

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:
21-217s000

STATISTICAL REVIEW(S)



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: 21-217/0000

Drug Name: EXALGO (Hydromorphone HCl Extended Release) Tablets

Indication(s): Management of moderate to severe pain

Applicant: Neuromed Pharmaceuticals Ltd.

Date(s): Letter 5/22/2009
Action 11/20/2009

Review Priority: Priority

Biometrics Division: Division of Biometrics II

Statistical Reviewer: Jonathan Norton, Ph.D.

Concurring Reviewers: Dionne Price, Ph.D.

Medical Division: Division of Anesthesia, Analgesia, and Rheumatology Products

Clinical Team: Elizabeth Kilgore, M.D.
Ellen Fields, M.D.

Project Manager: Diana Walker

Keywords: Analysis of covariance, benefit-risk, clinical studies, missing data, one study application, sensitivity analyses

Table of Contents

1. EXECUTIVE SUMMARY	3
1.1 CONCLUSIONS AND RECOMMENDATIONS	3
1.2 BRIEF OVERVIEW OF CLINICAL STUDIES	3
1.3 STATISTICAL ISSUES AND FINDINGS	4
2. INTRODUCTION	6
2.1 OVERVIEW.....	6
2.2 DATA SOURCES	8
3. STATISTICAL EVALUATION	9
3.1 EVALUATION OF EFFICACY	9
3.2 EVALUATION OF SAFETY	20
4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	20
4.1 GENDER, RACE AND AGE	20
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS	20
5. SUMMARY AND CONCLUSIONS	21
5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE	21
5.2 CONCLUSIONS AND RECOMMENDATIONS	22
5.3 REVIEW OF THE PROPOSED LABEL	22
APPENDIX A: REFERENCES.....	24
APPENDIX B: ADDITIONAL STATISTICAL DETAILS.....	25

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

Neuromed Pharmaceuticals seeks to have Exalgo, hydromorphone HCl extended release tablets, approved for management of moderate to severe chronic pain. In the previous Approvable letter for this NDA, the Applicant was told to submit an additional adequate and well-controlled study with multiple doses. The results of Study NMT-1077-301 provide support for a finding that Exalgo is efficacious among patients comparable to those in the enriched study population.

1.2 Brief Overview of Clinical Studies

FDA issued an Approvable letter for this NDA on October 27, 2000. One of the requirements listed in the letter was, "Conduct an adequate and well-controlled [AWC] study, with multiple dosing of the to-be-marketed formulation, in the setting of moderate to severe pain, to establish efficacy of the product." Since that time, the NDA has been owned by multiple firms who have met with the responsible division at FDA. During these meetings, the responsible division stated that the study should be 12 weeks in duration, should use a titrated dose of hydromorphone, and could employ a randomized withdrawal design. In September 2007, the Division of Anesthesia, Analgesia and Rheumatology Products (DAARP) granted a Special Protocol Agreement for Study NMT-1077-301, which I will refer to as Study 301.

Study 301, a double-blind randomized withdrawal trial of Exalgo for the treatment of chronic lower back pain, was the single confirmatory study included in this submission. It was conducted at 66 sites in the United States. The study enrolled 459 patients, of whom 268 were randomized into the double-blind phase.

The design of the study was as follows. After screening and enrollment, subjects were assigned an Exalgo dose that was equal to 75% of their prior opioid dose in morphine equivalents. During the Conversion and Titration Phase, an investigator could increase a patient's dose as often as every three days, to a maximum dose of 64 mg per day. Only one dosage decrease was allowed, with a minimum dose of 12 mg per day. For the first three days of this phase, subjects were allowed unlimited immediate release hydromorphone as rescue medication. After Day 3, subjects were allowed a maximum of two rescue tablets per day, with each tablet containing 5-15% of the daily Exalgo dose. Patients entered into the Double-blind Phase when they were able to maintain a stable dose of Exalgo for at least seven consecutive days.

In the Double-blind Phase, subjects were randomized in a 1:1 ratio to either the same stable dose of Exalgo or placebo. Those assigned to placebo had their dose tapered over a period of 14 days. After Day 15, rescue medication was restricted to a weekly average of two tablets per day, and patients who violated this requirement were discontinued.

The primary efficacy endpoint was the mean change from baseline to Week 12 in pain intensity, taking a weekly average of the diary pain scores. A related efficacy variable was the change from

baseline in the pain intensity recorded at each clinic visit. Additional endpoints included time to treatment failure, change from baseline in patient global assessment, change from baseline in the Roland-Morris disability questionnaire, proportion of dropouts, rescue tablets per day, and the cumulative count of rescue tablets used over time (mean cumulative function). All comparisons were between the Exalgo arm and the placebo arm.

1.3 Statistical Issues and Findings

Study 301 showed a positive result on the primary endpoint using the analysis which was included in the Special Protocol Agreement. However, there were two features of the study that complicated interpretation, namely the high dropout rate and the widespread use of rescue medication.

Of the subjects who entered the Conversation and Titration phase, 42% dropped out. Since these dropouts occurred before randomization, they do not affect the internal validity of the trial. They do, however, make the results less generalizable; the finding of superiority to placebo is only applicable to patients in the general disease population who respond to and tolerate Exalgo. The previous statement may seem circular, but a truly ineffective agent would not be expected to beat placebo in *any* population (assuming proper pre-specification and adjustment for multiplicity as needed).

Even among the patients who were randomized, 59% withdrew. As would be expected, the discontinuation rate was somewhat higher in the placebo group (67%) than in the active group (51%). These high discontinuation rates raise a concern about the internal validity of the trial. One cannot assume, for example, that patients who discontinued from the trial would have had similar pain scores to the patients who completed it. I addressed this concern by performing a variety of sensitivity analyses.

In regard to use of supplemental analgesia, there is no universally accepted method to simultaneously evaluate pain and rescue medication. The method of Silverman et al (1993), which showed Exalgo to be superior to placebo, is reasonable and appears adequate for the present purpose. Further research and discussion is called for on this topic.

As noted earlier, the Approvable letter stated that the Applicant needed to submit one AWC study with multiple doses. Since that time the Applicant has submitted two confirmatory studies, one which failed to show superiority on the primary endpoint (M03-644-05) and one which succeeded (301). On general principle, the combination of a failed study and a successful one provides a lower level of statistical evidence than a single successful study. If the type I error rate of a single study is set at .05, then picking the best result from two studies results in an overall type I error rate of about .10. To be precise, the latter error rate is the probability of rejecting the null hypothesis at least once when it is true for both studies.

In this case, the differing results must be interpreted in light of the fact that the two studies had substantially different designs. Study M03-644-05 used a parallel-group design with patients randomized to placebo, 8 mg Exalgo, or 16 mg Exalgo. Study 301 used a randomized withdrawal design in which patients were titrated to a stable dose as high as 64 mg.

A Joint Meeting of the Anesthetic and Life Support Drugs Advisory Committee and the Drug Safety and Risk Management Advisory Committee was convened on September 24, 2009. Although the stated topic of the meeting was risk mitigation, the committee reached a consensus that Exalgo is a “significantly efficacious drug for a group [i.e., subpopulation] of patients who are in pain.” My findings are consistent with this statement.

2. INTRODUCTION

2.1 Overview

The Applicant, Neuromed Pharmaceuticals, seeks to have hydromorphone HCl extended release tablets approved for management of moderate to severe chronic pain. The proposed tradename is Exalgo, and the proposed dosages are 8 mg, 12 mg, 16 mg, and 32 mg, taken daily.

FDA issued an Approvable letter for this NDA on October 27, 2000. One of the requirements listed in this letter was, "Conduct an adequate and well-controlled study, with multiple dosing of the to-be-marketed formulation, in the setting of moderate to severe pain, to establish efficacy of the product." The submitted study, DO-119, had failed to show a significant difference between the high and low doses of extended release hydromorphone. The sponsor at that time was Knoll Pharmaceutical Company.

On January 24, 2003, then-sponsor Abbott Laboratories met with the Division of Anesthetic, Critical Care, and Addiction Drug Products (DACCADP) to discuss the deficiencies cited in the Approval letter. DACCADP indicated that the adequate and well-controlled (AWC) study should be conducted on patients who need around-the-clock opiates for an extended period of time. The sponsor was encouraged to conduct a 12-week superiority trial using a titrated dose of hydromorphone. DACCADP confirmed that only one clinical trial would be needed to establish efficacy.

On November 1, 2005, then-sponsor Alza Corporation met with DACCADP to discuss the results of protocol M03-644. This trial had failed to achieve statistical significance using the planned analysis, which used baseline-observation-carried-forward (BOCF) imputation for dropouts. DACCADP noted that the trial was difficult to interpret because of the high rate of discontinuations due to adverse events in the active treatment arm. Dr. Thomas Permutt stated that the alternative analyses conducted by the sponsor assigned good scores to patients who dropped out, even though they were no longer benefitting from the drug. He proposed several alternatives to BOCF. There was an extensive discussion about how to deal with the high dropout rates in chronic pain trials. The sponsor was told that a randomized withdrawal design might be acceptable for an AWC trial, but that they would need to take care that the apparent treatment effect was not confounded with opioid withdrawal symptoms in the placebo arm.

Following a reorganization within FDA, this NDA fell within the purview of the Division of Anesthesia, Analgesia and Rheumatology Products (DAARP). On August 15, 2007, DAARP granted a Special Protocol Agreement (SPA) to the Applicant, Neuromed Pharmaceuticals, for Study NMT-1077-301. However, the SPA letter included the following statements:

Our determination of whether you demonstrated substantial evidence of efficacy depends on our concurrence with your classifications of patients who prematurely discontinue. That is a review issue.

Therefore, we recommend that you capture and submit all data that will permit us to assess your outcomes and dispositions. Such important data include but are not limited to:

- Full reporting of COWS and SOWS questionnaires including the individual question scores
- Detailed descriptions of adverse events, particularly adverse events leading to discontinuation and a focused physical exam to detect signs of opioid withdrawal
- Detailed accounting of drug administration over the past 7 days prior to discontinuation

DAARP had a pre-Complete Response meeting with the Applicant on August 8, 2008. There was extended discussion of the differing requirements for a 505(b)(1) vs. a 505(b)(2) application for this product. The Applicant was told to submit case report forms for patients who drop out due to “lack of efficacy” and show evidence of abuse or misuse. They agreed to develop an algorithm for identifying these subjects.

The Applicant ultimately chose to use the 505(b)(1) pathway. Table 1 summarizes the double-blind, Phase 3 studies.

Table 1: Phase 3 Chronic Pain Studies, Placebo- or Active-Controlled (Source: Reviewer)

Study Info	Design	Results
<p>DO-119 Sponsor: Knoll Dates conducted: 10/1998-6/1999</p>	<p>Patients titrated to stable HM IR dose, then randomized to following arms: same dose of HM IR, “full” (equivalent) dose of HM ER, half dose of HM ER.</p> <p>Duration of DB treatment: seven days</p> <p>Primary endpoint: Rescue medication use from last four days, full dose vs. half</p>	<p>Failure on pre-specified analysis (p=.42). FDA issued Approvable letter for NDA in December 2000.</p>
<p>M03-644-05 Sponsor: Alza Dates conducted: 11/2003-4/2005</p>	<p>Parallel design with following arms: HM ER 8 mg, HM ER 16 mg, placebo</p> <p>Primary endpoint: Area-under-the curve for pain, through Week 12.</p>	<p>Using planned analysis, neither HM dose found superior to placebo. Alza told by DACCADP that study did not demonstrate efficacy.</p>
<p>NMT-1077-301 Sponsor: Neuromed Dates conducted: 11/2007-1/2009</p>	<p>Randomized withdrawal design comparing titrated HM ER to placebo.</p> <p>Primary endpoint: Change in weekly pain score from randomization baseline to Week 12.</p>	<p>Neuromed reports that they found HM ER superior to placebo.</p>

HM = hydromorphone

IR = immediate release

ER = extended release

DACCADP = Division of Anesthetic, Critical Care, and Addiction Drug Products

2.2 Data Sources

The electronic version of this NDA can be found at \\cdsesub1\EVSPROD\NDA021217.

3. STATISTICAL EVALUATION

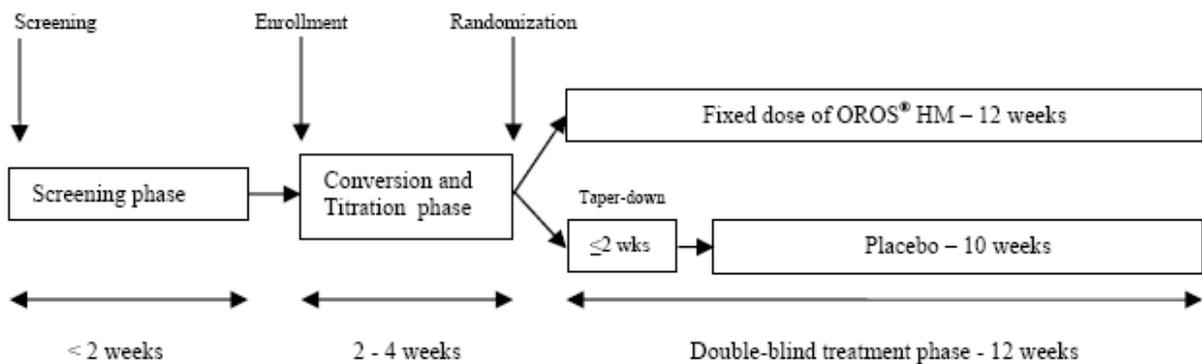
NMT-1077-301

3.1 Evaluation of Efficacy

Study Design and Endpoints

NMT-1077-301 was a double-blind randomized withdrawal study to evaluate the efficacy of Exalgo for the treatment of chronic lower back pain. It was conducted at 66 sites in the United States. The study enrolled 459 patients, of whom 268 were randomized into the double-blind phase of the study.

Figure 1: Study Design (Source: Figure 1, Clinical Study Report)



The design of the study is illustrated by Figure 1. During the Screening phase, patients were taught to report their pain scores using the phone diary and their compliance was assessed. After enrollment in the study, subjects were assigned an Exalgo dose that was equal to 75% of their prior opioid dose in morphine equivalents. During the Conversion and Titration Phase, they continued to record their daily pain intensity using a phone diary. The investigator could increase a patient's dose as often as every three days, to a maximum dose of 64 mg per day. Only one dosage decrease was allowed, with a minimum dose of 12 mg per day. For the first three days of this phase, subjects were allowed unlimited rescue medication. The rescue medication was HM IR, each tablet containing 5-15% of the daily Exalgo dose. After Day 3, subjects were allowed a maximum of two rescue tablets per day. Patients entered into the Double-blind Phase when they met the following criteria:

- Maintain stable dose for at least seven consecutive days
- During this period, average no more than two rescue tablets per day
- During this period, average a pain score of no more than four
- Indicate that medication helped their pain enough that they would continue to take it
- Have no intolerable side effects

In the Double-blind Phase, subjects were randomized in a 1:1 ratio to either the same stable dose of Exalgo or placebo. Those assigned to placebo had their dose tapered over a period of 14 days. Subjects were allowed up to six rescue tablets per day during Week 1 and up to four during Week 2. After Day 15, rescue medication was restricted to a weekly average of two tablets per day. Patients who averaged greater than two tablets per day over any seven-day period were discontinued.

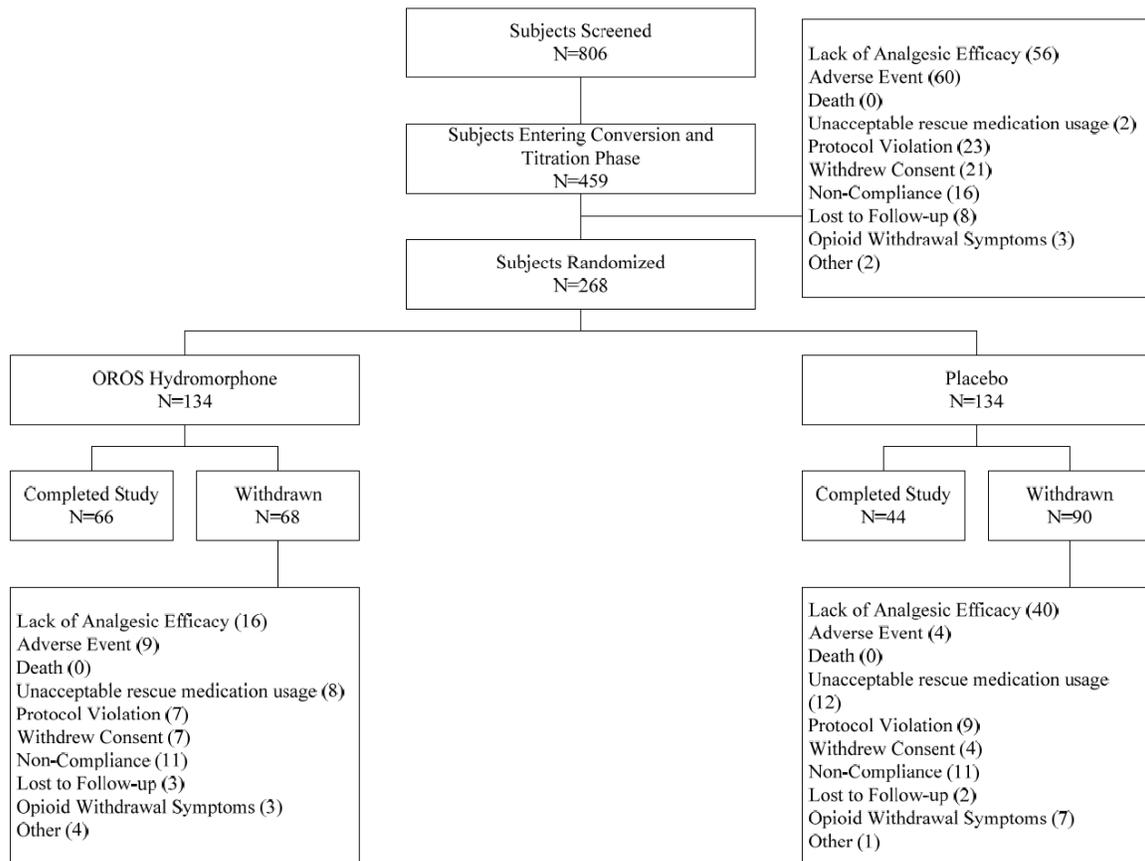
The nature of the blind in this trial was as follows. During the Conversion and Titration phase of the trial, the dose a subject was taking was clearly indicated by the color of the tablet(s). In the Double-Blind phase, patients received either the same active pills or matching placebo. In order to maintain the blind during the taper, the pills were over-encapsulated.

The primary efficacy endpoint was the mean change from baseline to Week 12 in pain intensity, taking a weekly average of diary pain scores. A related efficacy variable was the change from baseline in the pain intensity recorded at each clinic visit. Additional endpoints included time to treatment failure, change from baseline in patient global assessment, change from baseline in the Roland-Morris disability questionnaire, proportion of dropouts, rescue tablets per day, and the cumulative count of rescue tablets used over time (mean cumulative function). All comparisons were between the Exalgo arm and the placebo arm. There was no adjustment for multiple endpoints. The primary analysis population was intent-to-treat (ITT), defined as all randomized and treated patients.

Patient Disposition, Demographic and Baseline Characteristics

Figure 2 shows the disposition of subjects in the study. The figure was originally provided by the Applicant, but I verified the contents. Note the high rates of discontinuation even in the enriched population that entered the double-blinded phase: 51% in the Exalgo arm and 67% in the placebo arm. Overall, 76% of patients who entered the Conversion and Titration Phase eventually withdrew from the study.

Figure 2: Disposition of Subjects (Source: Figure 3, Clinical Study Report)



The Applicant excluded two subjects from the ITT population despite the fact that they received treatment in the double-blind phase. Subject 084014 was discontinued due to noncompliance. In response to an information request, the Applicant stated that they excluded this patient because they have “no data indicating that the subject took study medication after they were randomized on 7/28/08.” However, the subject’s case report form indicates that a blister card of 14 doses of double-blind study medication was given to the subject and 6 doses were missing when the card was returned. Moreover, the narrative states that the subject “was randomized and received the first dose of double-blind study drug on 28 July 2008 at a dose of 64 mg.” According to the final SAP, “The ITT population will include all subjects randomized to the double-blind phase who have received at least one dose of study medication after randomization.” Following the definition in the SAP, I consider him to be in the ITT population. Subject 084017 was discontinued due to unacceptable rescue medication use. The Applicant claims that this subject should be excluded from ITT because there are no baseline pain scores, but I disagree.

Table 2 shows the demographics of the patients in the Applicant’s ITT population. It was provided by the Applicant, but I verified the contents.

Table 2: Demographics, Applicant ITT (Source: Table 11, Clinical Study Report)

Characteristic	OROS [®] Hydromorphone N=133	Placebo N=133	Total N=266
Age, years			
Mean (SD)	47.8 (10.53)	49.4 (10.57)	48.6 (10.56)
Median	49.0	49.0	49.0
Range (min, max)	24, 75	23, 72	23, 75
Age Group, n (%)			
18 to 64	128 (96.2)	122 (91.7)	250 (94.0)
65 to 75	5 (3.8)	11 (8.3)	16 (6.0)
Gender, n (%)			
Male	72 (54.1)	60 (45.1)	132 (49.6)
Female	61 (45.9)	73 (54.9)	134 (50.4)
Race, n (%)			
American Indian	0	1 (0.8)	1 (0.4)
Black	14 (10.5)	9 (6.8)	23 (8.6)
Caucasian	108 (81.2)	117 (88.0)	225 (84.6)
Hispanic	9 (6.8)	5 (3.8)	14 (5.3)
Oriental	1 (0.8)	1 (0.8)	2 (0.8)
Other	1 (0.8)	0	1 (0.4)
Weight, kg			
Mean (SD)	90.43 (24.598)	93.10 (23.731)	91.77 (24.160)
Median	85.40	89.30	88.50
Range (min, max)	44.0, 164.3	56.7, 168.0	44.0, 168.0
Height, cm			
Mean (SD)	172.50 (11.178)	169.93 (10.486)	171.21 (10.893)
Median	171.50	170.20	170.70
Range (min, max)	142.2, 198.1	127.0, 194.3	127.0, 198.1

Protocol Deviations

Table 3 lists the protocol violations reported by the Applicant. Of the 37 patients listed as not meeting the double-blind inclusion criteria, in all cases but one, the reason was that they had a pain score greater than four at the randomization baseline. Since these violations preceded randomization, in principle they do not undermine the efficacy analysis. Moreover, they were balanced across treatment arms.

Table 3: Protocol Violations in Randomized Population (Table 8, Clinical Study Report)

Category ^a	OROS Hydromorphone	Placebo
	N=134 n (%) ^b	N=134 n (%) ^b
Inclusion Criteria Not Met	0	2 (1.5)
Exclusion Criteria Met	4 (3.0)	3 (2.2)
Double-blind Inclusion Criteria Not Met	19 (14.2)	18 (13.4)
Prohibited Medication	4 (3.0)	7 (5.2)
Suspected Diversion	0	0
Newly Found History	0	0
Other	3 (2.2)	3 (2.2)

^aPatients may be included in multiple categories.

^bPercentages based upon the number of patients randomized to each treatment group.

Statistical Methods

The Applicant proposed different versions of the primary efficacy analysis at different times. DAARP granted a SPA (Special Protocol Agreement) for version 2 of the SAP, in which the primary analysis was analysis of covariance (ANCOVA). The independent variables in the model were to be treatment, site, and baseline pain score. The baseline pain was defined as the average of the diary values in the week prior to randomization. Missing values due to premature withdrawal were to be imputed as follows:

- For discontinuations due to apparent opioid withdrawal symptoms, the randomization baseline score would be carried forward. Note that the pain score carried forward is relatively low.
- For discontinuations due to an adverse event (AE), the screening pain observation would be carried forward (SOCF). This is a relatively high score.
- For other discontinuations, last-observation-carried-forward (LOCF) imputation would be used. The last observation was defined as the average pain score over the final week that the patient is in the study. This category included patients who discontinued due to excessive use of rescue.

If the Week 12 pain scores were missing but the patient did not discontinue, then LOCF was to be used. The analysis did not adjust for use of rescue medication by patients who stayed in the trial.

The final SAP included a different primary analysis than the version that was granted a SPA. First of all, the effect of treatment center was removed from the ANCOVA model. Also, the analysis method used was to depend on whether the data violated certain assumptions. See Appendix B for details. If any of the assumptions were violated, then the Applicant planned to use a non-parametric methodology that their cited reference (Stokes et al, 2000) calls *rank ANCOVA*. The results of this analysis were included in the clinical study report, but the results of the original ANCOVA analysis were not. In response to an information request, the Applicant performed the ANCOVA analysis specified in the SPA and submitted the results. For reasons described in Appendix B, I prefer the ANCOVA analysis to the rank-based method.

Both the SPA version and the final version of the SAP state that the baseline pain score will be computed as the average of the diary scores in the week prior to randomization. In the clinical study report, however, the Applicant included the score from Double-Blind Visit 1 in the average. In response to an information request, the Applicant submitted baseline pain scores using the planned method. The baseline pain scores from the two methods are quite similar, having a mean difference of .02 (on an 11-point scale) and a correlation of .992.

The SPA version of the SAP included centers in the ANCOVA model and did not make any allowance for pooling. When the Applicant performed the SPA analysis at our request, they pooled centers with fewer than eight patients. The pooling algorithm was such that the combined centers had at least four subjects in each treatment arm. Pooling did not alter the finding of significance for the primary endpoint.

Results and Conclusions

Table 4 shows the results from the primary analysis, as originally submitted by the Applicant. In a later correction, the Applicant stated that footnote c should read “Baseline value is the mean of the patient diary measurements in the week prior to randomization and Double-Blind Visit 1.” (See explanation in previous section). I verified the contents from the analysis data, and further reproduced the analysis data using the Applicant’s code. The p-value is based on the rank ANCOVA test. The min and max are rounded to the nearest whole number.

Table 4: Primary Analysis Based on Final SAP (Source: Table 15, Clinical Study Report)

Statistic ^a	OROS [®] Hydromorphone	Placebo	P-value ^b
Baseline ^c			
N	133	133	
Mean	3.2	3.1	
Median	3.3	3.3	
Range (min, max)	0, 6	0, 6	
Visit 11/final visit (Week 12) ^d			
N	133	133	
Mean	3.8	4.8	
Median	3.6	4.8	
Range (min, max)	0, 9	0, 9	
Change from Baseline			0.000007
N	133	133	
Mean	0.6	1.7	
Median	0.2	1.6	
Range (min, max)	-5, 5	-3, 7	

^aThis is an 11-point scale ranging from 0 (no pain) to 10 (worst possible pain).

^bP-value from test for significant treatment difference using Cochran-Mantel-Haenszel chi-square test comparing change from Baseline after adjusting for Baseline value using ranks.

^cMean of the patient diary measurements in the week prior to randomization.

^dPatients with missing weekly patient diary data due to premature withdrawal had their value at final visit imputed based on the reason for discontinuation.

ITT=intent-to-treat; max=maximum; min=minimum

Reviewer's note: Applicant corrected footnote c.

At our request, the Applicant also carried out the SPA version of the primary analysis. As a reminder, the SPA analysis differed from that shown in Table 4 as follows: parametric ANCOVA was used, pooled study center was included as a factor, and the baseline pain was computed from diary entries only. Using the SPA analysis, the Applicant again found Exalgo to be statistically superior to placebo ($p < .001$). I verified this finding from the analysis data.

A questionable feature of the Applicant's handling of missing data was that LOCF imputation was used for patients who discontinued due to overuse of rescue medication. This is arguably inappropriate because the score carried forward could be artificially low due to rescue medication use. As a sensitivity analysis, I used SOCF for these patients, giving them high pain scores. The primary endpoint remained significant.

As noted earlier, the Applicant excluded two patients from the ITT whom I believe should have been included. One subject was missing post-randomization pain scores (084014) and other was missing baseline pain scores (084017). For my analysis, I used the corresponding placebo means

to impute the missing values for these subjects. Adding these additional subjects to the analysis population did not substantially change the results.

(b) (4)
 The Applicant's figure is shown in section 5.3, and the tabular data on which the figure is based are shown in Table 5. I was able to nearly reproduce the counts in each category using the analysis data. The exception was that I found that 70 placebo patients had a response $\geq 20\%$, not 71. Also, it is not clear to me why the Applicant considers the ITT population to have 132 in the Exalgo arm for this analysis, whereas they used 133 in the primary analysis.

Table 5: Applicant's Cumulative Responder Analysis
(Source: Table 14.2.12, Clinical Study Report)

Response Category ^a	OROS Hydromorphone N=132 n (%)	Placebo N=133 n (%)
>0% reduction in pain	101 (76.5)	93 (69.9)
$\geq 10\%$ reduction in pain	99 (75.0)	79 (59.4)
$\geq 20\%$ reduction in pain	89 (67.4)	71 (53.4)
$\geq 30\%$ reduction in pain	80 (60.6)	57 (42.9)
$\geq 40\%$ reduction in pain	64 (48.5)	48 (36.1)
$\geq 50\%$ reduction in pain	56 (42.4)	32 (24.1)
$\geq 60\%$ reduction in pain	32 (24.2)	20 (15.0)
$\geq 70\%$ reduction in pain	20 (15.2)	12 (9.0)
$\geq 80\%$ reduction in pain	9 (6.8)	8 (6.0)
$\geq 90\%$ reduction in pain	3 (2.3)	2 (1.5)
100% reduction in pain	1 (0.8)	1 (0.8)

^a Percent reduction in pain calculated as (Week 12/Final visit mean pain score from patient diary – screening pain score) / screening pain score *100. Patients who withdrew prematurely had their value at Week 12/Final visit imputed based on the reason for discontinuation.

As the footnote to the table indicates, the screening baseline pain score was used to determine the percentage of improvement. This is appropriate for a randomized withdrawal design, because the screening pain score is relatively high. In contrast, the randomization baseline is not appropriate for this purpose under the current design, because patients must have their pain under control at the time of randomization.

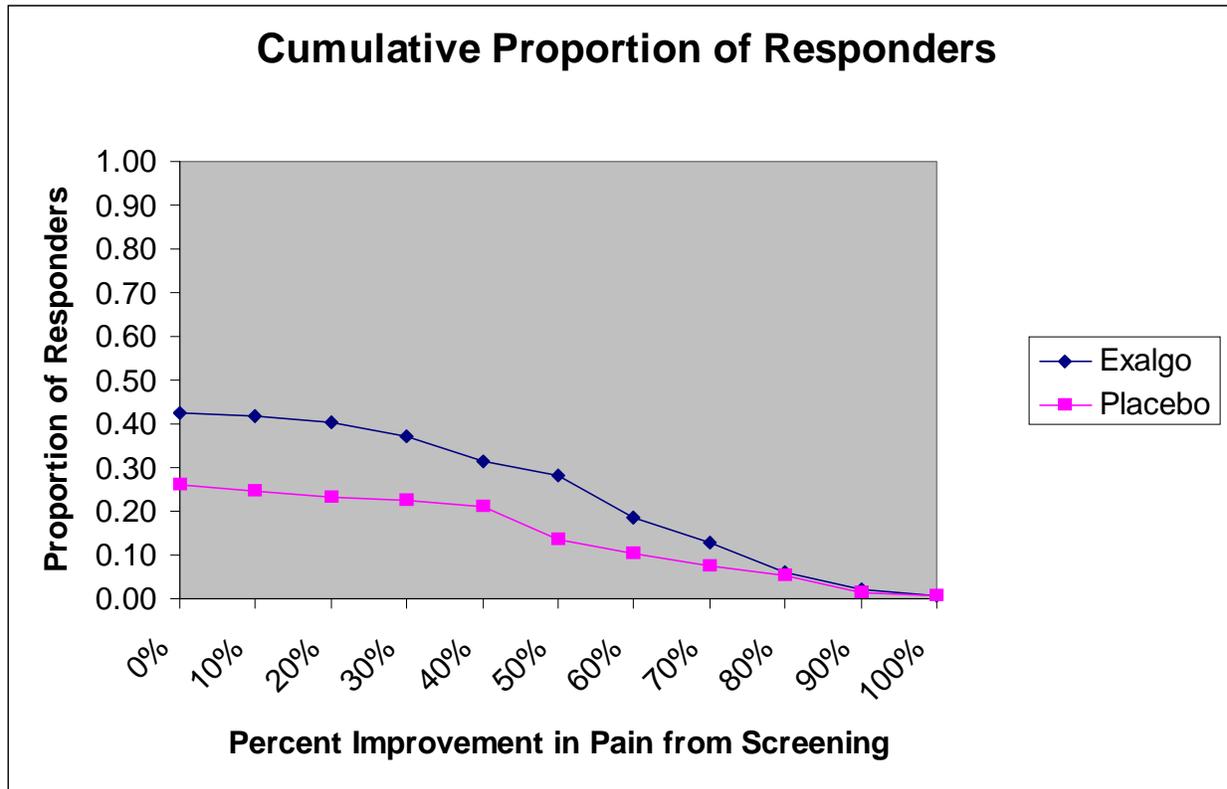
In the Applicant's figure, missing pain scores were imputed at Week 12. DAARP prefers to treat all patients who withdraw from analgesic trials as non-responders, based on the rationale that a withdrawal suggests that the patient was not benefitting from treatment. Figure 3 shows my version of the responder analysis, which follows this reasoning. The relatively low response rates reflect the high dropout rates. I consider each arm to have 134 patients in the ITT set, and calculated percentages this way. It is common to use 30% and 50% improvement from baseline as criteria for a clinically meaningful effect. Using 30% as the cutoff, patients in the active arm had a 37% (50/134) chance of response and those in the placebo arm had a 22% (30/134) chance of response. With a cutoff of 50%, the success rates were 28% (38/134) with Exalgo vs. 13% (18/134) with placebo. With either cutoff, Exalgo was statistically superior to placebo using a chi-square test ($p < .01$).

The large percentage of discontinuations in each group, however, makes the results of the study highly dependent on what type of outcome is imputed for dropouts. For simplicity, let the outcome of interest be a 30% improvement from the screening baseline. Under a "worst case" scenario, one could imagine that all of the dropouts in the active arm would have been failures, had they completed the trial, and the ones in the placebo arm would have been successes. Under this implausible scenario, the placebo would have had a 90% success rate, compared to 37% in the Exalgo arm. Note that the success rates in the previous paragraph are based on all dropouts being failures, so the Exalgo success rate is unchanged under this scenario.

Since the "worst case" is clearly unrealistic, one can ask how many of the 90 withdrawals in the placebo arm can be counted as successes before we no longer find Exalgo to be statistically superior to placebo. Again using a chi-square test, one could classify 5 of the 90 placebo withdrawals as successes and still find Exalgo superior at the .05 level. Further, one could count 7 placebo withdrawals as successes and find superiority at the .10 level. These calculations assume that all Exalgo withdrawals are failures. There are also numerous intermediate scenarios in which some Exalgo withdrawals are imputed as successes.

As the preceding paragraphs show, the interpretation of the study depends on what one is willing to assume about patients who withdraw during the double-blind phase. The responder analysis shown in Figure 3 has a certain appeal insofar as it treats withdrawals from the active treatment and placebo arms in the same way, namely as clinical failures. If 30% or 50% improvement are accepted as cutoffs, then Exalgo is statistically superior to placebo according to these post-hoc tests.

Figure 3: Cumulative Responder Analysis (Source: Reviewer)



In a trial of this sort with widespread use of rescue analgesics, there is a risk that the treatment effect could be either exaggerated or attenuated by differential use of rescue analgesia. Table 6 shows the mean number of rescue tablets per day, broken down by treatment arm and the visit at which the data were collected. As the table shows, use of rescue medication was roughly balanced across treatment arms throughout the trial. The fact that rescue analgesic use was balanced is reassuring, but it is also helpful to conduct a statistical test of efficacy that accounts for both pain and rescue medication. There is no definitive approach to this problem, however, because there is no unique way to put subjective pain ratings on the same scale as objective measures of analgesic use. One approach is to compare ranks, assuming that a given population percentile of pain is equivalent to the same percentile of rescue medication use, e.g., a subject who is in the 90th percentile for pain and the 75th percentile for rescue medication is interchangeable with a subject who has the same percentiles switched. Note that reported pain and rescue use are comparable in the sense that both are expected to be part of a subject's response to pain.

Table 6: Mean Rescue Tablets per Day, by Visit (Source: Reviewer)

Visit # (Time)	Exalgo			Placebo		
	N	Mean	SD	N	Mean	SD
2 (Day 4)	124	1.5	1.4	120	1.6	1.5
3 (Week 1)	112	1.6	1.5	106	1.8	1.5
4 (Day 11)	99	1.6	1.5	80	1.8	1.5
5 (Week 2)	100	1.5	1.2	73	1.7	1.4
6 (Week 3)	89	1.2	0.9	57	1.3	1.0
7 (Week 4)	81	1.1	0.8	52	1.1	1.1
8 (Week 6)	70	0.9	0.8	49	1.1	1.0
9 (Week 8)	66	0.9	0.8	47	1.0	1.0
10 (Week 10)	66	0.9	0.8	44	1.0	0.9
11 (Week 12)	66	0.9	0.8	44	1.0	0.9
Early Termin.	61	2.1	1.3	81	2.4	1.4

Silverman et al (1993) introduced such a rank-based method. Leaving aside calculations done for cosmetic purposes, they proposed the following procedure:

- 1) Rank the subjects according to pain.
- 2) Rank the subjects according to rescue medication use.
- 3) Sum the two ranks for each subject.
- 4) Compare the summed ranks between the treatment arms using a Wilcoxon rank sum method.

I carried out this procedure and again found Exalgo to be superior to placebo (exact $p < .0001$).

Exploratory Benefit-Risk

Chuang-Stein et al (1991) introduced a method for evaluating both the benefits and “risks” (i.e., harms) of a drug on a per-patient basis. As Table 7 illustrates, each subject is classified according to whether that subject experienced benefit, harm, neither, or both. Subjects who withdraw from a study are put in a separate category, regardless of whether they experienced benefit and/or harm. For the table, I defined ***benefit*** as a 30% reduction in pain from the screening baseline. All moderate or severe AE’s (as determined by the Applicant) are counted, regardless of whether they were attributed to treatment. These definitions are somewhat arbitrary and were not pre-specified; hence this analysis should be taken as exploratory.

Table 7: Individual Benefit-Risk (Source: Reviewer)

Category	Exalgo (n=134)	Placebo (n = 134)
1 – Benefit w/out AE	30%	16%
2 – Benefit with AE	7%	6%
3 – No benefit, no AE	10%	7%
4 – AE only	2%	4%
5 – Withdrawal	51%	67%
1 or 2 – Benefit	37%	22%

A common measure of benefit-risk is the number-needed-to-treat (NNT), defined as the inverse of the increase in the probability of clinical success for patients who use the experimental treatment compared to the control. This can be interpreted as the number of patients who would have to be given the experimental treatment in order for one patient to benefit. In this case the NNT is approximately seven, regardless of whether the criterion is simply a 30% improvement in pain (category 1 or 2) or the stricter requirement to get that level of improvement without a more-than-mild AE (category 1). Similar results are obtained when benefit is defined as a 50% improvement from screening.

3.2 Evaluation of Safety

The safety of tapentadol was reviewed by Elizabeth Kilgore, M.D.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

Table 8 shows the results for the primary endpoint by age, race, and sex. There was a marginally significant interaction between age and treatment effect ($p = .09$), with patients under 55 tending to show a larger effect. The treatment effect did not significantly differ between genders. It was not possible to perform a meaningful test for an interaction with race due to the small number of non-Caucasian subjects.

Table 8: Subgroup Analysis of Primary Endpoint (Source: Reviewer)

Subgroup	Exalgo Mean (SD, N)	Placebo Mean (SD, N)
Age		
Under 55	0.6 (1.8, 99)	1.8 (1.9, 92)
55 or Older	0.8 (1.9, 34)	1.3 (1.8, 41)
Gender		
Female	0.8 (1.9, 61)	1.7 (1.9, 73)
Male	0.4 (1.7, 72)	1.6 (1.9, 60)
Race		
Black	0.7 (1.7, 14)	1.1 (1.4, 9)
Caucasian	0.6 (1.6, 108)	1.7 (1.9, 117)
Other	0.8 (3.2, 11)	1.4 (1.9, 7)

4.2 Other Special/Subgroup Populations

I did not examine other subgroups.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

Study NMT-1077-301, which I will refer to as Study 301, showed a positive result on the primary endpoint using both the original and revised analysis plans. However, there were two features of the study that complicate interpretation, namely the high dropout rate and the widespread use of rescue medication.

As Figure 2 shows, 42% percent of subjects dropped out during the titration phase. Since these dropouts occurred before randomization, they did not affect the internal validity of the trial. They do, however, make the results less generalizable; the finding of superiority to placebo is only applicable to patients in the general disease population who respond to and tolerate Exalgo. The previous statement may seem circular, but a truly ineffective agent would not be expected to beat placebo in *any* population (assuming proper pre-specification and adjustment for multiplicity as needed).

Even among the patients who were randomized, 59% withdrew. As would be expected, the discontinuation rate was somewhat higher in the placebo group (67%) than in the active group (51%). These high discontinuation rates raise a concern about the internal validity of the trial. One cannot assume, for example, that patients who discontinued from the trial would have had similar pain scores to the patients who completed it. I addressed this concern by performing a variety of sensitivity analyses.

In regard to use of supplemental analgesia, there is no universally accepted method to simultaneously evaluate pain and rescue medication. The method of Silverman et al (1993), which showed Exalgo to be superior to placebo, is reasonable and appears adequate for the present purpose. Further research and discussion is called for on this topic.

The Approvable letter issued in October 2000 stated that the Applicant needed to submit one adequate-and-well-controlled study with multiple doses. Since that time the Applicant has submitted two confirmatory studies, one which failed to show superiority on the primary endpoint (M03-644-05) and one which succeeded (301). On general principle, the combination of a failed study and a successful one provides a lower level of statistical evidence than a single successful study. If the type I error rate of a single study is set at .05, then picking the best result from two studies results in an overall type I error rate of about .10. To be precise, the latter error rate is the probability of rejecting the null hypothesis at least once when it is true for both studies.

In this case, the differing results must be interpreted in light of the fact that the two studies had substantially different designs. Study 644 (M03-644-05) used a parallel-group design with patients randomized to placebo, 8 mg Exalgo, or 16 mg Exalgo. Study 301 used a randomized withdrawal design in which patients were titrated to a stable dose as high as 64 mg. This design was suggested by FDA and incorporated in a Special Protocol Agreement.

A Joint Meeting of the Anesthetic and Life Support Drugs Advisory Committee and the Drug Safety and Risk Management Advisory Committee was convened on September 24, 2009. Although the stated topic of the meeting was risk mitigation, the committee reached a consensus that Exalgo is a “significantly efficacious drug for a group [i.e., subpopulation] of patients who are in pain.” My findings are consistent with this statement.

5.2 Conclusions and Recommendations

In the previous Approvable letter for this NDA, the Applicant was told to submit an additional adequate and well-controlled study with multiple doses. The results of Study 301 provide support for a finding that Exalgo is efficacious among patients comparable to those in the enriched study population.

5.3 Review of the Proposed Label

EXALGO was investigated in a double-blind, placebo-controlled, randomized withdrawal study in opioid tolerant patients with moderate to severe low back pain. Patients were considered opioid tolerant if they were currently on opioid therapy that was ≥ 60 mg/day of oral morphine equivalent for at least 2 months prior to screening. Patients entered an open-label conversion and titration phase with EXALGO, were converted to a starting dose that was approximately 75% of their total daily morphine equivalent dose, and were dosed once daily until adequate pain control was achieved while exhibiting no intolerable side effects. Supplemental immediate release hydromorphone tablets were allowed during the conversion and titration phase. Patients who achieved a stable dose entered a 12-week, double-blind, placebo-controlled, randomized treatment phase. Mean daily dose at randomization was 37.8 mg/day (median 32.0 mg/day, range of 12 mg/day – 64 mg/day). Fifty eight percent of patients were successfully titrated to a stable dose of EXALGO during the open-label conversion and titration phase.

During the double-blind treatment phase, patients randomized to EXALGO continued with the stable dose achieved in the conversion and titration phase of the study. Patients randomized to placebo received, in a blinded manner, EXALGO and matching placebo in doses tapering from the stable dose achieved in conversion and titration. During the taper down period, patients were allowed immediate release hydromorphone tablets as supplemental analgesia to minimize opioid withdrawal symptoms in placebo patients. After the taper period, the number of immediate release hydromorphone tablets was limited to two tablets per day. Forty-nine (49) percent of patients treated with EXALGO and 33% of patients treated with placebo completed the 12-week treatment period.

EXALGO provided superior analgesia compared to placebo. There was a (b) (4) significant difference between the mean changes from Baseline to Week 12 or Final Visit in average weekly pain intensity NRS scores obtained from patient diary between the two groups ($p < 0.0001$ in ITT). (b) (4)

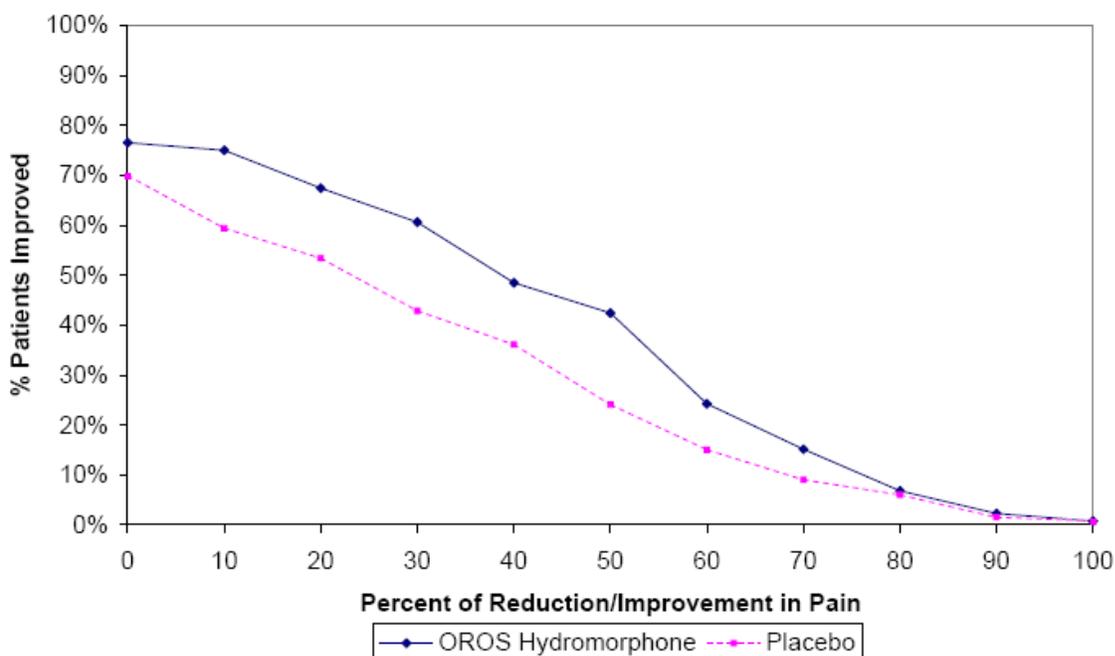
The last sentence in this paragraph is essentially redundant with the penultimate one. The medians may be provided if they are deemed helpful.

This statement is misleading in light of the fact that a common reason for withdrawal was a lack of analgesic effect. The clinical study report does not provide support for a value of (b) (4) and it was not a prospective endpoint. The sentence should be omitted.

This is not a prospective efficacy endpoint, and should be omitted.

The proportion of patients with various degrees of improvement from screening to Week 12 or Final Visit is shown in Figure 4.

Figure 4: Patients achieving various levels of pain relief



This figure uses the imputation method from the SAP for subjects who withdrew during the double-blind phase. Our preferred approach for the responder analysis is to treat these subjects as non-responders and impute zero improvement. Figure 3 in the body of the review shows my version of the responder analysis, which reflects the high dropout rates in each arm. I consider each arm to have 134 patients in the ITT set, and calculated percentages this way. That is the figure which should be included in the label.

APPENDIX A: References

Chuang-Stein, C, Mohberg, NR, Sinkula, MS, 1991, Three measures for simultaneously evaluating benefits and risks using categorical data from clinical trials, *Statistics in Medicine*, 10:1349-1359.

Farrar, JT, Dworkin, RH, and Max, MB, 2006, Use of the Cumulative Proportion of Responders Analysis Graph to Present Pain Data Over a Range of Cut-Off Points: Making Clinical Trial Data More Understandable, *Journal of Pain and Symptom Management*, 31:369-377.

Silverman, DG, O'Connor, TZ and Brull, SJ, 1993, Integrated Assessment of Pain Scores and Rescue Morphine Use During Studies of Analgesic Efficacy, *Anesthesia and Analgesia*, 77:168-170.

APPENDIX B: Additional Statistical Details

The Applicant submitted a revised SAP under which the primary endpoint was to be tested for homogeneity of variance and normality of the residuals. The residuals versus fitted values plot was also to be evaluated for “independent errors”. If the data violated any of these assumptions, then a rank-based ANCOVA method was to be used.

I prefer the ordinary ANCOVA analysis. First of all, this is the analysis that was approved as part of the SPA process, so it is prospective in a very strong sense. Secondly, I find the rank-based ANCOVA model inelegant and difficult to interpret.

The Applicant’s concerns about the assumptions of ANCOVA being violated are not to be lightly dismissed, however, so I will consider them individually. Considering homogeneity of variance, the residuals from the ANCOVA model have almost identical variances in each treatment arm. Furthermore, a scatter plot of the residuals vs. the baseline pain scores does not indicate that the variance is related to this covariate. In regard to normality, the issue is not whether individual observations follow a normal distribution (which the usual tests evaluate), but whether the sample size is large enough for the sample means to approximately follow a normal distribution. In order to evaluate the latter question, I took bootstrap samples (i.e., samples with replacement) of the distribution of residuals from the two arms and found that the means of the samples closely followed a normal distribution. Finally, a plot of standardized residuals vs. fitted values does not indicate any relationship. In summary, a reasonable suite of diagnostics indicates that the ordinary ANCOVA model is an appropriate primary analysis for this study.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-21217	ORIG-1	NEUROMED PHARMACEUTICA LS LTD	DILAUDID CR (HYDROMORPHONE HCL)8/16/32/6

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

/s/

JONATHAN D NORTON
10/28/2009

DIONNE L PRICE
10/28/2009
Concur

THOMAS J PERMUTT
10/29/2009
concur



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

Addendum to Statistical Review and Evaluation

CLINICAL STUDIES

NDA/Serial Number: 21-217
Drug Name: EXALGO (hydromorphone)
Indication(s): Moderate to severe pain in opioid tolerant patients
Applicant: Neuromed Pharmaceuticals
Date(s): Consultation request received: August 4, 2009
Completion date: August 28, 2009
Review Priority: S
Biometrics Division: DB VI
Statistical Reviewer: Ling Chen, Ph.D., Mathematical Statistician, Special Project Team.
Concurring Reviewers: Yi Tsong, Ph.D., Acting Division Director
Medical Division: Controlled Substance Staff
The CSS Team: John Gong, Ph.D., Pharmacologist, OD/CSS
Project Manager: Corrine Moody, OD/CSS
Keywords: Crossover design; Drug abuse potential study; Self-reported endpoint; Multiple endpoints

This is in addendum to the last bullet after Table 5 on page 13 of the review report on Study C-2004-022. This bullet should be replaced by the following:

- one may notice that the mean difference between OROS® 64 mg and IR 8 mg for Overall Liking VAS is 0.23. But the difference in ranks is -0.29 with a standard error of 0.131. Thus, the p-value 0.0347 shows the significantly higher median response in OROS® 64 mg than in IR 8 mg.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-21217	ORIG-1	NEUROMED PHARMACEUTICALS LTD	DILAUDID CR (HYDROMORPHONE HCL)8/16/32/6
NDA-21217	ORIG-1	NEUROMED PHARMACEUTICALS LTD	DILAUDID CR (HYDROMORPHONE HCL)8/16/32/6

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

/s/

LING CHEN
09/21/2009

YI TSONG
09/21/2009



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

Statistical Review and Evaluation

CLINICAL STUDIES

NDA/Serial Number: 21-217/SDN 30
Drug Name: EXALGO (hydromorphone)
Indication(s): Moderate to severe pain in opioid tolerant patients
Applicant: Neuromed Pharmaceuticals
Study Number: C-2004-022
Date(s): Consultation request received: August 4, 2009
Completion date: August 28, 2009
Review Priority: S
Biometrics Division: DB VI
Statistical Reviewer: Ling Chen, Ph.D., Mathematical Statistician, Special Project Team.
Concurring Reviewers: Stella Machado, Ph.D., Acting Team Leader, Division Director
Medical Division: Controlled Substance Staff
The CSS Team: John Gong, Ph.D., Pharmacologist, OD/CSS
Project Manager: Corrine Moody, OD/CSS
Keywords: Crossover design; Drug abuse potential study; Self-reported endpoint; Multiple endpoints

Table of Contents

LIST OF TABLES.....	2
LIST OF FIGURES.....	2
1. EXECUTIVE SUMMARY	3
2. OVERVIEW	4
2.1 OBJECTIVES OF THE STUDY	4
2.2 STUDY DESIGN	4
2.3 NUMBER OF SUBJECTS (PLANNED AND ANALYZED)	4
2.4 SUBJECTIVE ABUSE POTENTIAL MEASURES	5
2.5 STATISTICAL METHODS USED IN THE SPONSOR’S ANALYSIS	5
2.6 CONCLUSIONS FROM THE SPONSOR	6
3. DATA LOCATION	6
4. REVIEWER’S ANALYSES.....	6
4.1 ABUSE POTENTIAL MEASURES CONSIDERED IN THE REVIEWER’S ANALYSES.....	6
4.2 PRIMARY ANALYSIS	7
4.2.1 <i>Descriptive statistics</i>	7
4.2.2 <i>Statistical testing</i>	11
4.2.2.1 Study model and contracts of interest.....	11
4.2.2.2 Results.....	11
4.3 REVIEWER’S SECONDARY ANALYSIS.....	13
4.3.1 <i>Descriptive statistics</i>	13
4.3.2 <i>Statistical testing</i>	15
5. CONCLUSION.....	17

List of Tables

Table 1: Summary Statistics for Emmax of Abuse Potential Measures in Phase A.....	7
Table 2: Summary Statistics for Emmax of Abuse Potential Measures in Phase B.....	8
Table 3: Summary Statistics for Overall Drug Liking VAS at Hours 10 and 48.....	11
Table 4: p-values from the W-test	12
Table 5: Treatment Comparisons for Emmax of Abuse Potential Measures in the Primary Analysis.....	12
Table 6: Summary Statistics for Emmax of Other Abuse Potential Measures.....	14
Table 7: Treatment Comparisons for Emmax of Abuse Potential Measures in the	16

List of Figures

Figure 1: Mean Time Course Profiles for Drug Liking VAS	8
Figure 2: Mean Time Course Profiles for Good Effects VAS.....	9
Figure 3: Mean Time Course Profiles for High VAS.....	9
Figure 4: Mean Time Course Profiles for High VAS.....	9
Figure 5: Mean Time Course Profiles for Opiate Agonist Scale	10
Figure 6: Mean Time Course Profiles for Overall Drug Liking VAS	10

1. Executive Summary

For the review of Study C-2004-022 in NDA 21-217, this reviewer evaluated responses from opiate-experienced, non-dependent recreational drug users (30 in Phase A, and 28 in Phase B) to sixteen abuse potential measures. The results from this reviewer's analyses show that OROS® has large abuse potential, which is larger than IR 8 mg.

Study C-2004-22 was a single-center, single-dose, double-blind, double-dummy, placebo-controlled, randomized, crossover study in healthy subjects who had a history of polydrug use and moderate opiate use, but were not dependent on opiates (DSM-IV-TR).

For assessing the abuse potential of OROS®, IR 8 mg was compared to OROS® 16 mg, OROS® 32 mg, OROS® 8 mg Crushed in Phase A, and to OROS® 64 mg in Phase B. The maximum response during 15 hours after dosing was studied for fifteen abuse potential measures. For Overall Drug Liking, the sponsor only collected data at hours 10 and 48. Thus, data from hour 10 were analyzed for this measure.

Overall OROS® 16 mg had less abuse potential than IR 8 mg. OROS® 32 mg and OROS® 8 mg Crushed did not show significant difference from IR 8 mg for most abuse potential measures. However, OROS® 64 mg had consistently higher mean responses than IR 8 mg for all abuse potential measures in the study, and the differences were significant in 14 out of 16 measures.

Abuse potential of OROS® 32 mg and OROS® 64 mg is evident. Although in comparing OROS® with IR 8 mg, the peak of the mean response of OROS® is delayed to hour 15, a substantially higher response of OROS® than that of placebo could start around hour 6, and last to hour 24 in this study. Because the sponsor did not collect the data between hours 15 and 24, the maximum mean response of OROS® may not have been observed.

Based on the study results the reviewer concludes that the abuse potential of OROS® is higher than that of IR 8 mg. It is recommended that the sponsor provide a detailed risk management plan for OROS®, and the use of OROS® should be well controlled.

2. Overview

2.1 Objectives of the study

The primary objective of the study was to evaluate the abuse potential of single-doses of OROS[®] hydromorphone (controlled-release formulation, intact and crushed), hydromorphone IR (Dilaudid[®], immediate-release formulation), and placebo in opiate-experienced, non-dependent recreational drug users.

A secondary objective was to evaluate the pharmacokinetic/pharmacodynamic relationship of hydromorphone IR and OROS[®] hydromorphone on measures of abuse potential.

Reviewer's Comment: This review focuses on the primary objective.

2.2 Study design

This was a single-center, single-dose, double-blind, double-dummy, placebo-controlled, randomized, crossover study in healthy subjects who had a history of polydrug use and moderate opiate use, but were not dependent on opiates (DSM-IV-TR).

After the screening, subjects that tolerated the hydromorphone IR 8 mg treatment well and were able to discriminate the hydromorphone 8 mg IR dose from placebo (≥ 15 -mm difference in peak score on a 100-mm drug-liking VAS) were enrolled in the study.

There were two phases of the study as follows:

- In Phase A (5-period, 5-sequence, double-dummy, placebo-controlled, randomized, crossover design), each subject received single doses of OROS[®] hydromorphone 16 mg, OROS[®] hydromorphone 32 mg, OROS[®] hydromorphone 8 mg crushed, hydromorphone 8 mg IR (active control), and placebo.
- The subjects who could well tolerate (University of Wisconsin Hospital and Clinic [UWHC] sedation score ≤ 3 , respiratory rate ≥ 8 breaths/minute, vomiting ≤ 2 episodes) all the treatments in Phase A entered Phase B (2-period, 2-sequence, double-dummy, randomized, crossover design), in which the subjects received single doses of OROS[®] hydromorphone 64 mg and hydromorphone 8 mg IR (active control).

The washout period (7-14 days) began immediately after each treatment was administered. Subjects remained at the study site during each treatment period.

2.3 Number of subjects (planned and analyzed)

Screening: Treated n=64

Phase A: Planned n=50; Treated n=38; Completed n=30

Phase B: Treated n=29; Completed n=28

2.4 Subjective Abuse Potential Measures

The primary measure was Overall Drug Liking.

The other abuse potential measures were the following:

Subjective Drug Value (Crossover point, \$)

DRQS (Drug Rating Questionnaire-Subject)

Any Effects VAS, Good Effects VAS, Bad Effects VAS, High VAS, Take Drug Again VAS, Drug Liking VAS

Subjective-rated Opiate Agonist Scale
Composite score per single dose

Addiction Research Center Inventory (Cole/ARCI)

Cole/ARCI Abuse Potential
Cole/ARCI Sedation--Mental
Cole/ARCI Sedation--Motor
Cole/ARCI Stimulation--Euphoria
Cole/ARCI Stimulation--Motor
Cole/ARCI Unpleasantness--Dysphoria
Cole/ARCI Unpleasantness—Physical

Reviewer's Comments: The reviewer was told by the CSS that the Subjective Drug Value is not a valid measure to assess the abuse potential. Therefore, this measure will not be considered in the reviewer's analysis.

Notice that responses to all abuse potential measures listed above were collected at hours 0.5, 1, 2, 4, 6, 8, 12, 15, 24, and 48 after dosing except Overall liking and Crossover point (\$), which were collected at hour 10 and 48.

2.5 Statistical methods used in the sponsor's analysis

To evaluate if average response to any of the pharmacodynamic (PD) parameters was different between single doses of OROS® hydromorphone (controlled-release formulation, intact and crushed), hydromorphone IR, and placebo, a mixed-effects analysis of variance (ANOVA) model was used. This ANOVA model included the fixed-effect factors of treatment, sequence, and period, and the random effects of intersubject and intrasubject factors.

Reviewer's Comments: The statistical model used by the sponsor is different from that used by the reviewer. The difference is the random effects included in the model. The reviewer uses subject nested with sequence as a random effect in the analysis.

2.6 Conclusions from the Sponsor

1. Preference for OROS® hydromorphone 16 mg was significantly lower than hydromorphone 8 mg IR based on all analyses of the primary endpoint, Overall Drug Liking, and maximum scores on the Subjective Effects VAS (Any Effects, Good Effects, High, Take Drug Again, and Drug Liking) and the Cole/ARCI (Stimulation – Euphoria and Abuse Potential). When crushed, OROS® hydromorphone behaved similarly to hydromorphone 8 mg IR.
2. OROS® hydromorphone 32 mg and 64 mg were not significantly different from, but were generally lower than, hydromorphone 8 mg IR, based on the primary endpoint, Overall Drug Liking. Results demonstrated a generally lower drug liking with OROS® hydromorphone than with the IR formulation.
3. The controlled-release delivery of hydromorphone from the OROS® formulation delays effects leading to drug liking: with hydromorphone 8 mg IR, the maximum responses on items such as High and Drug Liking were seen approximately 2 hours after dosing. With OROS® hydromorphone, however, maximum responses occurred approximately 6 to 12 hours after dosing - which has the potential to lessen the appeal of this product to an abuser seeking a rapid high. In addition, doses 4- to 8-fold higher were needed with the OROS® formulation to achieve maximum responses similar to those seen with the 8 mg IR formulation.

3. Data Location

The following was the link of the data sets used in this review.

<\\CDSESUB1\EVSPROD\NDA021217>

4. Reviewer's analyses

4.1 Abuse potential measures considered in the reviewer's analyses

By consultation with primary reviewer Dr. John Gong at the CSS, the abuse potential variables were classified into two groups:

Group 1: Drug Liking VAS, Good Effects VAS, High VAS, Opiate Agonist Scale, Overall Liking VAS, and Take Drug Again VAS.

Group 2: Any Effects VAS, Bad Effects VAS, Cole/ARCI Abuse Potential, Cole/ARCI Sedation—Mental, Cole/ARCI Sedation—Motor, Cole/ARCI Stimulation—Euphoria, Cole/ARCI Stimulation—Motor, Cole/ARCI Unpleasantness—Dysphoria, Cole/ARCI Unpleasantness—Physical, and Composite Score Per Single Dose.

The abuse potential measures in these two groups were considered in the reviewer's primary and secondary analyses, respectively.

Emax (maximum of change from predose response during 15 hours after dosing or maximum response during 15 hours after dosing if the predose response is not assessable, for example, Drug

Liking VAS) for an abuse potential measure is used as a response variable in the model except Overall liking VAS, for which the response at hour 10 from each subject was used.

Notice that the reviewer’s analyses were based on different endpoints and a different model from the sponsor’s analysis. The model used in the reviewer’s analyses was used to assess drug abuse potential for all drug abuse potential studies submitted to the FDA for review. Emax was chosen during 15 hours postdose because OROS® is an extended release formulation, and the peak response is achieved around hour 15.

4.2 Primary Analysis

4.2.1 Descriptive statistics

Table 1-2 summarized the mean, standard deviation, minimum, the first quartile (Q₁), median, the third quartile (Q₃), and maximum for endpoints in Group 1 (excluding Overall Drug Liking VAS) for Phase A and Phase B respectively.

Table 1: Summary Statistics for Emax of Abuse Potential Measures in Phase A

Abuse Potential Variable	Treatment	N	Mean	Std	Min	Q1	Med	Q3	Max
Drug liking	IR 8 mg	30	77.90	16.05	50	64.75	77	92.50	100
	OROS 16 mg	30	64.33	19.76	0	51.00	65	76.25	100
	OROS 32 mg	30	73.30	15.31	50	62.00	72	85.25	100
	OROS 8 mg Crushed	30	75.67	15.46	50	66.00	76	88.25	100
	Placebo	30	42.77	21.06	0	50.00	50	51.25	66
Good effects	IR 8 mg	30	70.67	27.25	0	58.50	75	92.25	100
	OROS 16 mg	30	51.80	32.45	0	23.00	56.5	74.50	100
	OROS 32 mg	30	68.33	24.26	12	51.00	73.5	88.75	100
	OROS 8 mg Crushed	30	65.90	25.33	0	58.00	71	80.75	100
	Placebo	30	22.37	28.55	0	0.00	2.5	51.00	81
High	IR 8 mg	30	70.73	29.03	0	56.50	76.5	94.50	100
	OROS 16 mg	30	51.13	32.65	0	20.25	55.5	78.25	100
	OROS 32 mg	30	67.13	26.46	7	50.75	73	87.25	100
	OROS 8 mg Crushed	30	66.27	27.45	0	56.75	72	83.50	100
	Placebo	30	20.73	27.21	0	0.00	2	51.25	69
Opiate Agonist Scale	IR 8 mg	30	395.83	331.80	-404	82.25	390	691.75	1021
	OROS 16 mg	30	229.43	244.24	-442	115.75	181	333.25	964
	OROS 32 mg	30	384.67	259.78	-59	166.25	402.5	571.50	906
	OROS 8 mg Crushed	30	333.80	258.18	-129	154.25	269.5	541.25	815
	Placebo	30	60.40	205.47	-626	0.00	40.5	113.75	659
Take drug again	IR 8 mg	30	76.67	24.98	3	69.25	77	100.00	100
	OROS 16 mg	30	56.47	26.94	0	48.75	59.5	70.25	100
	OROS 32 mg	30	70.63	20.24	17	51.75	72	86.50	100
	OROS 8 mg Crushed	30	73.10	28.34	0	63.75	81	95.75	100
	Placebo	30	39.00	28.69	0	1.00	50	56.25	100

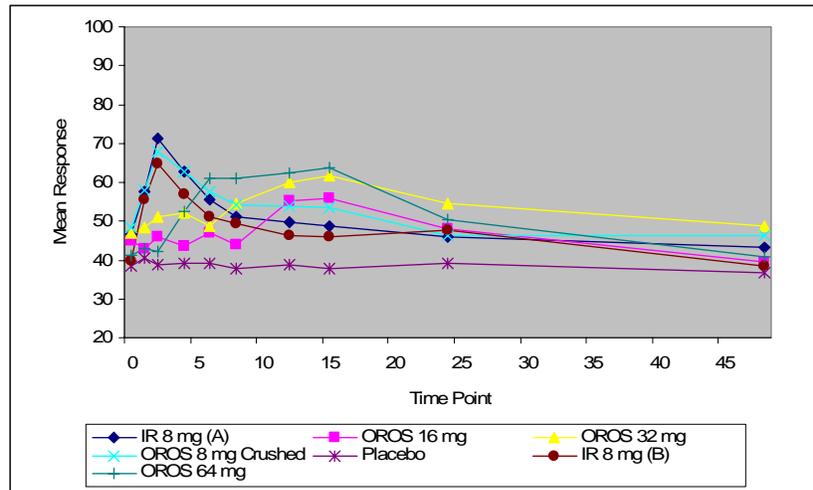
Table 2: Summary Statistics for Emax of Abuse Potential Measures in Phase B

Abuse Potential Variable	Treatment	N	Mean	Std	Min	Q1	Med	Q3	Max
Drug Liking	IR 8 mg	28	70.04	25.26	0	56.25	75.5	86.50	100
	OROS 64 mg	28	79.18	17.67	35	72.00	80	94.50	100
Good effects	IR 8 mg	28	66.71	29.99	0	60.25	73.5	83.75	100
	OROS 64 mg	28	78.75	26.30	7	70.25	83.5	100.00	100
High	IR 8 mg	28	67.82	29.93	0	61.50	76.5	87.25	100
	OROS 64 mg	28	80.43	25.11	5	71.50	88	99.75	100
Opiate Agonist Scale	IR 8 mg	28	343.43	250.98	15	145.25	258.5	493.25	989
	OROS 64 mg	28	484.25	312.94	33	264.25	400.5	753.50	1159
Take drug again	IR 8 mg	28	68.21	29.08	0	58.25	74.5	93.25	100
	OROS 64 mg	28	74.82	27.29	5	55.25	85.5	95.00	100

From Table 1 and 2, one may notice that the mean and median of Emax of potential measures from IR 8 mg group were greater than those from OROS[®] 16 mg and OROS[®] 32 mg groups. But the mean and median of Emax of potential measures from OROS[®] 64 mg were greater than those of IR 8 mg.

Figures 1 and 5 give the mean time course profiles for each abuse potential measures in Group 1.

Figure 1: Mean Time Course Profiles for Drug Liking VAS



In the graphs, IR 8 mg (A) and IR 8 mg (B) represent IR 8 mg in Phases A and B, respectively.

Figure 2: Mean Time Course Profiles for Good Effects VAS

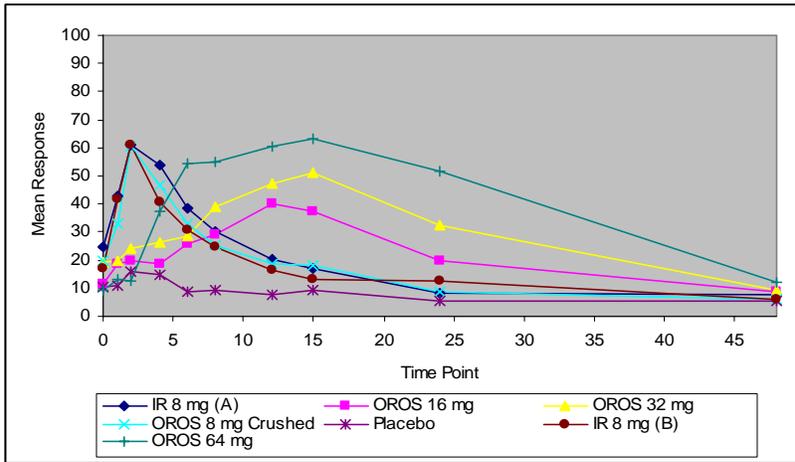


Figure 3: Mean Time Course Profiles for High VAS

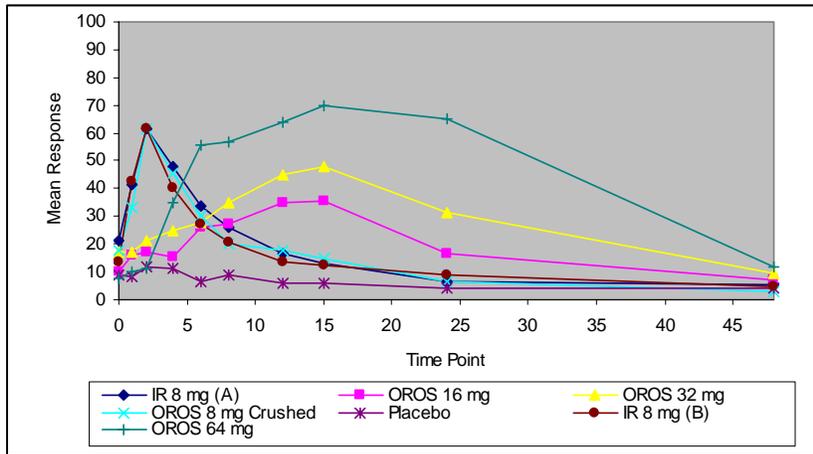


Figure 4: Mean Time Course Profiles for Opiate Agonist Scale

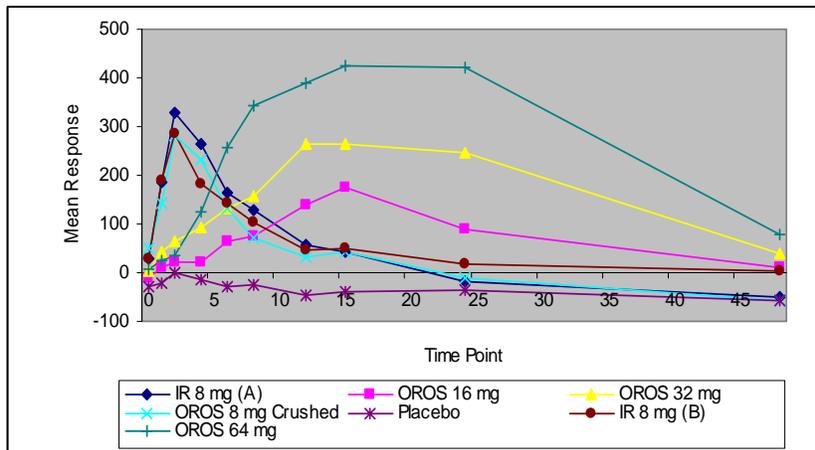
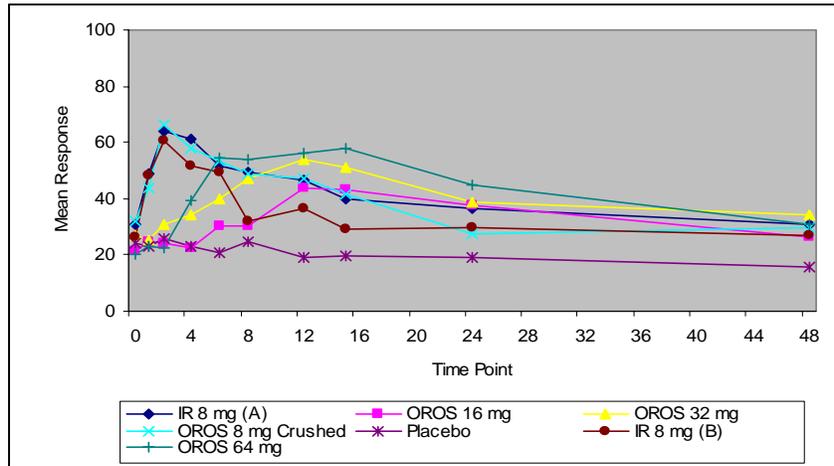


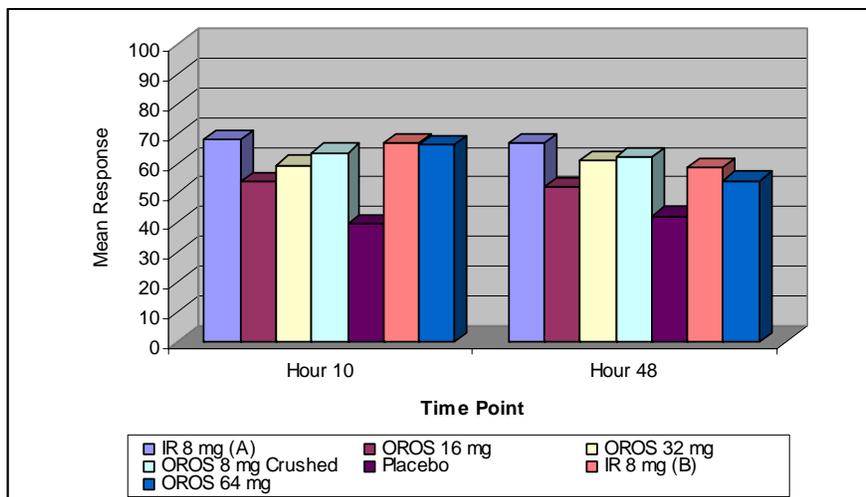
Figure 5: Mean Time Course Profiles for Take Drug Again



From Figures 1-5, one may see that all doses of OROS[®] started their higher mean responses to Drug Liking VAS, Good effects VAS, High VAS and Opiate Agonist Scale than IR 8 mg at hour 6, and lasted to hour 48. The peaks of mean response of all doses of OROS[®] were around hour 15. Data were not collected after hour 15 and before hour 24. It is possibly that the actual peak mean responses were achieved after hour 15. Consistently higher mean responses in the OROS[®] groups were observed between hour 6 and hour 24 for Good effects VAS, High VAS and Opiate Agonist Scale. Because there was no observation between hour 24 and hour 48, the reviewer would not be able to determine the exact time that the drug effects were going down. From those graphics, one may not see much difference in mean time profiles between IR 8 mg and IR 8 mg Crushed. Slight differences were observed between IR 8 mg in Phase A and IR 8 mg in Phase B, because the different treatments and the different subject numbers were in two phases of the crossover designed study.

Figure 6 shows the mean time profile for Overall Drug Liking VAS, which is the primary measure in the sponsor’s analysis.

Figure 6: Mean Time Course Profiles for Overall Drug Liking VAS



Summary statistics for Overall Drug Liking are listed in Table 3.

Table 3: Summary Statistics for Overall Drug Liking VAS at Hours 10 and 48

Abuse Potential Variable	Treatment	N	Mean	Std	Min	Q1	Med	Q3	Max
Hour 10	OROS 16 mg	30	53.77	24.79	0	50.00	54.5	70.75	100
	OROS 32 mg	30	59.53	20.87	0	50.00	62	72.50	93
	OROS 8 mg Crushed	30	63.67	21.04	0	50.75	68	76.50	95
	IR 8 mg (A)	30	68.03	17.00	35	52.75	66.5	78.25	100
	Placebo	30	39.77	20.23	0	43.00	50	50.25	54
	IR 8 mg (B)	28	66.68	23.86	0	54.00	73	81.50	100
	OROS 64 mg	28	66.32	24.23	9	51.25	69.5	82.50	100
Hour 48	OROS 16 mg	30	52.17	24.19	0	49.75	54	70.25	87
	OROS 32 mg	30	61.00	15.00	28	50.00	58.5	70.25	95
	OROS 8 mg Crushed	30	62.30	14.63	50	50.00	52	74.00	93
	IR 8 mg (A)	30	66.87	17.04	42	50.00	62.5	85.25	100
	Placebo	30	42.33	17.40	0	49.00	50	50.00	55
	IR 8 mg (B)	28	58.50	31.54	0	50.00	63.5	80.75	100
	OROS 64 mg	28	54.21	25.95	0	47.25	57.5	76.00	90

The summary statistics and the graphics did not take into account the possible effects due to treatment periods and sequences used in the crossover design study.

4.2.2 Statistical testing

4.2.2.1 Study model and contrasts of interest

The statistical model used in the reviewer's analysis includes sequence, treatment, and period as fixed effects, and subject nested within sequence as a random effect. The model assumption of the normality of error terms was checked using Shapiro-Wilk W-test on the residuals using $\alpha=0.05$. If the normal assumption was not satisfied, the rank data (ranking within subject) were used to obtain the p-value of the test for difference in medians between two treatments.

Five contrasts were studied by the reviewer. These contrasts are: IR 8 mg versus placebo (validating the study), IR 8 mg versus OROS® 16 mg, IR 8 mg versus OROS® 32 mg, and IR 8 mg versus OROS® 8 mg Crushed in Phase A of the study, and IR 8 mg versus OROS® 64 mg in Phase B of the study. Because OROS® has drug abuse potential whenever any dose of OROS® shows abuse potential, the dose response of OROS® is not of interest in this review.

4.2.2.2 Results

The statistical analyses are based on Emax of the abuse potential measures. Table 4 shows the results from the Shapiro-Wilk W-test on the residuals of the model used in the analysis.

Table 4: p-values from the W-test

Abuse Potential Measure	Phase A		Phase B	
	W statistic	p-value	W statistic	p-value
Drug Liking VAS	0.97324	0.0075	0.92544	0.0020
Good Effects VAS	0.98398	0.1067	0.95926	0.0561
High VAS	0.98056	0.0524	0.91998	0.0012
Opiate Agonist Scale	0.98619	0.1669	0.96977	0.1717
Overall Liking	0.96048	0.0005	0.94829	0.0178
Take Drug Again	0.97568	0.0155	0.91148	0.0006

Tables 5 list the statistical analysis results from the reviewer’s primary analysis. For the Emmax of abuse potential measures which satisfies the normality assumption of the model used in the analysis, the difference of least square means, standard error, p-values of the t-test, and the lower and upper 95 % confidence interval limits are listed. For those endpoints where the model assumption was not satisfied, the p-values from the rank test are listed. The least square means and standard error are presented in the original scales.

Table 5: Treatment Comparisons for Emmax of Abuse Potential Measures in the Primary Analysis

Measure	TRT (versus IR 8 mg)	N	Lsmean Diff	StdErr	p-value (t)	95% CI		p-value (Rank)
						Low Limit	Upper Limit	
Drug Liking VAS	OROS 16 mg	30	13.74	4.53				<.0001
	OROS 32 mg	30	5.08	4.53				0.0248
	OROS 8 mg Crushed	30	2.39	4.54				0.1054
	Placebo	30	35.81	4.52				<.0001
	OROS 64 mg	28	-9.18	3.62				<.0001
Good Effects VAS	OROS 16 mg	30	21.10	7.12	0.0038	6.98	35.22	
	OROS 32 mg	30	5.26	7.13	0.462	-8.88	19.40	
	OROS 8 mg Crushed	30	6.85	7.15	0.3405	-7.34	21.03	
	Placebo	30	50.92	7.12	<.0001	36.79	65.05	
	OROS 64 mg	28	-11.73	4.25	0.0104	-20.46	-2.99	
High VAS	OROS 16 mg	30	24.12	8.28	0.0045	7.68	40.56	
	OROS 32 mg	30	8.57	8.27	0.3029	-7.85	24.99	
	OROS 8 mg Crushed	30	8.87	8.25	0.285	-7.50	25.24	
	Placebo	30	54.78	8.22	<.0001	38.46	71.09	
	OROS 64 mg	28	-12.30	4.95				0.0021
Opiate Agonist Scale	OROS 16 mg	30	248.51	58.64	<.0001	132.22	364.80	
	OROS 32 mg	30	92.63	58.45	0.1161	-23.28	208.55	
	OROS 8 mg Crushed	30	146.24	58.70	0.0143	29.84	262.64	
	Placebo	30	426.72	58.42	<.0001	310.87	542.57	
	OROS 64 mg	28	-136.96	28.66	<.0001	-195.87	-78.05	

Table 5 continued.

Overall Liking VAS	OROS 16 mg	30	18.52	5.35			0.0002
	OROS 32 mg	30	13.19	5.35			0.0236
	OROS 8 mg Crushed	30	8.65	5.35			0.2344
	Placebo	30	33.04	5.34			<.0001
	OROS 64 mg	28	0.23	3.30			0.0347
Take Drug Again VAS	OROS 16 mg	30	25.72	6.65			<.0001
	OROS 32 mg	30	11.98	6.66			0.014
	OROS 8 mg Crushed	30	8.67	6.64			0.2935
	Placebo	30	43.01	6.63			<.0001
	OROS 64 mg	28	-6.46	4.52			0.0148

From Table 5, one may see that at $\alpha = 0.05$ for the endpoints considered in the reviewer's analyses

- the mean (or median) response of IR 8 mg is significantly greater than that of placebo for all six abuse potential measures (the study is validated);
- the mean (or median) response of OROS® 16 mg is significantly lower than that of IR 8 mg for all six abuse potential measures;
- there is no significant difference in mean (or median) between OROS® 32 mg and IR 8 mg in Good effects VAS, High VAS, and Opiate Agonist Scale;
- There is no significant difference in mean (or median) response between OROS® 8 mg Crushed and IR 8 mg, except in Opiate Agonist Scale, for which IR 8 mg shows significantly higher mean response than OROS® 8 mg Crushed;
- the mean (or median) response of OROS® 64 mg is significantly greater than that of IR 8 mg in all abuse potential measures;
- one may notice that the mean difference between OROS® 8 mg Crushed and IR 8 mg is 0.23. But the difference in ranks is -0.29 with a standard error of 0.131. Thus, the p-value 0.0347 shows the significantly higher median response in OROS® 8 mg Crushed than in IR 8 mg.

4.3 Reviewer's Secondary Analysis

4.3.1 Descriptive statistics

Table 6 summarizes the mean, standard deviation, minimum, the first quartile (Q1), median, the third quartile (Q3), and maximum for Emaxs of Any Effects VAS, Bad Effects VAS, Cole/ARCI Abuse Potential, Cole/ARCI Sedation – Mental, Cole/ARCI Sedation – Motor, Cole/ARCI Stimulation - Euphoria, Cole/ARCI Stimulation - Motor, Cole/ARCI Unpleasantness – Dysphoria, Cole/ARCI Unpleasantness – Physical, and Composite Score Per Single Dose.

Table 6: Summary Statistics for Emaxs of Other Abuse Potential Measures

Abuse Potential Variable	Treatment	N	Mean	Std	Min	Q1	Med	Q3	Max
Any effects	IR 8 mg (A)	30	69.67	27.11	0	62.75	68	93	100
	OROS 16 mg	30	52	31.32	0	27.75	57	76	100
	OROS 32 mg	30	67.4	27.15	9	47.5	74.5	92.25	100
	OROS 8 mg Crushed	30	65.57	25.61	0	57.75	72	81.5	100
	Placebo	30	23.7	30.82	0	0	2.5	50.25	88
	IR 8 mg (B)	28	10.79	19.38	0	0	0	9.5	57
	OROS 64 mg	28	69.86	26.81	0	61.5	74	90.8	100
Bad effects	IR 8 mg (A)	30	24.17	25.54	0	1	13	50.25	70
	OROS 16 mg	30	26	29.11	0	0	8.5	51	100
	OROS 32 mg	30	27.63	27.86	0	0	16.5	51	94
	OROS 8 mg Crushed	30	26.9	23.99	0	1.5	24	51.75	70
	Placebo	30	13.43	23.73	0	0	0	16.5	81
	IR 8 mg (B)	28	6.29	15.24	0	0	0	0	52
	OROS 64 mg	28	27.36	29.34	0	0	14.5	50	93
Cole/ARCI Abuse Potential	IR 8 mg (A)	30	7.17	4.4	-1	3	7	10	17
	OROS 16 mg	30	4.23	3.38	0	2	3.5	6	14
	OROS 32 mg	30	5.83	3.98	0	3	5	7.25	16
	OROS 8 mg Crushed	30	6.4	3.45	0	3	6.5	9	13
	Placebo	30	1.47	2.85	-3	0	1	3	10
	IR 8 mg (B)	28	-0.21	2.25	-6	-0.8	0	0	4
	OROS 64 mg	28	2.61	5.58	-8	0	3	6.5	16
Cole/ARCI Sedation--Mental	IR 8 mg (A)	30	10.33	5.84	2	6	8.5	15.25	23
	OROS 16 mg	30	8.6	6.2	-3	3	8.5	14	24
	OROS 32 mg	30	10.57	5.42	-1	6	11	15	20
	OROS 8 mg Crushed	30	8.63	5.58	0	5.5	8	11	24
	Placebo	30	3.73	4.43	-1	0	2	6.25	18
	IR 8 mg (B)	28	2.5	3.47	-4	0	2	4	11
	OROS 64 mg	28	8.46	7.36	-4	3	6	14.8	22
Cole/ARCI Sedation--Motor	IR 8 mg (A)	30	6.83	5.75	0	2	5	10.25	19
	OROS 16 mg	30	3.77	4.95	-2	0	2	6.5	20
	OROS 32 mg	30	5.3	4.99	0	1	4	9	22
	OROS 8 mg Crushed	30	4.43	5.06	0	0	3	8.25	21
	Placebo	30	1.73	2.74	0	0	0	3	13
	IR 8 mg (B)	28	0.21	2.04	-3	0	0	0	8
	OROS 64 mg	28	6.14	6.14	0	1	3	9	19
Cole/ARCI Stimulation--Euphoria	IR 8 mg (A)	30	14.9	9.91	-3	9	15.5	21.25	42
	OROS 16 mg	30	7.37	6.86	-1	1.75	6	12	23
	OROS 32 mg	30	10.53	9.77	-6	2.75	9.5	17.25	33
	OROS 8 mg Crushed	30	12.47	8.46	0	4.75	12	19	33
	Placebo	30	2.03	3.61	-2	0	0	4.25	11
	IR 8 mg (B)	28	1.46	4.25	-2	0	0	0.8	16
	OROS 64 mg	28	7.68	8.55	-4	0	6	13.3	30

Table 6 continued.

Cole/ARCI Stimulation--Motor	IR 8 mg (A)	30	3.43	2.19	-2	2.00	3	5.25	7
	OROS 16 mg	30	2.17	1.86	0	0.00	2	4.00	6
	OROS 32 mg	30	3.03	2.37	-1	1.00	3	4.25	9
	OROS 8 mg Crushed	30	3.23	2.08	0	2.00	3.5	5.00	8
	Placebo	30	0.83	1.53	-2	0.00	0	1.25	4
	IR 8 mg (B)	28	0.25	1.24	-2	0.0	0	0.0	4
	OROS 64 mg	28	2.71	2.57	-2	0.0	3	5.0	7
Cole/ARCI Unpleasantness-- Dysphoria	IR 8 mg (A)	30	3.93	2.60	0	2.00	4	5.00	12
	OROS 16 mg	30	2.87	2.79	0	0.75	2	4.00	10
	OROS 32 mg	30	2.93	2.13	-1	1.00	3	5.00	7
	OROS 8 mg Crushed	30	3.20	3.13	0	0.00	2	6.00	9
	Placebo	30	0.67	2.07	-4	0.00	0	2.00	7
	IR 8 mg (B)	28	0.71	1.46	0	0.0	0	1.0	6
	OROS 64 mg	28	3.57	3.12	0	2.0	3	5.0	15
Cole/ARCI Unpleasantness-- Physical	IR 8 mg (A)	30	2.70	3.02	0	0.00	2	4.00	13
	OROS 16 mg	30	2.00	2.41	0	0.00	2	3.00	11
	OROS 32 mg	30	2.13	2.08	-1	0.75	2	3.00	7
	OROS 8 mg Crushed	30	2.17	2.96	-1	0.00	1	2.50	10
	Placebo	30	0.70	1.66	-2	0.00	0	1.00	6
	IR 8 mg (B)	28	0.86	1.98	-1	0.0	0	0.8	7
	OROS 64 mg	28	2.79	3.98	-2	0.0	2	4.5	14
Composite score per single dose	IR 8 mg (A)	30	3.07	2.16	0	0.00	3.5	5.00	7
	OROS 16 mg	30	2.63	1.94	0	0.00	3	4.00	6
	OROS 32 mg	30	2.90	1.88	0	1.50	3	4.25	6
	OROS 8 mg Crushed	30	2.07	1.96	0	0.00	2	4.00	5
	Placebo	30	1.03	1.47	0	0.00	0	2.00	4
	IR 8 mg (B)	28	0.79	1.52	0	0.0	0	1.0	6
	OROS 64 mg	28	3.39	2.28	0	1.3	4	5.0	7

Note that the summary statistics did not take into account the possible effects from treatment periods and sequences used in a crossover design study.

4.4.2 Statistical testing

The same statistical testing procedures as the primary analysis were used in the secondary analysis. Tables 7 show the test results for the endpoints which satisfied, or did not satisfy the model assumption of normality. For those that did not satisfy the model assumption, the p-values of the rank tests are presented in the table. The least square means and standard error are presented in the original scales.

Table 7: Treatment Comparisons for Emmax of Abuse Potential Measures in the Secondary Analysis

Measure	TRT (versus IR 8 mg)	N	Lsmean Diff	StdErr	p-value (t)	95% CI		p-value (Rank)
						Lower	Upper	
Any effects	OROS 16 mg	30	17.41	6.36	0.0072	4.81	30.02	
	OROS 32 mg	30	1.90	6.36	0.7661	-10.71	14.51	
	OROS 8 mg Crushed	30	3.90	6.39	0.5431	-8.76	16.55	
	Placebo	30	46.04	6.39	<.0001	33.39	58.69	
	OROS 64 mg	28	-12.80	4.48				0.0007
Bad effects	OROS 16 mg	30	-1.07	5.21				0.7715
	OROS 32 mg	30	-3.19	5.21				0.6015
	OROS 8 mg Crushed	30	-2.07	5.23				0.1564
	Placebo	30	10.93	5.23				0.0100
	OROS 64 mg	28	-13.37	6.11	0.0378	-25.94	-0.81	
Cole/ARCI Abuse Potential	OROS 16 mg	30	2.98	0.85	0.0006	1.31	4.66	
	OROS 32 mg	30	1.44	0.85	0.0920	-0.24	3.11	
	OROS 8 mg Crushed	30	0.90	0.85	0.2928	-0.78	2.58	
	Placebo	30	5.92	0.85	<.0001	4.24	7.60	
	OROS 64 mg	28	-0.49	0.96				0.4612
Cole/ARCI Sedation--Mental	OROS 16 mg	30	1.75	1.14				0.0867
	OROS 32 mg	30	-0.19	1.14				0.6943
	OROS 8 mg Crushed	30	1.61	1.14				0.0856
	Placebo	30	6.62	1.14				<.0001
	OROS 64 mg	28	-3.58	1.42	0.0178	-6.50	-0.67	
Cole/ARCI Sedation--Motor	OROS 16 mg	30	3.10	1.00				0.0006
	OROS 32 mg	30	1.64	1.00				0.1522
	OROS 8 mg Crushed	30	2.43	1.00				0.0352
	Placebo	30	5.17	1.00				<.0001
	OROS 64 mg	28	-4.75	1.17	0.0004	-7.15	-2.35	
Cole/ARCI Stimulation--Euphoria	OROS 16 mg	30	7.49	1.77				0.0001
	OROS 32 mg	30	4.59	1.77				0.0007
	OROS 8 mg Crushed	30	2.67	1.78				0.4927
	Placebo	30	13.32	1.78				<.0001
	OROS 64 mg	28	-2.44	1.59	0.1382	-5.71	0.84	
Cole/ARCI Stimulation--Motor	OROS 16 mg	30	1.25	0.41	0.0027	0.45	2.06	
	OROS 32 mg	30	0.47	0.41	0.2569	-0.34	1.28	
	OROS 8 mg Crushed	30	0.23	0.41	0.5750	-0.58	1.04	
	Placebo	30	2.71	0.41	<.0001	1.90	3.52	
	OROS 64 mg	28	-0.77	0.53	0.1563	-1.85	0.31	

Table 7 continued.

Cole/ARCI Unpleasantness-- Dysphoria	OROS 16 mg	30	1.09	0.53				0.0002
	OROS 32 mg	30	0.98	0.53				0.0034
	OROS 8 mg Crushed	30	0.81	0.53				0.0132
	Placebo	30	3.29	0.53				<.0001
	OROS 64 mg	28	-1.96	0.76				0.0148
Cole/ARCI Unpleasantness-- Physical	OROS 16 mg	30	0.68	0.51				0.0681
	OROS 32 mg	30	0.58	0.51				0.2337
	OROS 8 mg Crushed	30	0.52	0.51				0.0548
	Placebo	30	1.95	0.51				<.0001
	OROS 64 mg	28	-2.82	0.68				<.0001
Composite score per single dose	OROS 16 mg	30	0.44	0.45	0.3272	-0.44	1.32	
	OROS 32 mg	30	0.20	0.45	0.6586	-0.68	1.08	
	OROS 8 mg Crushed	30	1.06	0.45	0.0195	0.17	1.94	
	Placebo	30	2.12	0.45	<.0001	1.24	3.01	
	OROS 64 mg	28	-1.49	0.33	0.0001	-2.17	-0.82	

The following study results are for the endpoints in the reviewer's secondary analysis.

- The mean (or median) response of IR 8 mg is significantly greater than that of placebo for all 10 abuse potential measures (the study is validated).
- The mean (or median) response of OROS® 16 mg is significantly lower than that of IR 8 mg for 6 out 10 abuse potential measures. No significant difference in mean (or median) between IR 8 mg and OROS® 16 mg is found for Bad Effects VAS, Cole/ARCI Sedation – Motor, Cole/ARCI Unpleasantness Physical, and Composite Score per Single Dose.
- The significantly higher mean (or median) response in IR 8 mg than OROS® 32 mg is found only for Cole/ARCI Stimulation – Euphoria, and Cole/ARCI Unpleasantness – Dysphoria.
- No significant difference is found between IR 8 mg and OROS® 8 mg Crushed for 7 abuse potential measures. For Cole/ARCI Sedation – Motor, Cole/ARCI Unpleasantness – Dysphoria and Composite Score per Single Dose, the IR 8 mg has higher mean (or median) response than OROS® 8 mg Crushed.
- OROS® 64 mg has higher mean (or median) response than IR 8 mg for all 10 abuse potential measures. The differences are significant in 8 out of 10 comparisons. The abuse potential measures that do not show the significance are Cole/ARCI Stimulation – Euphoria and Cole/ARCI Simulation – Motor.

5. Conclusion

Abuse potential of OROS® 32 mg and OROS® 64 mg is evident. Although comparing OROS® with IR 8 mg, the peak of mean response of OROS® is delayed to hour 15, a substantially higher response of OROS® than that of placebo could start around hour 6, and last to hour 24 in this study. Because the sponsor did not collect the data between hour 24 and 48, the maximum response of OROS® may not be observed. The reviewer concludes that the abuse potential of OROS® is higher than that of IR 8 mg. Therefore, it is recommended that the sponsor provide a detailed risk management plan for OROS®, and the use of OROS® should be well controlled.

Linked Applications	Submission Type/Number	Sponsor Name	Drug Name / Subject
NDA 21217	ORIG 1	NEUROMED PHARMACEUTICA LS LTD	DILAUDID CR (HYDROMORPHONE HCL)8/16/32/6
NDA 21217	ORIG 1	NEUROMED PHARMACEUTICA LS LTD	DILAUDID CR (HYDROMORPHONE HCL)8/16/32/6

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

/s/

LING CHEN
08/26/2009

STELLA G MACHADO
08/27/2009

Statistical Review and Evaluation

NDA 21-217

Name of drug: Dilaudid CR (hydromorphone HCl) controlled-release tablets

Sponsor: Knoll

Indication: pain

Documents reviewed: volumes 1.1, 1.118-169; electronic data

Dates: received 29 December 1999; user fee goal 29 October 2000 (10 months, type 3S)

Project manager: Judit Milstein

Medical officer: Douglas Kramer, M.D.

Reviewer: Thomas Permutt

INTRODUCTION

Hydromorphone is a pre-1938 opioid analgesic marketed in several oral and parenteral dosage forms. The subject NDA concerns a new, sustained-release tablet intended for dosing once a day. The statistical section reports four Phase III clinical trials. Studies 104 and 105 were nonrandomized trials designed mainly to determine the appropriateness of a proposed regimen for converting patients from other drugs and titrating the dose of hydromorphone. Study 119 was the only randomized trial. Study 109 was an open-label extension for patients from any of the other three studies.

In view of the existence of other approved formulations of hydromorphone for the same indication, the division and the applicant agreed that a single well-controlled trial might serve as the basis for a finding of efficacy of the test drug. Study 119 is clearly the only candidate. After consultation with clinical reviewers, therefore, I shall focus on the evidence of efficacy from that study.

Study 119 was a multi-center, randomized, double-blind, double-dummy, comparison of three treatments. After conversion from other drugs to immediate-release (IR) hydromorphone and titration to a stable dose, patients were assigned to one of three groups. One group continued to receive IR hydromorphone five times daily at the same dose. A second group received an approximately equal daily dose of the sustained-release (SR) product once a day, with placebo IR tablets at the other dosing times. The third group also had the sustained-release product but at half the dose of the second group. All groups had access to IR hydromorphone as rescue medication.

The study thus had the potential to provide several kinds of information about the effects of the test drug. It was an active-controlled comparison of the SR preparation to an IR formulation of the same drug, in which it was anticipated that there would be no substantial difference in outcome. Such a finding of no difference would be strengthened by finding a difference between the two SR doses, as this would be evidence of the trial's ability to detect a difference between active drugs if there was a difference. Furthermore, a difference in outcome between the two SR doses would be evidence in itself of the efficacy of the test drug: if the full dose was better than the half dose, it must also be better than nothing, unless the half dose were worse than nothing.

THE PROTOCOL

The protocol for study 119 identified the primary endpoint and statistical method. The primary variable of interest was the consumption of rescue medication. This was to be averaged over the last four days of the seven-day double-blind treatment period. A baseline measure was defined as the average consumption over the last two days of the titration period. The mean changes from baseline for the three treatments were to be compared by analysis of variance. The primary comparison was to be between the full and half dose sustained-release groups.

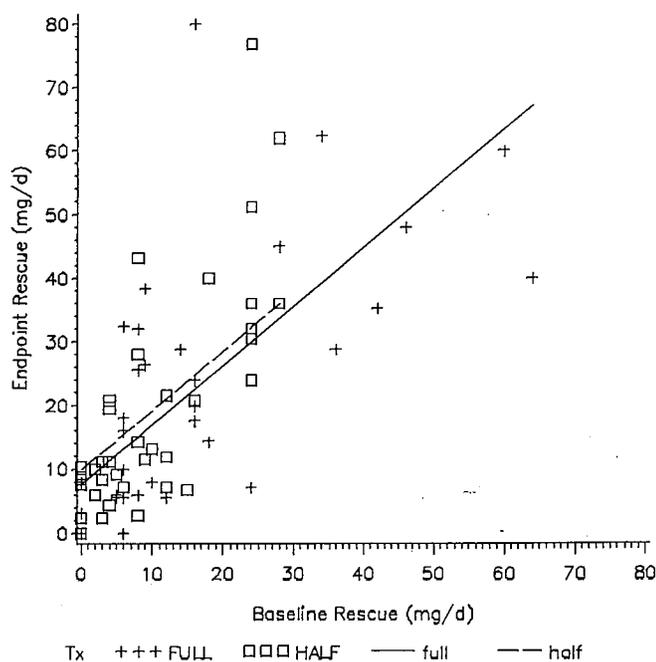
The protocol described an analysis of variance with terms for center and center-by-treatment interaction "if appropriate," with additional language about the handling of small centers. As it called for 25 patients per treatment group to be randomized and as there were 15 centers, it is hard to see under what circumstances this analysis would have been considered "appropriate." In any case, the study report silently dispenses with center effects and interactions.

Measurements of pain were considered a secondary outcome because it was anticipated that rescue medication would be used in such a way that the pain outcomes would not be very different between groups but the consumption of rescue medication would be. For the purposes of review, this particular distinction between primary and secondary outcomes is somewhat artificial. If a comparison of rescue use favored the test drug but pain outcomes went the other way, a finding of efficacy would be doubtful. Still, it was appropriate to identify in the protocol the variable in which favorable differences were expected.

THE DATA

The figure shows the primary outcome data (vertically) along with the baseline data (horizontally) for the two SR groups. Squares represent the half-strength and crosses the full-strength group. Parallel least-squares (ANCOVA) lines are also shown; these will be discussed later. The broken line is for the half-strength and the solid line for the full-strength group.

The groups were unbalanced at baseline. All six patients requiring more than 30 mg of rescue at baseline were randomized to the full-strength group while the half-strength group had 16 of the 19 patients with baseline values of 4 mg or less.



PRIMARY ANALYSIS

These data (average daily rescue dose of hydromorphone, mg) are summarized in the table below, compiled from tables 24 and 25 of the study report (v. 119, pp. 69, 71).

	SR/Full	SR/Half	IR
n	34	40	39
Endpoint:			
mean ± s.d.	23.2 ± 19.8	19.1 ± 17.5	21.4 ± 23.8
median	18.0	11.4	16
Baseline:			
mean ± s.d.	16.4 ± 16.3	10.7 ± 9.9	13.7 ± 14.5
median	9.0	8.0	8.0
Change:			
mean ± s.d.	6.6 ± 16.0	9.2 ± 12.0	7.1 ± 14.5
median	2.0	7.4	4.4

The use of rescue medication was actually higher in the full-strength than in the half-strength group. The difference between groups at endpoint was less than at baseline, however, so that the change from baseline was greater in the half-strength group.

According to the submission, “Some of the parametric analyses [including the primary analysis, apparently] planned in the protocol were found to be inappropriate due to the distribution of the results. In these instances, non-parametric methods were adopted.” In particular, a rank-sum test for the significance of the difference between SR treatments in change from baseline gave a two-sided p-value of 0.16. I am not sure the distribution was such as to make the prespecified analysis “inappropriate,” but there was a single fairly influential observation in each group (the two points at the top of the figure). In any case, I calculate the p-value for the prespecified analysis (t-test, assuming no center effects) to be 0.42.

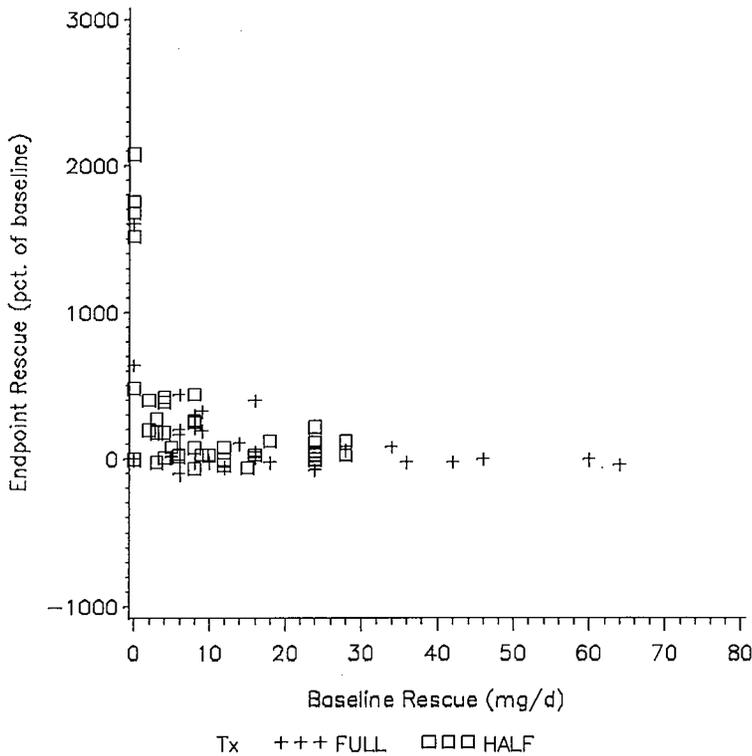
ALTERNATIVE ANALYSES (APPLICANT’S)

The applicant conducted a post-hoc alternative analysis of the same primary data. Changes from baseline were converted to percent changes by dividing the change by the baseline score and

multiplying by 100. In the 10 cases (7 half-strength, 3 full-strength) where the baseline score was zero, a baseline score of 0.5 mg was used as the denominator. Some sensitivity analysis was carried out by varying the 0.5 mg figure; details are not given, but the applicant asserts that the results did not change much. The results of this analysis are shown in the table below (percent change in rescue medication from baseline to endpoint, from applicant's table 27, v, 119, p.73).

	SR/Full	SR/Half	IR
n	33	38	39
mean \pm s.d.	133 \pm 309	330 \pm 583	386 \pm 1277
median	18	94	53

The numerical results seem impressively different, especially the medians. The rank-sum test applied to the two SR groups gives a two-sided p-value of 0.037.



The figure plots the calculated percent change against the baseline score. There are some very high percent changes from very low or zero scores to scores that are still rather low; for example, there is a change of about 2000 percent from 0 (replaced by 0.5 mg) to 10.4 mg, still less than the median rescue dose for all groups, which was 14 mg. These extreme percent changes are influential in computing the means, and there are enough of them to influence the median substantially as well, particularly for the half-strength group.

The sponsor also gives a normalized analysis, in which endpoint rescue doses were expressed as a fraction of the prescribed dose rather than of the baseline rescue dose. No significant difference was found ($p = 0.12$, rank-sum test).

Another secondary analysis concerned the number of times per day that rescue medication was taken. Like the normalized analysis, this weights the patients slightly differently than does the primary analysis. The size of each rescue dose was approximately proportional to the prescribed dose, so that patients with high prescribed doses count more heavily when rescue is analyzed by

milligrams than by number of rescue doses. The table below (number of times per day rescue medication was taken) was compiled from tables 24 and 29 of the report (v. 119, pp. 69, 76).

	SR/Full	SR/Half	IR
n	33	38	39
Endpoint:			
mean \pm s.d.	2.9 \pm 1.9	3.2 \pm 1.5	2.4 \pm 1.6
median	2.6	3.0	2.6
Baseline:			
mean \pm s.d.	2.1 \pm 0.9	1.8 \pm 1.1	1.7 \pm 1.2
median	2.0	2.0	2.0
Change:			
mean \pm s.d.	0.8 \pm 1.9	1.4 \pm 1.6	0.7 \pm 1.4
median	0.25	1.05	0.40

The applicant gives p-values for the rank-sum test for the two SR groups of 0.19 (endpoint) and 0.026 (change from baseline). In addition, a p-value of 0.0001 is reported by "Poisson regression." Details of the Poisson regression analysis are not given. A printout shows effects for treatment and day but not for subject, suggesting that the individual daily observations for each patient (rather than the total over the last four days) may have been used in this analysis and that the four resulting observations for each patient may have been incorrectly treated as independent. In this case the variability of the estimated treatment effect would have been underestimated and the p-value would be spuriously small.

Finally, there is an analysis using only the signs of the changes from baseline. Eleven of 33 patients in the full-strength group decreased their rescue dose from baseline to endpoint, 2 stayed the same and 20 increased. In the half-strength group, 4 of 38 decreased, 3 stayed the same and 31 increased. A p-value of 0.0259 is given "based on Cochran-Mantel-Haenszel test for nonzero correlation of two ordinal measures."

ALTERNATIVE ANALYSES (REVIEWER'S)

As I noted above, the treatment groups were unbalanced at baseline. Of the patients who got one of the SR treatments, the highest users of rescue medication were randomized exclusively to the full strength and the lowest users were randomized disproportionately to the low strength. There is no one correct way to adjust for baseline differences in analysis. The method proposed in the

protocol was an acceptable one. It controlled the probabilities of errors of both kinds at levels that the sponsor considered acceptable prospectively. It failed to detect a treatment effect. Either there was in fact no treatment effect, or a Type II error occurred; and it is impossible to know which. I believe precedence ought to be given to the results of this primary analysis. However, a number of secondary analyses are presented, which seem to suggest that the primary analysis may understate the treatment effect. Again, I believe this question cannot be answered definitively by post-hoc analyses. I think, however, that it is more likely that the primary analysis *overstates* the effect in this case and the alternative analyses do so even more.

If two groups different at baseline were assigned to the *same* treatment, the difference between the groups would be expected to be less at endpoint than at baseline. This is regression to the mean. The less strongly the baseline and endpoint measurements are correlated, the more regression is to be expected. Subtracting baseline values without allowing for regression, therefore, tends to exaggerate the effect of baseline differences: the group that was worse at baseline has an improvement attributed to treatment that is likely due in part to regression.

The method of adjusting for baseline that I generally recommend in reviewing protocols is analysis of covariance. This amounts to subtracting only a fraction of the baseline value, the fraction being determined by the correlation of the baseline and outcome measurements, along with their variances, which allows for the expected effect of regression to the mean. It can also be seen as fitting parallel, least-squares lines to the treatment groups.

Such parallel lines are shown in the first figure above. They are very close together, and the treatment effect is not statistically significant ($p = 0.52$). In other words, the differences at endpoint, as well as the differences in change from baseline, are very nearly what would have been expected from two *identically treated* groups that were different at baseline. The significance of these differences is probably exaggerated by the primary analysis. It is exaggerated even more by the percent change from baseline. There the smallest baseline scores, which would be expected to increase the most (and, in the case of zeroes, could only increase or stay the same), were especially influential. The same obviously applies to the findings on direction of change: zero or very small baseline scores were likely to go up regardless of treatment, and such scores were disproportionately randomized to the half-strength group.

CONCLUSIONS

Study 119 failed to show a difference between full and half strengths of the sustained-release drug according to the analysis specified in the protocol. Alternative analyses were presented that appear to indicate a difference. While objective evaluation of post-hoc analysis is difficult at best, I believe that the alternative analyses tend to obscure rather than to elucidate the main features of the data. The full-strength group used *more* rescue medication than the half-strength group. Allowing for baseline differences in the way I think most appropriate, this paradoxical dose effect goes away. The primary analysis, which overadjusts for baseline, still fails to show a statistically significant difference. The alternative analyses overcompensate still more for putative effects of baseline differences. They are heavily influenced by rather small changes from very small baselines. Some of these secondary analyses produce p-values in the range of 0.02 to 0.05. Such borderline results in

post-hoc analyses would be unpersuasive on grounds of multiplicity even if they were fundamentally sound. The only very small p-value, from "Poisson regression," appears to be incorrect.

Thomas Permutt 6/9/00

Thomas Permutt, Ph.D.
Mathematical Statistician (Team Leader)

Michael Welch 6/15/00

Concur: Michael Welch, Ph.D.
Acting Deputy Director, Division of Biometrics II

archival: NDA 21-217

cc:

HFD-715/Nevius, Welch

HFD-170/Milstein, Rappaport, McCormick, Permutt

HFD-170/division file