females, this study can be considered valid in a sense of exposing a sufficient number of animals for a sufficient length of time at a dose which appears to be close to the MTD. For the male mice the study was valid in the sense of exposing a sufficient number of animals for a sufficient length of time. However, the reduction in mean body weights of the HD animals compared to the controls was slight (0-6%), which seems to indicate that the HD was not close to the MTD for the male mice.

APPENDIX

Test for Dose-Tumor Positive Linear Trend

Source: Male Mouse Data

Organ Code	Organ Name	Tumor Code	Tumor Name	CTRL	LOW	MED	HIGH	pValue (Exact)	pValue (Asymp)	Natural Tumor # in control group	Natural Rate (in ctrl group)	Tumor type
1100	tongue	110013	papilloma	1	0	0	0	1.0000	0.7430	1	1%	IN
1800	liver	180001	adenoma	17	6	7	5	0.9201	0.9136	17	17%	IN
1800	liver	180002	carcinoma	10	2	4	4	0.7000	0.6966	10	10%	MX
1800	liver	180003	hemangioma	1	0	1	0	0.7182	0.7826	1	1%	MX
1800	liver	180004	hemangiosarcoma	1	3	5	2	0.5557	0.6351	1	1%	MX
2100	kidneys	210001	adenoma	0	1	0	0	0.6061	0.5929	0	.0%	IN
2100	kidneys	210006	adenocarcinoma	0	0	1	0	0.4531	0.6635	0	.0%	IN
2500	testes	250003	hemangioma	0	0	1	0	0.4574	0.6671	0	.0%	IN
2500	testes	250009	interstitial cell	1	1	1	0	0.7522	0.8383	1	1%	IN
3100	penis	310006	adenocarcinoma	0	0	0	1	0.2553	0.0449	0	.0%	FA
4400	adrenal glands	440001	adenoma	1	1	1	1	0.4250	0.4967	1	1%	IN
4400	adrenal glands	440015	ganglioneuroma	1	0	0	0	1.0000	0.6582	1	1%	IN
4600	spleen	460004	hemangiosarcoma	2	0	2	1	0.5042	0.5856	2	2%	MX
5000	thymus	500007	leiomyoma	1	0	0	0	1.0000	0.6994	1	1%	IN
5500	lacrimal gland	550001	adenoma	7	2	8	2	0.8556	0.8747	7	7%	IN
5700	skin	570010	lipoma	1	0	0	0	1.0000	0.7457	1	1%	IN
5900	bone	590021	chondroma	0	1	0	0	0.6124	0.7301	0	.0%	IN
5900	bone	590025	osteosarcoma	0	0	0	1	0.1983	0.0229	0	.0%	FA
6600	ear	660003	hemangioma	1	0	0	0	1.0000	0.7457	1	1%	IN
7400	tail	740004	hemangiosarcoma	0	0	0	1	0.2558	0.0451	0	.0%	IN
800	generalized	4500	lymphosarcoma	0	1	0	0	0.6094	0.7272	0	.0%	IN
8000	generalized	4500	lymphosarcoma	9	1	3	3	0.7305	0.7172	9	9%	MX
900	lung	90001	adenoma	29	16	10	11	0.9185	0.9134	29	29%	IN
900	lung	90006	adenocarcinoma	3	3	1	1	0.8047	0.7972	3	3%	MX
900	lung	90012	mesothelioma	1	0	0	0	1.0000	0.7326	1	1%	FA

Test for Dose-Tumor Positive Linear Trend

Source: Female Mouse Data

Organ Name	Organ Code	Tumor Name	Tumor Code	Natural Rate (in ctrl group)	CTRL	LOW	MED	HIGH	Tumor type	pValue (Exact)	pValue (Asymp)
brain	100	oligodendroglioma	10014	1%	1	0	0	0	FA	1.0000	0.7495
brain	100	astrocytoma	10019	1%	1	0	0	0	IN	1.0000	0.7495
liver	1800	adenoma	180001	3%	3	1	0	1	IN	0.7267	0.6246
liver	1800	carcinoma	180002	1%	1	0	0	1	IN	0.4525	0.2362
liver	1800	hemangioma	180003	1%	1	0	0	0	IN	1.0000	0.7495
liver	1800	hemangiosarcoma	180004	2%	2	1	0	2	FA	0.3042	0.2130
pancreas	2000	adenoma	200001	1%	1	0	1	0	IN	0.6858	0.7781
kidneys	2100	carcinoma	210002	.0%	0	0	0	1	IN	0.2589	0.0465
kidneys	2100	pheochromocytoma	210023	1%	1	0	0	0	IN	1.0000	0.7495
urinary bladder	2300	hemangiosarcoma	230004	.0%	0	0	1	0	FA	0.4155	0.6455
ovaries	3200	adenoma	320001	.0%	0	1	0	1	IN	0.2755	0.2278
ovaries	3200	hemangioma	320003	3%	3	0	0	0	MX	1.0000	0.8548
ovaries	3200	hemangiosarcoma	320004	1%	1	0	0	0	FA	1.0000	0.7471
ovaries	3200	leiomyoma	320007	.0%	0	0	1	0	IN	0.4414	0.6609
ovaries	3200	granulosa cell tum	320018	.0%	0	1	0	0	FA	0.6141	0.7106
uterus	3400	adenoma	340001	1%	1	0	0	0	IN	1.0000	0.7495
uterus	3400	hemangioma	340003	4%	4	2	0	2	MX	0.5418	0.4761
uterus	3400	leiomyoma	340007	4%	4	1	2	1	IN	0.7064	0.7360
uterus	3400	endometrial stroma	340008	3%	3	0	2	2	FA	0.3091	0.3454
uterus	3400	leiomyosarcoma	340024	.0%	0	0	0	1	FA	0.2165	0.0296
uterus	3400	Polyp	340038	1%	1	1	0	0	IN	0.8869	0.8199
cervix	3401	leiomyoma	340107	2%	2	0	1	0	IN	0.8292	0.8330
vagina	3500	leiomyoma	350007	1%	1	1	0	0	IN	0.8869	0.8199
pituitary gland	4100	adenoma	410001	4%	4	0	0	2	IN	0.3906	0.2906
adrenal glands	4400	adenoma	440001	.0%	0	1	0	0	IN	0.6607	0.7325
adrenal glands	4400	pheochromocytoma	440023	.0%	0	1	0	2	IN	0.0808	0.0425
spleen	4600	hemangioma	460003	.0%	0	1	1	1	IN	0.2517	0.3505
spleen	4600	hemangiosarcoma	460004	1%	1	1	0	1	FA	0.4446	0.3562
mesenteric lymph	5104	hemangioma	510403	1%	1	0	0	0	IN	1.0000	0.7558
lacrimal gland	5500	adenoma	550001	5%	5	2	1	1	IN	0.8766	0.8371
mammary gland	5600	adenocarcinoma	560006	4%	4	2	0	1	MX	0.8225	0.7537
mammary gland	5600	fibroadenoma	560011	1%	1	0	0	0	IN	1.0000	0.7447
skin	5700	carcinoma	570002	1%	1	0	0	1	MX	0.4212	0.2108
skin	5700	hemangiosarcoma	570004	.0%	0	1	0	0	FA	0.6273	0.7105
skin	5700	sarcoma	570005	.0%	0	1	0	0	FA	0.6607	0.7325

skin	5700	trichoepithelioma	570016	1%	1	1	0	0	IN	0.8134	0.7474
skin	5700	keratoacanthoma	570020	1%	1	0	0	0	IN	1.0000	0.7495
skin	5700	neurofibrosarcoma	570022	.0%	0	1	0	0	IN	0.5429	0.6723
skeletal muscle	5800	hemangiosarcoma	580004	.0%	0	0	0	1	FA	0.2178	0.0300
generalized	8000	lymphosarcoma	4500	15%	15	8	7	1	MX	0.9989	0.9967
generalized	8000	fibrous histiocyto	800017	.0%	0	0	0	1	FA	0.2189	0.0304
lung	900	adenoma	90001	18%	18	15	11	7	MX	0.9448	0.9445
lung	900	adenocarcinoma	90006	4%	4	2	2	1	MX	0.8012	0.8166

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

/s/

Roswitha Kelly
1/16/02 10:33:01 AM
BIOMETRICS

George Chi
1/25/02 10:30:59 AM
BIOMETRICS

Statistical Review and Evaluation

Review of Rat Carcinogenicity Data

NDA#: **21-164**

Applicant: **Organon Inc.**

Drug Name: **Gepirone HCl (Org 33062)**

Indication: **Depression**

Documents Received: **Electronic Submission of 05/18/01; Electronic Data Set Submitted 12/10/01**

Pharmacologist: **Linda Fossom, Ph.D. (HFD-120)**

Statistical Reviewer: **Roswitha Kelly, MS (HFD-710)**

Project Manager: **Paul David (HFD-120)**

Cc: Orig. NDA 21-164
HFD-120/Division Files
HFD-120/Mr. David
HFD-120/Dr. Fossom
HFD-120/Dr. Rosloff
HFD-700/Dr. Anello
HFD-710/Dr. Chi
HFD-710/Dr. Jin
HFD-710/Ms. Kelly
This review consists of 9 pages of text and 4 pages of appendix. Gepirone_rat.doc. 01/11/02

TABLE OF CONTENTS

TABLES

FIGURES

APPENDIX

1.0 Background

The electronic data sets for both the mouse and the rat carcinogenicity studies of the 05/18/01 submission contained errors and inconsistencies and the sponsor was requested to re-submit. On Dec. 10, 2001 the revised rat data were received which this review addresses. The revised mouse data were received 12/28/01 but were found inadequate (only 85 records submitted). Therefore, there will be a separate statistical review of the mouse carcinogenicity study when the appropriate data are received.

2.0 The Rat Study

2.1 Introduction

Five hundred rats were exposed for two years to gepirone in doses that after adjustment reached 0, 4, 16, and 48 mg/kg/day. The control groups consisted of 100 animals per gender; the treated groups were 50 animals per sex. All animals had complete histopathologic evaluation. In addition, the examining pathologist classified all tumors of animals dying before terminal sacrifice as fatal or incidental. Tumors observed in the terminally sacrificed animals were classified as incidental.

2.2 Sponsor's Findings

Statistical analyses using the methods of Peto and Pike were performed for individual tumor types with at least 2 tumor-bearing animals in the high dose, or at least 4 tumor-bearing animals in the combined intermediate and high dose groups, on selected combined tumors, on sets of all animals having tumors, and on all animals having malignant tumors. The sponsor formed weekly time intervals for determining observed and expected number of tumors (both for fatal and incidental tumors) and combined these to an overall Chi-square test derived by Tarone. Doses were weighed 0, 1, 2, 3 for control, low, medium, and high dose, respectively.

Mortality analyses showed no increase with dose for either male ($p=0.93$) or female ($p>0.99$) rats. The female high-dose animals lived longer than the control animals. Using the minimum number of tumor-bearing animals as mentioned above, the following tumors had sufficient numbers of occurrences for statistical analysis: adrenal pheochromocytoma (males), adrenal cortical adenomas (females), hepatocellular adenomas (females), mammary adenocarcinomas and fibroadenomas (females), pituitary adenomas (males and females), pancreatic islet cell adenomas (males), skin papillomas, squamous cell carcinomas, and keratoacanthomas (males), subcutis fibromas and lipomas (males), testicular interstitial cell adenomas (males), thyroid C-cell adenomas and follicular cell adenomas (males), and uterine polyps (females). In addition, various combinations of tumors were formed. Statistical analysis of the tumor data showed an increase for incidental testicular interstitial cell adenomas (no fatal occurrences) with $p=0.04$ and for incidental hemangiomas in males (no fatal occurrences) with $p=0.02$.

There were no statistically significant differences for the overall tumor- or malignancy rates.

2.3 Reviewer's Findings

As can be seen from Tables 1 a/b - 2 a/b and Figures 1 a/b, survival for both the male and female rats was better among the treated than the control animals. This differential reached statistical significance for the females ($p<0.002$). The sponsor's p-value of >0.99 is consistent with this reviewer's findings, as their test was based on a one-sided trend increasing with dose, whereas this reviewer's test was two-sided.

Table 1a: Number of Animals Dying during Time Interval

Species: Rat, Sex: Male

	Treatment Group				Total
	CTRL	LOW	MED	HIGH	
	N	N	N	N	N
Week 0-52	2	3	.	2	7
53-78	13	4	2	.	19
79-91	8	5	9	5	27
92-104	19	4	6	8	37
105-106	58	34	33	35	160
Total	100	50	50	50	250

Table 1b: Dose-Mortality Trend Tests

This test is run using Trend and Homogeneity Analyses of Proportions and Life Table Data

Version 2.1

by Donald G. Thomas, National Cancer Institute

Species: Rat, Sex: Male

Method	Time-Adjusted Trend Test	Statistic	P Value
Cox	Dose-Mortality Trend	1.66	0.1970
	Depart from Trend	1.08	0.5813
	Homogeneity	2.75	0.4319
Kruskal-Wallis	Dose-Mortality Trend	1.88	0.1701
	Depart from Trend	0.82	0.6624
	Homogeneity	2.71	0.4392

Table 2a:Number of Animals Dying During Time Interval

Species: Rat, Sex: Female

	Treatment Group				Total
	CTRL	LOW	MED	HIGH	
	N	N	N	N	N
Week					
0-52	3	1	2	.	6
53-78	10	7	4	2	23
79-91	21	9	6	4	40
92-104	21	6	6	8	41
105-106	45	27	32	36	140
Total	100	50	50	50	250

Table 2b: Dose-Mortality Trend Tests

This test is run using Trend and Homogeneity Analyses of Proportions and Life Table Data

Version 2.1

by Donald G. Thomas, National Cancer Institute

Species: Rat, Sex: Female

Method	Time-Adjusted Trend Test	Statistic	P Value
Cox	Dose-Mortality Trend	9.95	0.0016
	Depart from Trend	1.66	0.4364
	Homogeneity	11.61	0.0088
Kruskal-Wallis	Dose-Mortality Trend	10.39	0.0013
	Depart from Trend	1.09	0.5794
	Homogeneity	11.48	0.0094

Figure 1a : Survival Curves Male Rat

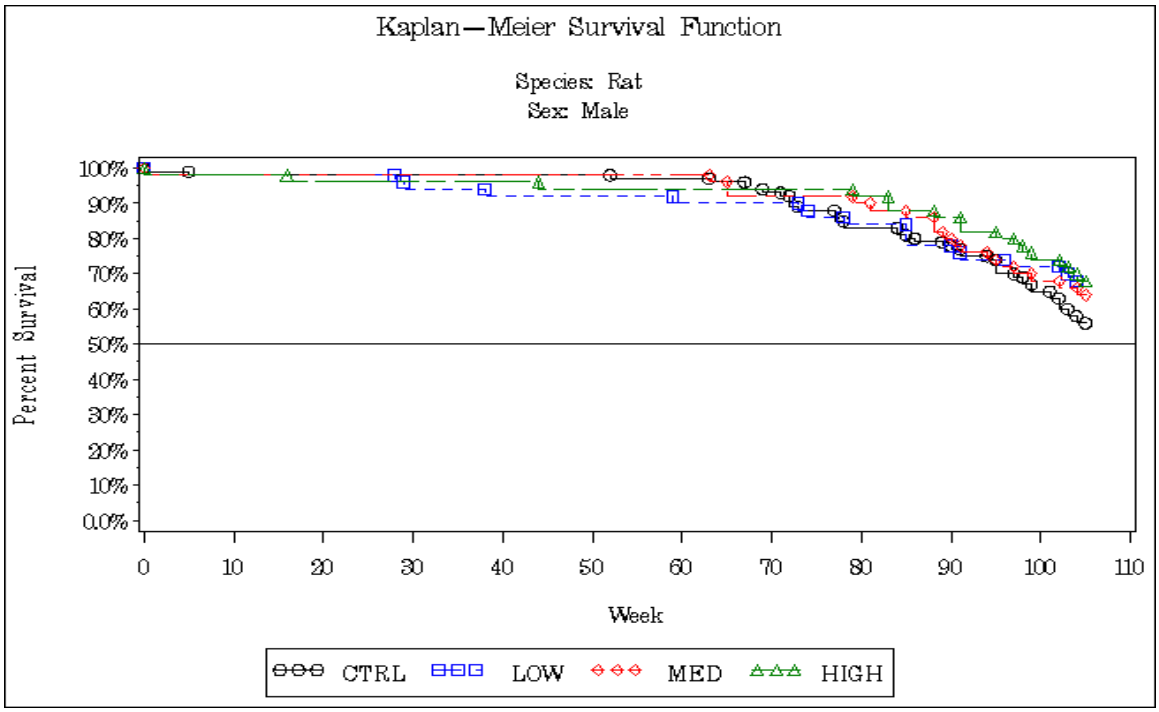
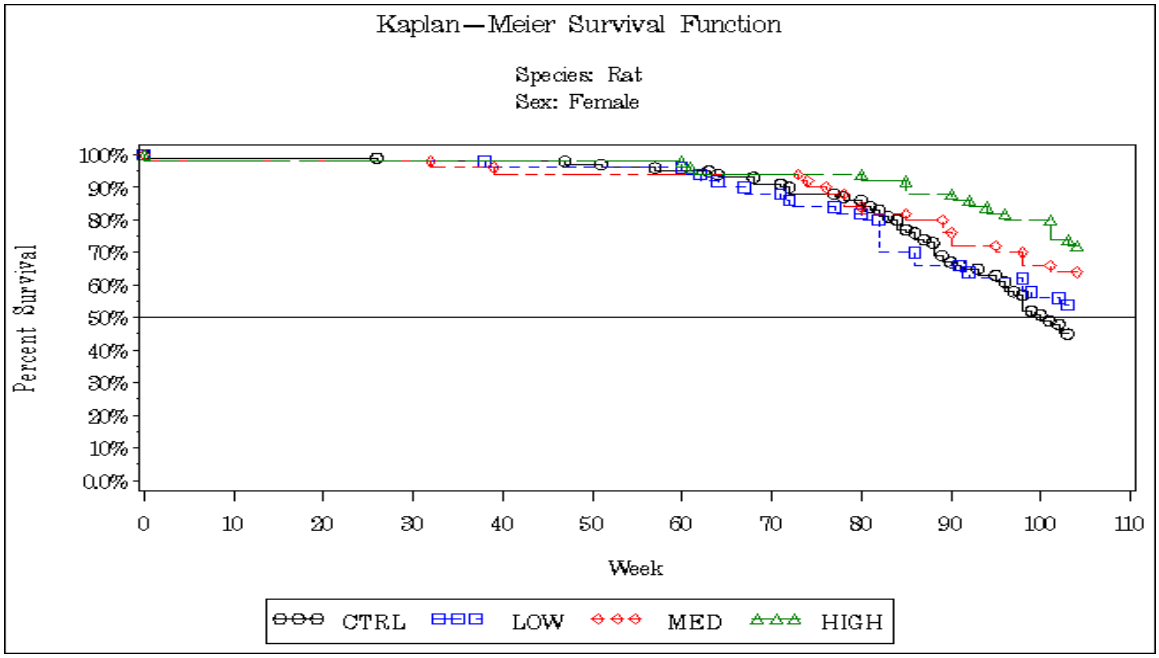


Figure 1b: Survival Curves: Female Rat



The sponsor analyzed tumor findings only if a minimum number (c.f. above) of tumor-bearing animals were observed, whereas all tumor findings were analyzed by this reviewer. Mortality adjusted exact permutation trend tests with increasing dose were computed for tumors observed in the incidental or fatal context. When a tumor presented itself in both contexts within a given time interval, the normal approximation is used and the 'asymptotic' p-value is more appropriate unless the number of tumors is small. In the latter case the true p-value is expected to be bounded by the exact and asymptotic calculations. As there were two two-year carcinogenicity studies in this submission, the appropriate α-levels for trend tests are 0.025 and 0.005 for rare and common tumors, respectively. Dose groups were weighed by the final doses, which is the approach used in all carcinogenicity review. Adjusting the doses according to the incremental increases would have produced weights of 15.17 and 45.5 instead of 16 and 48 and had negligible effect on the results. The sponsor's ordinal (0, 1, 2, 3) weighing should have minimal effect on p-values from exact test but may affect p-values of asymptotic tests to a greater extend.

All tumor findings are presented in the Appendix. Spot-checking of incidence numbers between the sponsor and this reviewer produced no discrepancies. Among the male rats, interstitial cell adenoma of the testes are considered rare based on the concurrent controls and statistically significant with p=0.0125. If these tumors are considered common, the criterion for statistical significance is not reached. The combined testicular cell adenomas and carcinomas resulted in p=0.0282, which would not be considered statistically significant. This reviewer did not group all hemangiomas regardless of site. If these

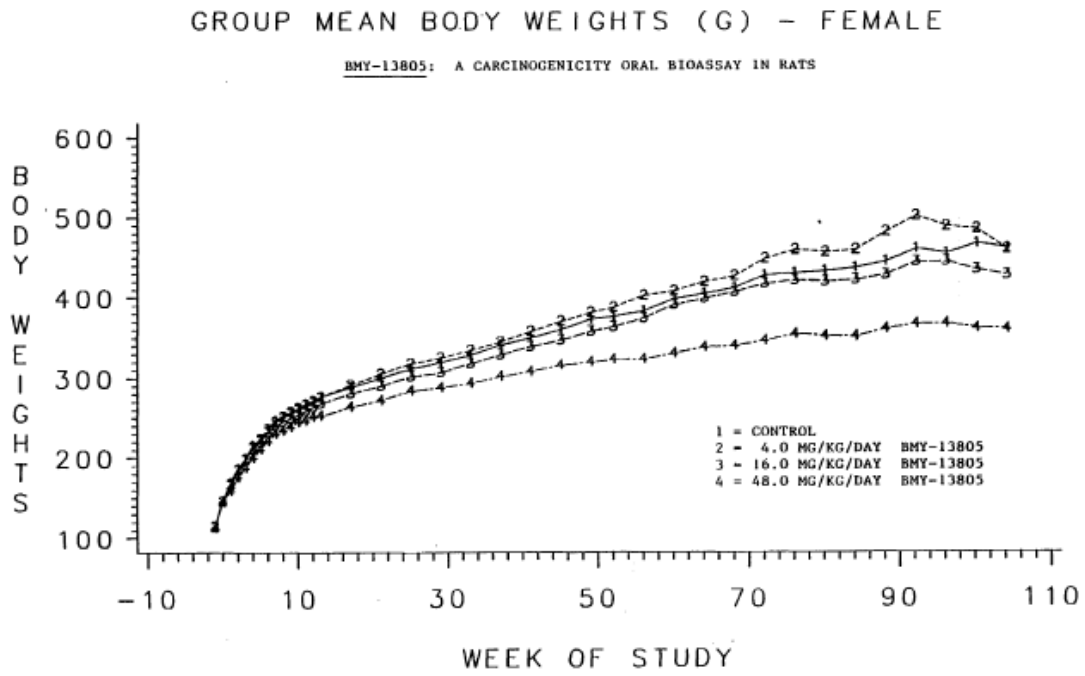
tumors are considered rare, the sponsor's p-value of 0.02 would be considered statistically significant.

Among the females, none of the tumor findings reached statistical significance.

3.0 Validity of the Female Rat Study

Survival was excellent (> 66% at 104 weeks) for all treated groups and in fact better than for the controls. Therefore, it is concluded that a sufficient number of animals lived long enough. As survival was better in the high dose than in the controls, this finding does not contribute to assessing whether the high dose presented a sufficient tumor challenge. From the sponsor's Toxicology Table 5 and their Figure reproduced below, it is apparent that mean body weights of the high dose group were below the controls early on. The reduction was observed from week one on and reached the 10% differential at about week 29. The differential steadily increased till it reached 22% at the end of the study. Therefore, based on early mean body weight data, the high dose appears to have been close to the MTD.

Figure 2: Group Mean Body Weights, Female Rat



4.0 Summary

In this two-year study of Crl:CD(SB)BR rats gepirone HCl was available in the diet at levels of 0, 4, 16, and 48 mg/kg/day after upward adjustment of the mid and high doses in week 19. Survival was very good, especially for the high-dose animals of either sex. This reviewer disagrees with the sponsor's conclusion of considering the increase in testicular

interstitial cell adenomas as not statistically significant. However, the argument revolves around whether the tumors are classified as rare or common. If they are considered common in these animals, this reviewer's findings would not reach the corresponding statistical criterion. The sponsor combined hemangiomas regardless of site. The observed increase in incidence with dose would be considered statistically significant among the male rats if these tumors are considered rare, and non-significant otherwise. There were no statistically significant increases in tumors among the females. Based on the reduction of mean body weights in the high dose females, this study can be considered valid in a sense of exposing a sufficient number of animals for a sufficient length of time at a dose which appears to be close to the MTD.

APPENDIX

Test for Dose-Tumor Positive Linear Trend

Source: Male Rat Data

Organ Name	Organ Code	Tumor Name	Tumor Code	Natural Rate (in ctrl group)	CTRL	LOW	MED	HIGH	Tumor type	pValue (Exact)	pValue (Asymp)
brain	100	malignant gl	10035	.0%	0	0	0	1	FA	0.2143	0.0371
liver	1800	hemangioma	180003	.0%	0	0	0	1	IN	0.2162	0.0365
liver	1800	hepatocellul	180029	2%	2	2	2	0	MX	0.8435	0.8535
liver	1800	cholangioma	180037	1%	1	0	0	0	IN	1.0000	0.7925
liver	1800	hepatocellul	180046	.0%	0	1	0	0	IN	0.6375	0.7257
spinal cord	200	malignant gl	20035	1%	1	0	0	0	FA	1.0000	0.7811
pancreas	2000	lipoma	200010	.0%	0	0	0	1	IN	0.2187	0.0386
pancreas	2000	acinar cell	200031	1%	1	0	0	0	IN	1.0000	0.7677
pancreas	2000	islet cell a	200034	8%	8	1	1	2	IN	0.8295	0.8271
urinary blad	2300	carcinoma	230002	.0%	0	0	0	1	IN	0.2187	0.0386
urinary blad	2300	polyp	230038	1%	1	0	0	0	IN	1.0000	0.7677
urinary blad	2300	cortical car	230053	.0%	0	0	1	0	IN	0.4250	0.4818
testis	2500	interstitial	250001	1%	1	0	0	0	IN	1.0000	0.7925
testis	2500	interstitial	250009	1%	1	2	1	5	IN	0.0125	0.0077
prostate	2700	adenoma	270001	1%	1	0	0	0	IN	1.0000	0.7909
prepuital gl	2900	squamous cel	290036	1%	1	0	0	0	FA	1.0000	0.7897
pituitary	4100	adenoma	410001	41%	41	16	13	19	MX	0.7329	0.7326
pituitary	4100	carcinoma	410002	1%	1	1	0	0	MX	0.8581	0.8334
thyroid	4200	follicular c	420026	4%	4	1	2	2	IN	0.4903	0.4957
thyroid	4200	c cell adeno	420032	4%	4	4	5	2	IN	0.6537	0.6610
thyroid	4200	c cell carci	420033	3%	3	0	1	0	MX	0.8834	0.8692
adrenal	4400	ganglioneuro	440015	.0%	0	0	1	0	IN	0.4250	0.4818
adrenal	4400	pheochromocy	440023	17%	17	2	7	9	IN	0.3288	0.3288
adrenal	4400	cortical ade	440028	5%	5	0	1	1	IN	0.7873	0.7813
spleen	4600	hemangiosarc	460004	1%	1	1	0	1	MX	0.2359	0.1983
thymus	5000	malignant ly	4500	.0%	0	1	0	0	FA	0.5913	0.7033
thymus	5000	thymoma	500049	.0%	0	1	0	0	IN	0.6190	0.7179
mesenteric	5104	hemangiosarc	510404	1%	1	0	0	0	IN	1.0000	0.7930
lacrimal gla	5500	adenoma	550001	1%	1	0	0	0	IN	1.0000	0.7925
lacrimal gla	5500	adenocarcino	550006	.0%	0	0	1	0	IN	0.3784	0.4562
mammary glan	5600	adenoma	560001	1%	1	1	1	1	IN	0.3808	0.4026
mammary glan	5600	fibroadenoma	560011	.0%	0	1	0	0	FA	0.6256	0.7261
skin	5700	hemangioma	570003	.0%	0	0	0	1	IN	0.2187	0.0386
skin	5700	lipoma	570010	1%	1	0	3	3	IN	0.0491	0.0326
skin	5700	papilloma	570013	4%	4	2	2	3	IN	0.2717	0.2694

skin	5700	trichoepithe	570016	.0%	0	1	0	0	IN	0.4865	0.6988
skin	5700	fibrohistioc	570017	.0%	0	1	0	2	FA	0.0692	0.0420
skin	5700	keratoacanth	570020	7%	7	7	3	5	IN	0.5140	0.5193
skin	5700	fibroma	570027	4%	4	1	2	4	MX	0.1483	0.1348
skin	5700	squamous cel	570036	1%	1	1	3	0	MX	0.6975	0.6830
skin	5700	sebaceous gl	570050	.0%	0	2	0	0	IN	0.7150	0.7972
skin	5700	fibrosarcoma	570051	.0%	0	1	1	1	FA	0.1923	0.2172
ear	6600	papilloma	660013	.0%	0	0	1	0	IN	0.3784	0.4562
ear	6600	chondroma	660021	.0%	0	0	1	0	IN	0.4250	0.4818
mesentery	6804	adenocarcino	680406	.0%	0	0	1	0	FA	0.4250	0.4818
mesentery	6804	lipoma	680410	.0%	0	2	0	0	IN	0.7150	0.7972
zimbal gland	6900	adenocarcino	690006	1%	1	0	0	0	IN	1.0000	0.7925
jaw	7001	odontoma	700130	1%	1	0	0	0	FA	1.0000	0.7899
jaw	7001	squamous cel	700136	1%	1	0	0	0	FA	1.0000	0.7879
jaw	7001	rhabdomyosar	700142	.0%	0	1	0	0	FA	0.6130	0.7188
tail	7400	papilloma	740013	1%	1	0	0	0	IN	1.0000	0.7677
tail	7400	keratoacanth	740020	.0%	0	0	0	1	IN	0.2187	0.0386
head	7500	squamous cel	750036	1%	1	0	0	1	FA	0.3884	0.2403

Test for Dose-Tumor Positive Linear Trend

Source: Female Rat Data

Organ Name	Organ Code	Tumor Name	Tumor Code	Natural Rate (in ctrl group)	CTRL	LOW	MED	HIGH	Tumor type	pValue (Exact)	pValue (Asymp)
brain	100	malignant gl	10035	.0%	0	0	1	0	FA	0.4141	0.4709
brain	100	malignant me	10045	1%	1	0	0	0	FA	1.0000	0.7953
stomach	1500	polyp	150038	.0%	0	1	0	1	IN	0.1357	0.0870
jejunum	1602	adenocarcino	160206	1%	1	0	0	0	IN	1.0000	0.8128
liver	1800	hepatocellul	180046	3%	3	0	1	2	IN	0.4282	0.4074
pancreas	2000	islet cell a	200034	4%	4	0	1	1	IN	0.7703	0.7594
pancreas	2000	islet cell c	200052	.0%	0	1	0	0	IN	0.6786	0.7541
kidney	2100	cortical ade	210028	.0%	0	0	1	0	IN	0.4857	0.5263
ovary	3200	leiyomyoma	320007	1%	1	0	0	1	IN	0.3217	0.1693
ovary	3200	theca cell a	320041	2%	2	0	1	0	IN	0.8058	0.8059
uterus	3400	polyp	340038	6%	6	1	0	3	IN	0.5425	0.5484
cervix	3401	leiyomyoma	340107	1%	1	0	0	0	FA	1.0000	0.7872
cervix	3401	polyp	340138	1%	1	0	0	0	IN	1.0000	0.7567
vagina	3500	leiomyosarco	350024	.0%	0	1	0	0	FA	0.6009	0.7133
vagina	3500	squamous cel	350036	.0%	0	0	1	0	IN	0.4820	0.5217
pituitary	4100	adenoma	410001	72%	72	28	31	26	MX	0.9991	0.9987
pituitary	4100	carcinoma	410002	6%	6	3	1	1	MX	0.9456	0.9336
thyroid	4200	follicular c	420026	1%	1	1	1	1	IN	0.3701	0.4011
thyroid	4200	c cell adeno	420032	1%	1	1	1	1	IN	0.3946	0.4008
thyroid	4200	c cell carci	420033	.0%	0	1	2	0	IN	0.6341	0.6723
thyroid	4200	parathyroid	420048	1%	1	0	0	0	IN	1.0000	0.8145
adrenal	4400	ganglioneuro	440015	1%	1	0	0	0	IN	1.0000	0.8145
adrenal	4400	pheochromocy	440023	4%	4	5	0	0	IN	0.9944	0.9795
adrenal	4400	cortical ade	440028	7%	7	4	4	4	IN	0.3752	0.3804
mediastinum	4999	squamous cel	499936	.0%	0	0	1	0	IN	0.4857	0.5263
thymus	5000	malignant ly	4500	1%	1	0	0	0	IN	1.0000	0.7402
lacrimal gla	5500	adenocarcino	550006	1%	1	0	0	0	IN	1.0000	0.7567
mammary glan	5600	adenoma	560001	3%	3	0	0	0	IN	1.0000	0.9195
mammary glan	5600	adenocarcino	560006	10%	10	5	6	3	MX	0.8562	0.8528
mammary glan	5600	fibroadenoma	560011	38%	38	19	23	5	MX	1.0000	0.9999
mammary glan	5600	papilloma	560013	3%	3	2	0	0	IN	0.9801	0.9530
mammary glan	5600	cystadenoma	560039	13%	13	2	3	0	IN	0.9997	0.9980

skin	5700	lipoma	570010	1%	1	1	0	0	IN	0.7717	0.7412
skin	5700	keratoacanth	570020	.0%	0	0	1	0	IN	0.4857	0.5263
skin	5700	fibroma	570027	1%	1	0	1	0	IN	0.7373	0.7464
skin	5700	squamous cel	570036	.0%	0	0	1	0	IN	0.4857	0.5263
skin	5700	mast cell tu	570040	1%	1	0	0	0	IN	1.0000	0.8145
skin	5700	rhabdomyosar	570042	1%	1	0	0	0	FA	1.0000	0.8169
skin	5700	basal cell a	570043	1%	1	1	0	0	IN	0.7717	0.7412
skeletal mus	5800	rhabdomyosar	580042	1%	1	0	0	0	FA	1.0000	0.8062
ear	6600	papilloma	660013	1%	1	0	0	0	IN	1.0000	0.7567
mesentery	6804	lipoma	680410	1%	1	0	0	1	IN	0.4495	0.3040
tail	7400	osteoma	740044	1%	1	0	0	0	IN	1.0000	0.7251

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

/s/

Roswitha Kelly
1/14/02 04:18:44 PM
BIOMETRICS

George Chi
1/14/02 05:07:29 PM
BIOMETRICS