FDA Briefing Document

Psychopharmacologic Drugs Advisory Committee (PDAC) Meeting

December 1, 2015

The attached package contains background information prepared by the Food and Drug Administration (FDA) for the panel members of the advisory committee. The FDA background package often contains assessments and/or conclusions and recommendations written by individual FDA reviewers. Such conclusions and recommendations do not necessarily represent the final position of the individual reviewers, nor do they necessarily represent the final position of the Review Division or Office. We are bringing NDA 21164, gepirone hydrochloride extended-release tablets for the treatment of major depressive disorder (MDD), to the advisory committee in order to gain the Committee's insights and opinions. The background package may not include all issues relevant to the final regulatory recommendation and instead is intended to focus on issues identified by the Agency for discussion by the advisory committee. The FDA will not issue a final determination on the issues at hand until input from the advisory committee process has been considered and all reviews have been finalized. The final determination may be affected by issues not discussed at the advisory committee meeting. All tables, figures, and graphics contained in this briefing document were created by FDA or have been electronically copied and reproduced from the sponsor's submission.

TABLE OF CONTENTS

1 DIVISION DIRECTOR/OND DIRECTOR MEMORANDUM 5

2 CRITICAL MILESTONES 12

3 EFFICACY EVALUATION 17

3.1 SHORT-TERM STUDIES 17
3.1.1 Brief Summary of the Four Studies in Question – Studies Deemed to be Negative by DPP/ODE-I; Deemed to be Failed by the Sponsor 21
3.1.2 Clarifications for the Division's Analysis of the Four Studies Whose Evidence is in Question 23
3.1.3 Evaluation of Strength of Evidence for Effectiveness 27
3.2 MAINTENANCE STUDY ORG28709 28

4 SUMMARY 33

5 APPENDICES 35

5.1 MEMO – RUSSELL KATZ - NON-APPROVAL RECOMMENDATION (03/08/2002) 36
5.2 MEMO – ROBERT TEMPLE - NON-APPROVAL RECOMMENDATION (03/14/2002) 44
5.3 LETTER – ROBERT TEMPLE – NOT APPROVABLE LETTER (03/15/2002) 49
5.4 MEMO – RUSSELL KATZ - NON-APPROVAL RECOMMENDATION (06/15/2004) 60
5.5 MEMO – ROBERT TEMPLE - NON-APPROVAL RECOMMENDATION (06/22/2004) 68
5.6 LETTER – ROBERT TEMPLE – NOT APPROVABLE LETTER (06/23/2004) 71
5.7 MEMO – THOMAS LAUGHREN - NON-APPROVAL RECOMMENDATION (10/25/2007) 79
5.8 LETTER – ROBERT TEMPLE – NOT APPROVABLE LETTER (11/02/2007) 88
5.9 MEETING MINUTES –THOMAS LAUGHREN (12/28/2011) 93
5.10 STATISTICS REVIEW –YEH FONG CHEN, PEILING YANG, HSIEN MING J HUNG (10/25/2013) 143
5.11 STATISTICS REVIEW –PEILING YANG AND HSIEN MING J HUNG (10/25/2013) 175
5.12 CLINICAL REVIEW –SILVANA BORGES (04/18/2014) 187
5.13 ADVICE LETTER –ROBERT TEMPLE (04/18/2014) 228
5.14 FORMAL DISPUTE RESOLUTION MEETING MINUTES –KHUSHBOO SHARMA (03/18/2015) 236
5.15 DISPUTE APPEAL GENERAL ADVICE –JOHN JENKINS (04/07/2015) 287
5.16 INTERIM RESPONSE TO APPEAL –JOHN JENKINS (06/01/2015) 291
5.17 STATISTICS MEMO –LISA LAVANGE (06/08/2015) 298
5.18 REFERENCE TO ARTICLE – "REGULATORY AND SCIENTIFIC ISSUES IN STUDIES TO EVALUATE SEXUAL DYSFUNCTION IN ANTIDEPRESSANT DRUG TRIALS" (08/2015)" 305
5.19 REFERENCE TO ARTICLE – "SUMMARY OF FINDINGS FROM THE FDA REGULATORY SCIENCE FORUM ON MEASURING SEXUAL DYSFUNCTION IN DEPRESSION TRIALS" (08/2015) 306
5.20 REFERENCE TO ARTICLE – "VILAZODONE: CLINICAL BASIS FOR THE US FOOD AND DRUG ADMINISTRATION'S APPROVAL OF A NEW ANTIDEPRESSANT" (11/2011) 307
5.21 GUIDANCE FOR INDUSTRY – PROVIDING CLINICAL EVIDENCE OF EFFECTIVENESS FOR HUMAN DRUG AND BIOLOGICAL PRODUCTS (05/1998) 308

LIST OF TABLES

TABLE 1: TWELVE SHORT-TERM STUDIES WITH HAMD-17 RESULTS……19
TABLE 2: SUMMARY OF PAIRWISE COMPARISONS FOR FOUR CONTROVERSIAL STUDIES…… 20
TABLE 3: PROBABILITIES OF FALSELY CONCLUDING EFFICACY BASED ON TWO POSITIVE TRIALS; VARIOUS NUMBERS OF NEGATIVE TRIALS……27
TABLE 4: SUMMARY OF PROPORTION OF PATIENTS WITH RELAPSE…… 29
TABLE 5: SUMMARY OF PROPORTION OF PATIENTS WHO HAD RELAPSED – PRIMARY ANALYSIS AND POST-HOC ANALYSES…… 32

LIST OF FIGURES

FIGURE 1: SPONSOR'S DIAGRAM OF WELL-CONTROLLED GEPIRONE ER STUDIES……17
FIGURE 2: SPONSOR'S CUMULATIVE PROPORTION OF PATIENTS WHO HAD RELAPSED OVER TIME…… 31

1 DIVISION DIRECTOR/OND DIRECTOR MEMORANDUM

M E M O R A N D U M

DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

DATE: October 13, 2015

FROM: Mitchell V. Mathis, M.D.
Director
Division of Psychiatry Products, HFD-130

John Jenkins, M.D.
Director
Office of New Drugs

TO: Members of the Psychopharmacologic Drugs Advisory Committee (PDAC)

SUBJECT: December 1, 2015 Meeting of the PDAC

This one-day PDAC meeting will focus on issues critical to the Center for Drug Evaluation and Research (CDER) assessment of whether the sponsor has provided substantial evidence of effectiveness in the pending NDA for gepirone HCL extended-release (ER) tablets for treatment of major depressive disorder (MDD). The issues are the subject of a pending request for Formal Dispute Resolution (FDRR) in which the sponsor is appealing to the Director of the Office of New Drugs (OND) prior decisions by the Division of Psychiatry Products (DPP) and the Office of Drug Evaluation 1 (ODE-1) that the available data do not provide the substantial evidence of effectiveness required under the Food Drug and Cosmetic Act (FD&CA) to support approval.

First, we will discuss the general issue of how clinical trial data should be interpreted in a development program that has accumulated a relatively large number of negative/failed trials along with the two positive, adequate and well-controlled trials normally required to meet the substantial evidence standard for approval. Although negative or failed trials are often observed in psychiatric drug development programs, at what point does the information provided by negative/failed trials undermine the evidence of effectiveness?

Second, as a specific example of the first issue, we will discuss the case of gepirone ER for the treatment of MDD where the sponsor has submitted results from a relatively large number of negative/failed trials along with two positive, adequate and well-controlled trials. This application has been the subject of three review cycles and received a “not-approvable” action on each cycle, primarily due to concerns about whether the sponsor has provided substantial

evidence of effectiveness. We are interested in the Committee's thoughts on how to weigh negative/failed trials in our determination of efficacy for a drug that, if approved, would be used to treat a serious, and potentially life-threatening, illness in a clinical environment where there are many effective treatment options.

Background

The effectiveness requirement for a drug was added to the Food Drug and Cosmetic Act (FD&CA, the Act) in 1962.[1] The 1962 amendments included a provision requiring manufacturers of drug products to establish a drug's effectiveness by "substantial evidence." *Substantial evidence* was defined in section 505(d) of the Act as:

> "…evidence consisting of adequate and well-controlled investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could be fairly and responsibly concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof."

It has been FDA's position, based on the language of the statute and the legislative history of the 1962 amendments, that Congress generally intended to require at least two adequate and well-controlled trials, each convincing on its own, to establish effectiveness.

In 1997 under the FDA Modernization Act (FDAMA) section 505(d) of the Act was amended to make it clear that the Agency may consider "data from one adequate and well-controlled clinical investigation and confirmatory evidence" to constitute substantial evidence if FDA determines that such data and evidence are sufficient to establish effectiveness.

The usual requirement for more than one adequate and well-controlled investigation reflects the need for independent substantiation of experimental results. Independent substantiation of experimental results greatly reduces the possibility that a biased, chance, site-specific, or fraudulent result will lead to an erroneous conclusion that a drug is effective.

Of note, the Act does not contemplate and provides no guidance with respect to situations in which countervailing evidence may undercut the evidence of effectiveness from the required adequate and well-controlled investigations. FDA's guidance on the clinical evidence required to support demonstration of effectiveness is also silent on this question. This means that FDA must carefully consider and apply scientific and regulatory judgment in considering the evidence submitted in a marketing application to determine whether effectiveness has been demonstrated.

[1] For a more complete discussion of FDA's thinking on this topic refer to the following:
US Food and Drug Administration. Guidance for Industry. Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products. http://www fda.gov/downloads/Drugs/.../Guidances/ucm078749.pdf. May 1998. Accessed November 2, 2015. ***(****See appendix 5.20, i.e., page 329 of FDA Briefing Document)*

Typically, in weighing evidence of efficacy, FDA pays much attention to the quality of the evidence, e.g., appropriateness of the primary endpoint, clinical meaningfulness of the effect size, study design features, study execution, extent of missing data, internal consistency, dose-response, statistical persuasiveness, and generalizability of the findings to the to-be-marketed U.S. population.

In MDD, it was appreciated some years ago that trials often fail to show treatment effects over placebo; even drugs with well-established efficacy fail to show a treatment effect in some trials. The reasons for this observation are not fully known but, because failure to demonstrate treatment effect is not uncommon in MDD, FDA has encouraged sponsors to include known effective drugs as active controls in their trials of drugs for treatment of MDD. For trials where the new drug shows no treatment effect compared to placebo, the inclusion of an active control helps to provide an indication of whether the trial was adequately designed and executed, such that it would have been capable of showing a treatment effect if one were actually present; i.e., the active control provides a means to assess the assay sensitivity of the trial.

The Division of Psychiatry Products has interpreted trials that fail to show a treatment effect of an active control drug and the test drug as "failed" trials, meaning that the trial itself was a failure because it could not demonstrate the treatment effect of a known effective drug; i.e., the trial lacked assay sensitivity. In contrast the Division has categorized trials that show no effect of the test drug and a positive effect of the active control drug as "negative." In essence, the positive results for the known effective drug demonstrate that the trial had assay sensitivity, and the negative results for the test drug can be interpreted, therefore, as a 'true' negative finding. In the absence of an active control, a study showing no effect of a test drug could be either a negative trial or a failed trial. Historically, for obvious reasons, negative trials have been more concerning than failed trials for demonstration of substantial evidence of effectiveness, but they have nevertheless been seen in development programs for drugs to treat MDD that have been considered effective.

Given the complexities of clinical trials and the myriad of possible outcomes, some have suggested that it might be reasonable to consider the totality of data in a development program in a formal way, after the 2 adequate and well-controlled studies have been successful, to assess whether the overall results are favorable, e.g., using meta-analysis or some other methodology, rather than merely counting the number of positive and negative trials. Clearly, much information is lost if trials are deemed to only be either positive or negative. Again, the Act provides no guidance on how to consider these complexities; i.e., two positive adequate and well-controlled trials along with negative/failed trials. We plan to seek the Committee members' advice on how FDA should interpret the substantial evidence standard in such cases.

Efficacy Data for Gepirone

As noted above, development programs for treatment of MDD often include negative or failed trials. In fact, the Division has seen two development programs for drugs that were approved with as many negative/failed trials as positive trials, but the development program for gepirone includes more negative/failed trials than positive trials. The importance and interpretation of negative trials will be one of the main points of the discussion we plan to have with the PDAC.

The gepirone development program has been a long one; the first NDA submission to FDA was in September, 1999. The clinical trials of gepirone were conducted over a 20-year period and have been managed by three different sponsors.

The gepirone ER development program in MDD consists of 12 short-term treatment trials and one maintenance trial for a total of 13 trials. The maintenance trial was considered negative by the review team. Of the twelve short-term trials, the FDA and the sponsor agree that two (ORG134001 and FD-GBE-007) are adequate and well-controlled and support a treatment effect for gepirone that is in the same range as observed for other approved anti-depressants; i.e., these are positive trials. Views on the interpretation and relevance of the remaining trials, and their importance to the demonstration of efficacy have differed between the FDA and the sponsor. Three of the remaining 10 short-term trials (CN105052, CN105078, and CN105083) were considered uninformative in the evaluation of efficacy, and the sponsor and the review team have agreed that these studies should not be considered further.[2] The remaining seven of the short-term trials (ORG134002, FKGBE008, ORG134023, ORG134004, ORG134006, CN134017, and CN105053) demonstrated no difference between gepirone ER and placebo. Four of these seven were designed with an active-control arm (ORG134004, ORG134006, CN134017, and CN105053), which allows for the distinction between negative and failed trials.

Each of the four trials with active control arms failed to demonstrate a difference between gepirone and placebo, or between active control drug and placebo, based upon their protocol-specified primary endpoints (HAMD-25 in trials ORG134004 and ORG134006, MADRS in trial ORG134017, HAMD-17 and CGI as co-primary endpoints in CN105-053); thus, they are failed trials based on the definition presented above. The review team, in an effort to further explore these failed active-controlled trials, conducted a re-evaluation of these trials using a commonly accepted measure of depression, the Hamilton Depression Rating Scale -17 item (HAMD-17). HAMD-17 was a pre-specified secondary endpoint in trials ORG134004, ORG134006 and ORG134017. Using the HAMD-17 analysis, gepirone ER did not separate from placebo, but the active control did in two of the four trials. In three trials (004, 017, and 006), the review team noted that the active control was nominally statistically superior to gepirone on HAMD-17, which led the review team to consider these to be negative rather than failed trials; i.e., the trials were interpreted as having demonstrated assay sensitivity to detect a difference between active drug and gepirone. Interpretation of the review team's HAMD-17 analyses, in particular the finding of "superiority" of the active controls to gepirone (when the active control failed to differentiate from placebo in any of the four trials based on the primary endpoint, and failed to differentiate in two of the four trials based on the secondary HAMD-17 endpoint), and a conclusion that this demonstrated that the trials had assay sensitivity has been the subject of debate between FDA and the sponsor and within FDA. We will be interested in hearing the Committee members' views regarding how or whether to incorporate these analyses into a determination of whether the standard for substantial evidence of effectiveness has been met.

The review team concluded that the number of negative short-term treatment trials was concerning and raised doubts about the strength of evidence provided from the two positive

[2] This conclusion is subject to reconsideration if a formal "totality of evidence" analytic approach is considered.

trials. The review team also interpreted the randomized withdrawal (maintenance of efficacy) trial as negative: such trials have rarely failed for known effective anti-depressants. Based on these concerns, a third "not-approvable" letter, based primarily on failure to demonstrate substantial evidence of efficacy, was issued on 11/02/07. This view was maintained in a general advice letter from the ODE1 Director to the sponsor on 04/18/2014 (see attachment 5.13).

In summary, the sponsor presented two short-term trials for gepirone ER in the treatment of MDD that DPP, ODE-1, and the sponsor agreed are adequate, well-controlled, and positive. DPP, ODE-1 and the sponsor disagreed on the analysis and interpretation of the remaining 10 short-term negative/failed trials in the gepirone development program and their impact on demonstration of efficacy. DPP and ODE-1 interpreted the large number of negative/failed trials to undermine the two positive trials. Based on this, as well as the negative maintenance study, the Division and ODE-I previously concluded that the statutory standard for substantial evidence of effectiveness had not been met.

The sponsor contends that the two positive trials meet the substantial evidence standard and they are requesting that the Director of OND concur in their conclusion and direct DPP and ODE-1 to work with the sponsor to resolve other outstanding issues, such as labeling and manufacturing issues, so the NDA can be approved. The Director of OND determined that the issues in question in this case were of sufficient complexity and importance to warrant seeking advice from a public AC meeting before reaching a decision on the sponsor's request for formal dispute resolution.

Safety Data for Gepirone

The review team has not identified safety concerns from the gepirone development program that are different from approved anti-depressants and no safety deficiencies were noted in the most recent not-approvable letter.

Sexual dysfunction is a common adverse reaction with serotonergic antidepressants and is a very common reason for patients to discontinue treatment. The sponsor of gepirone ER has proposed to label their drug, if approved, as having less sexual dysfunction than other drugs used to treat MDD. This has been a secondary issue of concern for the Division because, unless and until the drug has demonstrated substantial evidence of effectiveness from adequate and well-controlled investigations, there is no reason to consider a safety advantage over other drugs. As was mentioned above, the decision to not approve gepirone ER is currently being reconsidered, and so it seems appropriate here and at the AC meeting to discuss the Division's current thinking with respect to the evaluation of sexual dysfunction with serotonergic drugs.

Although sexual dysfunction is a known issue with serotonergic antidepressants, data relative to this important adverse reaction have been largely limited to spontaneous adverse event reporting in clinical trials (passive data collection). For patient sensitivity reasons, sexual dysfunction is likely to be underreported when data are collected passively. If a sponsor seeks a claim that their drug does not cause sexual dysfunction (no worse than placebo) or causes less sexual dysfunction than another drug (superior to another drug), then there are ways to assess this in clinical trials and the Division is willing to consider such data to support these claims.

Because an antidepressant with a claim of no sexual dysfunction or less sexual dysfunction will have a marketing advantage over the other approved antidepressants, the Division expects that sponsors seeking this claim complete a rigorous assessment of drug-induced sexual dysfunction before allowing the claim. The exact set of studies required by the Division to assess sexual dysfunction has not been established, but any assessment of an important safety risk with the goal of labeling absence or a significant decrease of that risk would require independent substantiation of the hypothesis being tested—in other words, two positive adequate and well-controlled trials. In addition, because the goal is to demonstrate the absence or a significant decrease of sexual dysfunction, each trial must have an active control that demonstrates sexual dysfunction in order to interpret the effect of the test drug.

Patients with depression can experience decreased interest in sex as a symptom of their illness. As depression improves, interest in sex is expected to improve as well; however, antidepressants can negatively impact desire and performance. Because the treatment of depression can both positively (via overall symptom improvement) and negatively (via adverse reactions) affect sexual function, it may be difficult to quantify the effect of a drug on sexual function, even in a controlled trial. For this reason, the Division has considered the concept of assessing sexual dysfunction in non-depressed patients, or even in healthy volunteers.

In the gepirone ER studies, SSRIs were used as active controls in several trials, and the Changes in Sexual Function Questionnaire (CSFQ) was used to collect data in one short-term trial (study 134017) and its extension phase (study 134506). The Division has accepted CSFQ data in our evaluation of data for claims of reduced or no sexual dysfunction. This has not been the focus of the review team, because the question of efficacy has been paramount—if we cannot agree that a drug has substantial evidence of effectiveness, then consideration of a claim for not worsening sexual function is a moot issue. In these trials evaluating sexual dysfunction, the percentages of patients who either dropped out or had at least one missing item score were large (> 50% for study 134017, for example); hence, many patients had the total score imputed at the planned endpoint visit. Although in study 134017, gepirone was shown statistically superior to fluoxetine (active comparator) in CSFQ at nominal significance level of 0.05, none of the sponsor's inferences about sexual dysfunction considered multiplicity adjustment arising from multiple comparisons, multiple sexual function scales, multiple measures (change from baseline to endpoint, average AUC, etc.) for each scale and each domain/subscale. We generally do not accept these types of retrospective analyses in support of a labeling claim. Even if the results are positive, they would at best be described in Section 6 Adverse Reactions and only after effectiveness has been demonstrated unless the overall study-wise type I error rate had been adequately controlled.

The primary question to be discussed at this AC meeting is how to interpret and apply the statutory standard of substantial evidence of effectiveness when there are negative/failed studies along with positive studies.

In the case of gepirone, the sponsor has submitted two positive adequate and well-controlled trials along with a number of negative/failed trials. The sponsor and DPP/ODE-I do not agree on the characterization of all of the negative/failed trials, and so the numbers of negative versus failed trials are in dispute. The disagreement is primarily related to differing interpretations of

the review team's analysis of the active-controlled trials using HAMD-17, which was not the prospectively planned primary endpoint in any of the trials. All four of the active-controlled trials failed to demonstrate an effect of gepirone or the active control on the pre-specified primary endpoint. This leads to an important question regarding the types of analyses that are valid in determining whether a trial is negative or failed.

Although many in the Division and ODE-I would simply interpret the various studies as positive, negative, or failed, integrate the results, and reach a conclusion on whether substantial evidence of effectiveness has been demonstrated, some would advocate a different approach: one that considers, at least for the studies beyond the two successful studies, the particular results of each study (beyond positive, negative, failed), along with its individual strengths and weaknesses. This Advisory Committee meeting will present an opportunity to consider and discuss both approaches.

Draft points to consider:

1. The Food Drug and Cosmetic Act requires a sponsor to provide substantial evidence of effectiveness to support approval of a new drug. The Act defines the level of evidence necessary as generally requiring two positive adequate and well-controlled trials. Please discuss the following questions related to substantial evidence:
    a. In the situation where two positive adequate and well-controlled trials have been completed, how much and what type of "negative evidence" from other negative or failed trials would it take to undermine a finding of substantial evidence of effectiveness?
    b. What approaches for synthesizing evidence across positive and negative/failed trials in a development program are useful for our decision-making?
2. Please discuss your views on the proper way to evaluate clinical trials for assay sensitivity. Please consider the following questions in your discussion:
    a. Is the protocol pre-specified primary endpoint for efficacy the only meaningful way to evaluate assay sensitivity?
    b. Should *post hoc* analyses of other efficacy endpoints or using other analysis methods contribute to the determination of assay sensitivity? If so, how does the multiplicity due to multiple endpoints or multiple analyses impact the interpretation of assay sensitivity?
3. Has the sponsor provided substantial evidence of effectiveness for gepirone in the treatment of MDD?
4. Is the safety profile of gepirone typical of this drug class?
5. Do the available data support a favorable benefit risk profile of gepirone ER to support approval?
6. What, if any, additional data are needed pre- or post-approval to address outstanding issues? Please describe and clearly state whether you believe any additional data are required prior to approval.

2 CRITICAL MILESTONES

Gepirone has not been approved for marketing in any country. It was originally developed by Mead Johnson and Bristol-Myers Company for the treatment of both anxiety and depression. An IND for the IR formulation was originally submitted in 1984; a second IND for the ER formulation was submitted in 1989. The shift in focus from IR to ER was based on poor tolerance of the IR formulation. These two INDs were originally held by Bristol-Myers Squibb (BMS), but BMS discontinued all trials in 1992. In 1993, rights to gepirone ER were transferred to Fabre-Kramer Pharmaceuticals, Inc. In 1998, Organon, Inc. executed an agreement with Fabre-Kramer Pharmaceuticals granting Organon rights to further develop and market gepirone. These transfers of ownership resulted in several disruptions in the flow of the development program. Several milestones in the development of this product are highlighted below. The details are complex, and are presented here for reference.

- [1994 March] **EOP2 meeting held;** discussion included requirements to establish efficacy of the ER formulation, given the existing database with IR.

- [1998 March] **pre-NDA meeting held.** FDA first became aware of the multiple negative trials for the IR formulation. The main focus of the meeting was on standard requirements for filing the NDA and on format issues.

- [1999 September] **Organon submitted the NDA, but FDA refused to file**. The submission included 16 placebo-controlled trials (6 with ER and 10 with IR), but the sponsor focused on only four of these as useful for assessing efficacy: one short-term ER study (CN105-053), two short-term IR studies (03A7A-003 & 03A7A-001-B), and one long-term IR study (03A7A-002).

  The ER Study CN105-53 was, in fact, a 2-center study in which the protocol-specified analysis called for analyzing all data for the two centers. This analysis was not positive overall, but the sponsor looked at each center independently, and submitted results for one center as a positive trial. Although there had been discussion of this trial at the pre-NDA meeting, the critical facts regarding the analysis plan had not been made clear. Consequently, FDA refused to file this application.

- [2001 May] **Organon successfully re-submitted the NDA.** The submission included 18 placebo-controlled trials (8 with ER and 10 with IR). The sponsor designated four of the 18 adequate and well-controlled trials as providing “proof of efficacy.” Of these four, only Organon’s phase 3 study 134001 was conducted with the ER dosage form, and three other were Phase 2 studies conducted by BMS with the older Immediate Release (IR) formulation (Studies 03A7A-003, 03A7C-001-B and 03A7A-002).

- [2002 March] **FDA issued a Not-Approvable (NA) Letter**, citing inadequate evidence of effectiveness and requesting an additional positive trial with the ER formulation for the claim of short-term efficacy for gepirone ER, although one short-term trial (ORG134001) with the ER formulation was acknowledged as positive. The studies with the IR drug product were found to be deficient to support a claim in MDD and the agency requested

another positive study with the ER formulation to demonstrate antidepressant efficacy. The letter also identified insufficient longer-term safety data.

Major Efficacy Concern: At that point in time, only the ER trial (study 134001) demonstrated statistical significance in favor of gepirone ER for the treatment of MDD. Of the three IR trials intended to provide support to the ER product, only Study 03A7A-003 reached statistical significance per the FDA statistical review. The validity of the findings from this positive IR study was questionable because the study was small in size (n=30 per arm), single center, suffered from high dropout rates, and may not have enrolled patients representative of the MDD patient population of interest (the patients had HAMD scores roughly 10 points lower than most patients in MDD trials, and were identified as having MDD with atypical features).

The NA Letter acknowledged that there had been 18 short-term studies (both IR and ER formulations) at that point, but also acknowledged that only four were sufficient by design/dose/duration/population to be considered relevant: study 134001 was positive and supported the claim, three other studies with the ER formulation (134002, CN105-078, and CN105-083) were considered (in the NA Letter) at the time to be "uninterpretable." The NA Letter stated that an additional trial with the ER formulation would be required to demonstrate efficacy in MDD. The Division acknowledged earlier discussions with the sponsor, at which time it had been agreed that a single positive short-term trial with the ER formulation, in the face of independent evidence for the efficacy of the IR formulation, would be sufficient to support a claim for the efficacy of the ER formulation. The Division concluded that the sponsor had not provided such evidence for the IR formulation, and stated so in the NA Letter.

- [2002 July] **Type B Meeting held to discuss issues related to Not-Approvable Action.** For the IR Study 03A7A-003, the sponsor presented a reasonably strong argument that most of the patients were diagnosed as having MDD by Research Diagnostic Criteria for depression (RDC criteria) and the efficacy results were consistent when those patients who didn't meet the criteria were excluded. The Division concluded that this study could be considered a positive study to provide additional support of efficacy, but noted that there had been enough negative trials with gepirone that an additional positive study with the ER formulation was necessary to support resubmission.

- [2003 July] **Type C meeting held.** At the time of this meeting, the agency had been informed of a then recently completed short-term trial of the ER formulation (Study 134004) using fluoxetine as an active comparator. This study did not demonstrate a statistically significant difference between gepirone ER and placebo, or fluoxetine and placebo on the pre-specified primary efficacy endpoint (HAMD-25), and the results for gepirone were numerically worse than placebo. There was discussion regarding the use of a gepirone ER maintenance trial (Study 28709) in place of a second short-term well-controlled trial to support efficacy of the ER formulation, and the Division agreed to consider those data in a re-submission. The Division did not commit to an assessment of sufficiency of such an approach and reemphasized that the accumulated negative trials remained a concern.

- [2003 December] **Organon re-submitted the NDA.** This re-submission did not include a second positive short-term ER study, but instead included a randomized withdrawal trial for the ER formulation (Study 28709) that the sponsor considered positive.

- [2004 June] **FDA issued a second Not-Approvable Letter.** The NA Letter cited lack of efficacy data to support gepirone ER. The second study intended to be confirmatory evidence of the ER formulation was a failed study, and the randomized withdrawal study submitted to supplement one positive IR study and one positive ER study, was considered by the review team to be negative because of the sponsor's exclusion of patients from their analysis. As a result, the agency stated that the application was not approvable and asked for a second, "robustly positive" study with the ER formulation along with a positive randomized withdrawal study.

- [2005 June] All rights to develop and market Gepirone were reacquired by Fabre-Kramer Pharmaceuticals.

- [2005 October] **NDA Re-submission Meeting.** Sponsor presented the results of a second positive short-term ER study (FKGBE007) and noted their intention to re-submit an NDA without a positive longer-term efficacy trial as advised in the FDA Not-Approvable letter. Although the Division indicated they were likely to file the application, they also indicated that the application would be taken to the PDAC for discussion, given the multiple negative efficacy results for this drug. The Division also indicated that the efficacy review would focus on individual study results rather than on an ISE (Integrated Summary of Efficacy).

  At this meeting, the sponsor also questioned the possibility of a claim for lower risk of sexual dysfunction. FDA indicated that this was possible if the sponsor could consistently show an advantage over other antidepressants on sexual dysfunction.

- [2007 May] **Fabre-Kramer resubmitted the NDA,** including the results of a second positive short-term ER study (FKGBE007) and various meta-analyses of the 12 short-term ER studies that the sponsor considered adequate from the standpoint of dose. This re-submission did not include the results of another randomized withdrawal study as FDA had advised, but instead included a post-hoc re-analysis of the results from maintenance study 28709. With these data the sponsor was seeking the additional claim of maintenance treatment of MDD for gepirone ER. The sponsor also proposed to include a labeling statement that patients treated with gepirone had statistically significantly better sexual functioning than SSRI-treated patients.

- [2007 November] **FDA issued a third Not-Approvable Letter**, with two major deficiencies noted: (1) lack of substantial evidence of effectiveness, despite two positive trials; (2) observed magnitude of treatment effect unacceptably small.

  Brief Summary: FDA acknowledged that some of the studies were terminated early for business reasons and, therefore, might not be expected to provide such evidence because they did not reach their planned sample sizes.

The NA Letter acknowledged two positive trials, FDGBE007 and 134001, from a pool of 12 short-term gepirone ER trials. The agency re-evaluated these 12 trials using a common endpoint, HAMD-17 Total Score, but acknowledged that this was not the protocol specified endpoint for three of the 12 trials. The results included three trials (134004, 134017, and 134006) with different protocol-specified endpoints (HAMD25 for 004 and 006, and MADRS for 017) that, when evaluated with the HAMD-17, were found to have statistically significant differences between active comparator and gepirone. The Division interpreted this as evidence of assay sensitivity and concluded that these studies did not demonstrate an effect of gepirone on depression. That is, the Division interpreted these as negative studies.

The maintenance study was interpreted as negative because the applicant eliminated 40 patients from the efficacy analysis several years after the blind was broken; this negative maintenance trial was cited in the letter as further evidence of lack of efficacy. The review team identified no safety findings that would preclude the approvability of gepirone ER. There were multiple CMC deficiencies identified as well, but those are not the focus of the PDAC meeting.

- [2008 January] **Post-Action Meeting Held.** The Sponsor disagreed with FDA's conclusions as stated in the Not-Approvable Letter.

  With regard to assessing the strength of evidence for effectiveness, the sponsor thought the Agency should not use all 12 trials. In addition, the sponsor argued that, in a few of these studies, HAMD-17 was not the pre-specified primary endpoint, but FDA used this endpoint to assess the strength of the overall evidence for effectiveness. FDA presented a counter-argument that the HAMD-17 was a common metric assessed in each of the 12 trials and so could be used in an effort to assess the entire set of studies together.

- [2011 April] **The Sponsor submitted *Request for Reconsideration*** with three arguments: (1) FDA analysis of totality of evidence flawed, (2) comparative effectiveness standard invalid, and (3) gepirone ER's benefit-risk profiles adequate for approval.

- [2011 November] **Type C Meeting held to address Sponsor's *Request for Reconsideration*.** FDA's responses included the following:

  (1) Disclaimed any reliance on a comparative efficacy standard.
  (2) Reemphasized that the Not-Approvable action was based only on a lack of substantial evidence of effectiveness despite two positive short-term trials.
  (3) Changed the view of two short-term trials (CN105-078 and CN105-083) from negative to failed because of early termination.
  (4) With regard to the longer-term maintenance trial (study 28709), FDA did not find the post-hoc analysis (by excluding 40 patients) credible or valid.
  (5) In summary, at this time in the consideration of the gepirone ER development program, four short-term trials and one maintenance trial remained controversial. FDA suggested that the sponsor provide additional justifications to support their arguments.

- [2012 December] **The Sponsor submitted an NDA Amendment in support of Informal Appeal.** This submission included a summary of efficacy results and arguments for how the results should be interpreted for each of 12 short-term studies and one longer-term study, as well as expanded exploratory analyses of sexual dysfunction data. If FDA would agree that gepirone was shown to provide substantial evidence of efficacy, the sponsor agreed to then prepare and submit the remainder of their complete response to the NDA.

- [2014 April] **FDA issued a General Advice Letter**, stating that all five controversial trials (four short-term, one longer-term) were considered negative. FDA concluded that "*Although two short-term trials favor gepirone ER for the treatment of MDD, the seven negative short-term studies and one negative maintenance trial raise considerable doubts about the effectiveness of gepirone in the acute or sustained treatment of depression. The 2 positive studies could represent chance findings, given the absent, negative, or minimal findings in 8 other studies*."

- [2015 January] **FDA accepted the sponsor's *Formal Dispute Resolution Request***. The Director of Office of Biostatistics, Dr. Lisa Lavange, was subsequently consulted by the Director of the Office of New Drugs, Dr. John Jenkins. Refer to Appendix for her memorandum.

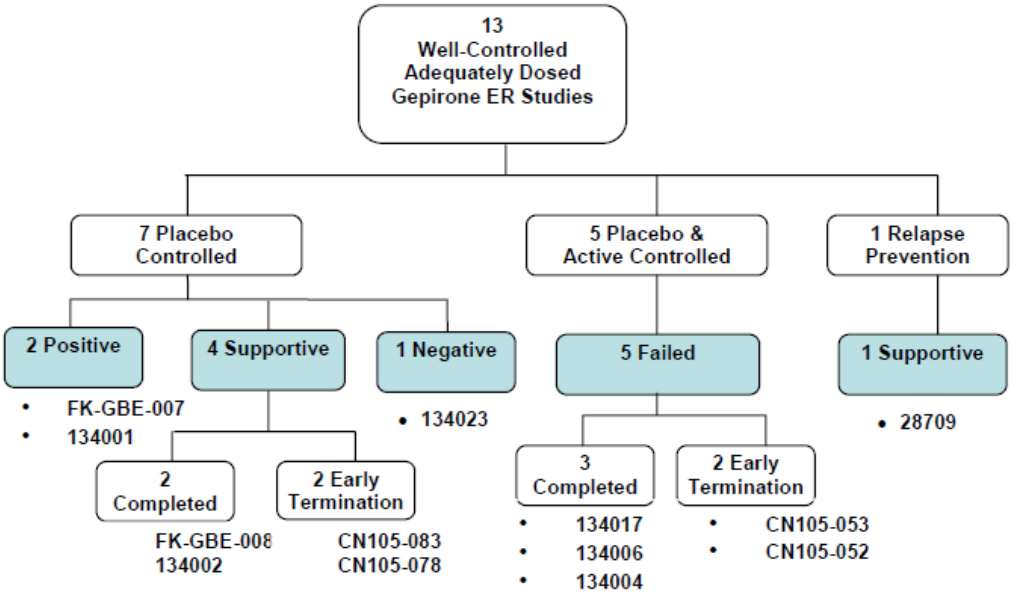
3 EFFICACY EVALUATION

This section focuses on the ER formulation.

In its third NDA re-submission in 2007, the sponsor indicated that there were 15 adequate and well-controlled ER studies that were randomized, double-blind and placebo-controlled. Among these, two studies investigated inadequate doses and the other 13 studies were conducted at relevant doses, so these 13 studies were the primary focus of the sponsor's efficacy evaluation in its third re-submission. Of these 13 studies (Figure 1), one was the maintenance study (28709), the basis of sponsor's second re-submission. The sponsor noted that some studies were terminated early due to business reasons; these studies are marked in the figure as "*early termination*." The sponsor classified some of these studies as *supportive* because gepirone separated from placebo in a statistically significant manner on one or more key outcome measures at one or more time points.

The 12 short-term trials are summarized in Section 3.1 with a focus on the four studies for which the evidence provided by each is controversial, and the maintenance trial is summarized in section 3.2.

Figure 1: Sponsor's Diagram of Well-Controlled Gepirone ER Studies



[**Source:** Figure 1 of "summary-2007.pdf" in sponsor's 2007-05-01 submission]

3.1 Short-Term Studies

Table 1 summarizes the treatment arms, efficacy outcomes, and study conclusions from the Division and the sponsor, respectively, for the 12 short-term trials. There were two gepirone arms in two failed studies CN105-078 and CN105-084, so the numbers of patients on gepirone were essentially twice the number on placebo.

From the table, studies FK-GBE-007 and ORG134001 were positive. Studies FK-GBE-008 and ORG134002 were viewed as negative by the review team because gepirone was not shown to be superior to placebo based on the prospectively planned primary endpoint HAMD-17. The sponsor classified both studies as supportive based on results of secondary efficacy measures evaluated at the planned time of assessment, or at an earlier visit. The findings of the four studies that included an active-comparator arm to assess assay sensitivity are disputed (shown in red font in the table). Table 2 summarizes the pairwise comparison results of these four studies.

Per the review team, gepirone was numerically worse than placebo in three of the four studies, whereas the active comparator was numerically better than placebo in all four studies. Of note, the active comparator was shown to be statistically significantly better than gepirone on the HAMD-17, which was not the pre-specified primary endpoint, in three of these studies.

The controversies in these four studies are briefly summarized in Section 3.1.1. Clarifications for the Division's re-analysis of data and interpretation of results from these four studies are highlighted in Section 3.1.2.

Table 1: **Twelve Short-Term Studies with HAMD-17 Results**

Trial No.	Number of Subjects		Active Comparator, Dose range (mean dose)	Gep Doses Range (mean daily dose in mg)	Pre-Specified Primary Endpoint(s)	HAMD-17 Results† (Gep.-Pbo.)		Conclusion	
	Gep.	Pbo.				LS Mean Diff[1]	*p*-value	Division	Sponsor
FK-GBE-007	116	122	None	20-80 (58.2)	HAMD-17	-2.45	0.018	**Positive**	
ORG 134001	101	101	None	20-80 (61.1)	HAMD-17	-2.47	0.013	**Positive**	
FK-GBE-008	96	99	None	20-80 (60.0)	HAMD-17	-1.38	0.20	Negative	Supportive
ORG 134002	102	103	None	20-80 (57.9)	HAMD-17	-0.71	0.42	Negative	Supportive
ORG 134023	123	123	None	20-80 (61.3)	HAMD-17	0.13	0.90	Negative	
ORG 134004	124	130	Fluoxetine 20-40 mg (34.1)	20-80 (67.1)	HAMD-25	1.04	0.18	**Negative[2]**	**Failed**
ORG 134006	140	143	Paroxetine 10-40 mg (28.2)	20-80 (55.3)	HAMD-25	0.22	0.76	**Negative[2,3]**	**Failed**
ORG 134017	159	159	Fluoxetine 20-40 mg (25.9)	40-80 (59.7)	MADRS	0.65	0.39	**Negative[2]**	**Failed**
CN105-053*	56	56	Imipramine 50-200 mg (145)	10-60 (50.4)	HAMD-17, CGI	-2.0	0.19	**Negative[3]**	**Failed**
CN105-052*	35	37	Fluoxetine 20-80 mg (23.3)	20-60 (43.4)	HAMD-17, CGI	-0.69	0.74	Failed	
CN105-078*	88	47	None	10-50 (30.4) 20-100 (52.6)	HAMD-17	-1.0	0.36	Failed	
CN105-083*	73	39	None	10-50 (30.4) 20-100 (57.1)	HAMD-17	-0.49	0.75	Failed	

†Results were from LOCF ANCOVA including treatment and center as factors and baseline score as a covariate, with active comparator arm included in the analyses.

*Trial was terminated early.

[1] A lower HAMD-17 total score indicates less severe symptoms; a negative difference indicates gepirone better than placebo.

[2] Active comparator beat gepirone on HAMD-17.

[3] Active comparator beat placebo on HAMD-17.

[Source: Sample sizes were from Table 24 and Information on dose rang from Table 11 and Table C4 of ISE-2007.pdf in sponsor's 2007 submission. Efficacy results are from FDA Statistical Review by Dr. Fanhui Kong.]

Table 2: Summary of Pairwise Comparisons for Four Controversial Studies

Trial	LS Mean Change from Baseline			LS Mean Diff**			*p*-value		
	Geprione (G)	Active Comparator (AC)	Placebo (P)	G-P	AC-P	G-AC	G-P	AC-P	G-AC
ORG134004	**N = 124**	**Fluoxetine N = 134**	**N = 130**						
Sponsor Results:* HAMD-25	-9.76 --	-- -11.66	-10.63 -10.61	(0.87)†	(-1.05)†	(1.91)†	0.42	0.33	(0.07)†
FDA Results: HAMD-17	-5.75	-7.46	-6.79	1.04	-0.68	1.71	0.18	0.38	**0.027**
ORG134006	**N = 140**	**Paroxetine N = 136**	**N=143**						
Sponsor Results:* HAMD-25	-10.94 --	-- -12.58	-11.00 -11.23	(0.06)†	(-1.35)†	(1.10)†	0.95	0.18	(0.27)†
FDA Results: HAMD-17	-7.09	-8.94	-7.31	0.22	-1.63	1.85	0.76	**0.026**	**0.012**
ORG134017	**N = 159**	**Fluoxetine N=159**	**N=159**						
Sponsor Results: MADRS	-12.23	-13.88	-12.73	(0.5)†	(-0.9)†	(1.65)†	0.65	0.30	0.14
FDA Results: HAMD-17	-10.37	-11.92	-11.02	0.65	-0.90	1.54	0.39	0.24	**0.042**
CN105-053	**N = 56**	**Imipramine N = 54**	**N = 56**						
Sponsor Results: HAMD-17	-9.7	-11.5	-9.0	(-0.7)†	(-2.5)†	(1.8)†	0.69	0.14	(0.29)†
FDA Results: HAMD-17	-10.16	-11.35	-8.16	-2.0	-3.19	1.2	0.19	**0.038**	0.44

*Separate models used for different pairwise comparisons, leading to separate estimates of placebo responses in the same trial.

**A lower HAMD-17 total score indicates less severe symptoms; a negative difference indicates the former better than the latter.

†Estimated LS Mean differences and p-values are provided by FDA statistician Dr. Jinglin Zhong (or former FDA statistician Dr. Yang Yang during her FDA employment) using sponsor's approach.

[Source: Sponsor's results are from study reports in sponsor's 2007-05-01 submission. FDA's results (except †) are from statistical review by Dr. Fanhui Kong.]

3.1.1 Brief Summary of the Four Studies in Question – Studies Deemed to be Negative by DPP/ODE-I; Deemed to be Failed by the Sponsor

Studies ORG134004 and ORG134006

Protocol Titles:

- **ORG134004:** *A double-blind, multi-center, randomized, placebo-controlled, efficacy and safety study of Org 33062 ER and Fluoxetine in subjects who suffer from major depressive disorder with atypical features*
- **ORG134006:** *A double-blind, multi-center, randomized, placebo-controlled, parallel group study of efficacy and safety of Org 33062 ER and paroxetine in subjects who suffer from major depressive disorder with atypical features*

These two studies were 3-arm, flexible-dose (20-80 mg/day) studies including an active-comparator arm. Both studies enrolled patients with Atypical Depression (i.e., MDD with Atypical Features or MDD-AF). In both studies, the protocol-specified primary endpoint was HAMD-25.

The major disagreement on the evidence provided by these two studies was due to different interpretations of assay sensitivity, in part because of a disagreement on the particular efficacy endpoint used to assess assay sensitivity. In the division's view, both trials (134004 and 134006) had assay sensitivity because the active comparators (fluoxetine and paroxetine, respectively) were statistically significantly superior to gepirone on HAMD-17. Moreover, in Study 134006, paroxetine was also statistically significantly superior to placebo on HAMD-17. In the sponsor's view, neither study had assay sensitivity, because neither active comparator was statistically significantly superior to placebo on HAMD-25, the pre-specified primary endpoint. Moreover, the sponsor asserted that assay sensitivity should not be judged based on the comparison between two drugs, particularly because the efficacy of these active comparators in Atypical Depression was unknown.

In addition to arguing against assay sensitivity, the sponsor provided additional reasons why these studies failed, including: (1) HAMD-25 is the more appropriate measure of efficacy in the MDD-AF population; (2) use of a comparator with unknown efficacy in atypical depression; (3) different population: lower severity of depression than most studies; (4) high placebo response rate; (5) inappropriate use of the comparator; (6) and significant treatment by site interaction ; (7) reasons for trends in HAMD favoring placebo over gepirone. The review team at the time determined that these reasons were not persuasive enough for the evidence provided by these trials – the lack of treatment effect – to not be taken into account in determining the overall evidence of effectiveness of gepirone ER.

Study ORG134017

Protocol Title: *A double-blind, multi-center, randomized, placebo-controlled, efficacy and safety trial of Org 33062 ER and fluoxetine in subjects with major depressive disorder*

This was a 3-arm, flexible-dose (40-80 mg/day) study. The protocol-specified primary endpoint was MADRS. The major disagreement about the evidence provided by this study involved the determination of assay sensitivity. The division concluded that this study had assay sensitivity because fluoxetine was statistically significantly superior to gepirone on HAMD-17. The sponsor disagreed because fluoxetine was not statistically superior to placebo on MADRS.

With regard to the statistical analysis approach, the only major difference was that ANCOVA was used for the review team's analysis and ANOVA was used for the sponsor's analysis. Regardless of whether ANCOVA or ANOVA was used, fluoxetine was statistically significantly superior to gepirone on HAMD-17; however, fluoxetine was not superior to placebo with respect to the pre-specified primary endpoint using either analysis method.

In addition to arguing against assay sensitivity, the sponsor provided additional reasons why this study failed, including: (1) inconsistency among sites; (2) high placebo response; (3) positive results from reliable investigators; (4) flaws in study conduct; (5) spurious trends favoring placebo. The review team was not convinced that these reasons were persuasive enough to not evaluate the data from this trial as part of the overall evidence of effectiveness.

Study CN105-053

Protocol Title: *A double-blind, multicenter trial of Org 33062 ER, imipramine, and placebo in the treatment of depressed outpatients*

This was a 3-arm, flexible-dose (10-60 mg/day) study. The study was conducted at two US sites (Feiger and Gelenberg) but was terminated prematurely. According to the sponsor, at the time of termination, a total of 170 patients had been randomized and 166 patients had post-baseline data: 120 patients at the Feiger site and 46 patients at the Gelenberg site (39% of the enrolled population). Of 46 patients at the Gelenberg site, only 57% completed the trial. There were two pre-specified co-primary endpoints: HAMD-17 and CGI responder (yes vs. no).

This study was the primary basis for the sponsor's initial NDA submission, which FDA refused to file. Gepirone was not shown to be effective in the overall population, but the sponsor argued that the results from a single site (Feiger site) supported gepirone's effectiveness.

The sponsor considers this study to have failed for several reasons, including early termination, lower mean modal dose of gepirone, and higher placebo response, all of which occurred at the Gelenberg site. In the Division's view, however, this study had assay sensitivity because imipramine, the active comparator, was statistically significantly superior to placebo on HAMD-17, but gepirone was not. This conclusion was based on a statistical model without the treatment-by-center interaction included, and the conclusion was the same whether using ANCOVA or the protocol-specified analysis method, ANOVA. The study did not have assay sensitivity, when the treatment-by-center interaction term was included in the model.

The review team noted that early termination of a large number of patients appeared to have negatively impacted the efficacy results at the Gelenberg site. However, the sponsor's results

showed that although the mean modal dose for gepirone was lower at the Gelenberg site, it was also lower for imipramine at that site, and the responses in the gepirone and imipramine groups differed only slightly from their counterparts in the other site. Although the differential placebo responses may have increased the variability of the efficacy outcome, it is difficult to ignore the efficacy outcome from the Gelenberg site, even though the sponsor concluded superiority of both gepirone and imipramine to placebo at the Feiger site (where enrollment was completed).

3.1.2 Clarifications for the Division's Analysis of the Four Studies Whose Evidence is in Question

Summarized below are the Division's clarifications for their re-analysis of data from the four trials for which the interpretation is in dispute.

[1] **Assay Sensitivity:**

- **Relevant Studies:** In the sponsor's view, all four studies failed because of lack of assay sensitivity, but in the division's view, all four had assay sensitivity.

- **Clarifications:** In the sponsor's opinion, superiority of the active comparator to gepirone does not provide evidence of assay sensitivity, given that no difference was detected between the active comparator and placebo. The Division's view was that the ability to detect a statistically significant difference between effective and ineffective treatments is the essence of assay sensitivity. Having shown a difference between the active comparator and gepirone, the Division asserted that all four studies had assay sensitivity.

  The Division acknowledges that the active comparator failed to reach statistical significance over placebo on the protocol-specified primary endpoint (HAMD-25 for studies ORG134004 and ORG134006; MADRS for study ORG134017). However, the treatment effect favored the active comparators on the HAMD-17, the primary endpoint in nine of the 12 well-controlled trials. The active comparator showed superiority to placebo and gepirone in ORG134006 and to gepirone in ORG134004 and ORG 134017. The review team also noted that in three of these four studies, gepirone was numerically worse than placebo whether based on the protocol-specified primary endpoint or HAMD-17.
  The Division recognized that depression trials of effective drugs have failure/negative rates of about 50%, but emphasized that this development program had an unusually low trial success rate in what appeared to be well-controlled studies. Based on a common endpoint, HAMD-17, superiority of the active comparator to placebo (ORG134006 and CN105053) and/or to gepirone (ORG134004, ORG134006, and ORG134017) was observed in all three trials of adequate size that had an active comparator, an outcome very unusual from what had been seen with approved antidepressant drugs.

[2] Primary Endpoint:

- **Relevant Studies:** ORG134004 and ORG134006 (protocol-specified: HAMD-25); ORG134017 (protocol-specified: MADRS).

- **Clarifications:** Because HAMD-17 is a widely used efficacy endpoint and was the protocol-specified primary endpoint in nine of the 12 trials, the Division considered it a reasonable endpoint to use as a common metric for each trial in an overall assessment of a depression program, regardless of the protocol-specified primary endpoint(s). It was noted that both HAMD and MADRS are highly correlated and have similar sensitivity in detecting antidepressant efficacy in drug trials.[3] Among the MDD trials included in NDAs, almost all had used HAMD as the primary endpoint; it was not until recently that the frequency of using MADRS increased.

  Atypical MDD Patient Population in Studies ORG134004 and ORG134006:
  Because MDD with atypical features was an entry criterion in these two studies, the sponsor argued that the patient populations differed significantly from the participants in the rest of the studies and, as such, HAMD-25 (not HAMD-17) is the appropriate efficacy measure. However, the review team had found a similar distribution of HAMD-25 total scores, HAMD-17 total scores, the sum of the eight items missing in the HAMD-17 total score (compared with the HAMD-25 total scores), and the sum of the five items from the HAMD-25 that measure atypical features in both positive studies (FKGBE007 and ORG134001), which enrolled all patients with MDD, and in studies ORG134004 and ORG134006, which enrolled patients with atypical depression. The distributions of these depression rating scales were also similar among treatment arms in all four studies. The review team noted that this indicated that the patient populations in all four studies were comparable with respect to the various depression rating scales commonly used in clinical trials (i.e. HAMD-17, HAMD-21, MADRS), and that any of these scales would have been able to differentiate an effective antidepressant from placebo. In addition, using HAMD-17 as the primary endpoint for studies ORG134004 and ORG134006, the *p*-values for the gepirone-placebo comparison were in fact smaller than those obtained using HAMD-25. Therefore, the HAMD-17 total score seemed to be at least as sensitive as the HAMD-25 total score at detecting a difference between gepirone and placebo in studies ORG134004 and ORG134006. The team also noted that, in the sponsor's analysis, the two positive trials (where HAMD-17 was the primary endpoint) also showed positive results on HAMD-25. In the review team's view, this provided further evidence that any of the above mentioned depression scales would have been sensitive in showing a drug effect.

[3] Use of a Comparator with Unknown Efficacy in Atypical Depression/Inappropriate Use of Comparator:

[3] Khin N, Chen, Y, Yang, Y, et al. Exploratory Analyses of Efficacy Data from Major depressive Disorder Trials Submitted to the US Food and Drug Administration in Support of New Drug Applications. *J Clin Psychiatry*. 2011; 72(4): 464 – 472.

- **Relevant Studies:** ORG134004 (comparator: fluoxetine) and ORG134006 (comparator: paroxetine).

- **Clarifications:** The sponsor stated that fluoxetine and paroxetine had not been thoroughly studied in patients with atypical depression and that the use of a comparator with unknown efficacy in the target population limits the value of the study to judge the efficacy of gepirone in that population. The review team pointed out that, if this were true, the sponsor was ill-advised to include inappropriate active comparators in these studies for assay sensitivity. In fact, in those studies, the two active comparators were statistically significantly superior to gepirone on HAMD-17. It was not reasonable to select active comparators for these trials, and in retrospect, having found a difference between the active comparator and placebo (or gepirone), declare that the comparator was inappropriate. Moreover, the review team pointed out that if the active comparator arms were ignored, neither study would provide evidence of efficacy, because gepirone was not statistically significantly better than placebo on either HAMD-17 or HAMD-25.

[4] High Placebo Response:

- **Relevant Studies:** ORG134004, ORG134006 and ORG134017.

- **Clarifications:** The sponsor asserted that the negative results of these studies were due in part to the high placebo response. In the review team's view, such substantial responses in the placebo arm are common in acute depression trials and undoubtedly contributed to the high failure rate with these trials. However, the placebo responses with these trials were not unusually high and did not appear related to success or failure.

  In studies ORG134004 and ORG134006, the placebo response (about -7 points in the HAMD-17 for both) was similar to that observed in the positive trials (FKGBE007 and ORG134001, about -8 and -7 points in the HAMD-17, respectively). The review team noted that the active comparators (fluoxetine and paroxetine) were consistently better numerically than placebo and were shown statistically significantly superior to gepirone on HAMD-17. Study ORG134017 had a large placebo response, with a 45% HAMD-17 responder rate in the placebo arm per sponsor's responder definition. Nonetheless, the HAMD-17 responder rate in the fluoxetine group was 57%, whereas it was only 42% in the gepirone ER group, and fluoxetine was statistically significantly superior to gepirone on HAMD-17. The trial was thus able to distinguish different rates of response from different drugs - despite the high responder rate in the placebo arm using a common measure of depressive symptoms that was different than the protocol pre-specified endpoint

[5] Inconsistency Among Sites

- **Relevant Studies:** ORG134004, ORG134006 and ORG134017.

- **Clarifications:** The sponsor argued that the gepirone effect was inconsistent across sites, with some sites favoring gepirone over the active comparator and others in the reverse direction. In the review team's view, it is not surprising to observe inconsistent results across sites if the overall treatment effect is relatively small, as it typically is in antidepressant trials. Even in the positive study FK-GBE-007, there appeared to be large variations in treatment effect (difference between gepirone and placebo) among sites (*p*-value = 0.092 for the treatment-by-center interaction according to the sponsor). Thus, if one were to hold the inconsistent findings across sites against those studies, the validity of the positive study FKGBE-007 would also be questionable. It was also noted that the *p*-value for the interaction term was > 0.3 in both studies ORG134004 and ORG134006, and > 0.1 in study ORG134017, which did not suggest significant variation across sites, as compared to the positive study FKGBE-007.

[6] Whether to Include Treatment-by-Center Interaction in the Statistical Model

- **Relevant Study:** CN105-053

- **Clarifications:** The sponsor's study results were derived from a statistical model including the treatment-by-center interaction terms. The review team re-analyzed the data excluding the treatment-by-center interaction terms. In this study, the *p*-value for the interaction term was relatively large (> 0.3) whether using ANOVA or ANCOVA, supporting a post-hoc analysis excluding the interaction.

[7] Common vs. Separate Statistical Models in Analyses

- **Relevant Studies:** 134004 and 134006

- **Clarifications:** In the sponsor's analyses, the statistical model did not include data from the active comparator when comparing gepirone with placebo. Likewise, it did not include gepirone data when comparing the active comparator with placebo. In FDA's analysis, the statistical model included all treatment arms, and the three pairwise comparisons were then generated based on the same model. This approach avoids generating different placebo responses within a single trial. In addition, the approach tends to make the variance estimate more precise because the common model includes a larger data set.

[8] ANOVA vs. ANCOVA

- **Relevant Studies:** ORG134017, CN105-053

- **Clarifications:** The sponsor's protocol-specified analysis was ANOVA, but the review team used ANCOVA to assess the efficacy for all 12 trials for consistency. Compared with ANOVA, ANCOVA tends to provide variation reduction in estimating treatment effects, thereby increasing power to detect treatment differences. The review team also noted that the sponsor used ANCOVA in their meta-analyses of the 12 trials. For these

two trials, the conclusions were the same regardless of whether ANOVA or ANCOVA was used.

3.1.3 Evaluation of Strength of Evidence for Effectiveness

As noted in FDA Guidance for Industry – Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products (May, 1998), substantial evidence of effectiveness is generally accepted to mean two positive adequate and well-controlled trials. The usual requirement for more than one adequate and well-controlled investigation reflects the need for independent substantiation of experimental results. Although the division agreed that the sponsor had provided two positive, short-term, adequate and well-controlled trials, the division remained concerned with the overall evidence of effectiveness, given the number of negative trials that accompanied the two positive trials.

Of the 10 non-positive trials, the division reached agreement with the sponsor that three trials (CN105-052, CN105-078, CN105083) should be considered failed because they were terminated early. However, there was disagreement on the evidence provided by four trials (ORG 134004, ORG134006, ORG134017, CN1050053), each of which had an active comparator. The sponsor considered these trials failed (hence, not interpretable), but the Division considered them negative because the active comparator was statistically significantly superior to either gepirone or placebo on HAMD-17. Consequently, the Division concluded that there were seven non-positive but interpretable trials, each considered to be a negative trial. In three of these seven trials (ORG134023, ORG134004, CN134017), gepirone was numerically worse than placebo on the primary endpoint and on many secondary endpoints. In the Division's experience, such conflicting findings would be highly unusual with effective antidepressants.

The Division concluded, given a total of nine interpretable trials (two positive + seven negative), that a finding of two positive trials among nine trials could occur by chance and, if so, would not represent substantial evidence of effectiveness.

In some situations, Office of Drug Evaluation 1 has utilized the binomial probability concept as an auxiliary approach to assessing the strength of collective evidence, as illustrated below.

Assume that for each trial, the false positive error rate is controlled at 2.5% (i.e., the typical one-sided significance level of 2.5% or two-sided significance level of 5%). Table 3 shows the chances of reaching a false conclusion of efficacy, based on having two positive trials and various numbers of negative trials, ignoring any other information about the results of the positive or negative trials:

Table 3: Probabilities of Falsely Concluding Efficacy Based on Two Positive Trials; Various Numbers of Negative Trials

Combination of Trials Conducted	Chance of Falsely Concluding Efficacy
2 positive; 0 negative	0.0625%
2 positive; 2 negative	0.36%

2 positive; 4 negative	0.85%
2 positive; 6 negative	1.5%
2 positive; 8 negative	2.3%
2 positive; 10 negative	3.2%

[Source: Provided by FDA statistician Dr. Fanhui Kong.]

The probabilities in the above table may be interpreted as the chances of falsely concluding that an ineffective drug is effective if all the trials have the same design and the same patient population and were conducted in the same manner and at the same time. The table shows that if a sponsor conducted only two trials and both trials were positive, the chance of falsely concluding that an ineffective drug was effective would be very slim, only 0.0625%. Likewise, if a sponsor conducted four trials and two turned out to be positive, then the chance of falsely concluding that an ineffective drug is effective would be only 0.36%.

For the gepirone case, Table 3 suggests that the chance of falsely concluding from two positive trials and seven negative trials is ~2.0%. Of note, however, this line of reasoning assumes that the results of a trial can only be interpreted in binary fashion, i.e., either positive or negative. There is no provision for considering effect sizes, statistical certainty, or design features, and all of this information is simply ignored.

3.2 Maintenance Study ORG28709

Because antidepressants are used chronically, it is important to obtain information on efficacy for long-term maintenance. Typically, new antidepressant drugs are approved on the basis of only short-term efficacy data. Thus, with the initial approval of an antidepressant, the Division negotiates a post-marketing commitment to study maintenance treatment. In practice, maintenance studies of antidepressants have rarely failed to show efficacy.

Study ORG28709, a long-term maintenance study, was the primary basis for sponsor's second resubmission in 2003.

This was a randomized withdrawal study conducted exclusively in Europe (Germany, France, Poland, Finland, and Turkey). Patients with a qualifying diagnosis of MDD were treated with open-label gepirone ER 40-80 mg/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 1:1 to gepirone ER (at the same dose) or to placebo for a double-blind continuation phase that lasted up to one year.

The primary endpoint was the proportion of patients who had relapsed at the end of the double-blind phase. Relapse was 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 Cochran-Mantel-Haenszel (CMH) test, adjusting for center. Time to relapse was a secondary endpoint and was analyzed by log-rank test.

A total of 250 patients were randomized and included in the primary analysis set, intent-to-treat (ITT) population (126 to gepirone and 124 to placebo). In both treatment arms, 44% of patients discontinued early. Of those patients who discontinued, the sponsor reported that 26 in the gepirone group and 35 in the placebo group discontinued because of a relapse. According to the sponsor's analysis, the results on the primary endpoint, the proportion of patents who had relapsed, reached statistical significance in favor of gepirone (top row of Table 4). The results of time to relapse did not reach statistical significance (sponsor's *p*-value = 0.065, FDA's *p*-value=0.089).

Table 4: Summary of Proportion of Patients with Relapse

	Proportion of Patients with Relapse		**CMH *p*-value**
	Gepirone	**Placebo**	
Sponsor's Analysis Results	29/126 (23%)	43/124 (35%)	0.024
FDA's Analyses Results			
(i) Grouping small centers	29/126 (23.0%)	43/124 (34.7%)	0.0971
(ii) Grouping centers into countries	29/126 (23.0%)	43/124 (34.7%)	0.0805
(iii) Grouping small centers and reclassifying 5 patients as relapses	34/126 (27.0%)	43/124 (34.7%)	0.3302
(iv) Grouping by country and reclassifying 5 patients as relapses	34/126 (27.0%)	43/124 (34.7%)	0.3145

[Source: Table 6 of FDA statistical review by Ms. Roswitha Kelly]

FDA's results (Table 4) differed from the sponsor's because of two important reasons:

[1] 5 patients on gepirone who appeared to have had relapses were not regarded as having relapses in the sponsor's analysis:

Prior to unblinding the data, the sponsor identified five gepirone patients who appeared to have met relapse criteria, as they appeared to have been discontinued due to worsening of depression. The sponsor considered them to not have had relapses, however, because had not been designated as having relapses on the case report forms. Re-classification of these five patients as relapses would have rendered the treatment effect not statistically significant.

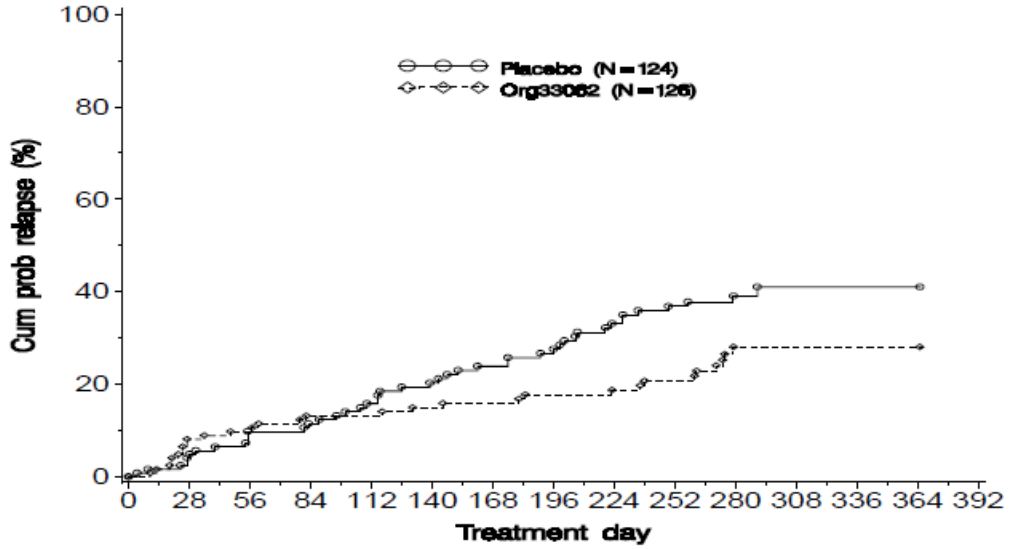
[2] Failure to include all ITT patients in the analysis:

The study was conducted at 29 centers, with the sample sizes per center ranging from 1 to 27 patients. Because of anticipated low numbers of patients at some of the centers, a center-pooling algorithm should have been utilized. But the sponsor did not take this approach; their analysis excluded centers where patients were assigned to only one treatment group as well as centers where there were no relapses. Consequently, 32 ITT patients were not included in the sponsor's primary analysis.

> In order to bring these 32 ITT patients back into the analysis population, FDA adopted two approaches to grouping: (a) pooling centers together to form a pseudo center where only one treatment group was represented, or where there were no relapses, or where there were no more than 4 patients; (b) grouping original centers into countries, i.e., CMH adjusted for country instead of center. After these re-groupings, the primary endpoint lost statistical significance.

For a randomized withdrawal study such as this, the review team noted that a time-to-relapse analysis is the more appropriate analysis, and the one that the Division recommends for this type of study. The CMH test comparing the proportions relapsed does not properly address censored patients, in that it implicitly treats them as successes. The result of time-to-relapse (secondary endpoint) also did not reach statistical significance with a p-value of 0.065 per the sponsor's analysis. Of course the p-value was even less significant when the additional five relapsed patients on gepirone were considered to have had a relapse in the analysis. Figure 2 displays the sponsor's cumulative proportion of patients who had relapsed. This figure does not include the five relapsed patients on gepirone identified by the FDA statistical reviewer. The reviewer noted that the results slightly favored placebo through the first 84 days of the double-blind period; this trend was more apparent when the relapse status was corrected for the additional five relapsed patients on gepirone (data not shown).

Figure 2: Sponsor's Cumulative Proportion of Patients Who Had Relapsed over Time



The presented days were with respect to the start of the continuation phase.
Log-Rank Test p-value = 0.065.
[Source: Figure 2 of "28709-2003.pdf" in sponsor's 2003-12-23 submission]

In its third NDA re-submission in 2007, the sponsor provided their re-analysis results and still sought a claim of maintenance treatment. They modified the original ITT population (N=250) by excluding a total of 40 subjects (22 on gepirone and 18 on placebo) who were considered protocol violators. The review team did not find this *post hoc* elimination of 40 patients from this study after blind breaking to be credible or valid. In addition, per FDA's analyses (per protocol results in Table 5 below), even this *post hoc* reevaluation did not demonstrate a statistically significant result, although the sponsor reported significant *p*-values. The fact that this trial was negative is significant, in that antidepressant trials of this design rarely fail to show a drug effect. Although it is not common to have maintenance data prior to approval, having a negative maintenance study prior to approval negatively impacts the assessment of efficacy.
In the sponsor's 2012 Informal Appeal, they agreed that antidepressant studies of this type are usually successful if the studies are designed and conducted properly. Importantly, they acknowledged the flaws in the analyses submitted by the former sponsor Organon in 2003, i.e., failure to count the five additional relapses on gepirone patients and failure to include 32 ITT patients in the analyses. They pointed out that this study had not been conducted properly by the former sponsor and that its negative results were caused by design flaws, poor protocol compliance, and careless mistakes by investigators – not lack of efficacy.

The sponsor further argued that not all patients randomized to the double-blind phase were "true" responders. As such, they re-analyzed the data using different definitions of responders. Although all of their re-analyses yielded significant *p*-values, the review team disagreed with sponsor's results for the following reasons:

[1] failure to correct the relapse status for the five relapsed patients on gepirone;

[2] failure to include 32 ITT patients who came from centers that had only one treatment arm represented or had no relapses;
[3] failure to remove all patients who should have been removed according to their various *post-hoc* definitions of true responders.

After these corrections were made, the *p*-values were not statistically significant (Table 5). Importantly, the review team questioned the validity of re-defining the analysis set or responders after data unblinding.

Table 5: Summary of Proportion of Patients Who Had Relapsed – Primary Analysis and Post-Hoc Analyses

	Sponsor's Analysis Results			FDA's Analysis Results		
	Gepirone	**Placebo**	***p*-value[4]**	**Gepirone**	**Placebo**	***p*-value[4]**
Original ITT (Primary Analysis)	29/126 (23%)	43/124 (35%)	0.024	34/126 (27%)	43/124 (35%)	0.36
Post-Hoc Analyses						
Per Protocol	25/104 (24%)	41/106 (39%)	0.023	25/104 (24%)	40/106 (37%)	0.11
Re-Defined Non-Responders[1]	22/126 (18%)	40/124 (32%)	0.007	26/118 (22%)	40/121 (33%)	0.17
Re-Defined Non-Responders[2]	22/126 (18%)	42/124 (34%)	0.003	26/118 (22%)	42/123 (34%)	0.11
Re-Defined Non-Responders[3]	25/126 (20%)	42/124 (34%)	0.013	29/121 (24%)	42/123 (34%)	0.25

[1] Excludes relapses on 1st visit after randomization, i.e., 11 patients were removed.
[2] Excludes relapses on 1st visit after randomization if response was confirmed prior to randomization, i.e., nine patients were removed.
[3] Includes subjects with 50% drop in HAMD-17 prior to randomization as responders, i.e., six patients were removed from the analysis.
[4] *p*-value based on CMH test adjusted for center.
[Source: Table 8 of FDA Statistical Review by Dr. Yeh-Fong Chen]

4 SUMMARY

The development program for gepirone ER for the treatment of MDD has occurred over two decades and three FDA regulatory review cycles. At the conclusion of each review cycle, DPP/ODE1 concluded that the sponsor had not presented data sufficient to meet the substantial evidence standard required under the Food Drug and Cosmetic Act to support approval. The sponsor of gepirone ER has appealed the most recent "not-approvable" decision to the Director of the Office of New Drugs based on their conclusion that they have provided substantial evidence of effectiveness of gepirone ER in the treatment of MDD and, therefore, the NDA should be approved once other issues, such as product manufacturing and product labeling, are resolved.

In his interim response to the request for formal dispute resolution (FDRR) the OND Director noted that the issues in dispute involve complex analyses and interpretations of the data provided in the NDA.[4] These issues include the fact that the application includes two "positive" adequate and well-controlled trials of gepirone ER in the treatment of MDD, and these trials demonstrated treatment effects compared to placebo similar to those seen in trials of approved anti-depressants. Despite the two positive adequate and well-controlled trials, the NDA also contains numerous trials that failed to demonstrate efficacy for both early immediate-release formulations and the proposed to-be-marketed extended-release (ER) formulation.

The OND Director noted that if he concurred with the sponsor's conclusion that substantial evidence of effectiveness had been provided this would effectively clear the way for NDA approval since the 2007 not-approvable letter did not include other clinical or pre-clinical deficiencies that needed to be addressed prior to approval of the NDA. The OND Director also noted that Section 918 of the FDA Amendments Act (FDAAA) of 2007 states that "prior to approval of a drug no active ingredient….of which has been approved in any other application under this section or section 351 of the Public Health Service Act, the Secretary shall…refer such drug to a Food and Drug Administration advisory committee for review at a meeting of such advisory committee." The OND Director, therefore, concluded that the NDA for gepirone ER should be presented to a public advisory committee meeting for review prior to his decision on the appeal.

The issues that will be presented to the AC for discussion, recommendations, and formal votes include the following:

- Interpretation of the statutory standard for substantial evidence of effectiveness in a situation where the sponsor has provided two positive adequate and well-controlled trials in the context of numerous other negative/failed trials. The FD&C Act and FDA guidance on providing clinical evidence of effectiveness do not address this situation. Although negative and failed trials have been seen in other development programs for

[4] John K. Jenkins. Interim Response To Appeal- Input Needed From Advisory Committee. June, 1, 2015. ***(****See appendix 5.16, i.e., page 299 of FDA Briefing Document)*

MDD, the gepirone ER case has been considered highly unusual in the experience of DPP/ODE1. The large number of negative/failed trials has led to a conclusion that the two positive trials could have occurred by chance and that substantial evidence of effectiveness has not been provided.

- Interpretation of the meaning of "assay sensitivity" of a trial and its use to conclude that a trial was "negative" or "failed." The current case has led to discussions within the FDA on whether a finding of assay sensitivity should be restricted to the pre-specified primary endpoint and analysis methodology or can be expanded to include secondary endpoints and analyses, even if the primary endpoint failed, and how to interpret a finding that the active drug is nominally superior to the test drug when both the active and test drugs fail to show superiority to placebo.
- Evaluation of methods to assess positive and negative/failed trials in a development program. Possible methods include an approach of counting trials (e.g., 2 positive out of 9 total trials) to assess the likelihood that the findings occurred by chance. Another approach could include application of some method to synthesize the information available from all trials, such as a meta-analysis. While FDA does not accept a meta-analysis across a development program as the basis for demonstration of substantial evidence of effectiveness, a meta-analysis may provide useful information to integrate the effects seen across all trials in interpreting a finding of two positive trials in a situation where numerous negative/failed trials are included in an NDA.
- Evaluation of the data provided by the sponsor for gepirone ER to assist the Agency in determining whether substantial evidence of effectiveness has been provided, whether the safety profile has been adequately evaluated, whether the benefit risk profile of gepirone ER is favorable for approval, and what, if any, additional data are required pre- or post-approval.

FDA looks forward to hearing the committee's advice on these challenging topics.

5 APPENDICES

5.1 Memo – Russell Katz - Non-Approval Recommendation (03/08/2002)

MEMORANDUM

DATE: March 7, 2002

FROM: Director
Division of Neuropharmacological Drug Products/HFD-120

TO: File, NDA 21-164

SUBJECT: Recommendation for Action on NDA 21-164, for the use of Gepirone hydrochloride Extended-Release Tablets in Patients with Major Depressive Disorder (MDD)

NDA 21-164, for the use of Gepirone hydrochloride Extended-Release Tablets in Patients with Major Depressive Disorder (MDD), was submitted by Organon, Inc., on 5/18/01. It had previously been submitted on 9/30/99, but the division refused to file the application for review; the reason for this action was detailed in a letter to the sponsor on 11/30/99. Specifically, prior to that submission, we had informed the sponsor that marketing of the ER formulation could be supported by a single "positive" controlled trial with this formulation if there were at least 2 "positive" controlled trials with the immediate release formulation, which the sponsor had asserted there were. In addition, in discussions prior to the submission, the sponsor asserted that they did have a "positive" trial with the ER formulation.

However, upon receipt of the application, the review team noted that the single "positive" trial was, in fact, one of two centers of a planned 2 center, multi-center trial, that, when analyzed as per protocol, failed to yield a statistically significant between-group difference, by the sponsor's admission. Because the division believed that this nominally significant post hoc finding at this single center did not constitute a bona fide "positive" trial, the application was not filed for review. The sponsor subsequently performed new trials with the ER formulation, and the application was re-submitted on 5/18/01, and filed for review.

The application has been reviewed by Dr. Earl Hearst, medical reviewer (review dated 2/19/02), Drs. Roswitha Kelly and Kooros Mahjoob, statisticians (review dated 3/4/02), Dr. Tarek Hammad, safety reviewer (review dated 2/26/02), Dr. Judy Racoosin, safety team leader (memo dated 2/28/02), Dr. Linda Fossom, pharmacologist, Dr. Sherita McLamore, chemist (reviews dated 1/31/02 and 2/26/02, Dr. Gerald Fetterly, Office of Clinical Pharmacology and Biopharmaceutics (review dated 2/19/02), and Dr. Thomas Laughren, Psychiatric Drugs Team Leader (memo dated 3/5/02).

While Dr. Hearst recommends that the application be considered Approvable, Drs. Kelly and Mahjoob have concluded that the application contains only one adequate and well-controlled trial that yielded a reliable statistically significant

between treatment difference, and Dr. Laughren recommends that the application be considered Not Approvable.

Dr. Laughren's memo provides a comprehensive, detailed, yet concise summary of the relevant data in the application.

Briefly, the sponsor presents the results of 18 controlled trials, 10 performed with the IR formulation, 8 with the ER formulation (the large number of studies is related to the fact that this drug was originally developed by Bristol Myers Squibb, who started many trials and then decided to abandon the project). Of these studies, the sponsor proposes that four are adequate and well-controlled trials that, taken together, provide substantial evidence of effectiveness for Gepirone as a treatment for MDD.

Study 001 was an 8 week trial in which patients were randomized to Gepirone ER 20-80 mg/ day or placebo. The primary outcome was change from baseline on HAMD-17. This study yielded a between-treatment p-value of 0.018 for the LOCF analysis.

Study 003 was an 8 week trial performed at a single US site in which patients were randomized to receive Gepirone IR 10-120 mg/day, given BID, or placebo. Patients in this trial were diagnosed with MDD with atypical features. Critically, the mean HAMD on entrance into the trial was about 13 (compared to >20 in the other trials in this application, as well in other studies in other anti-depressant development programs). The primary outcomes in this trial were change from baseline in HAMD-17 total score, and percent of responders, defined as patients with a score of 1 (very much improved) or 2 (much improved). The p-values for the drug-placebo contrasts were 0.009 and 0.009.

Study 001B was an 8 week study performed in 3 US sites in which patients were randomized to Gepirone IR 5-60 mg/day given BID, Gepirone 10-120 mg/day given BID, or placebo. The primary outcomes were as in Study 003. The p-values for the drug-placebo contrasts were significant for both dose-range groups on both outcomes. However, there was a very significant treatment-by-center interaction. The following display describes the results by center for the HAMD-17, LOCF analysis:

	Carmen (N=89)	Haggerty (N=89)	Cole (N=28)
Gepirone 5-60	-10.5	-11.3	-8.2
Gepirone 10-120	-10.1	-10.2	-13.6
Placebo	-8.5	-9.5	-1.1

Analyses for the HAMD-17 excluding data from the Cole site yielded p-values of 0.3 (Gep 5-60 vs placebo) and 0.8 (Gep 10-120 vs placebo).

Study 002 was a randomized withdrawal trial in which patients were treated with open-label Gepirone IR for 6 weeks. Responders were re-randomized to continue on their previous dose of Gepirone IR (10-90 mg/day given BID) or placebo for an additional 6 weeks. The protocol did not specifically identify primary outcomes, but did state that change from baseline in HAMD-17 and percent of responders on CGI were "important" endpoints. The p-values for the drug-placebo contrasts were 0.08 and 0.24, respectively. However, the sponsor presented the results of analyses of time to reaching 6 different exit criteria, all of which were apparently constructed after the data had been examined. The results of the drug-placebo contrasts for 4 of these outcomes were nominally significant (including one of the six that the sponsor considered most important).

As noted, the sponsor submitted the results of 14 short term (6-8 weeks) additional controlled trials. None of these studies yielded statistically significant between-treatment differences on their primary outcomes, save for one (2486), with a 70% dropout rate. Seven of these studies included an active control (4 with the IR formulation, 3 with the ER formulation; all studies utilized appropriate doses of the active control); in 5 of these studies, the active control group was not distinguished from placebo. In two studies (022 and 028), the active control was significantly superior to placebo; the mean Gepirone dose in the first was about 15 mg/day, and in the second about 13 mg/day.

As Dr. Laughren notes, many of these 14 studies employed relatively low doses. As he points out, any trial that limited the maximum daily Gepirone dose to 40 mg was "negative", and in most of these studies, the mean daily dose (they were all flexible dose range trials) was below 20 mg. In addition, a number of the studies were stopped by the sponsor before they reached their originally planned total sample size.

However, as Drs. Laughren, Kelly, and Mahjoob note, there were a number of studies that appeared to be well-designed and which utilized doses of Gepirone which were comparable to those used in the studies which the sponsor asserts demonstrate effectiveness. While it is not unexpected that a certain number of trials of effective anti-depressants will not distinguish drug from placebo, there are at least 4 such trials in this application.

I completely agree with Dr. Laughren's reasoning for concluding that the sponsor has not submitted substantial evidence of effectiveness for Gepirone ER as a treatment for MDD.

Specifically, Study 001 is a well-controlled trial that is one source of evidence that could contribute to a finding of substantial evidence of effectiveness.

However, the other studies, while ostensibly yielding significant between-treatment differences, do not, in my view, support the conclusion that gepirone is effective as a treatment for patients with MDD.

Study 003 enrolled patients that were clearly not comparable to the other patients enrolled in studies in this program (or in other anti-depressant development programs). While the sponsor asserts that these patients were appropriately diagnosed as having MDD, the extraordinarily low mean baseline HAMD scores would suggest otherwise. It is clear, given this data, that these patients were not typical MDD patients, and therefore I cannot consider this study as evidence that the treatment is effective as a treatment for MDD. I also agree with Dr. Laughren that the reasons cited by Drs. Kelly and Mahjoob for questioning the results of the trial (single center with a small sample size and about a 40% dropout rate), in and of themselves, are not sufficient for discounting the results of this trial.

Study 001B yielded "positive" results that were entirely due to the outcome at one of the 3 sites. This site was by far the smallest site (enrolling 1/3 the number of patients at each of the other 2 sites), 7/9 placebo patients discontinued, and, most critically, the placebo response seen at this site was extraordinarily and uniquely small; it was essentially non-existent. This highly anomalous placebo response makes the result at this site unreliable in my view. In such a case, I would argue that a reliable conclusion can be reached only on the basis of re-analysis of the 2 remaining centers; this analysis does not yield statistically significant between-treatment contrasts.

Finally, Study 002 was designed and conducted as a typical randomized withdrawal study of the sort ordinarily performed to demonstrate long-term effectiveness (although the treatment periods were fairly short). I believe that such a trial could reliably support a short-term effectiveness claim as well. However, in this case, the primary outcomes in the protocol (as best we can tell) were not of the sort normally employed in this type of trial. That is, the sponsor did not designate time to relapse as a primary measure of effectiveness; rather, the protocol suggests that the primary the outcomes were more typical of short-term studies, namely change from baseline in HAMD and proportion of CGI responders. Such a choice for this design is not inappropriate, but the results were not significant. The sponsor did, retrospectively, create 6 different failure criteria, all constructed and chosen after the original study had been unblinded and analyzed. Even in this case, only 4 reached nominal significance, and, as noted by Dr. Laughren any or all of these would likely have been considered appropriate had they been prospectively designated, so there is no valid reason to ignore the 2 negative contrasts. The fatal flaw in this trial, though, is the creation and analysis of "primary" outcomes after the trial data have been unblinded and the contrasts on the protocol specified primary outcomes were not significant.

Although I do not believe that the sponsor has provided substantial evidence of effectiveness at any dose, it is likely that if they ultimately do provide such evidence, the effective dose will probably be greater than 40 mg/day. As the review team has noted, the sponsor has not submitted sufficient long-term safety

data at this dose to adequately characterize the long-term safety. At the moment, there appear to be no adverse events seen at what might turn out to be appropriate doses to preclude ultimate approval, although the presentation of the safety data has been marred by a number of problems, including an inadequate presentation of the adverse event data separately for the ER and IR formulations. There are other issues that require clarification, and the sponsor should be requested to address these (see Dr. Laughren's and Dr. Racoosin's memos on this point). The one additional point I would add is that there were a number of patients whose final laboratory values were abnormal; I believe we should ask the sponsor to obtain follow-up for these patients.

There are additional points that need to be made in our action letter, including a number of CMC issues, the finding that the proposed tradename (Ariza) has been found to be unacceptable (similarity to Arava; we informed the sponsor of this in a letter dated 1/14/02), additional biopharmaceutics requirements, including the fact that the 40, 60, and 80 mg tablets have not been shown to be equivalent to the 20 mg tablet, and an additional pharmacology requirement (a phase 4 commitment to perform an additional in vitro chromosomal aberration test).

The clinical program presented in this application is remarkable in a number of aspects. The sponsor has provided results of 18 controlled trials, only one of which can be considered, in my view, unambiguously "positive" by the usual standards. While a number of these studies randomized patients to 2 dose groups, these groups were not fixed doses but flexible ranges, and often the ranges for the 2 dose groups overlapped with each other. This resulted in a very large development program with no useful dose response data; this serious deficiency in the development program might very well have been responsible for the lack of "positive" studies.

For the reasons cited above, then, I believe that the sponsor has not provided substantial evidence of effectiveness of gepirone ER in the treatment of patients with MDD. I recommend that the Agency issue a Not Approvable letter with an explicit requirement for at least one, fixed dose controlled trial of, ideally, the ER formulation. Dr. Laughren also recommends that we issue a Not Approvable letter, but that we also issue labeling. This would be quite unusual, and I would recommend that we not issue labeling with the letter, primarily because, if the drug is ultimately approved, the final label might be considerably different from what we wish to propose at this time (although I acknowledge that certain sections are unlikely to change significantly).

I wish to address one final point.

As Dr. Laughren has described, we had originally informed the sponsor that a single "positive" trial with the ER formulation, in the context of at least 2 such trials with the IR formulation, would support the approval of the ER tablet. He

notes in his memo that even if one of the 3 submitted IR studies were “positive”, in his view this would be inadequate to support approval of the ER tablet, because there would be no replicated finding for either of the dosage forms.

I believe I would entertain the possibility that a single “positive” study with each formulation could be considered sufficient to support the approval of both formulations. Such an outcome could be argued to provide independent replication of the finding that the moiety is effective. Indeed, there are precedents in this division for similar, though not identical, actions.

For example, we have granted claims for monotherapy and adjunctive therapy for drugs to treat patients with Parkinson’s Disease on the basis of a single study in each of these settings. I would be willing to consider an argument that a single study with each gepirone formulation would support approval of both as being analogous to the former situation. However, because I believe that the sponsor needs to provide at least one additional trial to support approval of gepirone ER, it would be least controversial for that study to be performed with the ER tablet.

Russell Katz, M.D.

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

/s/

Russell Katz
3/8/02 08:41:06 AM
MEDICAL OFFICER

5.2 Memo – Robert Temple - Non-Approval Recommendation (03/14/2002)

MEMORANDUM

DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

DATE: March 13, 2002

FROM: Director, Office of Drug Evaluation I, HFD-101

SUBJECT: NDA 21-164 (Gepirone, Organon)

TO: Director, Division of Neuropharmacologic Drug Products, HFD-120

I concur in the NA action, but I have a few comments (and have modified the letter somewhat):

1. I am troubled by the lack of any D/R information. Failure of studies with doses < 40 mg/day may give some idea of the dose needed for effectiveness (although there were many failures at higher doses as well and note that many patients in some of the "positive" studies didn't get to 40 mg), but we have little or no data within the dose range where gepirone might work. I therefore believe the additional ER study should study several fixed doses, such as 40, 80, 120. You could refer to ICH-E4, a U.S. adopted guideline. It also seems at least possible that the mixed success is a result of substantial PK variability. It would not be a bad idea to get trough blood levels to allow at least a retrospective look at C/R relationships. I could possibly be dissuaded from this view if there were some persuasive reason to titrate everyone from (say) 20 mg, but even in that case you need to know what dose to go to.

2. Whether a molecule is effective or not does not necessarily tell you that a particular dosage form and dose work, although assurance can perhaps be gained by PK modeling. In the present case, the marginal results suggest at a minimum that dose could matter a lot. I believe the additional study needed should therefore use the to-be-marketed product, i.e., the ER form. Note (Table 1) that some of the $\leq$ 40 mg studies (105-057, 105-078) had mean doses (30-34 mg) not so different from the higher dose studies [03A7A-003 (mean 41); 03A7C-001B (mean 33); 03A7A-002 (mean 40); 03A7C-001A-2490 (mean 47); 03A7C-001A-2486 (mean 47)]; so not much about dose seems clear, and not all the low dose (<40 mg) failures can be dismissed.

3. We need to pin down which of the Clin Pharm deficiencies are really needed. I have removed the request for an in-vitro study of 3A4 interactions (they already have a in – vivo study of ketoconazole) but left in a request to study 3A4 induction (the letter says in vitro but I don't believe there is such a method). Do we now insist on such a study of all drugs (I believe current guidance does not say this). As induction of 3A4 would lead to lower gepirone blood levels over time (it's 3A4 metabolized), perhaps blood levels over time in the further study would be sufficient.

Robert Temple, M.D.

cc:
Orig.
HFD-120
HFD-120/P David
HFD-101/R Behrman
HFD-101/R Temple
draft:sb/3/12/02
final:sb/3/13/02
filename:Gepirone_MM_Mar02.doc

Table 1
18 placebo-controlled

		(approx) n/gp	Dose	Result
	ER			
	CN 105-052, 053, 064	15-35	--	Active failed (053 NS as pooled, though one center SS)
	CN 105-057	150	2-4, 5-10, 10-20, 20-40 (mean 34)	D/R, but no dose worked
--	CN 105-078	45	10-50 (mean 30); 20-100 (mean 53)	Failed (no active)
--	CN 105-083	40	10-50 (mean 30); 20-100 (mean 57)	Failed (no active)
*	134001	100	20-80 (mean 70)	HamD 17 – p=0.018
--	134002	105	20-80 (mean 68)	Failed
	IR			
	CN 105-037, 029, 028	20-60	--	Active control failed
	CN 105-043, 022	60, 70	$\leq$ 40 mg (mean 17, 15)	Cpos for fluoxetine
	03A7C-001A-2486	40	5-45 (mean 22); 10-80 (mean 47)	"Pos," but very high D/O (>70%)
--	03A7C-001A-2496	40	5-45 (mean 25); 10-90 (mean 47)	Failed (no active)
*	03A7A-003	30	10-90 (mean 41)	Baseline HamD = 13.7-.8; Pos HamD 17, p=0.009; but ? patients
*	03A7C-001B	70	5-45 (mean 21) 10-90 (mean 33)	Pos but driven by 1 small study site
*	03A7A-002	35	10-90 (mean 40)	Rand WD; failed 1° endpoints

*Submitted as positive
-Failed (dose $\geq$ 40, no failed active control)

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

/s/

Sandra Benton
3/13/02 09:59:39 AM
TECHNICAL

Robert Temple
3/14/02 06:45:32 PM
MEDICAL OFFICER

5.3 Letter – Robert Temple – Not Approvable Letter (03/15/2002)

DEPARTMENT OF HEALTH & HUMAN SERVICES
Public Health Service

Food and Drug Administration
Rockville, MD 20857

NDA 21-164

Organon, Inc.
Attention: Edna Gilvary, Ph.D.
Regulatory Scientist II
375 Mount Pleasant Avenue
West Orange, NJ 07052

Dear Dr. Gilvary:

Please refer to your new drug application (NDA) dated September 30, 1999, submitted under section 505(b) of the Federal Food, Drug, and Cosmetic Act for Gepirone hydrochloride Extended-Release 20 mg, 40 mg, 60 mg, and 80 mg Tablets.

Reference is also made to an Agency letter dated November 30, 1999, refusing to file this application, and to your resubmission of the above referenced NDA dated May 18, 2001.

We acknowledge receipt of your submissions dated August 31, September 18, October 11, November 13, 21, December 10, 28, 2001, January 9, 15, 23, 24, 31, and February 15, 2002.

We have completed our review and find the information presented is inadequate. The critical deficiency that is the basis of this nonapproval action is the lack of adequate efficacy data to support a claim in major depressive disorder as well as the inadequate amount of long-term safety data. In addition, we have included in the letter other issues that, while not the basis for this action, would need to be addressed, in some cases prior to any final approval action, and in others, postapproval. Because of the lack of substantial evidence of effectiveness and the inadequate amount of long-term safety data, the application is not approvable under section 505(d) of the Act and 21 CFR 314.125(b). The deficiencies are summarized as follows:

NONAPPROVAL DEFICIENCIES

1. Inadequate Efficacy Data

You have not provided substantial evidence of effectiveness for the claim of short-term efficacy for gepirone ER in major depressive disorder (MDD). We acknowledge our earlier discussions with you, at which time we had agreed that a single positive short-term trial with the ER formulation, in the face of independent evidence for the efficacy of the IR formulation in MDD, would be sufficient to support a claim for the efficacy of the ER formulation in MDD. We have concluded, however, that you have not provided such evidence for the IR formulation.

We are in agreement with you regarding the short-term gepirone ER study (134001) and consider this a positive study in support of gepirone ER in MDD. However, we do not feel that any of the 3 gepirone IR studies proposed by you as support for the MDD claim can, in fact, be considered positive studies in support of this claim.

-Study 03A7A-003: While we do not dispute the fact that this study has a positive outcome on the protocol specified coprimary outcomes, we have a serious concern about the characterization of the population studied. While a majority of patients reportedly met RDC criteria for MDD (about 2/3), their baseline HAMD-17 scores belie this assertion. The mean baseline scores of 13-14 are roughly 10 points below the usual scores on this measure in MDD, and this finding raises a serious question of the diagnoses of these patients. While "atypical depression," as defined in this protocol, is in fact described as an accepted "specifier" for MDD in DSM-IV, there is still, in our view, much controversy about what population is actually captured when, as was the case here, "atypical features" are the primary criteria for selection of patients. Clearly, patients with a mood disorder other than MDD could have been recruited, and we think the relatively low HAMD-17 scores are reflective of the diversity of patients in this sample. In fact, many of these patients likely overlapped with patients having GAD, and if this were the case in this study, the outcome would not be surprising, given the fact that a very closely related drug, buspirone, has been shown to be effective in GAD. We therefore do not consider it a study that can support a claim of effectiveness in MDD. You may, of course, choose to challenge this objection. If you do, you will need to describe very clearly how you have determined that at least some of the patients did meet criteria for MDD. We may then ask for an additional analysis of the subgroup of patients who could have been reasonably considered to meet the criteria for MDD.

-Study 03A7C-001-B: While this study appears to have an overall positive outcome on the protocol specified coprimary outcomes, we explored the significant treatment by center interaction, and discovered that the overall positive results are coming from one center, i.e., the smallest center (Cole), while the other centers show little effect.

Efficacy Results by Center on HAMD-17 Total Score for 03A7C-001-B (LOCF)

Carmen (n=89)

	Baseline HAMD-17
Gepirone IR (5-60)	**-10.5**
Gepirone IR (10-120)	**-10.1**
Placebo	**-8.5**

Cole (n=28)

	Baseline HAMD-17
Gepirone IR (5-60)	**-8.2**
Gepirone IR (10-120)	**-13.6**
Placebo	**-1.1**

Haggerty (n=89)

	Baseline HAMD-17
Gepirone IR (5-60)	**-11.3**
Gepirone IR (10-120)	**-10.2**
Placebo	**-9.5**

In fact, an analysis conducted without the Cole center does not yield positive results on the HAMD-17 change from baseline (p=0.28 for the 5-60 mg group and p=0.84 for the 10-120 mg group). A critical consideration in our judgement is the finding of essentially no placebo effect in the Cole center. This is an extremely unusual finding in depression trials, and brings into question any results seen in this center.

-Study 03A7A-002: This study had the usual randomized withdrawal design ordinarily employed to document long-term effectiveness of antidepressants, although it was quite atypical for this design in that both the run-in period and the randomized withdrawal period were quite short. Despite this, it might have provided some indication of antidepressant activity, had there been an effect on the primary outcomes specified in the protocol. In fact, the protocol was not clear on the primary outcome but specified change from baseline in the HAMD-17 as an "important" endpoint. The study report also refers to 6 different definitions of relapse as key endpoints, but these endpoints were apparently defined long after the conduct and analysis of this trial. There were no statistically significant differences favoring gepirone IR over placebo on change from baseline in HAMD-17, or on CGI-I responder status, another outcome that would have been relevant to the claim, and that appears from the protocol also to have been considered an "important" outcome measure. The results on time to relapse favored gepirone 20-90 over placebo for 4 of the 6 definitions of relapse. Although none of the 6 definitions would have been rejected as unreasonable primary endpoints, their post-facto identification, together with failure to show an effect on the primary endpoints, leaves Study 03A7-002 unable to provide primary support for the claim of efficacy of gepirone IR in MDD.

An additional concern for all three gepirone IR studies was that interim analyses were planned in the protocols for these trials, but we have no information about whether or not such analyses were actually carried out. This is an additional source of potential bias for these trials that greatly complicates their interpretation, although we wish to be clear that, even in the absence of any interim analyses, the deficiencies described above make these latter three studies unacceptable.

In addition to the specific problems identified in the "positive" studies presented, there is a general concern about the many negative trials in the overall development program for both the IR and ER formulations. The 4 studies that you have proposed as positive in support of the efficacy claim arise from a total of 18 short-term placebo-controlled trials in depression. We agree with the explanations proposed to explain the negative findings for many of these studies, in particular, the dose likely being too low, and an active control arm that also failed to beat placebo. Nevertheless, 3 of the 14 remaining studies are negative without good explanation. These studies are as follows: 134002, CN105-078, and CN105-083, all for gepirone ER. All 3 studies were sufficiently adequate by design, including gepirone ER dose, duration, population studied, assessment methods, and study conduct otherwise, to be considered relevant to the efficacy of gepirone, and yet all 3 failed to show a benefit for gepirone ER. While we agree that these studies are not readily interpretable, we still think they need to be considered as negative trials for gepirone ER, and therefore, should be considered in the overall benefit assessment for this drug. Thus, in summary, of 4 studies that appear, on face, to be relevant for consideration of the efficacy of gepirone ER, only one yielded a positive result. Although the result seen here (i.e., 1 of 4 relevant studies being positive) is not a result that would lead us to conclude that gepirone ER is ineffective as an antidepressant, we believe one additional positive study for the ER formulation is needed to demonstrate that this formulation has antidepressant efficacy.

The additional study should be a dose-response study. There is at present little data pertinent to the question of dose response. Given the uniformly negative results in studies with a maximum dose of ≤ 40, it is tempting to conclude that doses lower than 40 mg are not likely to be effective. However, it would be highly preferable to have the finding of lack of effect below 40 mg/day come from the same study in which doses above 40 mg/day were shown to be effective, i.e., a dose response study, as well as to explore the effective dose range. If you choose to continue to develop this drug for MDD, you should design a study that looks at different fixed doses of gepirone ER in MDD. In addition, we recommend that you collect, through sparse sampling, plasma level data for gepirone and its two major metabolites, 1-pyrimidinyl-(2-piperazine) (1-PP) and 3'OH-gepirone. These data could help in understanding the relationship between exposure and clinical response.

2. Inadequate Safety Data

Long-Term Safety Data

As conveyed to you in a conference call dated October 11, 2001, your long-term exposure data is inadequate to assess the safety of gepirone. The ICH guidelines indicate that there should be 300-600 individuals exposed to an effective dose for 6 months and 100 individuals exposed to an effective dose for 1 year. While we have concluded that you have not provided substantial evidence of effectiveness for any dose, it is likely that, if the drug is effective, the effective dose is likely to be at least 40 mg/day. Your application contains data showing that there were only 124 subjects exposed to a modal dose ≥40 mg for approximately 6 months and only 35 subjects exposed to a modal dose ≥40 mg for approximately 1 year. At a minimum, you will need to augment this safety database to meet the ICH guidelines for long-term safety data exposure at clinically effective doses as described above.

OTHER REQUESTS AND COMMENTS

1. Safety Update

As part of any resubmission of this application, it will be necessary for you to provide an update on your safety database since your last safety update submission of September 18, 2001. We have already provided you detailed advice on what needs to be included in this safety update in a January 14, 2002 fax transmission. In addition to following that general advice, the safety update should address the following specific issues:

Dizziness

We feel that dizziness is a sufficiently common and important adverse event to justify additional analyses to better understand and characterize this event. At a minimum, we request that you characterize the course of dizziness over time; for example, how long does it last, does the severity change with subsequent episodes, do patients develop tolerance to this symptom. In addition, please provide a discussion of the relationship, or lack thereof, between gepirone-associated dizziness, syncope, objectively measured orthostatic changes, and accidental injury.

Adverse Events Associated with Discontinuation of Treatment

The NDA noted that 761 subjects had "adverse event" identified on the "End of Study" CRF page as the reason for discontinuing from a study; however, 18% (N=138) did not specify the adverse event resulting in discontinuation. Please attempt to re-evaluate those adverse events. One approach would be to go to the AE page in the CRF and identify the outcome for each specific AE. This process affords an additional opportunity to determine which AE led to discontinuation for those patients that didn't have it specified on the "End of Study" page in the CRF.

Adverse Event Tables in Proposed Labeling

In any resubmission of labeling, we ask that you recalculate table II for adverse events associated with discontinuation by lowering the threshold for inclusion of common events leading to discontinuation down to 1%.

For each of the tables in the Adverse Events section of the proposed labeling (II, III, and IV), we ask that you recalculate the placebo AE incidences based on the inclusion of only those patients treated with placebo in the gepirone ER controlled depression trials.

ECG Changes

For controlled trials of gepirone ER in depression, please provide a comparison of the frequency of outliers for QTc duration with a change from baseline of $\geq$30 msec and of $\geq$ 60 msec in each treatment group (gepirone IR, gepirone ER and placebo).

In addition, please clarify which trials were included in the ECG analysis you refer to in your proposed labeling.

Vital Sign Changes

We ask that you conduct additional analyses of the vital signs data. In particular, please provide the risk ratios for the incidence of orthostatic changes in the controlled trials of gepirone ER in depression.

In any resubmission of this NDA, please propose a section for vital signs in labeling.

Weight Changes

The data in the cited appendix in support of your proposed labeling statement regarding weight changes address only the changes in the BMS development program. Please recalculate changes in weight based on the gepirone ER data from short-term controlled depression trials.

Effect on Sexual Function

The data pertaining to sexual function that support your proposed labeling statement regarding sexual function appear to come from AEs reported in the placebo controlled depression trials. Please indicate what self-report instrument was used and provide the data from this analysis. Until you show that the approach used was sensitive to the detection of

sexual dysfunction, we do not believe you can include labeling language suggesting that gepirone is free of this adverse effect. We would be willing to discuss with you the design of studies that could definitively answer this question.

Suicidal ideation/ Suicide Attempt

We ask that you recalculate the rate of suicidal ideation and suicide attempt using person-time exposure in the denominator.

Allergic reaction/ Hypersensitivity syndrome

In order to more thoroughly identify patients whose symptoms may have represented an allergic reaction or hypersensitivity syndrome, we ask that you develop a case definition for allergic reaction/hypersensitivity syndrome, taking into account the multiple symptoms that may be part of such a syndrome. Using this case definition, you should identify patients whose AE profile fits that of an allergic reaction/ hypersensitivity syndrome. This investigation should also include an assessment of eosinophilia in the controlled clinical trials, including mean change from baseline and outlier analyses.

Follow-up for patient with evidence of hepatic dysfunction

We note that there was a patient, 0415 in ongoing study 28709, who had elevated hepatic enzymes. After three months on blinded treatment, the subject's AST, ALT, and alkaline phosphatase values increased to 52, 93, and 151, respectively, from normal values pretreatment. Seven days after stopping study drug, the values were 148, 393, and 485. There was no information about bilirubin values, diagnostic work-up, or outcome of this event. Should this patient turn out to have been on gepirone ER treatment, follow-up on this event will be needed.

Adverse Event Dose Response Analysis

You should provide a dose response analysis using all reported treatment-emergent AEs. In the NDA submission, you used only "treatment-related" AEs as judged by the investigator.

2. Regulatory Status Update

Before resubmitting this application, please provide any new information on the regulatory status of gepirone worldwide, i.e., information available subsequent to the regulatory status update provided in your May 18, 2001 resubmission.

3. Worldwide Literature Update

Before resubmitting this application, please provide an updated worldwide literature search. We note that you have in the NDA submission provided 286 literature references, including links to the full papers for 21 of these. This alone is inadequate. We request that you conduct a comprehensive review of all of the available literature pertinent to gepirone, including papers published since your original literature review, and provide commentary on the relevance, if any, of these published papers to the safety of gepirone.

4. Proposed Tradename

Please refer to our Agency letter dated January 14, 2002, informing you that your proposed tradename of Ariza was unacceptable. Please submit another proposed tradename for review by the Agency.

PRECLINICAL TOXICOLOGY

The *in vitro* chromosomal aberration assay was inadequate because although gepirone was negative for 5-hour treatment, with and without metabolic activation, this negative finding (without activation) should have been followed up with a study using continuous treatment with gepirone (without activation) for ~ 24 hours (1.5 cell doubling times) in accordance with current guidelines. Since the weight of evidence suggests that gepirone is neither genotoxic nor carcinogenic, we are not requiring that this study be repeated; however, it will have to be repeated if it is to be included in product labeling.

CMC

1. At several places in the manufacturing process section you use the term "or equivalent" to describe the equipment to be used. Please clarify.

2. Your specification for [redacted] is NMT [redacted]%. Please update your specification from NMT [redacted]% to NMT [redacted]% for these impurities.

3. The specification for [redacted] is NMT [redacted]%; however, the actual amount of these substances found in the drug product has not exceeded [redacted]%. Please tighten this specification to be more reflective of the data.

4. The specifications for the drug product do not include a specification for the [redacted] content. As the drug product is [redacted], for all future batches of the drug product, please provide a specification for and monitor the [redacted] content in the drug product on release and stability until a significant body of data is obtained to warrant excluding this test from the specifications.

5. DMF 2880 (Sud-Chemie Performance Packaging) is deficient for the [redacted] The DMF Holder is being notified of this by separate letter which includes a list of the deficiencies.

6. You indicate on page 167 of volume 1.10 in the table pertaining to the package insert that the requirements for the dimension and shipment marking are "conforms to specification" however it is not clear what these specifications are. Please clarify.

7. You indicate on page 168 of volume 1.10 in the table pertaining to the folding cartons that the requirements for the color, dimension and shipment marking are "conforms to specification" however it is not clear what these specifications are. Please clarify.

8. In your marketed stability protocol on page 179 of volume 1.13 you indicate that one batch of each strength will be stored annually and tested at intervals of 0, 6, 12, 24 and 36 months and then yearly thereafter if expiration dating extension is desired. This reduced testing is

not acceptable. Please commit to testing at intervals of 0, 3, 6, 9, and 12 months then yearly thereafter.

9. The proposed cartons and blister backing labels for the drug product has Ariza™ (Gepirone HCl) Extended-Release Tablets listed as the name of the drug product. As noted above, this name is not accepted by the Office of Post–Marketing Drug Risk Assessment (OPDRA). Please commit to submitting revised container carton labels information when a new name is agreed upon.

CLINICAL PHARMACOLOGY AND BIOPHARMACEUTICS

1. In the pivotal bioequivalence study comparing different strengths of gepirone ER, the 40 and 80 mg strengths were not shown to be bioequivalent to the 20 mg tablets, which was the tablet strength used in the clinical trials. No adequate data were provided comparing the 60 mg strength to the 20 mg strength of gepirone ER. Based on the available data, the 40, 60 and 80 mg strengths of gepirone ER are not approvable.

2. In the food effect study, the data showed that food has a significant effect on gepirone bioavailability (increase in Cmax approximately of 62%). Prior to resubmitting this application, you should conduct an *in vivo* study evaluating the effects of different meal compositions on gepirone ER pharmacokinetics.

3. A strong pharmacokinetic interaction was observed between gepirone and ketoconazole (5-fold increase in gepirone concentrations). The effect of other CYP3A4 inhibitors that can be potentially coadministered with gepirone ER has not been evaluated. Prior to resubmitting this application, you should conduct an *in vivo* drug interaction study to evaluate the effects of gepirone coadministration with an intermediate inhibitor of CYP3A4, such as verapamil.

4. Prior to resubmitting this application, you should conduct an *in vitro* assessment of (a) potential drug-drug interactions between gepirone and potent CYP3A4 inducers and (b) the drug's ability to induce CYP3A4 enzymes.

5. Prior to resubmitting this application, we recommend that you conduct an *in vivo* drug-drug interaction study with a potent CYP2D6 inhibitor to assess its effects on the pharmacokinetics of gepirone and its metabolites.

6. We request that you agree to change the dissolution specifications at 12 h and 20 h to 65-85% and >85%, respectively.

7. We request that you clarify whether or not plasma gepirone (and any metabolites) concentrations were measured in any of the pivotal efficacy trials following administration of the IR and/or ER formulations of gepirone.

Within 10 days after the date of this letter, you are required to amend the application, notify us of your intent to file an amendment, or follow one of your other options under 21 CFR 314.120. If you do not follow one of these options, we will consider your lack of response a request to withdraw the application under 21 CFR 314.65. Any amendment should respond to all the deficiencies listed. We will not process a partial reply as a major amendment nor will the review clock be reactivated until all deficiencies have been addressed.

The drug product may not be legally marketed until you have been notified in writing that this application is approved.

If you have any questions, call Paul David, R.Ph., Senior Regulatory Project Manager, at (301) 594-5530.

Sincerely,

{See appended electronic signature page}

Robert Temple, M.D.
Director
Office of Drug Evaluation I
Center for Drug Evaluation and Research

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

/s/

Robert Temple
3/15/02 12:50:08 PM

5.4 Memo – Russell Katz - Non-Approval Recommendation (06/15/2004)

MEMORANDUM

DATE: June 14, 2004

FROM: Russell Katz, M.D.
Director
Division of Neuropharmacological Drug Products/HFD-120

TO: File, NDA 21-164

SUBJECT: Recommendation for action on NDA 21-164, for the use of Gepirone Hydrochloride Extended Release (ER) Tablets in Patients with Major Depressive Disorder (MDD)

NDA 21-164, for the use of Gepirone Hydrochloride Extended Release (ER) Tablets in Patients with Major Depressive Disorder (MDD), was submitted by Organon, Inc., on 9/30/99. The Agency issued a Not Approvable letter on 3/15/02; the critical deficiency at that time was noted to be the lack of substantial effectiveness for the ER tablet. In the original application, a total of 18 controlled trials (utilizing either the IR or ER tablets) had been submitted. Out of this array of studies, only one study, Study 134001, a short term study of the ER formulation, was considered a positive study. The sponsor had proposed that three other controlled trials of the IR formulation were positive, but we disagreed (see, for example, my previous memo of 3/8/02 and Dr. Thomas Laughren's memo of 3/5/02). It should be noted that although we agreed that many of the other 14 controlled trials were not adequate trials (some were done at inappropriately low doses, others were discontinued by the previous sponsor), there were three trials of the ER formulation that were not positive for any obvious design or conduct reasons. In addition to this critical deficiency, we noted a lack of adequate long-term safety data at doses of 40 mg/day (the dose thought to be potentially effective). In the letter, we informed the sponsor that they would need to submit another positive study with the ER formulation. Finally, a number of other, non-critical deficiencies were identified.

The sponsor responded to the Not Approvable letter in a submission dated 12/23/03. This submission contained a report of a long-term, randomized withdrawal study. This submission needs some explanation.

After the Not Approvable letter was issued, the division met with the sponsor to discuss the further development of the drug. In one meeting (7/3/02), we determined that an additional one of the original studies submitted by the sponsor (Study 03A7A-003), a short term study of the IR formulation, and one we had originally concluded was not acceptable, was, in fact, acceptable as a positive study (for the IR formulation). In our original thinking, this study was considered unacceptable because it appeared that the sponsor did not enroll patients with MDD, although this study was clearly a positive study by protocol, a

fact we acknowledged in our letter. Specifically, the mean baseline HAM-D 17 scores were about 10 points lower than those of patients typically enrolled in MDD studies (and in the other MDD studies included in this application). These patients were characterized as having "atypical depression", and our letter expressed concern that patients with other diagnoses, in particular GAD, might have been enrolled. However, in our meeting with the sponsor (and based in part on their pre-meeting submission), they convinced us that at least 65-88% of patients in this study met diagnostic criteria for MDD, that the low baseline HAMD scores were related to a few items not relevant for atypical depression, but that they had MADRS scores in the more typical range, that they did not meet criteria for GAD, and that analysis of the subset with a bona fide diagnosis of MDD also was positive for gepirone. Therefore, we had concluded that the application contained one positive study with the ER formulation and one positive study with the IR formulation, but still required that the sponsor submit another clearly positive study with the ER formulation. Indeed, the sponsor noted that a short term trial with the ER formulation (Study 134004, a study that compared gepirone, fluoxetine, and placebo) was nearing completion.

However, the sponsor informed the division in a 4/17/03 submission that this study was a failed study. They asserted, however, that their randomized withdrawal study of the ER formulation, (Study 28709) was a positive study, and they argued that the panoply of positive results (one short term study each with the ER and IR formulations and the long term ER study) should be a sufficient basis for the demonstration of substantial evidence of effectiveness of the ER formulation. We informed the sponsor that we would review such an application (with obviously no assurance that this package would be acceptable).

The sponsor's resubmission has been reviewed by Dr. Earl Hearst, medical reviewer (review dated 5/20/04), Dr. Roswitha Kelly, statistician (review dated 5/6/04), Dr. Sherita McLamore, chemist (review dated 6/14/04), Dr. Sally Yasuda, Office of Clinical Pharmacology and Biopharmaceutics (review dated 5/25/04), Dr. Linda Fossom, pharmacologist (memo dated 4/7/04), Dr. Ni Khin, Division of Scientific Investigations (memo dated 5/21/04), Dr. Tia Harper-Velazquez, Division of Medication Errors and Technical Support (DMETS; review dated 4/30/04), and Dr. Tom Laughren, Psychiatry Drugs Team Leader (memo dated 5/20/04). The clinical review team recommends that the application not be approved. I will very briefly describe the results of Study 28709 and offer the division's recommendation for action on the application.

Study 28709 was a multi-center, multi-national study in which patients with MDD who met responder criteria at 8-12 weeks after treatment initiation (at doses between 40-80 mg/day) in an open-label phase were randomized to continue active treatment or placebo for 40-44 weeks. The primary outcome measure was the proportion of patients who met relapse criteria, analyzed by CMH with adjustment for centers. A relapse was considered to have occurred if the patient had a HAMD-17 of at least 16, or was discontinued because the investigator

determined that the treatment was not effective. Time to relapse was a secondary outcome.

According to the sponsor, the rate of relapse on gepirone was 29/126 (23%) compared to a rate of relapse on placebo of 43/124 (35%). By the sponsor's analysis, the p-value for the primary outcome was p=0.024.

However, as all of the reviewers have noted, two problems complicate this outcome as reported by the sponsor.

Prior to unblinding the data, the sponsor classified 5 additional gepirone patients as having met relapse criteria due to lack of effect, although they had not been so identified in the CFR (a company memo indicated that these patients would be classified as failures). When the data were unblinded, the sponsor noted that these patients had not met the HAMD criterion of at least 16, but at that point it was impossible to verify with the investigators the status of these patients, given that all parties had been unblinded.

When these five patients are included in the analysis, the p-value for the primary outcome becomes p=0.1.

Further, as the team notes, the sponsor excluded an additional 32 patients from the analysis because either 1) they were in centers that enrolled patients in only one treatment arm, or 2) there were no relapses in these centers. When these patients are included in appropriate analyses, the p-value for the between-treatment contrast is p=0.08 - 0.1. When the 5 additional patients described above are included in these latter analyses, the p-value becomes p=.31 - .33.

Analyses of Time to Relapse, a secondary outcome in this study, but the more traditional outcome considered primary in most other studies of this type, yielded p-values of p=0.09 (including the 32 patients at small centers) to p=0.28 (when the additional 5 patients are included).

As the team also points out, Study 134004, the short term trial comparing gepirone, fluoxetine, and placebo, not only failed to distinguish the active drugs from placebo, but, in fact, in this study, gepirone was numerically worse than placebo, fluoxetine was numerically superior to placebo, and the p-value for the fluoxetine-gepirone contrast was p=0.068; the chart below describes the relevant comparison:

Drug	Change from Baseline (Mean HAMD-25)
Gepirone (N=125)	-9.9
Fluoxetine (N=136)	-11.8
Placebo (N=136)	-10.6

Other issues

As Drs. Laughren and Hearst note, the sponsor has provided sufficient long-term safety data at doses of at least 40 mg/day. Further, they have provided adequate responses to the other specific safety issues raised in the Not Approvable letter (which were not reasons for the NA action), although several of the reviewers have additional questions for the sponsor (specifically there remain CMC and OCPB questions, and DMETS finds the sponsor's new proposed names, Variza and Alrize, unacceptable).

COMMENTS

Subsequent to the Not Approvable letter, the sponsor has submitted the results of two additional controlled trials examining the effectiveness of gepirone as a treatment for patients with MDD; Study 134004, a short term trial comparing gepirone, fluoxetine, and placebo, and Study 28709, a randomized withdrawal study. Study 134004 failed, as described by the sponsor, to distinguish gepirone (and fluoxetine) from placebo, but Study 28709, according to the sponsor, is a positive study. In addition, we have agreed with the sponsor that Study 03A7A-003, a short term study of the IR formulation, is acceptable as a positive study. Therefore, the sponsor argues that they have submitted data from three controlled studies (one with the IR, two [one short term and one randomized withdrawal] with the ER) that establish the effectiveness of gepirone ER as a treatment for patients with MDD.

Unfortunately, we cannot agree with the sponsor that Study 28709 is a positive study. As described above, the sponsor had identified 5 patients (all on gepirone) that they considered as having met relapse criteria, while the study was still blinded, but then, ***after the data were unblinded***, decided that these patients should not be included in the analysis, because they could not confirm the patients' status with the investigators. The team concludes, and I completely agree, that these patients appear to have met the criteria for relapse (recall that these criteria include physician decision that the patient had failed treatment, independent of any particular HAMD score), and that they should be included in the analyses; as we have seen, when these patients are included, the results are no longer significant.

Beyond this, the sponsor has also excluded data from 32 patients because they were at centers that enrolled patients only on one treatment, or there were no events at those centers. The team has concluded, and I again agree, that these patients should be included in the analyses, and Dr. Kelly has included these patients in analyses that are fairly typically performed under these circumstances. These analyses are not significant, and inclusion of the 5 gepirone patients discussed above in these analyses yield p-values for the

gepirone-placebo contrasts of about 0.3. For these reasons, then, I consider Study 28709 a negative study.

We are then left with two positive short term studies, one with the ER formulation, and one with the IR formulation. In our Not Approvable letter, we informed the sponsor that they would need an additional positive short term study with the ER tablet; they have not provided such a study. However, they have now presented the results of two positive studies. Is this sufficient to support a finding that there is substantial evidence of effectiveness for the ER formulation?

In answering this question, I would like to make an initial point.

Dr. Laughren has concluded that we have never approved an application for an ER product on the basis of one study with an ER, and one study with an IR, formulation, and that such an array of results should not support approval of the ER. I noted in my memo of 3/7/02 that we have approved applications on the basis of similar reasoning in other contexts, but agreed that we have never approved an application for one formulation based on two trials, one of which was with a different formulation (as is the situation here). I did, however, allow for the possibility that two studies of this sort might support approval of the ER (see my 3/7/02 my reasoning). I still believe that this would be a possibility, all other things being equal.

However, I do not believe that the data before us support approval, even though there are two positive trials.

We now have a total of 5 short term trials with the ER formulation that are capable of demonstrating effectiveness, only one of which is positive. Further, in the most recent such trial, fluoxetine is almost statistically significantly superior to gepirone. Additionally, the new randomized withdrawal study, when appropriately analyzed, does not yield a statistically significant difference in favor of gepirone (indeed, the most appropriate analyses yield p-values of about 0.3 for the gepirone-placebo differences).

I believe that this panoply of results raises serious doubts about the effectiveness of gepirone, and especially about the effectiveness of gepirone ER. Although there are two “positive” studies, the existence of many well designed and conducted studies that fail to find an effect is troubling, and I agree with Dr. Laughren that it is difficult to imagine what else the sponsor could do to convince us that the drug is effective. However, I suppose that an additional clearly positive short term study as well as a clearly positive long term maintenance, randomized withdrawal study would provide the sufficient additional data necessary to support the approval of gepirone ER.

For these reasons, then, we recommend that the attached Not Approvable letter be sent to the sponsor.

APPEARS THIS WAY ON ORIGINAL

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

/s/

Russell Katz
6/15/04 09:36:23 AM
MEDICAL OFFICER

5.5 Memo – Robert Temple - Non-Approval Recommendation (06/22/2004)

MEMORANDUM

DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

DATE: June 22, 2004

FROM: Director, Office of Drug Evaluation I, HFD-101

SUBJECT: Gepirone, NDA 21-164, ER Tablets for MDD

TO: File

I concur with the Division's view that NDA 21-164 is not approvable. The thorough failure of 4/5 ER studies, some at the dose that was successful in study 134001, is impressive and 134004 gave a result numerically worse than placebo and almost significantly worse than Prozac. [I can't find any reference to the dose used in this study, but it appears that no serious D/R study has been conducted.] Also, and given the history of success of studies of this design, study 28709, a randomized withdrawal study, also failed once the 37 excluded patients were put back (including 5 gepirone relapses). Such failures are very unusual for effective agents. There thus seems real doubt as to whether gepirone is effective and no doubt that its effectiveness has not been shown. It is certainly possible that the variable blood levels associated with gepirone's 3A4 metabolism are part of the difficulty. It is conceivably a case where a D/R study accompanied by blood levels and an attempt at a C/R analysis could be useful.

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

/s/

Robert Temple
6/22/04 04:20:40 PM
MEDICAL OFFICER

5.6 Letter – Robert Temple – Not Approvable Letter (06/23/2004)

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

Food and Drug Administration
Rockville, MD 20857

NDA 21-164

Organon, Inc.
Attention: Edna Gilvary, Ph.D.
Regulatory Scientist II
375 Mount Pleasant Avenue
West Orange, NJ 07052

Dear Dr. Gilvary:

Please refer to your new drug application (NDA) dated September 30, 1999, submitted under section 505(b) of the Federal Food, Drug, and Cosmetic Act for Gepirone hydrochloride Extended-Release 20 mg, 40 mg, 60 mg, and 80 mg Tablets.

We acknowledge receipt of your submissions dated December 23, 2003, February 9, February 27, March 19, April 5, and May 19, 2004.

The December 23, 2003 submission constituted a complete response to our March 15, 2002 action letter.

We have completed our review and find the information presented is inadequate. Therefore, the application is not approvable under section 505(d) of the Act and 21 CFR 314.125(b). The deficiencies are summarized as follows:

Nonapproval Deficiencies

You have still not, in our view, provided substantial evidence of the effectiveness of gepirone ER in the short-term or longer-term treatment of major depressive disorder (MDD). We acknowledge our earlier discussion with you on July 3, 2002, at which time we agreed that you had produced evidence of short-term efficacy for gepirone ER in study 134001 and for gepirone IR in study 03A7A-003. At that meeting, you indicated that the results from another short-term ER study (134004) would soon be available, as well as results from a longer-term randomized withdrawal study (28709). We indicated that, if study 134004 were robustly positive, you would have sufficient efficacy data for filing the application. We subsequently learned, however, in an April 17, 2003 package outlining your proposed response to our March 15, 2002 nonapprovable letter, that study 134004 failed to distinguish gepirone from placebo. Although at first glance, the failure of fluoxetine to show superiority to placebo suggests that study 134004 lacked assay sensitivity, fluoxetine was in fact nearly significantly superior to gepirone, a very unusual outcome in comparisons of two active antidepressants and a more troubling result. You indicated, however, that your randomized withdrawal study (28709) was positive, and argued

that this combination of two positive short-term studies, one for each of the two formulations, together with one positive gepirone ER long-term study, should be sufficient to file the application. In a July 14, 2003 teleconference, we agreed only that we would probably be willing to file such an application, without reaching any agreement that this combination of studies would represent sufficient support of the effectiveness of gepirone. We noted that part of the difficulty was a preponderance of negative studies for gepirone in the application, making the overall program very weak indeed.

Now that we have had an opportunity to review the data for study 28709, we do not agree that it is a positive study. It was a fairly typical randomized withdrawal study in adult depression (MDD), with an 8-12 week open label phase, followed by a double-blind phase in which patients responding to open treatment with gepirone (HAMD-17 total score $\leq$ 8, at week 8 or week 12) were randomized to either continuation of gepirone (at the same dose that was associated with a "response") or to placebo, with a 40-44 week period of observation for relapse (defined as a HAMD-17 $\geq$ 16, or discontinuation due to lack of efficacy as determined by the investigator). The primary outcome was the proportion of patients who had relapsed at the end of the observation period, and the primary analysis was the Cochran-Mantel-Haenszel test (CMH), with adjustment for centers. Time to relapse was a secondary endpoint. Survival curves were estimated using Kaplan-Meier methodology and the log-rank test was used to compare survival distributions. Both analyses were based on a modified intent-to-treat sample, i.e., all patients randomized who received at least 1 dose of assigned treatment and who had at least 1 post baseline efficacy evaluation.

According to your analysis, the results on the primary endpoint, rate of relapse at study end, favored gepirone:

Gepirone:	29/126 (23%)	
Placebo:	43/124 (35%)	p=0.024

However, there were two important problems with your analysis:

1. 5 patients on gepirone appeared to have had relapses, but were not included in your analysis:

   Prior to unblinding the data, you identified 5 gepirone patients who appeared to have met relapse criteria, since they were discontinued due to worsening of depression, but had not been so designated on the CRF (as noted, discontinuation due to lack of efficacy was one of two criteria for relapse). In fact, you have produced an internal memo, prepared prior to unblinding, indicating that these patients would be redefined as having relapsed, and they were so redefined. Subsequent to unblinding and analysis, you discovered that these patients' scores had not met the $\geq$ 16 criterion, but it was too late to query investigators, since they had also been unblinded. Nevertheless, you then decided to exclude the patients as relapsers, on the grounds that the redefinition had not been done by formal amendment and it had not been possible to query investigators. As it turns out, this redefinition is critical to the outcome, since, if these 5 patients are included as relapsers, the results are as follows:

Gepirone:	34/126 (27%)	
Placebo:	43/124 (35%)	p=0.101

We consider it inappropriate, after looking at the results of the analysis, to decide not to include these patients as relapsers, when they had already been quite reasonably reclassified as relapsers by you prior to unblinding the data. It seems obvious, on face, that these 5 patients who were discontinued for worsening depression should be counted as relapsers, given that their discontinuations were for "worsening of depression," whether or not there was an opportunity to query the investigators to try to verify this result. Thus, we believe the appropriate analysis is the one that includes these 5 patients as having relapsed.

2. Failure to include all ITT patients in the analysis

We also note that you excluded 32 patients from the analysis (CMH) because they came from centers that had patients in only 1 treatment arm, or had no relapses. A more appropriate analysis, grouping these centers, gives a nonsignificant result: p=0.10, with grouping of small centers; p=0.08, with grouping by country. These results are negative, even with your exclusion of the 5 relapsed patients we feel should be included. When these 5 patients are included in these analyses with appropriate grouping of centers, the results are far from significant (p=0.33 and p=0.31, respectively, for groupings by small center and by country).

The results on the secondary endpoint, time to relapse, also did not favor gepirone, with a p-value of 0.089. When the additional 5 patients are included, the p-value is 0.28. Generally for this type of study, we consider the time to relapse analysis as the more appropriate analysis, and the one that we always recommend. The CMH analysis of proportion relapsed does not properly address censored patients, in that it implicitly treats them as successes.

The finding of only two positive short-term trials for gepirone, one for IR (03A7A-003) and one for ER (134001), out of a total of 19 short-term placebo controlled trials is a significant concern. While we have accepted your explanation for the failure of many of these trials, we noted in our March 15, 2002 letter that 3 of the short-term ER studies appeared adequate, on face, and would have been expected to succeed. Thus, the finding of only one of four adequately designed short-term ER studies having a positive outcome was not reassuring. This concern was the basis for our indicating that an additional "robustly positive" ER study would be needed. The result of the most recent ER study (134004) adds to our concern. Although this study is on face a failure for both fluoxetine and gepirone, it actually favors fluoxetine over gepirone. The results for gepirone are actually numerically worse than placebo, while the results for fluoxetine are numerically superior to placebo, and the p-value for the fluoxetine/gepirone contrast on the primary outcome (HAMD-25) is strongly trending toward statistical significance (p=0.068). Thus, of 5 short-term seemingly well- designed ER studies, only one of the five was positive. Given this very marginal set of results for the gepirone short-term efficacy data, a robustly positive result for the randomized withdrawal study became even more pressing. As noted, we do not consider 28709 to be a positive study.

In summary, you have not provided substantial evidence that gepirone ER is effective in the treatment of MDD. While it is true that you have provided evidence of an effect from 2 short-

term studies, these used 2 different formulations (1 for IR and 1 for ER). We do not find these data sufficient to support the effectiveness of gepirone ER. The negative outcome on study 28709, which used a randomized withdrawal design that regularly is successful in showing the effectiveness of effective agents, further weakens what evidence there was to support an antidepressant claim for gepirone ER. Furthermore, as the negative data continue to accumulate, it becomes difficult to advise you on what further work you might do to address such a preponderance of negativity. At this point, we would want at least a "robustly positive" short-term trial with gepirone ER and a positive randomized withdrawal study. The short-term trial should look at different fixed doses of gepirone ER, and the randomized withdrawal study would need to involve a period of "response" for a minimum of 6 months before randomization.

Additionally, we have the following comments and requests that will need to be addressed in your resubmission.

Proposed Tradename

Our Division of Medication Errors and Technical Support (DMETS) has completed their review of your proposed tradenames of "Variza" and "Alrize". The Variza tradename was found unacceptable because there are existing drug names that sound like or look like Variza to a degree that potential confusion between drug names could occur under the usual clinical practice settings. The Alrize tradename was found unacceptable because the name is misleading.

Please submit another proposed tradename for review by the Agency.

Additionally, DMETS reviewed the labels and labeling from a safety perspective. DMETS has identified several areas of possible improvement that might minimize potential user error.

A. Blister Label (14 Count Patient Starter Kit)

    1. Please include product strength.

    2. If the starter kit is available in different strengths, ensure that the strengths are clearly differentiated by using contrasting color, boxing, or some other means.

    3. Add a net quantity statement.

    4. Please include the statement "Each tablet contains xx mg……..etc".

B. Container Label (Professional Sample – 7 count)

    See comments A-2 and A-4.

C. Container Label (20 mg, 40 mg, 60 mg, and 80 mg - 30 count and 500 count)

    1. Increase the prominence of the established name, relative to the dosage form.

2. Relocate the product strength so that it appears below the established name, and away from the net quantity.

3. Please ensure that the multiple strengths are clear differentiated by using contrasting color, boxing, or some other means.

4. Decrease the prominence of the net quantity statement for the 500 count label.

5. Please add a Child Resistant Closure (CRC) statement to the 30 count bottle.

D. Carton Labeling (20 mg, 40 mg, 60 mg, and 80 mg - 30 count and 500 count)

See comments C-1, C-3, and C-4.

Chemistry, Manufacturing, and Controls

1. Provide a stability protocol and updated stability data for the commercial drug product.

2. The Office of Compliance has informed us that the Organon Inc. Sub Akzona Inc. (West Orange, NJ; CFN #2211109) site will be closing in June 2004. As a result, this site will need to be withdrawn from the NDA. Please verify that Organon N.V. (OSS, NL; CFN #9610342) and Pliva USA Inc. (East Hanover, NJ; CFN #2243128) will continue to serve as the drug product release testing facilities.

Clinical Pharmacology and Biopharmaceutics

Your resubmission provided studies or explanations that addressed the recommendations of the Office of Clinical Pharmacology and Biopharmaceutics in our March 15, 2002 action letter. Although your responses were generally acceptable, there are several recently identified issues regarding the pharmacokinetics of gepirone as well as recent changes in the drug product that are of some concern. They are as follows:

1. You have now changed the tablet shape with a change in commercial tooling from the original biconvex tablet to the modified flat tablets. Dissolution has only been evaluated for the 20 mg and 80 mg strength tablets. Therefore, the 20 and 80 mg strength tablets (with the modified flat tablet shape) are acceptable, but the 40 and 60 mg strength tablets (with the modified flat tablet shape) are not acceptable.

   Please submit results of dissolution comparisons (of the original vs. modified shape) in multiple media with f2 comparisons to request a biowaiver for the 40 and 60 mg strength tablets.

2. Your proposed dissolution specifications are acceptable on an interim basis. The dissolution specifications are as follows:

   15-25% at 1 h
   40-85% at 5 h
   65-86% at 12 h
   > 86% at 20 h

   These specifications are acceptable provided that you adhere to your stated commitment to re-evaluate the specification at 12 hours after the manufacture of 20 batches of each strength.

3. We ask that you agree to conduct the following studies as a Phase 4 commitment:

   a) Measurement of the effect of intermediate inducers such as rifabutin.
   b) You previously agreed to conduct an *in vivo* study evaluating the effects of different meal compositions on gepirone ER pharmacokinetics as a Phase 4 commitment, and you have stated that the study has been initiated.
   c) We note your commitment to re-evaluate the dissolution specification at 12 hours after the manufacture of 20 batches of each strength.

Within 10 days after the date of this letter, you are required to amend the application, notify us of your intent to file an amendment, or follow one of your other options under 21 CFR 314.120. If you do not follow one of these options, we will consider your lack of response a request to withdraw the application under 21 CFR 314.65. Any amendment should respond to all the deficiencies listed. We will not process a partial reply as a major amendment nor will the review clock be reactivated until all deficiencies have been addressed.

The drug product may not be legally marketed until you have been notified in writing that this application is approved.

If you have any questions, call Paul David, R.Ph., Senior Regulatory Project Manager, at (301) 594-5530.

Sincerely,

{See appended electronic signature page}

Robert Temple, M.D.
Director
Office of Drug Evaluation I
Center for Drug Evaluation and Research

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

/s/

Robert Temple
6/23/04 03:36:42 PM

5.7 Memo – Thomas Laughren - Non-Approval Recommendation (10/25/2007)

M E M O R A N D U M **DEPARTMENT OF HEALTH AND HUMAN SERVICES**
PUBLIC HEALTH SERVICE
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

DATE: October 25, 2007

FROM: Thomas P. Laughren, M.D.
Director, Division of Psychiatry Products
HFD-130

SUBJECT: Recommendation for Non-Approval Action for
Gepirone ER for Major Depressive Disorder (MDD)

TO: File NDA 21-164
[**Note:** This overview should be filed with the 5-1-07 response to our 6-23-04 non-approvable letter.]

1.0 BACKGROUND

Gepirone ER is an extended release formulation of gepirone, an azapirone that is structurally and pharmacologically similar to buspirone, a drug marketed as Buspar for GAD. Like buspirone, gepirone acts primarily at 5HT1A receptors, as a full and partial agonist. Gepirone is not approved for any indications, and this NDA provides data in support of a claim for MDD, in a dose range of 20 to 80 mg/day. No other drugs with this particular pharmacological profile are approved for MDD.

IND 33,626 for gepirone ER was originally submitted 8-28-89. There is also an IND 23,952 for gepirone immediate release (IR) that was originally submitted 4-2-84. The shift in focus from the IR to the ER formulation was based on poor tolerance of the IR formulation, i.e., dizziness, nausea, and insomnia. It should be noted that these INDs were originally held by BMS, but they discontinued all trials in 1992. In 1993, rights to gepirone ER were transferred to Fabre-Kramer, and then Organon reached agreement with Fabre-Kramer to further develop and market gepirone ER in 1998. More recently, Fabre-Kramer has again taken over the NDA. These transfers of ownership resulted in several disruptions in the flow of the development program.

This NDA was originally submitted 9-30-99. Although there were a total of 16 placebo-controlled trials submitted (10 with IR and 6 with ER), there were notable problems with many of these, and in fact, the sponsor focused on only 4 of these as worthy of particular attention: 1 ST ER study (53); 2 ST IR studies (003 & 001B); and 1 randomized withdrawal study (002). We had previously agreed with the sponsor on the principle of bridging to a positive IR database with a single positive ER study. However, it had now become apparent that the study offered as a positive ER study, i.e., study 53, was in fact a 2-center study in which the protocol specified analysis called for a pooling of the data for the 2 centers. However, this analysis was not

positive overall, yet the sponsor had looked at each center independently, and submitted positive results for 1 center as a positive trial. Consequently we issued a RTF letter on 11-30-99.

The NDA was resubmitted 5-18-01. We had previously reached agreement with the sponsor that one additional positive short-term trial with the ER formulation would be sufficient, providing there was independent evidence for the IR formulation. We agreed with the sponsor that ER study 134001 could be considered a positive study. Upon review of the NDA, however, we had concluded that none of the 3 candidate IR studies could be considered positive. Thus, we issued a nonapprovable letter on 3-15-02. We subsequently modified our view on IR study 03A7A-003, and did consider that a positive study. However, given the accumulating negative data, we indicated that we would want at least 2 positive ER studies for support of this NDA.

The NDA was resubmitted for the second time on 12-23-03, but did not include the results from a second positive short-term ER study. Rather, the sponsor included the results of a randomized withdrawal study that they considered positive (28709). However, we did not agree with the sponsor's exclusion of patients and events from their analysis, and we had become increasingly concerned about the mounting number of negative trials. Thus, we issued a second nonapprovable letter on 6-23-04. We advised the sponsor that they would need both a second positive short-term ER study and a positive longer-term randomized withdrawal study.

This 5-1-07 third resubmission of this NDA did include the results of a second short-term ER study that the sponsor considered positive (FKGBE007), but did not include the results of another randomized withdrawal study. Rather, the sponsor presented the results of yet another analysis of study 28709 which they now again considered a positive study. This resubmission also included the results of various analyses of sexual dysfunction data which they felt provided evidence that gepirone ER is not associated with sexual dysfunction.

2.0 CHEMISTRY

The only definitive issue is a stability concern, and we will be asking for an update on this in our action letter. In addition, however, the only drug substance manufacturing site has been recently inspected and we have not yet received a recommendation from OC.

3.0 PHARMACOLOGY

The only new pharmacology/toxicology data submitted were results of a chromosomal aberration study. In the 3-15-02 nonapprovable letter, we had noted that their in vitro chromosomal aberration test was not adequate, however, we had not required them to repeat it. Nevertheless, the sponsor did repeat it, and submitted the results in this resubmission. Unfortunately, the second study was also inadequate, and the pharm/tox group has decided that an acceptable resolution is to remain silent in labeling regarding the issue of these two studies.

4.0 BIOPHARMACEUTICS

The resubmission included results of 2 food effect studies and a request for a biowaiver for a BE study for the proposed 40 and 60 mg tablet strengths. OCP considers the food effect studies acceptable, agrees with the requested biowaiver, and recommends a phase 4 commitment for a rifabutin interaction study. They also have comments for labeling.

5.0 CLINICAL DATA

5.1 Efficacy Data

As noted, the third resubmission of this NDA did include the results of a second short-term ER study that the sponsor considered positive (FKGBE007), but did not include the results of another randomized withdrawal study. Rather, the sponsor presented the results of yet another analysis of study 28709 which they now again considered a positive study. They have also done various meta-analyses of the 12 gepirone ER studies they consider adequate from the standpoint of dose. The details of these various analyses are provided in the clinical and statistical reviews of this resubmission, and I refer to these reviews by Drs. Kong, Hearst, and Khin for these details. I will provide an overview of what I consider to be the critical efficacy issues in this memo.

In my view, the major deficiency in this application continues to be a failure to provide substantial evidence for the effectiveness of gepirone ER in the short-term or longer-term treatment of major depressive disorder (MDD). Although I agree that the sponsor has provided evidence of short-term antidepressant effectiveness for gepirone ER from 2 adequate and well-controlled trials, i.e. from studies FKGBE007 and 134001, this evidence is derived from a total pool of 12 studies for which the remaining 10 studies do not provide evidence of the effectiveness of gepirone. I acknowledge that 4 of these remaining 10 studies were terminated early for business reasons, and, therefore, might not be expected to provide such evidence because they did not reach their planned sample sizes. However, there are other findings among these trials that amplify my concern about the potential value of gepirone ER as a treatment for MDD.

We have re-evaluated all 12 trials with a focus on the HAMD-17 total score as a common measure of efficacy. Although this was not the protocol specified primary endpoint for 3 of the 12 trials, we felt it was a reasonable common measure because it is so widely used as a primary endpoint in depression trials. In fact, the sponsor selected this as a common endpoint for their meta-analyses. Using this measure, we found the following:

In 3 of these 12 trials, an active comparator antidepressant was statistically superior to gepirone ER, as follows:

Trial	Active Comparator	Active Comparator vs Gepirone ER
ORG 134004	Fluoxetine	-1.71 (p=0.027)
ORG 134017	Fluoxetine	-1.54 (p=0.042)
ORG 134006	Paroxetine	-1.85 (p=0.012)

In 2 of these 12 trials (CN105-053 and ORG 134006), an active comparator was superior to placebo and gepirone ER was not, as follows:

Trial	Active Comparator	P-Values Act Comp vs Pbo	P-Values Gepirone ER vs Pbo
CN105-053	Imipramine	-3.19 (p=0.038)	-2.00 (p=0.190)
ORG 134006	Paroxetine	-1.63 (p=0.026)	0.22 (p=0.760)

I agree that meta-analysis is a reasonable approach to try to better understand the totality of the short-term efficacy evidence for gepirone ER, however, I don't agree with the sponsor's approach to looking at "supportive" subsets of the data, and I don't think the appropriate meta-analyses help their case. The major reason for conducting a meta-analysis in this situation is to determine if, among the remaining 10 trials that were not considered positive for gepirone ER, a meta-analysis would provide any support for gepirone ER. Using the sponsor's meta-analytic model for these 10 trials, we found an effect size of -0.09 (p= 0.62). The one reasonable alternative approach that the sponsor utilized, i.e., including all 12 trials, resulted in an effect size of -0.48 (p= 0.09). Thus, neither approach provides additional confidence that gepirone ER is an effective antidepressant therapy.

The negative outcome for the longer-term maintenance efficacy trial (study 28709) is also a concern for this drug. First, we disagree with the sponsor's approach to trying to repair this study by establishing rules for identifying so-called protocol violators, either because they did not technically meet criteria for randomization or were non-compliant in some manner during the trial. The protocol for this study stated that "All protocol violations will be determined by medical, clinical and biometrics personnel prior to breaking the blind…." Thus, we do not find this post hoc attempt to rescue this study by eliminating 40 patients several years after the blind was broken credible or valid. Second, the fact that this trial is negative is significant in that antidepressant trials of this design rarely fail to show a drug effect. Thus, this finding also brings into question the clinical value of this drug in the treatment of MDD.

As noted in our 6-23-04 not-approvable letter for this application, it is difficult to know how to advise the sponsor regarding any future work to salvage this program with such a preponderance of negative findings. Given that they have never, despite our repeated advice to conduct proper dose finding studies, addressed the issue of dose/concentration response, it is possible that better dose finding might help to better understand the multiple failures with this drug. I do not feel that simply conducting additional flexible dose studies is going to address my concern that this is a drug that, while it may have some marginal antidepressant efficacy, is likely inferior to other available antidepressants. I consider MDD a serious illness and it is hard to imagine what the

justification would be for approving an antidepressant drug that is demonstrably less effective than other available agents. Delaying effective treatment with the use of gepirone ER is not something I would be willing to support. The sponsor has provided some findings that suggest there might be a lesser risk of sexual dysfunction with gepirone ER than is seen with many other antidepressant agents (see 5.2, Safety Data). I am not yet convinced, however, that the data they have accumulated regarding sexual dysfunction consistently support this premise. Although additional analyses of the available sexual dysfunction data may help to convince us of the merit of this argument, I do not feel this would be a worthwhile effort, given the striking weakness of the efficacy data for gepirone ER at the present time. I would be willing to discuss the efficacy data with the sponsor, but I am not optimistic that there is a reasonable path forward for any further development of this drug as an antidepressant.

5.2 Safety Data

5.2.1 Overview

This resubmission of the NDA includes both an overview of safety data plus more detailed findings for new studies. There have been roughly 5000 patients exposed to gepirone in this development program up to this point (about 3000 for the ER formulation and about 2000 for IR). Safety concerns have not been a primary issue in previous nonapproval actions, and are not now. The most common and drug-related adverse events appear to be dizziness, nausea, and vomiting.

5.2.2 Safety Issues of Particular Interest

Two safety issues of particular interest have been sexual dysfunction and suicidality, and I will comment briefly on these issues.

<u>Lack of Sexual Dysfunction</u>

The sponsor has tried to make a case that gepirone ER is superior to other antidepressants regarding sexual dysfunction and no different than placebo in this regard. They conducted several different analyses because different information on sexual function was collected in different trials (DISF/DISF-SR; CSFQ; DSM-IV diagnoses reflecting sexual dysfunction, and adverse events suggesting sexual dysfunction). They evaluated both change from baseline and AUC data for these measures. We had met with the sponsor on 10-12-05 to discuss an approach to establishing this claim, and had stated they would have to show both a signal for sexual dysfunction for other antidepressants and a noninferiority of gepirone ER to placebo (using formal hypothesis testing), at doses relevant for efficacy. Dr. Kong has summarized these findings in detail in his review, and has concluded that there is not consistent evidence of an advantage for gepirone ER. It is true that the signal is not consistent for the trials involving DISF/DISF-SR and DSM-IV diagnoses. However, I think the analyses involving adverse events and those for the 2 trials involving the CSFQ measure are quite strong. We have endorsed the CSFQ as a valid measure, and gepirone ER meets our noninferiority standard in both of these trials. It is not as clear that the active control in those studies was shown to be worse than placebo, however, that may not be a reasonable standard to set. In any case, this apparent

advantage regarding sexual dysfunction would not be relevant if the drug has not been shown to have a benefit regarding depression.

Suicidal Ideation/Suicide Attempt

The sponsor conducted analyses for possibly suicide-related adverse events that were classified using the Columbia classification system. Two different pools were used: all phase 2-3 studies (presumably including open label extensions) and limited to the controlled phases of the trials. The first pooling yielded statistically significant results both for suicidality overall (p=0.022) and for suicidal behavior (p=0.048). These signals lost significance when limited to the controlled phases of these trials (p=0.08 for suicidality overall and p=0.12 for suicidal behavior), probably because the inclusion of open extension data biases the analyses against drug. Nevertheless, the data even for the controlled only phases numerically trend in the direction of a suicidality signal for gepirone ER. If this drug were to be approved for MDD, it would have the same strong warning language regarding suicidality as other antidepressants.

5.2.3 Conclusions Regarding Safety of Gepirone ER

There are no safety findings that would preclude the approvability of gepirone ER, however, I agree with the review team that the weak efficacy findings for gepirone ER represent an obstacle to the approval of this drug. The possible advantage of gepirone ER over other antidepressants on sexual dysfunction, even if demonstrated to be consistent, would not be sufficient to overcome the efficacy problem.

5.3 Clinical Sections of Labeling

Since the clinical/statistical review team is in agreement with recommending a nonapproval action for this NDA, we have not included a draft of labeling with the package.

6.0 WORLD LITERATURE

The sponsor provided a brief literature review in this resubmission. Dr. Hearst did not discover any new safety issues of concern for this drug in this material.

7.0 FOREIGN REGULATORY ACTIONS

To my knowledge, neither gepirone IR nor gepirone ER is marketed anywhere at this time.

8.0 PSYCHOPHARMACOLOGICAL DRUGS ADVISORY COMMITTEE (PDAC) MEETING

We decided not to take this application to the PDAC. We did discuss our findings for this application at an internal regulatory briefing on 10-19-07, and there was unanimous agreement that this application does not provide sufficient support for effectiveness in MDD to justify approval.

9.0 DSI INSPECTIONS

It is my understanding that an inspection has been conducted for 2 sites in the most recent positive study, FKGBE-007, and data from these sites were judged to be acceptable.

10.0 LABELING AND NONAPPROVAL LETTER

10.1 Labeling

As noted, we have not proposed labeling for this application.

10.2 Foreign Labeling

Gepirone ER is not marketed anywhere at this time.

10.3 Nonapproval Letter

The nonapproval letter explains the basis for the action.

11.0 CONCLUSIONS AND RECOMMENDATIONS

I believe that the current sponsor for this NDA has not submitted sufficient data to support the conclusion that gepirone ER is sufficiently effective in the treatment of MDD to justify an approval action. Thus, I recommend that we issue the attached nonapproval letter.

cc:
Orig NDA 21-164 (Gepirone ER)
HFD-130/TLaughren/MMathis/NKhin/EHearst/RGrewal/WBender
ODE-I/RTemple

DOC: Laughren_Gepirone MDD_NA3. Memo.doc

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

/s/

Thomas Laughren
10/25/2007 12:27:40 PM
MEDICAL OFFICER

5.8 Letter – Robert Temple – Not Approvable Letter (11/02/2007)

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

Food and Drug Administration
Rockville, MD 20857

NDA 21-164

Fabre-Kramer Pharmaceuticals, Inc.
Attention: Stephen J. Kramer M.D., Chief Executive Officer
5847 San Felipe, Suite 2000
Houston, Texas 77057

Dear Dr. Kramer:

Please refer to your new drug application dated September 30, 1999 received October 1, 1999, submitted under section 505(b) of the Federal Food, Drug, and Cosmetic Act for Gepirone Hydrochloride Extended Release 20mg, 40mg, 60mg, and 80mg tablets.

We acknowledge receipt of your submissions dated May 1, 2007, July 13, 2007, & October 8, 2007.

The May 1, 2007 submission constituted a complete response to our June 23, 2004 action letter.

We have completed our review of your resubmission and find the information presented is inadequate. Therefore, the application is not approvable under section 505(d) of the Act and 21 CFR 314.125(b). The deficiencies are summarized as follows:

Non-Approval Deficiencies:
There are two major deficiencies in the application. First you have failed to provide substantial evidence for the effectiveness of gepirone ER in the short-term or longer-term treatment of major depressive disorder (MDD). In addition, if the data could be considered as representing such evidence we also believe the effect size that would have been demonstrated is unacceptably small compared to alternative therapy for a serious illness. Although we agree that you have provided two well – controlled trials that show a significant effect, i.e. studies FKGBE007 and 134001, this evidence is derived from a total pool of 12 studies. The remaining 10 studies do not show evidence of an antidepressant effect. Indeed, they do not even show favorable trends in almost all cases. We acknowledge that 4 of these remaining 10 studies were terminated early for business reasons, and, therefore, might not be expected to provide such evidence because they did not reach their planned sample sizes, but in one of these an active control was effective and none of the studies show reasonably strong favorable trends. Moreover, there are other findings among these trials that amplify our concern about the potential value of gepirone ER as a treatment for MDD.

We have re-evaluated all 12 trials with a focus on the HAMD-17 total score as a common measure of efficacy. Although this was not the protocol specified primary endpoint for 3 of the 12 trials, we felt it was a reasonable common measure because it is so widely used as a primary endpoint in depression trials. In fact, you selected this as a common endpoint for your meta-analyses. Using this measure, we found the following:

In 3 of these 12 trials, an active comparator antidepressant was statistically superior to gepirone ER, as follows:

Trial	Active Comparator	Active Comparator vs Gepirone ER
ORG 134004	Fluoxetine	-1.71 (p=0.027)
ORG 134017	Fluoxetine	-1.54 (p=0.042)
ORG 134006	Paroxetine	-1.85 (p=0.012)

In 2 of these 12 trials (CN105-053 and ORG 134006), an active comparator was superior to placebo and gepirone ER was not, as follows:

Trial	Active Comparator	P-Values Act Comp vs Pbo	P-Values Gepirone ER vs Pbo
CN105-053	Imipramine	-3.19 (p=0.038)	-2.00 (p=0.190)
ORG 134006	Paroxetine	-1.63 (p=0.026)	0.22 (p=0.760)

Thus, among a total of 5 trials with active comparators, 3 clearly possessed assay sensitivity, yet failed to show an effect of gepirone. We have not seen such results with any effective drug.

We regularly use meta-analysis as an approach for evaluating demographic subsets within trials (the integrated summary of effectiveness) and for considering, at least at the exploratory level, other hypotheses that a single trial might not be powered to demonstrate, but we have not accepted pooled data as a substitute for a showing of effectiveness in individual trials. Nonetheless, we did examine the 10 failed studies to determine whether they might collectively suggest clinical effectiveness. This meta-analysis would of course need to exclude studies FKGBE007 and 134001 because the major reason for conducting it would be to determine if, among the remaining 10 non-supportive trials, there was any suggestion of an affect of gepirone ER. Using your meta-analytic model for these 10 trials we found an effect size of essentially zero, -0.09 (p= 0.62). Even your proposed approach of including all 12 trials, gives an effect size of -0.48 (p= 0.09), showing that the remaining trials weaken the effect of your two favorable studies. A finding of two positive trials among 12 could occur by chance (about 3.5%) and does not represent substantial evidence of effectiveness.

The negative outcome for the longer-term maintenance efficacy trial (study 28709) is also a concern for this drug. First, we disagree with your approach to trying to repair this study by establishing rules for identifying so-called protocol violators, either because they did not technically meet criteria for randomization or were non-compliant in some manner during the trial. Your protocol for this study stated that "All protocol violations will be determined by medical, clinical and biometrics personnel prior to breaking the blind…." Thus, we do not find this post hoc attempt to rescue this study by eliminating 40 patients several years after the blind was broken credible or valid. Second, the fact that this trial is negative is significant in that antidepressant trials of this design almost never fail to show a drug effect. Thus, this finding further casts doubt on the evidence that gepirone is effective.

As noted in our June 23, 2004, not approvable letter for this application, it is difficult to know how to advise you regarding any future work with gepirone ER. Even if another positive study were to provide further support of an effect, the overall results and 3 studies showing inferiority to an active control indicate that gepirone, if it has any effect at all, is far less effective than standard therapy. We consider MDD a serious illness and it is hard to see a basis for approving an antidepressant drug that was demonstrably and substantially less effective than other available agents, essentially leaving

patients untreated until they substitute effective therapy. The only possible basis for approval we can see at present would be is a determination that an inadequate dose was used in studies to date with studies at an appropriate dose showing an effect similar to other antidepressants. We note that you have provided some evidence of a lesser risk of sexual dysfunction with gepirone ER than is seen with many other antidepressant agents but this cannot support approval. First, the data you have accumulated regarding sexual dysfunction do not consistently support this premise. Second, we do not feel such a finding would overcome the disadvantage of the observed decreased effectiveness. We would be willing to discuss the efficacy data with you, but we are not optimistic that there is a reasonable path forward for any further development of this drug as an antidepressant.

Although not a reason for this not approvable action, you will need to also address the Chemistry Manufacturing and Controls deficiencies below:

1. Revise your acceptance criterion for Individual unspecified impurity to NMT (b) (4) % in accordance with ICH Q3B guideline. Any individual impurities at levels higher than identification threshold of (b) (4) % should be specified by name, relative retention time or some other suitable identifier.

2. The provided stability data for the original biconvex tablets and the Organon modified flat tablets is not sufficient to support your request for 36 month expiration date for the drug product. Please provide long-term and accelerated stability data for the commercial to-be-marketed drug product in each of the proposed packaging configurations.

Within 10 days after the date of this letter, you are required to amend the application, notify us of your intent to file an amendment, or follow one of your other options under 21 CFR 314.120. If you do not follow one of these options, we will consider your lack of response a request to withdraw the application under 21 CFR 314.65. Any amendment should respond to all the deficiencies listed. We will not process a partial reply as a major amendment nor will the review clock be reactivated until all deficiencies have been addressed.

Under 21 CFR 314.102(d), you may request a meeting or telephone conference with this division to discuss what steps need to be taken before the application may be approved.

The drug product may not be legally marketed until you have been notified in writing that this application is approved.

If you have any questions, contact LCDR Renmeet Grewal, Pharm.D., Senior Regulatory Project Manager, at (301) 796-1080.

Sincerely,

{See appended electronic signature page}

Robert Temple, M.D.
Director
Office of Drug Evaluation I
Center for Drug Evaluation and Research

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

/s/

Robert Temple
11/2/2007 02:23:04 PM

5.9 Meeting Minutes –Thomas Laughren (12/28/2011)

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration
Silver Spring MD 20993

NDA 21164

MEETING MINUTES

Fabre-Kramer Pharmaceuticals, Inc.
Attention: Martin Lobel, Esq.
Attorney
Law Offices of Lobel, Novins & Lamont, LLP
888 17th Street, N.W., Suite 810
Washington, D.C. 20006

Dear Mr. Lobel:

Please refer to your New Drug Application (NDA) submitted under section 505(b) of the Federal Food, Drug, and Cosmetic Act for gepirone hydrochloride extended release tablets.

We also refer to the meeting between representatives of your firm and the FDA on November 29, 2011. The purpose of the meeting was to further discuss reconsideration of the Agency's not approvable decision conveyed in the November 2, 2007 not approvable letter.

A copy of the official minutes of the meeting is enclosed for your information. Please notify us of any significant differences in understanding regarding the meeting outcomes.

If you have any questions, call Hiren Patel, Pharm.D., Regulatory Project Manager at (301) 796-2087.

Sincerely,

{See appended electronic signature page}

Thomas Laughren, M.D.
Director
Division of Psychiatry Products
Office of Drug Evaluation I
Center for Drug Evaluation and Research

ENCLOSURE:
Meeting Minutes

FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

MEMORANDUM OF MEETING MINUTES

Meeting Type: C

Meeting Date and Time: November 29, 2011; 4:00PM-5:00PM (EST)
Meeting Location: Building 22 Conference Room 1313

Application Number: NDA 21164
Product Name: Gepirone hydrochloride extended release tablets
Indication: Treatment of Major Depressive Disorder
Sponsor/Applicant Name: Fabre-Kramer Pharmaceuticals, Inc.

Meeting Chair: Thomas Laughren, M.D.

FDA ATTENDEES

Robert Temple, M.D.	Director, Office of Drug Evaluation I
Thomas Laughren, M.D.	Director, Division of Psychiatry Products (DPP)
Mitchell Mathis, M.D.	Deputy Director, DPP
Ni Khin, M.D.	Medical Team Leader, DPP
Silvana Borges, M.D.	Medical Reviewer, DPP
Barry Rosloff, Ph.D.	Supervisory Pharmacologist, DPP
Peiling Yang, Ph.D.	Biometrics Team Leader, Office of Biostatistics
Fanhui Kong, Ph.D.	Biometrics Reviewer, OB
Denise Esposito, J.D.	Deputy Associate Director, Office of Regulatory Policy (ORP)
Martha Nguyen, J.D.	Regulatory Counsel, ORP
Theodore Liazos, J.D.	Associate Chief Counsel, Office of Chief Counsel
Renmeet Grewal, Pharm.D., RAC	Senior Regulatory Project Manager, DPP
Mallory Makowsky	Pharmacy Student Intern, DPP
Nikunj Patel	Pharmacy Student Intern, DDMAC
Varun Vasudeva	Pharmacy Student Intern, DPP

FABRE-KRAMER PHARMACEUTICALS, INC., ATTENDEES

Louis F. Fabre, M.D.	Chairman of the Board
Stephen J. Kramer, M.D.	CEO
Edward H. Koehler, MBA	Executive VP
Mary F. Johnson, Ph.D.	Executive VP Biostatistics, PharmaNet
Mary K Pendergast, J.D.	Consultant
Martin Lobel, Esq.	Attorney, Lobel, Novins & Lamont, LLP
Lee Ellen Helfrich, Esq.	Attorney, Lobel, Novins & Lamont, LLP

1.0 BACKGROUND

During this meeting scheduled for November 29, 2011, Fabre-Kramer Pharmaceuticals, Inc. plans to discuss the statistical report provided by their consultant, Mary F. Johnson. It was submitted in support of the sponsor's request that FDA reconsider our non-approval decision conveyed in a November 2, 2007 letter for gepirone ER in the treatment of major depressive disorder (MDD). Fabre-Kramer Pharmaceuticals, Inc. was previously granted a Face-to-Face meeting on January 14, 2008 to discuss their responses to the November 2, 2007 NA letter. During the January 14, 2008 meeting we concluded that it was highly unlikely any additional analyses of the existing database would justify further review of the NDA.

Regulatory History

Gepirone was initially developed in an immediate release (IR) formulation under IND 23,952. Because of poor tolerability with the IR formulation, the development program was re-focused on an extended release (ER) formulation under IND 33,626. There were several ownership transfers for this drug development program. Since 2005, Fabre-Kramer Pharmaceuticals, Inc. has been the sponsor for Gepirone ER. The original NDA (NDA 21164) which was submitted on September 30, 1999 was refused for filing. The NDA was resubmitted on May 18, 2001 and an NA letter was issued on March 15, 2002. The sponsor resubmitted the NDA on December 23, 2003 and another NA letter was issued on June 23, 2004. In response to the NDA resubmission dated May 30, 2007, a third NA letter was issued on November 2, 2007. Inadequate efficacy was the primary basis for all the above actions.

The major deficiency cited in the November 2, 2007 NA letter was, as noted, a failure to provide substantial evidence of efficacy in the short-term and longer-term treatment of MDD. Although the letter noted that the available evidence suggested that gepirone appeared to be less effective than other available antidepressants, relative efficacy was not the basis for the NA decision. The NA action was based on the lack of substantial evidence of effectiveness. The finding of lesser efficacy for gepirone compared to several other antidepressants was considered additional evidence that gepirone may not be effective.

The NA letter acknowledged that studies FK-GBE-007 and ORG 134001 could be considered positive studies. It noted, however, that this evidence was derived from a total pool of 12 studies, and that the remaining 10 studies did not show evidence of an antidepressant effect, and that they did not even show favorable trends in most cases. We re-evaluated all 12 trials with a focus on the HAMD-17 total score as a common measure of efficacy, and used an analytical model that FDA often recommends (see Table below).

Study No.	Number of Subjects		Active Control, Dose range (mean dose)	Doses Range of gepirone ER (mean daily dose) in mg	LS mean difference using HAMD-17; p-value (gepirone vs. placebo)	Overall Study Results
	GepER	Pbo				
FK-GBE-007	116	122	None	20-80 (58.2)	-2.45; p=0.018	Positive
FK-GBE-008	96	99	None	20-80 (60.0)	-1.38; p=0.20	Negative
ORG 134001	101	101	None	20-80 (61.1)	-2.47; p=0.013	Positive
ORG 134002	102	103	None	20-80 (57.9)	-0.71; p=0.42	Negative
ORG 134004	124	130	Fluoxetine 20-40 mg (34.1)	20-80 (67.1)	1.04; p=0.18	Negative[1]
ORG 134006	140	143	Paroxetine 10-40 mg (28.2)	20-80 (55.3)	0.22; p=0.76	Negative
ORG 134017	159	159	Fluoxetine 20-40 mg (25.9)	40-80 (59.7)	0.65; p=0.39	Negative[1]
ORG 134023	123	123	None	20-80 (61.3)	0.13; p=0.90	Negative
CN105-052*	35	37	Fluoxetine 20-80 mg (23.3)	20-60 (43.4)	-0.69; p=0.74	Failed
CN105-053*	56	56	Imipramine 50-200 mg (145)	10-60 (50.4)	-2.0; p=0.19	Negative
CN105-078*	88	47	None	10-50 (30.4) 20-100 (52.6)	-1.0; p=0.36	Negative
CN105-083*	73	39	None	10-50 (30.4) 20-100 (57.1)	-0.49; p=0.75	Negative

Information on dose range was extracted from the integrated summary of efficacy submitted by Fabre-Kramer Pharmaceuticals (ise-2007.pdf, Table 11 on Page 43 and Table C4 on Page 552)
* terminated early due to business/administrative reasons by Bristol-Myers Squibb
[1] Ordinarily, a study for a new drug is considered negative for that drug if the new drug fails to beat placebo but an active comparator beats placebo. In this study, neither active comparator nor the new drug beat placebo, however, we still considered the study negative for gepirone because the active comparator beat gepirone on HAMD-17.

Using this measure and using an analytical model that FDA often recommends, we found that, in 3 of these 12 trials, an active comparator antidepressant was statistically superior to gepirone ER, i.e., in studies ORG 134004 (Fluoxetine, -1.71, p=0.027), ORG 134017 (Fluoxetine, -1.54, p=0.042), and ORG 134006 (Paroxetine, -1.85, p=0.012). In 2 of these 12 trials (CN105-053 and ORG 134006), an active comparator was superior to placebo and gepirone ER was not. Thus, we suggested that, among 4 trials with apparent assay sensitivity, gepirone was not found to be effective. We noted that we have not observed such results for any drug shown to be effective for MDD.

The NA letter also commented on the use of meta-analysis as an approach for evaluating effectiveness, and provided our results of an analysis of the 10 non-positive studies to determine whether they might collectively suggest clinical effectiveness. This meta-analysis excluded studies FK-GBE-007 and ORG 134001 because the major reason for conducting the meta-analysis was to see whether there would be further support for gepirone among the remaining 10 non-supportive trials. In fact, there was no suggestion of any effect of gepirone ER. Using the sponsor's meta-analytic model for these 10 trials, we found an effect size of essentially zero, -0.09 (p= 0.62). Even the sponsor's proposed approach of including all 12 trials gave an effect

size of -0.48 (p= 0.09), showing that the remaining trials weaken the effect of the two favorable studies.

We also commented on the negative outcome for the longer-term maintenance efficacy trial (study 28709), a randomized withdrawal design study, as a concern for this drug. First, we noted our disagreement with the sponsor's approach to trying to repair this study by establishing rules for identifying so-called "protocol violators," either because they did not technically meet criteria for randomization or were non-compliant in some manner during the trial. The protocol for this study stated that "All protocol violations will be determined by medical, clinical and biometrics personnel prior to breaking the blind…." Thus, we did not find this post hoc attempt to rescue this study by eliminating 40 patients several years after the blind was broken credible or valid. Second, the fact that this trial is negative is extraordinary. To date, of 11 trials of effective antidepressants using this same design, we have not observed a single failed study. This finding casts serious doubt on gepirone's effectiveness.

Arguments by the Sponsor

The sponsor proposed 3 arguments in this meeting package in support of FDA's re-consideration of this application for gepirone. First, they argued that FDA's analysis of the NDA was flawed and used approaches inconsistent with those applied to other NDAs. Second, they objected to what they considered FDA's reliance on a comparative effectiveness standard. Finally, they argued that gepirone should be approved because of its positive benefit risk profile.

Regarding FDA's analysis of the efficacy data, the sponsor argued that 7 of the 12 short-term studies for gepirone ER should not be included in the efficacy analysis: three studies (CN105-052, CN105-078 and CN105-083) because they were terminated prematurely and were therefore insufficiently powered; and four studies involving comparator drugs (CNl05-053, ORG 134004, ORG 134006 and ORG 134017) for lacking assay sensitivity in the sponsor's view. Therefore, the sponsor considers only 5 of the 12 studies appropriate for inclusion in the efficacy analysis: ORG134001; ORG 134002; FK-GBE-007; FK-GBE-008; and ORG 134023. Two of those studies established gepirone ER effectiveness (ORG 134001 and FK-GBE-007), as acknowledged by FDA. In the sponsor's argument, two additional studies (ORG 134002 and FK-GBE-008), although negative, provide support of gepirone ER effectiveness when secondary variables are considered, and only one of the adequate and well-controlled studies of gepirone ER (ORG 134023) offered no corroboration of its effectiveness. Therefore, the sponsor concludes that the ratio of positive to non-positive studies is 2 to 3 rather than the 2 to 10 figure cited in the FDA November 2, 2007 action letter, and that, under FDA policy and practice, substantial evidence of the effectiveness of gepirone ER exists to support approval of NDA 21-164. In addition, the sponsor believes that the standards the FDA used to evaluate gepirone ER were different from those used in the review of NDAs for other antidepressants.

2.0 DISCUSSION

FDA Preliminary Responses: *Regarding the argument that FDA relied on a comparative effectiveness standard, we agree that new drugs for MDD cannot be held to a comparative*

efficacy standard under the FD&C Act or the Clinton-Gore Reinvention guidance. We reemphasize that the basis for the NA action was a conclusion that there was a lack of substantial evidence of effectiveness. The findings from FDA's re-analyses of the data using a standard endpoint (HAMD-17) and a standard analytical model suggesting apparent inferiority of gepirone to several active standard antidepressants were considered additional evidence that gepirone may not be effective at all.

Regarding the sponsor's argument about FDA's analysis of the efficacy data for gepirone, we agree that not all of the 10 non-positive studies can be considered equally strong evidence of a lack of effectiveness in a judgment about the positive to negative ratio for these trials considered individually. Of the 4 trials terminated early, we agree that 3 of these should not be considered in such a judgment, because they were terminated early and did not have assay sensitivity: CN105-052; CN105-078; CN105-083. The fourth study terminated early, i.e., CN105-053, was clearly negative for gepirone, however, did show an effect for the imipramine control, and therefore did have assay sensitivity. We also disagree on another 3 of the 10 trials, which, in our view, also had assay sensitivity: ORG 134004, ORG 134017, and ORG 134006. We acknowledge that our approach to re-evaluating these 4 trials depended on using either a standard endpoint other than the protocol specified endpoint, or an analytical model often recommended by FDA. We also acknowledge that FDA does not ordinarily rely on endpoints and analytical models not specified in the protocol. In this instance, however, we feel that the endpoint used (HAMD-17) was appropriate, as was the analytical model used (one that did not use a treat-by-center interaction term). The HAMD-17 has been a standard endpoint used in MDD trials. As a general statistical principle, inclusion of the treatment-by-center interaction term in the model is strongly discouraged for the primary analysis. In the presence of heterogeneity of treatment effects among centers, the interpretation of the main treatment effect based on the model with the treatment-by-center interaction term is controversial. The problem is even more serious when treatment effect or center effect is non-significant in the main-effects model, which is one reason the Agency did not include the interaction term in our analyses. We included baseline score in ANCOVA in an attempt to improve the precision of the treatment effect estimates.

As noted in our background section, using this standard approach, we found that gepirone did not show efficacy in these 4 trials that had documented assay sensitivity. We rarely see even one such finding with an active control, and we feel it raises a strong concern about whether gepirone has any effectiveness at all. Thus, in making a judgment about the ratio of positive to negative trials that focuses on individual trials, we feel that the correct ratio should be considered 2 positive and 7 negative, a ratio that we still think is a cause for concern, even without considering assay sensitivity. Moreover, 4 of those negative trials had documented assay sensitivity, something not usually known for negative trials. Again, we have never observed such a failure rate with an effective agent.

We do not conduct meta-analyses to salvage failed trials. We did the 10 non-positive trials meta-analysis to see whether it would provide some support for the 2 positive trials. We also did a meta-analysis for the 5 trials that had an active comparator, using the sponsor's summary statistics of HAMD-17 obtained from their analyses of these trials. Results from this meta-analysis did not support the efficacy of gepirone ER with a treatment effect estimate 0.38

apparently favoring placebo, but they provided evidence for the effectiveness of the active comparators as a whole (estimated effect = -1.22, p = 0.003) in the short-term treatment of MDD. We also performed a meta-analysis using the patient-level data for the 5 trials. All the active comparator arms were coded as one 'active control' arm for the comparison between active control and placebo. Fixed-effects (treatment, baseline, study) and mixed-effects (study as a random effect) ANCOVA models were applied. Results based on both models appear consistent with those from the meta-analysis based on the summary statistics. Again, these analyses were seeking any further suggestions that gepirone might be an active antidepressant.

Finally, another factor that points to the lack of effectiveness of gepirone ER is the longer-term maintenance study results (study 28709). Although such studies are not required for initial approval, they are used to support maintenance claims for antidepressant agents. In fact, for 11 agents approved by the FDA for MDD indication based on short-term efficacy results since 1987 (i.e. since the approval of the first SSRI, fluoxetine), each has a positive longer-term maintenance study. We know of no negative result of such a trial with an effective antidepressant. The negative results of your maintenance study with gepirone ER greatly increase our doubt that gepirone ER is an effective antidepressant, further weakening any conclusion from the 2 positive studies.

No risk benefit assessment can be made to support a drug lacking evidence of effectiveness.

Meeting Discussion: *The sponsor reiterated the arguments presented in their document dated November 8, 2011 (see attachment), which was sent in response to FDA preliminary comments. We referred to our preliminary comments (see above), acknowledging that we, in part, concurred with the sponsor. Regarding the 12 short-term studies, we agreed that studies FK-GBE-007 and ORG 134001 can be considered positive studies, and that study ORG134023 can be considered a negative study. We also agreed that studies CN105-052, CN105-078, and CN105-083 can be considered failed studies, because they were terminated early, were underpowered, and did not have assay sensitivity. The sponsor reiterated its view that studies FK-GBE-008 and ORG 134002 should be considered supportive studies. We acknowledged that these 2 studies numerically favored gepirone over placebo, however, we still considered them negative because they did not show statistical superiority for gepirone over placebo on their primary endpoints.*

It was clarified that the main disagreement between FDA and the sponsor, in regard to the 12 short-term studies, involved the 4 remaining studies, i.e., ORG 134004, 134006, 134017, and CN105-053. All 4 were active-controlled trials, and the sponsor argued that these should be considered failed studies because, for all 4 studies, both gepirone and the active control failed to beat placebo on the protocol-specified endpoint using the protocol-specified analysis. FDA stated that it viewed all 4 as negative studies, because they have assay sensitivity when a standard endpoint, i.e., HAMD-17 and/or a standard, commonly used analytical model, is used and the statistical superiority of an active control drug compared to gepirone is evidence of assay sensitivity, even when the active control drug could not be shown to be statistically superior to placebo. The sponsor argued against relying on a non-protocol specified endpoint and analytical model to justify a conclusion of assay sensitivity, and also our reliance on what they considered an unusual definition of assay sensitivity. For studies ORG 134004 and 134006,

they further noted that the protocol specified endpoint, i.e., HAMD-25, was a more appropriate endpoint, given that having atypical depression was an entry criterion for these studies.

With regard to the longer-term maintenance study (study 28709), the sponsor noted that the responder criterion for randomization was HAMD-17 <u><8</u>. They also noted, however, that some patients were randomized, even though they failed to meet this criterion. Although the analysis including all randomized patients was negative, a post-hoc analysis including only patients who met the protocol specified entry criterion was positive. The sponsor argued that this post hoc analysis, although not sufficient to support a maintenance claim, should be sufficient to reject our argument that the negative results from the original analysis should be considered evidence against the efficacy of gepirone as an antidepressant.

We suggested that the sponsor provide these additional arguments in a formal submission to the NDA, including new analyses, if needed, to support their arguments. In particular, we asked that they provide detailed arguments for why the 4 short-term studies in question, and the negative maintenance study, should not be considered in an overall judgment about gepirone's antidepressant effectiveness. We agreed to review this information in a timely manner.

3.0 ISSUES REQUIRING FURTHER DISCUSSION

The sponsor stated their intention to send additional information in support of gepirone ER's antidepressant effectiveness to the FDA for review.

4.0 ACTION ITEMS

Upon receipt of the additional information on gepirone ER from the sponsor, the FDA will review the submitted information and will convey our response to the sponsor.

5.0 ATTACHMENTS AND HANDOUTS

Sponsor's response (dated November 8, 2011) to FDA preliminary comments is attached.

MARTIN LOBEL
LEE ELLEN HELFRICH
HENRY M. BANTA

LAW OFFICES OF
LOBEL, NOVINS & LAMONT, LLP
888 17TH STREET, N.W., SUITE 810
WASHINGTON, D.C. 20006
(202) 371-6626
TELECOPIER: (202) 371-6643
LNLlaw.com

OF COUNSEL
ARTHUR L. FOX II
ALAN S. NOVINS
(1937 - 2006)
WILLIAM JOHN LAMONT
(1916 - 1994)

November 8, 2011
By email

Hiren D. Patel, Pharm.D., M.S.
LCDR USPHS
Regulatory Health Project Manager
Division of Psychiatry Products
Center For Drug Evaluation and Research, FDA
Office of Drug Evaluation I

RE: NDA 21164; Reply of Fabre-Kramer Pharmaceuticals, Inc. to FDA Preliminary Response for Discussion at the Scheduled November 29, 2011 meeting with the Division of Psychiatry Products.

Dear Dr. Patel:

Thank you for sending me on October 27, 2011 the FDA Preliminary Response to the Meeting Package we submitted with our April 27, 2011 meeting request. We are submitting these comments in the hope that they will help focus the discussion at our meeting scheduled for November 29, 2011 and make it more productive. Please distribute copies of our Reply to all the FDA participants in our meeting.

Since our submission, we have read an article entitled "*Vilazodone: Clinical Basis for the US Food and Drug Administration's Approval of a New Antidepressant*" in the Journal of Clinical Psychiatry (72:9, September 2011)(hereafter Dr. Laughren's paper). We applaud FDA's recent policy to publish information about its approval process, and believe it would be useful for our discussion to compare the FDA's treatment of gepirone-ER with Viibryd (vilazodone), another drug for MDD, which the FDA recently approved.[1] The article was co-authored by many of those who attended the pre-meeting[2] that resulted in the Preliminary Response given to Fabre-Kramer Pharmaceuticals.

[1] For ease of use, we have attached Dr. Laughren's article, along with a line by line comparison of vilazodone's and gepirone-ER's treatment by the FDA. (Exhibits A and B)

[2] Thomas J. Laughren, MD; Robert J. Temple, MD; Mitchell Mathis, MD; H.M. James Hung, Ph.D; and Peiling Yang, PhD are listed as co-authors and as participants in the meeting.

There are remarkable similarities between gepirone-ER and vilazodone in the weight of clinical evidence serving as a basis for approvability. It is clear that the same regulatory standards were not applied to both new drug applications (NDAs).

Based on the differential treatment of vilazodone and gepirone-ER's NDAs, we have concluded that the Preliminary Response is based on regulatory standards that are inappropriately and inconsistently applied, namely: (i) the admittedly unlawful use of comparative efficacy, (ii) the use of statistical methods that violate the FDA's own directives to drug sponsors, (iii) the classification of inadequate and failed studies (lacking assay sensitivity) as negative for efficacy, (iv) an inaccurate assessment of gepirone-ER's side effects, and (v) the use of a "preponderance of the evidence" standard, as opposed to the "substantial evidence" standard required by law. In short, gepirone-ER received discriminatory treatment as compared to vilazodone and other approved MDD drugs for which failed studies were rejected as a basis for evaluating effectiveness and were not considered negative studies[3]. It appears as if the FDA is only willing to approve "me too" reuptake drugs (SSRI or SNRI) like vilazodone and refuses to acknowledge, despite the two positive studies, that a drug like gepirone-ER that uniquely targets only 5HT1a receptors can be effective. In the last 25 years, the FDA has only approved one new non-reuptake drug for MDD (mirtazapine).[4]

Comparative Effectiveness:

In the first paragraph of its Preliminary Response, the FDA agrees, as it must, that "new drugs for MDD cannot be held to a comparative efficacy standard under the FD&C Act or the Clinton-Gore Reinvention guidance," but this belated assertion is undermined by the FDA's non-approvable decision that was based, in large part, on a comparative efficacy argument. Indeed, the FDA is still using comparative effectiveness of gepirone-ER in the Preliminary Response. On page three of the Preliminary Response, the FDA notes that, in studies 134004 and 134017, "neither the active comparator nor the new drug beat placebo" (*i.e.*, the studies lack assay sensitivity) and then specifically compares the effectiveness of gepirone-ER to fluoxetine stating "we still considered the study negative for gepirone because the active comparator beat gepirone on HAMD-17."[5] Further, the entire analysis of the longer-term maintenance study is an exercise in comparative effectiveness.[6]

As noted earlier, the FDA classified studies 134004 and 134017 as negative solely because the comparator, fluoxetine, had a better response rate than gepirone-ER in a post-hoc analysis of a secondary endpoint, even though statistically the comparator should only be used to determine assay

---

[3] *See, e.g.*, the studies listed in footnote 37 of our Meeting Package in which the FDA determined that data from failed studies of antidepressant drugs should not be considered in its assessment of the drugs effectiveness.

[4] *See, e.g.*, the comments about 5HT1a in Dr. Laughren's paper.

[5] Preliminary Response, p. 3, fn1 to the table.

[6] Preliminary Response, p. 6.

sensitivity according to published statements by Dr. Laughren and Dr. Temple[7].

Dr. Laughren's paper states: "Moreover, none of the 3 active comparator trials showed the comparator to be statistically distinguishable from placebo, indicating that these trials lacked assay sensitivity for the primary trial endpoint (HDRS) and were uninformative about vilazodone's effect on that endpoint." *i.e.*, they were not considered negative studies. Yet, the five (now four since the FDA now agrees that 052 should not have been considered[8]) studies of gepirone-ER in which the active comparator was not statistically distinguishable from placebo on the primary trial endpoint were considered negative studies. Why? The explanation offered for two of these four studies is on page three of the Preliminary Response that:

> Ordinarily, a study for a new drug is considered negative for that drug if the new drug fails to beat placebo but an active comparator beats placebo. In [studies 134004 and 134017], neither active comparator nor the new drug beat placebo, however, we still consider the study negative for gepirone because the active comparator beat gepirone on HAMD-17.

This statement that a negative conclusion was reached by the FDA solely based on an assessment of comparative efficacy is directly contrary to the first sentence in FDA's Preliminary Response, to the statements in Dr. Laughren's paper, to the FDA's usual policy in dealing with antidepressant drugs as shown by the statements of Drs. Temple and Laughren listed in footnote 1 and the studies listed in footnote 37 of our Meeting Package and could only reached by changing the primary endpoint for the study, a disallowed practice per FDA guidelines.[9]

Statistical Methods:

The statistical approach used in the assessment of gepirone-ER differs from the approach used for vilazodone and other approved MDD drugs to the unfair disadvantage of gepirone-ER. In its review of the gepirone-ER NDA, the FDA engaged in a statistical re-analysis of selected studies in violation of the "Guidance for Industry: E9 Statistical Principles for Clinical Trials" (Sept. 1998), which the FDA follows, that warns against any post-hoc change to the primary endpoints and their method of analysis. In order to change endpoints, the decision must be made independently, without

---

[7] See footnote 1 in our Meeting Package for a list of published papers.

[8] Although the FDA has finally agreed that three of the four prematurely terminated dose-finding studies conducted by BMS (052, 078 and 083) should never have been considered in its assessment of gepirone-ER's efficacy, it still considers the fourth of these studies (053) as evidence of gepirone-ER's lack of efficacy, even though it suffers from the same faults as the other three and the comparator failed to separate from placebo on the primary efficacy endpoint.

[9] Interestingly, although the FDA doesn't deal directly with Dr. Johnson's analysis that only 5 of the submitted studies were suitable for an efficacy analysis, for the first time the FDA admits that 3 of the 12 studies should not have been considered.

access to the unblinded data[10], which the FDA certainly had before electing to change the gepirone-ER endpoints. If a sponsor tried to submit an analysis by changing protocol endpoints to support its position, the FDA would never allow it.

In general, FDA's methods for re-evaluating the efficacy of gepirone are inconsistent with their directives to drug sponsors and violate rules of good clinical and statistical practice, namely:

1. To assert that a study has assay sensitivity when the active control failed to demonstrate a significant treatment effect based on the primary efficacy variable stated in the protocol.

2. To select a secondary efficacy endpoint to judge assay sensitivity of a trial after the primary endpoint failed to demonstrate efficacy of the active control.

3. To apply a statistical method for the primary efficacy analysis that (A) was not specified in the protocol, and (B) was selected for use after the database was unblinded and study results were available.

4. To pool results across centers, ignoring conflicting, qualitative differences in the effects of treatment among centers in a multicenter study.

Designating Failed Studies as "Negative":

Using the questionable methods noted above, FDA presents a biased assessment of gepirone-ER's efficacy by classifying four active-controlled studies (004, 006, 017, and 053) as "negative" that should be designated as failed or uninterpretable. FDA now agrees that three studies are failed (052, 078 and 083), but includes all others in the tally of negative studies under consideration. In addition, FDA has designated two studies that support the efficacy of gepirone-ER (002 and 008) as negative. In short, FDA has used unacceptable methods to skew the ratio of positive to negative trials for gepirone-ER.

Two active control studies, CN105-053 and ORG134006, are still considered negative by the FDA, but the basis for this judgment violates the FDA's own directive to drug sponsors. In both trials, the comparator drug failed to separate from placebo on the primary efficacy endpoint and thus should have been considered a failed, rather than a negative trial. Instead, the FDA used both a post-hoc selected endpoint AND a post-hoc selected non-protocol analysis to find a single instance in each study in which the comparator was positive and gepirone-ER was not. Although the FDA tries to justify its actions by claiming that its methods are "commonly used" in antidepressant trials, this does not justify its violations of statistical principles and review precedent. Again, to prevent bias and maintain validity of hypothesis tests, decisions to deviate from protocol-defined methods should be

---

[10] Evans S, (2007) *When and How Can Endpoints Be Changed after Initiation of a Randomized Clinical Trial?* PloS Clin Trials 2(4): e 18, doi:10.1371/journal.pctr.0020018. (Exhibit C).

made prior to review of unblinded data. FDA also ignores evidence of significant treatment by center interactions that further compromises the interpretation of these studies, especially in CN105-053, a study terminated before both centers completed enrollment.

According to Dr. Laughren's paper, vilazadone was approved on the basis of two positive studies out of seven studies submitted for review; FDA designated the other five vilazadone trials as failed and dismissed them from further consideration.[11] If gepirone-ER were treated as vilazodone was, it would have been approved. The NDA for gepirone-ER presents twelve studies, of which seven should be disregarded as failed trials because they lack assay sensitivity (004, 006, 017, 053) or were terminated early with inadequate power (052, 078, and 083); of the remaining five trials, two are positive, two support the efficacy of gepirone-ER, and one is negative. Gepirone-ER should be approved on the basis of two positive and two supportive studies out of five adequate and well-controlled studies that merit consideration.

The Preliminary Response states that two positive and seven negative studies is a ratio "that we still think is a cause for concern." However, Dr. Johnson's statistical review and evaluation of the studies[12] shows, that is not an accurate number and can only be reached by employing comparative efficacy arguments, selecting post-hoc endpoints and using post-hoc analyses not specified in the trial protocol, all in violation of FDA's customary practice. A further example showing that, if gepirone-ER were treated as other antidepressant drugs, it would have been approved is Celexa (citalopram). Celexa's NDA was approved based on ten short term studies of which only one was positive and three were negative.[13] In the ten studies there were five active control studies in which, like gepirone-ER, all comparators failed to separate from placebo on the primary efficacy endpoint so, like in every other antidepressant other than gepirone-ER, the FDA treated them as failed studies and did not consider them. In fact, of the five most recent FDA antidepressant approvals (Celexa, Lexapro, Cymbalta, Pristiq and Viibryd), all SSRI or SNRI drugs, two of the five do not even have two positive studies (Celexa and Lexapro) according to protocol.

In summary, FDA's ranking of efficacy evidence from each of the twelve gepirone-ER studies is biased and flawed. To illustrate this point, the following table summarizes results of each study based on FDA's analysis of HAMD-17 vs. the protocol-defined analysis of the primary endpoint; we also note our position on assay sensitivity, adequacy of design, and other factors misjudged by the FDA.

---

[11] According to Cheri Lindberg, MD, who did the medical review of vilazodone's NDA, of the five "failed" studies, two are a placebo-controlled and are clearly negative (not even dose-related effects on HAMD!) and the other three are active controlled, in some cases favoring fluoxetine over vilazodone.

[12] Exhibit B to our Meeting Package

[13] Memorandum from Thomas Laughren MD, March 26, 1998, Recommendation for Approval Action for Celexa (citalopram) for the Treatment of Depression, NDA 20-822.

Study No.	Number of Subjects		Active Control	LS mean diff HAMD-17 p-value (gepER vs. Pbo)	Protocol defined endpoint† p-value		Overall Study Results	Sponsor vs. FDA position
	Gep ER	Pbo			Gep-ER vs. placebo	Active vs. placebo		
FK-GBE-007	116	122	None	-2.45; p=0.018	-2.26; p=0.032	--	Positive	Agreement
FK-GBE-008	96	99	None	-1.38; p=0.20	-1.5; p=0.159	--	Supportive	Negative per FDA: FDA ignores p<0.05 at earlier time points; positive trends for secondary vars. (Week 2,3,6 HAMD-17 and Week 2,3,4 MADRS p<0.05)
ORG 134001	101	101	None	-2.47; p=0.013	-2.47; p=0.013	--	Positive	Agreement
ORG 134002	102	103	None	-0.71; p=0.42	-0.67; p=0.446	--	Supportive	Negative per FDA: FDA ignores positive trends for all vars; mixed models mlHAMD-17, Bech-6, item 1, and MADRS all p<0.05
ORG 134004	124	130	Fluoxetine	1.04; p=0.18	0.87; p=0.416 HAMD-25	-1.03; p=0.325 HAMD-25	Failed	Negative per FDA: FDA deviates from protocol-defined-analysis and endpoint to test assay sensitivity (AS); compares efficacy (Fluox>Gep) **without AS**; ignores trmt x center interaction for HAMD-17 (p=0.05)
ORG 134006	140	143	Paroxetine	0.22; p=0.76	0.06; p=0.953 HAMD-25	-1.58; p=0.178 HAMD-25	Failed	Negative per FDA: FDA deviates from protocol-defined analysis and endpoint to claim AS and compare efficacy (Parox>Gep); ignores trmt x center interaction (p=0.024 for HAMD-25)
ORG 134017	159	159	Fluoxetine	0.65; p=0.39	0.50; p=0.650 MADRS	-1.15; p=0.299 MADRS	Failed	Negative per FDA: FDA deviates from protocol-defined analysis and endpoint to test AS; compares efficacy (Fluox>Gep) **without AS**; ignores trmt x center interaction (p=0.024 for HAMD-25)
ORG 134023	123	123	None	0.13; p=0.90	0.13; p=0.898	--	Negative	Agreement
CN105-052*	35	37	Fluoxetine	-0.69; p=0.74	-0.66; p=0.757	-0.5; p=0.798	Failed	**FDA now agrees failed study; Agreement**
CN105-053*	56	56	Imipramine	-2.0; p=0.19	-0.70; p=0.687	-2.50; p=0.144	Failed	FDA claimed AS based on covariate-adjusted analysis; ignores small sample size (study terminated early); AS not supported by co-primary (CGI); conflicting results by center
CN105-078*	88	47	None	-1.0; p=0.36	-0.9; p=0.451	--	Failed	FDA claimed 'negative' study; infers lack of efficacy despite inadequate power to show drug effect; ignores positive trends in secondary vars **FDA now agrees failed study**
CN105-083*	73	39	None	-0.49; p=0.75	-0.5; p=0.742	--	Failed	FDA claimed 'negative' study; infers lack of efficacy despite inadequate power to show drug effect; ignores conflicting results by center **FDA now agrees failed study**

†Based on primary efficacy endpoint and method of analysis pre-specified in the protocol; the primary endpoint is HAMD-17 unless otherwise stated.
*Terminated early due to business/administrative reasons by Bristol-Myers Squibb.

The Preliminary Response's comments about the need for long-term maintenance studies is another example of the FDA's disparate treatment of gepirone-ER. Long term maintenance studies are not required for an NDA's approval for MDD. They are usually done AFTER the drugs are approved in part because, as the FDA admits on page 6 of its Preliminary Response, MDD NDAs are approved "based on short term efficacy results." Thus, as is common, when vilazodone was approved, there was no long term data. However, vilazodone's sponsor has agreed to perform a post-approval long term maintenance study, as would Fabre-Kramer Pharmaceuticals.

Dr. Laughren's paper states "The 2 trials lacking an active control group could be considered 'negative' trials for vilazodone, but this is not certain in the absence of an active control to confirm assay sensitivity." *i.e.,* these results were not weighted heavily in the decision to approve vilazodone's NDA.

Yet, similar studies of gepirone-ER which lacked comparators (002, 008, 023) were considered negative studies when its approval was considered. This is another clear example of discriminatory treatment of gepirone-ER's NDA.

Finally, we object to FDA's misuse of meta-analyses to support its arguments against gepirone-ER. There is no evidence in available review documents that any similar meta-analyses were conducted on any antidepressants approved since 1996. FDA's meta-analyses of gepirone-ER trials, which were limited to subsets of studies (non-positive trials or active-controlled trials), fail to address the main purpose of a meta-analysis; that is, to summarize drug effects based on information from all adequate and well-controlled studies. The pooled results of these subsets of studies are uninformative and misleading because they selectively omit high-quality positive studies and restrict the meta-analysis to inadequate studies (incomplete, under-powered, and/or deficient in design), to judge the efficacy of gepirone-ER. In particular, pooling studies to evaluate assay sensitivity of individual studies seems pointless. A test of the 'active comparators as a whole' is irrelevant to the internal validity of individual studies in the gepirone NDA. Assay sensitivity should relate only to individual studies. An assay sensitive trial requires that the active control product demonstrate a significant effect on the primary endpoint, as this will ensure that the study is sensitive to true drug effects and has the potential to demonstrate efficacy of the experimental drug.

Assessment of Sexual Dysfunction:

Dr. Laughren's paper cites three reasons why vilazodone was not shown to have reduced sexual dysfunction: (1) the overall incidence of sexual adverse events was greater than placebo, (2) there was no active control, and (3) the sexual questionnaires did not show a consistent pattern. However, gepirone-ER showed: (1) sexual adverse events that were less than placebo, (2) active controls were used to show assay sensitivity, and (3) there was a consistent pattern of results from sexual function questionnaires. In short, had the FDA assessed gepirone-ER the same way it assessed vilazodone, it would have come to a conclusion that gepirone-ER has less sexual dysfunction side effects than the comparator SSRIs. That is an important benefit, since sexual dysfunction side effects are one of the most common reasons MDD patients stop taking their medication.[14]

Preponderance of the Evidence Standard:

Contrary to the law, the FDA has used the more onerous "preponderance of the evidence" standard in assessing gepirone-ER, rather than the statutory standard of "substantial effectiveness." [15] The preponderance of the evidence test was specifically rejected by Congress in 1962.[16] Yet, here the

---

[14] Kennedy, S; Rizvi, S Sexual Dysfunction, Depression, andd the Impact of Anitdepressants, 29 J Clinical Psychopharacology, 157-164 (2009) (Exhibit P to our submission of April 27, 2011).

[15] *See* footnotes 3, 10, 11 and 12 in Meeting Package.

[16] *See* footnotes 13,14, 15, 16 and 19 in Meeting Package. *See also*, Dr. Laughren's quote in footnote 3 of the Meeting Package.

FDA comments several times that the ratio of negative studies disfavors gepirone-ER, rather than focusing on the fact that there is substantial evidence of gepirone-ER's effectiveness.[17]

The FDA has never disputed that gepirone-ER has two positive studies, which ordinarily would result in the NDA's approval. Indeed, the effect size of gepirone-ER in its two positive studies were better than vilazodone's in its two positive studies.[18] The question is, if the MADRS sizes of gepirone-ER in the pivotal trials of -3.3 and -4.2 are unacceptably small, why are the effect sizes in the pivotal trials of vilazodone of -2.5 and -3.2 acceptable? The result of FDA's discriminatory treatment of gepirone-ER is that patients are being deprived of an effective, novel mechanism drug that may help them, particularly those who stop taking MDD drugs because of the sexual dysfunctional side effects which do not appear to be a side effect of gepirone-ER.

We look forward to our scheduled meeting on November 29th, receiving answers to our questions and, hopefully, a recommendation that NDA 21164 is approvable. Approval of gepirone-ER will give physicians, for the first time in at least 15 years, a truly novel weapon with minimal side effects in their armamentarium against depression. This is important because, as Drs. Laughren and Temple have written, approved medications for MDD only work about half the time[19].

If you have any questions, please let me know.

Sincerely,

Martin Lobel
Counsel for
Fabre-Kramer Pharmaceuticals, Inc.

Attachments

---

[17] *See* Report of the Senate Committee on the Judiciary on Amendments to S. 1552, Report 1744, Part 2, 87th Cong., 2d Sess., pp.6, 16 and 57.

[18] For vilazodone, in the two pivotal trials the effect sized for MADRS change from baseline LOCF are study 07 -2.5, and study 04 -3.2. The MADRS change from baseline LOCF for the gepirone-ER pivotal trials are: study 001 -3.3 and study 007 -4.2.

[19] For a list of such statements, *see* footnotes 7 and 8 in the Meeting Package.

Comparison of the FDA's Approval Process for Vilazodone and Gepirone-ER as Shown by a Line by Line Analysis of Dr. Laughren's Paper (J. Clinical Psychiatry 72:9, Sept.2011, p. 1166)

1. Vilazodone was approved by the US Food and Drug Administration (FDA) on January 21, 2011, for the treatment of major depressive disorder (MDD).

The FDA completed the review of our resubmission (of NDA 21,164) and found that "the information presented is inadequate. Therefore, the application is not approvable under section 505(d) of the Act and 21 CFR 314.125(b)" (FDA not approvable letter, November 2, 2007)

2. Vilazodone is an indolalkylamine that binds with high affinity to the serotonin reuptake site (K_i = 0.1 nM) but not to the norepinephrine (K_i = 56 nM) or dopamine (K_i = 37 nM) reuptake sites.

Gepirone is an azapirone that binds with significant affinity to the serotonin receptor site (5-HT_{1A}) (K_i = 38 nM) but not to the 5-HT_{1B} (K_i = 7800 nM), 5-HT_{1C} (K_i = 5000 nM), 5-HT_{1D} (K_i>100,000 nM), 5-HT_{2A} (K_i >6310 nM), 5-HT_{2C} (K_i >3333 nM,) norepinephrine α_1 adrenergic (K_i = 1960 nM), α_2 adrenergic (K_i = 10,000 nM), β adrenergic (K_i > 100,000 nM), dopamine D_1(K_i >1,000 nM), or dopamine D_2 (K_i = 1905 nM) reuptake sites. (Gepirone Investigator's Brochure 5th Edition (IB5), Table 1, pp. 30-31)

3. It potently and selectively inhibits reuptake of serotonin (IC_{50} = 1.6 nM); i.e., it is a selective serotonin (5-HT) reuptake inhibitor (SSRI).

Gepirone does not inhibit reuptake of serotonin.
(IB5, p. 29)

4. Vilazodone also binds with high affinity to 5-HT_{1A} receptors and is a 5-HT_{1A} receptor partial agonist at this receptor (IC_{50} = 2.1 nM).

Gepirone is a 5-HT_{1A} partial agonist (IC_{50} = 60 nM).
(IB5, p. 29)

5. As is the case for all antidepressants, the mechanism of vilazodone's antidepressant effect is not known but is thought to be related to enhancement of serotonergic activity in the central nervous system through its SSRI effect.

As is the case for all antidepressants, the mechanism of gepirone's antidepressant effect is not known. Its potential is to normalize 5-HT neurotransmission both in 5-HT-deficit disorders (depression) and 5-HT excess (anxiety). These effects may derive from its partial agonist properties. (IB5, p. 36)

6. Although vilazodone is a partial agonist at 5-HT_{1A} receptors, the net effect of this activity on serotonergic transmission is not known. Whether the partial agonist activity enhances vilazodone's antidepressant effect has not been explored, e.g., by studying vilazodone in nonresponders to SSRIs.

Gepirone is a partial agonist at 5-HT_{1A} receptors. This pharmacology suggests a potentiation of serotonergic neurotransmission in the central nervous system. However, like other antidepressants the exact mechanism of action is unknown. (Label p18)

7. Peak plasma vilazodone concentrations are reached at about 4 to 5 hours after dosing when the drug is taken with food.

Peak plasma gepirone concentrations are reached at about 4 hours after dosing when the ER formulation is taken with food. (IB5, Table 32, p. 85)

8. Vilazodone's elimination half-life is about 25 hours, and steady state is achieved in about 3 days.

Gepirone-ER's elimination half-life ranges from 6 - 11 hours, and steady state is achieved in about 2 - 4 days. (IB5, Table 35, p. 87)

9. Vilazodone has an accumulation factor of about 1.8; i.e., peak plasma vilazodone concentration at steady state is 1.8 times that observed after a single dose.

Gepirone-ER has an accumulation factor of about 1.7; i.e., peak plasma gepirone concentration at steady state is 1.7 times that observed after a single dose. (IB5, p. 87)

10. Vilazodone pharmacokinetics show a prominent food effect; i.e., peak plasma vilazodone concentration and area under the concentration time curve are increased about 2-fold when the drug is taken with food, even a light meal. For this reason, vilazodone should be taken with food to ensure adequate plasma concentrations.

Gepirone-ER pharmacokinetics show a food effect; i.e., peak plasma gepirone concentration and area under the concentration time curve are increased about 30 – 70% when the drug is taken with food. For this reason, gepirone should be taken with food to ensure adequate plasma concentrations. (Label, p. 2)

11. Vilazodone is extensively metabolized through cytochrome P450 (CYP) and non-CYP (probably by carboxylesterase) pathways, with only - 3% excreted unchanged in the urine and feces.

Gepirone is extensively metabolized through cytochrome P450 (CYP450-3A4) and minimally by CYP450-2D6, with only 3.5% excreted unchanged in the urine. (IB5, p. 99; Clinical Study Report (CSR) FKGBE005, Table 5 p. 35)

12. CYP3A contributes significantly to the metabolism via CYP pathways; CYP2C19 and 2D6 pathways have a minor role in vilazodone's metabolism.

Gepirone is extensively metabolized through cytochrome P450 (CYP450-3A4). (IB5, p. 99)

13. Genetic variation in CYP2C19 and 2D6 has not been observed to affect vilazodone's plasma levels significantly.

In one study in adults (CN105-025) and one study in children and adolescents (FKGBE009) poor and extensive metabolizers of CYP2D6 and CYP3A4 have not been observed to affect gepirone's plasma levels significantly. (CSR CN105-025 and FKGBE009)

14. Vilazodone has no active metabolites; i.e., its clinical effects are thought to be due primarily to the parent drug.

Gepirone has at least one active metabolite. 3-OH-gepirone. A second metabolite 1-PP is not thought to contribute to antidepressant activity .It is not known if other metabolites of gepirone (5-hydroxy-gepirone and 3,5 di-hydroxy-gepirone) are active. Preclinical results indicate that 3'-OH-gepirone possesses antidepressant-like properties that are likely to contribute to the efficacy of gepirone in the treatment of depression. (IB5, p. 36)

15. The studies of various intrinsic factors on vilazodone's pharmacokinetics suggest that no dosage adjustment is needed based on age, gender, or the presence of mild to moderate hepatic impairment.

The studies of various intrinsic factors on gepirone's pharmacokinetics suggest that no dosage adjustment is needed based on age, gender, or ethnicity. A hepatic impairment study which included severe hepatic impairment was completed. The metabolism of gepirone is reduced by one-half in these patients and the recommendation is that, "Caution should be exercised when giving gepirone-ER to patients with hepatic impairment (IB5, pp. 87 - 90, 95 – 96, and Label p2)

16. Of course, factors other than pharmacokinetics may necessitate more cautious dosing in certain populations, e.g., the elderly.

In elderly subjects ($\geq$ 65 years old), higher gepirone exposure levels (AUC and C_{max}) were reached than in adult subjects (18 - 40 years old) and no gender differences for gepirone exposure were present with gepirone-ER tablets in adults and elderly subjects. Since treatment with gepirone-ER involves titration to effective dose, no dosage adjustment is needed. (IB5, pp. 87-89)

17. Patients with severe hepatic impairment have not yet been studied; however, the pharmaceutical sponsor has agreed to perform a pharmacokinetic study as a post-marketing commitment.

FKGBE005 studied gepirone in severely hepatically impaired subjects. Metabolism of gepirone to 1-PP in hepatically impaired subjects is reduced by about one-half compared to healthy subjects, resulting in an approximate two-fold increase in both the AUC and C_{max} for gepirone. Gepirone-ER) should be used with caution in patients with mild to moderate hepatic impairment and should be used with extreme caution in patients with severe hepatic impairment. (Label, p. 2)

18. Because vilazodone is extensively metabolized and not renally excreted, renal impairment is not expected to have an important effect on vilazodone's clearance, and no effect was observed in patients with mild to moderate renal impairment.

Renal impairment was associated with a 2-fold increase in gepirone C_{max} compare to the healthy control group. Dosage modification is required for moderate to severe renal impairment. (Label, p. 2)

19. Effects of other drugs on vilazodone (extrinsic factors). In vitro studies of several drugs in combination with vilazodone revealed a significant pharmacokinetic interaction only with a strong CYP3A4 inhibitor (Figure 2), which increased plasma concentrations by about 50%. On the basis of the finding with ketoconazole, the vilazodone dose should be reduced to 20 mg when it is used in combination with strong CYP3A4 inhibitors.

Concomitant administration of gepirone-ER with ketoconazole to healthy subjects was associated with a substantial reduction of gepirone oral clearance demonstrated by an increase in mean plasma C_{max} (~5-fold) and AUC (~6-fold). The amount of gepirone excreted unchanged in the urine increased ~5.5-fold. Veraparmil caused a 2.6 fold increase in these gepirone parameters. Concomitant use of gepirone-ER with potent inhibitors of P450 3A4 should be avoided, as the oral bioavailability of gepirone and the metabolite 3'-OH-gepirone are increased. Concomitant use of drugs that induce P450 3A4 may markedly reduce the oral bioavailability of gepirone. (Label, p. 2)

20. Although the interaction of vilazodone with CYP3A4 inducers has not been evaluated, it can be expected that such inducers could result in decreased vilazodone concentrations and possible diminished effectiveness.

Co-administration of gepirone-ER with rifampicin, a CYP3A4 inducer, was associated with a substantial reduction of gepirone steady-state parameters C_{max} and AUC_{0-24} (20 and 29 times, respectively). For 1-PP, no drug interaction effect of rifampicin on C_{max} was found. For AUC_{0-24} the conclusion was indeterminate. For 3'-OH-gepirone, the AUC_{0-24} after the combination treatment was on average 3 times lower and the C_{max} on average 2.5 times lower than gepirone alone treatment, indicative of an interacting effect. Gepirone alone did not induce CYP3A4. The conclusion is that co-administration of gepirone-ER with a strong P450-3A4 inducer markedly reduces the oral bioavailability of gepirone... (IB5, p. 94; Label, p. 14)

21. The CYP2C19 and CYP2D6 isoenzymes are minor elimination pathways in the metabolism of vilazodone, and concomitant administration of inhibitors of these isoenzymes is not thought to be important.

Gepirone is minimally metabolized by CYP450-2D6. (IB5, p. 99)

22. In vitro studies have shown that CYP1A2, CYP2A6, CYP2C9, and CYP2El make minimal contributions to the metabolism of vilazodone.

In vitro studies have shown that CYP450-1A2, CYP450-2C9, CYP450-2C19, CYP450-2D6, and CYP450-2E1 make minimal contributions to the metabolism of gepirone. (IB5, p. 99)

23. Effect of Vilazodone on Other Drugs. *In vitro* data suggest that co administration of vilazodone with substrates for CYPIA2, CYP2C9, CYP3A4, or CYP2D6 is unlikely to result in clinically significant changes in the concentrations of substrates for these enzymes.

In vitro data suggest that co-administration of gepirone with substrates for CYP450-1A2, CYP450-2C9, CYP4502D6, CYP450-2E1, or CYP450-3A4 is unlikely to result in clinically significant changes in the concentrations of substrates for these enzymes. (IB5, p. 99)

PHASE 2 DOSE-FINDING TRIALS

24. Five trials l using doses from 5 to 100 mg/d were conducted as part of the phase 2 program for vilazodone, but only 2 used fixed-dose designs (trials 246 and 248) that were informative about dose-response.

There are no studies designated phase 2 in the gepirone-ER program. This is because the sponsor at the time, Bristol Myers Squibb, had already progressed gepirone-IR to phase III. Four early gepirone-ER studies (105-052, 105-053,105-078, and 105-083), were performed by Bristol Myers and all were prematurely terminated for business reasons. The object of these four trials was (as a phase II program should be) to determine the optimal dose and regimen for gepirone-ER. Doses from 5 mg to 100 mg were studied. Bristol Myers believed that the effective dose of gepirone-ER should be less than 40 mg/day, so most of these studies were under dosed. (IB5 p108). (ISE, p. 120)

25. The other 3 used flexible titration designs (trials 244, 245, and 247).

The other 4 trials used flexible dose designs. (CSR CN105-052, CN105-053, CN105-078, CN105-083)

26. All were double-blind, randomized, placebo-controlled, 8-week, parallel group trials in outpatients meeting *DSM-IV* criteria for MDD and had sample sizes ranging from 86 to 140 patients per treatment arm.

All were double-blind randomized placebo controlled 6 or 8 week parallel group trials in outpatients meeting DSM-III criteria and had sample sizes from 20 to 60 per group. (ISS, Table 1, pp. 33-40)

27. The Hamilton Depression Rating Scale (HDRS) was the primary efficacy measure in all 5 trials, and the drug effect was assessed by comparing change from baseline at week 8 on the sum of the scores for the first 17 items (HDRS) in the drug and placebo groups.

In trials *CN105-052, CN105-053, CN105-078, CN105-083 the primary efficacy parameter was the HDRS first 17 items, change from baseline at week 8 (week 6 in CN105-078). (ISE pp29-31)*

28. Missing data were imputed using a last observation carried forward (LOCF) approach.

The LOCF approach was used in gepirone-ER studies. (ISE, p. 32)

29. Three of these trials included an active comparator to assess for assay sensitivity (ability of the trial to detect the treatment effect of a drug known to be effective).

In all, there were 5 trials with an active comparator, CN105-052 and CN105-078, in the above early phase III program and 134004, 134006, and 134017 in a later Phase III program. .In the 134004 and 134006 trials HDRS-25 total score change from baseline was used as the primary efficacy parameter. Study 134017 used MADRS total score change from baseline as the primary efficacy parameter. (ISE, p. 26 and appendix C)

30. Dosing in these trials is provided in Table 1.

Dosing (and other parameters) in the gepirone-ER trials are as follows:

Trial	*number per arm*	*flexible dose*	*active comparator*	*Primary Endpoint*
CN105-052	*37*	*20-60 mg*	*fluoxetine*	*HDRS-17*
CN105-053	*58*	*10-60 mg*	*Imipramine*	*HDRS-17*
CN105-078	*30*	*10-100 mg*	*none*	*HDRS-17*
CN105-083	*45*	*10-100 mg*	*none*	*HDRS-17*
134004	*135*	*20-80 mg*	*fluoxetine*	*HDRS-25*
134006	*150*	*20-80 mg*	*paroxetine*	*HDRS-25*
134017	*165*	*20-80 mg*	*fluoxetine*	*MADRS*

31. None of the 5 trials demonstrated a statistically significant treatment effect of vilazodone on its primary endpoint.

None of the 7 trials demonstrated a statistically significant treatment effect of gepirone-ER on its primary endpoint. (ISE, pp. 21-22)

32. Moreover, none of the 3 active comparator trials showed the comparator to be statistically distinguishable from placebo, indicating that these trials lacked assay sensitivity for the primary trial endpoint (HDRS) and were uninformative about vilazodone's effect on that endpoint.

Moreover, none of the 5 active comparator trials showed the comparator to be statistically distinguishable from placebo on the primary efficacy parameter, indicating they lacked assay sensitivity for the trial and were uninformative about gepirone-ER's effect on that endpoint. (ISE, p. 22)

33. The 2 trials lacking an active control group could be considered "negative" trials for vilazodone, but this is not certain in the absence of an active control to confirm assay sensitivity.

The 2 trials lacking an active control group (CN105-078 and CN105-083) could be considered "negative trials" for gepirone-ER but this is not certain in the absence of an active control to confirm assay sensitivity. (ISE, p. 22)

34. Although these trials were unsuccessful on their primary endpoints, the 2 fixed-dose trials (246 and 248) showed dose responses for the difference between vilazodone and placebo on a secondary endpoint, the Montgomery-Asberg Depression Rating Scale (MADRS), suggestive of a treatment effect for the 20-mg/d vilazodone groups (the nominal unadjusted p values were 0.06 in both trials).

Data from adequate and well controlled studies were examined for a relationship between the total daily dose and the therapeutic effect. Evidence of efficacy was consistently associated with a dose of 40 mg/day or higher. Studies with mean maximum dose below 40 mg/day tended to fail... (IBp108)

35. The secondary endpoint data for the MADRS total score are summarized in Table 2.

Two of the gepirone-ER studies had positive results on secondary efficacy parameters and are supportive. For CN105-078, a post-hoc mixed model analysis was performed which showed that the high dose group (20-100mg/day) was statistically better than placebo for the mHDRS-17, Bech-6, and Item 1 (Table 22 and 2003 Summary of Benefits and Risks, Appendix A-2) For study CN105-083 there was significant site interaction between the two sites preventing pooling. In one site, in the high dose group (20-100 mg/day) statistically significant results were found for HDRS-17, CGI, HDRS item 1 and MADRS.(ISE pp59-62).

36. Two nearly identical efficacy trials were conducted in the phase 3 program for vilazodone (trials 04 and 07).

For gepirone-ER, five nearly identical trials where conducted in the late phase III program (134001, 134002, FKGBE007, FKGBE008, and 134023). (ISE, p. 23 – 25) (ISE p43-45)

37. These were multicenter (all US sites), randomized, double-blind, parallel group, placebo-controlled, short-term (8-week) trials of vilazodone in adult patients (ages 18 to 70 years) meeting *DSM-IV-TR* criteria for MDD, single episode or recurrent.

These were US multicenter randomized double blind parallel group controlled short term studies age 18-65 meeting DSM-IV criteria for MDD, single or recurrent. (ISS, Table 5, p. 43 – 45)

38. Patients were required to have, at screening and baseline visits, an HDRS score of greater than: 22 on the first 17 items of the 21-item HDRS, and an HDRS item 1 (depressed mood) score of at least: 2.

Patients were required to have an HDRS score on the first 17 items of at least 20 at screening and baseline. No restriction was made on the HDRS item 1. (CSR 134001, FKGBE007)

39. Although these trials included only a single fixed dose of 40 mg/d, in trial 04, patients who did not tolerate the 40-mg dose could be maintained on 20 mg/d.

The subjects were titrated from 20 mg up to a maximum of 80 mg/day. Subjects who could not tolerate at least 40 mg/day were discontinued. (CSR 134001, FKGBE007)

40. Randomization was 1:1 vilazodone 40 mg/d versus placebo in both trials.

Randomization was 1:1 in both trials. (ISE pp26-27)

41. Vilazodone was initiated at 10 mg/d for 7 days, then increased to 20 mg/d for 7 days, and maintained at 40 mg/d for weeks 3 through 8.

Gepirone-ER was initiated at 20 mg/day the first 3 days, then 40 mg/day for the next week, then 60 mg/day for the next week, then 80 mg/day. (ISE p28)

42. Vilazodone was taken with food in both trials.

Gepirone-ER was to be taken once a day in the morning consistently with food in all trials. (ISE p28, Label, p. 4)

43. Neither trial included an active comparator.

In all 5 studies there were no active comparators. (ISE p27)

44. Unfortunately, there was no additional dose finding in phase 3.

.In Study 134001, the final prescribed dose was 60 mg/day for 18.8% of subjects and 80 mg/day for 66.3% of subjects. In Study FK-GBE-007, the final prescribed dose was 60 mg/day for 22.4% of subjects and 80 mg/day for 65.5% of subjects. Based on these data, the recommended usual target dose for gepirone ER is 60-80 mg/day. (ISE p92)

45. The rationale for the single 40-mg/d dose was based on data that were less than persuasive.

For gepirone-ER, data from adequate and well controlled studies were examined for a relationship between the total daily dose and the therapeutic effect. Evidence of efficacy was consistently associated with a dose of 40 mg/day or higher. Studies with mean maximum dose below 40 mg/day tended to fail. (IBp108)

46. The 40-mg/d dose was not specifically evaluated in the phase 2 studies, but the 20-mg/d dose appeared to be at the lower margin of efficacy.

A, meta-analyses of five efficacy parameters(HDRS-17, MADRS, HDRS-14, Bech-6, and CGI responders) indicate that the difference between 80 mg gepirone ER dose and placebo is statistically significant in the LOCF dataset for all efficacy parameters evaluated. Using the LOCF approach, the difference between 60mg gepirone ER dose and placebo was statistically significant for the Bech-6 and showed a trend toward significance for the HDRS-14 and MADRS. Neither the 40mg nor the sub-therapeutic20 mg dose groups were significantly different from placebo on any measure using the LOCF or OC approach.(ISE p 105 and Appendix A, Tables 10.8, 10.9, and 10.10,)

47. Of note, however, because the phase 2 studies with an active control group failed to demonstrate assay sensitivity, they cannot be interpreted as providing adequate assessment of the 20-mg vilazodone dose (or any dose, for that matter).

Doses of Gepirone-ER below 40 mg/day have been shown to be ineffective. (IB p108).

48. In addition, 5-HT_{1A} receptor occupancy is thought to correlate with salutary drug effects, and a positron emission tomography study was conducted to assess 5-HT_{1A} receptor occupancy in vivo.

No PET study was done with gepirone-ER.

49. There was evidence of modest receptor occupancy with 40-mg vilazodone (15%-35%) but no evidence of occupancy with 20-mg findings interpreted as supportive of the 40-mg dose.

No PET study was done with gepirone-ER

50. The positron emission tomography study was conducted using only single doses of vilazodone.

51. Overall, for these phase 2 trials, depression symptoms tended to decrease as daily doses increased to 20 mg.

Doses of Gepirone-ER below 40 mg/day have been shown to be ineffective. (IB p108).

52. There was no useful information from these trials regarding dose response for efficacy at doses above 20 mg/d.

For gepirone-ER data from adequate and well controlled studies were examined for a relationship between the total daily dose and the therapeutic effect. Evidence of efficacy was consistently associated with a dose of 40 mg/day or higher. Studies with mean maximum dose below 40 mg/day tended to fail. (IB p108)

53. The frequency of adverse events increased with dose, and doses above 40 mg/d were poorly tolerated in studies 244 and 245.

For gepirone-ER the only adverse event which increases with dose is dizziness. Doses above 80 mg/day are poorly tolerated (ISS p434).

54. Thus, neither the phase 2 studies nor the positron emission tomography study provides conclusive evidence that vilazodone at doses less than 40 mg/d are ineffective.

The dose finding studies for gepirone-ER conducted by Bristol Myers Squibb were labeled as phase III, but should have been labeled as phase II. The major conclusion was gepirone-ER doses under 40 mg are not effective. (ISE p108)

55. The MADRS was the primary efficacy assessment, and it was conducted at baseline and weeks 1, 2, 4, 6, and 8. The primary endpoint was the difference between drug and placebo in change from baseline to week 8 on the MADRS total score.

The HDRS-17 was used as the primary efficacy assessment in the gepirone-ER studies and it was conducted at baseline and weeks 1, 2, 4, 6, and 8. The primary endpoint was the difference between drug and placebo in change from baseline to week 8 on the HDRS -17 total score. In most phase III studies the MADRS was also used as a secondary efficacy endpoint. (ISE p27 and appendix C))

56. The primary analysis was analysis of covariance, based on a modified intent-to-treat (ITT) population: patients who took at least 1 dose of their assigned treatment and who had baseline and at least 1 follow-up efficacy assessment.

The primary efficacy analysis variable for gepirone-ER studies was analysis of variance (ANOVA). The analysis was performed on the intent-to-treat (ITT) population: patients who took at least 1 dose of their assigned treatment and who had baseline and at least 1 follow-up efficacy assessment. (ISE p27 and appendix C)

57. Missing data were imputed using an LOCF approach. Clinical Global Impressions-Improvement (CGI-I) and Severity of Illness (CGI-S) scales were included among several secondary endpoints.

Missing data were imputed by the LOCF method. CGI-I and CGI-S were also measured as secondary efficacy endpoints. (ISE p27 and appendix C))

58. For trial 07, there were approximately 230 patients per group in the ITT population, with approximately 19% dropouts (roughly the same for both groups).

For trial 134001, there were approximately 103 subjects per treatment group in the ITT population, with approximately 27% gepirone-ER and 24% placebo dropouts. (CSR 134001)

59. Placebo had a higher percentage dropout for lack of efficacy (15% for placebo and 6% for vilazodone) and a lower percentage dropout for adverse events (9% for placebo and 17% for vilazodone), compared to vilazodone.

In study 13400, placebo and gepirone-ER treated subjects dropouts for lack of efficacy were essentially the same at 3.9% and 3.8%. Gepirone-ER treated subjects had 9.8% dropouts for adverse events compared to placebo 2.8%. (CSR 134001p67)

60. The mean age of the sample was 42 years, and 56% of patients were female. The mean baseline MADRS total score was 32, and the least squares mean changes from baseline to week 8 were -10.8 (placebo) and -13.3 (vilazodone).

In study 134001, the average age was 40 years and 60.6% were female. (CSR 134001) The mean baseline MADRS in study 134001 was 30 and the least squares mean change from baseline to week 8 are -9.22 placebo and -12.28 gepirone-ER (ISE p49,ISE p70)(CSR 134001 appendix F8.6.1.1-2, ISE p70

61. The difference between groups in change from baseline was -2.5 (standard error [SE] of the mean=0.96; 95% confidence interval [CI], -4.4 to -0.6; p = 0.009).

The difference between the groups in 134001 in change from baseline was -3.06 (SE of the mean.1.36, CI -5.95-0.57 p = 0.018. (ISE p70)

62. A mixed-effects model for repeated measures (MMRM) analysis, a widely used and generally preferred alternative to LOCF, also significantly favored vilazodone over placebo on the MADRS, as did analyses with CGI-Sand CGI-I

For 134001, the primary endpoint was the HAMD-17, p = 0.013. Secondary endpoints confirming the efficacy of gepirone-ER are: HAMD-17 remitters, p = 0.017; HAMD-21, p = 0.021; HAMD-28, p = 0.013; HAMD item 1, p = 0.005; CGI-S, p = 0.016; HAMD-25, p = 0.007; and HAMD-25 responders, p=0.014. No mixed model analysis was performed. (CSR 134001 p47)

63. In trial 04, there were approximately 204 patients per group in the ITT population, with approximately 25% dropouts (roughly the same for both groups).

For trial FKGBE007, there were 116 patients in the gepirone-ER treatment group and 122 in the placebo treatment group in the ITT population, with 22% dropouts for the gepirone-ER group and 18% dropouts for the placebo group. (CSR FKGBE007, Tables 9 and 11, pp. 66 – 67)

64. Placebo again had a higher percentage dropout for lack of efficacy (18% for placebo and 8% for vilazodone) and a lower percentage dropout for adverse events (20% for placebo and 36% for vilazodone) compared to vilazodone.

In FKGBE007, there were 3.2% dropout in the gepirone-ER group and 2.4% in the placebo group for lack of efficacy, and 4.0% in the gepirone-ER group and 2.4% in the placebo group for adverse events. (CSR FKGBE007 Table 9, p. 66)

65. The mean age of the sample was 40 years, and 63% of patients were female.

The mean age of the FBGBE007 sample was 38.1 years and 68.9% were female. (CSR FKGBE007, Table 12, p. 69)

66. The mean baseline MADRS total score was 31, and the least squares mean changes from baseline to week 8 were -9.7 (placebo) and -12.9 (vilazodone).

The mean baseline MADRS was 30.5 and the least squares change from baseline to week 8 were -9.9 placebo and -13.7 gepirone-ER. (CSR FKGBE007, Table 20, p. 85)

67. The difference between groups in change from baseline was -3.2 (SE =0.99; 95% CI, -5.1 to -1.2; p =0.001.

The difference in the two groups in change from baseline favored gepirone-ER -3.78, 95% CI -6.58 to -0.98, p = 0.008. (CSR FKGBE007, Table 20, p. 85)

68. The MMRM analysis also favored vilazodone over placebo on the MADRS, as did analyses with CGI-S and CGI-I.

In FKGBE007, the primary efficacy parameter was HDRS-17 significant at p = 0.032, and other secondary efficacy parameters: HDRS-21 p = 0.043; HDRS-25 p = 0.029; HDRS-28 p = 0.032; CGI p = 0.015; and the Bech 6 core depression items p = 0.016. No mixed model analysis was done on this study. (CSR FKGBE007, Table 18, p. 82)

Two other trial 134002 and FKGBE008 were supportive of the positive trials. While the primary endpoint was not significant, several secondary endpoints point toward efficacy with gepirone-ER. For 134002, mixed models analysis produced significant results for mHDRS-17, Bech-6, item 1 and MADRS, all p<.0.05. ForFKGBE008, weeks 2, 3, 6 HDRS-17 and weeks 2, 3, 4 MADRS, p<0.05 (2003 Summary of Benefits and Risks, Tables 3 and 7)

In study 134023, there were no significant differences between gepirone-ER and placebo, but there was no comparator group. (ISE, p. 25)

69. Forty-one patients could not be titrated to the highest dose because of intolerability, and they were maintained on the middle dose, i.e., 20 mg in the vilazodone group and 1 placebo pill for the placebo group.

In study FKGBE007, subjects that could not tolerate 40 mg/day 2 tablets of gepirone-ER or placebo were discontinued. (CSR FKGBE007, p. 38)

70. There were 28 such patients in the vilazodone group and 13 in the placebo group.

In study FKGBE007, subjects that could not tolerate 40 mg/day 2 tablets of gepirone-ER or placebo were discontinued. (CSR FKGBE007, p. 38)

71. Eleven of these patients completed the study, i.e., 6 took vilazodone and 5 took placebo.

Two subjects in study FKGBE007 taking 20 mg completed the study, in violation of the protocol.

72. On the basis of an exploratory analysis of this small subgroup taking 20 mg/d, the drug minus placebo difference in least squares mean change from baseline on the MADRS total score was estimated to be -4.3 (95% CI, -11.6 to 2.9), suggesting that explorations of this dose would be worthwhile.

The results of FKGBE007 indicate that no exploration of a 20 mg/day dose is warranted. (CSR FKGBE007)

73. The MMRM analyses for trials 07 and 04 provided results (drug-placebo differences in change from baseline in MADRS total score and unadjusted p values) by treatment visit over the course of these 8-week trials, as illustrated in Table 3.

No MMRM analyses were done in studies 134001 and FKGBE007.

For 134002, mixed models analysis produced significant results for mHDRS-17, Bech-6, item 1 and MADRS, all $p<.0.05$. ForFKGBE008, weeks 2, 3, 6 HDRS-17 and weeks 2, 3, 4 MADRS, $p<0.05$ (2003 Summary of Benefits and Risks, Tables 3 and 7) (ISE, p. 25)

74. There have been published suggestions that vilazodone may have a rapid onset. The available data do not support such a conclusion.

For gepirone-ER, the results for the primary efficacy parameter (HDRS-17 change from baseline LOCF) for study 134001 are week 1, $p = 0.052$; week 2 $p = 0.059$; week 3 $p = 0.01$; week 4 $p = 0.078$; week 6 $p = 0.051$; and week 8 $p = 0.018$. For the same efficacy endpoint in study FKGBE007, the results are week 2 NS; week 3 $p = 0.08$; week 4 $p = 0.004$; week 6 $p = 0.006$; and week 8 $p = 0.032$. These results would suggest that gepirone-ER has an early onset of action. However, the studies were not designed for this and FKP is not making such a claim. (CSR 134001, p. 78; CSR FKGBE007, p. 73)

75. First, none of the trials purporting to show an early effect included an active control.

For gepirone-ER, neither of the positive trials, CSR 134001 and FKGBE007, had an active control. (CSR 134001 and FKGBE007)

76. Second, the largest effects in both trials are observed only at 6 to 8 weeks, similar to what is seen with other antidepressants.

As mentioned above, the studies with gepirone-ER show efficacy at 1, 2, 3, and 4 weeks with large effects at week 3 in 134001 and week 4 for FKGBE007. (CSR 134001 and FKGBE007)

77. Whether the statistically significant vilazodone-placebo treatment differences observed at the week-1 time point in trial 04 represent a clinically relevant treatment effect is unknown, and it is noteworthy that this finding was not replicated in trial 07.

The early onset of antidepressant activity seen in 134001 was replicated in FKGBE007. However, in the absence of a control drug, no claim of early onset is being made. (CSR 134001 and FKGBE007)

78. The only longer term data were from a 1-year open-label trial that could not be interpreted because of the lack of a control group.

For gepirone-ER an extensive program of long term safety and efficacy studies, which included both placebo and standard antidepressant controls, were carried out in extension studies, 2080 patients received gepirone, 696 placebo, 400 fluoxetine, 105 paroxetine, 104 imipramine, and 49 anxiolytics. (ISS, p. 383)

Nine hundred five gepirone treated patients completed 6 months, 271 placebo, 220 fluoxetine, 115 paroxetine, and 48 imipramine. (ISS, p. 383)

Two hundred sixty one gepirone treated patients completed one year, 65 placebo, 28 fluoxetine and 25 imipramine. (ISS, p. 383)

. *Long term adverse event data indicate that no new adverse events occur after the initial 8 week studies. (ISS, p. 411)*

79. In evaluating the safety of a new drug, the FDA pays particular attention to all deaths, other serious adverse events, and adverse events occurring in patients who dropped out of a study (adverse dropouts).

In a total 4976 subject treated with gepirone, death occurred rarely. There were 8 treatment-emergent deaths (4 gepirone ER subjects, 1 gepirone IR subject, 1 paroxetine subject, 1 fluoxetine subject, and 1 placebo subject). One additional subject died during a placebo wash-out period prior to randomization. The gepirone deaths were not treatment-related. (ISS 264-267)

80 .Serious adverse events are defined as deaths, nonfatal life-threatening adverse events, events leading to inpatient hospitalization or prolongation of a hospitalization, a persistent or significant disability or incapacity, or a congenital anomaly or birth defect.

81. The FDA relies on data derived from placebo-controlled trials to establish the common adverse event profile for a new drug.

82. In addition, the FDA evaluates laboratory, vital sign, and electrocardiographic data and data from special studies, e.g., a thorough QT study, and assessments to further establish a drug's safety profile.

83. The development program for vilazodone in MDD included data from 24 phase 1 trials, 1 the 5 phase 2 trials l noted above, and 3 phase 3 trials (the 2 described above, and a I-year open-label trial, all in adults).

The safety data population for gepirone-ER included data from 34 phase I trials and 42 phase II/III trials, and 11 open label studies. (ISS, p.32)

84. These 32 trials included a total of 2,898 subjects exposed to 1 or more doses of vilazodone.

These 87 trials included a total of 5868 subjects exposed to 1 or more doses of gepirone. (ISS, p. 42)

85. The 24-trial phase 1 program included 721 subjects exposed to vilazodone doses ranging from 1 to 80 mg in single- and repeat-dose trials.

For gepirone-ER, in the 34 phase I trials, 786 subjects were exposed to gepirone in doses from 1 to 100 mg. (ISS, p. 33-39)

86. The FDA's safety review focused primarily on the 8 phase 2 and 3 trials, including 2,177 patients exposed to vilazodone.

The gepirone-ER safety review focused primarily on 25 phase II/III studies that were placebo controlled, including 3117 subjects exposed to gepirone. (ISS, p.42)

87. Seven of these were placebo-controlled, 8-week trials that included 1,578 patients with MDD exposed to vilazodone.

For gepirone-ER, all 3117 were in placebo controlled studies. (ISS, p. 42)

88. The eighth trial was an uncontrolled, open-label trial involving 599 patients exposed to vilazodone for up to 1 year.

In addition, in extension studies, 2080 patients received gepirone; 696 placebo; 400 fluoxetine; 105 paroxetine; 104 imipramine; and 49 anxiolytics studied for up to one year. (ISS, p. 383)

89. Overall, the vilazodone exposure was 552 patient-years.

Overall in the phase II/III gepirone treated group, exposure was 552.5 patient years. (ISS, p.115)

90. The phase 2 and 3 trials included only 309 patients aged 55 years or older and only 37 patients aged 65 years and older, respectively, a common pattern but one that largely omits potentially vulnerable populations (elderly and patients with concomitant illness).

In the gepirone program, the age cut off was 65 years old (not 55), and 337 subjects over 65 were given gepirone. A specific study of the elderly (over 65), study 28715, was completed with 137 elderly receiving gepirone. There was no increased incidence of difference in adverse events in the elderly compared to subjects under 65 years old. (ISS, 2003, -p. 663)

91. There were no pediatric patients exposed to vilazodone in this program.

The pediatric program for gepirone-ER has been completed. There are two pediatric pharmacokinetic studies (28721 and FKGBE009) exploring children 6-12 and adolescents 12-17. There are two finished depression efficacy studies 134019 and 134020 in the same populations. Extensions to these studies form the long term data in study 134507. (ISS, 2003, p. 668-669; IND Amendments 178 and 179)

AUC_{0-24} and C_{max} are higher in children than adolescents. No greater incidence or different adverse events were seen in the pediatric studies. (ISS, 2003, p. 670; IND Amendments 178 and 179)

92. There were no deaths in the vilazodone program in normal subjects or patients taking vilazodone.

In 5868 subjects receiving gepirone there were 5 deaths. None were attributable to gepirone. There were 2 deaths in subjects on placebo, one on fluoxetine, and one on paroxetine. (ISS, p. 264-265)

93. Overall, 81 patients experienced a nonfatal serious adverse event, including 5 in phase 1 studies and 76 in phase 2 and 3 studies.

In 3483 subjects treated with gepirone, there were 74 serious adverse events 2.2%, compared to placebo 25 events in 2325 subjects 1.1%; fluoxetine 22 events in 665 subjects 3.3%; and paroxetine 14 in 276 subjects 5.1%. (ISS, p. 273)

94. These events are difficult to assess for causality.

95. Overall, however, in the phase 2 and 3 studies, the proportions of patients experiencing nonfatal serious adverse events appeared to be similar for vilazodone and placebo patients.

Overall, the serious adverse events for gepirone are similar to placebo and better than the SSRIs fluoxetine and paroxetine. (ISS, p. 273)

96. Many of the serious adverse events were psychiatric in nature and probably represented worsening of the underlying condition being treated; such events are expected and observed in any psychiatric drug development program.

For gepirone, 27 of the 74 serious adverse events (36%) were psychiatric in nature. (ISS, p. 273)

97. The others were common background events, such as pneumonia, prostate cancer, and cholecystitis, with no pattern of findings suggesting that any particular event was more common in patients exposed to vilazodone than in those exposed to placebo.

For gepirone, the other serious adverse events were largely pregnancy, coronary heart disease, and infections. No pattern emerged that suggested that gepirone was involved. (ISS, p. 273)

98. In the placebo-controlled trials, approximately 7% of vilazodone-exposed patients discontinued because of an adverse event compared to 3% of placebo-exposed patients, although there was no single adverse event leading to discontinuation in >1% of patients.

For gepirone-ER, 16.3% had at least one adverse events leading to discontinuation compared to placebo 6.8%, fluoxetine 9.3%, paroxetine 11.6% and imipramine 23.8%. (ISS, p. 285)

99 The most common adverse events leading to discontinuation were diarrhea, nausea, palpitations, and fatigue.

For gepirone-ER, the most common adverse events leading to discontinuation are: dizziness, nausea, headache, insomnia, and anxiety. (ISS, p. 321)

100. Certain adverse events are expected in SSRI-treated patients and were of particular interest in the vilazodone development program: Serotonin syndrome: The New Drug Application database was searched using terms suggestive of possible serotonin toxicity.

101. Two patients were identified with such events, including 1 patient in the long term, open-label, and phase 3 study.

For gepirone, no subject in the ISS, database experienced an AE described as "serotonin syndrome" by the investigator or based on predefined criteria. (ISS, p. 159)

102. This patient took an overdose of approximately 240 mg of vilazodone.

103. The other patient had received a vilazodone dose of 80 mg in a phase 2 study.

104. Mania/hypomania: A search of the database discovered 6 patients with probable treatment-emergent mania or hypomania (5 taking vilazodone and 1 taking placebo).

In the gepirone total database, there were 7 subjects on gepirone reporting mania and 3 patients on placebo. (Label, p. 7)

105. Bleeding: Although bleeding events were identified in the vilazodone database, the proportions of patients experiencing such events were similar in vilazodone and placebo groups, i.e., approximately 3% for each.

In the entire database gastrointestinal bleed was found in 2 subjects taking gepirone and 1 taking placebo. A drug-drug interaction study of gepirone-ER with warfarin indicated no pharmacokinetic or pharmacodynamic interaction and no change in prothrombin time. Gepirone-ER has been used with aspirin and non-steroidal anti-inflammatory drugs with no indication of bleeding. (Risk Benefit, p. 51; Label, p. 24; ISS, Table 5.5.2)

Common, Non-serious Adverse Events

106. On the basis of a pooling of the 2 phase 3 trials, the FDA has identified 4 adverse events that it considers common and drug-related, i.e., occurring at a rate of at least 5% with vilazodone and having a rate at least twice the placebo rate.

For gepirone-ER, there are only two common drug related adverse events that are at least 5% and at least twice placebo: dizziness and nausea. (ISS, Table 32)

107. The rates for these 4 adverse events are as follows: diarrhea (28%, vilazodone vs. 9%, placebo), nausea (23%, vilazodone vs. 5%, placebo), vomiting (5%, vilazodone vs. 1%, placebo), and insomnia (6%, vilazodone vs. 2%, placebo).

The rates of these two adverse events are as follows: dizziness 31% gepirone, 13.3% placebo; and nausea 27% gepirone; and 11% placebo. (ISS, p. 119)

108. The gastrointestinal adverse events and insomnia tended to occur early in treatment.

The dizziness and nausea occur early in treatment, have a short duration and tolerance develops. (Label, p.9)

109. Additional adverse events identified from this pooling as having a rate of at least 2% and at least twice the placebo rate (an approach to judging possible causality) included the following: gastroenteritis, paresthesia, tremor, abnormal dreams, restlessness, decreased libido, abnormal orgasm, delayed ejaculation, erectile dysfunction, feeling jittery, palpitations, and increased appetite.

The remaining gepirone-ER adverse events that are at least 2% and twice placebo are: paresthesia, tremor, abnormal dreams, and middle insomnia. (Label, p.10)

110 There is strong evidence for dose relatedness for many of these common adverse events, with poor tolerability of doses above 40 mg for most patients.

For gepirone-ER, there is evidence of a dose relationship for only one adverse event, dizziness, with poor tolerability over 80 mg/day. (ISS, p. 434)

Sexual Dysfunction

111. Other SSRIs have been shown to cause male and female sexual dysfunction, and adverse events of sexual dysfunction (decreased libido, abnormal orgasm, delayed ejaculation, erectile dysfunction, and sexual dysfunction) were reported more frequently for vilazodone compared to placebo in the phase 2 and 3 studies.

112. The overall percentages of vilazodone treated patients spontaneously reporting sexual dysfunction were quite low (5% or less), but such low rates are observed with all SSRIs, and it is well-recognized that reporting of sexual dysfunction is severely underreported in trials that do not devote particular attention to eliciting such effects.

The overall incidence of sexual adverse events for gepirone-ER was 2.4%. (ISS, Table 14.1.2.1.1)

113. Indeed, such adverse events were specifically identified in the FDA's gender guidance as needing special attention, particularly in women.

Gepirone-ER's sexual adverse events of 2.4% were numerically fewer and not statistically different from placebo subjects 3.9% p = 0.121. (ISS, Table 14.1.2.1.1)

114. Placebo-controlled trials comparing bupropion, a non-SSRI antidepressant, with SSRIs (either sertraline or escitalopram) that used specific approaches to elicit sexual dysfunction demonstrated a high rate of sexual dysfunction with typical SSRIs.

In gepirone-ER studies, the SSRIs fluoxetine had asexual adverse event rate of 12.5% and paroxetine had a rate of 21.1%. (ISS, Table 14.1.2.1.1)

115. The pharmaceutical sponsor did collect data using specific sexual function scales (Arizona Sexual Experience Scale [ASEX and Changes in Sexual Function Questionnaire) in 2 of its trials, but the results were inconsistent, showing worsening on some items and improvement on others, with no clear pattern of change for vilazodone compared to placebo.

The Changes in Sexual Function Questionnaire was used in one gepirone-ER study and the Derogatis Inventory of Sexual Function was used in 7 gepirone-ER studies. The results were consistent showing gepirone-ER treated subjects with statistically significantly better sexual function than placebo or SSRI treated subjects. (ISS, p. 235)

116. Unfortunately, these trials did not include an active control SSRI known to cause dysfunction, so no evidence exists that these trials were capable of detecting such adverse events.

SSRI active controls were used in three gepirone-ER trials (134004, 134006, and 134017) and produced sexual adverse events, verifying assay sensitivity. (CSR 134004,134006, 134017)

117. Published articles commenting on the trial including the ASEX assessments suggested that vilazodone did not cause sexual dysfunction, but these conclusions are not supported by the data.

Four published articles in peer reviewed journals comment on the effect of gepirone-ER on sexual function and the comments are supported by the data. (JSM 2010 7:1998-2014, JSM 2011 8:1411-1419, JSM 2011 8: 2569-2581, and JSM 2011 Aug 30 e-published ahead of print)

118. Without an active control that is known to cause sexual dysfunction to demonstrate assay sensitivity, the results from these trials cannot support any conclusion as to the absolute or relative effect of vilazodone on sexual function.

There are active controls which support the assay sensitivity of studies 134004, 134006, and 134017 for sexual dysfunction. Therefore the positive effect of gepirone-ER compared to placebo in these trials is substantiated. (ISS, p. 362)

Laboratories, Vital Signs, Weight, Electrocardiograms

119. Vilazodone was not associated with any clear finding of drug-related changes in laboratory parameters, vital signs, or weight in the placebo-controlled trials.

Gepirone-ER was not associated with any clear finding of drug related changes in laboratory parameters, vital signs, or weight in placebo controlled studies. (Label, p. 11 - 12)

120. Laboratory testing included routine serum chemistries, thyroid testing, routine hematology testing, urinalysis, and electrocardiograms.

Laboratory testing for gepirone-ER included routine serum chemistries, routine hematology testing, urinalysis, and electrocardiograms. No thyroid testing was routinely performed. (Label, p. 12)

121. Vital signs included blood pressure, pulse, and respiratory rate.

Gepirone-ER studies included blood pressure, pulse and respiratory rate. No drug related abnormalities were found. (Label, p.12)

122. A thorough QT study with vilazodone revealed that it does not prolong the QTc interval or cause any other important changes in electrocardiogram parameters.

No specific QT study was done with gepirone-ER; however ECGs were done in 4771 subjects at the clinical sites and 2614 centrally read. In Phase II/III studies 2006 ECGs were read at the sites and 2614 centrally. Examination of all of these ECGs revealed no evidence of a QTc effect for gepirone. (ISS, pp. 339 – 362)

123. Vilazodone was tested in this study at doses up to 80 mg.

Gepirone-ER doses were up to 100 mg/day. (ISS, Table 1, CSR 030L1-0001-1700, CN105-078, CN105-083)

124. The study demonstrated appropriate assay sensitivity, and the baseline corrected QTc interval was <10 milliseconds for vilazodone, below the threshold generally considered to indicate clinical concern.

There is consistent data that gepirone-ER does not increase the QTc interval. In fact, a lowering of QTc was found. These results are verified by the fact that subjects treated with imipramine, which could be considered a positive control, were significantly elevated. (ISS, p. 362, Risk Benefit p. 49, 50)

Overdose

125. Overdose experience with vilazodone is limited to observations in 5 patients who received doses in the range of 200 to 280 mg.

Six gepirone (IR or ER) overdoses occurred with doses up to 500 mg (although the exact dose in 2 of the cases was unknown). (Risk Benefit, pp. 44 - 45)

126. Clinical findings included serotonin syndrome, lethargy, restlessness, hallucinations, and disorientation.

The patient who took 500 mg of gepirone-ER with alcohol had no serious adverse events and no EKG abnormalities. One subject who injected 200 mg had right bundle branch block. Another patient who took 200 mg and a great quantity of other unknown drugs had a 60 second seizure. No deaths occurred. (Risk Benefit, pp.44 - 45)

Risk Information for Vilazodone's Label

127. Although no signal for excess suicidal ideation or behavior was observed in the development program for vilazodone, its labeling bears the box warning for suicidality, which all antidepressants have, based on findings from meta-analyses for drugs in this class.

Although no signal for excess suicidal ideation or behavior was observed in the development program for gepirone-ER, its labeling bears the box warning for suicidality, which all antidepressants have, based on findings from meta-analyses for drugs in this class. (Label, p. 5, 6)

128. As vilazodone's pharmacologic profile shows a prominent SSRI effect, it is expected to have the potential for other typical SSRI risks, including serotonin syndrome or neuroleptic malignant syndrome-like reactions, abnormal bleeding, activation of mania, discontinuation symptoms, hyponatremia, and seizures, and has warning language for all of these risks on its label.

Gepirone-ER is not an SSRI but does have serotonergic activity and therefore bears some of the warnings about potential serotonergic risks such as: activation of a bipolar patient to mania or hypomania, seizures, and possible interference with cognitive and motor performance. There are no warnings for neuroleptic malignant syndrome, abnormal bleeding, or hyponatremia. (Label, p.7, 8)

129. Given the concern for serotonin syndrome, vilazodone will need to be used cautiously with other drugs having serotonergic effects, including other SSRIs, serotonin-norepinephrine reuptake inhibitors, triptans, tramadol, buspirone, and tryptophan products.

Gepirone-ER should not be used with an MAOI or in patients that have recently discontinued MAOIs. However, gepirone-ER can be used with other SSRIs and SNRIs and with buspirone. Use with triptans, tramadol, and tryptophan have not been explored and gepirone-ER should be used cautiously with these drugs. (Label, p. 7, 8)

130. The concern for bleeding suggests that vilazodone will need to be used cautiously with aspirin or other non-steroidal anti-inflammatory drugs, or with warfarin and other anticoagulants.

A drug-drug interaction study of gepirone-ER with warfarin indicated no pharmacokinetic or pharmacodynamic interaction and no change in prothrombin time. Gepirone-ER has been used with aspirin and non-steroidal anti-inflammatory drugs with no indication of bleeding. (Risk Benefit, p. 51; Label, p. 24)

131. It also carries a contraindication for use with monoamine oxidase inhibitors (MAOIs).

Gepirone-ER should not be used with an MAOI or in patients that have recently discontinued MAOIs. (Label, p. 7)

132. There was no systematic assessment for discontinuation-emergent signs and symptoms, but given that such symptoms are expected with SSRIs, vilazodone labeling recommends tapering when it is discontinued.

Animal and human studies of gepirone-ER did not reveal any potential for drug seeking behavior or physical dependence. Therefore, no tapering is needed when gepirone-ER is discontinued. (Label, p. 16)

133. Standard animal reproduction studies of vilazodone did not indicate teratogenic potential.

Standard animal studies of gepirone-ER at doses approximately 36-48 times the maximum recommended human dose did not produce teratogenic effects. There are no well controlled studies in pregnant women. (Label, p. 15)

134. Nevertheless, because there are no controlled human data regarding vilazodone use during pregnancy, its label advises use during pregnancy only if the potential benefits outweigh the potential risks.

There are no well controlled studies of gepirone-ER in pregnant women. Therefore gepirone-ER should be used only if the potential benefit justifies the potential risk to the fetus. (Label pp 14-15)

135. Similarly, because there are no human data regarding vilazodone concentrations in breast milk, women are advised to breast-feed only if the potential benefits outweigh the potential risks.

Gepirone-ER is known to be excreted in the milk of laboratory animals. It is not known with regard to human milk. Therefore, caution should be exercised when gepirone-ER is administered to nursing women. Patients are advised to notify their physician if they are breastfeeding or plan to breastfeed. (Label, p. 15)

136. Vilazodone also has standard labeling language regarding possible neonatal complications associated with use of serotonergic antidepressants late in the third trimester.

Gepirone-ER has no such labeling language. However, the following language is proposed. "Gepirone-ER should be used in pregnancy only if the potential benefit justifies the potential risk to the fetus". (Label, p. 15)

137. Vilazodone is not classified as a controlled substance.

Gepirone-ER is not classified as a controlled substance. (Label, p. 16)

138. Animal studies did not reveal abuse or dependence potential; however, labeling acknowledges that its abuse potential has not been systematically evaluated in humans.

Animal studies and human clinical trials with gepirone did not reveal any potential for drug seeking behavior or physical dependence. (Label, p. 16)

MAJOR ISSUES IN APPROVAL DECISION

139. Although the FDA was satisfied that vilazodone has been shown to be an effective antidepressant in 8-weekstudies at a dose of 40 mg/d, there were no systematically collected longer-term controlled data to address the question of maintenance efficacy. These data are needed, as all antidepressants are used for long-term maintenance, and are ordinarily obtained after initial approval through randomized withdrawal trials in patients successfully treated for several months.

The FDA disallowed a post hoc per protocol analysis of study 28709. The analysis done by Organon was flawed and failed. The per protocol analysis of 28709 was provided to demonstrate that if the study had been done correctly it would have been a positive study. The per protocol analysis of study 28709 indicated that the efficacy of gepirone-ER was maintained up to one year. Further long term efficacy studies may be presented in phase IV. (ISEp112, table 48)

140There is also a question about the optimal dose for vilazodone. Some data in the program are suggestive of efficacy for the 20-mg dose, but more definitive data are needed. This is particularly relevant given that vilazodone is to be incrementally dose adjusted to the target of 40 mg and that it is most likely some patients will not receive the 40-mg dose because of side effects, notably gastrointestinal effects.

A, meta-analyses of five efficacy parameters(HDRS-17, MADRS, HDRS-14, Bech-6, and CGI responders) indicate that the difference between 80 mg gepirone ER dose and placebo is statistically significant in the LOCF dataset for all efficacy parameters evaluated. Using the LOCF approach, the difference between 60mg gepirone ER dose and placebo was statistically significant for the Bech-6 and showed a trend toward significance for the HDRS-14 and MADRS. Neither the 40mg nor the sub-therapeutic20 mg dose groups were significantly different from placebo on any measure using the LOCF or OC approach.(ISE p 105 and Appendix A, Tables 10.8, 10.9, and 10.10,)

141. Doses higher than 40 mg are poorly tolerated, and the incidence of certain adverse events is substantial even at 40 mg.

For doses of gepirone-ER greater than 80 mg, the adverse event of dizziness is poorly tolerated. (ISSp434)

142. Vilazodone has a tolerability profile similar to that observed with other SSRIs, however, and there were no safety problems identified that precluded an approval for this drug.

In FDA evaluations of gepirone-ER, no safety problems have been found that would preclude an approval for this drug. (FDA letter 11-02-2007)

PHASE 4 COMMITMENTS AND REQUIREMENTS

143. It is common for the FDA to obtain commitments from a sponsor to procure additional data for a new drug following the initial approval. In some instances, where trials involve a safety concern, such data can be required. The FDA has obtained the pharmaceutical sponsor's agreement to conduct the following studies post-approval:

Study of 20-mg dose: A study evaluating the efficacy of vilazodone at a dose of 20 mg/d. This is important because of the dose-related adverse events for vilazodone, particularly at doses of 40 mg and higher, and the limited study of the efficacy of 20 mg/d from the phase 2 trials (expected date for final report, January 31, 2014).

No post-approval clinical studies have been discussed with the FDA.

144. Maintenance study-A longer term randomized withdrawal trial to assess maintenance efficacy for vilazodone in MDD (expected date for final report, January 31, 2016).

The per protocol analysis of study 28709 provides evidence of long term efficacy (ISE p110)

145. Pediatric program-A pediatric program (ages 7-17 years) in MDD (expected date for final report, January 31, 2016).

The pediatric program for gepirone-ER has been completed. There are two pediatric pharmacokinetic studies (28721 and FKGBE009) exploring children 6-12 and adolescents 12-17. There are two finished depression efficacy studies 134019, 134020 in the same populations. Extensions to these studies form the long term data in study 134507. (IND serial numbers 178, 179)

146. CYP3A4 inducer study-A drug-drug interaction study of vilazodone with a 3A4 inducer, e.g., carbamazepine (expected date for final report, January 31, 2013).

The study of a CYP 450-3A4 inducer used rifampicin. The conclusion is that co-administration of gepirone-ER with a strong P450-3A4 inducer results in markedly reduced gepirone blood levels. (Label, p. 14)

147 Severe hepatic impairment-A pharmacokinetic study in patients with severe hepatic impairment (expected date for final report, February 28, 2013).

Gepirone-ER was administered to subjects with hepatic impairment in study FKGBE005. Patients with severe hepatic impairment (Child Pugh score of 10-12) were included. No further hepatic impairment studies are anticipated. (ISS, p.464)

148. P-glycoprotein interaction - An in vitro study to evaluate whether vilazodone is a substrate or inhibitor of P-gp (expected date for final report, December 31, 2011).

No p-glycoprotein studies have been conducted with gepirone-ER.

LESSONS LEARNED

There are always lessons to be learned from completed drug development programs, and the vilazodone program is no exception. For vilazodone, there are 3 issues in particular that deserve comment.

Dose Response

149 The dose-response characterization of vilazodone might have been handled differently.

The dose-response characterization of gepirone-ER might have been handled differently.

150. It appears that the phase 2 dose-response trials were neither optimally designed nor fully utilized in the design of the phase 3 trials.

The dose finding studies for gepirone-ER conducted by Bristol Myers Squibb were labeled as phase III, but should have been labeled as phase II. The studies were terminated prematurely for business reasons and had insufficient sample sizes to provide useful information. The major conclusion was gepirone-ER doses under 40 mg are not effective. Further analysis of the pertinent studies (134001 and FKGBE007) indicated that 60-80 mg was the most effective dose. (ISE, p. 106)

151. Early fixed dose trials should be the optimal source for selecting doses for subsequent studies.

More adequate early gepirone-ER fixed dose trials should have been carried out. It should be noted that Study 134015 evaluated the tolerability of a 40 mg titration schedule. The results of this trial suggest that gepirone-ER, at a dose of 40 mg per day for 7 days, is safe for subjects with MDD. On the other hand, tolerability of a starting dose of 40 mg/day appears limited by a high incidence of nausea and dizziness. (IB5, p. 100)

152. Only 2 of the 5 phase 2 trials, however, had fixed-dose designs.

There were no adequate fixed dose studies in the gepirone program. (ISE p106

153. Although 2 of the remaining 3 phase 2 trials did explore doses up to 100 mg/d, the flexible dose designs did not provide an opportunity for useful dose-response information to guide phase 3 dose selection.

Doses of gepirone-ER from 5 mg/day to 120 mg/day were explored. Better design would have resulted in better selection of later phase III trials.

154. Nevertheless, the 2 fixed-dose phase 2 trials did provide a rationale for evaluating a dose of 40 mg/d in phase 3 trials.

Early phase III gepirone-IR trials did provide evidence that 40 mg/day was the minimum effective dose. (ISE, p.92)

155. The other source of support for the 40-mg dose was the positron emission tomography trial; however, that trial was also not optimal, as it did not look at receptor occupancy at steady state plasma concentrations.

No positron emission tomography study was conducted with gepirone-ER.

156. A steady-state approach might have been more useful given that receptor occupancy was marginal after a single dose, i.e., 15%-35%, and might have been improved with the doubling of concentration at steady state.

A steady state PET scan for gepirone-ER would have been useful, but this technique was not available when gepirone was in Phase II in the late 1980s.

157. A repeat dose trial would also have provided an opportunity to explore a relationship between receptor occupancy and clinical response.

A steady state PET scan for gepirone-ER would have been useful, but this technique was not available when gepirone was in Phase II in the late 1980s.

158. In any case, although it was reasonable to study 40 mg, the phase 2 data gave no reason not to further study 20 mg (given the failure of active controls) and the sponsor missed an opportunity to definitively study the 20-mg dose by not including such an arm in these trials.

Lower doses of gepirone-ER have been adequately studied and have been found to be ineffective.

159. Recent publications have implied that vilazodone might have an early onset; but the phase 3 trials were not designed to examine this possibility.

For gepirone-ER, the results for the primary efficacy parameter (HDRS-17 change from baseline LOCF) for study 134001 are week 1, $p = 0.052$; week 2, $p = 0.059$; week 3, $p = 0.013$; week 4, $p = 0.078$; week 6, $p = 0.051$; and week 8, $p = 0.018$. The same efficacy endpoint was used in study FKGBE007. The results are week 2, NS; week 3, $p = 0.08$, week 4, $p = 0.004$; week 6, $p = 0.006$, and week 8, $p = 0.032$. These results would suggest that gepirone-ER has an early onset of action. However, the studies were not designed for this and FKP is not making such a claim. (CSR 134001, p78; CSR FKGBE007, p. 73)

160. These trials did not test hypotheses on time of onset, nor did they include other antidepressant treatment arms for comparison in time of onset.

The gepirone-ER trials did not test hypotheses on time of onset nor did they include antidepressant arms for comparison of time of onset. (CSR 134001; CSR FKGBE007)

161 Both would have been necessary to support such a claim.

Appropriate studies to determine time of efficacy onset could be done in the future...

162. In addition, it would have been necessary to replicate a finding of early onset of response.

The early response was shown in 134001 and FKGBE007, but appropriate studies as described above would have to be done to corroborate this finding...

163. As noted, it has also been suggested that vilazodone may have advantages over other antidepressants with regard to causing sexual dysfunction.

Gepirone-ER may have an advantage over other antidepressants with regard to causing sexual dysfunction. (Label, p.26)

164. The sponsor's goal was to demonstrate *no difference* for vilazodone versus placebo on sexual dysfunction.

Gepirone-ER had statistically significant better sexual function than placebo in 5 of 6 trials. (Label, p.26)

165. Importantly, the inability to find a statistically significant difference between the vilazodone and placebo groups on sexual dysfunction would not prove that there is no difference.

The difference in gepirone-ER, SSRI comparator and placebo groups on sexual dysfunction provides strong evidence that gepirone does not cause sexual dysfunction... (Label, p. 26)

166. A true demonstration of "no difference" would require a "non-inferiority" approach, and possibly a larger sample size.

The gepirone-ER studies were not designed for a non-inferiority analysis.

167. In this case, a study intended to demonstrate "no difference" would also need to include a group randomized to a drug that is known to cause sexual dysfunction (i.e., an active control) in order to establish assay sensitivity, the ability to detect sexual dysfunction.

Gepirone-ER studies did employ antidepressant arms with drugs that did cause sexual dysfunction. Gepirone-ER had statistically significantly better sexual function than fluoxetine or paroxetine in 5 of 6 studies. (Label, p. 26)

168. The larger sample size might have been accomplished by a prior hypothesis-testing strategy involving pooling data for the 2 phase 3 trials.

Although no a priori hypothesis testing strategy involving pooling of data was established, pooling of gepirone-ER data results in increased evidence of superiority of gepirone-ER over placebo. (JSM 2011 8:1411; JSM 2011 e-published ahead of print)

169 All of this ideally would have involved prior discussions with the Division of Psychiatry Products on how best to plan the studies to accomplish this important goal.

Specific discussions with the FDA have not addressed this directly.

CLINICAL SUMMARY

170. Vilazodone was recently approved by the FDA for the treatment of MDD.

Gepirone-ER was given a non-approval letter. (FDA not approvable letter, November 2, 2007)

171. Vilazodone is available in 10-, 20- and 40-mg immediate release tablets. The recommended dose is 40 mg per day, and vilazodone needs to be taken with food to ensure adequate plasma concentrations.

Gepirone-ER is available in 20, 40, 60 and 80 mg extended release tablets. (Label, p. 27)

172. The dose is incrementally adjusted to 40 mg, i.e., 10 mg once daily for 7 days, then 20 mg once daily for the next 7 days, and finally 40 mg once daily, to allow for adaptation to gastrointestinal symptoms.

The dose of gepirone is incrementally adjusted to 60-80 mg/day: starting with 20 mg/day, increasing to 40 mg/day at day 4, 60 mg/day after one week, and, if necessary, 80 mg/day after an additional week. (Label, p. 4)

173. Although there are as yet no randomized controlled trials that demonstrate longer term efficacy with vilazodone, it is standard practice with antidepressants to continue treatment of patients who have improved during short-term treatment.

Although disallowed, the per protocol analysis of study 28709 provides evidence that a placebo substitution study conducted correctly would be positive... (ISE p112).

174. Vilazodone's profile of adverse events is similar to that seen with other SSRIs, and it is unknown whether it has any advantages compared to other drugs in the antidepressant class.

Gepirone-ER is not an SSRI and its side effects differ from those of the SSRIs. Gepirone-ER is in a different class of drugs – the 5-HT_{1A} agonists. (Label, p. 18)

175 Consistent with the FDA's practice of including class effects in labeling, vilazodone has warning language in its label for serious adverse events observed with other antidepressants, including

suicidal ideation and behavior, serotonin syndrome, abnormal bleeding, activation of mania, and a contraindication for use with MAOIs.

Consistent with labeling for other antidepressants, gepirone-ER has warnings and precautions for: suicide risk, bipolar disorder, activation of mania/hypomania, use with MAOIs, seizures, and possible interference with cognitive and motor performance. (Label, p.7-8)

There has been no indication that gepirone-ER causes bleeding and no warning has been suggested. (Am J Med. 2006. 119:719-27).

176. It takes about 3 days for vilazodone to reach steady-state concentrations, and when stopping therapy, the dose should be tapered gradually to avoid discontinuation symptoms.

Steady state plasma levels of gepirone are reached in 2-4 days. There is no need to taper gepirone-ER since there are no discontinuation symptoms. (Label p. 13, 16)

177. Pharmacokinetic findings suggest that dose adjustment is not needed based on age, gender, or renal or hepatic impairment.

Pharmacokinetic findings with gepirone-ER suggest that no dose adjustment is needed based on age, gender, renal, or hepatic impairment, although caution should be used when treating patients with renal and hepatic impairments. (Label, p. 4)

178. Other factors, however, may, of course, lead to more cautious dosing in certain patients.

Severe renal or hepatic impairment patients and the extreme elderly should received gepirone-ER dosing with caution. (Label, p.4, 5)

179. It is recommended that the vilazodone dose be reduced to 20 mg when it is taken with strong CYP3A4 inhibitors, e.g., ketoconazole.

Concomitant use of gepirone-ER with potent inhibitors of CYP450 3A4 such as ketoconazole should be avoided. Concomitant use of gepirone-ER with drugs than induce P450-3A4 such as rifampicin reduces oral availability. (Label, p. 2)

180. Vilazodone is not expected to have important effects on the clearance of other drugs that are CYP450 substrates.

Gepirone-ER has been shown not to be an inhibitor or inducer of any of the cytochrome P450 enzymes, and is not expected to have important effects on the clearance of other drugs. (Label, p. 13)

OPEN ACCESS Freely available online

PLOS CLINICAL TRIALS

Essay

When and How Can Endpoints Be Changed after Initiation of a Randomized Clinical Trial?

Scott Evans

Introduction

Endpoints are outcome measures used to address the objectives of a clinical trial. The primary endpoint is the most important outcome and is used to assess the primary objective of a trial (e.g., the variable used to compare the effect difference of two treatment groups). A fundamental principle in the design of randomized trials involves setting out in advance the endpoints that will be assessed in the trial [1], as failure to prespecify endpoints can introduce bias into a trial and creates opportunities for manipulation. However, sometimes new information may come to light that could merit changes to endpoints during the course of a trial. This new information might include, for example, results from other trials or identification of better biomarkers or surrogate outcome measures. Such changes can allow incorporation of up-to-date knowledge into the trial design. However, changes to endpoints can also compromise the scientific integrity of a trial. Here I discuss some of the issues and decision-making processes that should be considered when evaluating whether to make changes to endpoints, and discuss the documentation and reporting of clinical trials that have revised endpoints.

Changing Endpoints

Many trials have changed their study endpoints after trial initiation. For example, a recent study [2] concluded that pioglitazone was associated with a significantly longer period in which patients remained free from death, myocardial infarction, or stroke, which was a composite endpoint. Conclusions from this report were questioned [3] because it was believed that the composite endpoint was not prespecified. This belief was based upon a previous publication [4] which listed the trial endpoints, but did not identify this specific composite. Authors of the original article responded to this criticism [5], stating that after initiation of the trial, the study executive committee recognized that the composite endpoint was not part of the original statistical analysis plan and thus the composite was subsequently added. The trialists also noted that the composite endpoint was documented in a revised analysis plan before unblinding the trial data.

More generally, Chan et al. [6] compared published articles with protocols for 102 randomized trials approved by the Scientific-Ethical Committees for Copenhagen and Frederiksberg, Denmark in 1994–1995, and reported that 62% of the trials had at least one primary endpoint that had been changed, introduced, or omitted. Chan et al. [7] compared published articles with protocols for 48 randomized trials approved for funding by the Canadian Institutes of Health Research in 1990–1998, and reported that primary endpoints differed between protocols and publications in 40% of the trials. Given that changes to endpoints are so frequent, it's important to evaluate when such changes are appropriate and how they should be reported.

Guiding Principles

The principle consideration when evaluating whether to modify an endpoint is whether the decision is independent of the data obtained from the trial to date. If the decision to revise endpoints is independent of the data from the trial, then such revisions may have merit. In fact, Wittes [8] encourages consideration of changes in long-term trials, as medical knowledge evolves or when assumptions made in design of the trial appear questionable. Wittes further argues that researchers "may consider changes to the primary endpoint when the trial has airtight procedures to guarantee separation of the people involved in making such changes from data that could provide insight into treatment effect" [8].

Some trials have successfully changed endpoints after the trials began by maintaining independence between the decision and the trial data. For example, the randomized Post-CABG (Post Coronary Artery Bypass Graft) trial [9] compared two lipid-lowering regimens in patients who had coronary artery bypass surgery. The investigators explicitly did not identify a primary endpoint when they designed the trial. An angiogram to assess lipid deposition in the coronary arteries was conducted at entry and then again five years later. The researchers planned to compare changes in lipid deposition over the five-year interval between the two regimens. Because by design no endpoint would be available for five years after randomizing the first participant, the protocol team used this period to define the endpoint and to develop methods for analyses. Although the endpoint was not prespecified in the design phase, a practice that is not generally recommended, trial leadership ensured that the selection of the endpoint was independent of data from the trial.

The Essay section contains opinion pieces on topics of broad interest to a general medical audience.

Funding: The author received no funding for this article.

Competing Interests: The author is a member of the *PLoS Clinical Trials* Editorial Board and has interacted with *PLoS Clinical Trials* personnel in his capacity as an Editor.

Citation: Evans S (2007) When and how can endpoints be changed after initiation of a randomized clinical trial? PLoS Clin Trials 2(4): e18. doi:10.1371/journal.pctr.0020018

Abbreviations: DMC, data monitoring committee

Scott Evans is with the Center for Biostatistics in AIDS Research, Department of Biostatistics, at the Harvard School of Public Health, Harvard University, Boston, Massachusetts, United States of America. E-mail: evans@sdac.harvard.edu

Box 1. Proposal for Handling Changes in Endpoints in Clinical Trials

Questions to Ask:

- What is the source of the new information that triggers consideration of a change in endpoints?
- Have interim data on the endpoint (or related data) been reviewed?
- Who is making the decision to change endpoints? Are trial sponsors involved, or is there an independent external advisory committee?

Documenting the Endpoint Changes:

- Update the protocol in a formal protocol amendment.
- Update the clinical trial registry record.
- Revise the statistical analysis plan.

Reporting the Trial Results:

- Include a clear statement describing the changes in endpoints, and the information obtained after the start of the trial that led to these changes.
- Include a description of the reasons for these changes and the decision-making procedure.
- Consider the potential biases that may have come about as a result of the change in endpoints.
- Consider including a disclaimer that the results should be interpreted with caution and may need to be confirmed in future trials.
- Report the reasons for excluding endpoints from the analyses and whether this was independent of trial data.

If, however, the decision to change the endpoint is not independent of the trial data, then "cherry-picking" is a serious concern. New endpoints may be selected because they displayed a trend towards significance, while other candidate endpoints may have been examined but not selected or reported because they failed to display a desirable trend; this increases the chance of false positive (type 1) errors. In the Physicians' Health Study [10,11], the trial's data monitoring committee (DMC) recommended termination of the study because interim data seemed unlikely to show any benefit of aspirin with respect to the primary endpoint, total mortality. At the time this decision was made, there was evidence of benefit with respect to myocardial infarction. However, the United States Food and Drug Administration did not approve an indication for aspirin for the prevention of myocardial infarction, because this was not the prespecified primary endpoint.

When Is a Decision Independent of Data?

To evaluate whether a change in endpoint is independent of data from the trial, investigators and reviewers should ask three important questions. First, what is the source of the new information that elicits consideration of the change in endpoints? If the source is external to the trial in question, for example arising from results from another trial, then the revision of endpoints may be credible. Second, have interim data on the endpoint (or related data) from a trial been reviewed? If trial data have not been reviewed, then the revision of endpoints may again be credible. Third, and most importantly, who is making the decision regarding endpoint revision (e.g., trial sponsors or an independent external advisory committee)? Appropriate decision makers should have no knowledge of the endpoint (or related trial data) results. In particular, if interim analyses have been conducted, the decision makers should not have knowledge of those data. Note, however, that even if no formal interim analysis has been conducted, any impressions that the investigators may have of the trial to date may influence decisions regarding changes in endpoints. For example, investigators may have a "sense" of the endpoint result or a related variable even though formal analysis of the endpoint has not been conducted. An investigator may notice changes in certain patients at his or her site and may attribute these changes to the investigational medication. This can be particularly problematic in unblinded trials. For these reasons, study sponsors, investigators, and DMCs may not be appropriate decision makers for endpoint revisions.

Appropriate Decision Makers

Since the decision to revise endpoints should be independent of the trial data, a DMC that has reviewed interim data may not be appropriate for making decisions regarding endpoint revisions. Even DMC review of pooled data can suggest treatment effects (e.g., in a two-group comparison study of response rates, a very high pooled response implies a relatively high response rate in both groups). In this case, trial leadership may wish to convene an external advisory committee that has not reviewed data from the trial to assess the potential impact on the integrity of the trial and to make recommendations regarding endpoint revision.

Scientific Relevance

It is also important to consider the scientific relevance of the endpoints in question. Does the current state of knowledge make the results of the current trial uninformative or inefficient? Is the trial now scientifically uninteresting or irrelevant? If so, then changing endpoints may be constructive, and perhaps even ethically necessary, to ensure that the study generates a scientific contribution. For example, new scientific questions may arise after recently completed trials have already answered the original question of interest. Also, better biomarkers or surrogates may have been identified, or there may have been changes in regulatory oversight.

One should be cautious of potential operational bias induced by the revision of endpoints. Operational bias is created when the conduct of clinical investigators or participants is changed by knowledge (or perceived knowledge) of trial data. Knowledge of revisions to endpoints may influence the actions of clinical investigators or participants as they anticipate the reasons for such revisions. For example, if a decision to change the primary endpoint is made, then participating clinicians and patients may believe that such a change was made due to a lack of efficacy of the intervention. This belief may affect their willingness to participate, affecting accrual and retention.

Documentation and Reporting

If the trial leadership decides to modify endpoints, then appropriate documentation is crucial. Changes should be described in amendments to the protocol and the analysis plan. The registry record for the trial should also be updated.

Changes in endpoints should also be declared when submitting a manuscript to a journal, so that the results can be properly evaluated. Reporting of a clinical trial with any modified endpoint should include: (1) a clear statement describing the fact that information obtained after trial initiation led to the change in endpoint; (2) a description of the reasons (e.g., whether the endpoint was suggested by the data) and decision procedure (e.g., who made the decision and whether data were unblinded); (3) a discussion of the potential biases induced

by the change of the endpoints; (4) if warranted, (i.e., if the decision to add endpoints was not independent of the data), a disclaimer that the results should be interpreted with caution and should be confirmed in future trials; and (5) a report of the reasons for excluding endpoints from the analyses and whether this was independent of trial data. Addressing these items will help ensure clarity and transparency of the analyses, enable the evaluation of the independence of the endpoint revision and trial data as well as the potential for selective reporting, allow assessment of the ramifications of the endpoint revision, and help avoid overinterpretation of the data. Researchers may further consider focusing on descriptive analyses using confidence intervals rather than hypothesis testing to avoid overstating the significance of the results.

Hawkey [12] suggests that journals require submission of the protocol alongside manuscripts describing clinical trial results, to help ensure that the reported endpoints indeed reflect what was defined at the start of the trial. Several journals have adopted this policy, including *PLoS Clinical Trials, The Lancet,* and the *British Medical Journal.* Other journals are considering a requirement to submit raw data (see the Harvard School of Public Health's Workshop on Assuring the Integrity of Reporting and Patient Safety in Therapeutic Trials; http://www.biostat.harvard.edu/events/schering-plough/agenda.html). Notably, for industry-sponsored studies, the *Journal of the American Medical Association* is requiring that analyses be conducted by an independent statistician at an academic institution, in part to protect against post hoc endpoint revisions.

Precise Definitions

Often, prespecified endpoints are defined vaguely or ambiguously. For example, a protocol designed to study the effects of 24 weeks of a new investigational drug on immune function might specify "CD4 count" as an endpoint. This endpoint could be interpreted in many different ways, including, for example: (1) CD4 count at week 24; (2) changes from baseline in CD4 count at week 24; (3) the occurrence of a doubling of CD4 count from baseline; or (4) the occurrence of at least a 50-cell increase in CD4 count from baseline. If a precise definition and analysis for each endpoint are not specified in advance, it is possible for many different versions of the endpoint to be examined, followed by selection and reporting of the most desirable result. This form of "cherry-picking" inflates the false positive error rate and leads to an underreporting of negative evidence. Thus it is critical to prespecify the precise definition of the primary endpoint together with the method of statistical analysis that will be applied [1].

An Alternative for Large Trials

In certain cases, it may be appropriate to change or identify endpoints after initiation of a trial, even when the decision is based on data from the trial. For example, if a trial is very large and of long duration, then investigators may divide the trial into two stages: a hypothesis-generating stage in which endpoints are identified, and a subsequent hypothesis-testing stage. In this case, statistical testing would be based only on data collected after the first stage was complete.

Conclusions

Revisions to endpoints (particularly primary endpoints) should be uncommon. If not appropriately evaluated, such revisions lead to misguided research and suboptimal patient care. If, however, important scientific knowledge has been gained after a trial begins, then this knowledge should be carefully and responsibly evaluated for incorporation into the trial. We should be open-minded and flexible in situations that may warrant the revision of endpoints and apply appropriate decision-making and reporting procedures when such situations arise. ■

REFERENCES

1. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (1998) ICH harmonized tripartite guideline: Statistical principles for clinical trials, E-9. Available: http://www.ich.org/LOB/media/MEDIA485.pdf. Accessed 6 March 2007.
2. Dormandy J, Charbonnel B, Eckland D, Erdmann E, Massi-Benedetti M, et al. (2006) Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive study (PROspective pioglitAzone Clinical Trial In macroVascular Events): A randomized controlled trial. Lancet 366: 1279–1289.
3. Skyler JS (2005) PROactive results overstated and misleading. DOC News 2: 4.
4. Charbonnel B, Dormandy J, Erdmann E, Massi-Benedetti M, Skene A (2004) The prospective pioglitazone clinical trial in macrovascular events (PROactive): Can pioglitazone reduce cardiovascular events in diabetes? Study design and baseline characteristics of 5238 patients. Diabetes Care 27: 1647–1653.
5. Dormandy J, Charbonnel B, Eckland D, Erdmann E, Massi-Benedetti M, et al. (2006) PROactive study. Lancet 367: 982.
6. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG (2004) Empirical evidence for selective reporting of outcomes in randomized trials. JAMA 291: 2457–2465.
7. Chan AW, Krleza-Jeric K, Schmid I, Altman DG (2004) Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. Can Med Assoc J 171: 735–740.
8. Wittes J (2002) On changing a long-term trial midstream. Stat Med 27: 2789–2795.
9. [No authors listed] (1997) The effect of aggressive lowering of low-density lipoprotein cholesterol levels and low dose anticoagulation on obstructive changes in saphenous-vein coronary artery bypass grafts. The Post Artery Bypass Graft Trial Investigators. N Engl J Med 336: 153–162.
10. Young F, Nightingale S, Temple R (1988) The preliminary report of the findings of the aspirin component of the ongoing Physicians' Health Study. The FDA perspective on aspirin for the primary prevention of myocardial infarction. JAMA 259: 3158–3160.
11. Cairns J, Cohen L, Colton T, DeMets D, Deykin D, et al. (1991) Issues in the early termination of the aspirin component of the Physicians' Health Study. Data Monitoring Board of the Physicians' Health Study. Ann Epidemiol 1: 395–405.
12. Hawkey CJ (2001) Journals should see original protocols for clinical trials. BMJ 323: 1309.

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

/s/

THOMAS P LAUGHREN
12/28/2011

5.10 Statistics Review –Yeh Fong Chen, Peiling Yang, Hsien Ming J Hung (10/25/2013)

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH
OFFICE OF TRANSLATIONAL SCIENCES
OFFICE OF BIOSTATISTICS

STATISTICAL REVIEW AND EVALUATION

Clinical Studies

NDA/Serial Number:	NDA21164 Amendment in Response to Information Request
Drug Name:	Gepirone ER
Indication:	Major Depression Disorder
Applicant:	Fabre-Kramer Pharmaceuticals, Inc.
Dates:	Date of Document: 12/7/2012
	PDUFA Due Date: None
Review Priority:	Not Available
Biometrics Division:	Division of Biometrics I, HFD-710
Statistical Reviewer:	Yeh-Fong Chen, Ph.D.
Concurring Reviewers:	Peiling Yang, Ph.D. & H.M. James Hung, Ph.D.
Medical Division:	Division of Psychiatry Products
Clinical Team:	Silvana Borges, M.D., HFD-130
	Ni Khin, M.D., HFD-130
Project Manager:	Hiren Patel, Pharm. D., HFD-130

Table of Contents

1. EXECUTIVE SUMMARY 3

2. INTRODUCTION 4

2.1 OVERVIEW 4
2.2 DATA SOURCES 4

3. STATISTICAL EVALUATION 5

3.1 EVALUATION OF EFFICACY 5
3.1.1 Overview of Trial Results 5
3.1.2 Summary of Sponsor's Reasons Why the Four Trials Should be Treated "Failed" 6
3.2 EVALUATION OF SAFETY 7

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS 7

5. SUMMARY AND CONCLUSIONS 7

5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE 7
5.2 CONCLUSIONS AND RECOMMENDATIONS 12

6. APPENDIX: DESCRIPTION OF SPONSOR'S REASONS WHY THE TRIALS FAILED 14

6.1 SPONSOR'S REASONS OF FAILURE FOR STUDY ORG134004 14
6.2 SPONSOR'S REASONS OF FAILURE FOR STUDY ORG134006 19
6.3 SPONSOR'S REASONS OF FAILURE FOR STUDY ORG134017 24
6.4 SPONSOR'S REASONS OF FAILURE FOR STUDY CN105-053 27
6.5 SPONSOR'S EVALUATION AND REANALYSIS FOR STUDY ORG28709 28

1. EXECUTIVE SUMMARY

To demonstrate the efficacy of geprione ER in treating MDD patients, the sponsor's program contained 12 short-term parallel-group trials and one long-term relapse-prevention trial with a randomized withdrawal design. FDA and the sponsor had reached the agreement that among the 12 short-term trials, 2 were positive, 3 were negative and 3 were failed trials. A decision is to be made as to whether the remaining 4 short-term trials are failed trials and whether the long-term trial would have been positive had the study been conducted correctly based on the sponsor's arguments and re-analyses.

After evaluation, in this reviewer's view, for the remaining four short-term studies (i.e., Studies ORG134004, ORG134006, CN134017 and CN105-053), most of the sponsor's reasons for failure were not sufficient to conclude that these trials are "failed" trials. First of all, it is difficult to point to the apparent treatment by site interactions based on the post hoc explorations as a reason for "failure"; in addition, an apparent treatment of site interaction was also suggested in the positive trial, FK-GBE-007.

The sponsor's argument that HAMD-25 total score is a better measurement than HAMD-17 total score for assessing efficacy in the MDD-AF (atypical features) patients in ORG134004 and ORG134006 is deferred to the medical division's determination. It is unclear, however, what "better" means. HAMD-25 total score had a larger standard deviation than HAMD-17 total score did. From the statistical power consideration, HAMD-25 may be harder to show a statistically significant treatment difference, unless the treatment effect is relatively larger based on HAMD-25.

Regarding the lack of assay sensitivity raised by the sponsor as a main reason why these trials should be considered "failed", historically, it has been a difficult issue as to how to determine whether a trial has assay sensitivity or not. The suggestions for assay sensitivity may be different, depending on the models for statistical analyses, such as including only two treatment arms under comparison or all three treatment arms, which endpoint to analyze, etc. To be more specific, for instance, if the pre-specified primary efficacy endpoint (HAMD-25 for ORG134004 and ORG134006, MADRS for ORG134017) is used and only two treatment arms (i.e., the active comparator and placebo) are included in the analysis models, there does not seem to be a support for assay sensitivity. Also noted is that the sponsor argues that use of the respective active comparators in ORG134004 and ORG134006 is inappropriate because their efficacy is unknown in the study patient population (atypical depression). Given this argument, it is unclear why these active comparators are included in the trials. There seems to be a conflict here. In all these three trials, if HAMD-17 (a commonly used primary endpoint) is analyzed and all three treatment arms are included (i.e., including the test treatment arm) in the model (a common practice), the results seem to suggest that there is assay sensitivity. On one hand, use of the pre-specified primary endpoint is arguably preferable; on the other hand, use of HAMD-17 may be reasonable for assay sensitivity assessment, unless there are gigantic differences between HAMD-17 and HAMD-25 (the two endpoints should be highly correlated).

Regarding the issue raised by the sponsor that the trends favoring placebo are exaggerated by the use of LOCF, for Study ORG134004, the MMRM analyses performed by this reviewer on

both HAMD17 and HAMD-25 indeed show smaller differences between gepirone and placebo. However, fluoxetine still performs numerically better than placebo, regardless of the model or the HAMD total score. Fluoxetine seems to be superior to gepirone.

Regarding the long-term relapse prevention trial, ORG28709, this reviewer found that the sponsor's re-analyses yielded statistically significant p-values, but the data of approximately 30 patients who came from centers with a single treatment arm or did not relapse was still removed and the five patients who indeed relapsed were not captured in re-analyses. Applying proper corrections, this reviewer's analyses do not show a statistically significant p-value to support the efficacy of gepirone in this trial.

2. INTRODUCTION

2.1 OVERVIEW

This NDA for gepirone ER (Org 33062 ER) as a treatment for major depression disorder (MDD) was originally submitted on September 30, 1999. After it was refused to file, the sponsor resubmitted it on May 18, 2001. FDA concluded that only one ER study was positive (i.e., ORG134001) and issued a non-approval (NA) letter on March 15, 2002 to convey that one additional positive ER study is required. On December 23, 2003, the sponsor resubmitted the NDA with data from a randomized withdrawal study. FDA determined that the results for the submitted randomized withdrawal trial results were problematic and issued a second NA letter on June 23, 2004, which stated that the sponsor was required to submit another positive short-term study and a positive randomized withdrawal study for gepirone ER. The sponsor then conducted two short-term studies (i.e., FKGBE007 and FKGBE008) and submitted the application with their analysis results on 5/1/2007. Although Study FKGBE007 was a positive study, FDA suspected that the observed effect size in the two positive studies seemed very small and that the positive results might be due to a play of chance, where only two studies were positive out of 12 short-term studies. Therefore, another non-approval action letter was issued on November 2, 2007.

On November 29, 2011, the sponsor (i.e., Fabre-Kramer Pharmaceuticals, Inc.) met with FDA for a follow-up meeting after the non-approval action. They presented arguments in support of their request for FDA to reconsider its non-approval decision for gepirone-ER as a treatment of MDD. FDA expressed willingness to reconsider the NDA and raised several questions. To respond to FDA's questions, the sponsor submitted an amendment dated February 3, 2012 with further arguments. After evaluation, FDA sent the sponsor a letter with a request of additional information which was included in this application for the re-consideration of its non-approval action.

2.2 DATA SOURCES

The sponsor's past submissions including data files and clinical study reports are stored in the following link: \\fdswa150\nonectd\N21164. The current submission is stored in \\cdsesub4\NONECTD\NDA021164\5196098.

3. STATISTICAL EVALUATION

3.1 EVALUATION OF EFFICACY

3.1.1 Overview of Trial Results

In this submission, the sponsor included 12 short-term placebo controlled trials and one long-term randomized withdrawal trial for evaluating Gepirone-ER as a treatment of MDD. Table 1 contains a brief summary from the sponsor about the primary efficacy findings of the 12 short-term trials. According to the sponsor, of the 12 short-term studies, the sponsor and FDA had been in an agreement on 8 studies: 1) FK-GBE-007 and ORG134001 are positive, 2) FK-GBE-008, ORG134002 and ORG134023 are negative, 3) CN105-052, CN105-078 and CN105-083 are failed studies. The sponsor argued that the remaining four studies (ORG134004, ORG134006, ORG134017 and CN105-053) are all failed studies for the reasons listed in Section.

Table 1. Summary of Sponsor's Short-Term Clinical Studies for Gepirone-ER

Study Number	Number gep. ER	Number Placebo	Active	Dose Range of gepirone ER (mean daily dose in mg)	LS mean diff HAMD-17 p-value (gep.-ER vs. Placebo)	Overall Results
ORG 134001	101	101	None	20-80 (61.1)	-2.47; p=0.013	Positive
FK-GBE-007	116	122	None	20-80 (58.2)	-2.45; p=0.018	Positive
FK-GBE-008	96	99	None	20-80 (60.0)	-1.38; p=0.20	Negative
ORG 134002	102	103	None	20-80 (57.9)	-0.71; p=0.42	Negative
ORG 134023	123	123	None	20-80 (61.3)	0.13; p=0.90	Negative
ORG 134004	124	130	Fluoxetine 20-40mg (34.1)	20-80 (67.1)	1.04; p=0.18	Pending
ORG 134006	140	143	Paroxetine 10-40mg (28.2)	20-80 (55.3)	0.22; p=0.76	Pending
ORG 134017	159	159	Fluoxetine 20-40mg (25.9)	40-80 (59.7)	0.65; p=0.39	Pending
CN105-053*	56	56	Imipramine 50-200mg (145)	10-60 (50.4)	-2.0; p=0.19	Pending
CN105-052*	35	37	Fluoxetine 20-80mg (23.3)	20-60 (43.4)	-0.69; p=0.74	Failed
CN105-078*	88	47	None	10-50 (30.4) 20-100 (52.6)	-1.0; p=0.36	Failed
CN105-083*	73	39	None	10-50 (30.4) 20-100 (57.1)	-0.49; p=0.75	Failed

Source: Sponsor's Table 1 of the submission dated February 3, 2012.

3.1.2 Summary of Sponsor's Reasons Why the Four Trials Should be Treated "Failed"

For Trial ORG134004, which was titled "A double-blind, multi-center, randomized, placebo-controlled, efficacy and safety study of Org 33062 ER and Fluoxetine in subjects who suffer from major depressive disorder with atypical features", the following is the list of the sponsor's reasons why this trial failed.

- No Assay Sensitivity
- HAMD-25 is the More Appropriate Measure of Efficacy in the MDD-AF Population
- Use of a Comparator with Unknown Efficacy in Atypical Depression
- Different Population: low Severity of Depression-Variable Severity Criterion
- High Placebo Response Rate
- Inappropriate Use of the Comparator
- Significant Treatment by Site Interaction
- Reasons for Trends in HAMD Favoring Placebo Over Gepirone-ER

For Trial ORG134006, which was titled "A double-blind, multi-center, randomized, placebo-controlled, parallel group study of efficacy and safety of Org 33062 ER and paroxetine in subjects who suffer from major depressive disorder with atypical features", the following is the list of the sponsor's reasons why this trial failed.

- No Assay Sensitivity
- HAMD-25 is the Appropriate Measure of Efficacy in the MDD-AF Population
- Use of a Comparator with Unknown Efficacy in Atypical Depression
- Different Population: Low Severity of Depression-Variable Severity Criterion
- Low Beck Depression Inventory Scores II
- High Placebo Response Rate
- Inappropriate Use of the Comparator
- Significant Treatment by Site Interaction

For Trial ORG134017, which was titled "A double-blind, multi-center, randomized, placebo-controlled, efficacy and safety trial of Org 33062 ER and fluoxetine in subjects with major depressive disorder", the following is the list of the sponsor's reasons why this trial failed.

- No Assay Sensitivity
- Inconsistency Among Sites
- High Placebo Response
- Positive Results from Reliable Investigators
- Flaws in Study Conduct
- Spurious Trends Favoring Placebo

For Trial CN105-053, which was titled "A double-blind, multicenter trial of Org 33062 ER, imipramine, and Placebo in the Treatment of Depressed Outpatients", the following is the list

of the sponsor's reasons why this trial failed. This study involved two sites; Feiger site was completed but Gelenberg site was terminated early.

- Early Termination
- By Protocol, the two sites should not be pooled
- Inappropriate FDA Analysis

For Trial ORG28709, which was titled "A multicenter, placebo-controlled study of relapse prevention during long-term treatment with Org33062 in outpatients with recurrent major depressive disorder", the sponsor listed factors indicating poor design and conduct of the trial:

- Investigator did not fully understand the protocol or the primary endpoint, as evidenced by a significant number of protocol violations.
- A high proportion of subjects received CNS drugs during the double-blind period, which can influence HAMD-17 ratings.
- Response criteria to qualify for randomization were not clearly defined and confirmed during the open-label period.
- Post hoc analyses restricted to qualified, protocol-compliant subjects show positive results for gepirone-ER.
- Post hoc analyses do not prove that this study shows efficacy for gepirone-ER. However, they do show had the study were done properly, the results would have been positive for gepirone-ER.

For all these four short term studies and the long term relapse prevention study, the sponsor's detailed reasons why these four trials should be treated as failed trials are provided the Appendix.

3.2 EVALUATION OF SAFETY

No relevant issues pertain to safety evaluation in this review.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

No relevant issues pertain to subgroup analysis in this review.

5. SUMMARY AND CONCLUSIONS

5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE

1. One of the sponsor's reasons why the four short-term trials (i.e., ORG134004, ORG134006, ORG134017 and CN105053) should be treated "failed" is the presence of treatment by site interaction or inconsistency among sites. According to this reviewer's observations, depending

on which endpoint (HAMD-17, HAMD-25, or MADRS) is analyzed and which model (either including only the gepirone-placebo comparison or including the active comparator) is used, the nominal p-value of the treatment by site interactions can be above or below 0.05. In Study CN105053, the early termination of one site may have added more uncertainty to the exploration of treatment-by-site interactions; that is, for whatever the reason is, gepirone performed similarly in both sites but placebo response was much larger in the early-terminated site. Even in the positive Study FK-GBE-007, the large variation in the gepirone effect (relative to placebo) among sites seems to render an impression of inconsistency among sites ($p = 0.092$ for treatment-by-site interactions). Thus, in this reviewer's view, it is difficult to point to the apparent treatment by site interactions as a reason for the trials to be treated "failed", as the exploration of such interactions is post hoc and can be controversial.

Table 2: Sponsor's HAMD-17 Total Score for Change from Baseline at End of Treatment by Center in Study FK-GBE-007

Center	Number of Subjects		Adjusted Mean Change		Treatment Difference		
	gepirone-ER	Placebo	gepirone-ER	Placebo	gepirone-ER – Placebo (95% CI)	SE	p-value
Center 701	21	23	-11.49	-5.34	-6.15 (-10.33, -1.97)	2.07	0.005
Center 702	14	16	-12.33	-8.02	-4.31 (-11.30, 2.67)	3.40	0.216
Center 704	22	20	-10.38	-11.63	1.25 (-2.63, 5.13)	1.92	0.519
Center 705	14	15	-5.98	-8.35	2.36 (-4.88, 9.61)	3.53	0.509
Center 706	10	9	-12.89	-5.34	-7.55 (-14.85, -0.24)	3.45	0.044
Center 708	12	10	-10.91	-10.51	-0.40 (-6.85, 6.05)	3.08	0.898
Center 709	8	11	-11.55	-4.24	-7.31 (-15.41, 0.79)	3.82	0.074
Center 999	15	18	-7.20	-7.94	0.74 (-5.28, 6.75)	2.95	0.804
All Centers	116	122	-10.24	-7.79	-2.45 (-4.47, -0.43)	1.02	0.018
Treatment by Center Interaction p-value = 0.092							
Analysis is based on ANCOVA with terms for treatment, center, and baseline value as a covariate; Center 999 is a pooled center that combines centers 703 and 707 [Source: ISE Table 16, Appendix A Statistical Table 3.2]							

Source: Sponsor's Table 5 and Table 9 of the file: summary-7dec2012.pdf in current submission

2. The sponsor argued that HAMD-25 total scores is a better measurement than HAMD-17 total scores (primary efficacy endpoint in the two positive studies, ORG134001 and FKGBE007) for assessing efficacy in the MDD-AF (atypical features) patients in ORG134004 and ORG134006. This is deferred to the medical division to determine. The following three tables might be helpful.

Table 4. Baseline HAMD-17 Total Scores (reported are mean and SD)

Study ORG134001		Study FKGBE007		Study ORG134004		Study ORG134006	
Org33062	22.7 (2.5)	Org33062	23.9 (2.7)	Org33062	19.6 (3.8)	Org33062	19.0 (3.5)
Placebo	22.8(2.5)	Placebo	24.2 (2.9)	Placebo	19.3 (3.8)	Placebo	18.8 (3.4)

Source: Values were extracted, for Study 134001 from Page 1079, for Study FKGBE007, from Page 3633, for Study 134004, from Page 79 and for Study 134006, from Page 91 of gepirone-nda-21164-summary-7dec2012.pdf.in the current submission.

Table 5. Baseline HAMD-25 Total Scores (reported are mean and SD)

Study ORG134001		Study FKGBE007		Study ORG134004		Study ORG134006	
Org33062	28.33 (3.88)	Org33062	29.5 (4.17)	Org33062	27.9 (4.9)	Org33062	27 (4.4)
Placebo	27.75(3.84)	Placebo	29.7 (4.30)	Placebo	27.6 (5.0)	Placebo	26.9 (4.3)

Source: Values were extracted, for Study 134001 from Page 1112, for Study FKGBE007, from Page 3812, for Study 134004, from Page 18272 and for Study 134006, from Page 23575 of gepirone-nda-21164-summary-7dec2012.pdf.in the current submission. …

Table 6. Change from Baseline to End Visit on HAMD-25 Total Scores by Sponsor's LOCF (reported are mean and SE)

Study ORG134001		Study FKGBE007		Study ORG134004		Study ORG134006	
Org33062	-11.57 (1.01)	Org33062	-12.65 (0.91)	Org33062	-9.76 (0.77)	Org33062	-10.94 (0.74)
Placebo	-8.19 (0.99)	Placebo	-9.85 (0.89)	Placebo	-10.63 (0.75)	Placebo	-11.0 (0.75)
Difference and P-value	-3.38 0.007	Difference and P-value	-2.80 0.029	Difference and P-value	0.87 0.416	Difference and P-value	0.06 0.953

Source: Values were extracted, for Study 134001, from Page 1114, for Study FKGBE007, from Page 3814, for Study 134004, from Page 18270 and for Study 134006, from Page 23577/87779 of gepirone-nda-21164-summary-7dec2012.pdf.in the current submission..

Based on Tables 4 and 5, the standard deviation of HAMD-25 total score seems larger than that of HAMD-17 total score; thus, from the statistical power consideration, HAMD-25 would tend to be harder to show a statistically significant difference unless the treatment effect is relatively larger based on HAMD-25. This does not seem to support the sponsor's argument that HAMD-25 is a better measurement for assessing efficacy in MDD patients with atypical features.

3. Lack of assay sensitivity is a common reason used in the sponsor's argument that ORG134004, ORIG134006 and ORG134017 should be "failed" trials. In ORG134004 and ORG134006, the sponsor's analyses using only the comparison between the active comparator (fluoxetine in -004 study and paroxetine in -006 study) and placebo on the pre-specified primary endpoint, HAMD-25, does not show that the respective active comparators are efficacious. Furthermore, the sponsor argues that use of the respective active comparators is inappropriate because their efficacy is unknown in the study patient population (atypical depression). Given this argument, it is unclear why these active comparators are included in the trials. There seems to be a conflict here. Likewise, in ORG134017, the sponsor's analysis using only the comparison between the active comparator (fluoxetine) and placebo on the pre-specified primary endpoint, MADRS, does not show the efficacy of the active comparator. In all three trials, FDA's analyses including three treatment arms in the models on HAMD-17 suggest that there may be assay sensitivity in that the respective active comparators beat gepirone. How to determine that a trial has assay sensitivity has been a difficult issue historically. On one hand, use of the pre-specified primary endpoint is arguably preferable. On the other hand, use of HAMD-17 may be reasonable for assay sensitivity assessment, unless there are gigantic differences between HAMD-17 and HAMD-25 (the two endpoints should be highly correlated).

4. For Study ORG134004, the sponsor noted that the observed trend in HAMD favoring placebo over gepirone-ER were affected not only by the placebo effect but also by the dropout rate. They further stated "*As a result, trends favoring placebo are exaggerated by the use of LOCF analyses, which carries forward final values for drop-outs. By contrast, numerical differences between gepirone-ER and placebo are negligible based on the Observed Case (OC) analysis, which does not impute values from prior visits for drop-outs". At Week 8, for example, the LS mean reduction from baseline in HAMD-25 (based on OC analysis) was 11.3 in the gepirone-ER group compared to 11.4 on placebo*."

Based on Table 7, the MMRM analyses performed by this reviewer on both HAMD17 and HAMD-25 indeed show smaller differences between gepirone and placebo. Fluoxetine still performs numerically better than placebo, regardless of the model or the HAMD total score. Fluoxetine seems to be superior to gepirone.

Table 7: FDA MMRM Analysis* Results for Trial ORG 134004

MMRM with Baseline covariate	HAMD 25 Total Scores	HAMD 17 Total Scores
Org33062 LS mean (SE)	-10.52 (0.83)	-6.24 (0.62)
Fluoxetine LS mean (SE)	-12.13 (0.77)	-7.90 (0.57)
Placebo LS mean (SE)	-11.04 (0.79)	-6.92 (0.59)
Org33062 minus Placebo LS mean (SE) & p-value	0.52 (1.14) 0.65	0.68 (0.85) 0.43
Fluoxetine minus Placebo LS mean (SE) & p-value	-1.09 (1.10) 0.32	-0.98 (0.82) 0.23
Org33062 minus Fluoxetine LS mean (SE) & p-value	1.61 (1.13) 0.15	1.65 (0.84) 0.05
MMRM w/o Baseline covariate	**HAMD 25 Total Scores**	**HAMD 17 Total Scores**
Org33062 LS mean (SE)	-10.45 (0.85)	-6.20 (0.64)
Fluoxetine LS mean (SE)	-12.18 (0.79)	-7.97 (0.59)
Placebo LS mean (SE)	-10.96 (0.81)	-6.82 (0.61)
Org33062 minus Placebo LS mean (p-value)	0.51 (1.17) 0.67	0.63 (0.88) 0.48
Fluoxetine minus Placebo LS mean (p-value)	-1.22 (1.12) 0.28	-1.14 (0.85) 0.18
Org33062 minus Fluoxetine LS mean (p-value)	1.73 (1.16) 0.14	1.77 (0.87) 0.04

* Besides the Baseline HAMD17 as the covariate or not, the analysis model included three treatment arms and center as factors.

5. For Relapse-Prevention Study ORG28709, the sponsor included their Table 62 in the re-submission to argue that had they had correctly identified the true drug responders before patient randomization, the study would have been positive. This reviewer noticed that in their analyses, (1) there were 5 patients who had relapsed but their relapse events were not counted, (2) the data of approximately 30 patients (depending on the type of analyses, the numbers varied a bit) from the centers which had only a single treatment arm or had no relapse were

removed from their CMH analyses with centers as strata, (3) in the sponsor's re-analyses, patients who were incorrectly identified as true drug responders were not removed from the analysis, but they were treated as non-relapsers. Including the aforementioned 5 patients' events and the data described in (2) above in the analysis by combining all the single armed or no-relapse centers, the reviewer's analyses yield very different results (Table 8). Noticeably, all the statistical reviewer's p-values are much larger than 0.05.

Table 8 Results of the Primary Analysis and Sponsor-Proposed Sensitivity Analyses

	Sponsor's Analysis Results			FDA Analysis Results		
	Gepirone-ER	Placebo	p-value	Gepirone-ER	Placebo	p-value
Original ITT	29/126 (0.23)	43/124 (0.35)	0.024	34/126 (0.27)	43/124(0.35)	0.36
Per Protocol	25/104 (0.24)	41/106 (0.39)	0.023	25/104 (0.24)	40/106(0.37)	0.11
Re-Defined Non-Responders(1)	22/126 (0.18)	40/124 (.32)	0.007	26/118 (0.22)	40/121(0.33)	0.17
Re-Defined Non-Responders(2)	22/126 (0.18)	42/124 (0.34)	0.003	26/118 (0.22)	42/123(0.34)	0.11
Re-Defined Non-Responders(3)	25/126 (0.20)	42/124 (0.34)	0.013	29/121 (0.24)	42/123(0.34)	0.25

(1) Excludes relapses on 1st visit after randomization, i.e., 11 patients were removed.
(2) Excludes relapses on 1st visit after randomization if response was confirmed prior to randomization, i.e., 9 patients were removed.
(3) Includes subjects with 50% drop in HAMD-17 prior to randomization as responders, i.e., 6 patients were removed from the analysis.

It is worth pointing out that in the sponsor's analysis for the primary endpoint, patients who discontinued from the double-blind continuation phase due to all reasons except 'Relapse Criteria Fulfilled' were treated "no relapse". As seen from Table 9, the proportion of patients who discontinued from the study due to either 'Unwilling or Uncooperative' or 'Reasons Not Mentioned' was much higher in the Gepirone-ER group than in the Placebo group (19.8% versus 11.2%). It remains unclear whether the sponsor's analyses by treating all dropouts as "no relapse" are seriously biased in favor of gepirone.

Table 9. Disposition of Subjects for Relapse-Prevention Study ORG28709

Screened				
435				
Open-Label Phase				
Number of Subjects		**Gepirone-ER**		
Selected to Participate		428		
Treated (Open-Label)		420		
Discontinued Prematurely	Total	117 (27.9%)		
	Adverse Events	46 (11.0%)		
	Insufficient Therapeutic Effect	26 (6.2%)		
	Other Reasons	19 (4.5%)		
Completed		303		
In Remission at end of OL Phase		250		
Double-Blind Continuation Phase				
Number of Subjects		**Gepirone-ER**	**Placebo**	**Overall**
Randomized		126	124	250
Treated with Double-Blind Medication		126	124	250
Discontinued Prematurely	Total	55 (43.7%)	54 (43.5%)	109 (43.6%)
	Adverse Events	4 (3.2%)	5 (4.0%)	9 (3.6%)
	Relapse Criteria Fulfilled	26 (20.6%)	35 (28.2%)	61 (24.4%)
	Unwilling or Uncooperative	11 (8.7%)	7 (5.6%)	18 (7.2%)
	Reason Not Mentioned	14 (11.1%)	7 (5.6%)	21 (8.4%)
Completed		71 (56.3%)	70 (56.5%)	141 (56.4%)

Source: Sponsor's Table 59 of gepirone-nda-21164-summary-7dec2012.pdf

5.2 CONCLUSIONS AND RECOMMENDATIONS

Short-term trials: ORG134004, ORG134006, ORG134017, and CN105-053

As articulated in [1] of Section 5.1, it is difficult to point to the apparent treatment by site interactions as a reason for these trials to be treated "failed", as the exploration of such interactions is post hoc and can be controversial.

The sponsor's argument that HAMD-25 total scores is a better measurement than HAMD-17 total scores for assessing efficacy in the MDD-AF (atypical features) patients in ORG134004 and ORG134006 is deferred to the medical division. It is unclear, however, what "better" means. This reviewer notes that HAMD-25 total score has a larger standard deviation than HAMD-17 does. From the statistical power consideration, HAMD-25 should be harder to show a statistically significant treatment difference, unless the treatment effect is relatively larger based on HAMD-25. ; See [3] of Section 5.1.

The lack of assay sensitivity issue is raised by the sponsor as a main reason why these trials should be considered "failed". Historically, it has been a difficult issue as to how to determine whether a trial has assay sensitivity or not. The suggestions for assay sensitivity may be different, depending on the models for statistical analyses, such as including only two treatment arms under comparison or all three treatment arms, which endpoint to analyze, etc. To be more specific, for instance, if the pre-specified primary efficacy endpoint (HAMD-25 for ORG134004 and ORG134006, MADRS for ORG134017) is used and only the two treatment arms (i.e., the active comparator and placebo) are included in the analysis models, there does not seem to be a support for assay sensitivity. Also noted is that the sponsor argues that use of the respective active comparators in ORG134004 and ORG134006 is inappropriate because their efficacy is unknown in the study patient population (atypical depression). Given this argument, it is unclear why these active comparators are included in the trials. There seems to be a conflict here. In all these three trials, if HAMD-17 (a commonly used primary endpoint) is analyzed and the three treatment arms (i.e., including the test treatment arm)are included in the model (a common practice), the results seem to suggest that there is assay sensitivity. As articulated in [3] of Section 5.1, on one hand, use of the pre-specified primary endpoint is arguably preferable; on the other hand, use of HAMD-17 may be reasonable for assay sensitivity assessment, unless there are gigantic differences between HAMD-17 and HAMD-25 (the two endpoints should be highly correlated).

Regarding the sponsor's raised issue that the trends favoring placebo are exaggerated by the use of LOCF, for Study ORG134004, the MMRM analyses performed by this reviewer on both HAMD17 and HAMD-25 indeed show smaller differences between gepirone and placebo. However, fluoxetine still performs numerically better than placebo, regardless of the model or the HAMD total score. Fluoxetine seems to be superior to gepirone. See [4] of Section 5.1.

Long-term relapse prevention trial: ORG28709

The sponsor's apparently statistically significant p-values in support of gepirone relative to placebo are based on their re-analyses that exclude the data of approximately 30 patients who

came from centers with a single treatment arm and do not correct the five patients who indeed relapsed. Applying proper corrections, this reviewer's analyses do not show a statistically significant p-value to support the efficacy of gepirone in this trial; see [5] of Section 5.1.

_____
Yeh-Fong Chen, Ph.D.
Mathematical Statistician

cc: NDA 21164
HFD-130/Dr. Mathis
HFD-130/Dr. Khin
HFD-130/Dr. Borges
HFD-130/Mr. Patel
HFD-700/Ms. Patrician
HFD-710/Dr. Mahjoob
HFD-710/Dr. Hung
HFD-710/Dr. Yang

6. Appendix: Description of Sponsor's Reasons why the Trials Failed

The following subsections include the details of sponsor's reasons why the four trials (ORG134004, ORG134006, ORG134017, CN105-053) failed and why they should not be used to discount the gepirone-ER's efficacy for treating MDD patients. They are mostly directly extracted from the sponsor's submission.

6.1 Sponsor's Reasons of Failure for Study ORG134004

6.1.1. No Assay Sensitivity

The protocol-specified analysis of the primary endpoint, HAMD-25, showed no statistically significant differences between either of the active treatments and placebo, or between fluoxetine and gepirone-ER. The trial lacks assay sensitivity -- the ability to distinguish between known effective and ineffective drugs (in this case, fluoxetine vs. placebo). Without this property, any differences between gepirone-ER and placebo, or between the active treatment arms, are uninterpretable. The study showed no assay sensitivity and is therefore a failed trial.

6.1.2. HAMD-25 is the More Appropriate Measure of Efficacy in the MDD-AF Population

Subjects were selected for this study with MDD-AF as established by the ADDS. Entry criteria required the presence of Atypical Features Specifier according to DSM-IV criteria. To fulfill these criteria, subjects had to maintain mood reactivity while depressed and have at least 2 of the following features: interpersonal rejection sensitivity throughout adulthood, weight gain/increased appetite, hypersomnia, and leaden paralysis; all as assessed using the ADDS.

It is inappropriate to evaluate efficacy in this trial with the HAMD-17, which has no items for assessing reverse neuro-vegetative symptoms. HAMD-25 (the total score for items 1-18 and 22-28 of the HAMD 31-item scale) was the primary efficacy variable chosen by the Sponsor and designated as such in the protocol. It is superior to the HAMD-17 as a measure of MDD-AF because it captures signs/symptoms that are an essential part of atypical depression and of the diagnostic scale (ADDS) used to enroll subjects into the trial. HAMD-25 items such as Q22 Hypersomnia (time in bed), Q23 Oversleeping, Q24 Napping, Q25 Increased Appetite, and Q26 Weight Gain, measure symptoms distinctive to MDD-AF that are not measured by the HAMD-17.

Table 43 shows the percent of change from baseline (CFB) in HAMD-25 total score that is accounted for by the items not contained in the HAMD-17 scale for MDD-AF subjects in study ORG134004 compared to MDD subjects in other studies of gepirone-ER.

Table 43: Percent of HAMD-25 CFB Not Counted in HAMD-17 CFB – Study ORG134004

Study	HAMD items 18 & 22-28[1]			HAMD items 22-26[2]			Study Classification
	Gep-ER	Active Control	Placebo	Gep-ER	Active Control	Placebo	
ORG134001	22	--	18	17	--	14	Positive
FK-GBE-007	19	--	19	9	--	10	Positive
ORG134004	**42**	**36**	**38**	**29**	**22**	**26**	Failed
ORG134017	15	15	15	6	6	6	Failed

[1] These are all the non-HAMD-17 items; percent calculated as [(HAMD-25 CFB – HAMD-17 CFB) / HAMD-25 CFB]*100
[2] These are the "hyper" non-HAMD-17 items; percent calculated as [(HAMD items 22-26 CFB) / (HAMD-25 CFB)]*100

42% of the change in HAMD-25 for the gepirone-ER group comes from the 8 items not contained in the HAMD-17 scale. In normally depressed populations, those 8 items only represent 15-22% of the change in HAMD-25 score. A similar range was seen in other gepirone-ER studies in MDD. By using the HAMD-17 score to measure response in MDD-AF subjects, we miss a substantial portion (42%) of the change in HAMD-25 scores, compared to only 14-25% for MDD patients across all gepirone-ER studies in MDD.

Analyzing only the 5 "hyper" items listed above, 29% of the change in the HAMD-25 total score for MDD-AF subjects in study 134004 comes from these 5 items, compared to 6-17% in normally depressed subjects. These differences are consistent across all other gepirone-ER MDD studies, including the supportive trials and other failed trials that were not fully enrolled.

These data indicate that HAMD-25 is more useful than HAMD-17 for measuring treatment efficacy in MDD-AF subjects of study ORG134004, because HAMD-17 captures a smaller portion of improvement in symptoms of MDD-AF than it does in MDD. Clearly the extra items in the HAMD-25 are important in the diagnosis and treatment of atypical depression and must be considered in assessing efficacy in this trial.

It is worth noting that HAMD-25 was also pre-specified as the basis for assessing gepirone-ER's effect on MDD-AF symptoms in studies ORG134001, ORG134002, FKGBE007 and FKGBE008. In each case, however, too few patients were enrolled with MDD-AF to permit such an analysis.

In summary, subjects in study 134004 were selected to have atypical depression by the ADD. The HAMD-25 scale was selected as the primary efficacy parameter because this scale includes items that measure the characteristic symptoms of atypical depression. To test treatment effects on scales that do not measure these symptoms would not fulfill the objectives of this study.

6.1.3. Use of a Comparator with Unknown Efficacy in Atypical Depression

Fluoxetine and other SSRIs have not been thoroughly or frequently studied in exclusively MDDAF populations. Use of a comparator with unknown efficacy in the target population limits the value of the study to judge the efficacy of gepirone-ER in that population.

6.1.4. Different Population: Low Severity of Depression – Variable Severity Criterion

Subjects entered into this trial had a low degree of depression. Entry scores on the Beck Depression Inventory II (BDI II) scale were ≤ 30 in over half of subjects (52% gepirone-ER, 51% fluoxetine, and 60% placebo). Low baseline scores limit the ability of the study to detect drug-placebo differences. Khan et al. (2002) indicate that studies with higher baseline HAMD scores are more likely to detect drug-placebo differences.

There was no minimum entry criterion for illness severity, resulting in low and highly variable baseline scores. Table 44 illustrates the difference in distribution of baseline HAMD-17 scores between this failed study and positive trials in MDD (ORG134001 and FK-GBE-007):

Table 44: Summary Statistics for Baseline HAMD-17 Total Score — Study ORG134004

Baseline HAMD-17	ORG134004		Positive Studies in MDD			
			ORG134001		FK-GBE-007	
	Gep-ER	Placebo	Gep-ER	Placebo	Gep-ER	Placebo
N	125	130	101	103	116	122
Mean	19.6	19.3	22.7	22.8	23.9	24.2
SD	3.8	3.8	2.5	2.5	2.7	2.9
Range	10-31	9-28	20-31	20-31	20-33	20-33

Note: All other gepirone-ER trials except ORG134004 and ORG134006 had minimum HAMD17 entry criterion of at least 20.

As shown, baseline mean HAMD-17 values were, on average, 3-5 points lower and more variable in this study compared to successful gepirone-ER studies. This difference is evident in comparison to all other gepirone-ER studies in MDD, for which mean baseline HAMD-17 scores ranged from 22.3 to 25.2 (Table 2).

Histograms in Appendix I-A further illustrate the wide dispersion in baseline HAMD-17 scores for each study in atypical depression (ORG134004 and ORG134006) relative to scores in the positive MDD studies. This variability in scores indicates a heterogeneous population and reduces the power of statistical tests to detect treatment effects.

Subjects selected for this study (and in ORG-134006) were drawn from a different population than all other gepirone-ER trials. MDD trial populations might include a small percentage of patients with atypical depression, but not 100% as in this trial. HAMD-17 scores were, on average, lower and more variable for these patients than in other MDD trials. In positive studies and all other studies of gepirone-ER (except ORG-134006), the entry criteria specified a HAMD- 17 total score of at least 20. Thus, the subjects selected for this study do not represent a normal MDD trial population.

6.1.5. High Placebo Response Rate

This study had a high placebo response rate: 36% on HAMD-25 and 42% on CGI (protocol defined responders). This further compromises the study's ability to detect drug-placebo differences. As noted by Khan et al. (2003), only 21% of antidepressant treatment arms used in trials with a high placebo response (> 30% mean change from baseline HAMD score)

showed statistical superiority over placebo. Calculated by Khan's method, the ORG134004 study had a placebo response rate of 39%, giving it low odds of success.

Placebo response rates in each of the 4 failed gepirone-ER studies under evaluation by FDA, based on the definition in Khan et al. (2003), are shown in Table 45.

Table 45: Placebo Response Rates in Failed Studies

Study	Placebo Response Rate*
ORG134004	38.8%
ORG134006	42.0%
ORG134017	45.1%
CN105-053**	44.8%

*Kahn definition: % mean change in HAMD total score from baseline to end of study in placebo group; calculations were based on HAMD-25 for studies ORG 134004 and ORG134006, and HAMD-17 for all others. Raw mean change (not LS means) were used in all cases.
**Gelenberg site only; Feiger's site (positive findings) had a 28% placebo response rate.

Given placebo response rates well above 30%, the trials were highly likely to fail. Out of 52 studies of approved antidepressants that Khan et al examined in 2003, only 7 had placebo response rates of 38.8% or higher. All 7 of these trials (which included venlafaxine, mirtazepine, nefazodone, bupropion, and citalopram) failed.

6.1.6. Inappropriate Use of the Comparator

FDA classified Study ORG134004 as "negative" because the comparator, fluoxetine, produced a greater response than gepirone-ER in a post-hoc analysis of a secondary endpoint. We disagree with this assessment based on scientific principles espoused by FDA itself:

a. The result is not based on the protocol-defined primary efficacy variable or pre-planned analysis, in violation of FDA's own "Guidance for Industry: E9 Statistical Principles for Clinical Trials" (Sept. 1998), which warns against any post-hoc change to the primary endpoints and their method of analysis. It also raises the potential for bias and false-positive findings from multiple comparisons.

b. ICH Guidance E10 emphasizes that in order to conduct a potentially useful and valid active control study, the comparator drug used must demonstrate Historical Evidence of Drug Effect, HESDE. Fluoxetine has not consistently shown effectiveness in similarly designed trials in atypical depression. Using a comparator that itself isn't consistently (i.e. more than 80% of the time) able to demonstrate superiority to placebo is inappropriate. See Page 10 and footnotes 25-30 of Fabre-Kramer's April 27, 2011 request for reconsideration.

c. It is inappropriate to judge assay sensitivity based solely on a comparison of the test vs. active control, and even less appropriate when the effect of the active control is uncertain. As noted by Dr. Temple in published statements, assay sensitivity is a property of a clinical trial defined by its "ability to distinguish effective from ineffective drugs". This property is demonstrated by validating the efficacy of the active control product (known to be effective) versus the placebo (known to be ineffective). Failing this, the active control's effect in

relation to the test product has no meaning (without a presumption that the test product is ineffective).

FDA has agreed that new drugs for MDD cannot be held to a comparative efficacy standard under current regulations and the Clinton-Gore Reinvention guidance.

Even if the test product were significantly better (or worse) than placebo in this 3-arm trial, the fact that the active control product is no different from placebo would invalidate the study as a basis for judging the efficacy of the test product.

6.1.7. Significant Treatment by Site Interaction

There is a significant treatment by site interaction for the HAMD-17 CFB analysis at endpoint, week 8 ($p = 0.050$), indicating that the gepirone-ER's effect on this variable was not consistent across sites. This was apparently not considered in FDA's re-analysis. The presence of interaction confounds interpretation of the treatment difference, so the finding should be examined in more detail. Further review of the HAMD-17 data shows the following:

a. ***Individual Site Results:*** Of the 10 sites, two favored gepirone-ER over fluoxetine (site 2 and site 4); and 8 favored fluoxetine over gepirone-ER (sites 1, 3, 5, 6, 7, 8, 9, and 10). This variation in the direction of treatment differences, and the fact that 2 of the sites showed opposite results, appear to account for the interaction.

b. ***Trimming the best and worst sites:*** If the best site for fluoxetine (site 5) and the worst site for fluoxetine (site 2) are removed, pooled results of the remaining 8 sites show that fluoxetine is no longer statistically significantly better than gepirone-ER on HAMD-17 CFB, $p=0.099$. This suggests that the significance of the finding is sensitive to site selection and is not particularly robust.

c. ***Analysis of data at a visit for which there was no interaction:*** An ANOVA of the HAMD-17 CFB data at Visit 5 (Week 6), when there was no treatment by site interaction ($p=0.223$), reveals that fluoxetine is not statistically significantly better than gepirone-ER, $p = 0.780$.

Thus, the FDA finding of a difference favoring fluoxetine over gepirone-ER for HAMD-17 CFB is not well-substantiated. The significant treatment by site interaction indicates a lack of consistency in the size and direction of treatment effects among study sites. No such interaction was evident for HAMD-25 ($p = 0.554$ for the fluoxetine vs. placebo comparison, and $p=0.323$ for the gepirone-ER vs. placebo comparison).

6.1.8. Reasons for Trends in HAMD Favoring Placebo over gepirone-ER

The FDA requested further explanation for why some efficacy scores for gepirone-ER were numerically worse than placebo in this study. The Sponsor believes that this is in large part due to the fact that there are at least 5 questions in the HAMD scale (Q4-Q8) that are biased against an effective drug when used in studies of atypical depression. These are items that

cover hypophagia and hyposomnia, symptoms not present to the same degree in atypical subjects as in "typical" depressed subjects. In ORG134004, gepirone-ER is +0.77 points relative to placebo on these 5 items of the HAMD, which accounts for about 90% of the amount by which gepirone-ER is numerically worse than placebo at EOT on the primary endpoint of HAMD-25, (+0.87).

This bias, combined with high variability of baseline scores within and between treatment groups and the high placebo response discussed above, likely accounts for gepirone-ER "leaning the wrong way" in this study. We note that this phenomenon also occurs in studies of other approved antidepressants: for example, in studies 244 and 245 of the vilazodone program, vilazodone was numerically worse than placebo on the HAMD-17 (a scale with a smaller range) by +0.8 and +0.5, respectively.

Numerical, non-significant trends should be interpreted cautiously in this trial. The observed differences are affected not only by the placebo effect, but also by the drop-out rate, which was higher in the gepirone-ER group (36%) than in the fluoxetine (18%) or placebo (21%) groups, mainly due to "other reasons" (22% vs. 12% and 17%) including lost to follow-up, noncompliance, and withdrawn consent. As a result, trends favoring placebo are exaggerated by the use of LOCF analysis, which carries forward final values for drop-outs. By contrast, numerical differences between gepirone-ER and placebo are negligible based on the Observed Case (OC) analysis, which does not impute values from prior visits for drop-outs. At Week 8, for example, the LS mean reduction from baseline in HAMD-25 (based on OC analysis) was 11.3 in the gepirone-ER group compared to 11.4 on placebo. In the Sponsor's view, it would be inappropriate to ascribe meaning to the size and direction of non-significant differences observed in this failed trial.

6.2 Sponsor's Reasons of Failure for Study ORG134006

6.2.1. No Assay Sensitivity

The protocol-specified analysis of the primary endpoint, HAMD-25, showed no statistically significant differences between either of the active treatments and placebo, or between paroxetine and gepirone-ER. The trial lacks assay sensitivity -- the ability to distinguish between known effective and ineffective drugs (in this case, paroxetine vs. placebo). Without this property, any differences between gepirone-ER and placebo, or between the active treatment arms, are uninterpretable. The study showed no assay sensitivity and is therefore a failed trial.

6.2.2. HAMD-25 is the Appropriate Measure of Efficacy in the MDD-AF Population

Subjects were selected for this study with MDD-AF as established by the ADDS. Entry criteria required the presence of Atypical Features Specifier according to DSM-IV criteria. To fulfill these criteria, subjects had to maintain mood reactivity while depressed and have at least 2 of the following features: interpersonal rejection sensitivity throughout adulthood, weight gain/increased appetite, hypersomnia, and leaden paralysis; all as assessed using the ADDS.

assessing reverse neuro-vegetative symptoms. HAMD-25 (the total score for items 1-18 and 22-28 of the HAMD 31-item scale) was the primary efficacy variable chosen by the Sponsor and designated as such in the protocol. It is superior to the HAMD-17 as a measure of MDD-AF because it captures signs/symptoms that are an essential part of atypical depression and of the diagnostic scale (ADDS) used to enroll subjects into the trial. HAMD-25 items such as Q22 Hypersomnia (time in bed), Q23 Oversleeping, Q24 Napping, Q25 Increased Appetite, and Q26 Weight Gain, measure symptoms distinctive to MDD-AF that are not measured by the HAMD-17.

Table 50 shows the percent of change from baseline (CFB) in HAMD-25 total score that is accounted for by the items not contained in the HAMD-17 scale for MDD-AF subjects in study ORG134006 compared to MDD subjects in other studies of gepirone-ER.

Table 50: Percent of HAMD-25 CFB Not Counted in HAMD-17 CFB — Study ORG134006

Study	HAMD items 18 & 22-28[1]			HAMD items 22-26[2]			Study Classification
	Gep-ER	Active Control	Placebo	Gep-ER	Active Control	Placebo	
ORG134001	22	--	18	17	--	14	Positive
FK-GBE-007	19	--	19	9	--	10	Positive
ORG134006	37	28	35	25	17	22	Failed
ORG134017	15	15	15	6	6	6	Failed

[1] These are all the non-HAMD-17 items; percent calculated as [(HAMD-25 CFB − HAMD-17 CFB) / HAMD-25 CFB]*100
[2] These are the "hyper" non-HAMD-17 items; percent calculated as [(HAMD items 22-26 CFB) / (HAMD-25 CFB)]*100

37% of the change in HAMD-25 for the gepirone-ER group comes from the 8 items not contained in the HAMD-17 scale. In normally depressed populations, those 8 items only represent 15-22% of the change in HAMD-25 score. A similar range was seen in other gepirone-ER studies in MDD. By using the HAMD-17 score to measure response in MDD-AF subjects of study ORG134006, we miss a substantial portion (37%) of the change in HAMD-25 scores, compared to only 14-25% for MDD patients across all gepirone-ER studies in MDD.

Analyzing only the 5 "hyper" items listed above, 25% of the change in the HAMD-25 total score for MDD-AF subjects in study 134006 comes from these 5 items, compared to 6-17% in normally depressed subjects. These differences are consistent across all other gepirone-ER MDD studies, including the supportive trials and other failed trials that were not fully enrolled.

These data indicate that HAMD-25 is more useful than HAMD-17 as an efficacy measure of treatment of MDD-AF subjects in study ORG134006, because HAMD-17 captures a smaller portion of improvement in symptoms of MDD-AF than it does in MDD. Clearly the extra items in the HAMD-25 are important in the diagnosis and treatment of atypical depression and must be considered in assessing efficacy in this trial.

It is worth noting that HAMD-25 was also pre-specified as the basis for assessing gepirone-ER's effect on MDD-AF symptoms in studies ORG134001, ORG134002, FKGBE007 and FKGBE008. In each case, however, too few patients were enrolled with MDD-AF to permit such an analysis.

In summary, subjects in study 134006 were selected to have atypical depression by the ADD. The HAMD-25 scale was selected as the primary efficacy parameter because this scale includes items that measure the characteristic symptoms of atypical depression. To test treatment effects on scales that do not measure these symptoms would not fulfill the objectives of this study.

6.2.3. Use of a Comparator with Unknown Efficacy in Atypical Depression

Paroxetine and other SSRIs have not been thoroughly or consistently studied in exclusively MDD-AF populations. Use of a comparator with unknown efficacy in the target population limits the value of the study to judge the efficacy of gepirone-ER in that population.

6.2.4. Different Population: Low Severity of Depression – Variable Severity Criterion

There was no minimum entry criterion for depression severity, so more than half of subjects (56%) had baseline HAMD-17 scores below 20, with 35% having scores below 18. Dr. Laughren reminded Fabre-Kramer that 18 is the minimum score to define MDD. Without a requirement for HAMD-17 scores of 20 or more, the study population did not include sufficiently depressed subjects. Antidepressant efficacy is difficult to detect in subjects with mild depression, as the potential effect size is limited.

Baseline severity of illness was low and also highly variable. Table 51 illustrates the difference in distribution of baseline HAMD-17 scores between this failed study and positive trials in MDD (ORG134001 and FK-GBE-007).

Table 51: Summary Statistics for Baseline HAMD-17 Total Score — Study ORG134006

Baseline HAMD-17	ORG134006		Positive Studies in MDD			
			ORG134001		FK-GBE-007	
	Gep-ER	Placebo	Gep-ER	Placebo	Gep-ER	Placebo
N	143	143	101	103	116	122
Mean	19.0	18.8	22.7	22.8	23.9	24.2
SD	3.5	3.4	2.5	2.5	2.7	2.9
Range	10-27	11-27	20-31	20-31	20-33	20-33

Note: All other gepirone-ER trials except ORG134004 and ORG134006 had minimum HAMD17 entry criterion of at least 20.

As shown, baseline mean HAMD-17 scores were, on average, 3-5 points lower and more variable in this study compared to successful gepirone-ER studies. This difference is evident in comparison to all other gepirone-ER studies in MDD, for which mean baseline HAMD-17 scores ranged from 22.3 to 25.2 (Table 2).

Histograms in Appendix I-A further illustrate the wide dispersion in baseline HAMD-17 scores for each study in atypical depression (ORG134004 and ORG134006) relative to scores in positive MDD studies. This variability in scores indicates a heterogeneous study population and reduces the power of statistical tests to detect treatment effects.

Baseline scores were also variable among the centers: For the 13 centers, paroxetine baseline scores ranged from 15.1 (centers 9 and 11) to 25.0 (center 1); gepirone-ER baseline scores ranged from 14.4 (center 11) to 21.0 (center 8); and placebo baseline scores ranged from 13.8 (center 11) to 22.3 (center 8).

Subjects selected for this study (and in ORG-134006) were drawn from a different population than all other gepirone-ER trials. MDD trial populations might include a small percentage of patients with atypical depression, but not 100% as in this trial. HAMD-17 scores were, on average, lower and more variable for these patients than in other MDD trials. In positive studies and all other studies of gepirone-ER (except ORG-134006), the entry criteria specified a HAMD-17 of at least 20. Thus, the subjects selected for this study do not represent a normal MDD trial population.

6.2.5. Low Beck Depression Inventory Scores II

To judge the impact of depression severity on efficacy results, we also examined BDI II scores at entry into the study: 16% of subjects had values below 20, and 48% below 30 points at entry. Thus, almost half of the subjects did not have enough depression to benefit from antidepressants. For subjects entered below 30, the percent responders (≥ 50% drop in BDI II) were: 47% placebo, 37% gepirone-ER, and 52% paroxetine. For those with scores above 30 at entry, response rates were: 37% placebo, 60% gepirone-ER, and 56% paroxetine. Thus, the subgroup with enough depression did in fact benefit from active drugs; drug-placebo differences in BDI response rates were statistically significant for both gepirone-ER ($p = 0.009$) and paroxetine ($p = 0.02$).

6.2.6. High Placebo Response Rate

This study had a high placebo response rate: 42% on HAMD-25 and 46% on CGI. This severely compromises the interpretation of this study. As noted by Khan et al (2003), only 21% of antidepressant treatment arms used in trials with a high placebo response (>30% mean change from baseline HAMD score) showed statistical superiority over placebo. Using this same response criterion, the ORG134006 study had a placebo response rate of 42%, giving it very low odds of success.

Placebo response rates in each of the 4 failed gepirone-ER studies still in contention with FDA, based on the definition in Khan et al. (2003), are shown in Table 52.

Table 52: Placebo Response Rates in Failed Studies

Study	Placebo Response Rate*
ORG134004	38.8%
ORG134006	42.0%
ORG134017	45.1%
CN105-053**	44.8%

*Kahn definition: % mean change in HAMD total score from baseline to end of study in placebo group; calculations were based on HAMD-25 for studies ORG 134004 and ORG134006, and HAMD-17 for the others. Raw mean change (not LS means) were used in all cases.
**Gelenberg site only; Feiger's site (positive findings) had a 28% placebo response rate

Placebo response rates exceeded 30% in all of these studies, greatly increasing the likelihood of failure. Out of 52 studies of approved antidepressants that Khan et al examined in 2003, only 7 had placebo response rates of 38.8% or higher. All 7 of these trials (which included venlafaxine, mirtazepine, nefazodone, bupropion, and citalopram) failed.

6.2.7. Inappropriate Use of the Comparator

FDA's reliance on a re-analysis of HAMD-17 data from this study violates ICH (E9 and E10) guidances, which warn against post-hoc analysis and use of an active comparator that does not have consistently established efficacy in a similar patient population. Specifically:

a. The result is not based on the protocol-defined primary efficacy variable or pre-planned analysis, in violation of FDA's own "Guidance for Industry: E9 Statistical Principles for Clinical Trials" (Sept. 1998), which warns against any post-hoc change to the primary endpoints and their method of analysis. It also raises the potential for bias and false-positive findings from multiple comparisons.

b. ICH Guidance E10 emphasizes that in order to conduct a potentially useful and valid active control study, the comparator drug used must demonstrate Historical Evidence of Drug Effect, HESDE. Paroxetine has not consistently shown effectiveness in similarly designed trials in atypical depression. Using a comparator that itself isn't consistently (i.e. more than 80% of the time) able to demonstrate superiority to placebo is inappropriate. See Page 10 and footnotes 25-30 of Fabre-Kramer's April 27, 2011 request for reconsideration.

c. It is inappropriate to judge assay sensitivity based on a comparison of the test drug vs. active control. As noted by Dr. Temple in published statements, assay sensitivity is a property of a clinical trial defined by its "ability to distinguish effective from ineffective drugs". This property is demonstrated by validating the efficacy of the active control product (known to be effective) versus the placebo (known to be ineffective). Failing this, the active control's effect in relation to the test product has no meaning (without a presumption that the test product is ineffective). FDA has agreed that new drugs for MDD cannot be held to a comparative efficacy standard under current regulations and the Clinton-Gore Reinvention guidance. Even if the test product were significantly better (or worse) than placebo in this 3-arm trial, the fact that the active control product is no different from placebo would invalidate the study as a basis for judging the efficacy of the test product.

6.2.8. Significant Treatment by Site Interaction

There is significant treatment by center interaction ($p = 0.024$) in the protocol-defined analysis of HAMD-25 CFB at endpoint, week 8/EOT, indicating that paroxetine's effect was not consistent for this variable. Though not tested statistically, a similar interaction is evident for HAMD-17, a problem overlooked by the FDA in their re-analysis of the data. We explore site-specific results for HAMD-17 below:

Analysis of HAMD-17 by Study Site

a. **paroxetine vs. placebo:** Based on HAMD-17 CFB at week 8/EOT, 9 sites (sites 2, 3, 5, 6, 7, 8, 11, 12, and 13) favored paroxetine over placebo; the differences were statistically significant in two sites (site 3, $p=0.0001$; site 6, $p=0.014$). The remaining 4

sites favored placebo over paroxetine (sites 1, 4, 9, and 10); the difference was statistically significant in one site (site 9, p=0.001). Thus, two positive sites 3 (n=60) and 6 (n=14) appear to drive the group of 13 sites to a significant outcome.

As shown in Appendix I-B, a Forest plot of HAMD-17 CFB values for paroxetine and placebo groups at 12 sites (sites 1 and 12 were pooled) illustrates inconsistent differences among sites, with one of the largest sites (site 3, n=60) driving the pooled finding. Removing this site, the p-value is no longer significant, p = 0.118.

b. **paroxetine vs. gepirone-ER:** Based on HAMD-17 CFB at week 8/EOT, 9 sites (sites 1, 3, 4, 5, 6, 7, 8, 11, and 13) favored paroxetine over gepirone; the differences were statistically significant for two sites (site 3, p=0.0001 and site 6, p=0.037). Three 3 sites (sites 2, 9, and 10) favored gepirone-ER over paroxetine, and one site (12) showed equivalence. Again, two sites, 3 and 6, drive the group of 13 sites to a significant outcome.

   As shown in Appendix I-B, a Forest plot of HAMD-17 CFB values for paroxetine and gepirone-ER groups at 12 sites (sites 1 and 12 were pooled) shows no consistent pattern in the differences, with site 3, again, driving the pooled finding. Removing this site, 8 sites favor paroxetine over gepirone, and 3 favor gepirone-ER over paroxetine; the pvalue is no longer significant, p = 0.081.

In summary, FDA claims that significant pairwise differences favor paroxetine over placebo and gepirone-ER for HAMD-17 CFB, but fails to take into account the treatment by center interaction. Further evaluation of the data shows differences among sites in the direction of paroxetine's effect. This qualitative interaction confounds interpretation of the pooled results and invalidates the treatment effect. Lacking assay sensitivity, the study is not a valid basis for evaluating the efficacy of gepirone-ER.

6.3 Sponsor's Reasons of Failure for Study ORG134017

6.3.1. No Assay Sensitivity

Based on the protocol-specified analysis of the primary endpoint, MADRS, no statistically significant pairwise differences were detected between gepirone-ER and placebo or between fluoxetine and placebo.

Fabre-Kramer believes that assay sensitivity should be demonstrated on the basis of the active control in relation to placebo. While directional trends favor fluoxetine over placebo at a few timepoints, these small differences were not sustained over the course of the treatment period and did not achieve statistical significance. Marginal effects of the active control on secondary variables are not sufficient to validate the assay sensitivity of the trial. If fluoxetine had demonstrated a clear and convincing advantage over placebo, the lack of significant finding for gepirone-ER could be interpreted as true lack of efficacy. Instead, the study provides no evidence that either drug performed significantly better than placebo. Lacking assay sensitivity, ORG134017 is a failed trial.

6.3.2. Inconsistency Among Sites

The primary efficacy parameter, MADRS, shows no consistent pattern in each of the 9 study sites: Gepirone-ER is favored over placebo in 5 sites (4, 7, 8, 11, and 12), whereas placebo is better than gepirone-ER in 4 sites (1, 5, 6, 9). A similar pattern is noted for the active comparator: fluoxetine is favored over placebo in 5 sites (1, 4, 8, 9, and 12), and placebo is better than fluoxetine in 4 sites (5, 6, 7, and 11).

This inconsistency in site results calls into question any reliance on non-significant trends in group means favoring one treatment or another.

6.3.3. High placebo response

This study had an extremely high placebo response rate: 53% of placebo-treated subjects were CGI responders (much or very much improved) at Week 8/Endpoint. At two of the study sites, placebo response rates were 90% and 100% based on CGI. The mean % change in MADRS score was 43% in the placebo group. Based on HAMD-17 responders (a 50% decrease from baseline score), only 4 of the 9 sites had placebo response rates less than 50%. This level of improvement on placebo severely compromises the interpretation of this study.

In MDD, studies with placebo response rates over 30% are unlikely to yield statistically significant results (Khan et al, 2003). As measured by Khan's method, the ORG134017 study had a placebo response rate of 45.1%, giving it very low odds of success. In fact, FDA has often used high placebo response as a valid justification for dismissing results of negative or failed trials; e.g. citalopram studies 89303 and 89306.

Placebo response rates in each of the 4 failed gepirone-ER studies still in contention with FDA, based on the definition in Khan et al. (2003), are shown in Table 56.

Table 56: Placebo Response Rates in Failed Studies

Study	Placebo Response Rate*
ORG134004	38.8%
ORG134006	42.0%
ORG134017	45.1%
CN105-053**	44.8%

*Khan definition: % mean change in HAMD total score from baseline to end of study in placebo group; calculations were based on HAMD-25 for studies ORG 134004 and ORG134006, and HAMD-17 for the others. Raw mean change (not LS means) were used in all cases.
**Gelenberg site only; Feiger's site (positive findings) had a 28% placebo response rate

Placebo response rates exceeded 30% in all of these studies, increasing their likelihood of failure. Out of 52 studies of approved antidepressants that Khan et al examined in 2003, only 7 had placebo response rates of 38.8% or higher. All 7 of these trials (which included venlafaxine, mirtazepine, nefazodone, bupropion, and citalopram) failed.

A study with a 53% placebo response rate on CGI, 45% using Khan's method, cannot produce data that can be considered valid.

6.3.4. Positive Results from Reliable Investigators

Given the extremely high placebo response, Fabre-Kramer examined data from two sites conducted by investigators (Arif Khan MD and Nick Vitakis MD) who they know and respect. Placebo response rates were only 36% and 12% at these sites, and mean differences in HAMD-17 scores showed favorable results for both active drugs: gepirone-ER vs. placebo -4.6 points, p=0.001; fluoxetine vs. placebo -4.8 points p=0.001.

6.3.5. Flaws in Study Conduct

This was a poorly conducted trial with major protocol compliance issues. The original criterion for entry was a HAMD-17 score of 18 or greater. After 9 months, it was changed to require more severe depression (HAMD-17 of at least 22). Consequently, the study included a substantial number of subjects without severe disease; baseline HAMD-17 scores were below 22 in 107 of 495 subjects enrolled (22%). Of 190 patients enrolled prior to the protocol amendment, 71 (37%) had HAMD-17 scores below 22; of 305 enrolled after the amendment, 36 (12%) had HAMD-17 scores below 22.

It is noteworthy that, after the HAMD-17 entry criterion was increased, subjects who qualified for the study showed more positive efficacy results than those who did not, as shown in Table 57 and Table 58.

Table 57: MADRS -- CFB for subjects with baseline HAMD-17 scores ≥ 22 who were enrolled after the entry criterion was raised (Study ORG134017)

Treatment	n	mean	SD	p-value	
				vs placebo	fluoxetine vs. gepirone-ER
gepirone-ER	93	-13.08	8.92	0.234	
placebo	88	-11.42	10.26		
fluoxetine	88	-13.40	9.99	0.182	0.826

For the 36 protocol violators, the results are much different and show both fluoxetine and placebo better than gepirone-ER:

Table 58: MADRS -- CFB for subjects with baseline HAMD-17 scores < 22 who were enrolled after the entry criterion was raised (Study ORG134017)

Treatment	n	mean	SD	p-value	
				vs placebo	fluoxetine vs. gepirone-ER
gepirone-ER	9	-5.22	10.05	0.037 (favors placebo)	
placebo	12	-13.83	7.14		
fluoxetine	15	-16.53	8.80	0.444	0.005 (favors fluoxetine)

As these findings illustrate, directionally positive results were produced from subjects appropriately entered into the study; illogical results were produced from subjects inappropriately entered.

6.3.6. Spurious Trends Favoring Placebo

The FDA requested further explanation for why gepirone-ER was numerically worse than placebo in this study. Trends in HAMD-17 favoring placebo over gepirone-ER reflect the marked placebo response that dominates results of this study. In the placebo group (with a baseline mean of 23.5), HAMD-17 scores dropped an average of 4.79 points (20%) at Week 1, 6.91 points (29%) at Week 2, and 8.30 points (35%) at Week 3. By Week 8/EOT, the placebo group had improved by 10.96 points, or 46.6%. Numerical differences in HAMD-17 actually favored placebo over fluoxetine on Weeks 1 and 3, but this random fluctuation is not informative. Statistically, the effect of fluoxetine on HAMD-17 was indistinguishable from placebo throughout the trial.

In Fabre-Kramer's view, the direction of non-significant differences in this failed trial should not be over-interpreted. The observed differences are affected not only by the placebo effect and the low severity of depression in subjects enrolled early in the trial, but also by the pattern of dropouts. More subjects dropped out of the gepirone-ER group (32%) than the fluoxetine (24%) or placebo (21%) groups, mainly due to "other reasons" (18% vs. 17% and 16%) including lost to follow-up, non-compliance, and withdrawn consent. These drop-outs, coupled with high placebo response, cause the LOCF results to favor placebo. The OC analysis, which only uses values for subjects who return for each study visit, showed trends favoring gepirone-ER over placebo. At Week 8, for example, LS mean reductions from baseline in MADRS (based on the OC analysis) were 14.7 in the gepirone-ER group compared to 13.8 on placebo; for HAMD-17, the LS mean reductions were 11.8 for gepirone-ER vs. 11.7 for placebo.

The high placebo response in the study, combined with the drop-out rate and the inconsistency among sites noted above, likely account for the non-significant result of gepirone-ER being 0.5 points worse than placebo at endpoint in this study. Further, among patients with adequate levels of depression enrolled after the protocol amendment, trends in HAMD-17 were comparable for gepirone-ER and fluoxetine, both favoring placebo.

6.4 Sponsor's Reasons of Failure for Study CN105-053

6.4.1. Early Termination – Sample Size Not Achieved

BMS terminated the program after 170 subjects were randomized (71% of the required number), reducing the power to approximately 63%. When the study was terminated, the Feiger site had exceeded enrollment (n=123), but only 47 of the 120 required subjects (39%) had been randomized at the Gelenberg site.

6.4.2. By Protocol, the two sites should not be pooled

Data from the two study sites show conflicting efficacy results and should not be pooled. The protocol did not mandate a pooled analysis of sites unless there is appropriate justification. The inconsistent efficacy results at the two sites may be due to the following factors:

a. There were major differences between study sites in demographic and baseline characteristics. Compared to the Gelenberg site, subjects in the Feiger site were younger (39 vs. 43 years), p=0.027; had fewer previous depressive episodes (60% vs. 81%), p=0.011; and had lower HAMD-17 baseline scores (23.7 vs. 25.1), p=0.006.

b. At the time of study termination, even the limited number of subjects at the Gelenberg site had not reached optimal doses of study drug: the mean modal dose for the Feiger site was 53.4 mg/day, compared to 42.2 mg/day for the Gelenberg site (p = 0.006). In positive studies of gepirone-ER, the mean modal doses were 70.3 mg/day (ORG134001) and 71.7 mg/day (FK-GBE-007). Thus, the Feiger site approached doses that are antidepressant; the Gelenberg site did not.

c. Study termination prevented an adequate number of subjects from being enrolled at the Gelenberg site. With only 15 or 16 subjects per group, the sample size at the Gelenberg site is too small to provide reliable estimates of treatment effects.

d. The placebo response rate at the Gelenberg site was high compared to the Feiger site, as judged by both the CGI-Improvement Scale (Gelenberg 56% vs. Feiger 33%) and the HAMD-17 responders (Gelenberg 50% vs. Feiger 33%). The calculation using the method in the Khan paper is 45% for Gelenberg and 28% for Feiger.

The typical multicenter study is centrally randomized and set up to combine sites unless there is a reason not to, such as significant treatment by site interaction. In this study, separate randomization schedules were used at each study site (page 29 of CSR), and there was no default rule to combine data from the two sites. The protocol stated: “Data for the two centers could be pooled if appropriate. Justification for pooling should be provided.” Given the differences between study sites in sample size, demographics, dose of study drug, and response to treatment, there is no justification for pooling.

6.4.3. Inappropriate FDA Analysis

The FDA re-analysis is inappropriate for a number of reasons:

- The analysis was done post-hoc, after unblinding;
- The two sites, only one of which completed enrollment, showed conflicting results; pooling was not warranted by the protocol.
- The ANCOVA analysis was not specified in the protocol.

6.5 Sponsor’s Evaluation and Reanalysis for Study ORG28709

Summary: Fabre-Kramer agrees antidepressant studies of this type are usually successful if the studies are designed and conducted properly. However, study ORG28709 was not conducted properly and its negative results are due to design flaws, poor protocol compliance, and careless mistakes by investigators, not lack of efficacy of gepirone-ER.

Critique of Clinical Trial Performance: For a placebo substitution study to be successful certain conditions must be met:

- the subjects entered into the open label portion must have a depressive illness,
- the subjects entered into the double-blind portion must be responders to medication, and
- the definition of relapse must be well understood and implemented by the investigators

Fabre-Kramer spoke with individual investigators after the start-up meeting in Brussels and found that most did not understand the protocol or the definition of relapse. Fabre-Kramer warned Organon that the study would not be successful, but Organon proceeded with their plans.

The study design was flawed in several respects:

1. The protocol did not prohibit use of concomitant benzodiazepines and other antidepressants that would confound the results. A total of 60 subjects took concomitant benzodiazepines or other CNS drugs, including other antidepressants, during the double blind period.

2. Numerous subjects randomized to the double-blind period were entered outside of the 8-12 week window or had other protocol violations. Eliminating these subjects in a 'per protocol' analysis, the difference in relapse rates is statistically significant favoring gepirone-ER. A full description of this analysis is found in the 2007 NDA. FDA rejected this approach as a post hoc analysis. However, this finding illustrates that, had the study been done correctly, it would have yielded positive results.

3. Investigators did not understand the primary endpoint (relapse criteria). This is evidenced by subjects who continued treatment for weeks to months after they met criteria for relapse, at which time they should have been discontinued.

4. The Sponsor believes that only true responders are appropriate for the study (i.e., response to gepirone-ER should have been confirmed on more than 1 visit prior to randomization). Unfortunately, subjects were randomized as soon as HAMD scores reached 8. Eleven subjects qualified for randomization (HAMD-17 score 8 or less) at one visit and relapsed (HAMD-17 score 16 or more) at the very next visit. While this could happen theoretically, it should be rare. Appendix I-C lists data for these subjects.

To eliminate non-responders, here we present another re-analysis under the following premise: *Subjects who relapse immediately after randomization are not true responders.* This analysis demonstrates that the difference in relapse rates among true responders is statistically significant favoring gepirone-ER.

The numbers of such occurrences make the results very suspicious: 11 subjects meet these criteria (8 gepirone-ER and 3 placebo subjects). All were relapses in the ITT analysis. If these non-responders are removed from the calculations, relapse rates are: gepirone-ER 22/126 (17.5%) and placebo 40/124 (32.3%), with $p = 0.007$.

The above subjects were classified as relapses on their first visit after randomization, but only 2 had values of 8 or less for more than 1 visit prior to randomization, both in the placebo group (257 and 636). Adding these subjects back into the analysis, the relapse rates are: gepirone-ER 22/126 (17.5%) and placebo 42/124 (33.9%), with p = 0.003.

The FDA advised us (November 29, 2011 meeting) that a HAMD-17 score of 8 or less is not necessary for randomization in a placebo substitution study, that a 50% drop in baseline HAMD-17 is adequate to define remission. If we include subjects with a 50% drop in HAMD-17 for more than one visit, the results include 5 more subjects: 3 gepirone-ER subjects (429,573,321) and 2 placebo subjects (636,257). The re-calculated relapse rates are: 25/126 gepirone-ER (19.8%) vs. 42/124 placebo (33.9%), with p = 0.013 (see February 3, 2012 Submission).

Table 62 summarizes results based on the various alternative definitions for remission.

Table 62: Relapse Rates in Study ORG28709 (Alternate Analysis)

Randomized	No. (%) Relapse		p-value*
	gepirone-ER N = 126	placebo N = 124	
Original ITT	29/126 (23.0%)	43/124 (34.7%)	**0.024**
Corrected ITT	34/126 (27.0%)	43/124 (34.7%)	0.101
Per Protocol	25/104 (24.0%)	41/106 (24.0%)	**0.023**
Excluding non-responders[1]	22/126 (17.5%)	40/124 (32.3%)	**0.007**
Excluding non-responders[2]	22/126 (17.5%	42/124 (33.9%)	**0.003**
Including 50% drop in HAMD[3]	25/126 (19.8%)	42/124 (33.9%)	**0.013**

*Chi-square test of proportions (two-sided); relapse=HAMD-17 ≥ 16
[1] Excludes relapses on 1st visit after randomization.
[2] Excludes relapses on 1st visit after randomization if response was confirmed prior to randomization
[3] Includes subjects with 50% drop in HAMD-17 prior to randomization as responders.

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

/s/

YEH FONG CHEN
10/25/2013

PEILING YANG
10/25/2013
Refer to my memo for additional information that I think is important but may not be captured in Dr. Chen's review.

HSIEN MING J HUNG
10/25/2013

5.11 Statistics Review –Peiling Yang and Hsien Ming J Hung (10/25/2013)

Memorandum

NDA #:	NDA 21-164 Amendment in Response to Information Request
Letter Date:	December 7, 2012
Applicant:	Fabre Kramer Pharmaceutical, Inc.
Drug Name:	Gepirone HCI Extended-Release Tablets
Indication:	Major Depression Disorder
Submission Location:	\\cdsesub4\NONECTD\NDA021164\5196098

The purpose of this memorandum is to supply certain information that I consider important and relevant but might not be completely captured in the primary statistical review by Dr. Chen, whose review mainly focused on certain aspects.

1 Background

Geperione ER (Extended Release) has not been approved as a treatment for major depressive disorder since its original NDA submission in September 1999. In FDA's most recent non-approval letter (November 2007), it indicated that the sponsor failed to provide substantial evidence of efficacy, where the evidence for short-term treatment was based on 12 trials and for long-term treatment it was based on one trial, all of which were included in the sponsor's relevant submission in May 2007.

It is noted that more than 12 short-term studies had actually been conducted over past many years and that these studies were managed by different sponsors. According to Organon's clinical data summary enclosed in 18 May 2001 and 23 December 2003 submissions for NDA21-164/N000, there were 18 "adequate and well controlled" studies with either ER or IR formulation[1]. Among those, Organon considered 8 ER studies "adequate and well controlled": ORG-134001 and ORG-134002 (both sponsored by Organon), and CN105-052, -053, -057, -064, -078, -083 (all these 6 sponsored by BMS).

In the May 2007 submission, Fabre-Kramer considered 15 out of 20 ER studies "adequate and well-controlled", but commented that two (CN105-064 and CN105-057) of these 15 studies investigated inadequate doses and not surprisingly failed. It further commented that the remaining 13 of the 15 studies (12 short-term trials and 1 long-term trial as diagramed in Table 1 below) were conducted at relevant doses and thus were the primary focus of efficacy evaluation of the ER formulation by Fabre-Kramer[2]. Indeed, Fabre-Kramer included these 12 short-term studies in their meta-analysis, where HAMD-

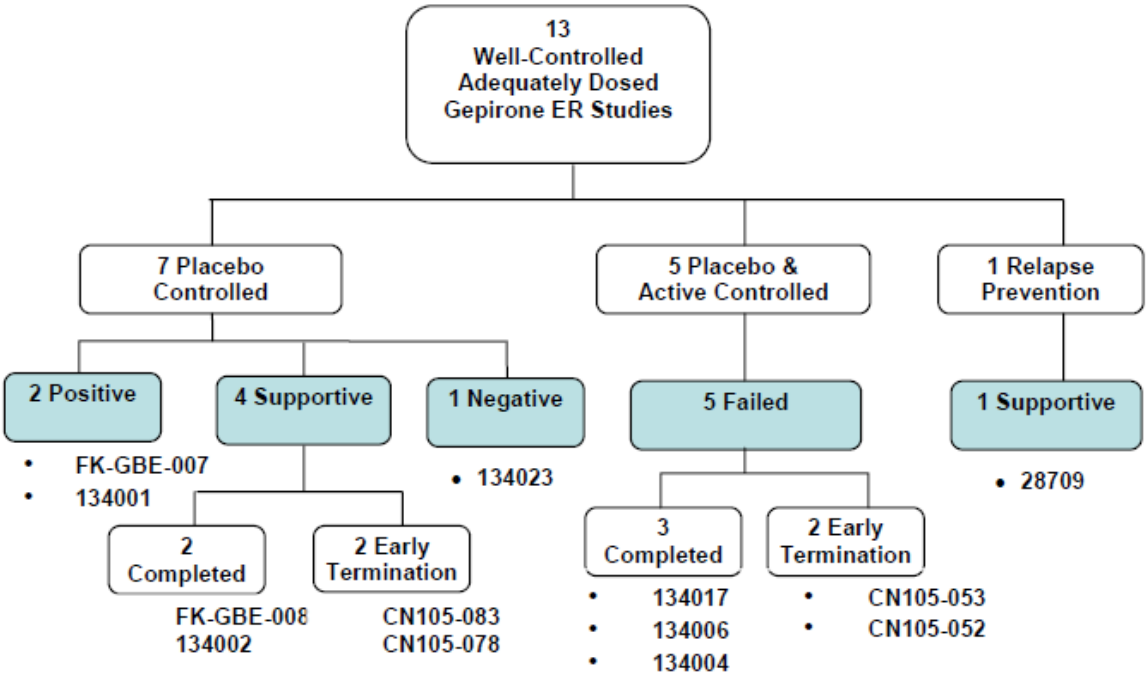
[1] Table 1 on Page 17 of ISE \\fdswa150\nonectd\N21164\N_000\2003-12-23\clinstat\ise\ise.pdf.
[2] Page 33 of \\fdswa150\nonectd\N21164\N_000\2007-05-01\Summary\summary-2007.pdf.

17 was used as the primary efficacy endpoint in order to integrate results from all 12 studies together.

It is unclear why all these 12 trials were included in the Fabre-Kramer's meta-analysis if some of the trials should have not been considered. Despite the apparently inconsistent views from the sponsor itself over past many years, it was clarified at the face-to-face meeting with Fabre-Kramer (November 29, 2011) that the main disagreement between FDA and Fabre-Kramer, in regard to the 12 short-term studies, involved the 4 remaining studies, i.e., ORG-134004, 134006, 134017, and CN105-053.

The next section highlights the key elements in each of these studies.

Figure 1: Fabre-Kramer's Diagram of Well-Controlled Gepirone ER Studies



[Source: Sponsor's "Figure 1 Well-Controlled Gepirone ER Studies" on page 34 of \\fdswa150\nonectd\N21164\N_000\2007-05-01\Summary\summary-2007.pdf]

2 Short-term Studies

A total of 12 short-term studies were included in the sponsor's meta-analysis in the 2007 submission. Similar to what the sponsor did, FDA's meta-analysis was based on the same efficacy endpoint (HAMD-17), the same statistical model and the same analysis across all 12 studies, regardless of the pre-specified primary endpoint and analysis by the sponsor. The same approach was used by FDA to evaluate assay sensitivity in individual

studies. This is where disagreement mainly lies in the remaining four studies (ORG-134004, 134006, 134017, and CN105-053).

2.1 Study CN105-053

Study Title: A double-blind, multicenter trial of Org 33062 ER, imipramine, and placebo in the treatment of depressed outpatients

Sponsor's Reasons for Study Failure:

- early termination;
- by protocol, the two sites should not be pooled;
- inappropriate FDA Analysis.

Brief Description: This was a 3-arm, flexible-dose study. According to Fabre-Kramer, this study was conducted at 2 US sites, but was terminated prematurely when BMS discontinued development of the ER formulation. Per the sponsor's submission, at the time of termination, one site (Feiger) had completed enrollment but the other site (Gelenberg) was incomplete (only 39% enrolled). When the study was terminated, a total of 170 patients had been randomized and 166 patients had post-baseline data: 120 patients at the Feiger site and 46 patients at the Gelenberg site. Of 46 patients at the Gelenberg site, only 57% completed the trial. The sponsor commented that patients at the Gellenberg's site barely reached the minimum dose, while patients at the Feiger site had a reasonable antidepressant dose. They also noted differential placebo responses observed between these two sites. See Appendix of this memorandum for sponsor's summary results.

My Comments:

[1] In my view, the major disagreement is on whether or not there was assay sensitivity in this study. The non-approval letter (issued in year 2007) indicated that this trial had assay sensitivity because the active comparator imipramine was statistically superior to placebo, but gepirone was not. There appear major differences between FDA's and sponsor's analyses, resulting in the disagreement:

- **FDA Analysis**: (a) all three arms were included in the model regardless of comparisons; (b) treatment-by-site interaction was not included although both main factors were; (c) primarily relied on ANCOVA.
- **Sponsor Analysis**: (a) only two arms (under comparison) at a time were included in the model; (b) treatment-by-site interaction was included; (c) relied on pre-specified ANOVA.

Per FDA analysis, imipramine was statistically superior to placebo whether based on ANCOVA (p-value = 0.038) or ANOVA (p-value = 0.043). FDA's analysis was based on a common practice. Regarding the treatment-by-site interaction, as a

general statistical principle, unless both the treatment and the center effects are statistically significant in the model without the interaction term, the interpretation of the interaction when added to the model can be very problematic, particularly if the trial was exclusively conducted in one region. For example, if the treatment works in US, it should work in all states, in which scenario the presence of the interaction could be due to chance. In addition, if the overall treatment effect is very small, it is very likely to observe positive treatment effect in some sites and negative treatment effect in other sites.

[2] I agree that early termination had impact on efficacy outcome at the Gelenberg site – at least, the variation from a small sample (approximately 15 completers per arm at this site) would tend to be larger and the smaller average modal dose at this site might have also impacted the efficacy outcome. However, per your own results (included in Appendix of this memorandum), although the mean modal dose for gepirone was lower at the early terminated site, it was also lower for imipramine at the same site, yet neither the observed gepirone responses (i.e., before subtracting the placebo response) nor the observed imipramine responses differed much numerically between these two sites. While the differential placebo responses may have added variation in efficacy outcome, it's difficult to ignore the efficacy outcome from the early terminated site even though you concluded superiority of both gepirone and imipramine to placebo at the Feiger site (where enrollment was completed).

[3] For a clarification purpose, although FDA's analysis is post-hoc, FDA used the same analysis across all 12 short-term trials for consistency. This was also sponsor's approach (individual study statistics obtained based on ANCOVA with the terms of treatment and center, and baseline score as a covariate) to their meta-analysis of 12 trials[3].

[4] **In summary**, whether this trial should be considered failed as requested by the sponsor is not purely a statistical call and is deferred to the medical division.

2.2 Study ORG134017

Study Title: A double-blind, multi-center, randomized, placebo-controlled, efficacy and safety trial of Org 33062 ER and fluoxetine in subjects with major depressive disorder

Sponsor's Reasons for Study Failure:

- no assay sensitivity;
- inconsistency among sites;
- high placebo response;
- positive results from reliable investigators;

[3] Section 5.1 in sponsor's "ise-2007.pdf" stored in \\fdswa150\nonectd\N21164\N_000\2007-05-01\clinstat\ise.

- flaws in study conduct;
- spurious trends favoring placebo.

Brief Description: This was also a 3-arm, flexible-dose study. This study was conducted at 9 US sites. The pre-specified primary endpoint was MADRS. According to the protocol, there was a plan to test the treatment-by-site interaction and the interaction was to be removed from the model if it was not statistically significant at the pre-specified level of 10%.

My Comments:

[1] In my view, the major disagreement is also on assay sensitivity. The same non-approval letter indicated that this trial had assay sensitivity because the active comparator fluoxetine was statistically superior to gepirone. There also appear to be major differences between FDA's and Sponsor's analyses, resulting in the disagreement:

- **FDA Analysis**: (a) all three arms were included in the model regardless of comparisons; (b) HAMD-17 (instead of the pre-specified primary endpoint MADRS) was used to conclude assay sensitivity; (c) primarily relied on ANCOVA.
- **Sponsor Analysis**: (a) only two arms (under comparison) at a time were included in the model; (b) MADRS (pre-specified primary endpoint) was used to draw conclusions; (c) relied on pre-specified ANOVA.

In FDA's analysis where HAMD-17 was used, fluoxetine was statistically superior to gepirone whether based on ANCOVA (p-value = 0.042) or ANOVA (p-value = 0.046). The treatment-by-site was not included in the sponsor's final model because the interaction was not statistically significant. Based on the sponsor's analysis on MADRS, fluoxetine was not statistically superior to gepirone.

[2] The sponsor considered treatments effect across sites inconsistent. Refer to my comment [1] in Section 2.1 on the treatment-by-site issue. The sponsor's final model did not include the treatment-by-site interaction because it was not considered statistically significant. There are also a few other conjectures, which the sponsor considered to have impacted the efficacy outcome, but in my view these do not appear to be of sufficient relevancy in justifying why this study should be considered failed.

[3] The sponsor considered this a failed study mainly because there was no statistically significant difference between placebo and any other treatment arm (gepirone or fluoxetine) based on their pre-specified primary endpoint (MADRS) on their pre-specified model ANOVA. They argued that assay sensitivity should be demonstrated on the basis of the active comparator to placebo.

[4] **In Summary,** whether this trial should be considered failed as requested by the sponsor is not purely a statistical call and is deferred to the medical division.

2.3 Studies ORG134004 and ORG134006

Study Titles:

- **ORG134004:** A double-blind, multi-center, randomized, placebo-controlled, efficacy and safety study of Org 33062 ER and Fluoxetine in subjects who suffer from major depressive disorder with atypical features
- **ORG134006:** A double-blind, multi-center, randomized, placebo-controlled, parallel group study of efficacy and safety of Org 33062 ER and paroxetine in subjects who suffer from major depressive disorder with atypical features

Sponsor's Reasons for Study Failure: .

ORG134004	ORG134006
– No Assay Sensitivity	– Same as left
– HAMD-25 is the More Appropriate Measure of Efficacy in the MDD-AF Population	– HAMD-25 is the Appropriate Measure of Efficacy in the MDD-AF Population
– Use of a Comparator with Unknown Efficacy in Atypical Depression	– Same as left
– Different Population: low Severity of Depression-Variable Severity Criterion	– Same as left
– High Placebo Response Rate	– Same as left
– Inappropriate Use of the Comparator	– Same as left
– Significant Treatment by Site Interaction*	– Same as left
– Reasons for Trends in HAMD Favoring Placebo Over Gepirone-ER	– Low Beck Depression Inventory Scores II

*referring to the HAMD-17, not the HAMD-25 endpoint.

Brief Description: These two studies were 3-arm, flexible-dose studies, but included respective active comparators. Both studies enrolled patients with Atypical Depression (i.e., MDD with Atypical Features or MDD-AF). Study ORG134004 was conducted at 10 US sites, and Study ORG134006 at 12 US sites and 1 Canada site. In both studies, the pre-specified primary endpoint was HAMD-25. In study ORG134004, the pre-specified statistical model was ANOVA with a plan to test the treatment-by-site interaction, but the sponsor's final model did not include the interaction because the p-values were between 0.3 and 0.4. In Study ORG134006, the pre-specified statistical model was ANCOVA without the treatment-by-site interaction.

My Comments:

[1] In my view, the major disagreement is also on assay sensitivity, which was greatly impacted by the efficacy endpoint that was analyzed. The same non-approval letter

indicated that both trials had assay sensitivity because the respective active comparators were statistically superior to gepirone based on HAMD-17. There appear to be major differences between FDA's and Sponsor's analyses, resulting in the disagreement:

- **FDA Analysis**: (a) all three arms were included in the model regardless of comparisons; (b) HAMD-17 (instead of the pre-specified primary endpoint HAMD-25) was used to conclude assay sensitivity.
- **Sponsor Analysis**: (a) only two arms (under comparison) at a time were included in the model; (b) HAMD-25 (pre-specified primary endpoint) was used to draw conclusions.

In FDA's analysis, the respective active comparators were statistically superior to gepirone based on HAMD-17 whether using ANOVA or ANCOVA when all 3 treatment arms were included. FDA used HAMD-17 for this assessment because it is a commonly used primary endpoint and it was the endpoint the sponsor used in meta-analysis of 12 trials. As I also pointed out in Section 2, to be consistent across all 12 studies, FDA used the same efficacy endpoint for this assessment.

[2] In the sponsor's view, both studies did not have assay sensitivity. A major reason is that neither active comparator was statistically superior to placebo based on their pre-specified primary analysis on the pre-specified endpoint (HAMD-25). They think that assay sensitivity should not be judged based on the comparison between two drugs, particularly because of the unknown efficacy in Atypical Depression of these active comparators. If the efficacy of these two active comparators is unknown on MDD-AF, it is not clear why the sponsor included inappropriate active comparators in these studies. These appear to be conflicting with each other.

[3] **In Summary,** there were more controversial clinical issues involved in these two studies. Whether one may conclude assay sensitivity based on the comparison of HAMD-17 between two drugs is deferred to the medical division.

3 Relapse-Prevention Study ORG28709

Study Title: A multicenter, placebo-controlled study of relapse prevention during long-term treatment with Org 33062 in outpatients with recurrent major depressive disorder

Sponsor's Reasons for Study Being Poorly Designed and Conducted:

- Investigator did not fully understand the protocol or the primary endpoint, as evidenced by a significant number of protocol violations.
- A high proportion of subjects received CNS drugs during the double-blind period, which can influence HAMD-17 ratings.

- Response criteria to qualify for randomization were not clearly defined and confirmed during the open-label period.
- Post hoc analyses restricted to qualified, protocol-compliant subjects show positive results for gepirone-ER.
- Post hoc analyses do not prove that this study shows efficacy for gepirone-ER. However, they do show had the study were done properly, the results would have been positive for gepirone-ER.

Brief Description: This was a randomized withdrawal study conducted in Europe. A total of 250 patients were qualified as responders during the 8-12 week open-label phase and thus were randomized to the 40-44 week double-blind phase. The sponsor's primary analysis was to compare relapse rates between gepirone and placebo using the CMH test, stratified by center. Analysis of time to first occurrence of relapse was considered as supportive analysis by the sponsor. Based on the sponsor's summary, a total of 109 patients (55 from the gepirone and 54 from the placebo arm) were discontinued from the double-blind phase. Of those 109 who discontinued, 61 (26 from the gepirone and 35 from the placebo arm) discontinued because relapse criteria were fulfilled. The sponsor's primary analysis based on stratified CMH test yielded statistically significant superiority of gepirone over placebo (p-value = 0.024), but this was not supported by the sponsor's supportive analysis of time to relapse based on logrank test (p-value = 0.065).

This study was included in the NDA Amendment submitted by Organon (a former sponsor) in December 2003. During the review process, the FDA statistician Ms. Roswitha Kelly noticed two major problems in the sponsor's analyses:

(a) Failure to count 5 patients from the gepirone arm who had relapses. It appears that these patients had relapses 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.

(b) Failure to include 32 patients who came from centers that had only 1 treatment arm represented or had no relapses. One way for these patients to become part of the analysis is by pooling all these patients together to form a pseudo center.

After the above two problems were corrected, Ms. Kelly found that the primary analysis result was no longer statistically significant (p-value = 0.101)[4].

My Comments:

[1] In the current submission, Fabre-Kramer considered that this study was not adequately designed or conducted, and that these inadequacies accounted for its failure. In specific, they argued that not all patients randomized to the double-blind phase were "true" responders. Hence, they re-analyzed data using different

[4] Refer to statistical review by Ms. Roswitha Kelly of NDA 21-164/N000 Amendment, submitted by Organon with the letter date December 23, 2003. Sponsor's submission is stored in \\fdswa150\nonectd\N21164\N_000\2003-12-23.

definitions of true responders. Although all of their re-analyses yielded statistically significant p-values in favor of gepirone (p-values < 0.05), Dr. Yeh-Fong Chen, the primary statistical reviewer of the current submission, disagrees to Fabre-Kramer's re-analyses results for the following reasons: (a) the same problems as identified by Ms. Kelly remained despite the acknowledgement of these problems by Fabre-Kramer in the current submission, (b) failure to remove all patients who should have been removed according to their various definitions of true responders. After these problems were corrected, the p-values were no longer statistically significant.

[2] The primary analysis for a randomized withdrawal study, such as this, has been typically based on time-to-relapse analysis. The result based on the time-to-relapse analysis appeared to be negative, consistent with the result from the CMH test when the aforementioned problems were corrected.

[3] Fabre-Kramer considered that their post-hoc analyses do show that the results would have been positive for gepirone had the study been done properly. These post-hoc analyses were performed on selected analysis sets, which were likely to violate the randomization principle, an important assumption for a valid statistical analysis. Hence, I do not consider their justification to be adequate.

[4] **In summary,** in my view, this study was negative regardless of re-analyses on various analysis sets explored by Fabre-Kramer. In a typical clinical trial whether the trial turns out to be positive or negative, there are often some flaws, such as non-compliance by patients or by investigators, patients who did not meet inclusion criteria are randomized, poorly trained investigators. Whether the severity (or the extent) of flaws is substantial for this trial to be discarded is deferred to the medical division.

Appendix – Extracted from Sponsor's Results of Study CN105-053

Table 35: Maximum Modal Dose by Treatment Group and Study Site — CN105-053 (ITT Population)

Site	Gepirone-ER 10-60 mg/day		Imipramine 50-200 mg/day	
	Mean ± SD	N	Mean ± SD	N
Site 0001 (Feiger)	53.4 ± 11.5	41	173 ± 48.5	39
Site 0002 (Gelenberg)	42.0 ± 16.6	15	157 ± 49.5	15
Combined Sites	50.5 ± 14.7	56	161 ± 50.2	54

[*Source: Appendix F Tables 6.10-3 ,6.10-4, and 6.10-5*]

A 40 mg/day dose of gepirone-ER is considered the minimum effective antidepressant dose, and the proposed label recommends a target dose of 60 to 80 mg/day. Gellenberg's subjects barely reached the minimum dose. Feiger's subjects had a reasonable antidepressant dose.

[Source: Table 35 of sponsor's current submission \\cdsesub4\NONECTD\NDA021164\5196098]

Table 36: HAMD-17 Total Score: Change from Baseline at Each Visit (ITT/LOCF) – Study CN105-053

Treatment		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
Feiger (Site 001)								
gepirone-ER	n	41	40	41	41	41	41	41
	Mean	23.7	-4.7	-5.5	-7.2	-8.2	-10.8	-10.1
	SE	0.52	0.7	0.9	1.0	1.0	1.1	1.2
imipramine	n	39	39	39	39	39	39	39
	Mean	23.6	-4.5	-7.0	-9.3	-9.4	-11.8	-10.9
	SE	0.48	0.7	0.9	1.0	1.0	1.2	1.2
placebo	n	40	38	40	40	40	40	40
	Mean	23.9	-3.9	-4.9	-6.8	-6.0	-7.7	-6.8
	SE	0.46	0.7	0.9	1.0	1.0	1.2	1.2
gepirone-ER vs pbo	p-value		0.455	0.649	0.746	0.129	0.063	0.049
imipramine vs. pbo	p-value		0.572	0.096	0.092	0.024	0.014	0.017
Gelenberg (Site 002)								
gepirone-ER	n	15	15	15	15	15	15	15
	Mean	24.6	-3.1	-7.3	-9.0	-9.1	-10.3	-9.3
	SE	0.62	1.3	1.5	1.7	2.0	2.5	2.5
imipramine	n	15	15	15	15	15	15	15
	Mean	25.7	-3.1	-7.5	-7.9	-10.9	-10.9	-12.2
	SE	0.54	1.3	1.5	1.7	2.0	2.5	2.5
placebo	n	16	15	16	16	16	16	15
	Mean	25.0	-2.8	-4.7	-6.4	-9.1	-10.2	-11.2
	SE	0.75	1.3	1.1	1.7	2.0	2.4	2.5
gepirone-ER vs. pbo	p-value		0.837	0.226	0.289	0.998	0.966	0.589
imipramine vs. pbo	p-value		0.837	0.182	0.534	0.527	0.843	0.776
Combined Sites								
gepirone-ER	n	56	55	56	56	56	56	56
	Mean	23.9	-3.9	-6.4	-8.1	-8.7	-10.5	-9.7
	SE	0.41	0.7	0.8	1.0	1.0	1.2	1.2
imipramine	n	54	54	54	54	54	54	54
	Mean	24.2	-3.8	-7.3	-8.6	-10.2	-11.3	-11.5
	SE	0.39	0.7	0.9	1.0	1.1	1.2	1.2
placebo	n	56	53	56	56	56	56	56
	Mean	24.2	-3.3	-4.8	-6.6	-7.6	-8.9	-9.0
	SE	0.39	0.7	0.8	1.0	1.0	1.2	1.2
gepirone-ER vs. pbo	p-value		0.568	0.195	0.274	0.448	0.343	0.687
imipramine vs. pbo	p-value		0.633	0.041	0.153	0.080	0.160	0.144

LS Means and p-values for combined sites is based on ANOVA with effects for treatment, center and treatment by center interaction.
[*Source: CSR CN105-053 Table 17, Appendix F Tables 7.1.1-3 and 7.1.1-6*]

[Source: Table 36 of sponsor's current submission \\cdsesub4\NONECTD\NDA021164\5196098]

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

/s/

PEILING YANG
10/23/2013

HSIEN MING J HUNG
10/25/2013

5.12 Clinical Review –Silvana Borges (04/18/2014)

CLINICAL REVIEW

Application Type	NDA
Application Number(s)	21164
Priority or Standard	Sponsor's informal appeal of non-approval decision
Submit Date(s)	December 7, 2012
Received Date(s)	December 10, 2012
PDUFA Goal Date	Not applicable
Division / Office	DPP/ODE1
Reviewer Name(s)	Silvana Borges, M.D
Review Completion Date	August 30, 2013
Established Name	Gepirone Hydrochloride
Trade Name	Travivo®
Therapeutic Class	Antidepressant
Applicant	Fabre Kramer Pharmaceuticals Inc.
Formulation(s)	20 mg, 40 mg, 60 mg and 80 mg extended-release tablets
Dosing Regimen	40 to 80 mg QD
Indication(s)	Treatment of major depressive disorder
Intended Population(s)	Adults

Table of Contents

1 RECOMMENDATIONS/RISK BENEFIT ASSESSMENT 3

1.1 Recommendation on Regulatory Action 3
1.2 Risk Benefit Assessment 3

2 INTRODUCTION AND REGULATORY BACKGROUND 5

2.1 Product Information 5
2.2 Currently Available Treatments for Proposed Indication 5
2.3 Summary of Presubmission Regulatory Activity Related to Submission 5

3 SIGNIFICANT EFFICACY/SAFETY ISSUES RELATED TO OTHER REVIEW DISCIPLINES 7

3.1 Chemistry Manufacturing and Controls 7

4 SOURCES OF CLINICAL DATA 7

4.1 Tables of Studies/Clinical Trials 7
4.2 Review Strategy 9

5 REVIEW OF EFFICACY 10

Efficacy Summary 10
5.1 Studies Pertinent to the Efficacy Claim of gepirone ER for the Treatment of MDD (study start years are included in brackets) 11
5.1.1 Positive studies 11
5.1.1.1 Study ORG134001 (1999) 11
5.1.1.2 Study FK-GBE-007 (2003) 11
5.1.2 Negative studies 13
5.1.2.1 Study ORG134002 (1999) 13
5.1.2.2 Study FK-GBE-008 (2003) 15
5.1.2.3 Study ORG134023 (2003) 16
5.1.2.4 Studies ORG134004 (2000) and ORG134006 (2000) 18
5.1.2.5 Study ORG134017 (2002) 27
5.1.2.6 Study CN105053 (1991) 30
5.1.2.7 Study ORG28709 (1999) 33
5.1.3 Failed studies 36
5.1.3.1 Study CN105-052 (1991) 36
5.1.3.2 Study CN105-078 (1991) 36
5.1.3.3 Study CN105-083 (1991) 37
5.2 Conclusions Regarding the Efficacy Claim of gepirone ER for the Treatment of MDD 38

1 Recommendations/Risk Benefit Assessment

1.1 Recommendation on Regulatory Action

It is recommended that a letter reiterating our non-approval action be issued in response to this informal appeal. In this reviewer's opinion, the sponsor has failed again to demonstrate the efficacy of gepirone ER for the treatment of adults with major depressive disorder.

1.2 Risk Benefit Assessment

Since the original NDA submission in 1999, the efficacy data for gepirone ER were reviewed by FDA clinical and statistical reviewers in three different occasions (2001, 2003, and 2007). In each of these submissions the sponsor included new clinical data to support their application. Although the review team changed over time, the conclusions on the merits of gepirone ER as an antidepressant did not.

This new re-submission, fourth in the sequence, was made in 2012 as an informal appeal to the last FDA non-approval action (2007) and contained no new clinical data. Therefore, in this submission the gepirone ER efficacy data submitted in 2007 were re-examined by a new clinical and statistical team.

The sponsor presented 12 short-term clinical trials and 1 maintenance trial with gepirone ER. Of these studies, only 2 short-term trials (ORG134001 and FK-GBE-007) support the efficacy of gepirone ER for the treatment of MDD.

It has been an FDA policy to approve antidepressants on the basis of 2 positive short-term trials. The rationale behind the request of a second positive trial has been the need for replication of the findings, i.e. the need to increase the possibility that the positive findings are the result of a true antidepressant effect rather than a random event. In this context, the 2 gepirone ER positive trials would support its approval for marketing. That is, if we were to ignore the rest of the available data on this drug. When we consider gepirone ER efficacy data as a whole, the possibility that these 2 positive trials are the evidence of a true antidepressant effect becomes quite distant.

Three of the 13 studies (CN105052, CN105078, and CN105083) were not informative in the evaluation of the efficacy of gepirone ER for the treatment of MDD and were not considered any further.

The remaining 7 short-term trials (ORG134002, FKGBE008, ORG134023, ORG134004, ORG134006, CN134017 and CN105053) showed no difference between gepirone ER and placebo. Four of these 7 studies included an active-control arm (ORG134004,

ORG134006, CN134017 and CN105053), in which the active controls performed consistently better than gepirone ER and placebo, reaching statistical significance over placebo in study CN105053, over gepirone ER in studies ORG134004 and CN134017, and over gepirone ER and placebo in study ORG134006.

The sponsor has argued that superiority of active control over gepirone ER is not evidence of assay sensitivity. However, in our view, findings of a statistically significant difference between two treatment arms show that the study was good enough to detect a difference, which is the essence of assay sensitivity. Interpreting superiority of the active control over the test drug as proof of assay sensitivity has been a long-standing FDA position and has been used in the past in the evaluation of efficacy data of other compounds seeking a claim for the treatment of MDD. In addition, in our decades-long experience with antidepressant development programs, we have found very few trials in which an effective antidepressant drug shows no effect while the active control does.

Another puzzling finding is that, in 4 of the 7 negative trials (ORG134023, ORG134004, ORG134006, and CN134017), gepirone ER performed worse than placebo on the primary endpoint and on many secondary variables. Again, in our experience, this is a very infrequent scenario with effective antidepressant drugs.

Finally, the negative maintenance gepirone ER trial (ORG28709) is an important piece of evidence against the efficacy of gepirone ER for the treatment of MDD. Study ORG28709 was an adequately designed trial, with an adequate number of patients enrolled, with response and relapse criteria similar to those used in other maintenance studies with antidepressants, and with a sufficient number of relapse events to detect a difference between treatment arms. In our review of all maintenance trials with antidepressants submitted to the FDA since the approval of the first second-generation antidepressant, every single maintenance trial with these characteristics has shown positive results. In this context, the negative results of this maintenance trial with gepirone ER are difficult to ignore.

In conclusion, although two short-term trials favor gepirone ER for the treatment of MDD, seven negative short-term studies and one maintenance trial with gepirone ER provide compelling evidence against the efficacy of gepirone ER as an antidepressant.

Major depressive disorder is a chronic condition, the leading cause of disability and an important cause of death worldwide, and represents a significant emotional and financial burden for patients and their families. This reviewer believes that the approval of gepirone ER for the treatment of MDD would have negative effects on the well-being of patients suffering from depression by delaying a much needed effective antidepressant treatment and increasing the risk of fatal outcomes.

2 Introduction and Regulatory Background

2.1 Product Information

Gepirone hydrochloride is a new molecular entity. A member of the azapirone class of compounds, it is an analog of buspirone. Buspirone, approved for the treatment of anxiety disorders, has been used as an augmentation strategy in patients with SSRI-resistant MDD. However, buspirone has shown low efficacy as monotherapy for treatment of MDD [1].

2.2 Currently Available Treatments for Proposed Indication

Antidepressant drugs are the mainstay of treatment for major depressive disorder. Numerous antidepressant medications are currently approved and available including: Tricyclic antidepressants (e.g. imipramine, desipramine, amitryptiline, nortryptiline, doxepin, amoxapine, trimipramine and maprotiline), Monoamine oxidase inhibitors (e.g. tranylcypromine, isocarboxazid, and selegiline patch), Selective serotonin reuptake inhibitors (e.g. fluoxetine, sertraline, paroxetine, citalopram, and escitalopram), Serotonin and norepinephrine reuptake inhibitors (e.g. venlafaxine, desvenlafaxine, and duloxetine), Other antidepressants (e.g. bupropion, trazadone, nefazodone, and mirtazapine).

2.3 Summary of Presubmission Regulatory Activity Related to Submission

Gepirone has not been approved for marketing in any country. According to the clinical review by Dr. Hearst, dated October 24, 2007, gepirone was originally developed by Mead Johnson and Bristol-Myers Company for the treatment of both anxiety and depression. In 1992, a business decision by Bristol-Myers Squibb (BMS) led to termination of the gepirone development program, resulting in discontinuation of ongoing Phase III studies in MDD. In 1993, Fabre-Kramer Pharmaceuticals, Inc. (Houston, Texas) acquired the rights to gepirone ER. In May 1998, Organon, Inc. executed an agreement with Fabre-Kramer Pharmaceuticals granting Organon rights to further develop and market gepirone.

The original New Drug Application (NDA) submitted by Organon on **September 30, 1999** was not filed by the Food and Drug Administration (FDA), which issued a refusal to file letter dated **November 30, 1999**.

Organon resubmitted the NDA on **May 18, 2001** and included one Phase III trial conducted by Organon with the ER dosage form (Study 134001), and three Phase II studies conducted by BMS with the older Immediate Release (IR) formulation (Studies 03A7A-003, 03A7C-001-B and 03A7A-002). According to the biometrics review dated

March 04, 2002, statisticians Roswitha Kelly, M.S. and Kooros Mahjoob, Ph.D. found that only the ER trial (study 134001) showed statistical significance in favor of Gepirone. 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, was 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. Further concerns with submission related to the fact that 14 other adequate and well-controlled trials with either dosage form had failed. Several of these studies were very similar in design and conduct to the four identified in support of the claim. In the biometrics reviewer's opinion these studies should have been included in the evaluation of the evidence of gepirone ER efficacy. The FDA issued a nonapprovable letter dated **March 15, 2002** citing inadequate evidence of effectiveness.

Organon amended the NDA on **December 23, 2003** with additional clinical data from a long-term relapse/prevention study (Study 28709). The interpretation of this study was confounded by issues related to reclassification of relapsed subjects after unblinding and definition of the intent-to-treat (ITT) population. These deficiencies were outlined in a second nonapprovable letter issued to Organon on **June 23, 2004**.

In June 2005, all rights to develop and market gepirone were reacquired by Fabre-Kramer Pharmaceuticals. On **May 3, 2007**, Fabre-Kramer resubmitted the NDA with 12 short-term trials and one maintenance trial in patients with MDD. On **October 19, 2007,** DPP held a regulatory briefing to discuss the gepirone ER case with the leadership of the Center for Drug Evaluation and Research (CDER). The CDER leadership agreed with DPP in that the analysis of the available data as a whole did not support the efficacy of gepirone ER. The FDA issued a non-approval letter on **November 2, 2007.**

At the sponsor’s request, a face-to-face guidance meeting between the FDA and the sponsor was held on **January 14, 2008**. At this meeting, the sponsor posed their arguments in support of gepirone ER efficacy, and the FDA reiterated that the data submitted by the sponsor as a whole did not provide sufficient evidence of gepirone ER efficacy for the treatment of MDD.

On **April 27, 2011** the sponsor requested reconsideration of the 2007 non-approval action. On **December 10, 2012,** following meetings, communications and exchanges of documents with the FDA, the sponsor submitted an NDA amendment providing information for an informal review of the gepirone ER efficacy data along with their current arguments in support of its efficacy. The efficacy data contained in this submission were the same as those reviewed in 2007 in their entirety. No new efficacy data was submitted.

It is of note that in the non-approval letter dated November 2, 2007, the sponsor was asked to address several Chemistry Manufacturing and Controls (CMC) deficiencies. In an email dated August 10, 2012, the FDA stated the following: “Your amendment must

be deemed a complete response in order to be considered acceptable for review. Therefore, all CMC deficiencies identified in the November 2, 2007 not approvable letter must be addressed. Furthermore, your submission of new CMC information could present additional issues that may impact the approvability of your application. For your new manufacturing site you will need to provide, in tabulated form, all changes comparable to the initial site including:
1. equipment changes
2. process changes
3. batch records
4. analytical methods
5. batch release data
6. stability data
7. container packaging/closure system
8. cGMP statement of readiness for inspection
Please note that we will evaluate the provided data and the expiry will be set accordingly."

However, at the sponsor's request, in an email dated October 02, 2012, the FDA agreed to review their submission containing the efficacy studies submitted in 2007 and their current arguments in support of gepirone efficacy as an informal appeal, without requiring the submission of the CMC information mentioned above as a prerequisite for review. Therefore, at the time this clinical review was completed, there was not enough CMC information to assert the sponsor's ability to manufacture their product in acceptable conditions.

3 Significant Efficacy/Safety Issues Related to Other Review Disciplines

3.1 Chemistry Manufacturing and Controls

This submission does not address pending CMC deficiencies described in the non-approval letter dated November 2, 2007 (refer to section 2.5 "Summary of Presubmission Regulatory Activity Related to Submission" for further description of these issues).

4 Sources of Clinical Data

4.1 Tables of Studies/Clinical Trials

The primary sources of data for this review are twelve short-term studies and one longer-term study (Table 1-3), which examined the efficacy of gepirone ER for the acute and maintenance treatment of MDD.

Table 1: Positive trials	
Study Number	Study Description[1]
ORG134001	8-week, randomized, double-blind, placebo-controlled trial of the efficacy and safety of gepirone ER (20 mg to 80 mg) for the treatment of MDD: gepirone ER (n=101); placebo (n=101).
FK-GBE-007	8-week, randomized, double-blind, placebo-controlled trial of the efficacy and safety of gepirone ER (40 mg to 80 mg) for the treatment of MDD: gepirone ER (n=116); placebo (n=122).

Table 2: Negative trials	
Study Number	Study Description[1]
ORG134002	8-week, randomized, double-blind, placebo-controlled trial of the efficacy and safety of gepirone ER (20 mg to 80 mg) for the treatment of MDD: gepirone ER (n=102); placebo (n=103).
FK-GBE-008	8-week, randomized, double-blind, placebo-controlled trial of the efficacy and safety of gepirone ER (40 mg to 80 mg) for the treatment of MDD: gepirone ER (n=96); placebo (n=99).
ORG134023	9-week, randomized, double-blind, placebo-controlled trial of the efficacy and safety of gepirone ER (20 mg to 80 mg) for the treatment of MDD: gepirone ER (n=123); placebo (n=123).
ORG134004	8-week, randomized, double-blind, placebo and active-controlled trial of the efficacy and safety of gepirone ER (40 mg to 80 mg) for the treatment of MDD with atypical features (active control: fluoxetine 20 mg to 40 mg): gepirone ER (n=124); placebo (n=130).
ORG134006	8-week, randomized, double-blind, placebo and active-controlled trial of the efficacy and safety of gepirone ER (20 mg to 80 mg) for the treatment of MDD with atypical features (active control: paroxetine 10 mg to 40 mg): gepirone ER (n=140); placebo (n=143).
ORG134017	8-week, randomized, double-blind, placebo and active-controlled trial of the efficacy and safety of gepirone ER (40 mg to 80 mg) for the treatment of MDD (active control: fluoxetine 20 mg to 40 mg): gepirone ER (n=159); placebo (n=159).
CN105-053	8-week, randomized, double-blind, placebo and active-controlled trial of the efficacy and safety of gepirone ER (10 mg to 60 mg) for the treatment of MDD (active control: imipramine 50 mg to 200 mg): gepirone ER (n=56); placebo (n=56).
ORG28709	MDD maintenance trial with an 8-12-week OL phase (gepirone ER 20-80 mg) followed by a 40-44-week randomized, double-blind, placebo-controlled phase: gepirone ER (n=126); placebo (n=124).

1 n represents number of subjects in the intent-to-treat population

Table 3: Failed trials	
Study Number	Study Description[1]
CN105-052	8-week, randomized, double-blind, placebo and active-controlled trial of the efficacy and safety of gepirone ER (20 mg to 60 mg) for the treatment of MDD (active control: fluoxetine 20 mg): gepirone ER (n=35); placebo (n=37).
CN105-078	6-week, randomized, double-blind, placebo-controlled, three-arm trial of the efficacy and safety of gepirone ER (10 mg to 100 mg) for the treatment of MDD: gepirone ER 10-50 mg (n=48); gepirone ER 20-100 mg (n=40); placebo (n=47).
CN105-083	6-week, randomized, double-blind, placebo-controlled, three-arm trial of the efficacy and safety of gepirone ER (10 mg to 100 mg) for the treatment of MDD: gepirone ER 10-50 mg (n=36); gepirone ER 20-100 mg (n=37); placebo (n=39).

4.2 Review Strategy

This review consisted of an examination of relevant background clinical information and efficacy data from the 2001, 2003, and 2007 submissions for NDA 21164, clinical reviews by Dr Hearst (2007), statistical reviews by Ms. Kelly and Dr. Mahjoob (2002), Ms. Kelly (2004), Dr. Kong (2007), Dr. Chen (2013), new statistical analysis by Dr. Zhong (no review on file), and the sponsor's current arguments in support of the efficacy of gepirone ER contained in this submission. Given that this informal appeal involves no new clinical data, that the safety data was reviewed in previous submissions, and that the sponsor's arguments focus mainly on gepirone ER efficacy, no review of safety data was performed at this time.

In addition to their arguments on gepirone ER efficacy contained in this submission, the sponsor included post-hoc analyses of the effects of gepirone ER on sexual dysfunction, arguing that this adverse event occurs less frequently with gepirone ER than with other antidepressants. However, in this reviewer's opinion, the sponsor's argument on the advantage of gepirone ER safety profile over other antidepressants is a moot point, since the efficacy data does not provide sufficient evidence of gepirone ER efficacy. Furthermore, the absence of adverse events generally associated with antidepressants would not surprise in the absence of an antidepressant effect. Therefore, the sponsor's arguments on sexual dysfunction are not discussed in this review.

Clinical Review
Silvana Borges, M.D.
NDA 21164/Informal appeal
Travivo® (gepirone hydrochloride ER tablets)

5 Review of Efficacy

Efficacy Summary

The sponsor presented 12 short-term clinical trials and 1 maintenance trial with gepirone ER, of which only 2 short-term studies support the efficacy of gepirone ER for the treatment of MDD. Three of the 13 studies were considered not informative. The remaining 7 short-term trials showed no difference between gepirone ER and placebo, with active controls (and sometimes placebo) performing consistently better than gepirone ER (Table 4). The negative findings of the gepirone ER maintenance trial were also puzzling. Therefore, the available data considered as a whole does not support gepirone ER efficacy for the treatment of MDD.

Table 4. 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 LS Means (SE) p-values	101 -6.57	101 -9.04	NA	 -2.47 (0.98) p=0.013	NA	NA
FKGBE007 N LS Means (SE) p-values	122 -7.79	116 -10.24	NA	 -2.45 (1.02) P=0.018	NA	NA
ORG 134002 N LS Means (SE) p-values	103 -9.24	102 -9.95	NA	 -0.71 (0.88) P=0.42	NA	NA
FKGBE008 N LS Means (SE) p-values	99 -8.48	96 -9.86	NA	 -1.38 (1.06) P=0.20	NA	NA
CN105-078 N LS Means (SE) p-values	47 -6.42	88 -7.42	NA	 -1.0 (1.10) P=0.36	NA	NA
CN105-083 N LS Means (SE) p-values	39 -8.97	73 -9.46	NA	-0.49 (1.53) P=0.75	NA	NA
ORG 134023 N LS Means (SE) p-values	123 -8.05	123 -7.93	NA	 0.13(0.97) P=0.90	NA	NA
CN105-052 N LS Means (SE) p-values	37 -10.29	35 -10.98	(Fluoxetine) 36 -10.95	-0.69 (2.05) P=0.74	0.02 (2.06) P=0.99	-0.67 (2.03) P=0.74
CN105-053 N LS Means (SE) p-values	56 -8.16	56 -10.16	(Imipramine) 54 -11.35	-2.0 (1.51) P=0.19	1.2 (1.53) P=0.44	-3.19 (1.52) 0.038
ORG 134004 N LS Means (SE) p-values	130 -6.79	124 -5.75	(Fluoxetine) 134 -7.46	1.04 (0.78) P=0.18	1.71 (0.77) P=0.027	-0.68 (0.76) P=0.38
ORG 134017 N LS Means (SE) p-values	159 -11.02	159 -10.37	(Fluoxetine) 159 -11.92	0.65 (0.76) P=0.39	1.54 (0.76) P=0.042	-0.90 (0.76) P=0.24
ORG 134006 [b] N LS Means (SE) p-values	143 -7.31	144 -7.09	(Paroxetine) 136 -8.94	0.22 (0.72) P=0.76	1.85 (0.73) P=0.012	-1.63 (0.73) P=0.026

From Dr. Kong's review (2007)
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 ORG134006 when centers 1 and 12 are pooled together.

5.1 Studies Pertinent to the Efficacy Claim of gepirone ER for the Treatment of MDD (study start years are included in brackets)

5.1.1 Positive studies

5.1.1.1 Study ORG134001 (1999)

This was an 8-week, randomized, double-blind, flexible-dose, placebo-controlled trial of the efficacy and safety of gepirone ER (20 mg to 80 mg) for the treatment of MDD: gepirone ER (n=101); placebo (n=101).

- Inclusion criteria

- Males or females, 18-70 years of age with a diagnosis of MDD per DSM-IV
- Depressive symptoms for at least 4 weeks before study entry.
- HAM-D 17 total score ≥20 at screening and baseline.

- Primary efficacy measure**:** change from baseline in HAMD-17 total score at endpoint (Week 8 or last visit).

- Results:

Table 5: Sponsor's HAMD-17 Total Score for Change from Baseline at End of Treatment by Center

Center	Number of Subjects		Adjusted Mean Change		Treatment Diff (95% CI) gepirone-ER – Placebo	SE	p-value
	gepirone-ER	Placebo	gepirone-ER	Placebo			
Center 1	33	32	-10.46	-6.59	-3.88 (-6.82, -0.94)	1.47	0.011
Center 2	24	25	-7.84	-8.71	0.87 (-2.90, 4.63)	1.87	0.645
Center 3	25	24	-11.58	-8.31	-3.27 (-7.68, 1.15)	2.19	0.143
Center 4	15	15	-9.57	-5.03	-4.54 (-11.44, 2.37)	3.36	0.189
Center 5	4	5	-6.03	-3.77	-2.26 (-10.73, 6.22)	3.46	0.539
All Centers	101	101	-9.04	-6.57	-2.47 (-4.41, -0.53)	0.98	0.013
Treatment by Center Interaction p-value = 0.385							
Analysis is based on ANCOVA with effects for treatment, center, and baseline value as a covariate. [*Source: Statistical Table 3.1 in the ISE*]							

5.1.1.2 Study FK-GBE-007 (2003)

This was an 8-week, randomized, double-blind, flexible-dose, placebo-controlled trial of the efficacy and safety of gepirone ER (20 mg to 80 mg) for the treatment of MDD: gepirone ER (n=116); placebo (n=122).

➢ Inclusion criteria

- Males or females, 18-70 years of age with a diagnosis of MDD per DSM-IV
- Depressive symptoms for at least 4 weeks before study entry.
- HAM-D 17 total score ≥20 at screening and baseline.

➢ Primary efficacy measure**:** change from baseline in HAMD-17 total score at endpoint (Week 8 or last visit).

➢ Results:

Table 6: Sponsor's HAMD-17 Total Score for Change from Baseline at End of Treatment by Center

Center	Number of Subjects		Adjusted Mean Change		Treatment Difference		
	gepirone-ER	Placebo	gepirone-ER	Placebo	gepirone-ER – Placebo (95% CI)	SE	p-value
Center 701	21	23	-11.49	-5.34	-6.15 (-10.33, -1.97)	2.07	0.005
Center 702	14	16	-12.33	-8.02	-4.31 (-11.30, 2.67)	3.40	0.216
Center 704	22	20	-10.38	-11.63	1.25 (-2.63, 5.13)	1.92	0.519
Center 705	14	15	-5.98	-8.35	2.36 (-4.88, 9.61)	3.53	0.509
Center 706	10	9	-12.89	-5.34	-7.55 (-14.85, -0.24)	3.45	0.044
Center 708	12	10	-10.91	-10.51	-0.40 (-6.85, 6.05)	3.08	0.898
Center 709	8	11	-11.55	-4.24	-7.31 (-15.41, 0.79)	3.82	0.074
Center 999	15	18	-7.20	-7.94	0.74 (-5.28, 6.75)	2.95	0.804
All Centers	116	122	-10.24	-7.79	-2.45 (-4.47, -0.43)	1.02	0.018
Treatment by Center Interaction p-value = 0.092							

Analysis is based on ANCOVA with terms for treatment, center, and baseline value as a covariate; Center 999 is a pooled center that combines centers 703 and 707
[Source: ISE Table 16, Appendix A Statistical Table 3.2]

As shown in the table 5 and 6, for Study ORG134001, the final positive results for all centers appeared to be driven by Center 1, and for Study FK-GBE-007, the final positive results for all centers appeared to be driven by Center 701 and Center 706. In particular, it is worth noting that some centers even had a larger improvement in placebo patients than in drug patients. For example, for Study ORG134001, in Center 2 (n=49), placebo patients' change was -8.71 in HAMD-17 total score but gepirone patients' change was only -7.84. This wrong trend was also observed in Center 704 and 705 for Study FK-GBE-007.

According to Dr. Kong's review, 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. However, further *post hoc* subgroup analyses suggest that the treatment effect was mainly driven by Caucasians, and female patients, as shown in Table 7.

Table 7. Treatment Effect by Sex and Age Groups on the effect size in Study FK-GBE-007 (LOCF Analysis)

	Placebo	Org 33062 ER	Difference
Sex Effect			**(p-value*)**
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.

Although some weakness in the efficacy results has been found, this reviewer agrees with the sponsor and previous FDA reviewers that studies ORG134001 and FK-GBE-007 are positive gepirone ER trials.

5.1.2 Negative studies

5.1.2.1 Study ORG134002 (1999)

This was an 8-week, randomized, double-blind, flexible-dose, placebo-controlled trial of the efficacy and safety of gepirone ER (20 mg to 80 mg) for the treatment of MDD: gepirone ER (n=102); placebo (n=103).

- Inclusion criteria

- Males or females, 18-70 years of age with a diagnosis of MDD per DSM-IV
- Daily dysphoria for at least 4 weeks before study entry.
- HAM-D 17 total score ≥20 at screening and baseline.

- Primary efficacy measure**:** change from baseline in HAMD-17 total score at endpoint (Week 8 or last visit).

- Results:

By the sponsor's own writing, study 134002 was adequately designed, properly conducted in the appropriate population of MDD patients, and employed doses of gepirone-ER in the correct therapeutic range, but it failed to meet its primary objective (Table 8). The sponsor argues this is due at least in part to the high placebo response.

Clinical Review
Silvana Borges, M.D.
NDA 21164/Informal appeal
Travivo® (gepirone hydrochloride ER tablets)

Table 8: Sponsor's HAMD-17 Total Score for Change from Baseline at End of Treatment by Week

Treatment Group		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
gepirone-ER	n	107	101	102	102	102	102	102
	Mean	23.96	-4.34	-6.68	-8.71	-9.50	-9.92	-9.96
	SE	0.27	0.38	0.46	0.55	0.63	0.67	0.65
Placebo	n	104	101	103	103	103	103	103
	Mean	24.11	-3.73	-6.12	-8.04	-8.66	-9.74	-9.29
	SE	0.29	0.38	0.46	0.55	0.63	0.67	0.65
Difference: (gep-ER – placebo)			-0.61	-0.56	-0.67	-0.84	-0.18	-0.67
p-value			0.235	0.375	0.370	0.322	0.841	0.446

Based on ANOVA with effects for treatment and center; ET=End of Treatment; LS=Least Squares; SE=Standard Error of the Mean
[*Source: CSR 134002 Table 12, Appendices F8.6.1.1-2 and F8.6.1.1-4*]

The placebo response (average reduction from baseline at study end expressed as a percentage of the average baseline value) in Study 134002 was 38.5%. As noted in the statistical review by Ms. Kelly and Dr. Mahjoob, in the absence of an active control, it is subjective to judge 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.

In addition, the sponsor interprets this study to be supportive of gepirone ER efficacy, arguing that treatment effects consistently favored gepirone-ER over placebo for each of the secondary efficacy variables. It is of note that none of these measures were pre-specified endpoints and, as shown in Table 9, per the sponsor's own analysis, this is not the case. For instance, gepirone ER performed worse than placebo for HAMD-17 remitters or CGI Improvement. Additionally, with the exception of HAMD-25 responders and HAMD-Item 1, none of these secondary variables reached statistical significance. In agreement with previous clinical and statistical reviews, this reviewer considers study ORG134002 to be a negative gepirone ER trial.

Table 9: Sponsor's Summary of Secondary Efficacy Results at Endpoint (ITT/LOCF)

Parameter	End-of-Treatment Outcome		Difference	p-Value
	gepirone-ER	Placebo		
HAMD-17 Responders (%)	40.2%	32.0%	8.2%	0.225
HAMD-17 Remitters (%)	15.7%	17.5%	-1.8%	0.731
HAMD-21 CFB	-10.77	-9.90	-0.87	0.365
HAMD-25 CFB	-12.54	-11.00	-1.54	0.153
HAMD-25 Responders (%)	42.2%	29.1%	13.1%	**0.014**
HAMD-28 CFB	-14.32	-12.51	-1.81	0.150
HAMD-Item 1 CFB	-1.30	-1.01	-0.18	**0.036**
HAMD-Factor 1 CFB	-3.14	-2.77	0.37	0.277
CGI Severity CFB	-1.09	-0.88	-0.21	0.130
CGI Improvement	2.51	2.73	-0.22	0.145
CGI Responders (%)	52.0%	44.7%	7.3%	0.297
MADRS CFB	-11.55	-9.21	-2.34	**0.078**
Bech-6 CFB	-5.90	-4.95	-0.95	**0.076**

[*Source: 134002 CSR Table 21, Appendix F8.6.2*]
CFB=Change from baseline; LS means from ANOVA model, including effects for treatment and center; % responders and % remitters were compared using CMH test. CGI Responders are subjects much or very much improved based on the CGI improvement score at a post-baseline assessment. HAMD-17 Responders are subjects with ≥ 50% reduction from baseline at a post-baseline assessment. HAMD-25 Responders are subjects with ≥ 50% reduction from baseline at the endpoint assessment.

5.1.2.2 Study FK-GBE-008 (2003)

This was an 8-week, randomized, double-blind, flexible-dose, placebo-controlled trial of the efficacy and safety of gepirone ER (20 mg to 80 mg) for the treatment of MDD: gepirone ER (n=96); placebo (n=99).

- Inclusion criteria

- Males or females, 18-64 years of age with a diagnosis of MDD per DSM-IV
- Daily dysphoria for at least 4 weeks before study entry.
- HAM-D 17 total score ≥20 at screening and baseline.

- Primary efficacy measure**:** change from baseline in HAMD-17 total score at endpoint (Week 8 or last visit).

- Results:

Table 10: Sponsor's HAMD-17 Total Score for Change from Baseline at End of Treatment by Week

Treatment		Baseline	LS Mean Change from Baseline				
			Week 2	Week 3	Week 4	Week 6	Week 8/ET
gepirone-ER	n	99	93	96	96	96	96
	Mean	24.2	-5.62	-7.31	-8.52	-9.46	-9.87
	SE	0.30	0.44	0.55	0.62	0.69	0.75
Placebo	n	100	93	98	98	98	98
	Mean	24.0	-4.10	-5.81	-7.17	-7.52	-8.37
	SE	0.27	0.44	0.54	0.61	0.68	0.75
Difference: (gep-ER – placebo)			-1.52	-1.51	-1.35	-1.94	-1.50
p-value			0.016	0.053	0.123	0.046	0.159

ET=End of treatment; LS=Least squares; SE=Standard error
[*Source: CSR FK-GBE-008 Table 16, Supportive Table 21*]

In the sponsor's own words, the FK-GBE-008 study was adequately designed and executed, and employed doses in the appropriate therapeutic range for gepirone-ER. As with study ORG134002, the sponsor interprets this study to be supportive of gepirone ER efficacy, arguing that treatment effects consistently favored gepirone-ER over placebo for each of the secondary efficacy variables. This reviewer acknowledges that the directional trend of the primary endpoint and the secondary variables favors gepirone ER. However, none of the secondary measures were pre-specified endpoints and, as shown in Table 11, per the sponsor's own analysis, almost none of them reached statistical significance at endpoint. In agreement with previous clinical and statistical reviews, this reviewer considers study FK-GBE-008 to be a negative gepirone ER trial.

Table 11: Sponsor's Summary of Secondary Efficacy Results at Endpoint (ITT/LOCF)

Parameter	Week8/End of Treatment		Difference	p-value
	gepirone-ER	Placebo		
HAMD-17 responders (%)	39.6%	32.7%	6.9%	0.293
MADRS CFB	-11.73	-9.87	-1.86	0.208
MADRS responders (%)	39.6%	29.6%	10.0%	0.128
HAMD-21 CFB	-10.67	-9.14	-1.52	0.209
HAMD-21 responders (%)	40.6%	32.7%	7.9%	0.231
HAMD-25 CFB	-11.51	-10.14	1.37	0.281
HAMD-25 responders (%)	44.8%	30.6%	14.2%	0.035
HAMD-28 CFB	-13.47	-11.94	-1.53	0.319
HAMD-28 responders (%)	42.7%	33.7%	9.0%	0.178
HAMD Item 1 CFB	-1.10	-0.99	-0.11	0.469
CGI severity CFB	-1.27	-1.08	-0.19	0.275
CGI responders (%)	47.9%	37.8%	10.1%	0.147
HAMD-17 remitters (%)	22.9%	15.3%	7.6%	0.156

[*Source: FK-GBE-008 CSR Appendix 15 Supportive Tables 22, 29-30, 43-44, 51-52, 59-60, 67, 73-74 and 103*]
CFB=Change from baseline; LS means are from ANOVA model, including effects for treatment and center; % responders and % remitters were compared using CMH test. For HAMD and MADRS scales, responders are subjects with ≥ 50% reduction from baseline at any post-baseline assessment; HAMD-17 remitters are subjects with a HAMD-17 total score of ≤ 7. For CGI, responders are much or very much improved on the CGI improvement score at any post-baseline assessment.

5.1.2.3 Study ORG134023 (2003)

This was a 9-week, randomized, double-blind, flexible-dose, placebo-controlled trial of the efficacy and safety of gepirone ER (20 mg to 80 mg) for the treatment of MDD: gepirone ER (n=123); placebo (n=123).

- Inclusion criteria

- Males or females, 18-70 years of age with a diagnosis of MDD per DSM-IV
- Current episode of MDD for a minimum of 1 month
- MADRS total score ≥30 at screening and baseline
- Dysphoria for most days over the past 4 weeks

- Primary efficacy measure**:** change from baseline in HAMD-17 total score at endpoint (Week 9 or last visit).

- Results:

The sponsor acknowledges that this is a negative gepirone ER trial. In their own words, no statistically significant treatment effects were detected for gepirone ER based on the primary or secondary efficacy variables (Tables 12 and 13). However, the sponsor cites

a high placebo response (an 8-unit improvement in HAMD-17 and 39% CGI responder rate) as a possible contribution to the study's failure.

In this regard, it is worth noting that placebo patients showed an improvement of 7.79 units in the HAMD-17 in one of the positive trials (study FKGBE007) and the CGI responder rate was 35.6% in the placebo group in the other positive trial (study ORG134001). It is also noteworthy that in study ORG134023, placebo performed numerically better on 4 of the 6 study visits on the primary endpoint and in several secondary measures at endpoint (Table 12 and 13).

Table 12: Sponsor's HAMD-17 Total Score for Change from Baseline at End of Treatment by Week

Treatment Group		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 5	Week 7	Week 9/ET
gepirone-ER	n	123	123	123	123	123	123	123
	Mean	22.9	-2.8	-4.7	-6.2	-7.6	-8.3	-8.0
	SE	0.32	0.34	0.44	0.55	0.61	0.65	0.72
Placebo	n	123	123	123	123	123	123	123
	Mean	22.8	-3.1	-4.9	-6.8	-7.5	-8.5	-8.0
	SE	0.33	0.34	0.45	0.56	0.61	0.66	0.72
Difference: (gep-ER – placebo)			0.3	0.2	0.6	-0.1	0.2	0.0
p-value			0.558	0.815	0.433	0.903	0.829	0.947

[*Source: CSR 134023 Table 13*]

Table 13: Sponsor's Summary of Secondary Efficacy Results at Endpoint (ITT/LOCF)

Parameter	End-of-Treatment Outcome		Difference	p-Value
	gepirone-ER	Placebo		
HAMD-17 Responders (%)	33%	37%	-4%	0.518
HAMD-17 Remitters (%)	22%	24%	-2%	0.731
HAMD-25 CFB	-10.4	-10.0	-0.4	0.739
HAMD-Item 1 CFB	-1.1	-1.2	0.1	0.438
CGI Severity CFB	-1.2	-1.3	0.1	0.467
CGI Improvement	2.7	2.7	0.0	0.864
CGI Responders (%)	43%	39%	4%	0.624
MADRS CFB	-13.9	-13.1	-0.8	0.572
Bech-6 CFB	-4.6	-4.5	-0.1	0.906

[*Source: CSR 134023 Tables 14-22*]
CFB=Change from baseline; LS means are from ANOVA model, including effects for treatment and center; % responders and % remitters were compared using CMH test. For HAMD-17 responders are subjects with ≥ 50% reduction from baseline at any post-baseline assessment; HAMD-17 remitters are subjects with a HAMD-17 total score ≤ 7. For CGI, responders are much or very much improved on the CGI improvement score at any post-baseline assessment.

In agreement with previous clinical and statistical reviews, this reviewer considers study ORG134023 to be a negative gepirone ER trial.

5.1.2.4 Studies ORG134004 (2000) and ORG134006 (2000)

Given the similarities in study design, study population and sponsor's arguments, studies ORG134004 and ORG134006 are discussed together.

Study ORG134004 was an 8-week, randomized, double-blind, flexible-dose, placebo and active-controlled trial of the efficacy and safety of gepirone ER (20 mg to 80 mg) for the treatment of MDD with atypical features (active control: fluoxetine 20 mg to 40 mg): gepirone ER (n=124); placebo (n=130).

Study ORG134006 was an 8-week, randomized, double-blind, flexible-dose, placebo and active-controlled trial of the efficacy and safety of gepirone ER (20 mg to 80 mg) for the treatment of MDD with atypical features (active control: paroxetine 10 mg to 40 mg): gepirone ER (n=140); placebo (n=143).

- Inclusion criteria

- Males or females, 18-65 years of age with a diagnosis of MDD with Atypical features specifier per DSM-IV with either a single episode or recurrent episodes of moderate to severe depression.
- No more than 20% decrease on the HAMD-25 total score between screening and baseline
- Current episode of MDD with atypical features lasting at least 3 months.

- Primary efficacy measure**:** change from baseline in HAMD-25 total score at endpoint (Week 8 or last visit).

- Results:

The sponsor's arguments for considering trials ORG134004 and ORG134006 as failed studies (and this reviewer's interpretation of the data) are as follows:

- No Assay Sensitivity

Although in both trials the active control fails to reach statistical significance to show superiority over placebo on the primary endpoint (HAMD-25), the treatment effect favors both fluoxetine and paroxetine (i.e. the active control showed a better outcome than placebo) for both HAMD-25 and HAMD-17 (Tables 4, 14 to 17). That is not the case for gepirone ER. Using HAMD-17 as the endpoint, fluoxetine beats gepirone ER in study ORG134004 and paroxetine beats both gepirone ER and placebo in study ORG134006 (Table 4). In addition, the equal or smaller dropout rate due to lack of efficacy (Table 20) does support the efficacy of active control drugs. That is not the case for gepirone ER.

Table 14. Sponsor's change from baseline in HAMD-25 Total Score per visit (study ORG134004)

Treatment Group		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
gepirone-ER (N=125)	N	125	124	124	124	124	124	124
	Mean	27.9	-3.72	-5.99	-7.99	-8.96	-9.61	-9.76
	SE	0.44	0.43	0.54	0.64	0.70	0.74	0.77
placebo† (N=130)	N	130	130	130	130	130	130	130
	Mean	27.6	-3.94	-6.75	-8.16	-10.12	-10.94	-10.63
	SE	0.44	0.42	0.53	0.63	0.68	0.72	0.75
gepirone-ER vs. pbo	p-value		0.706	0.316	0.845	0.230	0.193	0.416
fluoxetine (N=136)	N	136	134	134	134	134	134	134
	Mean	28.1	-4.07	-5.77	-8.60	-9.93	-10.38	-11.66
	SE	0.46	0.43	0.53	0.60	0.64	0.73	0.75
placebo† (N=130)	N	130	130	130	130	130	130	130
	Mean	27.6	-3.84	-6.64	-8.12	-10.14	-10.91	-10.61
	SE	0.44	0.43	0.53	0.61	0.65	0.74	0.76
fluoxetine vs. pbo	p-value		0.699	0.242	0.576	0.817	0.605	0.325

ET = end of trial; LS = least squares; SE = standard error of the mean
†LS means and p-values from separate ANOVA models, with effects for treatment (2 groups: active drug vs. placebo) and center; preliminary tests of treatment x center interaction were not significant (p > 0.10) at any visit.
[Source: *CSR 134004 Table 16, Appendix F Tables 6.1.1.4 and 6.1.1.4A*]

Table 15: Sponsor's secondary efficacy results at endpoint: study ORG134004

Efficacy Variable	End of Treatment Outcome			Pairwise Tests (p-values)		
	Gep-ER	Fluoxetine	Placebo	G vs. P	F vs. P	G vs. F
HAMD-17 CFB	-5.67	-7.5	-6.55	0.282†	NR	NR
HAMD-28 CFB	-11.54	-14.0	-12.52	0.438	NR	NR
HAMD-Item1 CFB	-0.97	-1.2	-1.11	0.328	NR	NR
CGI (severity) CFB	-0.98	-1.2	-1.11	0.392	NR	NR
CGI (global improvement)	2.98	2.70	2.76	0.142	NR	NR
% Responders (HAMD-25)	33.87%	NR	36.15%	0.765	NR	NR
% Responders (CGI)	34.68%	NR	42.42%	0.224	NR	NR
% Remitters (HAMD-25)	16.94%	NR	23.85%	0.178	NR	NR
HAMD-31 (items 22-26) CFB	-2.82	-2.57	-2.80	0.948	NR	0.459
HAMD Factor 1 CFB (anxiety)	-2.09	-2.56	-1.93	0.594	NR	0.154
HAMA total score CFB	-4.08	-5.68	-4.95	0.226	NR	**0.025**

[*Source: ORG134004 CSR Appendix F Tables 6.1.1.1-6.1.2.8.1*]
LS means and p-values from ANOVA, with treatment (2 groups) and site as factors; Cochran-Mantel-Haenszel (CMH) test for % responders.
†Significant treatment-by-site interaction (p=0.05) for HAMD-17 CFB (gepirone-ER vs. placebo).
NR = Not Reported; p-values are presented for all protocol-defined statistical comparisons.

Table 16. Sponsor's change from baseline in HAMD-25 Total Score per visit (study ORG134006)

Treatment Group		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
gepirone-ER (N=143)	N	143	140	140	140	140	140	140
	Mean	27.0	-4.63	-6.24	-8.51	-9.00	-10.58	-10.94
	SE	0.38	0.48	0.53	0.57	0.63	0.69	0.74
placebo† (N=143)	N	143	143	143	143	143	143	143
	Mean	26.9	-4.87	-7.23	-8.37	-9.46	-10.63	-11.00
	SE	0.36	0.48	0.53	0.57	0.63	0.69	0.75
gepirone-ER vs. placebo		p-value	0.698	0.154	0.850	0.578	0.957	0.953
paroxetine (N=136)	N	136	136	136	136	136	136	136
	Mean	26.7	-4.95	-6.40	-9.18	-10.70	-12.13	-12.58
	SE	0.41	0.51	0.53	0.62	0.66	0.73	0.79
placebo† (N=143)	N	143	143	143	143	143	143	143
	Mean	26.9	-5.01	-7.47	-8.72	-9.88	-10.99	-11.23
	SE	0.36	0.49	0.51	0.60	0.64	0.70	0.76
paroxetine vs. placebo		p-value	0.917	0.114	0.566	0.326	0.220	0.178

ET = end of trial; LS = least squares; SE = standard error of the mean
† LS means and p-values from separate ANCOVA models, with effects for treatment (2 groups: active drug vs. placebo) and center; preliminary tests of treatment x center interaction were not significant (p > 0.10) at any visit for gepirone-ER vs. Placebo; the interactions were statistically significant for paroxetine vs. placebo at Weeks 1 (p=0.007), 2 (p=0.043), 3 (p=0.050), 6 (p=0.034) and 8/ET (p=0.024).
[Source: *CSR 134006 Table 19, Appendix F Tables 6.1.1.4A and 6.1.1.4AA*]

Table 17: Sponsor's secondary efficacy results at endpoint: study ORG134006

Efficacy Variable	End of Treatment Outcome			Pairwise Tests (p-values)*		
	Gep-ER	Paroxetine	Placebo	G vs. P	Px vs. P	G vs. Px
HAMD-17 CFB	-6.92	-9.1	-7.15	0.750	NR	NR
HAMD-28 CFB	-12.68	-14.8	-12.57	0.927	NR	NR
HAMD-Item1 CFB	-1.11	-1.4	-1.07	0.720	NR	NR
CGI (severity) CFB	-1.10	-1.4	-1.21	0.423	NR	NR
CGI (global improvement)	2.68	2.3	2.63	0.726	NR	NR
HAMD Bech-6 score CFB	-4.60	-6.0	-4.55	0.918	NR	NR
% Responders (HAMD-25)	42.86%	NR	41.96%	0.894	NR	NR
% Responders (CGI)	45.71%	NR	46.85%	0.808	NR	NR
% Remitters (HAMD-25)	30.71%	NR	32.87%	0.652	NR	NR
HAMD-31 (items 22-26) CFB	-2.76	-2.10	-2.47	0.253	NR	**0.012**

[*Source: ORG134006 Final Report Appendix F Tables 6.1.1.4AA-6.1.2.3.3*]
*LS means and p-values from ANCOVA, with treatment (gepirone-ER vs. placebo) and center as factors, baseline value as covariate; ANCOVA applied to 3 treatments for HAMD-31 (items 22-26) only. Cochran-Mantel-Haenszel (CMH) test used for % responders.
NR=Not Reported; p-values are presented for all protocol-defined statistical comparisons.

- High Placebo Response Rate

As seen in Table 18, the HAMD-17 change from baseline for placebo was actually smaller in studies 134004 and 134006 than in the positive trials (134001 and FKGBE007). This is also the case for the CGI-I score. Other variables show slightly higher values for placebo in studies 134004 and 134006 than in the positive trials.

However, it is worth noting that the active controls (fluoxetine and paroxetine) were consistently better numerically than placebo, while gepirone ER was consistently worse.

Table 18: Primary and secondary variables for positive studies ORG134001 and FKGBE007, and negative studies ORG134004 and ORG134006

Efficacy variable	**End of treatment outcome Study 134001**		**End of treatment outcome Study FKGBE007**		**End of treatment outcome Study 134004**			**End of treatment outcome Study 134006**		
	Gep-ER	**Placebo**	**Gep-ER**	**Placebo**	**Gep-ER**	**Fluoxetine**	**Placebo**	**Gep-ER**	**Paroxetine**	**Placebo**
HAMD-25 CFB	-11.57	**-8.19**	-12.65	**-9.85**	-9.76	-11.66	**-10.63**	-10.94	-12.58	**-11.00**
HAMD-17 CFB	-9.04	**-6.57**	-10.24	**-7.79**	-5.67	-7.5	**-6.55**	-6.92	-9.1	**-7.15**
HAMD-28 CFB	-13.27	**-9.60**	-15.04	**-11.83**	-11.54	-14.00	**-12.52**	-12.68	-14.80	**-12.57**
HAMD-Item 1 CFB	-1.16	**-0.78**	-1.22	**-0.97**	-0.97	-1.2	**-1.11**	-1.11	-1.4	**-1.07**
CGI-S CFB	-1.19	**-0.79**	-1.30	**-0.92**	-0.98	-1.2	**-1.11**	-1.10	-1.4	**-1.21**
CGI- I	2.82	**3.10**	NR	NR	2.98	2.7	**2.76**	2.68	2.3	**2.63**

Source: Reviewer, based on sponsor's data from CSRs.
CFB = change from baseline; NR = not reported

In addition, as seen in Tables 19 and 20, in studies 134004 and 134006 the dropout rates (both total and by lack of efficacy only) tended to be larger in the gepirone ER groups than in the placebo groups. That is not the case for active control arms.

Table 19: All Studies Dropout Rates based on Total Patient Population

Study (%)	ORG	active	Placebo	Outcome
ORG134001	27		24	positive
FK-GBE-007	22		18	positive
ORG134002	32		29	negative
FK-GBE-008	25		21	negative
ORG134023	26		21	negative
CN105-053	41	37	61	negative
ORG134004	36	18	21	negative
ORG134006	31	29	24	negative
ORG134017	32	24	21	negative
CN105-078	42		31	failed
CN105-083	37		34	failed
CN105-052	43		50	failed

Table 20: All Studies Dropout Rates only for Lack of Efficacy Category

Study (%)	ORG	active	Placebo	Outcome
ORG134001	3.9		3.8	positive
FK-GBE-007	3.2		2.4	positive
ORG134002	2.7		2.8	negative
FK-GBE-008	1.0		1.9	negative
ORG134023	3.9		5.5	negative
CN105-053	21	1.9	50	negative
ORG134004	3.7	2.9	2.9	negative
ORG134006	6.1	2.8	4.7	negative
ORG134017	5.5	2.4	4.3	negative
CN105-078	1.1		10.2	failed
CN105-083	4.3		4.9	failed
CN105-052	15		21	failed

- HAMD-25 is the More Appropriate Measure of Efficacy in the MDD-AF Population

The sponsor claims that the patient population enrolled in studies ORG134004 and ORG134006 are significantly different from the participants in the rest of the studies and as such, HAMD-25 (not HAMD-17) is the appropriate efficacy measure.

As seen in Fig. 1, there is a similar distribution of HAMD-25 total scores, HAMD-17 total scores, the sum of the 8 items missing in the HAMD-17 scale (compared with the HAMD-25), and the sum of the 5 items from the HAMD-25 that measure atypical features in both positive studies (134001 and FKGBE007), which enrolled all patients with MDD, and in studies 134004 and 134006, which enrolled patients with atypical depression. These values are also comparable among treatment groups in all four studies. In this reviewer's opinion, this shows that the patient populations in all four studies are comparable. It follows that any of the depression scales commonly used in clinical trials (i.e. HAMD-17, HAMD-21, MADRS) should be sensitive to show the effect of an antidepressant agent.

In this regard, it is interesting to note that studies 134001 and FKGBE007 still showed positive results when HAMD-25 was used as the endpoint, with drug placebo differences of -3.38 and -2.8 and p-values equal to 0.007 and 0.029, respectively (Table 21 and 22). Furthermore, the treatment differences were even larger with HAMD-25 than those with HAMD-17 (drug-placebo difference -2.47 and -2.45 and p-values of 0.013 and 0.018 for Studies 134001 and FKGBE007, respectively).The same could be said when the HAMD-21 and HAMD-28 scales were used as endpoints. In this reviewer's opinion this is evidence that any of the above mentioned depression scales are sensitive to show drug effect, therefore invalidating the sponsor's argument on the choice of rating scales.

Clinical Review
Silvana Borges, M.D.
NDA 21164/Informal appeal
Travivo® (gepirone hydrochloride ER tablets)

Figure 1: Baseline Values Comparison by Treatment Group for ITT Population

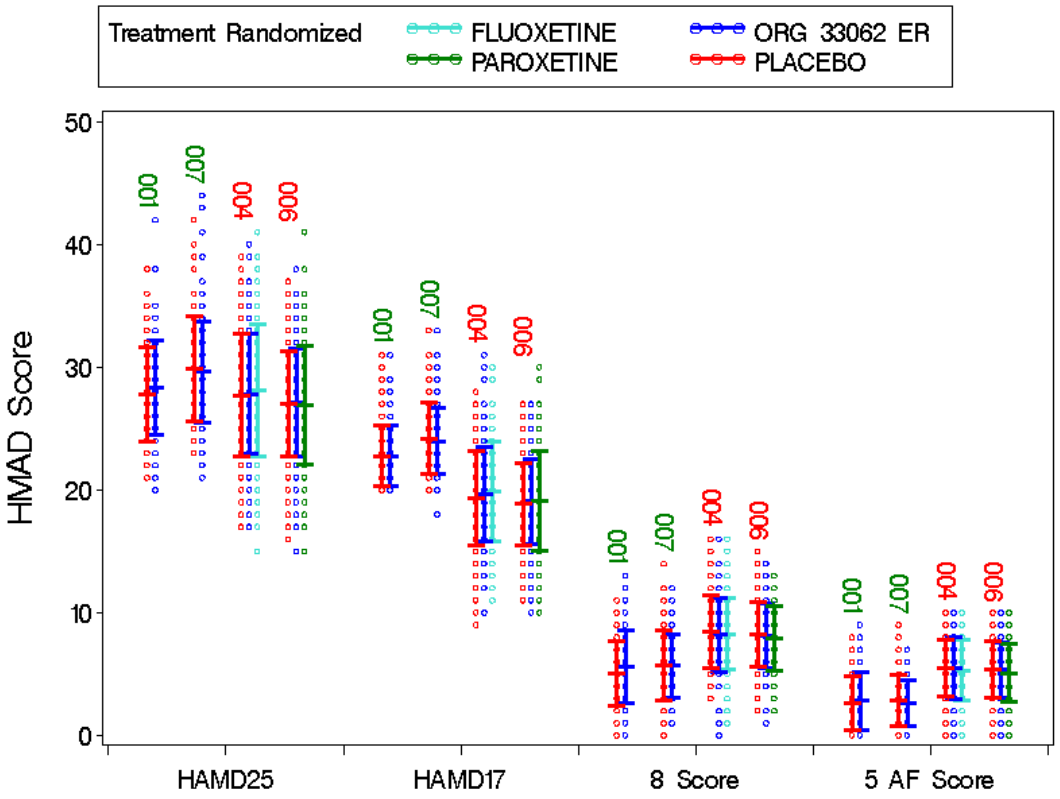


Table 21: Sponsor's secondary efficacy results at endpoint: study ORG134001 (positive trial)

Parameter	End of Treatment Outcome		Difference	p-value
	gepirone-ER	Placebo		
HAMD-17 Responders (%)	43.6%	30.7%	12.9%	0.059
HAMD-17 Remitters (%)	28.7%	14.9%	13.8%	0.017
HAMD-21 CFB	-10.01	-7.49	-2.51	0.021
HAMD-28 CFB	-13.27	-9.60	-3.68	0.013
MADRS CFB	-12.28	-9.22	-3.06	0.023
HAMD-Item 1 CFB	-1.16	-0.78	0.39	0.005
CGI Responders (%)	43.6%	35.6%	7.92	0.251
CGI severity CFB	-1.19	-0.79	-0.39	0.016
CGI improvement	2.82	3.10	0.29	0.072
HAMD Factor 1 CFB	-2.67	-2.08	0.58	0.124
HAMD-Item 12 CFB	-0.92	-0.49	0.43	0.001
HAMD-25 CFB	-11.57	-8.19	-3.38	0.007
HAMD-25 Responders (%)	45.5%	28.7%	16.8%	0.014

[*Source: 134001 CSR Table 21, Appendix F8.6.2*]
CFB=Change from baseline; LS means from ANOVA model, including effects for treatment and center; % responders and % remitters were compared using CMH test. For HAMD and MADRS scales, responders are subjects with ≥ 50% reduction from baseline; for CGI, responders are much or very much improved on the CGI improvement score.

Table 22: Sponsor's secondary efficacy results at endpoint: study FK-GBE-007 (positive trial)

Parameter	Week8/End of Treatment		Difference	p-value
	gepirone-ER	Placebo		
HAMD-17 responders (%)	45.7%	29.5%	16.2%	**0.014**
HAMD-17 Remitters (%)	34.5%	20.5%	14%	**0.019**
HAMD-21 CFB	-11.07	-8.79	-2.28	**0.043**
HAMD-21 responders (%)	46.6%	32.0%	14.6%	**0.031**
HAMD-25 CFB	-12.65	-9.85	-2.80	**0.029**
HAMD-25 responders (%)	48.3%	30.3%	18.0%	**0.007**
HAMD-28 CFB	-15.04	-11.83	-3.21	**0.032**
HAMD-28 responders (%)	49.1%	32.8%	16.3%	**0.015**
CGI severity CFB	-1.30	-0.92	-0.38	**0.015**
CGI responders (%)	48.3%	34.7%	13.6%	**0.045**
HAMD Item 1 CFB	-1.22	-0.97	-0.32	0.101
MADRS CFB	-13.72	-9.94	-3.78	**0.008**
MADRS responders (%)	50.9%	32.2%	18.7%	**0.005**

[*Source: FK-GBE-007 CSR Table 22, Appendix 15 Supportive Tables 22, 29, 30, 43, 44, 51, 52, 59, 60, 67, 73,74, 103*]
CFB=Change from baseline; LS means are from ANOVA model, including effects for treatment and center; % responders and % remitters were compared using CMH test. For HAMD and MADRS scales, responders are subjects with ≥ 50% reduction from baseline at any post-baseline assessment; HAMD-17 Remitters are subjects with a HAMD-17 total score of ≤ 7. For CGI, responders are much or very much improved on the CGI improvement score at any post-baseline assessment.

In addition, the treatment effects for HAMD-25 are 0.87 and 0.06 and p-values are 0.416 and 0.953 for studies 134004 and 134006, respectively. Using HAMD-17 (as shown on Table 4), the treatment effects are not larger but p-values are smaller. Therefore, contrary to the sponsor's assertion, the HAMD-17 total score appears to be a more sensitive measure to detect the difference between the study drug and placebo than the HAMD-25 total score in studies 134004 and 134006.

In this reviewer's opinion, all these data support the notion that the patient populations in all four studies (134001, FKGBE007, 134004, and 134006) are comparable and that HAMD-17 is a valid and sensitive measure to evaluate efficacy in all these trials.

- Use of a Comparator with Unknown Efficacy in Atypical Depression/Inappropriate Use of the Comparator

The sponsor argues that fluoxetine and paroxetine have not been thoroughly studied in patients with atypical depression and that the use of a comparator with unknown efficacy in the target population limits the value of the study to judge the efficacy of gepirone-ER in that population.

In this reviewer's opinion, if we consider the sponsor's argument to be true and conclude that studies 134004 and 134006 included an "inactive control", it follows that these studies had no control and that gepirone ER should be tested against placebo alone, making these trials conclusively negative.

It is also of note that this argument clearly contradicts the sponsor's first argument of assay sensitivity.

- Different Population: low Severity of Depression-Variable Severity Criterion

As seen in Table 23, based on HAMD-17 total scores, patients in studies 134004 and 134006 indeed had on average smaller baseline values than those in the two positive studies (134001 and FKGBE007). However, these differences are much smaller when HAMD-25 total scores are considered (Table 24). In this scenario, it is difficult to judge the relevance of the difference in baseline scores among these studies.

Table 23: Baseline HAMD-17 Total Scores (mean and SD)

Study 134001		Study FKGBE007		Study 134004		Study134006	
Gepirone ER	22.7 (2.5)	Gepirone ER	23.9 (2.7)	Gepirone ER	19.6 (3.8)	Gepirone ER	19.0 (3.5)
Placebo	22.8 (2.5)	Placebo	24.2 (2.9)	Placebo	19.3 (3.8)	Placebo	18.8 (3.4)

Table 24: Baseline HAMD-25 Total Scores (mean and SD)

Study 134001		Study FKGBE007		Study 134004		Study134006	
Gepirone ER	28.3 (3.9)	Gepirone ER	29.5 (4.2)	Gepirone ER	27.9 (4.9)	Gepirone ER	27 (4.4)
Placebo	27.8 (3.8)	Placebo	29.7 (4.3)	Placebo	27.6 (5.0)	Placebo	26.9 (4.3)

- Significant Treatment by Site Interaction

The sponsor argues that there is a significant treatment by site interaction for the HAMD-17 analysis at endpoint, week 8 ($p = 0.050$), indicating that the gepirone-ER's effect on this variable was not consistent across sites. Of the 10 sites, two favored gepirone-ER over fluoxetine (site 2 and site 4); and 8 favored fluoxetine over gepirone-ER (sites 1, 3, 5, 6, 7, 8, 9, and 10).

According to Dr. Chen's review, the inconsistency in the findings among sites has been observed in several short-term studies, including the two positive studies. As shown in the table 5 and 6, for Study ORG134001, the final positive results for all centers appeared to be driven by Center 1, and for Study FK-GBE-007, the final positive results for all centers appeared to be driven by Center 701 and Center 706. Also, as mentioned before, some centers even had a larger improvement in placebo patients than in drug patients. In this reviewer's opinion, if the site interaction issue is an argument to consider studies 134004 and 134006 as failed trials, it should also be an argument to consider studies ORG134001 and FK-GBE-007 (the two positive studies) as negative trials.

- Reasons for Trends in HAMD Favoring Placebo Over Gepirone-ER

For study ORG134004, the sponsor noted that the observed trend in HAMD favoring placebo over gepirone-ER were affected not only by the placebo effect but also by the dropout rate. They further stated that "As a result, trends favoring placebo are exaggerated by the use of LOCF analyses, which carries forward final values for drop-outs. By contrast, numerical differences between gepirone-ER and placebo are negligible based on the Observed Case (OC) analysis, which does not impute values from prior visits for drop-outs. At Week 8, for example, the LS mean reduction from baseline in HAMD-25 (based on OC analysis) was 11.3 in the gepirone-ER group compared to 11.4 on placebo."

To further assess whether the observed difference between gepirone-ER and placebo is indeed exaggerated by using the LOCF data, Dr. Chen performed the analysis by using the mixed-effect model for repeated measures (MMRM) with and without patients' baseline values as a covariate for HAMD-25 total scores and also HAMD-17 total Scores. The LOCF analysis showed a difference between gepirone-ER and placebo of about 0.9 for both HAMD-25 and HAMD-17. The differences obtained by the MMRM are indeed smaller compared with the LOCF analysis, and fluoxetine performed better than placebo regardless of the model or measure considered. In particular, we noted that for HAMD-17, fluoxetine beats gepirone-ER statistically significantly whether the baseline HAMD-17 is included or not in the model (Table 25). Again, these results do not support the sponsor's argument.

Table 25: FDA MMRM Analysis Results for Trial ORG134004

MMRM with Baseline covariate	HAMD 25 Total Scores	HAMD 17 Total Scores
Org33062 LS mean (SE)	-10.52 (0.83)	-6.24 (0.62)
Fluoxetine LS mean (SE)	-12.13 (0.77)	-6.92 (0.59)
Placebo LS mean (SE)	-11.04 (0.79)	-7.90 (0.57)
Org33062 minus Placebo LS mean (SE) & p-value	0.52 (1.14) 0.65	0.68 (0.85) 0.43
Fluoxetine minus Placebo LS mean (SE) & p-value	-1.09 (1.10) 0.32	-0.98 (0.82) 0.23
Org33062 minus Fluoxetine LS mean (SE) & p-value	1.61 (1.13) 0.15	1.65 (0.84) 0.05
MMRM w/o Baseline covariate	HAMD 25 Total Scores	HAMD 17 Total Scores
Org33062 LS mean (SE)	-10.45 (0.85)	-6.20 (0.64)
Fluoxetine LS mean (SE)	-12.18 (0.79)	-6.82 (0.61)

Placebo LS mean (SE)	-10.96 (0.81)	-7.97 (0.59)
Org33062 minus Placebo LS mean (p-value)	0.51 (1.17) 0.67	0.63 (0.88) 0.48
Fluoxetine minus Placebo LS mean (p-value)	-1.22 (1.12) 0.28	-1.14 (0.85) 0.18
Org33062 minus Fluoxetine LS mean (p-value)	1.73 (1.16) 0.14	1.77 (0.87) 0.04

For all the reasons expressed above and in agreement with previous clinical and statistical reviews, this reviewer considers studies ORG134004 and ORG134006 to be negative gepirone ER trials.

5.1.2.5 Study ORG134017 (2002)

This was an 8-week, randomized, double-blind, flexible-dose, placebo and active-controlled trial of the efficacy and safety of gepirone ER (40 mg to 80 mg) for the treatment of MDD (active control: fluoxetine 20 mg to 40 mg): gepirone ER (n=159); placebo (n=159).

➢ Inclusion criteria

- Males or females, 18-65 years of age with a diagnosis of MDD per DSM-IV
- Dysphoria for most days over the past 4 weeks
- HAM-D 17 total score ≥18 (amended to ≥22 during study) at baseline

➢ Primary efficacy measure**:** change from baseline in MADRS total score at endpoint (Week 8 or last visit).

➢ Results:

The sponsor's arguments for considering trial ORG134017 as a failed study (and this reviewer's interpretation of the data) are as follows:

- No Assay Sensitivity

Although both gepirone ER and fluoxetine failed to beat placebo on the primary endpoint (MADRS), as seen on Table 26, the FDA analysis using HAMD-17 (Table 4) showed a statistically significant difference favoring fluoxetine over gepirone-ER (-1.54, p=0.042).

It is also of note that fluoxetine consistently performed better than placebo, regardless of the efficacy measure. That is not the case for gepirone ER, which performed worse than placebo on the primary endpoint and on several secondary measures. This pattern

in which the directional trends show gepirone ER performing worse than placebo in several studies further increase our doubts on gepirone ER efficacy as an antidepressant.

Table 26. Sponsor's MADRS change from baseline per visit.

Treatment Group		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
gepirone-ER (N=160)	n	159	159	159	159	159	159	159
	Mean	29.5	-3.86	-7.78	-9.46	-11.56	-12.29	-12.23
	SE	0.37	0.50	0.61	0.70	0.74	0.80	0.84
placebo (N=159)	n	159	159	159	159	159	159	159
	Mean	29.6	-4.94	-7.35	-9.60	-11.23	-12.47	-12.73
	SE	0.41	0.50	0.61	0.69	0.73	0.79	0.84
fluoxetine (N=161)	n	159	159	159	159	159	159	159
	Mean	29.1	-4.59	-7.54	-9.66	-12.27	-12.56	-13.88
	SE	0.41	0.50	0.61	0.70	0.74	0.80	0.84
Treatment (Overall)	p-value		0.244	0.865	0.975	0.547	0.965	0.310
gep-ER vs. placebo	p-value		0.100	0.591	0.879	0.730	0.860	0.650
fluoxetine vs. placebo	p-value		0.600	0.809	0.947	0.282	0.934	0.299
gep-ER vs. fluoxetine	p-value		0.263	0.768	0.827	0.465	0.795	0.136

LS means and p-values from ANOVA model, with treatment and site as main effects; treatment x site interactions were not significant (p > 0.10).
ET = end of trial; LS = least squares; SE = standard error of the mean
[*Source: CRS ORG134017 Table 15, Appendix F Table 6.1-1.3*]

Table 27. Sponsor's secondary efficacy variables at endpoint.

Efficacy Variable	End of Treatment Outcome			Pairwise Tests (p-values)*		
	Gep-ER (N=160)	Fluoxetine (N=161)	Placebo (N=159)	G vs. P	F vs. P	G vs. F
HAMD-17 CFB	-10.23	-11.76	-10.96	--	--	--
HAMD-25 CFB	-12.03	-13.77	-12.86	--	--	--
HAMD-Item1 CFB	-1.33	-1.50	-1.24	--	--	--
CGI (severity) CFB	-1.24	-1.41	-1.32	--	--	--
CGI (global improvement)	2.48	2.29	2.44	--	--	--
% Responders (HAMD-17)	42.14%	57.23%	45.28%	0.607	NR	0.006
% Responders (CGI)	54.09%	62.26%	52.83%	0.829	NR	0.152
% Remitters (HAMD-17)	22.01%	32.08%	31.45%	0.060	NR	0.044
HAMD (items 22-24) hypersomnia	-0.59	-0.43	-0.65	--	--	--
HAMD (items 25-26) hyperphagia	-0.10	-0.38	-0.06	0.742	NR	0.037

[*Source: CSR ORG134017 Tables 16-25*]
*P-values and LS means from ANOVA for continuous variables, with treatment and center as the main effects; -- indicates that a pairwise test was not performed because the overall treatment effect was not statistically significant. CFB=Change from Baseline.
CMH test was used for % responders (gepirone-ER vs each other group). NR = Not Reported; this was not a protocol-defined statistical comparison.

- Inconsistency Among Sites

As mentioned for studies ORG134004 and ORG134006, this argument could also be used to consider the two positive trials as negative.

- High Placebo Response

As seen in Tables 19 and 20, the overall dropout rate and the dropout rate by lack of efficacy in study ORG134017 is higher in the gepirone ER group than in the placebo group. That is not the case for fluoxetine.

The sponsor states that the percentage of HAMD-17 responders in the placebo group was 45.28%, an arguably high rate. However, the HAMD-17 responders rate in the fluoxetine group was 57.23%, while it was 42.14% in the gepirone ER group. Furthermore, the HAMD-17 responder and remitter rates were significantly higher for fluoxetine than for gepirone-ER (p=0.006 and p=0.044, respectively). Therefore, even though the placebo response could be considered high in a study without an active control arm, the fact that the active control performed better than placebo on the primary endpoint and most secondary variables (while gepirone ER performed worse than placebo) invalidates the argument.

- Positive Results from Reliable Investigators

The sponsor examined data from two sites conducted by investigators (Arif Khan MD and Nick Vitakis MD) who, in their own words, “they know and respect”. Placebo response rates were only 36% and 12% at these sites, and mean differences in HAMD-17 scores showed favorable results for both active drugs: gepirone-ER vs. placebo -4.6 points, p=0.001; fluoxetine vs. placebo -4.8 points p=0.001.

In this reviewer’s opinion, picking some sites and discarding others based on sponsor’s preference of investigators is not a valid approach for the evaluation of efficacy.

- Flaws in Study Conduct

The sponsor states that study ORG134017 was poorly conducted trial with protocol compliance issues, citing that after 9 months the HAMD-17 entry criterion was changed from ≥18 to ≥22 and that, as a result, the studied included a substantial number of subject with less severe depression. However, per the sponsor’s own account, only 20% of the patients enrolled in the study had a baseline HAMD-17 score below 22. In this reviewer’s opinion, 20% does not constitute a substantial number of patients.

In addition, it is important to consider that protocol deviations occur in all clinical trials and, as mentioned before, this study with its possible variability in the patients’ baseline HAMD-17 scores was good enough to show the active control performing consistently better than placebo in many variables (while gepirone ER performed worse that placebo). This study was also good enough to detect a difference between treatment arms (i.e. the active control beats gepirone ER).

The sponsor does not present any other piece of evidence to support their claim of flaws in study conduct. In this reviewer's opinion, this claim has no basis in the study data.

- Spurious Trends Favoring Placebo

In elaborating on why gepirone ER was numerically worse than placebo in this study, the sponsor argues that it can be explained by:

- High placebo response (this reviewers refers to the same argument above).
- Low severity of depression in enrolled subjects (this reviewer refers to the same argument under "Flaws in study conduct" above).
- High dropout rate in the gepirone ER group: as shown in Table 19, the dropout rate for all reasons in the gepirone-ER group was 32%, compared to 24% and 21% in the fluoxetine and placebo groups, respectively. The sponsor cites this as favoring placebo in study results. As mentioned before, despite these differences, the active control performed better than placebo while gepirone ER did not. In addition, it is of note that the dropout rate for lack of efficacy was 5.5% for gepirone ER, 2.4% for fluoxetine, and 4.3% for placebo, another variable in which gepirone ER performed worse than placebo while the active control did better.

For all the reasons expressed above and in agreement with previous clinical and statistical reviews, this reviewer considers study ORG134017 to be a negative gepirone ER trial.

5.1.2.6 Study CN105053 (1991)

This was an 8-week, randomized, double-blind, flexible-dose, placebo and active-controlled trial of the efficacy and safety of gepirone ER (10 mg to 60 mg) for the treatment of MDD (active control: imipramine 50 mg to 200 mg): gepirone ER (n=56); placebo (n=56).

- Inclusion criteria

- Males or females, >18 years of age with a diagnosis of MDD per DSM-III-R
- HAMD-17 total score ≥20 at screening and baseline.

- Primary efficacy measure**:** change from baseline in HAMD-17 total score at endpoint (Week 8 or last visit).

➢ Results:

Table 28. Sponsor's HAMD-17 Change from baseline per visit

Treatment		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
Feiger (Site 001)								
gepirone-ER	n	41	40	41	41	41	41	41
	Mean	23.7	-4.7	-5.5	-7.2	-8.2	-10.8	-10.1
	SE	0.52	0.7	0.9	1.0	1.0	1.1	1.2
imipramine	n	39	39	39	39	39	39	39
	Mean	23.6	-4.5	-7.0	-9.3	-9.4	-11.8	-10.9
	SE	0.48	0.7	0.9	1.0	1.0	1.2	1.2
placebo	n	40	38	40	40	40	40	40
	Mean	23.9	-3.9	-4.9	-6.8	-6.0	-7.7	-6.8
	SE	0.46	0.7	0.9	1.0	1.0	1.2	1.2
gepirone-ER vs pbo	p-value		0.455	0.649	0.746	0.129	0.063	0.049
imipramine vs. pbo	p-value		0.572	0.096	0.092	0.024	0.014	0.017
Gelenberg (Site 002)								
gepirone-ER	n	15	15	15	15	15	15	15
	Mean	24.6	-3.1	-7.3	-9.0	-9.1	-10.3	-9.3
	SE	0.62	1.3	1.5	1.7	2.0	2.5	2.5
imipramine	n	15	15	15	15	15	15	15
	Mean	25.7	-3.1	-7.5	-7.9	-10.9	-10.9	-12.2
	SE	0.54	1.3	1.5	1.7	2.0	2.5	2.5
placebo	n	16	15	16	16	16	16	15
	Mean	25.0	-2.8	-4.7	-6.4	-9.1	-10.2	-11.2
	SE	0.75	1.3	1.1	1.7	2.0	2.4	2.5
gepirone-ER vs. pbo	p-value		0.837	0.226	0.289	0.998	0.966	0.589
imipramine vs. pbo	p-value		0.837	0.182	0.534	0.527	0.843	0.776
Combined Sites								
gepirone-ER	n	56	55	56	56	56	56	56
	Mean	23.9	-3.9	-6.4	-8.1	-8.7	-10.5	-9.7
	SE	0.41	0.7	0.8	1.0	1.0	1.2	1.2
imipramine	n	54	54	54	54	54	54	54
	Mean	24.2	-3.8	-7.3	-8.6	-10.2	-11.3	-11.5
	SE	0.39	0.7	0.9	1.0	1.1	1.2	1.2
placebo	n	56	53	56	56	56	56	56
	Mean	24.2	-3.3	-4.8	-6.6	-7.6	-8.9	-9.0
	SE	0.39	0.7	0.8	1.0	1.0	1.2	1.2
gepirone-ER vs. pbo	p-value		0.568	0.195	0.274	0.448	0.343	0.687
imipramine vs. pbo	p-value		0.633	0.041	0.153	0.080	0.160	0.144

LS Means and p-values for combined sites is based on ANOVA with effects for treatment, center and treatment by center interaction.
[*Source: CSR CN105-053 Table 17, Appendix F Tables 7.1.1-3 and 7.1.1-6*]

The sponsor's arguments for considering trial CN105053 as a failed study (and this reviewer's interpretation of the data) are as follows:

➢ Early Termination

The sponsor states BMS terminated the program after 170 subjects were randomized (71% of the required number), reducing the power to approximately 63%. However, as seen in Table 4, the active control was better than placebo on the primary endpoint (-3.19, p=0.038), thus showing that the study was sufficiently powered to detect a difference between treatment arms. In addition, per the sponsor's own analysis shown in Tables 28, the active control was consistently better than placebo and gepirone ER

on the secondary variables, in some cases even reaching statistical significance over placebo. That was not the case for gepirone ER.

Table 29. Sponsor's secondary variables analysis

Efficacy Variable	End of Treatment Outcome			P-values	
	Gepirone-ER	Imipramine	Placebo	G vs P	I vs P
Feiger	N=41	N=39	N=40		
HAMD-28 CFB	-17.3	-17.6	-10.4	**0.006**	**0.005**
HAMD-25 CFB	-14.4	-14.6	-8.9	**0.007**	**0.006**
HAMD-Item1 CFB	-1.2	-1.4	-0.6	**0.010**	**0.001**
HAMD-17 Responders	44%	49%	33%	0.294	0.145
CGI (severity) CFB	-1.5	-1.7	-0.9	**0.031**	**0.004**
CGI (global improvement)	2.4	2.1	3.1	**0.012**	**0.001**
MADRS CFB	-11.9	-14.3	-8.7	0.179	**0.020**
Gelenberg	N=15	N=15	N=16		
HAMD-28 CFB	-13.5	-17.3	-14.8	0.798	0.614
HAMD-25 CFB	-11.3	-14.3	-12.9	0.709	0.724
HAMD-Item1 CFB	-0.7	-1.1	-1.3	0.163	0.550
HAMD-17 Responders	40%	67%	50%	0.582	0.355
CGI (severity) CFB	-1.1	-1.8	-1.4	0.529	0.340
CGI (global improvement)	2.5	2.4	2.7	0.718	0.638
MADRS CFB	-12.7	-16.5	-15.9	0.502	0.900
Combined	N=56	N=54	N=56		
HAMD-28 CFB	-15.4	-17.5	-12.6	0.266	0.055
HAMD-25 CFB	-12.9	-14.5	-10.9	0.330	0.084
HAMD-Item1 CFB	-1.0	-1.2	-0.9	0.929	0.200
HAMD-17 Responders	43%	54%	38%	0.551	0.084
CGI (severity) CFB	-1.3	-1.8	-1.1	0.535	**0.021**
CGI (global improvement)	2.4	2.2	2.9	0.110	**0.031**
MADRS CFB	-12.3	-15.4	-12.3	0.987	0.197

[*Source: CN105-053 Final Report Appendices F7.1, 7.2, and 7.4*]
LS means and p-values from ANOVA, with treatment and center as factors; CMH test for % responders

- By Protocol, the two sites should not be pooled

The sponsor argues that when the study was terminated, the Feiger site had exceeded enrollment (n=123), but only 47 of the 120 required subjects (39%) had been randomized at the Gelenberg site. This argument about site interaction has been addressed previously.

This reviewer acknowledges that, in isolation, study CN105053 could be interpreted as a weak piece of evidence of gepirone ER lack of efficacy. However, the results in this study are consistent with many other studies in not supporting the efficacy of gepirone ER for the treatment of MDD. In this reviewer's opinion, study CN105-053 simply adds evidence (weak or not) against gepirone ER.

- Inappropriate FDA Analysis

As seen on Table 4, the FDA analysis found that the mean reduction in HAMD-17 at endpoint was significantly greater for imipramine compared to placebo (p = 0.038), but not for gepirone-ER (p = 0.190). The sponsor questions the accuracy of these results. This reviewer has no comments in this regard.

5.1.2.7 Study ORG28709 (1999)

This was a maintenance trial with an 8-12-week OL phase (gepirone ER 20-80 mg) followed by a 40-44-week randomized, double-blind, placebo-controlled phase: gepirone ER (n=126); placebo (n=124).

- Inclusion criteria

- Males or females, 18-70 years of age with a diagnosis of recurrent MDD by DSM-IV criteria.
- HAMD-17 total score ≥20 at screening and baseline.

- Response criteria

- HAMD-17 total score ≤8 within an 8-12 week window.

- Relapse criteria

- HAMD-17 total score ≥16 at any office visit, or
- Discontinuation due to 'Relapse criteria fulfilled' on the End of Trial form.

- Primary efficacy measure**:** relapse rate (time to relapse was a secondary measure).

- Results:

Although the sponsor initially reported this study as a positive trial, the sponsor's efficacy analysis was flawed by the fact that 5 subjects who were obvious gepirone-ER relapses were not counted and 32 subjects were deleted from the database because they came from centers that had only 1 treatment arm represented, or had no relapses. After adding these patients back into the analysis, Gepirone ER did not statistically significantly reduced the rate of relapse over placebo in the ITT population (Table 30).

The sponsor currently agrees with the FDA approach to the efficacy analysis and does not dispute that the study results do not support gepirone ER efficacy.

However, the sponsor has the following comments about the design and conduct of study ORG28709:

- Investigators did not fully understand the protocol or the primary endpoint

The sponsor argues that a significant number of protocol violations and subjects continuing treatment after meeting criteria for relapse are evidence of investigators not understanding the study protocol. However, the sponsor does not provide any data to support this argument.

In this reviewer's opinion, the experience of investigators in clinical trials varies and so does their performance in trial conduct. However, since not all sites or studies are audited, it is not possible to assess the level of understanding investigators have of study protocols. That is the case for every clinical trial we evaluate. Study ORG28709 is not different than other clinical trials in that regard.

- A high proportion of subjects received CNS drugs during the double-blind period, which can influence HAMD-17 ratings.

Again, the sponsor presents this argument as pure speculation without any data to support their belief.

- Response criteria to qualify for randomization were not clearly defined and confirmed during the open-label period.

This reviewer disagrees with the sponsor in that the response criteria were not clearly defined. On the contrary, there was only one response criterion (HAMD-17 total score ≤8 within an 8-12 week window), quite simple compared with the more complex criteria used in many of the maintenance trials with antidepressants submitted to the FDA in the last 25 years.

- Post hoc analyses restricted to qualified, protocol-compliant subjects show positive results for gepirone-ER.

For Study ORG28709, the sponsor included Table 62 in this submission trying to demonstrate that if they have had correctly identified the correct patients (i.e. true responders or relapsers by the sponsor's post-hoc redefinition), then the study would have been positive. The statistical reviewer has tried to confirm the sponsor's analysis results and found that in their analyses, the 5 patients who had relapsed but their relapse events were not counted and data of around 30 patients (depending on the type of analyses, the numbers varied a bit) coming from the centers which had only single treatment arm, or had no relapses were again removed from the analysis due to the use of CMH method with centers as strata. After adding the 5 patients' events and the data of the aforementioned type of patients in the analysis by combining all the single armed

centers, the statistical reviewer had extremely different results. Table 30 shows both the sponsor's analysis and the statistical reviewer's results.

Table 30. Results for the Primary Analysis and Sponsor-Proposed Sensitivity Analyses

	Sponsor's Analysis Results			FDA Analysis Results		
	Gepirone-ER	Placebo	p-value	Gepirone-ER	Placebo	p-value
Original ITT	29/126 (0.23)	43/124 (0.35)	0.024	34/126 (0.27)	43/124(0.35)	0.36
Per Protocol	25/104 (0.24)	41/106 (0.39)	0.023	25/104 (0.24)	40/106(0.37)	0.11
Re-Defined Non-Responders(1)	22/126 (0.18)	40/124 (.32)	0.007	26/118 (0.22)	40/121(0.33)	0.17
Re-Defined Non-Responders(2)	22/126 (0.18)	42/124 (0.34)	0.003	26/118 (0.22)	42/123(0.34)	0.10
Re-Defined Non-Responders(3)	25/126 (0.20)	42/124 (0.34)	0.013	29/121 (0.24)	42/123(0.34)	0.25

(1) Excludes relapses on 1st visit after randomization, i.e., 11 patients were removed.
(2) Excludes relapses on 1st visit after randomization if response was confirmed prior to randomization, i.e., 9 patients were removed.
(3) Includes subjects with 50% drop in HAMD-17 prior to randomization as responders, i.e., 6 patients were removed from the analysis.

Therefore, study ORG28709 shows negative results even when redefining response and relapse criteria.

- Post hoc analyses do not prove that this study shows efficacy for gepirone-ER. However, they do show that had the study been conducted properly, the results would have been positive for gepirone-ER.

Again, this is pure speculation. The data show this study as a negative one. In our review of all maintenance trials with antidepressants submitted to the FDA in the last 25 years, we have found that response and relapse criteria varied greatly among trials, as did the response stabilization period and the length of the OL and DB phases. However, almost every study succeeded in showing superiority of drug over placebo, regardless of the efficacy measure, whether they use stricter or looser definitions of response or relapse, or whether they had patients stabilized in their response before randomization.

In our view, study ORG28709 was an adequately designed trial, with an adequate number of patients enrolled, with response and relapse criteria similar to those used in other maintenance studies with antidepressants, and with a sufficient number of relapse

events (higher than in many maintenance trials with antidepressants) to detect a difference between treatment arms. Every single antidepressant maintenance trial with these characteristics submitted to the FDA since the approval of the first second-generation antidepressant has shown positive results. It is in this context that the negative results of this maintenance trial with gepirone ER stand out.

5.1.3 Failed studies

The FDA has agreed with the sponsor that the following studies can be considered failed studies and do not provide enough evidence in favor or against gepirone ER for the treatment of MDD. Since there is agreement on their lack of merit, only a brief summary of these studies are presented below.

5.1.3.1 Study CN105-052 (1991)

This was an 8-week, randomized, double-blind, flexible-dose, placebo and active-controlled trial of the efficacy and safety of gepirone ER (20 mg to 60 mg) for the treatment of MDD (active control: fluoxetine 20 mg): gepirone ER (n=35); placebo (n=37).

Table 31: Sponsor's HAMD-17 Total Score for Change from Baseline at End of Treatment by Week

Treatment		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
Gep-ER	n	35	34	35	35	35	35	35
	Mean	25.5	-2.6	-6.0	-8.4	-8.9	-10.0	-11.0
	SE	0.54	0.7	0.9	1.1	1.2	1.4	1.5
Fluoxetine	n	36	33	36	36	36	36	36
	Mean	25.2	-2.4	-4.4	-7.6	-9.2	-10.7	-11.0
	SE	0.43	0.7	0.9	1.0	1.2	1.3	1.5
Placebo	n	37	37	37	37	37	37	37
	Mean	25.2	-2.7	-5.6	-7.7	-9.2	-9.7	-10.5
	SE	0.54	0.7	0.9	1.0	1.2	1.3	1.4
Gepirone-ER vs. Pbo	p-value		0.928	0.736	0.660	0.880	0.905	0.825
Fluoxetine vs. Pbo	p-value		0.744	0.322	0.934	0.996	0.600	0.798

LS Means and p-values for combined sites is based on ANOVA with effects for treatment, center and treatment by center interaction.
[Source: *CSR CN105-052 Table 16, Appendix F Tables 7.1.1-1 and 7.1.1-3*]

5.1.3.2 Study CN105-078 (1991)

This was a 6-week, randomized, double-blind, flexible-dose, placebo-controlled, three-arm trial of the efficacy and safety of gepirone ER (10 mg to 100 mg) for the treatment of MDD: gepirone ER 10-50 mg (n=48); gepirone ER 20-100 mg (n=40); placebo (n=47).

Table 32: Sponsor's HAMD-17 Total Score for Change from Baseline at End of Treatment by Week

Treatment		Baseline	LS Mean Change from Baseline				
			Week 1	Week 2	Week 3	Week 4	Week 6/ET
Combined Sites							
Gep-ER 10-50	n	48	46	48	48	48	48
	Mean	22.7	-2.3	-4.9	-6.4	-7.3	-7.5
	SE	0.42	0.5	0.7	0.8	0.8	0.9
Gep-ER 20-100	n	40	37	40	40	40	40
	Mean	21.9	-2.8	-4.8	-5.9	-7.4	-7.5
	SE	0.28	0.5	0.8	0.9	0.9	1.0
Gep-ER Total	n	88	83	88	88	88	88
	Mean	22.3	-2.5	-4.8	-6.1	-7.2	-7.4
	SE	0.27	0.3	0.5	0.6	0.6	0.7
Placebo	n	47	46	47	47	47	47
	Mean	22.9	-3.1	-5.1	-6.5	-6.5	-6.5
	SE	0.35	0.5	0.7	0.8	0.9	0.9
Gep 10-50 vs Pbo	p-value		0.256	0.830	0.931	0.488	0.460
Gep 20-100 vs Pbo	p-value		0.717	0.807	0.667	0.506	0.473
Gep Total vs. Pbo	p-value		0.362	0.759	0.702	0.514	0.451

[*Source: CSR CN105-078 Table 23 and Appendix 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*]
P-values and LS Means from ANOVA model with factors for treatment and center, including the interaction term (combined sites), or one-way ANOVA model with a factor for treatment (single site analysis).

5.1.3.3 Study CN105-083 (1991)

This was a 6-week, randomized, double-blind, flexible-dose, placebo-controlled, three-arm trial of the efficacy and safety of gepirone ER (10 mg to 100 mg) for the treatment of MDD: gepirone ER 10-50 mg (n=36); gepirone ER 20-100 mg (n=37); placebo (n=39).

Table 33: Sponsor's HAMD-17 Total Score for Change from Baseline at End of Treatment by Week

Treatment		Baseline	LS Mean Change from Baseline				
			Week 1	Week 2	Week 3	Week 4	Week 6/ET
Combined Sites							
Gep-ER 10-50	n	36	36	36	36	36	36
	Mean	24.8	-4.0	-7.5	-8.6	-10.3	-9.8
	SE	0.81	0.8	1.0	1.1	1.1	1.3
Gep-ER 20-100	n	37	36	37	37	37	37
	Mean	23.1	-4.6	-7.6	-8.4	-9.0	-9.2
	SE	0.33	0.8	1.0	1.1	1.1	1.3
Gep-ER Total	n	73	72	73	73	73	73
	Mean	23.9	-4.3	-7.5	-8.5	-9.6	-9.4
	SE	0.40	0.5	0.7	0.8	0.8	0.9
Placebo	n	39	38	39	39	39	39
	Mean	24.0	-4.3	-7.3	-9.0	-8.7	-8.9
	SE	0.56	0.8	1.0	1.1	1.1	1.3
Gep 10-50 vs Pbo	p-value		0.792	0.920	0.784	0.311	0.646
Gep 20-100 vs Pbo	p-value		0.764	0.846	0.684	0.830	0.902
Gep Total vs. Pbo	p-value		0.998	0.864	0.691	0.484	0.743

[*Source: CSR CN105-083 Table 23 and Appendix F Tables 7.1.1-1, 7.1.1-3, 7.1.1-1A, and 7.1.1-3A*]
P-values and LS Means from ANOVA model with factors for treatment and center, including the interaction term (combined sites), or one-way ANOVA model with a factor for treatment (single site analysis).

5.2 Conclusions Regarding the Efficacy Claim of gepirone ER for the Treatment of MDD

The sponsor presented 12 short-term clinical trials and 1 maintenance trial with gepirone ER, of which only 2 short-term studies (ORG134001 and FK-GBE-007) support the efficacy of gepirone ER for the treatment of MDD.

Three of the 13 studies (CN105052, CN105078, and CN105083) were not informative in the evaluation of the efficacy of gepirone ER for the treatment of MDD and were not considered any further.

Seven short-term trials (ORG134002, FKGBE008, ORG134023, ORG134004, ORG134006, CN134017 and CN105053) showed no difference between gepirone ER and placebo. Four of these 7 studies included an active-control arm (ORG134004, ORG134006, CN134017 and CN105053), in which the active controls performed consistently better than gepirone ER and placebo, reaching statistical significance over placebo in study CN105053, over gepirone ER in studies ORG134004 and CN134017, and over gepirone ER and placebo in study ORG134006. In our decades-long experience with antidepressant development programs, we have found very few trials in which an effective antidepressant drug shows no effect while the active control does.

Another puzzling finding is that, in 4 of the 7 negative trials (ORG134023, ORG134004, ORG134006, and CN134017), gepirone ER performed worse than placebo on the primary endpoint and on many secondary variables. Again, in our experience, this is a very infrequent scenario with effective antidepressant drugs.

Finally, the negative maintenance gepirone ER trial (ORG28709) is an important piece of evidence against the efficacy of gepirone ER for the treatment of MDD. Study ORG28709 was an adequately designed trial, with an adequate number of patients enrolled, with response and relapse criteria similar to those used in other maintenance studies with antidepressants, and with a sufficient number of relapse events to detect a difference between treatment arms. In our review of all maintenance trials with antidepressants submitted to the FDA since the approval of the first second-generation antidepressant, every single maintenance trial with these characteristics has shown positive results. In this context, the negative results of this maintenance trial with gepirone ER are difficult to ignore.

In conclusion, although two short-term trials favor gepirone ER for the treatment of MDD, seven negative short-term studies and one maintenance trial with gepirone ER provide compelling evidence against the efficacy of gepirone ER as an antidepressant.

Reference List

1. Celada P, Bortolozzi A, Artigas F. Serotonin 5-HT Receptors as Targets for Agents to Treat Psychiatric Disorders: Rationale and Current Status of Research. CNS Drugs 2013;

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

/s/

SILVANA BORGES
04/18/2014

5.13 Advice Letter –Robert Temple (04/18/2014)

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration
Silver Spring MD 20993

NDA 21164

GENERAL ADVICE

Fabre-Kramer Pharmaceuticals, Inc.
Attention: Martin Lobel, Esq.
Attorney
Law Offices of Lobel, Novins & Lamont, LLP
888 17th Street, N.W., Suite 810
Washington, D.C. 20006

Dear Mr. Lobel:

Please refer to your New Drug Application (NDA) submitted under section 505(b) of the Federal Food, Drug, and Cosmetic Act for gepirone hydrochloride extended release tablets (gepirone ER).

We also refer to the following: 1) not approvable (NA) letter dated November 2, 2007; 2) meeting minutes dated January 21, 2008; 3) meeting minutes dated December 28, 2011; and 4) December 10, 2012 amendment, containing information for an informal review of gepirone ER efficacy data along with your current arguments in support of its efficacy.

The major deficiency cited in the November 2, 2007 NA letter was a failure to provide substantial evidence of efficacy in the short-term and longer-term treatment of major depressive disorder (MDD). Although the letter noted that the available evidence suggested that gepirone appeared to be less effective than other available antidepressants, relative efficacy was not the basis for the NA decision. The NA action was based on the lack of substantial evidence of effectiveness.

A face-to-face meeting was held on January 14, 2008 to discuss your responses to the November 2, 2007 NA letter and we concluded that it was highly unlikely any additional analyses of the existing database would justify further review of the NDA. On April 27, 2011 you requested reconsideration of the 2007 non-approval decision. Another face-to-face meeting was held on November 29, 2011 to discuss the statistical report provided by your consultant, Mary F. Johnson. You submitted the report to support your request that we reconsider our non-approval decision conveyed in the November 2, 2007 NA letter.

On December 10, 2012, following meetings, communications, and exchanges of documents, you submitted an NDA amendment providing information for an informal review of the gepirone ER efficacy data along with your current arguments in support of its efficacy. The efficacy data contained in the submission were the same as those reviewed in 2007 in their entirety.

We have reviewed the referenced material dated December 10, 2012 and have the following comments:

You presented 12 short-term clinical studies (ORG134001, FKGBE007, ORG134023, CN105052, CN105078, CN105083, ORG134002, FKGBE008, ORG134004, ORG134006, CN134017 and CN105053) and 1 maintenance study (ORG28709) with gepirone ER. All of these studies have been previously reviewed with your 2007 submission. No new clinical data were submitted at this time. Of these 12 short-term studies:

- 2 studies (ORG134001 and FKGBE007), as stated in your submission, support the efficacy of gepirone ER for the treatment of MDD.
- Another study, ORG134023, is a negative gepirone ER trial, as you acknowledged in your submission.
- 3 other studies (CN105052, CN105078, and CN105083), which you considered failed studies in your submission, were probably not informative in the evaluation of the efficacy of gepirone ER for the treatment of MDD, largely because of their overall size, and were not considered further. Only Study CN105052 had an active control showing no effect, representing stronger evidence of a failed study.

The remaining 6 short-term studies (ORG134002, FKGBE008, ORG134004, ORG134006, CN134017 and CN105053) showed no statistically significant difference between gepirone ER and placebo on the primary endpoint or on HAMD-17 in the 3 studies with a primary endpoint of HAMD-25 (ORG 134004, 134006) or MADRS (CN134017). In 4 of those studies an active control did show a significant effect on HAMD-17 compared to placebo, gepirone, or both. You presented several arguments as to why, in your view, these studies should not be considered negative studies and were either supportive of gepirone ER efficacy or failed studies. Your arguments and our responses to them follow:

- High Placebo Response

You assert that the negative results of studies ORG134002, ORG134004, ORG134006, and ORG134017 are due in part to the high response in the placebo group. Such substantial responses in the placebo group are common in acute depression trials and no doubt contribute to the high failure rate with these trials. But the responses in the placebo groups with these trials are not unusually high and did not appear related to success or failure. The failure rate of gepirone, however, exceeds the rate observed for any approved drug.

In study 134002 the change from baseline in the HAMD-17 total score in the placebo group was about -9 points. Similar values were seen in study CN105053 (about -8 points in the HAMD-17). In study CN105053, however, the active control showed superiority over placebo despite the placebo group response, but gepirone ER did not. At the same time, gepirone ER showed a statistically significant effect in the positive study FKGBE007, despite a placebo group response of about -8 points in the HAMD-17.

In studies ORG134004 and ORG134006, the placebo group response (about -7 points in the HAMD-17 for both) was similar to that observed in the positive trials (ORG134001 and FKGBE007, about -7 and -8 points in the HAMD-17, respectively). However, it is worth noting that the active controls (fluoxetine and paroxetine) were consistently better numerically than placebo and were shown significantly superior to gepirone on the HAMD-17.

Study ORG134017 had a large placebo group response, with a 45% rate of HAMD-17 responders in the placebo group. Nonetheless, the HAMD-17 responder rate in the fluoxetine group was 57%, while it was only 42% in the gepirone ER group, and fluoxetine was significantly superior to gepirone for the HAMD-17 total score. The trial was thus able to distinguish fluoxetine from a less effective treatment (gepirone) despite the high placebo group responder rate.

- No Assay Sensitivity

We acknowledge that in studies ORG134004, ORG134006, and ORG134017, the active control failed to reach statistical significance over placebo on the primary endpoint (HAMD-25 for studies ORG134004 and ORG134006; MADRS for study ORG134017). However, the treatment effect favored the active controls on the HAMD-17, the primary endpoint for most of your controlled trials, by showing superiority to placebo and gepirone in ORG 134006 and to gepirone in ORG 134004 and ORG134017. The superiority to gepirone was made possible by gepirone's inferiority to placebo.

You have commented on our reliance on what you considered an unusual definition of assay sensitivity. Specifically, you argued that superiority of the active control to gepirone ER is not evidence of assay sensitivity. That is incorrect. Finding a statistically significant difference between two treatment arms shows that the study was able to detect a difference between effective and ineffective treatments, which is the essence of assay sensitivity. It is also in our experience very unusual to see statistically significant superiority of the active control to the test drug, and this is a worrisome finding.

You have also argued against relying on a non-protocol specified endpoint to justify a conclusion of assay sensitivity, and we acknowledge some concern here with multiple comparisons. We are, however, dealing with an extraordinarily low study success rate in what appear to be well-controlled studies (i.e. 7 of 9). We recognize that depression trials of effective drugs have failure/negative rates of about 50% and believe that active controls can be informative as to whether it is the drug or the study that failed. Superiority of the active control to placebo (CN 105053 and ORG 134006) and/or gepirone (ORG 134004, ORG 134006, ORG 134017) was observed in all 4 trials of adequate size with a comparator, an outcome very far from what we have seen with approved drugs. We utilized HAMD-17, a most widely used efficacy endpoint and the endpoint in 9 of your 12 controlled trials, to compare the effect of gepirone ER across studies.

- Inconsistency among Sites

You argue that, in studies ORG134004, ORG134006, and ORG134017, the gepirone ER effect was inconsistent across sites, with some sites favoring gepirone ER over the active control and others favoring the active control over gepirone ER.

In general, it is not surprising to observe inconsistent results across sites if the overall treatment effect is relatively small. Even in the positive trial FK-GBE-007, large variations in treatment effect (difference between gepirone ER and placebo) among sites were seen (p-value = 0.092 for the treatment-by-center interaction based on your own result). If we were to hold the

inconsistent findings across sites against those studies, the validity of the positive study FK-GBE-007 in support of gepirone ER efficacy would also become questionable.

Study CN105053 was conducted at two sites only. You consider this study to have failed for several reasons, including early termination at one of the sites, lower mean modal dose of gepirone ER, and higher placebo response at this early terminated site. This is of course possible but such after-the-fact explanations of study failure are rarely persuasive. Study CN10503 remains a negative study. The pooled data showed an effect of imipramine but not gepirone.

- Studies ORG134004 and ORG134006 Enrolled Atypical MDD Patient Population

You argue that the patient populations enrolled in studies ORG134004 and ORG134006, which had MDD with atypical features as an entry criterion, are significantly different from the participants in the rest of the studies and as such, HAMD-25 (not HAMD-17) is the appropriate efficacy measure.

However, we have found a similar distribution of HAMD-25 total scores, HAMD-17 total scores, the sum of the 8 items missing in the HAMD-17 total score (compared with the HAMD-25 total scores), and the sum of the 5 items from the HAMD-25 that measure atypical features in both positive studies (ORG134001 and FKGBE007), which enrolled all patients with MDD, and in studies ORG134004 and ORG134006, which enrolled patients with atypical depression. These values are also similar among treatment groups in all four studies. In our view, this shows that the patient populations in all four studies are comparable and that any of the depression rating scales commonly used in clinical trials (i.e. HAMD-17, HAMD-21, MADRS) would be able to differentiate an effective antidepressant agent from placebo.

In addition, using HAMD-17 as the primary endpoint for studies ORG134004 and ORG134006, the p-values for the gepirone ER-placebo comparison are in fact smaller than those obtained using HAMD-25. Therefore, the HAMD-17 total score seems to be at least as sensitive as the HAMD-25 total score at detecting a difference between gepirone ER and placebo in studies ORG134004 and ORG134006.

We also note that, in your own analysis, the positive trials (studies ORG134001 and FKGBE007), which had HAMD-17 as the primary endpoint, also showed positive results on the HAMD-25. In our view, this is further evidence that any of the above mentioned depression scales would be sensitive to showing a drug effect.

- Use of a Comparator with Unknown Efficacy in Atypical Depression/Inappropriate Use of the Comparator in Studies ORG134004 and ORG134006

You state that fluoxetine and paroxetine have not been thoroughly studied in patients with atypical depression and that the use of a comparator with unknown efficacy in the target population limits the value of the study to judge the efficacy of gepirone ER in that population. In fact however, in those studies, the two drugs were significantly superior to gepirone on a valid measure of depression.

- Studies ORG134002 and FKGBE008 Support the Efficacy of Gepirone ER

As you acknowledged in your submission, studies ORG134002 and FKGBE008 were adequately designed, properly conducted, and employed doses of gepirone ER in the correct therapeutic range, but these two studies did not show any difference between gepirone ER and placebo on the primary endpoints. Nonetheless, you interpreted these studies to be supportive of gepirone ER efficacy, stating that treatment effects consistently favored gepirone ER over placebo for each of the secondary efficacy variables. We acknowledge that the directional trend of the primary endpoint and the secondary variables favor gepirone ER in these studies. However, per your own analysis, gepirone ER did not reach statistical significance over placebo either on the primary endpoint or on almost every secondary variable. We continue to interpret these studies as negative gepirone ER trials.

In summary, 7 out of 9 short-term trials (ORG134002, FKGBE008, ORG134023, ORG134004, ORG134006, CN134017 and CN105053) showed no difference between gepirone ER and placebo. Four of these 7 studies included an active-control arm (ORG134004, ORG134006, CN134017 and CN105053), in which the active control performed statistically significantly better than gepirone ER or placebo on the HAMD-17 scale. Statistical significance was reached over placebo in study CN105053, over gepirone ER in studies ORG134004 and CN134017, and over both gepirone ER and placebo in study ORG134006, based on statistical models without the treatment-by-center interaction term, but where the treatment factor included all treatment arms. In our decades-long experience with antidepressant development programs, we have found few trials in which an effective antidepressant drug shows no effect while the active control does.

Another unusual finding is that, in 3 of the 7 negative trials (ORG134023, ORG134004, and CN134017), gepirone ER performed numerically worse than placebo on the primary endpoint and on many secondary variables. In our experience, this too is a very infrequent occurrence with effective antidepressant drugs.

The negative maintenance gepirone ER trial (ORG28709) is an additional piece of important evidence against the efficacy of gepirone ER for the treatment of MDD. Study ORG28709 was an adequately designed randomized withdrawal trial, with an adequate number of patients enrolled, with response and relapse criteria similar to those used in other maintenance studies with approved antidepressants, and with a sufficient number of relapse events to detect a difference between treatment arms. In our review of all maintenance trials with approved antidepressants, every single maintenance trial with these characteristics has shown positive results. In this context, the negative results of this maintenance trial with gepirone ER are difficult to ignore. You argued that not all patients randomized to the double-blind phase were "true" responders; hence, you re-analyzed data using different definitions of true responders. Although all of your re-analyses yielded significant p-values, we disagree with your results for the following reasons: (1) failure to count 5 patients that were gepirone ER relapses; (2) failure to include approximately 30 patients who came from centers that had only 1 treatment arm represented or had no relapses; (3) failure to remove all patients who should have been removed according to your various definitions of true responders. After these corrections were made, the p-values were no longer statistically significant. The negative findings with gepirone ER in a maintenance trial in patients with MDD are also worrisome from a public health point of view, since MDD is a chronic disorder which can lead to fatal outcomes.

In conclusion, although two short-term trials favor gepirone ER for the treatment of MDD, the seven negative short-term studies and one negative maintenance trial with gepirone ER raise considerable doubts about the effectiveness of gepirone in the acute or sustained treatment of depression. The 2 positive studies could represent chance findings, given the absent, negative, or minimal findings in 8 other studies.

We are amenable to meeting with you should you decide to continue the development program of gepirone for the treatment of MDD. If you have any questions, contact Hiren Patel, Pharm.D., Regulatory Project Manager, at hiren.patel@fda.hhs.gov or (301) 796-2087.

Sincerely,

{See appended electronic signature page}

Robert Temple, MD
Deputy Director
Office of Drug Evaluation I
Center for Drug Evaluation and Research

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

/s/

ROBERT TEMPLE
04/18/2014

5.14 Formal Dispute Resolution Meeting Minutes –Khushboo Sharma (03/18/2015)

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration
Silver Spring MD 20993

NDA 21164

MEETING MINUTES

Fabre-Kramer Pharmaceutical
5847 San Felipe
Suite 2000
Houston, TX 77057

Attention: Stephen J. Kramer
CEO

Dear Dr. Kramer:

Please refer to your New Drug Application (NDA) submitted under section 505(b) of the Federal Food, Drug, and Cosmetic Act for gepirone hydrochloride extended release tablets (gepirone ER).

We also refer to the meeting between representatives of your firm and the FDA on February 23, 2015. The purpose of the meeting was to discuss the issues raised in your request for formal dispute resolution received on January 27, 2015.

A copy of the official minutes of the meeting is enclosed for your information. Please notify us of any significant differences in understanding regarding the meeting outcomes.

If you have any questions, call me at (301) 796-1270.

Sincerely,

{See appended electronic signature page}

Khushboo Sharma, M.B.A, R.A.C
CDER Formal Dispute Resolution Project Manager
Center for Drug Evaluation and Research

Enclosure:
Meeting Minutes

MEMORANDUM OF MEETING MINUTES

Meeting Date and Time:	February 23, 2015 3:00-4:30pm EST
Meeting Location:	White Oak Campus, Building 22, Rm 1421
Application Number:	NDA 21164
Product Name:	gepirone hydrochloride extended release tablets
Sponsor/Applicant Name:	Fabre-Kramer Pharmaceuticals
Meeting Chair:	John Jenkins, M.D.
Meeting Recorder:	Khushboo Sharma, M.B.A, RAC

FDA ATTENDEES

Office of New Drugs

John Jenkins, M.D.	Director
Amy Bertha	Team Leader, Regulatory Affairs Team
Khushboo Sharma, M.B.A, RAC	CDER Formal Dispute Resolution Project Manager

Office of Drug Evaluation I (ODE I)

Ellis Unger, M.D.	Director
Robert Temple, M.D.	Deputy Director
Colleen Locicero	Associate Director for Regulatory Affairs

ODE I/Division of Psychiatry Products

Mitchell Mathis, M.D.	Director
Tiffany Farchione, M.D.	Deputy Director
Silvana Borges, M.D.	Medical Officer

Office of Biostatistics

Lisa LaVange, Ph.D.	Director
Peiling Yang, Ph.D.	Biometrics Team Leader
Hsein Ming Hung, Ph.D.	Biometrics Reviewer
Jinglin Zhong, Ph.D.	Biometrics Reviewer

Office of Center Director

Virginia Behr	CDER Ombudsman

Office of Chief Counsel

Linda Epstein, J.D.	Chief Counsel

SPONSOR ATTENDEES

Stephen Kramer	CEO, Fabre-Kramer Pharmaceuticals
Edward Koehler	EVP, Fabre-Kramer Pharmaceuticals
Gerald Masoudi	Partner, Covington & Burling
Lewis Grossman	Of Counsel, Covington & Burling

Michael Thase	Consultant
Mary Johnson	Consultant
Eve Stoffel	Paralegal, Covington & Burling
Colleen Kelly	Associate, Covington & Burling

BACKGROUND

Fabre-Kramer Pharmaceuticals submitted a request for formal dispute resolution (FDRR) to the Office of New Drugs, received on January 27, 2015, concerning the Not Approvable (NA) letter dated November 2, 2007, and the General Advice letter dated April 18, 2014. The request specifically appeals the Office of Drug Evaluation I's (ODE I) conclusion that Fabre-Kramer had not provided "substantial evidence" of gepirone ER's effectiveness (the other deficiencies communicated in the November 2, 2007 NA letter were not part of the appeal). Dr. John Jenkins, Director, OND is the deciding authority. In Fabre-Kramer's dispute resolution submission, the company requested a meeting with the deciding authority before he renders his decision on the matter. The meeting was granted and was held on February 23, 2015.

MEETING OBJECTIVES

The objective of this meeting was to discuss the issues surrounding the appeal.

DISCUSSION

The discussion between FDA and Fabre-Kramer focused on the following issues:

- There were 12 short-term studies to support the efficacy of geprione ER. The sponsor and the FDA are in agreement regarding the analysis for 8 of the 12 studies. Of these 8 studies, the FDA and Fabre Kramer agree that there were 2 positive studies (ORG134001, FKGBE007) that supported the efficacy of gepirone ER, 3 negative studies (ORG134023, ORG134002 and FKGBE008), and 3 failed studies (CN105052, CN105078 and CN105083) that were not informative largely because of the overall size. Two out of the three negative studies (ORG134002 and FKGBEE008) had a positive directional trend favoring gepirone ER; however, the analysis did not show statistical significance compared to placebo on the primary endpoint. The sponsor and FDA disagree on the interpretation of results from 4 of the 12 studies. The studies in dispute are: ORG134004, ORG134006, ORG134017 and CN105053. These 4 studies were 3-arm studies with gepirone ER, active control and placebo. HAMD25 was the predefined primary endpoint for studies ORG134004 and ORG134006, MADRS was the predefined primary endpoint for study ORG134017, and HAMD17 was the predefined primary endpoint for CN105053. The FDA performed post-hoc analysis on these 4 studies using HAMD17 as the primary endpoint. The FDA analysis showed a significant difference in HAMD17 between active control and either gepirone, placebo, or both favoring active control in all 4 studies, and therefore the FDA considers these studies to be negative studies. The sponsor argues that the analysis on the 4 studies in dispute showed no assay sensitivity on the predefined primary endpoint, and 2 studies (ORG134004 and ORG134017) did not show assay sensitivity on the post-hoc analysis using HAMD17 as

the primary endpoint; therefore, Fabre-Kramer believes these should be considered failed studies.

- Fabre-Kramer presented the summary of the ratio of positive to negative short-term studies for gepirone ER. According to the FDA, the ratio is 2 positive to 7 negative studies, and the probability that the findings from the two positive studies are due to chance is 2.0%. The FDA argued that in considering whether there is substantial evidence of effectiveness for gepirone ER it is difficult to ignore 7 negative studies against 2 positive studies. The sponsor stated, however, that 4 out of the 7 studies identified as negative were deemed negative based on flawed post-hoc analyses and should not be considered negative. According to the sponsor, the ratio is 2 positive to 3 negative studies, which shows there is 0.59% likelihood that positive findings were due to chance. The sponsor asserted that the division recently approved antidepressants such as Celexa, Cymbalta, and Pristiq with more negative studies than positive studies.

- The FDA stated that the negative long-term maintenance study is important evidence against the efficacy of gepirone ER. The sponsor asserted that the FDA should not consider a long-term maintenance study when evaluating the substantial evidence of efficacy for an indication in short-term treatment of major depressive disorder (MDD). The sponsor noted that the FDA approved Fetzima in 2013 with a negative long-term study, and that Fabre-Kramer has repeatedly made the same argument for the gepirone ER long-term maintenance study. FDA stated that we would look into the administrative record for the approval of Fetzima.

- FDA asked if gepirone ER is approved anywhere in the world. Fabre-Kramer stated that the drug was not approved anywhere in the world, and they have recently initiated the process for submitting a marketing application for gepirone ER to the European Union.

- FDA stated that Fabre-Kramer has offered in the past to conduct another long term maintenance study for gepirone ER, but noted that they have not yet begun such a study. FDA asked why that was the case. Fabre-Kramer replied that they do not have the funding to begin another pre-marketing study for gepirone ER.

- FDA suggested that an advisory committee (AC) meeting may be beneficial to help evaluate the substantial evidence of efficacy for gepirone ER, given the disagreement on the analysis of the short-term studies. FDA stated that as part of the formal dispute resolution process, the deciding authority could request an AC meeting, if the deciding authority needed additional input in order to reach a final decision. FDA asked if Fabre-Kramer would be open to discussion of these issues at an AC meeting. Fabre-Kramer was not prepared to respond to that question. In general however, Fabre-Kramer believes that an advisory committee is not necessary because they have shown substantial evidence of efficacy based on 2 positive adequate and well-controlled trials. Also, Faber-Kramer asserts that the safety profile of Gepirone ER is as favorable as other approved antidepressants with fewer sexual side effects and lower risks than other products; therefore, an advisory committee meeting is not necessary.

DECISION (AGREEMENTS) REACHED:

This meeting was not conducted with the expectation that decisions would be made or agreements would be reached at the meeting. The issues discussed will be taken into careful consideration when reaching a decision regarding the formal dispute resolution request.

ATTACHMENTS/HANDOUTS:

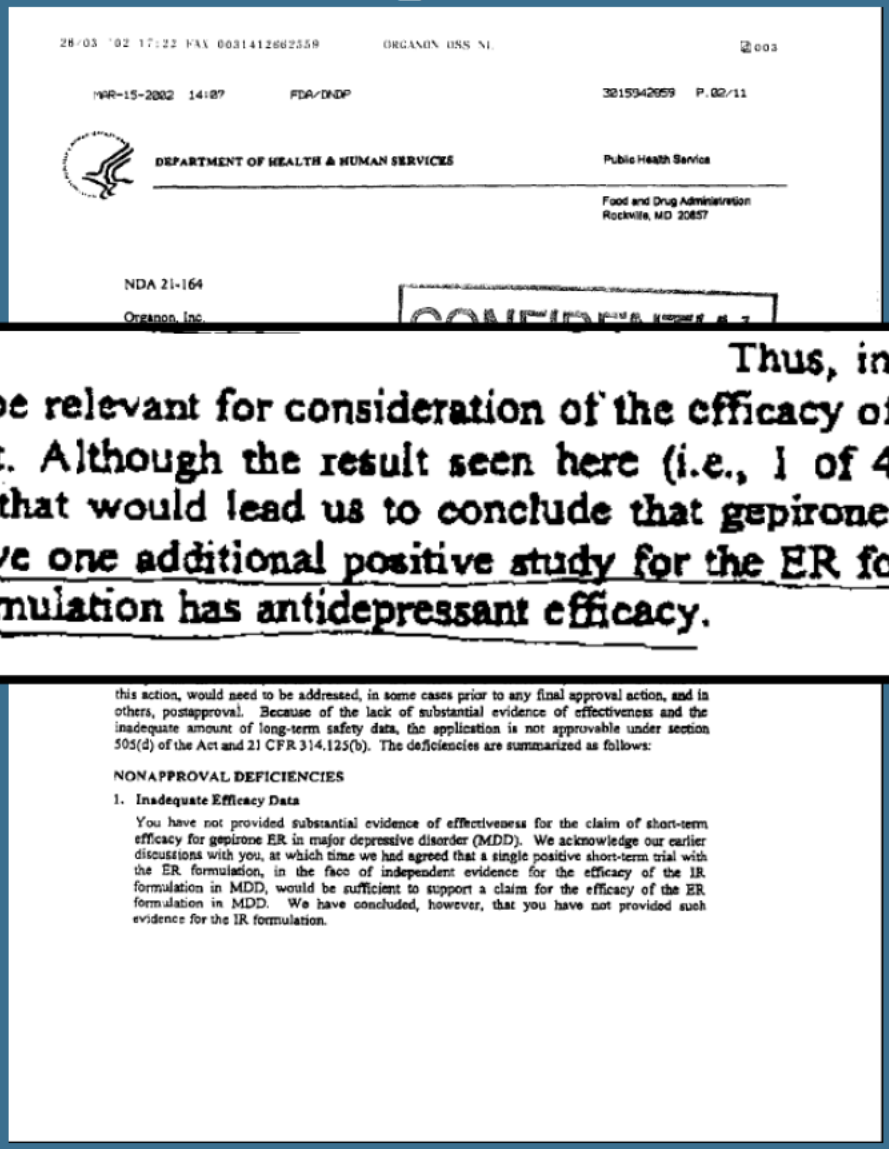
Handout from Fabre-Kramer's presentation.

Fabre-Kramer Pharmaceuticals, Inc.

Appeal before Dr. John Jenkins
Gepirone ER Efficacy
February 23, 2015

Regulatory History

2002 NA Letter: One Additional Study Required



28/03 '02 17:22 FAX 0031412662559 ORGANON OSS NL 003

MAR-15-2002 14:07 FDA/DNDP 3215942659 P.02/11

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

Food and Drug Administration
Rockville, MD 20857

NDA 21-164

Organon, Inc.

Thus, in summary, of 4 studies that appear, on face, to be relevant for consideration of the efficacy of gepirone ER, only one yielded a positive result. Although the result seen here (i.e., 1 of 4 relevant studies being positive) is not a result that would lead us to conclude that gepirone ER is ineffective as an antidepressant, we believe one additional positive study for the ER formulation is needed to demonstrate that this formulation has antidepressant efficacy.

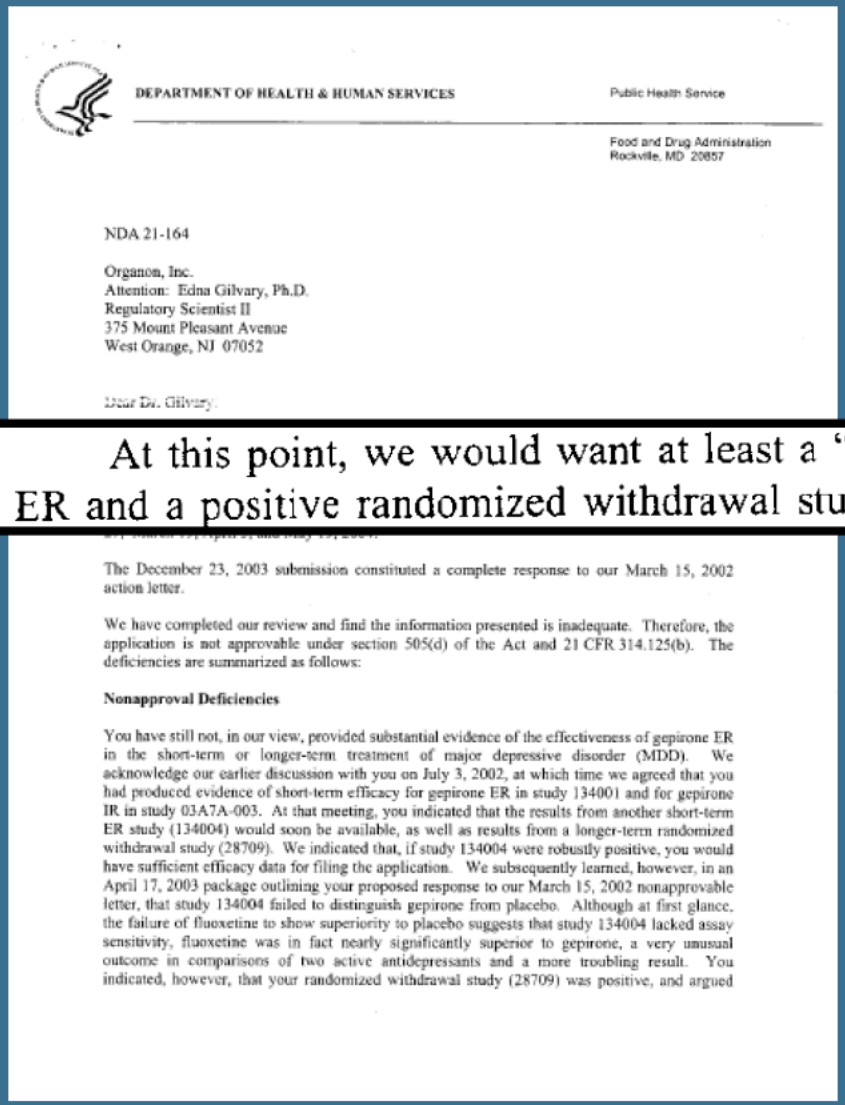
this action, would need to be addressed, in some cases prior to any final approval action, and in others, postapproval. Because of the lack of substantial evidence of effectiveness and the inadequate amount of long-term safety data, the application is not approvable under section 505(d) of the Act and 21 CFR 314.125(b). The deficiencies are summarized as follows:

NONAPPROVAL DEFICIENCIES

1. **Inadequate Efficacy Data**

You have not provided substantial evidence of effectiveness for the claim of short-term efficacy for gepirone ER in major depressive disorder (MDD). We acknowledge our earlier discussions with you, at which time we had agreed that a single positive short-term trial with the ER formulation, in the face of independent evidence for the efficacy of the IR formulation in MDD, would be sufficient to support a claim for the efficacy of the ER formulation in MDD. We have concluded, however, that you have not provided such evidence for the IR formulation.

2004 NA Letter: Requirement of Long-Term Study Added



DEPARTMENT OF HEALTH & HUMAN SERVICES — Public Health Service

Food and Drug Administration
Rockville, MD 20857

NDA 21-164

Organon, Inc.
Attention: Edna Gilvary, Ph.D.
Regulatory Scientist II
375 Mount Pleasant Avenue
West Orange, NJ 07052

Dear Dr. Gilvary:

At this point, we would want at least a "robustly positive" short-term trial with gepirone ER and a positive randomized withdrawal study.

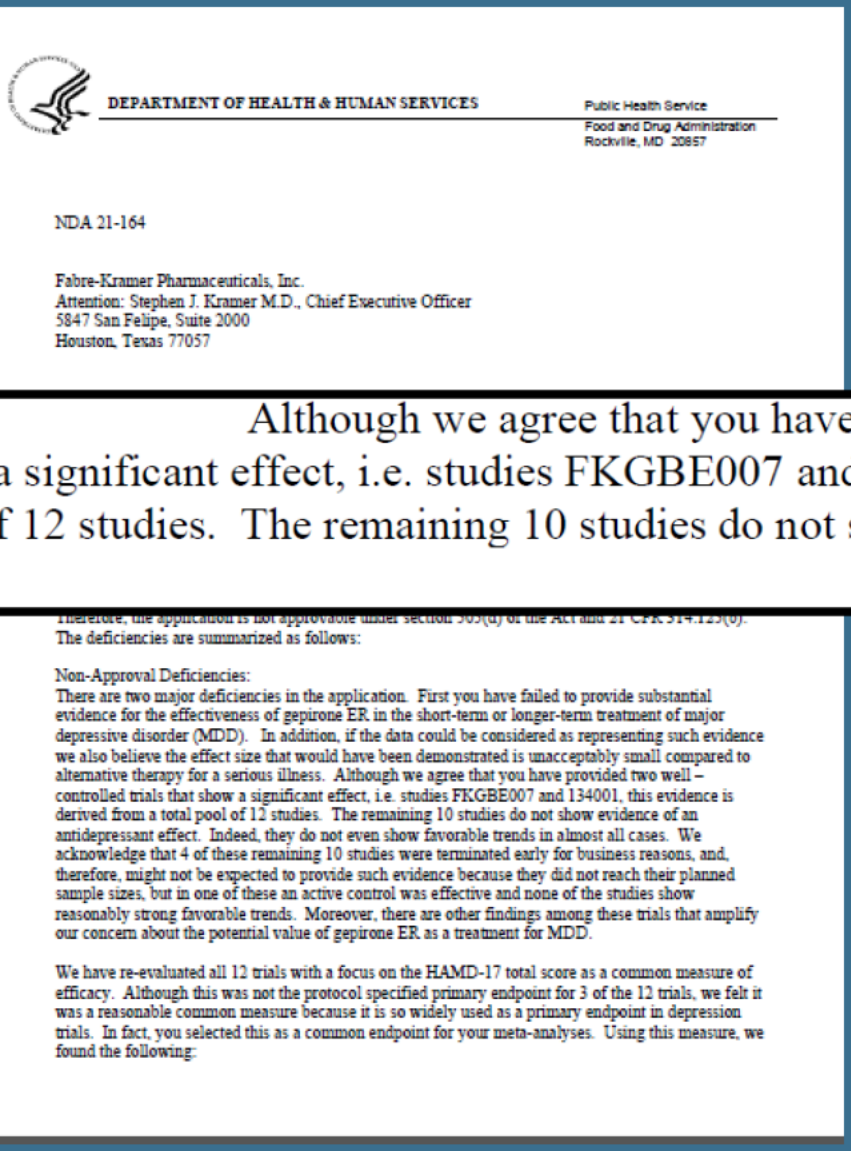
The December 23, 2003 submission constituted a complete response to our March 15, 2002 action letter.

We have completed our review and find the information presented is inadequate. Therefore, the application is not approvable under section 505(d) of the Act and 21 CFR 314.125(b). The deficiencies are summarized as follows:

Nonapproval Deficiencies

You have still not, in our view, provided substantial evidence of the effectiveness of gepirone ER in the short-term or longer-term treatment of major depressive disorder (MDD). We acknowledge our earlier discussion with you on July 3, 2002, at which time we agreed that you had produced evidence of short-term efficacy for gepirone ER in study 134001 and for gepirone IR in study 03A7A-003. At that meeting, you indicated that the results from another short-term ER study (134004) would soon be available, as well as results from a longer-term randomized withdrawal study (28709). We indicated that, if study 134004 were robustly positive, you would have sufficient efficacy data for filing the application. We subsequently learned, however, in an April 17, 2003 package outlining your proposed response to our March 15, 2002 nonapprovable letter, that study 134004 failed to distinguish gepirone from placebo. Although at first glance, the failure of fluoxetine to show superiority to placebo suggests that study 134004 lacked assay sensitivity, fluoxetine was in fact nearly significantly superior to gepirone, a very unusual outcome in comparisons of two active antidepressants and a more troubling result. You indicated, however, that your randomized withdrawal study (28709) was positive, and argued

2007 NA Letter: Acknowledges 2 Positive Trials



DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service
Food and Drug Administration
Rockville, MD 20857

NDA 21-164

Fabre-Kramer Pharmaceuticals, Inc.
Attention: Stephen J. Kramer M.D., Chief Executive Officer
5847 San Felipe, Suite 2000
Houston, Texas 77057

Although we agree that you have provided two well – controlled trials that show a significant effect, i.e. studies FKGBE007 and 134001, this evidence is derived from a total pool of 12 studies. The remaining 10 studies do not show evidence of an antidepressant effect.

Therefore, the application is not approvable under section 505(d) of the Act and 21 CFR 314.125(b).
The deficiencies are summarized as follows:

Non-Approval Deficiencies:
There are two major deficiencies in the application. First you have failed to provide substantial evidence for the effectiveness of gepirone ER in the short-term or longer-term treatment of major depressive disorder (MDD). In addition, if the data could be considered as representing such evidence we also believe the effect size that would have been demonstrated is unacceptably small compared to alternative therapy for a serious illness. Although we agree that you have provided two well – controlled trials that show a significant effect, i.e. studies FKGBE007 and 134001, this evidence is derived from a total pool of 12 studies. The remaining 10 studies do not show evidence of an antidepressant effect. Indeed, they do not even show favorable trends in almost all cases. We acknowledge that 4 of these remaining 10 studies were terminated early for business reasons, and, therefore, might not be expected to provide such evidence because they did not reach their planned sample sizes, but in one of these an active control was effective and none of the studies show reasonably strong favorable trends. Moreover, there are other findings among these trials that amplify our concern about the potential value of gepirone ER as a treatment for MDD.

We have re-evaluated all 12 trials with a focus on the HAMD-17 total score as a common measure of efficacy. Although this was not the protocol specified primary endpoint for 3 of the 12 trials, we felt it was a reasonable common measure because it is so widely used as a primary endpoint in depression trials. In fact, you selected this as a common endpoint for your meta-analyses. Using this measure, we found the following:

2007 NA Letter: Reliance on Comparative Efficacy

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service
Food and Drug Administration
Rockville, MD 20857

NDA 21-164

Fabre-Kramer Pharmaceuticals, Inc.
Attention: Stephen J. Kramer M.D., Chief Executive Officer
5847 San Felipe, Suite 2000
Houston, Texas 77057

There are two major deficiencies in the application. First you have failed to provide substantial evidence for the effectiveness of gepirone ER in the short-term or longer-term treatment of major depressive disorder (MDD). In addition, if the data could be considered as representing such evidence we also believe the effect size that would have been demonstrated is unacceptably small compared to alternative therapy for a serious illness.

Non-Approval Deficiencies:
There are two major deficiencies in the application. First you have failed to provide substantial evidence for the effectiveness of gepirone ER in the short-term or longer-term treatment of major depressive disorder (MDD). In addition, if the data could be considered as representing such evidence we also believe the effect size that would have been demonstrated is unacceptably small compared to alternative therapy for a serious illness. Although we agree that you have provided two well – controlled trials that show a significant effect, i.e. studies FKGBE007 and 134001, this evidence is derived from a total pool of 12 studies. The remaining 10 studies do not show evidence of an antidepressant effect. Indeed, they do not even show favorable trends in almost all cases. We acknowledge that 4 of these remaining 10 studies were terminated early for business reasons, and, therefore, might not be expected to provide such evidence because they did not reach their planned sample sizes, but in one of these an active control was effective and none of the studies show reasonably strong favorable trends. Moreover, there are other findings among these trials that amplify our concern about the potential value of gepirone ER as a treatment for MDD.

We have re-evaluated all 12 trials with a focus on the HAMD-17 total score as a common measure of efficacy. Although this was not the protocol specified primary endpoint for 3 of the 12 trials, we felt it was a reasonable common measure because it is so widely used as a primary endpoint in depression trials. In fact, you selected this as a common endpoint for your meta-analyses. Using this measure, we found the following:

2007 NA Letter: Post-Hoc Endpoint Analysis

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service
Food and Drug Administration
Rockville, MD 20857

NDA 21-164

Fabre-Kramer Pharmaceuticals, Inc.
Attention: Stephen J. Kramer M.D., Chief Executive Officer
5847 San Felipe, Suite 2000
Houston, Texas 77057

Dear Dr. Kramer:

We have re-evaluated all 12 trials with a focus on the HAMD-17 total score as a common measure of efficacy. Although this was not the protocol specified primary endpoint for 3 of the 12 trials, we felt it was a reasonable common measure

We have completed our review of your resubmission and find the information presented is inadequate. Therefore, the application is not approvable under section 505(d) of the Act and 21 CFR 314.125(b). The deficiencies are summarized as follows:

Non-Approval Deficiencies:
There are two major deficiencies in the application. First you have failed to provide substantial evidence for the effectiveness of gepirone ER in the short-term or longer-term treatment of major depressive disorder (MDD). In addition, if the data could be considered as representing such evidence we also believe the effect size that would have been demonstrated is unacceptably small compared to alternative therapy for a serious illness. Although we agree that you have provided two well – controlled trials that show a significant effect, i.e. studies FKGBE007 and 134001, this evidence is derived from a total pool of 12 studies. The remaining 10 studies do not show evidence of an antidepressant effect. Indeed, they do not even show favorable trends in almost all cases. We acknowledge that 4 of these remaining 10 studies were terminated early for business reasons, and, therefore, might not be expected to provide such evidence because they did not reach their planned sample sizes, but in one of these an active control was effective and none of the studies show reasonably strong favorable trends. Moreover, there are other findings among these trials that amplify our concern about the potential value of gepirone ER as a treatment for MDD.

We have re-evaluated all 12 trials with a focus on the HAMD-17 total score as a common measure of efficacy. Although this was not the protocol specified primary endpoint for 3 of the 12 trials, we felt it was a reasonable common measure because it is so widely used as a primary endpoint in depression trials. In fact, you selected this as a common endpoint for your meta-analyses. Using this measure, we found the following:

2007 NA Letter: Reliance on Long-Term Study Despite 2005 PDAC Decision

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service
Food and Drug Administration
Rockville, MD 20857

NDA 21-164

Fabre-Kramer Pharmaceuticals, Inc.
Attention: Stephen J. Kramer M.D., Chief Executive Officer
5847 San Felipe, Suite 2000
Houston, Texas 77057

Dear Dr. Kramer:

Please refer to your new drug application dated September 29, 1999 received October 1, 1999

The negative outcome for the longer-term maintenance efficacy trial (study 28709) is also a concern for this drug.

We have completed our review of your resubmission and find the information presented is inadequate. Therefore, the application is not approvable under section 505(d) of the Act and 21 CFR 314.125(b). The deficiencies are summarized as follows:

Non-Approval Deficiencies:
There are two major deficiencies in the application. First you have failed to provide substantial evidence for the effectiveness of gepirone ER in the short-term or longer-term treatment of major depressive disorder (MDD). In addition, if the data could be considered as representing such evidence we also believe the effect size that would have been demonstrated is unacceptably small compared to alternative therapy for a serious illness. Although we agree that you have provided two well – controlled trials that show a significant effect, i.e. studies FKGBE007 and 134001, this evidence is derived from a total pool of 12 studies. The remaining 10 studies do not show evidence of an antidepressant effect. Indeed, they do not even show favorable trends in almost all cases. We acknowledge that 4 of these remaining 10 studies were terminated early for business reasons, and, therefore, might not be expected to provide such evidence because they did not reach their planned sample sizes, but in one of these an active control was effective and none of the studies show reasonably strong favorable trends. Moreover, there are other findings among these trials that amplify our concern about the potential value of gepirone ER as a treatment for MDD.

We have re-evaluated all 12 trials with a focus on the HAMD-17 total score as a common measure of efficacy. Although this was not the protocol specified primary endpoint for 3 of the 12 trials, we felt it was a reasonable common measure because it is so widely used as a primary endpoint in depression trials. In fact, you selected this as a common endpoint for your meta-analyses. Using this measure, we found the following:

October 2011 Preliminary Comments

FDA Preliminary Response
NDA 21164 – gepirone HCl extended-release tablets
Fabre-Kramer Pharmaceuticals, Inc
Type C Meeting
Face to Face

Regarding the argument that FDA relied on a comparative effectiveness standard, we agree that new drugs for MDD cannot be held to a comparative efficacy standard under the FD&C Act or the Clinton-Gore Reinvention guidance.

Thomas Laughren, M.D.	Director, Division of Psychiatry Products (DPP)
Mitchell Mathis, M.D.	Deputy Director, DPP
Ni Khin, M.D.	Medical Team Leader, DPP
Silvana Borges, M.D.	Medical Reviewer, DPP
H.M. James Hung, Ph.D.	Director, DB1, Office of Biostatistics (OB)
Kooros Mahjoob, Ph.D.	Deputy Director, DB1, OB

Of the 4 trials terminated early, we agree that 3 of these should not be considered in such a judgment, because they were terminated early and did not have assay sensitivity: CN105-052; CN105-078; CN105-083.

Inc. plans to discuss the statistical report provided by their consultant, Mary P. Johnson. It was submitted in support of the sponsor's request that FDA reconsider our non-approval decision conveyed in a November 2, 2007 letter for gepirone ER in the treatment of major depressive disorder (MDD). Fabre-Kramer Pharmaceuticals, Inc. was previously granted a Face-to-Face meeting on January 14, 2008 to discuss their responses to the November 2, 2007 NA letter. During the January 14, 2008 meeting we concluded that it was highly unlikely any additional analyses of the existing database would justify further review of the NDA.

Regulatory History

1

Reference ID: 3035351

2014 General Advice (GA) Letter: Continued Reliance on Post-Hoc Endpoints to Show Assay Sensitivity & Inferior Efficacy

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration
Silver Spring MD 20993

NDA 21164

GENERAL ADVICE

Fabre-Kramer Pharmaceuticals, Inc.
Attention: Martin Lobel, Esq.
Attorney
Law Offices of Lobel, Novins & Lamont, LLP
888 17th Street, N.W., Suite 810
Washington, D.C. 20006

We acknowledge that in studies ORG134004, ORG134006, and ORG134017, the active control failed to reach statistical significance over placebo on the primary endpoint HAMD-25 for studies ORG134004 and ORG134006; MADRS for study ORG134017 . However, the treatment effect favored the active controls on the HAMD-17

The major deficiency cited in the November 2, 2007 NA letter was a failure to provide substantial evidence of efficacy in the short-term and longer-term treatment of major depressive disorder (MDD). Although the letter noted that the available evidence suggested that gepirone appeared to be less effective than other available antidepressants, relative efficacy was not the basis for the NA decision. The NA action was based on the lack of substantial evidence of effectiveness.

You have also argued against relying on a non-protocol specified endpoint to justify a conclusion of assay sensitivity, and we acknowledge some concern here with multiple comparisons. We are, however, dealing with an extraordinarily low study success rate in what appear to be well-controlled studies (i.e. 7 of 9).

efficacy data along with your current arguments in support of its efficacy. The efficacy data contained in the submission were the same as those reviewed in 2007 in their entirety.

We have reviewed the referenced material dated December 10, 2012 and have the following comments:

Reference ID: 3491931

2014 GA Letter: Continued Reference to Long-Term Study

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration
Silver Spring MD 20993

NDA 21164

GENERAL ADVICE

Fabre-Kramer Pharmaceuticals, Inc.
Attention: Martin Lobel, Esq.
Attorney
Law Offices of Lobel, Novins & Lamont, LLP
888 17th Street, N.W., Suite 810
Washington, D.C. 20006

Dear Mr. Lobel:

Please refer to your New Drug Application (NDA) submitted under section 505(b) of the Federal

efficacy data along with your current arguments in support of its efficacy.

The major deficiency cited in the November 2, 2007 NA letter was a failure to provide substantial evidence of efficacy in the short-term and longer-term treatment of major depressive disorder (MDD). Although the letter noted that the available evidence suggested that gepirone appeared to be less effective than other available antidepressants, relative efficacy was not the basis for the NA decision. The NA action was based on the lack of substantial evidence of effectiveness.

A face-to-face meeting was held on January 14, 2008 to discuss your responses to the November 2, 2007 NA letter and we concluded that it was highly unlikely any additional analyses of the existing database would justify further review of the NDA. On April 27, 2011 you requested reconsideration of the 2007 non-approval decision. Another face-to-face meeting was held on November 29, 2011 to discuss the statistical report provided by your consultant, Mary F. Johnson. You submitted the report to support your request that we reconsider our non-approval decision conveyed in the November 2, 2007 NA letter.

On December 10, 2012, following meetings, communications, and exchanges of documents, you submitted an NDA amendment providing information for an informal review of the gepirone ER efficacy data along with your current arguments in support of its efficacy. The efficacy data contained in the submission were the same as those reviewed in 2007 in their entirety.

We have reviewed the referenced material dated December 10, 2012 and have the following comments:

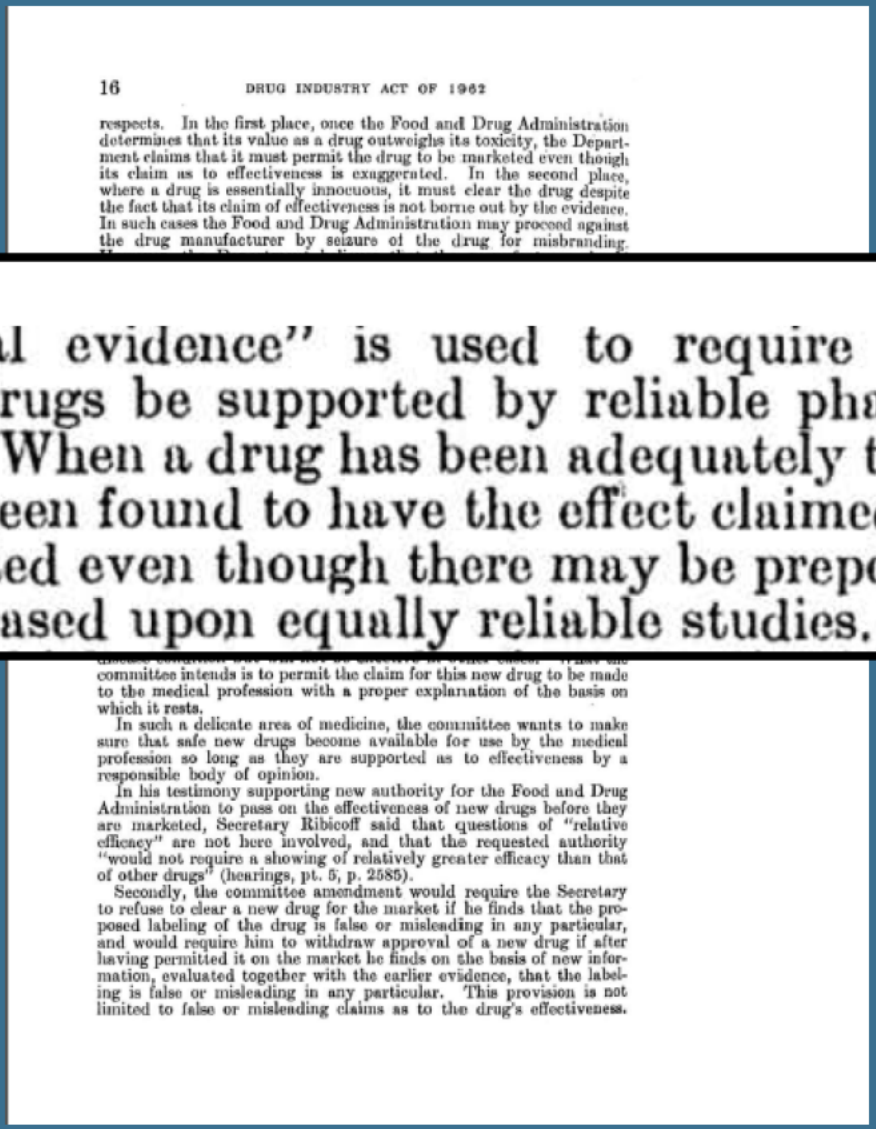
Reference ID: 3491931

The negative maintenance gepirone ER trial (ORG28709) is an additional piece of important evidence against the efficacy of gepirone ER for the treatment of MDD.

Also, numerous invalid rejections of Fabre-Kramer's arguments in NDA Amendments.

Legal Standards

1962 Drug Amendments: Senate Report



16 DRUG INDUSTRY ACT OF 1962

respects. In the first place, once the Food and Drug Administration determines that its value as a drug outweighs its toxicity, the Department claims that it must permit the drug to be marketed even though its claim as to effectiveness is exaggerated. In the second place, where a drug is essentially innocuous, it must clear the drug despite the fact that its claim of effectiveness is not borne out by the evidence. In such cases the Food and Drug Administration may proceed against the drug manufacturer by seizure of the drug for misbranding.

The term "substantial evidence" is used to require that therapeutic claims for new drugs be supported by reliable pharmacological and clinical studies. When a drug has been adequately tested by qualified experts and has been found to have the effect claimed for it, this claim should be permitted even though there may be preponderant evidence to the contrary based upon equally reliable studies.

committee intends is to permit the claim for this new drug to be made to the medical profession with a proper explanation of the basis on which it rests.

In such a delicate area of medicine, the committee wants to make sure that safe new drugs become available for use by the medical profession so long as they are supported as to effectiveness by a responsible body of opinion.

In his testimony supporting new authority for the Food and Drug Administration to pass on the effectiveness of new drugs before they are marketed, Secretary Ribicoff said that questions of "relative efficacy" are not here involved, and that the requested authority "would not require a showing of relatively greater efficacy than that of other drugs" (hearings, pt. 5, p. 2585).

Secondly, the committee amendment would require the Secretary to refuse to clear a new drug for the market if he finds that the proposed labeling of the drug is false or misleading in any particular, and would require him to withdraw approval of a new drug if after having permitted it on the market he finds on the basis of new information, evaluated together with the earlier evidence, that the labeling is false or misleading in any particular. This provision is not limited to false or misleading claims as to the drug's effectiveness.

Reference ID: 3717852

Independent Substantiation

Guidance for Industry

Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products

The usual requirement for more than one adequate and well-controlled investigation reflects the need for *independent substantiation* of experimental results.

Independent substantiation of a favorable result protects against the possibility that a chance occurrence in a single study will lead to an erroneous conclusion that a treatment is effective.

Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
May 1998
Clinical 6

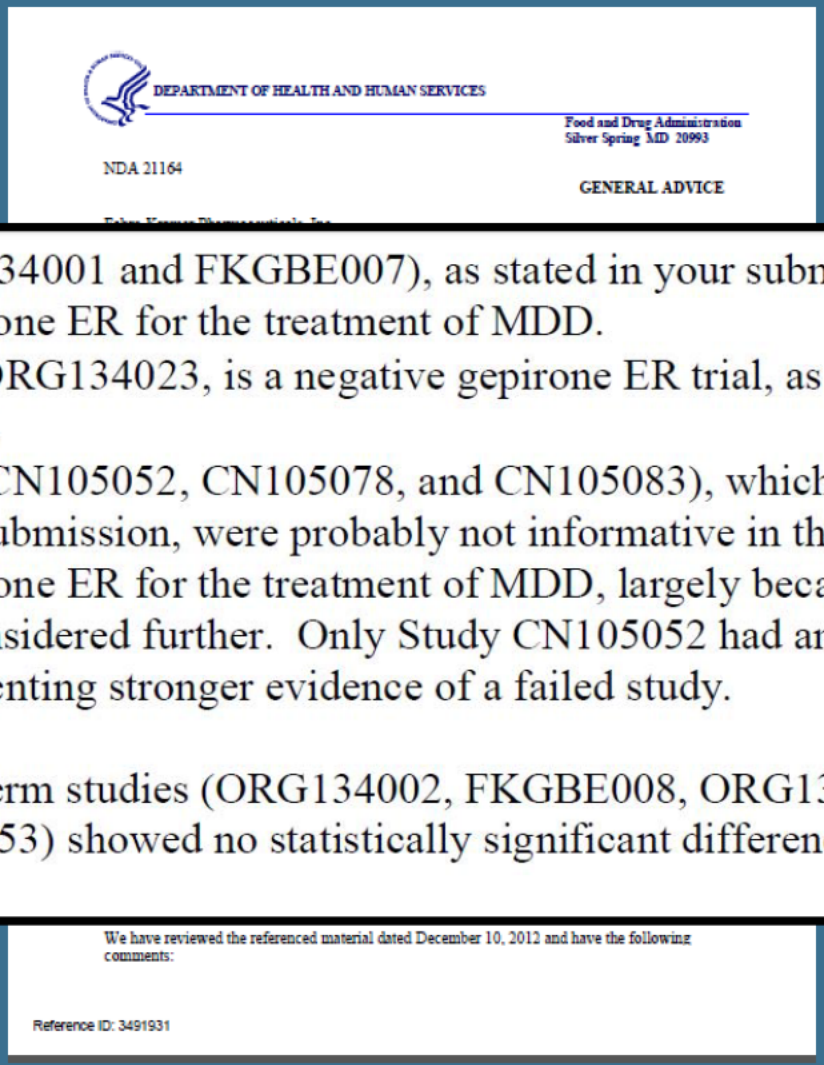
2 Adequate and Well-Controlled Studies = Substantial Evidence of Efficacy

The possibility of 2 positive AWC studies being outweighed by negative studies is not even mentioned in:

- The legislative history of the 1962 Drug Amendments
- The legislative history of the 1997 Food and Drug Administration Modernization Act (FDAMA)
- FDA's 1998 Guidance *Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products*

Analysis of Studies

2014 GA Letter: Summary of Studies



DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration
Silver Spring MD 20993

NDA 21164

GENERAL ADVICE

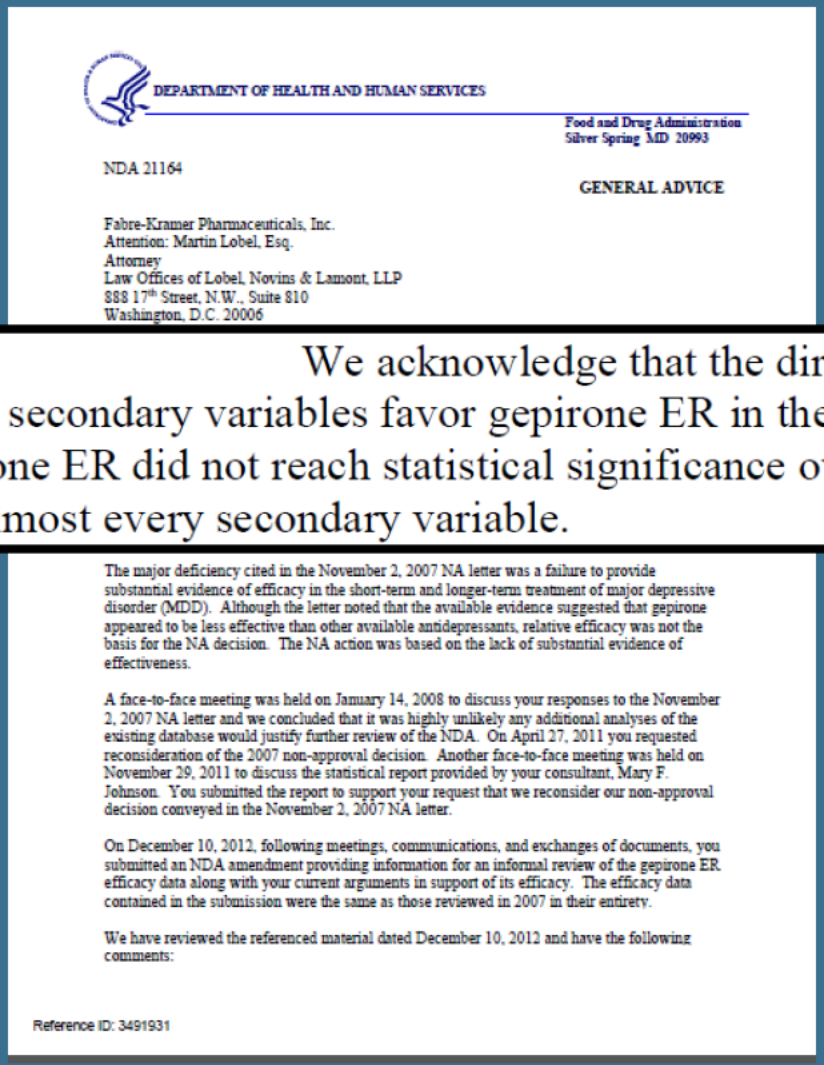
- 2 studies (ORG134001 and FKGBE007), as stated in your submission, support the efficacy of gepirone ER for the treatment of MDD.
- Another study, ORG134023, is a negative gepirone ER trial, as you acknowledged in your submission.
- 3 other studies (CN105052, CN105078, and CN105083), which you considered failed studies in your submission, were probably not informative in the evaluation of the efficacy of gepirone ER for the treatment of MDD, largely because of their overall size, and were not considered further. Only Study CN105052 had an active control showing no effect, representing stronger evidence of a failed study.

The remaining 6 short-term studies (ORG134002, FKGBE008, ORG134004, ORG134006, CN134017 and CN105053) showed no statistically significant difference between gepirone ER and placebo

We have reviewed the referenced material dated December 10, 2012 and have the following comments:

Reference ID: 3491931

ORG134002 and FKGBE008: Positive Directional Trends



DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration
Silver Spring MD 20993

NDA 21164

GENERAL ADVICE

Fabre-Kramer Pharmaceuticals, Inc.
Attention: Martin Lobel, Esq.
Attorney
Law Offices of Lobel, Novins & Lamont, LLP
888 17th Street, N.W., Suite 810
Washington, D.C. 20006

We acknowledge that the directional trend of the primary endpoint and the secondary variables favor gepirone ER in these studies. However, per your own analysis, gepirone ER did not reach statistical significance over placebo either on the primary endpoint or on almost every secondary variable.

The major deficiency cited in the November 2, 2007 NA letter was a failure to provide substantial evidence of efficacy in the short-term and longer-term treatment of major depressive disorder (MDD). Although the letter noted that the available evidence suggested that gepirone appeared to be less effective than other available antidepressants, relative efficacy was not the basis for the NA decision. The NA action was based on the lack of substantial evidence of effectiveness.

A face-to-face meeting was held on January 14, 2008 to discuss your responses to the November 2, 2007 NA letter and we concluded that it was highly unlikely any additional analyses of the existing database would justify further review of the NDA. On April 27, 2011 you requested reconsideration of the 2007 non-approval decision. Another face-to-face meeting was held on November 29, 2011 to discuss the statistical report provided by your consultant, Mary F. Johnson. You submitted the report to support your request that we reconsider our non-approval decision conveyed in the November 2, 2007 NA letter.

On December 10, 2012, following meetings, communications, and exchanges of documents, you submitted an NDA amendment providing information for an informal review of the gepirone ER efficacy data along with your current arguments in support of its efficacy. The efficacy data contained in the submission were the same as those reviewed in 2007 in their entirety.

We have reviewed the referenced material dated December 10, 2012 and have the following comments:

Reference ID: 3491931

Areas of Agreement (8 of 12 studies)

Positive Trials (2)	ORG 134001, FKGBE007
Negative Trials (3)	ORG134023, ORG134002,* FKGBE008*
Failed Trials (3)	CN105052, CN105078, CN105083
* with positive directional trend	

Studies in Dispute (4 of 12 Studies)

	Endpoint	Mode of Analysis
ORG134004		
Pre-specified	HAMD25	ANOVA
Post-hoc FDA	HAMD17	ANCOVA
ORG134006		
Pre-specified	HAMD25	ANCOVA (separate models)
Post-hoc FDA	HAMD17	ANCOVA (common model)
ORG134017		
Pre-specified	MADRS	ANOVA
Post-hoc FDA	HAMD17	ANCOVA
CN105053		
Pre-specified	HAMD17 & CGI Responder (co-primary)	ANOVA
FDA	HAMD17 only	??? (FK could not reproduce FDA result)

ORG134004, ORG134006, ORG134017: Failed Studies

- 3-arm studies (gepirone ER, active control, placebo).
- Primary endpoint predefined as HAMD-25 (004 & 006) or MADRS (017).
- For efficacy analysis, active control was included in each study for purpose of establishing assay sensitivity.
- On the respective pre-defined primary endpoints, there was no statistically significant difference between placebo and active control.

ORG134004, ORG134006, ORG134017: No Assay Sensitivity on Primary Endpoint

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration
Silver Spring MD 20993

NDA 21164

GENERAL ADVICE

Fabre-Kramer Pharmaceuticals, Inc.
Attention: Martin Lobel, Esq.
Attorney
Law Offices of Lobel, Novins & Lamont, LLP
888 17th Street, N.W., Suite 810
Washington, D.C. 20006

We acknowledge that in studies ORG134004, ORG134006, and ORG134017, the active control failed to reach statistical significance over placebo on the primary endpoint (HAMD-25 for studies ORG134004 and ORG134006; MADRS for study ORG134017). However, the treatment effect favored the active controls on the HAMD-17

substantial evidence of efficacy in the short-term and longer-term treatment of major depressive disorder (MDD). Although the letter noted that the available evidence suggested that gepirone appeared to be less effective than other available antidepressants, relative efficacy was not the basis for the NA decision. The NA action was based on the lack of substantial evidence of effectiveness.

A face-to-face meeting was held on January 14, 2008 to discuss your responses to the November 2, 2007 NA letter and we concluded that it was highly unlikely any additional analyses of the existing database would justify further review of the NDA. On April 27, 2011 you requested reconsideration of the 2007 non-approval decision. Another face-to-face meeting was held on November 29, 2011 to discuss the statistical report provided by your consultant, Mary F. Johnson. You submitted the report to support your request that we reconsider our non-approval decision conveyed in the November 2, 2007 NA letter.

On December 10, 2012, following meetings, communications, and exchanges of documents, you submitted an NDA amendment providing information for an informal review of the gepirone ER efficacy data along with your current arguments in support of its efficacy. The efficacy data contained in the submission were the same as those reviewed in 2007 in their entirety.

We have reviewed the referenced material dated December 10, 2012 and have the following comments:

Reference ID: 3491931

Assay Sensitivity Claim Based on Post-Hoc Endpoint Analysis Is Invalid

Guidance for Industry

E9 Statistical Principles for Clinical Trials

The primary variable should be specified in the protocol, along with the rationale for its selection. Redefinition of the primary variable after unblinding will almost always be unacceptable, since the biases this introduces are difficult to assess.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
September 1998
ICH

Roundtable Discussion, Clinical Trials (2005)

CLINICAL TRIALS

ROUNDTABLE DISCUSSION

Clinical Trials 2005; **2**: 373–378

FDA senior management perspectives

Janet Woodcock, Robert Temple, Karen Midthun, Daniel Schultz and Stephen Sundlof

Dr Temple: Well, let me make a distinction between information that gets incorporated before the results are known and after the results are known. If you already have the trial data in hand and then dig up some reason to modify the analysis plan, I would say that approach would have zero credibility.

ORG134004, ORG134017: No Assay Sensitivity – *Even on HAMD-17*

- In these 2 studies, active control did not statistically separate from placebo even on HAMD-17.

- October 2011 Preliminary Comments*:

> In this study, neither active comparator nor the new drug beat placebo, however, we still considered the study negative for gepirone because the active comparator beat gepirone on HAMD-17.

* This statement applies to both ORG134004 and ORG134017

Assay Sensitivity in 3-Arm Trials

Guidance for Industry

E 10 Choice of Control Group and Related Issues in Clinical Trials

three-arm trials including an active control as well as a placebo-control group can readily assess whether a failure to distinguish test treatment from placebo implies ineffectiveness of the test treatment or is simply the result of a trial that lacked the ability to identify an active drug. The comparison of placebo to standard drug in such a trial provides internal evidence of assay sensitivity.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2001
ICH

No Assay Sensitivity if Active Control Does Not Beat Placebo

Guidance for Industry

Non-Inferiority Clinical Trials

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document contact Robert Temple at 301-796-2270 or Robert O'Neill at 301-796-1700 (CDER), or the Office of Communication, Outreach, and Development (CBER) at 301-800-835-4709 or 301-827-1800.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

March 2010
Clinical/Medical

No Assay Sensitivity if Active Control Does Not Beat Placebo

Guidance for Industry

b. Assessing assay sensitivity of a placebo-controlled study

Although a successful superiority trial (e.g., placebo-controlled) is readily interpreted, a failed trial of this design is not. Failure to show superiority to placebo can mean that the drug is ineffective or that the trial lacked assay sensitivity. To distinguish between these two possibilities, it is often useful to include an active control in placebo-controlled studies of drugs in a class or condition where known effective drugs often cannot be distinguished from placebo (e.g., depression, allergic rhinitis, angina, and many other symptomatic conditions). If the active control is superior to placebo but the test drug is not, one can conclude that the test drug lacks effectiveness (or at least is less effective than the active control). If neither the active control nor the test drug is superior to placebo, the trial lacked assay sensitivity and is uninformative about the effect of the test drug.

Center for Biologics Evaluation and Research (CBER)

March 2010
Clinical/Medical

No Assay Sensitivity if Active Control Does Not Beat Placebo

Guidance for Industry

b. Assessing assay sensitivity of a placebo-controlled study

Although a successful superiority trial (e.g., placebo-controlled) is readily interpreted, a failed trial of this design is not. Failure to show superiority to placebo can mean that the drug is ineffective or that the trial lacked assay sensitivity. To distinguish between these two possibilities, it is often useful to include an active control in placebo-controlled studies of drugs in a class or condition where known effective drugs often cannot be distinguished from placebo (e.g., depression, allergic rhinitis, angina, and many other symptomatic conditions).

If neither the active control nor the test drug is superior to placebo, the trial lacked assay sensitivity and is uninformative about the effect of the test drug.

Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

March 2010
Clinical/Medical

CN105053: No Assay Sensitivity on HAMD-17

Table 36: HAMD-17 Total Score: Change from Baseline at Each Visit (ITT/LOCF) – Study CN105-053

Treatment		Baseline	LS Mean Change from Baseline					
			Week 1	Week 2	Week 3	Week 4	Week 6	Week 8/ET
Combined Sites								
gepirone-ER	**n**	56	55	56	56	56	56	56
	Mean	23.9	-3.9	-6.4	-8.1	-8.7	-10.5	-9.7
	SE	0.41	0.7	0.8	1.0	1.0	1.2	1.2
imipramine	**n**	54	54	54	54	54	54	54
	Mean	24.2	-3.8	-7.3	-8.6	-10.2	-11.3	-11.5
	SE	0.39	0.7	0.9	1.0	1.1	1.2	1.2
placebo	**n**	56	53	56	56	56	56	56
	Mean	24.2	-3.3	-4.8	-6.6	-7.6	-8.9	-9.0
	SE	0.39	0.7	0.8	1.0	1.0	1.2	1.2
gepirone-ER vs. pbo	p-value		0.568	0.195	0.274	0.448	0.343	0.687
imipramine vs. pbo	p-value		0.633	0.041	0.153	0.080	0.160	0.144

LS Means and p-values for combined sites is based on ANOVA with effects for treatment, center and treatment by center interaction.
[*Source: CSR CN105-053 Table 17, Appendix F Tables 7.1.1-3 and 7.1.1-6*]

Fabre-Kramer Response to FDA Information Request Letter of 5/8/2012

CN105053: FDA Claim of Assay Sensitivity on HAMD-17

Trial	Active Comparator	P-Values Act Comp vs Pbo	P-Values Gepirone ER vs Pbo
CN105-053	Imipramine	-3.19 (p=0.038)	-2.00 (p=0.190)

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service
Food and Drug Administration
Rockville, MD 20857

NDA 21-164

Fabre-Kramer Pharmaceuticals, Inc.
Attention: Stephen J. Kramer M.D., Chief Executive Officer
5847 San Felipe, Suite 2000
Houston, Texas 77057

Therefore, the application is not approvable under section 505(d) of the Act and 21 CFR 314.125(b). The deficiencies are summarized as follows:

Non-Approval Deficiencies:
There are two major deficiencies in the application. First you have failed to provide substantial evidence for the effectiveness of gepirone ER in the short-term or longer-term treatment of major depressive disorder (MDD). In addition, if the data could be considered as representing such evidence we also believe the effect size that would have been demonstrated is unacceptably small compared to alternative therapy for a serious illness. Although we agree that you have provided two well – controlled trials that show a significant effect, i.e. studies FKGBE007 and 134001, this evidence is derived from a total pool of 12 studies. The remaining 10 studies do not show evidence of an antidepressant effect. Indeed, they do not even show favorable trends in almost all cases. We acknowledge that 4 of these remaining 10 studies were terminated early for business reasons, and, therefore, might not be expected to provide such evidence because they did not reach their planned sample sizes, but in one of these an active control was effective and none of the studies show reasonably strong favorable trends. Moreover, there are other findings among these trials that amplify our concern about the potential value of gepirone ER as a treatment for MDD.

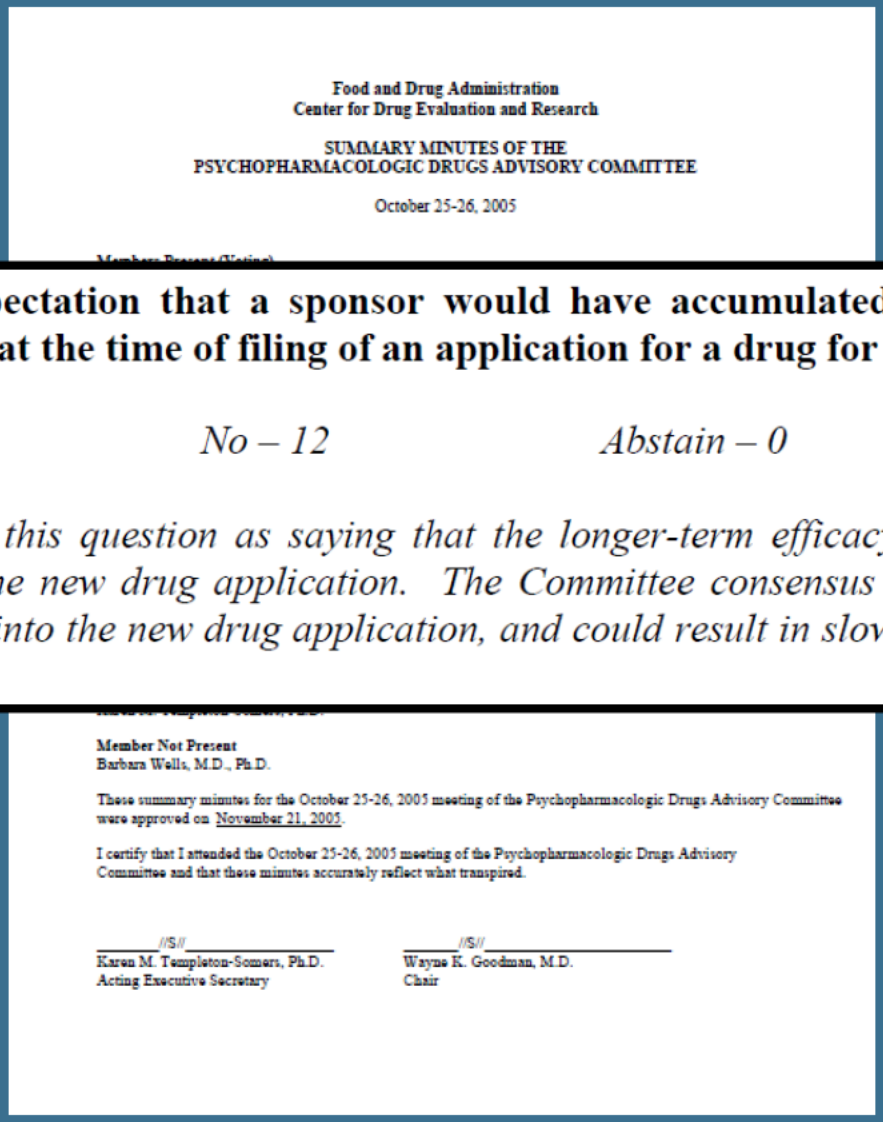
We have re-evaluated all 12 trials with a focus on the HAMD-17 total score as a common measure of efficacy. Although this was not the protocol specified primary endpoint for 3 of the 12 trials, we felt it was a reasonable common measure because it is so widely used as a primary endpoint in depression trials. In fact, you selected this as a common endpoint for your meta-analyses. Using this measure, we found the following:

Summary of Disputed Short-Term Trials

- **ORG134006: FAILED STUDY**
  - No Assay Sensitivity on Primary Endpoint
- **ORG134004 and ORG134017: FAILED STUDIES**
  - No Assay Sensitivity on Primary Endpoints
  - No Assay Sensitivity Even After Post-Hoc Analysis
- **CN105053: FAILED STUDY**
  - No Assay Sensitivity on Co-Primary Endpoint

Study in Dispute: Long-Term Trial

2005 Psychopharmacologic Drugs Advisory Committee Meeting



Food and Drug Administration
Center for Drug Evaluation and Research

SUMMARY MINUTES OF THE
PSYCHOPHARMACOLOGIC DRUGS ADVISORY COMMITTEE

October 25-26, 2005

1. **Is it a reasonable expectation that a sponsor would have accumulated data for both acute and longer-term efficacy trials at the time of filing of an application for a drug for the treatment of MDD?**

Yes – 0 *No – 12* *Abstain – 0*

The Committee interpreted this question as saying that the longer-term efficacy data would be completed, analyzed and included in the new drug application. The Committee consensus was that this is too strict a requirement to incorporate into the new drug application, and could result in slowing the process of new drug development.

Member Not Present
Barbara Wells, M.D., Ph.D.

These summary minutes for the October 25-26, 2005 meeting of the Psychopharmacologic Drugs Advisory Committee were approved on November 21, 2005.

I certify that I attended the October 25-26, 2005 meeting of the Psychopharmacologic Drugs Advisory Committee and that these minutes accurately reflect what transpired.

_____//S//_____
Karen M. Templeton-Somers, Ph.D.
Acting Executive Secretary

_____//S//_____
Wayne K. Goodman, M.D.
Chair

2014 GA Letter: Statement on Long-Term Studies

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration
Silver Spring MD 20993

NDA 21164

GENERAL ADVICE

Fabre-Kramer Pharmaceuticals, Inc.
Attention: Martin Lobel, Esq.
Attorney
Law Offices of Lobel, Novins & Lamont, LLP
888 17th Street, N.W., Suite 810
Washington, D.C. 20006

Dear Mr. Lobel:

In our review of all maintenance trials with approved antidepressants, every single maintenance trial with these characteristics has shown positive results. In this context, the negative results of this maintenance trial with gepirone ER are difficult to ignore.

substantial evidence of efficacy in the short-term and longer-term treatment of major depressive disorder (MDD). Although the letter noted that the available evidence suggested that gepirone appeared to be less effective than other available antidepressants, relative efficacy was not the basis for the NA decision. The NA action was based on the lack of substantial evidence of effectiveness.

A face-to-face meeting was held on January 14, 2008 to discuss your responses to the November 2, 2007 NA letter and we concluded that it was highly unlikely any additional analyses of the existing database would justify further review of the NDA. On April 27, 2011 you requested reconsideration of the 2007 non-approval decision. Another face-to-face meeting was held on November 29, 2011 to discuss the statistical report provided by your consultant, Mary F. Johnson. You submitted the report to support your request that we reconsider our non-approval decision conveyed in the November 2, 2007 NA letter.

On December 10, 2012, following meetings, communications, and exchanges of documents, you submitted an NDA amendment providing information for an informal review of the gepirone ER efficacy data along with your current arguments in support of its efficacy. The efficacy data contained in the submission were the same as those reviewed in 2007 in their entirety.

We have reviewed the referenced material dated December 10, 2012 and have the following comments:

Reference ID: 3491931

2013: Fetzima Approved with Negative Long-Term Study

CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:
204168Orig1s000

SUMMARY REVIEW

In one other short-term flexible dose US study (LVM-MD-02) and the longer-term maintenance study (LVM-MD-05), the drug did not separate from placebo for efficacy.

Comment: It is very unusual for an antidepressant with multiple positive short-term studies to not demonstrate a difference from placebo in a maintenance study. I believe the primary reason this study was negative is because the patients were only stable for 2 weeks prior to randomization. We have asked the sponsor to repeat this study with a better design post-marketing and they have agreed. Labeling will have our usual advice about continuing treatment past the initial response.

Fabre-Kramer repeatedly made the same argument about inadequate criteria for randomization in the gepirone ER long-term maintenance trial. See, e.g., December 2012 NDA Amendment at 96-97.

Summary: Ratio of Positive to Negative Short-Term Studies

- 2007 NA letter claimed ratio of 2 positive to 10 negative studies (3.5% likelihood that positive findings were due to chance).
  - BUT Division later acknowledged 3 of these "negative" studies were "probably not informative" and would "not [be] considered further."
- 2014 GA letter claimed that the ratio was 2 positive to 7 negative studies (2.0% likelihood that positive findings were due to chance).
  - BUT 4 of the 7 claimed "negative" studies were deemed negative based on flawed post-hoc endpoint analysis.
- The real ratio is 2 positive to 3 negative studies (0.59% likelihood that positive findings were due to chance).
  - GA letter acknowledged that 2 of the negative studies showed a "directional trend" that "favor[s] gepirone ER."

Recently-Approved Antidepressants with More Negative Studies than Positive Studies

1998	Celexa (citalopram)
2004	Cymbalta (duloxetine)
2008	Pristiq (desvenlafaxine)

Very Few 3-Arm Trials for FDA-Approved MDD Drugs Are Positive

Investigational Drug	Year Approved	Short term 3 arm Comparator Studies	Failed*	Positive**	Negative**	Percentage Failed or Negative
citalopram	1998	5	5	0	0	100
escitalopram	2002	2	1	1	0	50
duloxetine	2004	6	5	1	0	83
desvenlafaxine	2008	2	1	0	1	100
vilazodone	2011	3	3	0	0	100
TOTALS		**18**	**15**	**2**	**1**	**89**
gepirone ER		5	5	0	0	100
ADJUSTED TOTAL		**23**	**20**	**2**	**1**	**91**

* on primary efficacy endpoint

** on primary efficacy endpoint with assay sensitivity

Dr. Michael Thase

Professor of Psychiatry
Perelman School of Medicine
University of Pennsylvania
Philadelphia VAMC

Advisory Committee Not Necessary

- Division informed Fabre Kramer in June 2012 that FDA "very unlikely" to request advisory committee (June 25, 2012 email from H. Patel to M. Lobel)
- Advisory committees are not normally convened to assess a finding of substantial evidence of efficacy based on two indisputably positive adequate and well controlled studies.
- Gepirone ER presents no increased risk relative to other products against which to balance its efficacy profile – indeed, it presents lower risks than other products.
- No recent antidepressant NDA has been referred to an advisory committee
    - Including NME drugs, e.g., Brintellix (vortioxetine), 2013; Viibryd (vilazodone), 2011.

Overall Summary

- Efficacy of gepirone ER established by 2 unequivocally positive trials
- The effect size in these 2 trials is similar to the pivotal trials of other approved antidepressants.
- The determination that gepirone ER was not effective was based on methods of analysis that FDA would not accept from sponsors.
- 2 positive trials out of 5 adequate and well-controlled trials is consistent with other approved antidepressants
- Gepirone ER presents a safety profile at least as favorable as those of approved antidepressants and does not have their sexual side effects.
- More treatment options for MDD are needed.

A Plea for Consistent Treatment

- Use the same approach to "substantial evidence" as with other approved antidepressants
- Use only the protocol-specified endpoints and methods in assessing submitted studies
- Determine assay sensitivity in 3-arm trials based on comparison of active control to placebo
- Do not rely on comparative efficacy analysis
- Do not use unsuccessful long-term maintenance study as evidence against approval of short-term indication
- Do not refer NDA to Advisory Committee

Fabre Kramer Remains Ready to Work with FDA

- Labeling
- Post-Approval Commitments

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

/s/

KHUSHBOO SHARMA
03/18/2015

5.15 Dispute Appeal General Advice –John Jenkins (04/07/2015)

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration
Silver Spring MD 20993

NDA 21164

DISPUTE APPEAL- GENERAL ADVICE

Fabre-Kramer Pharmaceutical
Attention: Stephen J. Kramer, CEO
5847 San Felipe
Suite 2000
Houston, TX 77057

Dear Dr. Kramer:

Please refer to your New Drug Application (NDA) submitted under section 505(b) of the Federal Food, Drug, and Cosmetic Act for gepirone hydrochloride extended release tablets (gepirone ER).

I also refer to the following:

- Your June 13, 2014, request for formal dispute resolution appealing the November 2, 2007, Not Approvable (NA) letter and the April 18, 2014, General Advice letter in which the Office of Drug Evaluation I (ODE I) concluded that Fabre-Kramer had not demonstrated "substantial evidence" of gepirone ER's effectiveness.
- The June 30, 2014, Dispute Not Accepted letter stating that the June 13, 2014, request for formal dispute resolution was not accepted because it contained new information/re-analysis of previously submitted information, that was not previously reviewed by the original deciding authority. The Dispute Not Accepted letter stated that Fabre-Kramer could appeal the November 2, 2007, NA letter and the re-analysis submitted after the NA action would not be considered as part of the appeal. The Dispute Not Accepted letter also offered Fabre-Kramer an Advisory Committee (AC) meeting to discuss the clinical issues and the re-analysis submitted after the NA action.
- Your November 12, 2014, letter to Elizabeth Dickinson, J.D., Chief Counsel, FDA, in which you requested that the Office of New Drugs (OND) accept Fabre-Kramer's formal dispute resolution request submitted on June 13, 2014, regarding the November 2, 2007, NA letter and the April 18, 2014, General Advice letter from ODE I.
- The January 27, 2015, Acknowledgement and Meeting Granted letter stating that your request for formal dispute resolution, dated June 13, 2014, was accepted, and that the FDA receipt date for the request for formal dispute resolution was January 27, 2015.
- Meeting between FDA and Fabre Kramer held on February 23, 2015
- Your March 6, 2015, letter providing follow-up information on issues that were raised at the February 23, 2015, meeting between the FDA and Fabre-Kramer.
- The March 18, 2015 Interim Response stating that I require discussion with internal FDA experts, prior to me being able to render a final decision on the appeal. I have requested Dr. LaVange, Director, Office of Biostatistics, CDER, and her staff to re-review the available data from the twelve short-term trials and the one long-term maintenance trial.
- Your March 26, 2015 letter, received on March 31, 2015 where you raised several points regarding the consultation to Dr. LaVange and her staff, and provided a comment to the meeting minutes from the February 23, 2015 meeting between the FDA and Fabre-Kramer.

In response to the points raised in items I(a) and I(b) of your March 26, 2015 letter, when I review an appeal, I review the entire case and all aspects of the issue(s) in dispute between the sponsor and the FDA. The appeal in front of me is concerning the November 2, 2007, NA letter. The General Advice letter dated April 18, 2014, reiterated the conclusions stated in the November 2, 2007, NA letter. Fabre-Kramer is specifically appealing the conclusion that ODE I believes that Fabre-Kramer had not demonstrated "substantial evidence" of gepirone ER's effectiveness. Therefore, in order for me to evaluate the totality of evidence of gepirone ER's effectiveness, I have requested Dr. LaVange to review and opine on all the twelve short term-studies and the one long-term maintenance study. As Dr. LaVange is an FDA employee, she will have access to the entire administrative file related to this application.

In Item II of your March 26, 2015, letter, you state that the draft guidance on Formal Dispute Resolution: Appeals Above the Division Level (March 2013) states (p. 8) that when the decision-maker in a Formal Dispute Resolution Process requires limited discussions with internal experts, that discussion will be held in a meeting involving "all parties ... on a mutually acceptable date and time." As the surrounding text makes clear, "all parties" in the guidance refers to the deciding authority and the consulted experts. The language in the draft guidance on Formal Dispute Resolution ensures that the discussions with the consulted experts are conducted in a timely manner so as to not unreasonably delay a final response to the appeal. "All parties" does not include discussions with the sponsor. Additionally, as Dr. LaVange is a FDA employee, her review is not publically available. Reviews conducted by FDA staff during the review of an application are not publically available, since they are pre-decisional.

In response to Item III of your March 26, 2015, letter, we note your comment to the meeting minutes from the February 23, 2015, meeting between the FDA and Fabre-Kramer. The FDA meeting minutes issued on March 18, 2015, is the official meeting record and will remain unchanged. Your March 26, 2015, letter is also part of the administrative record.

If you have any questions, call Khushboo Sharma at (301) 796-1270.

Sincerely,

{See appended electronic signature page}

John Jenkins, M.D.
Director
Office of New Drugs
Center for Drug Evaluation and Research

cc:
Covington & Burling LLP
Gerald Masoudi
1201 Pennsylvania Ave, NW
Washington, DC 20004

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

/s/

JOHN K JENKINS
04/07/2015

5.16 Interim Response to Appeal –John Jenkins (06/01/2015)

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration
Silver Spring MD 20993

NDA 21164

INTERIM RESPONSE TO APPEAL–-
INPUT NEEDED FROM ADVISORY COMMITTEE

Fabre-Kramer Pharmaceutical
Attention: Stephen J. Kramer, CEO
5847 San Felipe
Suite 2000
Houston, TX 77057

Dear Dr. Kramer:

Please refer to your New Drug Application (NDA) submitted under section 505(b) of the Federal Food, Drug, and Cosmetic Act for gepirone hydrochloride extended release tablets (gepirone ER or gepirone).

I also refer to the following:

- Your June 13, 2014, request for formal dispute resolution appealing the November 2, 2007, Not Approvable (NA) letter and the April 18, 2014, General Advice letter in which the Office of Drug Evaluation I (ODE I) concluded that Fabre-Kramer had not demonstrated substantial evidence of gepirone ER's effectiveness in the treatment of major depressive disorder (MDD).
- The June 30, 2014, Dispute Not Accepted letter stating that the June 13, 2014, request for formal dispute resolution was not accepted because it contained new information/re-analysis of previously submitted information, that was not previously reviewed by the original deciding authority. The Dispute Not Accepted letter stated that Fabre-Kramer could appeal the November 2, 2007, NA letter and the re-analysis submitted after the NA action would not be considered as part of the appeal. The Dispute Not Accepted letter also offered Fabre-Kramer an advisory committee (AC) meeting to discuss the clinical issues and the re-analysis submitted after the NA action.
- Your November 12, 2014, letter to Elizabeth Dickinson, J.D., Chief Counsel, FDA, in which you requested that the Office of New Drugs (OND) accept Fabre-Kramer's formal dispute resolution request submitted on June 13, 2014, regarding the November 2, 2007, NA letter and the April 18, 2014, General Advice letter from ODE I.
- The January 27, 2015, Acknowledgement and Meeting Granted letter stating that your request for formal dispute resolution, dated June 13, 2014, was accepted, and that the FDA receipt date for the request for formal dispute resolution was January 27, 2015.
- The meeting between FDA and Fabre Kramer held on February 23, 2015.
- Your March 6, 2015, letter providing follow-up information on issues that were raised at the February 23, 2015, meeting between the FDA and Fabre-Kramer.
- The March 18, 2015, Interim Response letter stating that I required discussion with internal FDA experts, prior to reaching a final decision on the appeal. I requested that Dr.

LaVange, Director, Office of Biostatistics, CDER, and her staff re-review the available data from the twelve short-term trials and the one long-term maintenance trial.

- Your March 26, 2015, letter, received on March 31, 2015, in which you raised several points regarding the consultation to Dr. LaVange and her staff, and provided a comment to the minutes from the February 23, 2015, meeting between the FDA and Fabre-Kramer.
- The April 7, 2015, General Advice letter in response to your March 26, 2015, letter.

I have reviewed your appeal and conclude that additional input from an expert advisory committee is needed before I reach my decision on your appeal.

Accordingly, I will direct the Division of Psychiatry Products and the Office of Drug Evaluation I to convene a public meeting of the Psychopharmacologic Drugs Advisory Committee to review the available data supporting the use of gepirone for the treatment of major depressive disorder (MDD). While the primary issue in the current dispute relates to whether you have provided substantial evidence of effectiveness for the proposed use, I believe the advisory committee should be asked to review the entirety of the efficacy and safety data for gepirone and to opine on the full range of issues on which FDA typically requests input from advisory committees during pre-marketing review of an NDA; i.e., demonstration of efficacy, review of the available safety data, whether the available data support a favorable benefit risk profile to support approval, and what, if any, additional data are needed pre- or post-approval to address outstanding issues.

We will notify you when the meeting is scheduled and work with you on the planning, as appropriate. I will direct FDA staff to schedule the advisory committee meeting in a timely manner in order to minimize the delay in reaching a decision on your appeal.

I believe discussion at an Advisory Committee meeting is needed to reach a decision for the following reasons:

First, the issues in dispute involve complex analyses and interpretations of the data provided in support of demonstration of the efficacy of gepirone for the treatment of MDD. Although you have provided two adequate and well-controlled trials that were positive, the positive trials were a relatively small part of an overall development program for gepirone to treat MDD that included numerous adequate and well-controlled trials that failed to demonstrate efficacy for both early immediate-release formulations and the proposed to-be-marketed extended release (ER) formulation. You state that the likelihood of obtaining two positive trials out of a pool of 12 short-term trials of the ER formulation, assuming gepirone is in fact ineffective, is approximately 2%. You conclude the probability of a false positive finding of efficacy for gepirone is below the Agency's traditional threshold for the p-value of an individual trial of 0.05 (which suggests less than a 5% chance of a false positive result). However, your analysis oversimplifies what the Agency evaluates in determining whether the statutory standard of substantial evidence has been met.

An exact definition of the level of assurance the Agency normally requires to protect against approving a truly ineffective drug is not stated in the Agency's 1998 *Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products* (the

"Effectiveness Guidance"), but the guidance does note that when many trials are conducted to demonstrate efficacy there is a reasonable chance that some will be "effective" at conventional levels of statistical significance by chance alone, even if the test drug is truly ineffective. The guidance also notes that "independent substantiation of a favorable result protects against the possibility that a chance occurrence in a single study will lead to an erroneous conclusion that a treatment is effective." The guidance specifically refers to the Agency's usual standard of "at least two adequate and well-controlled studies, each convincing on its own, to establish effectiveness." However, the guidance is silent on how this standard should be applied in a situation, like the one for gepirone, where there are two adequate and well-controlled trials supporting efficacy but also a large number of adequate and well-controlled trials that fail to demonstrate efficacy. Although there have been other development programs where some adequate and well-controlled trials have failed to demonstrate efficacy and the Agency nonetheless concluded the standard for substantial evidence had been met, the gepirone case represents an extreme with respect to the number of negative adequate and well-controlled trials included in the development program along with the two positive trials.

It is not clear to me that your proposed approach of counting trials as "positive" or "negative" and then calculating an overall "p-value" for the development program is the most appropriate approach to understanding the false positive risk for gepirone, as the analysis ignores the level of evidence provided by each trial, reducing that evidence to a 1-0 outcome for positive or negative. Another approach could be to synthesize the information available from all the trials using appropriate statistical methods for meta-analysis to explore potential inconsistency of treatment effects across the trials. While FDA does not accept a meta-analysis as the basis for demonstration of substantial evidence of effectiveness, a meta-analysis can provide useful information to integrate the effects observed across all trials in interpreting the findings from the two positive trials in conjunction with other negative trials. One of the challenges in using a meta-analysis for this purpose is to determine which trials to include and which to exclude. In assessing gepirone's evidence of effectiveness, these decisions have a profound impact on both the point estimate of the overall effect size of gepirone and also the nominal p-value obtained.

The 12 adequate and well-controlled short-term trials submitted in support of gepirone ER were conducted over an extended period of time and by three different sponsors. While many of the trials were similarly designed, some specifically targeted a population of patients with atypical symptoms of depression and pre-specified primary endpoints other than HAMD-17, which has traditionally been the primary endpoint for demonstration of efficacy of drugs for MDD. These factors further complicate interpretation of the data supporting the efficacy of gepirone.

I believe that an expert advisory committee, composed of psychiatrists, experts in clinical trial design, conduct, and analysis, and statisticians will provide valuable advice to assist me in reaching my decision on whether the available data constitute substantial evidence of effectiveness.

Second, it is logical to invoke the requirement for a public advisory committee meeting under Section 918 of the Food and Drug Administration Amendments Act (FDAAA) of 2007 during this formal dispute process. Section 918 of FDAAA states that "prior to the approval of a drug no active ingredient (including any ester or salt of the active ingredient) of which has been approved

in any other application under this section or section 351 of the Public Health Service, the Secretary shall— refer such drug to a Food and Drug Administration advisory committee for review at a meeting of such advisory committee."

The November 12, 2007, NA letter listed two clinical deficiencies that precluded approval. The first, and most important, was a conclusion that you had not provided substantial evidence of effectiveness for gepirone in the treatment of MDD.[1] If I reach a conclusion that you have provided substantial evidence of effectiveness, and grant your appeal, the application would be favorably situated for approval on resubmission.[2] It would be illogical to present this application to an advisory committee after my review if I grant your appeal because the issues the AC would normally be asked to opine on would have already been decided. FDAAA Section 918 states "….if the Secretary does not refer a drug to a Food and Drug Administration advisory committee prior to approval of the drug, (the Secretary shall) provide in the action letter on the application for the drug a summary of the reasons why the Secretary did not refer the drug to an advisory committee prior to approval." Although the Agency has used this provision to justify not referring a new molecular entity NDA to an AC on numerous occasions, I do not find this provision to be applicable here. The gepirone NDA has received 3 NA letters and has been the subject of a formal dispute resolution related to the issue of whether substantial evidence of effectiveness has been provided. The administrative record includes strongly held views from Agency staff, including Dr. Robert Temple, stating that substantial evidence of effectiveness has not been provided. Before I reach a decision on your appeal, which could effectively approve your application if the appeal were granted, I believe this application must be presented for review before a public advisory committee.

During our meeting on February 23, 2015, you argued that convening an advisory committee to discuss this application was "not necessary." I do not agree with the reasons you presented in that meeting and note that you failed to address the FDAAA Section 918 requirement for an advisory committee. You also expressed concerns about the cost of preparing for an advisory committee meeting, but cost is not a factor FDA considers in administering FDAAA Section 918.

I will respond to your appeal within 30 calendar days after the Advisory Committee meeting.

If you have any questions, call Khushboo Sharma, CDER Formal Dispute Resolution Project Manager at (301) 796-1270.

[1] The second clinical deficiency related to the "unacceptably small" effect size of gepirone compared to other approved drugs to treat MDD. Dr. Temple clarified in his April 18, 2014, General Advice letter that "relative efficacy was not the basis for the NA action;" therefore, I do not address this issue here. I do note, however, that aspects of the benefit of a drug, such as effect size, durability, etc., are components of the benefit-risk decision the Agency must reach in determining whether to approve a drug.

[2] You would be required to satisfactorily address other outstanding issues such as CMC, GMP compliance, labeling, any requests for postmarketing required studies or postmarketing commitments, etc., prior to an approval action.

Sincerely,

{See appended electronic signature page}

John K. Jenkins, M.D.
Director
Office of New Drugs
Center for Drug Evaluation and Research

cc:
Covington & Burling LLP
Lewis Grossman
Michael Labson
1201 Pennsylvania Ave, NW
Washington, DC 20004

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

/s/

JOHN K JENKINS
06/01/2015

5.17 Statistics Memo –Lisa Lavange (06/08/2015)

Date of the Memorandum: June 8, 2015

Memorandum May 1, 2015

To: John Jenkins, MD, Director, Office of New Drugs
From: Lisa LaVange, PhD, Director, Office of Biostatistics
Re: Consult for the FDRR of Gepirone-ER

In response to your request for a consult from the Office of Biostatistics regarding the dispute resolution for gepirone-ER, I reviewed the following documents from the Division of Biometrics I related to the product in question:

1. Statistical review of NDA 21-164 by Dr. Fanhui Kong, May 1, 2007
2. Statistical review of NDA 21-164 Amendment by Dr. Yeh-Fong Chen, Dec. 7, 2012
3. Secondary statistical review of NDA 21-164 Amendment by Dr. Peiling Yang, Dec. 7, 2012
4. Summary of additional meta-analyses generated by Dr. Fanhui Kong, March 21, 2015
5. Summary of gepirone-ER clinical trial results generated by Dr. Peiling Yang and updated on March 31, 2015

In addition, I consulted Dr. Tom Permutt, Division Director for Biometrics II. Based on the Feb. 23, 2015 meeting with the sponsor, Fabre-Kramer Pharmaceuticals, Inc. and several internal meetings held surrounding the meeting with the sponsor, there appeared to be three broad areas of concern regarding interpretation of the evidence provided in support of gepirone-ER, (i) determination of assay sensitivity for the negative trials, (ii) appropriate methods for synthesizing evidence across trials, and (ii) evidence provided by the maintenance trial. Each of these areas is addressed in the following sections, and my conclusions are provided at the end of this document.

Determination of assay sensitivity: The addition of an active control arm in anti-depressant trials has been recommended by the review division for the past several years due to the high rate of placebo response observed in anti-depressant trials. The need for trial sensitivity, which inclusion of an active control arm is intended to address, is related to the reason that including a placebo arm is considered ethical, namely that treatment effects can be modest in this disease area. Specific advice about the analysis strategy to be used for assessing trial sensitivity is not always provided, but the approach taken here was to compare the active control to placebo relative to the primary endpoint of the trial, with the understanding that this sensitivity assessment is only needed when the hypothesis test comparing the test drug to placebo, also with respect to the primary endpoint, fails to reach statistical significance.

From the draft guidance on non-inferiority studies (Guidance for Industry: Non-Inferiority Clinical Trials, 2010), we have the following definition:

> "Assay sensitivity is an essential property of a [non-inferiority] NI trial. Assay sensitivity is the ability of the trial to have detected a difference between treatments of a specified size. Stated in another way, assay sensitivity means that had the study included a placebo, a control drug-placebo difference of at least M_1 would have been present."

M_1 in this context refers to the difference between the active control and placebo expected to occur in a clinical trial based on placebo-controlled studies of the active control conducted in the past. While the non-inferiority guidance does not deal explicitly with three-arm trials such as those under consideration here, this definition of assay sensitivity is useful for its emphasis on the importance of the size of the active control effect and not just its statistical significance.

Because statistical significance was the criterion both the sponsor and the FDA reviewers used to determine assay sensitivity, the ability to interpret the p-value is important. The sponsor's tests of assay sensitivity were performed relative to the primary endpoint in each trial. The performance of the active control with respect to other endpoints is also of interest, but assigning significance to those comparisons is problematic. Similarly, the test of assay sensitivity was performed using the same analysis method that was used for the primary comparison (test drug vs. placebo), and it would be difficult to justify changing the analysis method for reasons other than it being incorrect. Assessing the significance level of post-hoc analyses requires a multiplicity adjustment, which is not straightforward to calculate when there are numerous methods that could potentially have been applied.

The question of whether a significant difference between the active control and test drug, when neither is significantly different from placebo, constitutes evidence of assay sensitivity is questionable due to the multiplicity considerations discussed above. Two of the trials under dispute fall into this category and are discussed further below.

Specific comments on the assay sensitivity of the gepirone trials are as follows:

- CN105-078 and CN105-083 were terminated early by the sponsor at that time (Bristol-Meyers-Squibb) for business reasons. Both trials have sample sizes approximately ½ that of other completed studies, and neither includes an active control. The sponsor and the agency have agreed that both trials fail to have assay sensitivity due to early termination. I view these trials as having relevant information about the effect of gepirone but without sufficient power to detect a significant difference. It should also be noted that when a trial is terminated early based on an interim futility analysis, there is a tendency for the effect at interim to be under-estimated, just as there is a tendency for the effect at interim of a trial that is terminated early for evidence of efficacy to be over-estimated. When summarizing evidence from these trials through a proper meta-analysis, this tendency should be taken into consideration.
- CN105-053 was also terminated early by the sponsor but does include an active control (imipramine). By the primary analysis method, assay sensitivity was not established (active control effect on the HAMD-17 scale = -2.5; p = 0.14), however, the appropriateness of that analysis method is questionable. The sponsor's method included treatment by site interaction terms in the model, and without those terms, the results are quite different (active control effect = -3.19; p = 0.038). The primary comparison (of gepirone versus placebo) also changes between methods (from -0.70 with p = 0.69 to -2.0 with p = 0.19). In this case, a change of analysis method appears to be justified for the following reason. Including treatment by site interactions in a model and then averaging across sites to produce comparisons between treatment groups, using weights proportional to the number of patients in each group, is a valid approach and similar to the averaging that occurs implicitly when the interaction terms are

omitted. In this trial, there are only two sites, one has twice as many patients as the other, and the responses within treatment groups vary between the two. It appears that the sponsor's analysis gave equal weight to the two sites in computing effects for both gepirone and imipramine, which resulted in under-estimates of both effects, though I am unable to confirm this. I prefer an analysis that gives equal weight to patients rather than sites, and FDA's review accomplished this by omitting the interaction terms. Applying this method yields both assay sensitivity and a positive trend in favor of gepirone relative to placebo.

- ORG 134004, 134006, and 134017 each included active controls (paroxetine or fluoxetine), and each used a different primary endpoint than the positive and other failed trials (HAMD-25 or MADRS). Based on the primary endpoint and primary analysis method, none of these three trials had assay sensitivity. Additional details for each trial are as follows:
    - FDA's re-analysis of 134006 was based on a secondary endpoint (HAMD-17) and different analysis method, and this combination of changes yielded a larger difference between active control and placebo of -1.63 on the HAMD-17 scale ($p = 0.026$) compared to the primary sensitivity comparison (-1.37 on the HAMD-25 scale; $p = 0.17$). The sponsor's analysis based the comparisons on just those patients in the two groups being compared, while the FDA approach based the comparisons on the three groups combined, resulting in a different variance estimate (under the assumption of equal variances among groups). Unlike the CN105-53 case, the sponsor's analysis method seems reasonable (e.g., one could take the view that the active control patients are only of interest if the test drug fails to show superiority to placebo, therefore, including them in the analysis occurred in a staged fashion). It is difficult to interpret the p-value for the comparison based on a secondary endpoint with an alternative analysis method as providing evidence of assay sensitivity.
    - The analysis method differs between FDA and the sponsor in 134004 as well, but neither method yielded a significant difference between active control and placebo with respect to the primary or secondary endpoint (HAMD-25 or HAMD-17), indicating a lack of assay sensitivity. FDA's analysis of the secondary endpoint yielded a difference between gepirone and fluoxetine of 1.71 ($p = 0.027$), but this result is difficult to interpret due to multiplicity issues, given that neither was distinguishable from placebo.
    - In 134017, the only difference in the sponsor's and FDA's analysis is the use of the primary (MADRS) versus secondary (HAMD-17) endpoint. The analysis methods were the same, and neither analysis yielded a significant difference between active control and placebo, indicating a lack of assay sensitivity. The FDA's analysis yielded a difference between gepirone and fluoxetine of 1.55 ($p = 0.042$), but ambiguity remains as to how to interpret this difference for reasons given above.

In summary, the three trials with primary endpoint other than HAMD-17 (134004, 134006, and 134017) showed no evidence of assay sensitivity based on the sponsor's analysis, i.e., the active control was not different from placebo based on the primary endpoint and primary analysis method. Secondary endpoints or analyses yielded additional results that differ somewhat, but none provided unambiguous evidence of assay sensitivity. In contrast, trial CNS105-53 had assay sensitivity with respect to the

primary endpoint under a more appropriate analysis method. Trials CNS105-078 and -083 are considered failed trials due to early stopping, which resulted in insufficient power to detect a positive treatment effect and probable under-estimation of the effect at interim, but neither included an active control, so determining assay sensitivity was not possible.

In the above discussion, I have provided my interpretation of the results of the various assay sensitivity analyses conducted for the gepirone trials. It should be noted, however, that while assay sensitivity may be useful in trying to discern why some trials were able to show positive effects of gepirone and others were not, I do not believe it should be the sole basis for deciding which trials to include in a synthesis of the information provided by the trials. This is the topic of the next section.

Counting trials:

Much of the sponsor's dispute rests on the interpretation of a meta-analysis in which multiple trials are treated as independent Bernoulli trials with an outcome of success or failure. The result is used to assign a 'Type I error' to the sequence of trials, which is then compared to FDA's standard for substantial evidence (i.e., a one-sided p-value less than 0.025 that is replicated across two independent trials). The number of trials to include in the Bernoulli trial meta-analysis is disputed due to disagreement on which trials had assay sensitivity. In my view, the use of such a meta-analysis for this purpose is not particularly helpful in evaluating evidence of effectiveness, as explained in the following paragraphs.

Classifying studies as positive or negative and counting successes and failures, from a statistical point of view, represents an inefficient use of the information available from those trials. This method was common practice 40 years ago, and the widespread criticism of it motivated the development of a wide array of meta-analytic methodology to take its place. In a Bernoulli trial meta-analysis, all positive results, whether barely or highly significant, are considered the same, losing important information about the effect of a drug. Further, all non-significant results are considered equivalently negative, even though nearly-significant results could be seen by proper meta-analytic methods to add to, rather than detract from, the evidence from positive studies that a drug is effective. Summarizing a series of trials in this way can misrepresent the evidence of effectiveness provided by the trials, making the approach not particularly useful in evaluating new drugs.

It is correct that by applying the Bernoulli meta-analysis technique, we can say that 2 positive studies out of 2 gives a very small (<0.001) "overall p-value" whereas 2 out of 12 gives an "overall p-value" that is not so dramatically small (< 0.05). I agree with the sponsor that based on this approach, the hypothesis that gepirone has a true effect of zero in all of the trials is implausible. I also believe that a proper synthesis of the results from the trials provides more relevant information for evaluating gepirone than a simple counting of positive and negative trials. Further, determining which trials to include in such a synthesis should be based on factors other than assay sensitivity, namely, whether the candidate trials are sufficiently homogenous to support combining their data or results with appropriate statistical methods.

Meta-analyses of the trials that take into account the magnitude of the observed effects rather than just reducing the results to success or failure provide some indication of the expected effect gepirone will have, if approved. In the 2007 submission, the sponsor provided results of a meta-analysis of all 12 trials

under consideration as well as for the subset that included two positive and five negative trials (seven trials total). The average treatment effect was estimated as -1.22 (SE = 0.39; p = 0.002) for the latter and -0.48 (SE = 0.28; p= 0.093) for the former. The FDA statistical reviewer conducted an additional meta-analysis of the 9 trials that specified HAMD-17 as the primary endpoint, regardless of their assay sensitivity status. In addition to the seven trials included in the sponsor's meta-analysis, CNS105-052 and -053 were included. The average effect, based on these nine trials was similar to that based on the seven trials, i.e., -1.26 (SE = 0.37; p = 0.0093).

With the 7-trial and 9-trial meta-analyses, both of which could be argued as relevant, the summary effect is approximately half that observed in the two positive trials, which is not surprising, given that several negative and at least one failed trial were included. My interpretation of this synthesis is that gepirone is effective, but the effectiveness varies under some conditions of use or in some patients.

Maintenance trial: The failure of the maintenance trial appeared to weigh heavily in the letter upholding the agency's non-approval decision. Based on the information provided, it appears the trial suffered from a number of quality issues that negatively impacted study conduct and the ability of the study to be successful. Several sites included patients in only one of the treatment arms, and several investigators appeared to have difficulty implementing the protocol, resulting in a large number of protocol violations. The sponsor's primary analysis was incorrect due to exclusion of patients from the sites with only one treatment arm as well as a few patients known to have relapsed but excluded for other reasons. FDA's re-analysis that includes all patients from all of the sites shows a favorable trend in the primary endpoint, but this evidence requires support from another trial to justify use of gepirone as a maintenance therapy.

Conclusion: Gepirone has been shown to be effective in two positive phase 3 trials of reasonable size and quality and with similar results. But our statute requires substantial evidence that the drug "will have the effect it purports or is represented to have" [Food Drug and Cosmetic Act 21 USC § 355(d)]. The inconsistency in effect across all of the trials seems to provide evidence that at least some of the time, gepirone will not have its claimed effect, but the sponsor has not provided any reasonable argument for identifying patients for whom the drug will more reliably work. The difference in average effect between the meta-analysis of all trials and the meta-analyses excluding the three trials with HAMD-25 as the primary endpoint may point to other aspects of these three trials, e.g., different patient populations or use of different clinical assessment methods, that may have negatively impacted the results. This information may be useful in identifying the patient population most likely to benefit from gepirone, and under what conditions the drug should be used.

This revised memorandum is an addendum to the original memo dated June 8, 2015. The original date of the memo still stands and is valid. The new revised memo is archived in CDER's document archiving, reporting and regulatory tracking system (DARRTS) on Nov. 10, 2015.

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

/s/

KHUSHBOO SHARMA
11/10/2015

5.18 Reference to Article – "Regulatory and Scientific Issues in Studies to Evaluate Sexual Dysfunction in Antidepressant Drug Trials (08/2015)

Khin N, Kronstein P, Yang P, et al. Regulatory and scientific issues in studies to evaluate sexual dysfunction in antidepressant drug trials. *J Clin Psychiatry*. 2015; 76(8): 1060 – 1063.

5.19 Reference to Article – "Summary of Findings from the FDA Regulatory Science Forum On Measuring Sexual Dysfunction in Depression Trials (08/2015)

Kronstein P, Ishida E, Khin N, et al. Summary of Findings From the FDA Regulatory Science Forum On Measuring Sexual Dysfunction. *J Clin Psychiatry*. 2015; 76(8): 1050 – 1059.

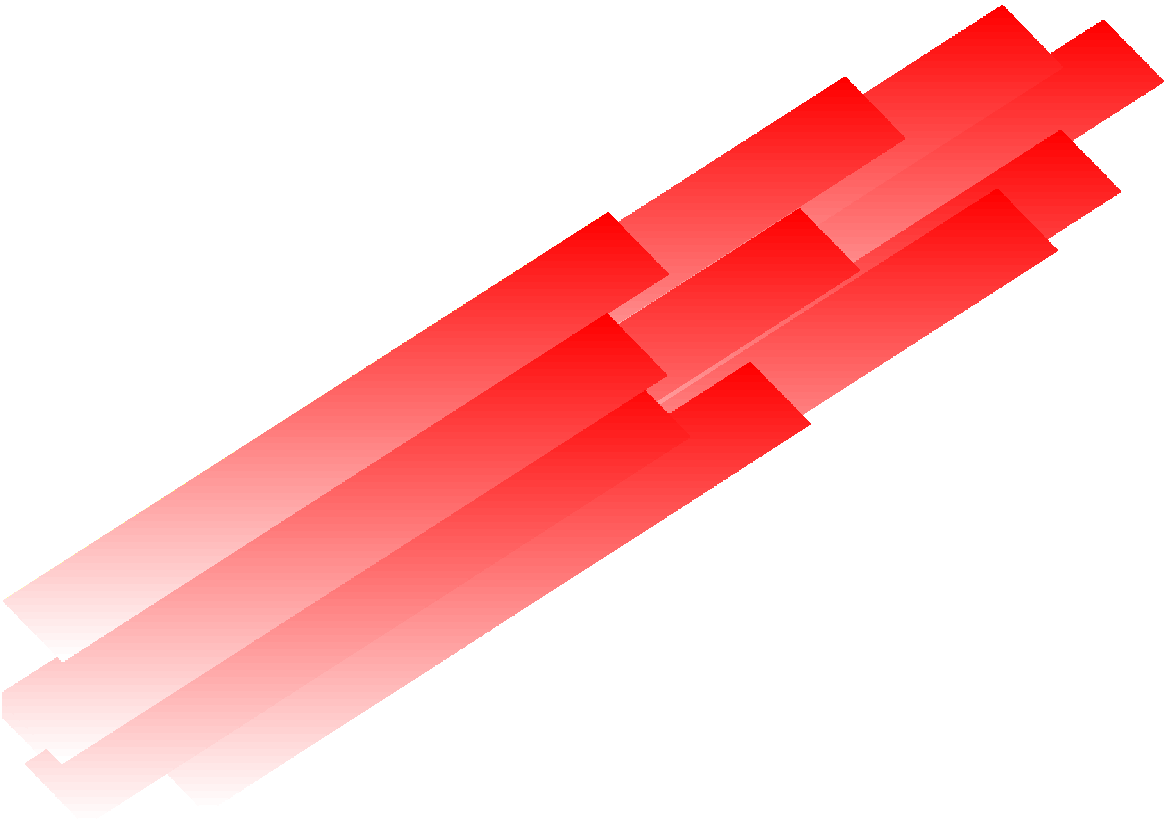
5.20 Reference to Article – "Vilazodone: Clinical Basis for the US Food and Drug Administration's Approval of a New Antidepressant (11/2011).

Laughren T, Gobburu J, Temple R, et al. Vilazodone: Clinical Basis for the US Food and Drug Administration's Approval of a New Antidepressant. *J Clin Psychiatry*. 2011; 72(9): 1166 – 1173.

5.21 Guidance for Industry – Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products (05/1998)

Guidance for Industry

Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products



U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
May 1998
Clinical 6

Guidance for Industry

Providing Clinical Evidence of Effectiveness for Human Drugs and Biological Products

Additional copies are available from:
the Drug Information Branch (HFD-210),
Center for Drug Evaluation and Research (CDER),
5600 Fishers Lane, Rockville, MD 20857 (Tel) 301-827-4573
Internet at http://www.fda.gov/cder/guidance/index.htm
or
Office of Communication,
Training, and Manufacturers Assistance (HFM-40)
Center for Biologics Evaluation and Research (CBER)
1401 Rockville Pike, Rockville, MD 20852-1448
http://www.fda.gov/cber/guidelines.htm
(Fax) 888-CBERFAX or 301-827-3844
(Voice Information) 800-835-4709 or 301-827-1800

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
May 1998
Clinical 6

TABLE OF CONTENTS

I. INTRODUCTION 1

II. QUANTITY OF EVIDENCE NECESSARY TO SUPPORT EFFECTIVENESS 2

A. Legal Standards for Drug and Biological Products 2

B. Scientific Basis for the Legal Standard 4

C. The Quantity of Evidence to Support Effectiveness 6

III. DOCUMENTATION OF THE QUALITY OF EVIDENCE SUPPORTING AN EFFECTIVENESS CLAIM 16

A. Reliance on Less Than Usual Access to Clinical Data or Detailed Study Reports 17

B. Reliance on Studies with Alternative, Less Intensive Quality Control/On-Site Monitoring 20

GUIDANCE FOR INDUSTRY[1]

Providing Clinical Evidence of Effectiveness[2] for Human Drug and Biological Products

I. INTRODUCTION

This document is intended to provide guidance to applicants planning to file new drug applications (NDAs), biologics license applications (BLAs), or applications for supplemental indications on the evidence to be provided to demonstrate effectiveness.

This document is also intended to meet the requirements of subsections 403(b)(1) and (2) of the Food and Drug Administration Modernization Act (the Modernization Act) of 1997 for human drug and biological products (P.L. 105-115).[3] Subsection 403(b)(1) directs FDA to provide guidance on the circumstances in which published matter may be the basis for approval of a supplemental application for a new indication. Section III of this guidance satisfies this requirement by describing circumstances in which published matter may partially or entirely support approval of a supplemental application. Subsection 403(b)(2) directs FDA to provide guidance on data requirements that will avoid duplication of previously submitted data by recognizing the availability of data previously submitted in support of an original application to support approval of a supplemental application. Section II of this guidance satisfies this requirement by describing a range of circumstances in which related existing data, whether from an original application or other sources, may be used to support approval of a supplemental application.

In 1962, Congress amended the Federal Food, Drug, and Cosmetic Act to add a requirement that, to obtain marketing approval, manufacturers demonstrate the effectiveness of their products through the conduct of adequate and well-controlled studies. Since then, the issue of what constitutes sufficient evidence of effectiveness has been debated by the Agency, the scientific community, industry, and others. Sound evidence of effectiveness is a crucial component of the Agency's benefit-risk assessment of a new product or use. At the same time, the demonstration of effectiveness represents a major component of drug development time and cost; the amount

---

[1] This guidance document represents the agency's current thinking on providing clinical evidence of effectiveness for human drug and biological products. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statute, regulations, or both.

[2] As used in this guidance, the term efficacy refers to the findings in an adequate and well-controlled clinical trial or the intent of conducting such a trial and the term effectiveness refers to the regulatory determination that is made on the basis of clinical efficacy and other data.

[3] The Modernization Act requirements in Section 403 also apply to animal drugs and medical devices. These products will be addressed in separate guidances.

and nature of the evidence needed can therefore be an important determinant of when and whether new therapies become available to the public. The public health is best served by the development of sound evidence of effectiveness in an efficient manner.

The science and practice of drug development and clinical evaluation have evolved significantly since the effectiveness requirement for drugs was established, and this evolution has implications for the amount and type of data needed to support effectiveness in certain cases. As a result of medical advances in the understanding of pathogenesis and disease staging, it is increasingly likely that clinical studies of drugs will be more narrowly defined to focus, for example, on a more specific disease stage or clinically distinct subpopulation. As a consequence, product indications are often narrower, the universe of possible indications is larger, and data may be available from a number of studies of a drug in closely related indications that bear on a determination of its effectiveness for a new use. Similarly, there may be studies of a drug in different populations, studies of a drug alone or in combination, and studies of different doses and dosage forms, all of which may support a particular new use of a drug. At the same time, progress in clinical evaluation and clinical pharmacology have resulted in more rigorously designed and conducted clinical efficacy trials, which are ordinarily conducted at more than one clinical site. This added rigor and scope has implications for a study's reliability, generalizability, and capacity to substantiate effectiveness.

Given this evolution, the Agency has determined that it would be appropriate to articulate its current thinking concerning the quantitative and qualitative standards for demonstrating effectiveness of drugs and biologics. FDA hopes that this guidance will enable sponsors to plan drug development programs that are sufficient to establish effectiveness without being excessive in scope. The guidance should also bring greater consistency and predictability to FDA's assessment of the clinical trial data needed to support drug effectiveness.

Another major goal of this guidance is to encourage the submission of supplemental applications to add new uses to the labeling of approved drugs. By articulating how it currently views the quantity and quality of evidence necessary to support approval of a new use of a drug, FDA hopes to illustrate that the submission of supplements for new uses need not be unduly burdensome.

II. QUANTITY OF EVIDENCE NECESSARY TO SUPPORT EFFECTIVENESS

A. Legal Standards for Drug and Biological Products

Drugs: The effectiveness requirement for drug approval was added to the Federal Food, Drug, and Cosmetic Act (the Act or the FDC Act) in 1962. Between passage of the Act in 1938 and the 1962 amendments, drug manufacturers were required to show only that their drugs were safe. The original impetus for the effectiveness requirement was Congress's growing concern about the misleading and unsupported claims being made by pharmaceutical companies about their drug products coupled with high drug prices. After two years of hearings on these issues, Congress adopted the 1962 Drug Amendments,

which included a provision requiring manufacturers of drug products to establish a drug's effectiveness by "substantial evidence." *Substantial evidence* was defined in section 505(d) of the Act as "evidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof."

Since the 1962 Amendments added this provision to the statute, discussions have ensued regarding the quantity and quality of the evidence needed to establish effectiveness. With regard to quantity, it has been FDA's position that Congress generally intended to require at least two adequate and well-controlled studies, each convincing on its own, to establish effectiveness. (See e.g., Final Decision on Benylin, 44 FR 51512, 518 (August 31, 1979); *Warner-Lambert Co. V. Heckler*, 787 F. 2d 147 (3d Cir. 1986)). FDA's position is based on the language in the statute[4] and the legislative history of the 1962 amendments. Language in a Senate report suggested that the phrase "adequate and well-controlled investigations" was designed not only to describe the quality of the required data but the "quantum" of required evidence. (S. Rep. No. 1744, Part 2, 87th Cong. 2d Sess. 6 (1962))

Nevertheless, FDA has been flexible within the limits imposed by the congressional scheme, broadly interpreting the statutory requirements to the extent possible where the data on a particular drug were convincing. In some cases, FDA has relied on pertinent information from other adequate and well-controlled studies of a drug, such as studies of other doses and regimens, of other dosage forms, in other stages of disease, in other populations, and of different endpoints, to support a single adequate and well-controlled study demonstrating effectiveness of a new use. In these cases, although there is only one study of the exact new use, there are, in fact, multiple studies supporting the new use, and expert judgment could conclude that the studies together represent substantial evidence of effectiveness. In other cases, FDA has relied on only a single adequate and well-controlled efficacy study to support approval — generally only in cases in which a single multicenter study of excellent design provided highly reliable and statistically strong evidence of an important clinical benefit, such as an effect on survival, and a confirmatory study would have been difficult to conduct on ethical grounds.

In section 115(a) of the Modernization Act, Congress amended section 505(d) of the Act to make it clear that the Agency may consider "data from one adequate and well-controlled clinical investigation and confirmatory evidence" to constitute substantial

[4] Section 505(d) of the Act uses the plural form in defining "substantial evidence" as "adequate and well-controlled investigations, including clinical investigations." See also use of "investigations" in section 505(b) of the Act, which lists the contents of a new drug application.

evidence if FDA determines that such data and evidence are sufficient to establish effectiveness. In making this clarification, Congress confirmed FDA's interpretation of the statutory requirements for approval and acknowledged the Agency's position that there has been substantial progress in the science of drug development resulting in higher quality clinical trial data.

Biologics. Biological products are approved under authority of section 351 of the Public Health Service Act (PHS Act) (42 U.S.C.§ 262). Under section 351, as in effect since 1944, licenses for biologics have been issued only upon a showing that the products meet standards designed to ensure the "continued safety, purity, and potency" of the products. *Potency* has long been interpreted to include effectiveness (21 CFR 600.3(s)). In 1972, FDA initiated a review of the safety and effectiveness of all previously licensed biologics. The Agency stated then that proof of effectiveness would consist of controlled clinical investigations as defined in the provision for "adequate and well-controlled studies" for new drugs (21 CFR 314.126), unless waived as not applicable to the biological product or essential to the validity of the study when an alternative method is adequate to substantiate effectiveness (21 CFR 601.25 (d) (2)). One such adequate alternative was identified to be serological response data where a previously accepted correlation with clinical effectiveness exists. As with nonbiological drug products, FDA has approved biological products based on single, multicenter studies with strong results.

Although section 123(a) of the Modernization Act amended section 351 of the PHS Act to make it clear that separate licenses are not required for biological products and the establishments at which the products are made, the evidentiary standard for a biological product was not changed: the product must be shown to be "safe, pure, and potent" (section 351 (a)(2) of the PHS Act as amended). In the Modernization Act (section 123(f)) Congress also directed the agency to take measures to "minimize differences in the review and approval" of products required to have approved BLAs under section 351 of the PHS Act and products required to have approved NDAs under section 505(b)(1) of the FDC Act.

B. Scientific Basis for the Legal Standard

The usual requirement for more than one adequate and well-controlled investigation reflects the need for *independent substantiation* of experimental results. A single clinical experimental finding of efficacy, unsupported by other independent evidence, has not usually been considered adequate scientific support for a conclusion of effectiveness. The reasons for this include the following.

- Any clinical trial may be subject to unanticipated, undetected, systematic biases. These biases may operate despite the best intentions of sponsors and investigators, and may lead to flawed conclusions. In addition, some investigators may bring conscious biases to evaluations.

- The inherent variability in biological systems may produce a positive trial result by chance alone. This possibility is acknowledged, and quantified to some extent, in the statistical evaluation of the result of a single efficacy trial. It should be noted, however, that hundreds of randomized clinical efficacy trials are conducted each year with the intent of submitting favorable results to FDA. Even if all drugs tested in such trials were ineffective, one would expect one in forty of those trials to "demonstrate" efficacy by chance alone at conventional levels of statistical significance.[5] It is probable, therefore, that false positive findings (i.e., the chance appearance of efficacy with an ineffective drug) will occur and be submitted to FDA as evidence of effectiveness. Independent substantiation of a favorable result protects against the possibility that a chance occurrence in a single study will lead to an erroneous conclusion that a treatment is effective.

- Results obtained in a single center may be dependent on site or investigator specific factors (e.g., disease definition, concomitant treatment, diet). In such cases, the results, although correct, may not be generalizable to the intended population. This possibility is the primary basis for emphasizing the need for independence in substantiating studies.

- Rarely, favorable efficacy results are the product of scientific fraud.

Although there are statistical, methodologic, and other safeguards to address the identified problems, they are often inadequate to address these problems in a single trial. Independent substantiation of experimental results addresses such problems by providing consistency across more than one study, thus greatly reducing the possibility that a biased, chance, site-specific, or fraudulent result will lead to an erroneous conclusion that a drug is effective.

The need for independent substantiation has often been referred to as the need for replication of the finding. Replication may not be the best term, however, as it may imply that precise repetition of the same experiment in other patients by other investigators is the only means to substantiate a conclusion. Precise replication of a trial is only one of a number of possible means of obtaining independent substantiation of a clinical finding and, at times, can be less than optimal as it could leave the conclusions vulnerable to any systematic biases inherent to the particular study design. Results that are obtained from studies that are of different design and independent in execution, perhaps evaluating different populations, endpoints, or dosage forms, may provide support for a conclusion of effectiveness that is as convincing as, or more convincing than, a repetition of the same study.

[5] p-value = 0.05, two-tailed, which implies an error rate in the efficacy (false positive) tail of 0.025 or one in forty.

C. The Quantity of Evidence to Support Effectiveness

The following three sections provide guidance on the quantity of evidence needed in particular circumstances to establish substantial evidence of effectiveness. Section 1 addresses situations in which effectiveness of a new use may be extrapolated entirely from existing efficacy studies. Section 2 addresses situations in which a single adequate and well-controlled study of a specific new use can be supported by information from other related adequate and well-controlled studies, such as studies in other phases of a disease, in closely related diseases, of other conditions of use (different dose, duration of use, regimen), of different dosage forms, or of different endpoints. Section 3 addresses situations in which a single multicenter study, without supporting information from other adequate and well-controlled studies, may provide evidence that a use is effective.

In each of these situations, it is assumed that any studies relied on to support effectiveness meet the requirements for adequate and well-controlled studies in 21 CFR 314.126. It should also be appreciated that reliance on a single study of a given use, whether alone or with substantiation from related trial data, leaves little room for study imperfections or contradictory (nonsupportive) information. In all cases, it is presumed that the single study has been appropriately designed, that the possibility of bias due to baseline imbalance, unblinding, post-hoc changes in analysis, or other factors is judged to be minimal, and that the results reflect a clear prior hypothesis documented in the protocol. Moreover, a single favorable study among several similar attempts that failed to support a finding of effectiveness would not constitute persuasive support for a product use unless there were a strong argument for discounting the outcomes in the studies that failed to show effectiveness (e.g., study obviously inadequately powered or lack of assay sensitivity as demonstrated in a three-arm study by failure of the study to show efficacy of a known active agent).

Whether to rely on a single study to support an effectiveness determination is not often an issue in contemporary drug development. In most drug development situations, the need to find an appropriate dose, to study patients of greater and lesser complexity or severity of disease, to compare the drug to other therapy, to study an adequate number of patients for safety purposes, and to otherwise know what needs to be known about a drug before it is marketed will result in more than one adequate and well-controlled study upon which to base an effectiveness determination.

This guidance is not intended to provide a complete listing of the circumstances in which existing efficacy data may provide independent substantiation of related claims; rather, it provides examples of the reasoning that may be employed. The examples are applicable whether the claim arises in the original filing of an NDA or BLA, or in a supplemental application.

1. Extrapolation from Existing Studies

In certain cases, effectiveness of an approved drug product for a new indication, or effectiveness of a new product, may be adequately demonstrated without additional adequate and well-controlled clinical efficacy trials. Ordinarily, this will be because other types of data provide a way to apply the known effectiveness to a new population or a different dose, regimen or dosage form. The following are examples of situations in which effectiveness might be extrapolated from efficacy data for another claim or product.

a. Pediatric uses

The rule revising the Pediatric Use section of product labeling (21 CFR 201.57(f)(9)(iv)) makes allowance for inclusion of pediatric use information in labeling without controlled clinical trials of the use in children. In such cases, a sponsor must provide other information to support pediatric use, and the Agency must conclude that the course of the disease and the effects of the drug are sufficiently similar in the pediatric and adult populations to permit extrapolation from adult efficacy data to pediatric patients. Evidence that could support a conclusion of similar disease course and similar drug effect in adult and pediatric populations includes evidence of common pathophysiology and natural history of the disease in the adult and pediatric populations, evidence of common drug metabolism and similar concentration-response relationships in each population, and experience with the drug, or other drugs in its therapeutic class, in the disease or condition or related diseases or conditions. Examples in which pediatric use labeling information has been extrapolated from adult efficacy data include ibuprofen for pain and loratidine for seasonal allergic rhinitis.

b. Bioequivalence

The effectiveness of alternative formulations and new dosage strengths may be assessed on the basis of evidence of bioequivalence.

c. Modified-release dosage forms

In some cases, modified release dosage forms may be approved on the basis of pharmacokinetic data linking the new dosage form to a previously studied immediate-release dosage form. Because the pharmacokinetic patterns of modified-release and immediate-release dosage forms are not identical, it is generally important to have some understanding of the relationship of blood concentration to response, including an understanding of the time course of that relationship, to extrapolate the immediate-release

data to the modified-release dosage form.

d. Different doses, regimens, or dosage forms

Dose-response relationships are generally continuous such that information about the effectiveness of one dose, dosage regimen, or dosage form is relevant to the effectiveness of other doses, regimens, or dosage forms. Where blood levels and exposure are not very different, it may be possible to conclude that a new dose, regimen, or dosage form is effective on the basis of pharmacokinetic data alone. Even if blood levels are quite different, if there is a well-understood relationship between blood concentration and response, including an understanding of the time course of that relationship, it may be possible to conclude that a new dose, regimen, or dosage form is effective on the basis of pharmacokinetic data without an additional clinical efficacy trial. In this situation, pharmacokinetic data, together with the well-defined pharmacokinetic/pharmacodynamic (PK/PD) relationship, are used to translate the controlled trial results from one dose, regimen, or dosage form to a new dose, regimen, or dosage form (See also section II.C.2.a).

2. Demonstration of Effectiveness by a Single Study of a New Use, with Independent Substantiation From Related Study Data

The discussion that follows describes specific examples in which a single study of a new use, with independent substantiation from study data in related uses, could provide evidence of effectiveness. In these cases, the study in the new use and the related studies support the conclusion that the drug has the effect it is purported to have. Whether related studies are capable of substantiating a single study of a new use is a matter of judgment and depends on the quality and outcomes of the studies and the degree of relatedness to the new use.

a. Different doses, regimens, or dosage forms

As discussed in Sections II.C.1.d, it may be possible to conclude that a new dose, regimen, or dosage form is effective on the basis of pharmacokinetic data without an additional clinical efficacy trial where blood levels and exposure are not very different or, even if quite different, there is a well-understood relationship between blood concentration and response. Where the relationship between blood concentration and response is not so well understood and the pharmacokinetics of the new dose, regimen, or dosage form differ from the previous one, clinical efficacy data will likely be necessary to support effectiveness of a new regimen. In this case, a single additional efficacy study should ordinarily be sufficient. For example, a single controlled trial was needed to support the recent approval of a once

daily dose of risperidone because the once daily and twice daily regimens had different pharmacokinetics and risperidone's PK/PD relationship was not well understood.

b. Studies in other phases of the disease

In many cases, therapies that are effective in one phase of a disease are effective in other disease phases, although the magnitude of the benefit and benefit-to-risk relationship may differ in these other phases. For example, if a drug is known to be effective in patients with a refractory stage of a particular cancer, a single adequate and well-controlled study of the drug in an earlier stage of the same tumor will generally be sufficient evidence of effectiveness to support the new use.

c. Studies in other populations

Often, responses in subsets of a particular patient population are qualitatively similar to those in the whole population. In most cases, separate studies of effectiveness in demographic subsets are not needed (see also discussion of the pediatric population in section II.C.1.a) However, where further studies are needed, a single study would ordinarily suffice to support effectiveness in age, race, gender, concomitant disease, or other subsets for a drug already shown to be generally effective in a condition or to be effective in one population. For example, a single study was sufficient to support tamoxifen use in breast cancer in males.

d. Studies in combination or as monotherapy

For a drug known to be effective as monotherapy, a single adequate and well-controlled study is usually sufficient to support effectiveness of the drug when combined with other therapy (as part of a multidrug regimen or in a fixed-dose combination). Similarly, known effectiveness of a drug as part of a combination (i.e., its contribution to the effect of the combination is known) would usually permit reliance on a single study of appropriate design to support its use as monotherapy, or as part of a different combination, for the same use. For example, a single study of a new combination vaccine designed to demonstrate adequate immune response will ordinarily provide sufficient evidence of effectiveness if the new combination contains products or antigens already proven to be effective alone or in other combinations. These situations are common for oncologic and antihypertensive drugs, but occur elsewhere as well.

e. Studies in a closely related disease

Studies in etiologically or pathophysiologically related conditions, or studies of a symptom common to several diseases (e.g., pain) can support each other, allowing initial approval of several uses or allowing additional claims based on a single adequate and well-controlled study. For example, certain anti-coagulant or anti-platelet therapies could be approved for use in two different settings based on individual studies in unstable angina/acute coronary syndrome and in the postangioplasty state. Because the endpoints studied and the theoretical basis for use of an anti-coagulant or anti-platelet drug are similar, each study supports the other for each claim. Similarly, single analgesic studies in several painful conditions would ordinarily be sufficient to support either a general analgesic indication or multiple specific indications. The recent approval of lamotrigine for treatment of Lennox-Gastaut Syndrome (a rare, largely pediatric, generalized seizure disorder) was based on a single adequate and well-controlled trial, due in part to related data showing efficacy of the drug in partial-onset seizures in adults.

f. Studies in less closely related diseases, but where the general purpose of therapy is similar

Certain classes of drug therapy, such as antimicrobials and antineoplastics, are appropriate interventions across a range of different diseases. For therapies of this type, evidence of effectiveness in one disease could provide independent substantiation of effectiveness in a quite different disease. For example, it is possible to argue that evidence of effectiveness of an antimicrobial in one infectious disease setting may support reliance on a single study showing effectiveness in other settings where the causative pathogens, characteristics of the site of infection that affect the disease process (e.g., structure and immunology) and patient population are similar.[6] Similarly, for an oncologic drug, evidence of effectiveness in one or more tumor types may support reliance on a single study showing effectiveness against a different kind of tumor, especially if the tumor types have a common biological origin.

g. Studies of different clinical endpoints

Demonstration of a beneficial effect in different studies on two different clinically meaningful endpoints could cross-substantiate a claim for

[6] See Division of Anti-Infective Drug Products: Points to Consider in the Clinical Development and Labeling of Anti-Infective Drug Products, October 1992.

effectiveness for each outcome. For example, the initial claim for effectiveness of enalapril for heart failure was supported by one study showing symptom improvement over several months and a second study showing improved survival in a more severely ill population. The two different findings, each from an adequate and well-controlled study, led to the conclusion that enalapril was effective in both treating symptoms and improving survival.

h. Pharmacologic/pathophysiologic endpoints

When the pathophysiology of a disease and the mechanism of action of a therapy are very well understood, it may be possible to link specific pharmacologic effects to a strong likelihood of clinical effectiveness. A pharmacologic effect that is accepted as a validated surrogate endpoint can support ordinary approval (e.g., blood pressure effects, cholesterol-lowering effects) and a pharmacologic effect that is considered reasonably likely to predict clinical benefit can support accelerated approval under the conditions described in 21 CFR 314 Subpart H and 21 CFR 601 Subpart E (e.g., CD4 count and viral load effects to support effectiveness of anti-viral drugs for HIV infection). When the pharmacologic effect is not considered an acceptable effectiveness endpoint, but the linkage between it and the clinical outcome is strong, not merely on theoretical grounds but based on prior therapeutic experience or well-understood pathophysiology, a single adequate and well-controlled study showing clinical efficacy can sometimes be substantiated by persuasive data from a well-controlled study or studies showing the related pharmacologic effect.

For example, a single clearly positive trial can be sufficient to support approval of a replacement therapy such as a coagulation factor, when it is combined with clear evidence that the condition being treated is caused by a deficiency of that factor. Demonstration of physical replacement of the deficient factor or restoration of the missing physiologic activity provides strong substantiation of the clinical effect. The corrective treatment of an inborn error of metabolism could be viewed similarly. In the case of preventive vaccines, one adequate and well-controlled clinical trial may be supported by compelling animal challenge/protection models, human serological data, passive antibody data, or pathogenesis information. The more evidence there is linking effects on the pharmacologic endpoint to improvement or prevention of the disease, the more persuasive the argument for reliance on a single clinical efficacy study.

Note, however, that plausible beneficial pharmacologic effects have often not correlated with clinical benefit, and, therefore, caution must be observed in relying on a pharmacologic effect as contributing to evidence

of effectiveness. For example, pharmacologic effects such as arrhythmia suppression by Type 1 antiarrhythmics and increased cardiac output by phosphodiesterase inhibitors or beta adrenergic inotropes resulted in increased mortality, rather than, as was expected, decreased sudden death and improved outcome in heart failure. The reasons for the absence of an expected correlation between pharmacologic and clinical effects are diverse and can include an incompletely understood relationship between the pharmacologic effect and the clinical benefit and the presence of other pharmacologic effects attributable to a drug in addition to the effect being measured and thought to be beneficial. Generally, the utility of pharmacologic outcomes in providing independent substantiation will be greatest where there is prior experience with the pharmacologic class. Even in this case, however, it is difficult to be certain that a pharmacologic effect that correlates with a clinical benefit accounts for all the clinical benefit or that other effects are not present and relevant.

3. Evidence of Effectiveness from a Single Study

When the effectiveness requirement was originally implemented in 1962, the prevailing efficacy study model was a single institution, single investigator, relatively small trial with relatively loose blinding procedures, and little attention to prospective study design and identification of outcomes and analyses. At present, major clinical efficacy studies are typically multicentered, with clear, prospectively determined clinical and statistical analytic criteria. These studies are less vulnerable to certain biases, are often more generalizable, may achieve very convincing statistical results, and can often be evaluated for internal consistency across subgroups, centers, and multiple endpoints.

The added rigor and size of contemporary clinical trials have made it possible to rely, in certain circumstances, on a single adequate and well-controlled study, without independent substantiation from another controlled trial, as a sufficient scientific and legal basis for approval. For example, the approval of timolol for reduction of post-infarction mortality was based on a single, particularly persuasive (low p-value), internally consistent, multicenter study that demonstrated a major effect on mortality and reinfarction rate. For ethical reasons, the study was considered unrepeatable. The Center for Biologics Evaluation and Research has also approved a number of products based upon a single persuasive study. The Agency provided a general statement in 1995 describing when a single, multicenter study may suffice (60 FR 39181; August 1, 1995), but the Agency has not comprehensively described the situations in which a single adequate and well-controlled study might be considered adequate support for an effectiveness claim, or the characteristics of a single study that could make it adequate support for an effectiveness claim.

Whether to rely on a single adequate and well-controlled study is inevitably a matter of judgment. A conclusion based on two persuasive studies will always be more secure than a conclusion based on a single, comparably persuasive study. For this reason, reliance on only a single study will generally be limited to situations in which a trial has demonstrated a clinically meaningful effect on mortality, irreversible morbidity, or prevention of a disease with potentially serious outcome and confirmation of the result in a second trial would be practically or ethically impossible. For example, sequential repetition of strongly positive trials that demonstrated a decrease in post-infarction mortality, prevention of osteoporotic fractures, or prevention of pertussis would present significant ethical concerns. Repetition of positive trials showing only symptomatic benefit would generally not present the same ethical concerns.

The discussion that follows identifies the characteristics of a single adequate and well-controlled study that could make the study adequate support for an effectiveness claim. Although no one of these characteristics is necessarily determinative, the presence of one or more in a study can contribute to a conclusion that the study would be adequate to support an effectiveness claim.

a. Large multicenter study

In a large multicenter study in which (1) no single study site provided an unusually large fraction of the patients and (2) no single investigator or site was disproportionately responsible for the favorable effect seen, the study's internal consistency lessens concerns about lack of generalizability of the finding or an inexplicable result attributable only to the practice of a single investigator. If analysis shows that a single site is largely responsible for the effect, the credibility of a multicenter study is diminished.

b. Consistency across study subsets

Frequently, large trials have relatively broad entry criteria and the study populations may be diverse with regard to important covariates such as concomitant or prior therapy, disease stage, age, gender or race. Analysis of the results of such trials for consistency across key patient subsets addresses concerns about generalizability of findings to various populations in a manner that may not be possible with smaller trials or trials with more narrow entry criteria. For example, the timolol postinfarction study randomized patients separately within three severity strata. The study showed positive effects on survival in each stratum supporting a conclusion that the drug's utility was not limited to a particular disease stage (e.g., relatively low or high severity).

c. Multiple *studies* in a single study

Properly designed factorial studies may be analyzed as a series of pairwise comparisons, representing, within a single study, separate demonstrations of activity of a drug as monotherapy and in combination with another drug. This model was successfully used in ISIS II, which showed that for patients with a myocardial infarction both aspirin and streptokinase had favorable effects on survival when used alone and when combined (aspirin alone and streptokinase alone were each superior to placebo; aspirin and streptokinase in combination were superior to aspirin alone and to streptokinase alone). This represented two separate (but not completely independent) demonstrations of the effectiveness of aspirin and streptokinase.

d. Multiple endpoints involving different events

In some cases, a single study will include several important, prospectively identified primary or secondary endpoints, each of which represents a beneficial, but different, effect. Where a study shows statistically persuasive evidence of an effect on more than one of such endpoints, the internal weight of evidence of the study is enhanced. For example, the approval of beta-interferon (Betaseron) for prevention of exacerbations in multiple sclerosis was based on a single multicenter study, at least partly because there were both a decreased rate of exacerbations and a decrease in MRI-demonstrated disease activity — two entirely different, but logically related, endpoints.

Similarly, favorable effects on both death and nonfatal myocardial infarctions in a lipid-lowering, postangioplasty, or postinfarction study would, in effect, represent different, but consistent, demonstrations of effectiveness, greatly reducing the possibility that a finding of reduced mortality was a chance occurrence. For example, approval of abciximab as adjunctive treatment for patients undergoing complicated angioplasty or atherectomy was supported by a single study with a strong overall result on the combined endpoint (decreased the combined total of deaths, new infarctions, and need for urgent interventions) and statistically significant effects in separate evaluations of two components of the combined endpoint (decreased new infarctions and decreased need for urgent interventions). In contrast, a beneficial effect on multiple endpoints that evaluate essentially the same phenomenon and correlate strongly, such as mood change on two different depression scales or SGOT and CPK levels postinfarction, does not significantly enhance the internal weight of the evidence from a single trial.

Although two consistent findings within a single study usually provide reassurance that a positive treatment effect is not due to chance, they do not protect against bias in study conduct or biased analyses. For example, a treatment assignment not well balanced for important prognostic variables could lead to an apparent effect on both endpoints. Thus, close scrutiny of study design and conduct are critical to evaluating this type of study.

e. Statistically very persuasive finding

In a multicenter study, a very low p-value indicates that the result is highly inconsistent with the null hypothesis of no treatment effect. In some studies it is possible to detect nominally statistically significant results in data from several centers, but, even where that is not possible, an overall extreme result and significance level means that most study centers had similar findings. For example, the thrombolysis trials of streptokinase (ISIS II, GISSI) had very sizable treatment effects and very low p-values, greatly adding to their persuasiveness. Preventive vaccines for infectious disease indications with a high efficacy rate (e.g., point estimate of efficacy of 80% or higher and a reasonably narrow 95% confidence interval) have been approved based on a single adequate and well-controlled trial.

4. Reliance on a Single, Multicenter Study — Caveats

While acknowledging the persuasiveness of a single, internally consistent, strong multicenter study, it must be appreciated that even a strong result can represent an isolated or biased result, especially if that study is the only study suggesting efficacy among similar studies. Recently, the apparent highly favorable effect of vesnarinone, an inotropic agent, in heart failure (60% reduction of mortality in what appeared to be a well-designed, placebo-controlled, multicenter trial with an extreme p-value) has proven to be unrepeatable. In an attempt to substantiate the finding, the same dose of the drug that seemed lifesaving in the earlier study significantly increased mortality (by 26%), and a lower dose also appeared to have a detrimental effect on survival. Although the population in the second study was, on the whole, a sicker population than in the first, the outcomes in similarly sick patients in each study were inconsistent so this factor does not explain the contradictory results.

When considering whether to rely on a single multicenter trial, it is critical that the possibility of an incorrect outcome be considered and that all the available data be examined for their potential to either support or undercut reliance on a single multicenter trial. In the case of vesnarinone, there were other data that were not consistent with the dramatically favorable outcome in the multicenter study. These data seemed to show an inverse dose-response relationship, showed no suggestion

of symptomatic benefit, and showed no effect on hemodynamic endpoints. These inconsistencies led the Agency, with the advice of its Cardio-Renal Advisory Committee, to refuse approval — a decision borne out by the results of the subsequent study.

This example illustrates how inadequacies and inconsistencies in the data, such as lack of pharmacologic rationale and lack of expected other effects accompanying a critical outcome, can weaken the persuasiveness of a single trial. Although an unexplained failure to substantiate the results of a favorable study in a second controlled trial is not proof that the favorable study was in error — studies of effective agents can fail to show efficacy for a variety of reasons — it is often reason not to rely on the single favorable study.

III. DOCUMENTATION OF THE QUALITY OF EVIDENCE SUPPORTING AN EFFECTIVENESS CLAIM

When submitting the requisite quantity of data to support approval of a new product or new use of an approved product, sponsors must also document that the studies were adequately designed and conducted. Essential characteristics of adequate and well-controlled trials are described in 21 CFR 314.126. To demonstrate that a trial supporting an effectiveness claim is adequate and well-controlled, extensive documentation of trial planning, protocols, conduct, and data handling is usually submitted to the Agency, and detailed patient records are made available at the clinical sites.

From a scientific standpoint, however, it is recognized that the extent of documentation necessary depends on the particular study, the types of data involved, and the other evidence available to support the claim. Therefore, the Agency is able to accept different levels of documentation of data quality, as long as the adequacy of the scientific evidence can be assured. This section discusses the factors that influence the extent of documentation needed, with particular emphasis on studies evaluating new uses of approved drugs.

For the purposes of this section, the phrase *documentation of the quality of evidence* refers to (1) the completeness of the documentation and (2) the ability to access the primary study data and the original study-related records (e.g., subjects' medical records, drug accountability records) for the purposes of verifying the data submitted as evidence. These interrelated elements bear on a determination of whether a study is adequate and well-controlled.

In practice, to achieve a high level of documentation, studies supporting claims are ordinarily conducted in accordance with good clinical practices (GCPs). Sponsors routinely monitor all clinical sites, and FDA routinely has access to the original clinical protocols, primary data, clinical site source documents for on-site audits, and complete study reports.

However, situations often arise in which studies that evaluate the efficacy of a drug product lack the full documentation described above (for example, full patient records may not be available) or in which the study was conducted with less monitoring than is ordinarily seen in commercially sponsored trials. Such situations are more common for supplemental indications because postapproval studies are more likely to be conducted by parties other than the drug sponsor and those parties may employ less extensive monitoring and data-gathering procedures than a sponsor. Under certain circumstances, it is possible for sponsors to rely on such studies to support effectiveness claims, despite less than usual documentation or monitoring. Some of those circumstances are described below.

A. Reliance on Less Than Usual Access to Clinical Data or Detailed Study Reports

FDA's access to primary data has proven to be important in many regulatory decisions. There are also reasons to be skeptical of the conclusions of published reports of studies. Experience has shown that such study reports do not always contain a complete, or entirely accurate, representation of study plans, conduct and outcomes. Outright fraud (i.e., deliberate deception) is unusual. However, incompleteness, lack of clarity, unmentioned deviation from prospectively planned analyses, or an inadequate description of how critical endpoint judgments or assessments were made are common flaws. Typically, journal article peer reviewers only have access to a limited data set and analyses, do not see the original protocol and amendments, may not know what happened to study subjects that investigators determined to be non-evaluable, and thus may lack sufficient information to detect critical omissions and problems. The utility of peer review can also be affected by variability in the relevant experience and expertise of peer reviewers. FDA's experiences with the Anturane Reinfarction Trial, as well as literature reports of the efficacy of tacrine and the anti-sepsis HA-1A antibody, illustrate its concerns with reliance on the published medical literature.

Notwithstanding these concerns, the presence of some of the factors discussed below can make it possible for FDA to rely on studies for which it has less than usual access to data or detailed study reports to partially or entirely (the so-called *paper* filing) support an effectiveness claim. FDA's reliance on a literature report to support an effectiveness claim is more likely if FDA can obtain additional critical study details. Section 1 below describes additional information that, if available, would increase the likelihood that a study could be relied on to support an effectiveness claim. Section 2 describes factors that may make efficacy findings sufficiently persuasive to permit reliance on the published literature alone. Note that the factors outlined in Section 2 are relevant to an assessment of the reliability of literature reports generally, whether alone, or accompanied by other important information as discussed in Section 1.

1. Submission of Published Literature or Other Reports in Conjunction with Other Important Information that Enhances the Reliability of the Data

If a sponsor wishes to rely on a study conducted by another party and cannot obtain the primary data from the study, for most well-conducted studies it is possible to obtain other important information, such as a protocol documenting the prospective plans for the trial, records of trial conduct and procedures, patient data listings for important variables, and documentation of the statistical analysis. FDA has considerable experience evaluating large multicenter outcome studies sponsored by U.S. and European government agencies (NIH, British Medical Research Council) and private organizations (the ISIS studies, the SAVE study) for which there was limited access to primary study data, but for which other critical information was available. Providing as many as possible of the following important pieces of information about a study, in conjunction with the published report, can increase the likelihood that the study can be relied on to support an effectiveness claim:

a. The protocol used for the study, as well as any important protocol amendments that were implemented during the study and their relation to study accrual or randomization.

b. The prospective statistical analysis plan and any changes from the original plan that occurred during or after the study, with particular note of which analyses were performed pre- and post-unblinding.

c. Randomization codes and documented study entry dates for the subjects.

d. Full accounting of all study subjects, including identification of any subjects with on-treatment data who have been omitted from analysis and the reasons for omissions, and an analysis of results using all subjects with on-study data.

e. Electronic or paper record of each subject's data for critical variables and pertinent baseline characteristics. Where individual subject responses are a critical variable (e.g., objective responses in cancer patients, clinical cures and microbial eradications in infectious disease patients, death from a particular cause), detailed bases for the assessment, such as the case report, hospital records, and narratives, should be provided when possible.

f. Where safety is a major issue, complete information for all deaths and drop-outs due to toxicity. For postapproval supplemental uses, however, there is generally less need for the results of lab tests or for details of adverse event reports and, consequently, much more limited documentation may be sufficient (e.g., only for unexpected deaths and previously undescribed serious adverse effects). Exceptions to this

approach would include situations in which the population for the supplemental use is so different that existing safety information has limited application (e.g., thrombolysis in stroke patients versus myocardial infarction patients) or where the new population presents serious safety concerns (e.g., extension of a preventive vaccine indication from young children to infants).

2. Submission of Published Literature Reports Alone

The following factors increase the possibility of reliance on published reports alone to support approval of a new product or new use:

a. Multiple studies conducted by different investigators where each of the studies clearly has an adequate design and where the findings across studies are consistent.

b. A high level of detail in the published reports, including clear and adequate descriptions of statistical plans, analytic methods (prospectively determined), and study endpoints, and a full accounting of all enrolled patients.

c. Clearly appropriate endpoints that can be objectively assessed and are not dependent on investigator judgment (e.g., overall mortality, blood pressure, or microbial eradication). Such endpoints are more readily interpreted than more subjective endpoints such as cause-specific mortality or relief of symptoms.

d. Robust results achieved by protocol-specified analyses that yield a consistent conclusion of efficacy and do not require selected post hoc analyses such as covariate adjustment, subsetting, or reduced data sets (e.g., analysis of only responders or compliant patients, or of an "eligible" or “evaluable” subset).

e. Conduct of studies by groups with properly documented operating procedures and a history of implementing such procedures effectively.

There have been approvals based primarily or exclusively on published reports. Examples include the initial approval of secretin for evaluation of pancreatic function and recent approvals of bleomycin and talc for malignant pleural effusion and doxycycline for malaria.

B. Reliance on Studies with Alternative, Less Intensive Quality Control/On-Site Monitoring

Industry-sponsored studies typically use extensive on-site and central monitoring and auditing procedures to assure data quality. Studies supported by other sponsors may employ less stringent procedures and may use no on-site monitoring at all. An International Conference on Harmonisation guideline on good clinical practices,[7] recently accepted internationally, emphasizes that the extent of monitoring in a trial should be based on trial-specific factors (e.g., design, complexity, size, and type of study outcome measures) and that different degrees of on-site monitoring can be appropriate. In recent years, many credible and valuable studies conducted by government or independent study groups, often with important mortality outcomes, had very little on-site monitoring. These studies have addressed quality control in other ways, such as by close control and review of documentation and extensive guidance and planning efforts with investigators. There is a long history of reliance on such studies for initial approval of drugs as well as for additional indications. Factors that influence whether studies with limited or no monitoring may be relied on include the following:

1. The existence of a prospective plan to assure data quality.

2. Studies that have features that make them inherently less susceptible to bias, such as those with relatively simple procedures, noncritical entry criteria, and readily assessed outcomes.

3. The ability to sample critical data and make comparisons to supporting records (e.g., hospital records).

4. Conduct of the study by a group with established operating procedures and a history of implementing such procedures effectively.

[7] International Conference on Harmonisation Guidance for Industry E6, *Good Clinical Practice: Consolidated Guideline*, April 1996.