CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

21-745

STATISTICAL REVIEW(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:	21-745
Name of drug:	Ryzolt (tramadol hydrochloride) extended-release tablets
Indication:	pain
Applicant:	Labopharm
Dates:	Received 2 July 2008; user fee goal (class 2 resubmission, 6 months) 2 January 2009
Review priority:	S
Biometrics division:	Division of Biometrics II
Statistical reviewer:	Thomas Permutt, Director
Medical division:	Anesthesia, Analgesia and Rheumatology Products
Project manager:	Kathleen Davies
Keywords:	NDA review, clinical studies, missing data, data imputation

1 INTRODUCTION

The subject NDA was submitted 25 November 2005, resubmitted 18 December 2006 and found to be Approvable 30 May 2007. The action letter noted:

> You have failed to demonstrate the efficacy of Ryzolt for your proposed indication of the management of moderate to moderately severe pain. Your conclusion of efficacy is dependent on the use of imputation strategies for missing data that do not adequately address the problem of good scores being assigned to subjects who dropped out because they were unable to tolerate the product. Provide substantial evidence of efficacy from at least one adequate and well-controlled clinical trial.

This resubmission is a reanalysis of study MDT3-005. The primary analysis of that study was described in the statistical review of the 2006 resubmission by Yongman Kim, Ph.D. as follows:

> The primary efficacy outcome variable of the study was Pain Intensity Numerical Rating Scale over 4 visits on Week 3, Week 6, Week 9, and Week 12. The score ranges from 0 (= no pain) to 10 (= worst possible pain) discretely. The score at Week 12 was prespecified as the primary endpoint to be used in statistical inference.

In 24 percent of cases in the active treatment group and 22 percent of cases in the placebo group, the primary datum was not observed because the patient discontinued treatment. The protocol specified that data were to be imputed by last observation carried forward (LOCF). In the case of dropout for lack of efficacy, a bad pain measurement would usually be captured and carried forward. In the case of dropout for toxicity, the last pain score might be good. This would result in the imputation of a good outcome to a patient who, by her own choice of dropping out, indicated that not taking the drug was preferable to taking it.

The protocol was the subject of oral and written discussions between the applicant and the Agency. Agency personnel clearly expressed concern about the imputation of favorable outcomes to patients who chose not to continue treatment. Nevertheless, we also agreed to consider the proposed analysis as primary provided it were supported by the outcomes of sensitivity analyses. This proviso somewhat vitiates the meaning of a *primary* analysis, for it makes the conclusion depend on the results of other analyses as well. Furthermore, the sensitivity analyses were not specified.

Dr. Kim conducted various analyses consistent with not imputing good scores for bad outcomes. He tried baseline observation carried forward (BOCF); since patients entered the study needing treatment, this technique reliably imputed bad scores to dropouts. He also graphed the proportion of responders at various levels of response (empirical cumulative distribution function) with dropouts considered to be nonresponders at all levels. Dr. Kim tested for a difference between the resulting curves with the van der Waerden test, a standard, appropriate (but not uniquely so) nonparametric test. Dr. Kim found no statistically significant differences between treatments in these sensitivity analyses, and he concluded that there was a lack of substantial evidence of effectiveness.

The applicant pursued several levels of appeal requesting that the Approvable action be overturned and the application approved. Curtis Rosebraugh, M.D. declined to do so. John Jenkins, M.D. also declined to do so, but suggested an alternative analysis. Douglas Throckmorton, M.D. confirmed Dr. Jenkins's refusal but also conveyed to the applicant that a favorable result in Dr. Jenkins's suggested analysis would be sufficient for approval.

2 THE "JENKINS METHOD"

As usual, the action letter expressed the reasons for nonapproval tersely. The applicant's appeals focused on the putatively pivotal role played by Dr. Kim's BOCF analysis. They argued that carrying forward the baseline score assigned an unrealistically bad outcome to patients who dropped out. Even without treatment, those patients might be expected to improve, for osteoarthritis is a variable condition. Patients needed to be in a relatively bad phase to require the treatment offered in the trial, and they might have been expected to regress toward a more average state over its course.

Dr. Jenkins suggested instead a procedure that would impute to dropouts scores comparable on average to those of the completers in the placebo group:

> Missing data at end of study will be imputed by scores drawn randomly from the placebo observations at end of study rather than by baseline scores. Specifically:
>
> 1. Stratify the placebo completers in tertiles with respect to outcome: upper third, middle third, lower third.
> 2. Stratify the combined active and placebo groups by tertile with respect to baseline score.
> 3. For each missing observation at week 12, substitute a random score from the placebo completers, drawn from the same tertile that the baseline score for that individual fell into.
> 4. Conduct the protocol-specified primary analysis on the now complete data set.

The applicant elaborated or modified this procedure in two ways. First, the definition of *tertile* was operationalized as follows. The lower tertile, for example, would be the lowest third of the scores in order. The scores in question, however, are discrete: integers from 0 to 10. Among the placebo completers, 25 percent of the scores were 0, 1 or 2, and a further 15 percent were 3. Either the equal scores of 3 must be assigned partly to the lower and partly to the middle tertile, or the lower tertile must be defined as either including them, thus comprising 40 percent rather than 33 percent of the scores, or excluding them, thus comprising 25 percent. The applicant reports analyses both dividing the boundary cases (a score of 3 in this instance) at random, and assigning them all to the lower or upper tertile rather than the middle. The results are similar, and I focus below on the latter approach, which seems to me slightly preferable.

Second, the applicant performed the imputation procedure multiple times and averaged the results, also accounting for the variation between repetitions in estimating standard errors. I view this as a desirable improvement on Dr. Jenkins's suggested analysis. The number of repetitions appears not to have been decided in advance, which causes a minor problem in interpreting the results.

3 RESULTS

The results of the applicant's analysis are summarized in the table below, copied from the submission (p. 7).

Multiple imputation results: No overlapping tertiles

	t-statistic				
Imputations	Mean	SE	Value	df	p-value
10	0.359	0.189	1.90	165	0.059
20	0.372	0.189	1.97	124	0.051
30	0.374	0.190	1.97	94	0.052
40	0.373	0.190	1.97	107	0.052
50	0.371	0.190	1.96	115	0.053
75	0.374	0.189	1.97	138	0.050
100	0.372	0.189	1.96	162	0.051
125	0.373	0.189	1.97	187	0.050
150	0.373	0.189	1.97	208	0.050
200	0.372	0.190	1.96	253	0.051
250	0.373	0.189	1.97	306	0.050
300	0.374	0.190	1.97	355	0.049
350	0.375	0.189	1.98	406	0.049

The seed used to select the imputed values for the i^{th} dataset was defined as: $36 + 10 \times i$.

There is a minor problem of multiplicity, as in trials with interim analyses. The figures, including the p-value, based on more data are more reliable than those based on fewer, so that more weight should be given to the last line of the table than the first. On the other hand, one might suppose that if the results had dipped under 0.05 sooner, the applicant might have stopped and reported a statistically significant result. If this is so, then the chance of getting a p-value below 0.05 sometime, and stopping, and so making a Type I error if the treatment actually had no effect, would be very slightly more than 0.05. Accordingly, the results should be taken as significant at a level very slightly *above* 0.05.

Although the usual Agency practice has been to consider 0.05 (two-sided) the dividing line between statistically significant and nonsignificant results, it has also been our practice not to

apply such a test too sharply. I do not think there is much to be gained by trying to establish post hoc precisely how much more or less than 0.05 the observed p-value really is. The signal conclusion is that it is 0.05, very nearly: an arguably statistically significant result, but of the weakest kind we would usually accept.

4 CONCLUSIONS AND RECOMMENDATIONS

The statistical method proposed by Dr. Jenkins addresses the Agency's concern about the attribution of good scores to patients with bad outcomes. Dr. Jenkins and Throckmorton suggested strongly that a positive result in this analysis would be sufficient for approval. The result can be taken as positive at the weakest level usually accepted.

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

/s/

Thomas Permutt
9/15/2008 03:35:08 PM
BIOMETRICS

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

COMPLETE RESPONSE/CLINICAL STUDIES

NDA:	21-745
Drug Name:	Ryzolt (formerly Tramadol Contramid® OAD)
	(tramadol hydrochloride extended-release tablets)
Indication(s):	For the management of moderate to moderately severe pain
Applicant:	Labopharm, Inc.
Date(s):	Submitted: December 18, 2006
	PDUFA Due Date: June 18, 2007
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Yongman Kim, Ph.D.
Concurring Reviewers:	Dionne Price, Ph.D.
	Thomas Permutt, Ph.D.
Medical Division:	Division of Anesthesia, Analgesia, and Rheumatology Products
Clinical Team:	Jin Chen, M.D.
Project Manager:	Paul, Balcer

Keywords: missing data, sensitivity analysis

1 Background

Labopharm originally submitted ———— on November 28, 2005. The applicant proposed the use of Ryzolt for the management of moderate to moderately severe pain. In my review of the NDA, I concluded that the two pivotal studies MDT3-003 and MDT3-005 failed to provide substantial evidence of efficacy of Ryzolt because efficacy shown by the applicant's analyses was not supported by my sensitivity analyses. In the two studies, the analyses using the last observation carried forward (LOCF) imputation strategy were successful in showing a statistically significant difference between Ryzolt and placebo. However, my sensitivity analyses employing a baseline observation carried forward (BOCF) strategy and a continuous responder analysis failed to show a difference.

b(4)

Upon completion of the review, the Agency issued an Approvable Letter. In the action letter dated September 28, 2006, the Agency requested that Labopharm conduct an additional clinical study to demonstrate the efficacy of Ryzolt (formerly Tramadol Contramid® OAD). The following excerpt is from the action letter:

> You have not provided substantial evidence that Ryzolt is effective for your proposed indication of the management of moderate to moderately severe pain. Your conclusion that efficacy has been demonstrated in studies MDT3-003 and MDT3-005 depends on the use of a last observation carried forward imputation methodology for patients who dropped out of the studies. We consider this method of imputing missing data inappropriate, and efficacy was not confirmed when other methods, such as baseline observation carried forward (BOCF) or continuous responder analysis (of the patient's status at the end of the study) were employed. Provide substantial evidence of efficacy from at least one adequate and well-controlled clinical trial. Ryzolt produced at your commercial manufacturing site should be used in future clinical trials.

In a post-action teleconference on October 20, 2006, the applicant maintained that the positive result from Study MDT3-005 based on the LOCF approach provided substantial evidence of efficacy. The applicant stated that the LOCF strategy was pre-specified as the primary method for handling missing data in the study protocol as agreed upon through a special protocol assessment and the statistical analysis plan. During the teleconference, the Agency clarified that the efficacy findings based on the LOCF analysis should have been supported by appropriate sensitivity analyses. The Agency further stated, "Although some sensitivity analyses performed by the applicant appeared to confirm the efficacy results, these analyses all shared the common flaw of attributing good scores to patients who were unable to tolerate Ryzolt and subsequently discontinued treatment."

In a subsequent End of Review meeting on November 27, 2006, the applicant proposed to conduct new sensitivity analyses using the placebo mean trajectory carried forward or placebo median trajectory carried forward approaches to impute dropout missing values. The Agency agreed to review the additional analyses once they were submitted. On December 18,

2006, the new analyses were submitted as a complete response to the action letter. After a preliminary review of the submission, the Agency additionally requested the following:

1. For the median placebo observation, use the median of all patients rather than of those with valid measures, counting those without valid data as bad scores. For example, if there are 99 patients in the placebo group and 20 drop out, use as the median, the 50th best of the 79 valid scores rather than the 40th.

2. For the mean placebo observation, use LOCF imputation within the placebo group.

In response to the request, the final analyses were submitted as an amendment to the complete response on February 9, 2007. The new approaches are the subject of the current review. The review has been formulated based on the submissions and discussions arising from numerous interactions outlined in Table 1.

Table1. Timeline of Post-Action Interactions

Date	Correspondence
28 November 2005	NDA 21-745 submitted by Labopharm, Inc.
28 September 2006	Approvable letter issued
20 October 2006	Teleconference requested by Labopharm to receive clarification from the division on clinical and statistical issues.
27 November 2006	End of Review meeting. The division agrees to consider alternative approaches to handle missing data based on discussions occurring during the teleconference.
19 December 2006	Complete Response was submitted.
9 February 2007	Response to FDA request for information was submitted.

2 Review

2.1 Study Design

MDT3-005 was 12-week, multi-center, double-blind study of the safety and efficacy of Ryzolt compared to placebo in patients with osteoarthritis of the knee. Patients were randomized to Ryzolt or placebo in a 2:1 ratio. Patients randomized to Ryzolt were titrated to Ryzolt 200 mg or 300 mg depending on the patient's tolerability of the treatment prior to the double-blind 12-week maintenance period. The primary objective of the study was to show efficacy of therapy with Ryzolt when compared to placebo.

The primary efficacy variable of the study was pain intensity at Week 12. Pain intensity was assessed via the Pain Intensity Numerical Rating Scale (PI-NRS) over four visits at Week 3, Week 6, Week 9, and Week 12. The score ranges from 0 (no pain) to 10 (worst possible pain) discretely.

2.2 Statistical Methodologies

In Study MDT3-005, pain intensity was compared at Week 12 between Ryzolt and placebo using an analysis of covariance (ANCOVA) model with a term for treatment and baseline value as covariate.

In this submission, the applicant conducted new analyses using the placebo mean trajectory carried forward and the placebo median trajectory carried forward imputation strategies. The applicant provided the following rationale for the methodology:

> **A sensitivity analysis for missing data seeks to derive plausible explanations for the treatment effect that would have been seen had subjects continued in the study. The LOCF and BOCF approaches are two possible explanations along a wide range. LOCF assumes that subjects who discontinue would experience pain similar to the pain they experienced when they were taking study medication, while BOCF assumes that subjects who discontinue would experience pain similar to what they experience in the absence of medication. BOCF, in this situation, dos not account for the regression to the mean that is caused by requiring subjects to have a PI-NRS score of 4 or higher to participate in the double-blind portion of the study.**
>
> **Along the philosophical lines of the BOCF method, but accounting for regression to the mean, are the Placebo Mean Trajectory Carried Forward ("Placebo Mean") and the Placebo Median Trajectory Carried Forward ("Placebo Median") approaches. Both estimate the trajectory of PI-NRS scores seen among subjects in the placebo group and use that trajectory along with the last PI-NRS value recorded at a planned study visit to project a value likely to have been observed at Visit 9 if the patients remained off study medication.**

The placebo mean trajectory (PGMEANT) method is implemented by first imputing the last observation for dropouts in the placebo group. The mean of the placebo group is then computed at each visit. Next, the differences between placebo mean values at each visit is calculated. Once a patient discontinues, the pain score at dropout is then incrementally decreased by subtracting the differences attained from the placebo mean values at each visit.

For example, the mean of the pain scores for the placebo group at Visit 4 is 7.16 and 5.37 at Visit 5. Thus, the difference between the two visits is -1.79. Consider a subject with a PI-NRS score of 9 at Visit 4 and missing data beyond. For this subject, the value at Visit 5 using the approach is the pain score at Visit 4 less the difference in the mean pain scores for the placebo group at Visit 4 and Visit 5, which is 9 – 1.79, or 7.21.

The placebo median trajectory (PGMEDT) method follows the same implementation scheme described above; however, instead of first imputing the last observation, the lowest ranks are assigned to dropouts in the placebo group.

In addition to the PGMEANT and PGMEDT methods, the applicant also performed two additional time-weighted average analyses. In the new analyses, missing values were imputed using either the LOCF or BOCF methods prior to the computation of the time-weighted average.

2.3 Results

The following tables and figures present results for Study MDT3-005 submitted from the original NDA submission, my sensitivity analyses from the original submission, and analyses from the current submission. An ANCOVA model with a term for treatment and baseline value as covariate was used in each analysis. The Appendix contains a table of the patient disposition and a graph of the dropout rate over the visits.

Results from the original NDA submission:

As shown in Table 2, the statistically significant difference between Ryzolt and placebo was shown in the analysis using a LOCF imputation strategy. However, a treatment difference between Ryzolt and placebo was not shown in the analysis using a BOCF approach in Table 3.

Table 2. Applicant's Analysis of Pain Intensity Score: LOCF

Pain Intensity Score at Endpoint (Week 12)		
	Ryzolt (n=431)	**Placebo (n=214)**
Baseline Mean (SD)	7.2 (1.6)	7.2 (1.6)
Endpoint LSMean (SE)	4.3 (0.1)	4.8 (0.2)
Difference vs PBO (95% CI)	-0.5 (-0.9, -0.1)	
p-value vs. placebo	**0.0157**	

LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.

Table 3. Reviewer's Analysis of Pain Intensity Score: BOCF

Pain Intensity Score at Endpoint (Week 12)		
	Ryzolt (n=431)	**Placebo (n=214)**
Baseline Mean (SD)	7.2 (1.6)	7.2 (1.6)
Endpoint LSMean (SE)	4.8 (.1)	5.0 (.2)
Difference vs PBO (95% CI)	-0.2 (-0.7, 0.1)	
p-value vs. placebo	0.2134	

LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.

Table 4 presents the time-weighted average (TWA) sensitivity analysis conducted by the sponsor. The results showed a statistically significant difference. The analysis used observed data only and included patients having at least two post-baseline assessments.

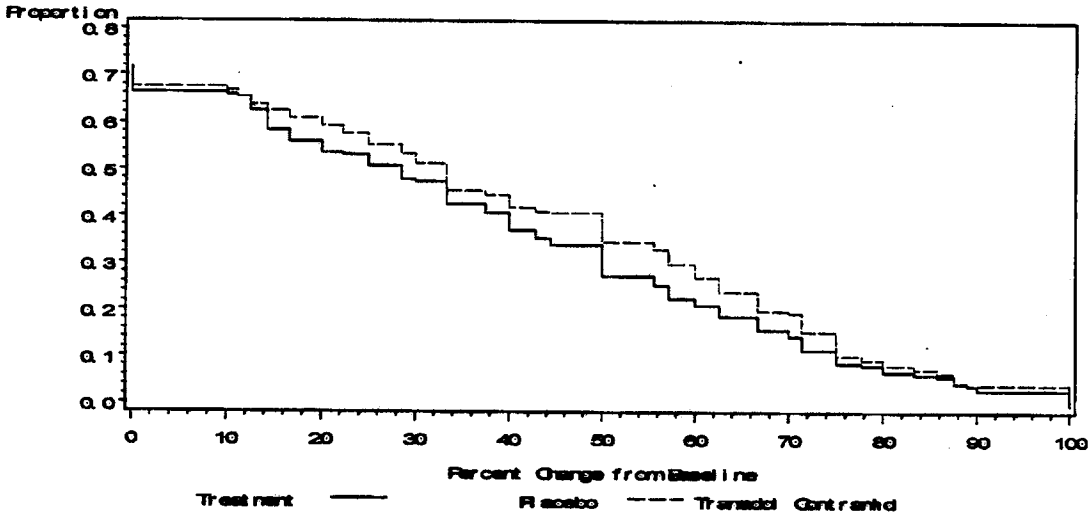
Table 4. Applicant's Analysis of Pain Intensity Score: TWA

Time-Weighted Average of Pain Intensity Scores		
	Ryzolt (n=431)	**Placebo (n=214)**
Baseline Mean (SD)	7.2 (1.6)	7.2 (1.6)
TWA Mean (SD)	4.2 (2.0)	4.9 (1.9)
Difference vs PBO (95% CI)	-0.7 (-1.0, -0.4)	
p-value vs. placebo	**<0.0001**	

p-value calculated from ANCOVA model: Y = trt + baseline.

During the original review, I additionally conducted a continuous responder analysis. In the continuous responder analysis, the proportion of responders was calculated using multiple definitions (or cut-offs) of treatment response ranging from 0% to 100% improvement. All discontinuations or drop-outs were classified as non-responders in the analysis. The analysis was graphically depicted via a plot of the proportion of responders against the multiple cut-offs (i.e. cumulative distribution function) shown in Figure 1. In order to statistically compare responder curves from each treatment, the van der Waerden normal score test was used on the change from baseline in pain intensity score.

Figure 1. Reviewer's Continuous Responder Analysis of Pain Intensity Score



Absolute Change from Baseline to Endpoint (Week 12)		
	Ryzolt (n=431)	Placebo (n=226)
p-value vs. placebo*	0.3466	

*p-values calculated from van der Waerden non-parametric test.

Results from the current submission:

Tables 5 and 6 present results from the current submission. Analyses using the placebo mean and median trajectory methods showed a statistically significant difference between Ryzolt and placebo.

Table 5. Applicant's Analysis of Pain Intensity Score: Placebo Mean Trajectory Carried Forward

Pain Intensity Score at Endpoint (Week 12)		
	Ryzolt (n=431)	**Placebo (n=214)**
Baseline Mean (SD)	7.2 (1.6)	7.2 (1.6)
Endpoint LSMean (SE)	4.13 (0.11)	4.57 (0.15)
Difference vs PBO (95% CI)	-0.45 (-0.81, -0.09)	
p-value vs. placebo	**0.0155**	

LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.

Table 6. Applicant's Analysis of Pain Intensity Score: Placebo Median Trajectory Carried Forward

Pain Intensity Score at Endpoint (Week 12)		
	Ryzolt (n=431)	**Placebo (n=214)**
Baseline Mean (SD)	7.2 (1.6)	7.2 (1.6)
Endpoint LSMean (SE)	4.22 (0.11)	4.66 (0.15)
Difference vs PBO (95% CI)	-0.44 (-0.81, -0.08)	
p-value vs. placebo	**0.0172**	

LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.

In this submission, the sponsor also conducted time-weighted average analyses using LOCF and BOCF imputation strategies as shown in Tables 7 and 8. Both analyses showed a

statistically significant difference in pain intensity over the 12 weeks between Ryzolt and placebo.

Table 7. Applicant's Analysis of Pain Intensity Score: TWA with LOCF

Time-Weighted Average of Pain Intensity Scores		
	Ryzolt (n=431)	**Placebo (n=214)**
Baseline Mean (SD)	7.2 (1.6)	7.2 (1.6)
TWA Mean (SD)	4.6 (2.0)	5.2 (1.9)
Difference vs PBO	-0.6	
p-value vs. placebo	**<0.0001**	

p-value calculated from ANCOVA model: Y = trt + baseline.

Table 8. Applicant's Analysis of Pain Intensity Score: TWA with BOCF

Time-Weighted Average of Pain Intensity Scores		
	Ryzolt (n=431)	**Placebo (n=214)**
Baseline Mean (SD)	7.2 (1.6)	7.2 (1.6)
TWA Mean (SD)	4.8 (2.1)	5.3 (1.9)
Difference vs PBO	-0.5	
p-value vs. placebo	**0.0015**	

p-value calculated from ANCOVA model: Y = trt + baseline.

Reviewer's new analyses related to the current submission:

At the request of the clinical review team, I conducted an additional sensitivity analysis using the mean of the placebo group (MPG) to impute missing data due to adverse events. To implement the MPG approach, the LOCF method was initially used to impute missing values in the placebo group, and then the mean pain intensity at week 12 for the placebo group was calculated. The mean of the placebo group was subsequently imputed for missing values due to adverse events. LOCF was used to impute dropouts for all other reasons. Table 9 shows a statistically significant difference between Ryzolt and placebo. However, the MPG method is concerning from a statistical perspective. Patients withdrawing should be assigned bad scores indicating an unfavorable outcome. However using the MPG method, patients withdrawing are assigned average scores. Inclusion of these average scores along with good scores (from non-dropouts) may systematically improve the mean outcome of the active treatment.

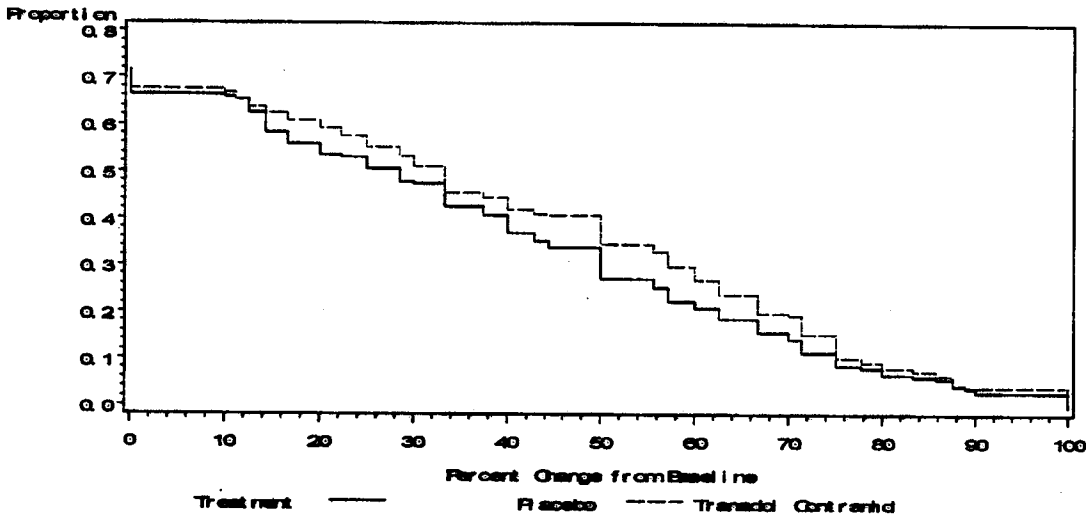
Table 9. Mean Placebo Group/LOCF Analysis of Pain Intensity Score

Pain Intensity Score at Endpoint (Week 12)		
	Ryzolt (n=431)	**Placebo (n=214)**
Baseline Mean (SD)	7.2 (1.6)	7.2 (1.6)
Endpoint LSMean (SE)	4.39 (0.11)	4.77 (0.16)
Difference vs PBO (95% CI)	-0.38 (-0.75, -0.001)	
p-value vs. placebo	**0.0490**	

LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.

Because the continuous responder analysis treats all the dropout subjects as non-responders, it is conservative like the BOCF analysis in which '0' (or 0 percent improvement) is assigned for missing values due to dropouts. In order to make the continuous responder analysis less conservative, I performed an additional analysis treating only dropouts due to adverse events or lack of efficacy as non-responders. As shown in Figure 2, this analysis does not show a statistically significant difference.

Figure 2. Reviewer's Continuous Responder Analysis of Pain Intensity Score (treating only adverse events or lack of efficacy dropouts as non-responders)



Absolute Change from Baseline to Endpoint (Week 12)		
	Ryzolt (n=431)	Placebo (n=226)
p-value vs. placebo*	0.3416	

*p-values calculated from van der Waerden non-parametric test.

Summary on all the analyses from the original and current submissions:

Table 10 summarizes the analyses submitted from the original NDA submission (LOCF, TWA analyses), my sensitivity analyses during the original review of the NDA (BOCF, Continuous Responder analyses), analyses from this submission (Completers, PGMEANT, PGMEDT analyses), and my analyses related to this submission (MPG, modified Continuous Responder analyses).

Table 10. Summary of All Analyses of Pain Intensity Score

	LSMean Difference (Ryzolt - PBO)	P-value*
Applicant's analyses from the original NDA		
LOCF	-0.479	**0.016**
TWA	-0.7#	**<0.001**
Reviewer's analyses from the original NDA		
BOCF	-0.251	0.220
Continuous Responder Analysis		0.347
Applicant's analyses from this submission		
TWA/LOCF	-0.6#	**<0.001**
TWA/BOCF	-0.5#	**0.002**
Completers	-0.453	0.053
PGMEANT	-0.480	**0.015**
PGMEDT	-0.478	**0.016**
Reviewer's analyses from this submission		
LOCF/BOCF	-0.272	0.177
MPG/LOCF	-0.377	**0.049**
Continuous Responder Analysis **		0.342

*Bold p-values denote a statistically significant difference at the level of 0.05.
**It treats as non-responders only subjects dropping out due to adverse events or lack of efficacy.
#Mean difference

3 Conclusions and Recommendations

The applicant has conducted several sensitivity analyses to demonstrate the efficacy of Ryzolt. Although many approaches may be used, the division has specific interest in strategies that do not assign a treatment benefit to patients who cannot tolerate the treatment for 12 weeks. In addition, the division's current approach to pain trials favors primary analyses that assess the treatment at the end of the study since the 12-week endpoint can be viewed as a surrogate for chronic treatment.

The placebo mean trajectory carried forward and the placebo median trajectory carried forward methods do not address the division's concern regarding missing data since the methods may result in favorable outcomes for dropouts due to adverse events. Furthermore, the methods give more benefit to early dropouts and could assign even better scores than the last observation carried forward method.

Although the time-weighted analysis may provide supportive information, the analysis averages results across the duration of the trial and may give treatment benefit to those patients who were unable to tolerate the treatment. Thus, the time-weighted analyses do not address the division's concerns.

Based on the collective evidence using several analysis strategies, the data from study MDT3-005 along with the data from study MDT3-003 do not provide substantial evidence of efficacy of Ryzolt for the indication of moderate to moderately severe pain.

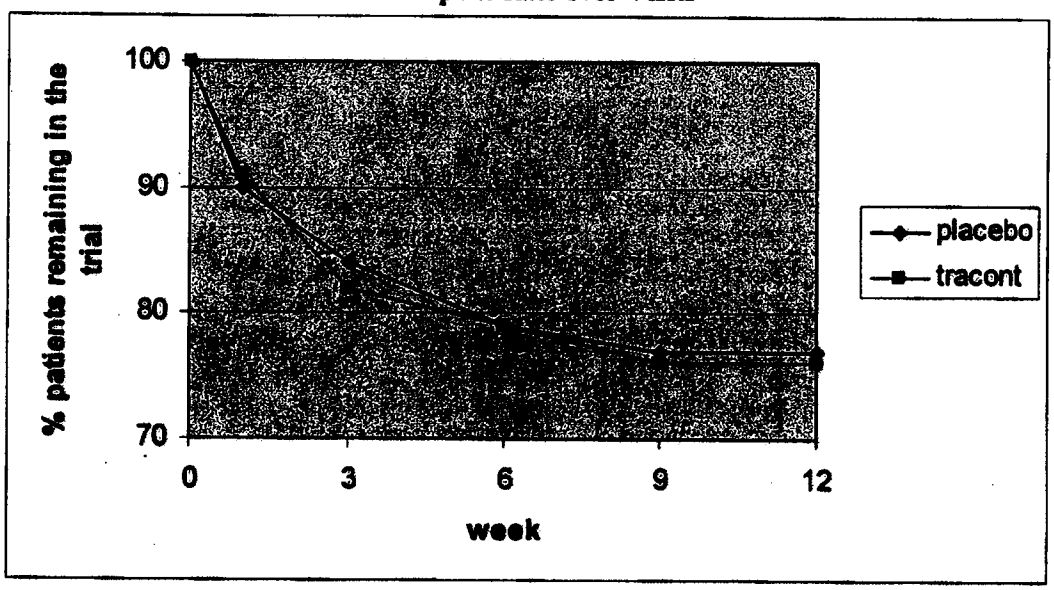
4 Appendix

Patient Disposition: Study MDT3-005

	Ryzolt	Placebo	Total
Open-label enrolled			1028
Double-blind - All Randomized	432	214	646
Full Analysis Population	431	214	645
Per Protocol	378	192	570
Completers	326 (75%)	165 (77%)	491 (76%)
Dropouts	106 (25%)	49 (23%)	155 (24%)
Adverse Events	44 (10%)	11 (5%)	55 (9%)
Lack of Efficacy	34 (8%)	22 (10%)	56 (9%)
Other	28	16	44

Full Analysis Population was defined as all randomized patients who received at least one dose of the randomized study medication regardless of the status of the post-dosing assessment..

Dropout Rate over Visits



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

/s/

Yongman Kim
4/24/2007 02:05:09 PM
BIOMETRICS

Dionne Price
4/24/2007 03:29:26 PM
BIOMETRICS

Thomas Permutt
4/24/2007 04:43:29 PM
BIOMETRICS
concur

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:	NDA 21-745
Drug Name:	Tramadol Contramid® OAD
Indication(s):	For the management of moderate to moderately severe pain
Applicant:	Labopharm, Inc.
Date(s):	Submitted: November 28, 2005
	PDUFA: September 28, 2006
Review Priority:	Standard
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Yongman Kim, Ph.D.
Concurring Reviewers:	Thomas Permutt, Ph.D.
Medical Division:	Division of Anesthesia, Analgesia, and Rheumatology Products
Clinical Team:	Jin Chen, M.D.
Project Manager:	Paul Balcer

Keywords: NDA review, clinical studies, endpoint analysis/LOCF, missing data, sensitivity analyses

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

The study MDT3-002 with knee osteoarthritis (OA) patients failed to show a statistically significant difference on WOMAC Pain Subscale Score, percent change from baseline to Week 12, as primary outcome variable at all dose levels (100 mg, 200 mg, and 300 mg) of tramadol Contramid once a day (OAD) compared to placebo in the last observation carried forward (LOCF) analysis on the full analysis set (FAS).

The study MDT3-003 with knee OA patients showed a statistically significant difference on WOMAC Pain Subscale Score, percent change from baseline to Week 12, at dose of 300 mg of tramadol Contramid OAD compared to placebo in the LOCF analysis on the FAS.

The study MDT3-005 with knee OA patients showed a statistically significant difference on Pain Intensity Numerical Rating Scale at Week 12 at dose of 200 mg or 300 mg of tramadol Contramid OAD compared to placebo in the LOCF analysis on the FAS. However, the statistically significant difference shown in the studies MDT3-003 and MDT3-005 was sensitive to dropout handling methods and to choice of analysis set. The statistically significant difference was not shown in the baseline observation carried forward (BOCF) analysis, in the continuous responder analysis, or in the ITT LOCF analysis in both studies. (See Tables 3 – 15 and Figures 2 – 4.)

Overall, although the evidence of efficacy was replicated in two well controlled studies MDT3-003 and MDT3-005, the submitted data of studies with high dropout rates failed to provide substantial evidence supporting pain indication of tramadol Contramid once a day formulation because the efficacy shown might be driven by imputation of missing data due to dropouts.

1.2 Brief Overview of Clinical Studies

The sponsor submitted the results and data from four efficacy studies of tramadol Contramid OAD in patients with OA of the knee. Three studies MDT3-002, MDT3-003, and MDT3-005 were superiority trials and the study MDT3-001/E1 was a non-inferiority trial. The three studies MDT3-001/E1, MDT3-002, and MDT3-003 were included in the original submission and the study MDT3-005 was submitted after midway of the review cycle. I reviewed only the three superiority studies – MDT3-002, MDT3-003, and MDT3-005. These were 12-Week, double-blind, placebo-controlled, multi-center studies to investigate the safety and analgesic effect of tramadol Contramid once a day (OAD) in patients with OA of the knee.

In study MDT3-002, five-hundred sixty-five patients were randomized to tramadol Contramid OAD 100 mg (n = 110), tramadol Contramid OAD 200 mg (n = 113), tramadol Contramid OAD 300 mg (n = 115), and placebo (n = 227) in 1:1:1:2 ratio.

In study MDT3-003, five-hundred fifty-two patients were randomized to tramadol Contramid OAD 100 mg (n = 106), tramadol Contramid OAD 200 mg (n = 111), tramadol Contramid OAD 300 mg (n = 108), and placebo (n = 227) in 1:1:1:2 ratio.

The primary objective of the two studies was to show efficacy of therapy with tramadol Contramid OAD 100 mg, tramadol Contramid OAD 200 mg, or tramadol Contramid OAD 300 mg when compared to placebo.
The primary efficacy outcome variable of the two studies was WOMAC (Western Ontario and McMaster Universities) Pain Subscale Score over 4 visits on Week 0, Week 3, Week 6, and Week 12. The subscale is the sum of five items with 0 – 100 mm VAS score. Therefore, the score ranges from 0 to 500 continuously. The percent change of the score from baseline to Week 12 was prespecified as the primary endpoint to be used in statistical inference.
The secondary efficacy variables were WOMAC Physical Function Subscale Score, Patient Global Rating of Pain Relief.

In study MDT3-005, six-hundred forty-six patients were randomized to tramadol Contramid OAD 200 mg or 300 mg (n = 432) and placebo (n = 214) in 2:1 ratio.

The primary objective of the study was to show efficacy of therapy with tramadol Contramid OAD when compared to placebo.
The primary efficacy outcome variable of the study was Pain Intensity Numerical Rating Scale over 4 visits on Week 3, Week 6, Week 9, and Week 12. The score ranges from 0 (= no pain) to 10 (= worst possible pain) discretely. The score at Week 12 was prespecified as the primary endpoint to be used in statistical inference.
The secondary efficacy variables were WOMAC Pain Subscale Score, WOMAC Physical Function Subscale Score, Patient Global Rating of Pain Relief, Physician's Global Impression of Changes.

1.3 Statistical Issues and Findings

For the efficacy analysis, the sponsor based their inferences on full analysis set (FAS) data with LOCF for a statistically significant difference on WOMAC Pain Subscale in studies MDT3-002 and MDT3-003 and on Pain Intensity Numerical Rating Scale in study MDT3-005 comparing tramadol Contramid OAD with placebo. Sponsor's FAS population was defined as all randomized patients who received at least one dose of study medication and at least one post baseline pain assessment. However, since ITT population defined as all randomized patients is more appropriate as the primary analysis set, I conducted the same analysis as the sponsor based on ITT population as a sensitivity analysis.

The sponsor proposed LOCF in order to impute 'missing data' due to dropouts in their primary analysis. However, FDA requested a sensitivity analysis using alternative imputation methods. I construe the agreement between the sponsor and FDA that sensitivity analyses should support a successful LOCF analysis for a study to be a success. I think that non-existence of the outcome variable at target endpoint does not necessarily mean 'missing data' and a method assigning bad scores to 'missing' outcomes due to dropout is desirable. In this sense, LOCF is not a good way of handling dropouts because it could assign good scores and BOCF is a more acceptable method because it usually assigns bad scores.

The sponsor provided post hoc sensitivity analyses assessing their conclusion with respect to imputation methods for missing data due to dropout. The sponsor conducted BOCF analysis, Time-Weighted Average analysis, and Repeated Measures ANOVA. I also conducted a continuous responder analysis, in which a continuous responder curve for each group is generated by changing responder criterion from 0% to 100% improvement from baseline to Week 12 and the curves are compared. Along with the continuous responder analysis, I used van der Waerden normal score test to compare tramadol groups and placebo. Since the test is also post hoc, it is subject to multiplicity. The sponsor also conducted a responder analysis with 30% response criterion. However, the analysis was problematic because they used LOCF for dropouts before determining the response.

The sponsor prespecified ANCOVA model with terms for treatment, center, and baseline pain score as a covariate as the primary analysis. However, they dropped the center term from the model with the reason that many centers had too few subjects. I conducted an analysis including the center term in the model as a sensitivity analysis.

The sponsor proposed Holm's sequentially rejective method to adjust for multiple comparisons in the studies MDT3-002 and MDT3-003, which is acceptable.

Based on my review of study results, I reached the following conclusions:

Data from study MDT3-002 failed to show the superiority of any dose of tramadol Contramid OAD to placebo in terms of WOMAC Pain Subscale Score, percent change from baseline to Week 12, in patients with OA of the knee.

Data from study MDT3-003 showed the superiority of tramadol Contramid OAD 300 mg to placebo in WOMAC Pain Subscale Score, percent change from baseline to Week 12, in patients with OA of the knee. However, the statistically significant difference was not supported by sensitivity analyses - BOCF and continuous responder analysis - with respect to methods of handling dropouts, implying that the statistically significant difference with LOCF could be driven by imputation considering high dropout rates. Unlike the sponsor's analysis on FAS with LOCF, my ITT with LOCF analysis with the same model as the sponsor did not show statistically significant difference. Also, unlike the sponsor's ANCOVA without the center term, my ANCOVA with the center term on FAS LOCF did not show statistically significant difference.

Data from study MDT3-005 showed the superiority of tramadol Contramid OAD to placebo in Pain Intensity Numerical Rating Scale at 12-week landmark in patients with OA of the knee. However, the statistically significant difference was not supported by BOCF and continuous responder analysis, implying that the statistically significant difference with LOCF could be driven by imputation considering high dropout rates.

2. INTRODUCTION

2.1 Overview

2.1.1 Drug class and regulatory history

The following are quotes from the submission regarding drug class.

> Tramadol HCl is a synthetic, centrally acting analgesic that has been shown to be effective in a variety of acute and chronic pain states. In particular, tramadol has been demonstrated to reduce pain attributed to OA.
>
> Tramadol HCl was developed by the Grunenthal Company of Germany. It has been marketed in Germany as Tramal™ since 1977, and in the United States as Ultram since 1998. Over ———— doses of tramadol HCl have been administered since its introduction in Germany, where it is the largest selling prescription analgesic. Tramadol HCl is also marketed in several other European countries, as well as Japan. b(4)
>
> After oral administration, tramadol HCl is rapidly and almost completely absorbed, and it is extensively metabolized. The major metabolic pathways appear to be N- and O-demethylation and glucuronidation or sulfonation in the liver. Only one metabolite, mono-O-desmethyltramadol (M1), is pharmacologically active, which has an approximate 200-fold higher affinity for the μ–opioid receptor than racemic tramadol. In healthy humans, tramadol is demethylated by the polymorphic enzyme cytochrome P450 2D6 (CYP2D6) to the M1 metabolite.
>
> The mechanism of action of tramadol HCl is not completely understood. Animal models indicate that the drug (and its active M1 metabolite) acts as opiate agonists, apparently by selective activity at the μ–opioid receptor. In addition to opiate agonist activity, tramadol HCl inhibits reuptake of certain monoamines (norepinerphrine, serotonin) which appears to contribute to the drug's analgesic effect.
>
> Cotramid® is a technology recently developed by Labopharm for the formulation and manufacturing of solid, sustained-release oral dosage forms. Contramid™ is obtained by a proprietary cross-linking of high-amylose starch to produce a three-dimensional structure, which is then combined with an active ingredient. Once in the stomach, gastric juices turn the tablet surface to a gel, and the active drug diffuses at an even rate, Because of its simplicity of formulation and flexibility, Contramid can be adapted to fit the specific requirements of oral drugs to produce tailored or difficult-to-replicate-by-other-means pharmacokinetic profiles. Labopharm has developed a once a day formulation of Tramadol HCl using its Contramid® drug delivery technology.

The following are quotes relevant to my review from the submission regarding regulatory history and interactions between the sponsor and FDA prior to NDA, which were confirmed from DFS by me.

3. INDICATION/PAIN MODEL

September 2002:

The Division stated that the indication would be [redacted] b(4)
and not [redacted]

The Division also indicated that replicated positive study results would be required for support [redacted]

The Division stated that another possible indication [redacted] could be obtained using three models: OA, low back pain and fibromyalgia (two replicate studies in each). b(4)

The Division noted that sponsors could no longer attain a global chronic pain indication.

April 2004:

The Division agreed that Labopharm could file for the indication moderate to moderately severe chronic pain, as was previously agreed in November 2001 and in May 2002.

May 2005:

The Division advised Labopharm that clinical data (at least one adequate and well-controlled clinical study) would be required to support [redacted] b(4)

4. STATISTICS

February 2004:

The Division did not agree that LOCF should be the only method of imputation and requested a sensitivity analysis using alternative imputation methods be performed.

It was agreed that no adjustment would be made for center effect (although it was specified in protocols MDT3-003 and MDT3-002) due to the very high number of centers and the very small number of patients per centers. In fact, more than a third of the centers had enrolled less than 5 patients. However, some analysis adjusting for center effect would be presented as exploratory.

2.1.2 Proposed Indication for Tramadol Contramid® OAD

Tramadol Contramid® OAD is indicated for the management of moderate to moderately severe pain.

2.2 Data Sources

The original paper submission on November 28, 2005 can be found on the FDA, CDER document room.
The electronic SAS data submission on November 25, 2005, on May 2, 2006, and May 25, 2006 can be found on the FDA, CDER electronic document room (EDR).

Data sets:

\\Cdsesub1\n21745\N_000\2005-11-25\datasets
\\Cdsesub1\n21745\N_000\2006-05-02\Datasets
\\Cdsesub1\n21745\N_000\2006-05-25\datasets

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Study Design and Endpoints

The studies MDT3-002 and MDT3-003 were of identical design with 12-week, multi-center, double-blind study of the safety and efficacy of tramadol Contramid OAD 100 mg, tramadol Contramid OAD 200 mg, tramadol Contramid OAD 300 mg compared to placebo in patients with OA of the knee. Patients were randomized to tramadol Contramid OAD 100 mg, tramadol Contramid OAD 200 mg, tramadol Contramid OAD 300 mg, or placebo in 1:1:1:2 ratio. Patients randomized to tramadol were titrated to the randomized dose before the double blind 12 week maintenance period.

The study MDT3-005 was 12-week, multi-center, double-blind study of the safety and efficacy of tramadol Contramid OAD compared to placebo in patients with OA of the knee. Patients were randomized to tramadol Contramid OAD or placebo in 2:1 ratio. Patients randomized to tramadol were titrated to tramadol Contramid OAD 200 mg or 300 mg depending on tolerability of the patients before the double blind 12 week maintenance period.

Figure 1 in Appendix shows schematic of study design for studies MDT3-002, MDT3-003, and MDT3-005.

Seventy-five investigators enrolled subjects from US sites and participated in the clinical study MDT3-002.

Seventy-four investigators enrolled subjects from US sites and participated in the clinical study MDT3-003.

One hundred-eight investigators enrolled subjects from US (67), Canada (14), Romania (9), and France (18) sites and participated in the clinical study MDT3-005.

The primary efficacy endpoint for studies MDT3-002 and MDT3-003 was WOMAC Pain Subscale Score, percent change from baseline to Week 12.

The primary efficacy endpoint for study MDT3-005 was Pain Intensity Numerical Rating Scale at Week 12.

In studies MDT3-002 and MDT3-003, the percent change from baseline in WOMAC Pain Subscale Score was compared at Week 12 between tramadol Contramid OAD doses of 100 mg, 200 mg, and 300 mg and placebo using ANCOVA model with terms for treatment and baseline value as covariate. Holm's sequentially rejective procedure was employed to adjust for the multiple comparisons between each of three doses of tramadol Contramid OAD and placebo.

In study MDT3-005, Pain Intensity Numerical Rating Scale was compared at Week 12 between tramadol Contramid OAD and placebo using ANCOVA model with terms for treatment and baseline value as covariate.

3.1.2 Patient Disposition and Demographics

As shown in Table 1 in Appendix, about 43%, 44%, and 24% of the patients discontinued from studies MDT3-002, MDT3-003, and MDT3-005, respectively. However, there was some degree of imbalance in dropout rates among treatment groups in the study MDT3-002 and MDT3-003. In the study MDT3-002, the placebo group had lower dropout rates (37%) compared to other groups (42-53%). In the study MDT3-003, the placebo group had lower dropout rate (41%) compared to tramadol contramid OAD 300 mg group (54%).

Table 2 in Appendix shows patient demographics by treatment groups for the studies MDT3-002, MDT3-003, and MDT3-005, respectively. There were no noticeable imbalances among treatment groups with respect to demographic variables of age, race, and BMI (body mass index).

The table also shows baseline values for the primary efficacy variables by treatment groups for the studies MDT3-002, MDT3-003, and MDT3-005, respectively. Mean baseline values for the primary efficacy variables were comparable among treatment groups.

3.1.3 Statistical Methodologies

In studies MDT3-002 and MDT3-003, the percent change from baseline in WOMAC Pain Subscale Score was compared at Week 12 between three doses – 100 mg, 200 mg, and 300 mg – of tramadol Contramid OAD and placebo using ANCOVA model with terms for treatment and baseline value as covariate. Holm's sequentially rejective procedure was employed to adjust for the multiple comparisons between each of three doses of tramadol Contramid OAD and placebo.

In study MDT3-005, Pain Intensity Numerical Rating Scale was compared at Week 12 between tramadol Contramid OAD and placebo using ANCOVA model with terms for treatment and baseline value as covariate.

3.1.4 Results and Conclusions

Tables 3 – 15 and Figures 2 – 4 present the statistical analyses done by the sponsor and me. Following are review results of the analyses.

Study MDT3-002:

Data from the study failed to show the superiority of tramadol contramid OAD 100 mg, tramadol Contramid OAD 200 mg, or tramadol Contramid OAD 300 mg to placebo in reduction of primary endpoint, WOMAC Pain Subscale Score in patients with OA of the knee. In their analysis, the sponsor used ANCOVA model with terms for treatment and baseline pain as covariate on FAS with LOCF (36% - 41% for tramadol doses vs. 38% for placebo; $p > 0.0167$ = Holm's significance level). My analysis on ITT population, analysis with center term in the ANCOVA model, and BOCF analysis showed no statistically significant difference. My continuous responder curves showed separation between tramadol dose groups and placebo to some degree with the response curve for placebo being mostly above those of tramadol groups. However, van der Waerden test did not show a statistically significant separation. (See Tables 3 – 6 and Figure 2 below.)

Table 1 Sponsor Analysis of WOMAC Pain (with site term not included in the model): Study MDT3-002 FAS with LOCF

Percent Change from Baseline to Endpoint (Week 12)				
	TRA100 mg (n=109)	TRA200 mg (n=110)	TRA300 mg (n=113)	PBO (n=226)
LSMean Percent Change (SE)	36% (4%)	37% (4%)	41% (4%)	38% (3%)
Diff. from PBO (95% CI)	-2% (-12%, 8%)	-1% (-11%, 9%)	3% (-7%, 13%)	
p-value vs. placebo*	(0.7173)	(0.7710)	0.5631	
Holm's adjusted level	0.025 (=0.05/2)	0.05	0.0167 (=0.05/3)	

LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.
*p-values are compared with adjusted significance levels for multiplicity based on Holm's sequentially rejective method. The sequential testing procedure stops prior to calculating p-values in the parenthesis.

Table 2 Reviewer Analysis of WOMAC Pain (with site term not included in the model): Study MDT3-002 FAS with BOCF

Percent Change from Baseline to Endpoint (Week 12)				
	TRA100 mg (n=109)	TRA200 mg (n=110)	TRA300 mg (n=113)	PBO (n=226)
LSMean Percent Change (SE)	25% (4%)	23% (4%)	26% (4%)	31% (3%)
Diff. from PBO (95% CI)	-5% (-14%, 4%)	-8% (-17%, 1%)	-4% (-13%, 5%)	
p-value vs. placebo*	(0.2479)	0.0807	(0.3408)	
Holm's adjusted level	0.025 (=0.05/2)	0.0167 (=0.05/3)	0.05	

LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.
*p-values are compared with adjusted significance levels for multiplicity based on Holm's sequentially rejective method. The sequential testing procedure stops prior to calculating p-values in the parenthesis.

Table 3 Reviewer Analysis of WOMAC Pain (with site term not included in the model): Study MDT3-002 FAS with LOCF

Absolute Change from Baseline to Endpoint (Week 12)				
	TRA100 mg (n=109)	TRA200 mg (n=110)	TRA300 mg (n=113)	PBO (n=226)
LSMean Change (SE)	110 (12)	115 (12)	128 (12)	113 (9)
Diff. from PBO (95% CI)	-3 (-33, 26)	2 (-28.4, 30.8)	15 (-15, 43)	
p-value vs. placebo*	(0.7980)	(0.9363)	0.3336	
Holm's adjusted level	0.025 (=0.05/2)	0.05	0.0167 (=0.05/3)	

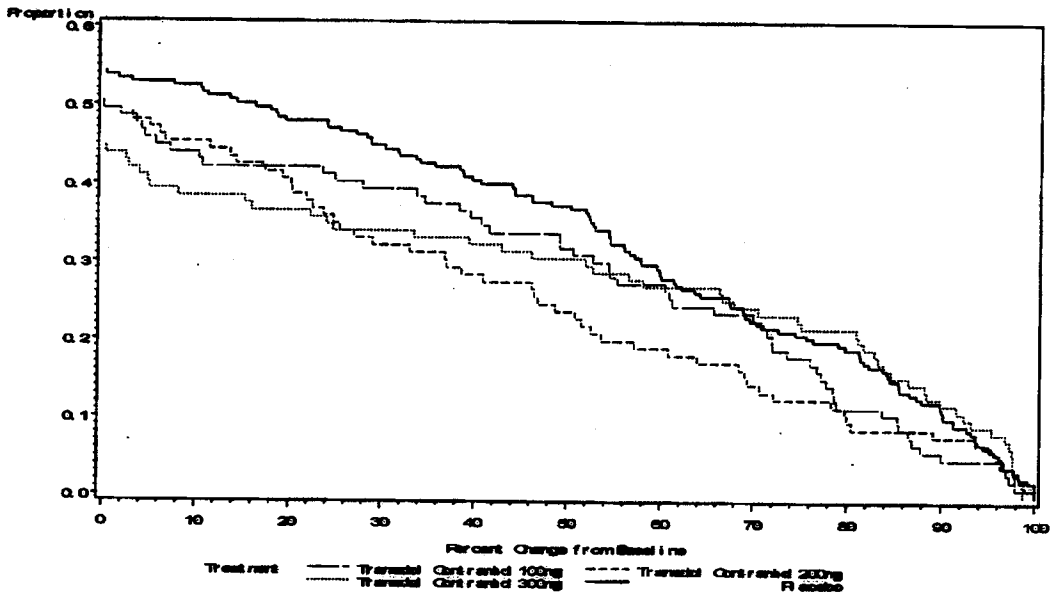
LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.
*p-values are compared with adjusted significance levels for multiplicity based on Holm's sequentially rejective method. The sequential testing procedure stops prior to calculating p-values in the parenthesis.

Table 4 Reviewer Analysis of WOMAC Pain (with site term not included in the model): Study MDT3-002 ITT with LOCF

Percent Change from Baseline to Endpoint (Week 12)				
	TRA100 mg (n=106)	TRA200 mg (n=111)	TRA300 mg (n=108)	PBO (n=227)
LSMean Percent Change (SE)	35% (4%)	34% (4%)	40% (4%)	37% (3%)
Diff. from PBO (95% CI)	-2% (-12%, 8%)	-3% (-13%, 7%)	3% (-7%, 12%)	
p-value vs. placebo*	(0.6648)	0.5365	(0.6104)	
Holm's adjusted level	0.05	0.0167 (=0.05/3)	0.025 (=0.05/2)	

LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.
*p-values are compared with adjusted significance levels for multiplicity based on Holm's sequentially rejective method. The sequential testing procedure stops prior to calculating p-values in the parenthesis.

Figure 1 Continuous Responder Analysis: Study MDT3-002 FAS



Absolute Change from Baseline to Endpoint (Week 12)				
	TRA100 mg (n=109)	**TRA200 mg (n=110)**	**TRA300 mg (n=113)**	**PBO (n=226)**
p-value vs. placebo*	0.4107	0.1359	0.1160	

*p-values calculated from van der Waerden Nonparametric test.

Study MDT3-003:

Data from the study showed the statistical significance of tramadol contramid OAD 300 mg when compared to placebo in reduction of primary endpoint, WOMAC Pain Subscale Score in patients with OA of the knee. The statistically significant difference was shown by the sponsor on FAS with LOCF analysis using ANCOVA model with terms for treatment and baseline pain as covariate (46% for tramadol vs. 32% for placebo; p = 0.0165 < 0.0167 = Holm's significance level).

However, my analysis on ITT population did not show statistical significant difference between tramadol 300 mg dose group and placebo (44% for tramadol vs. 32% for placebo; p = 0.0253 > 0.0167 = Holm's significance level). My analysis with center term in the ANCOVA model did not show statistical significant difference between tramadol 300 mg dose group and placebo (44% for tramadol vs. 31% for placebo; p =0 .0227 > 0.0167 = Holm's significance level). Both ANCOVA with center term and ANCOVA without center term used type-3 sum of squares when comparing tramadol dose groups to placebo. My BOCF analysis did not show statistical significant difference between tramadol 300 mg dose group and placebo (31% for tramadol vs. 29% for placebo; p =0 .7064 > 0.0167). The dose-response relationship appears to be reversed in the BOCF analysis (36% for tramadol 100 mg, 32% for tramadol 200 mg, 31% for tramadol 300 mg, and 29% placebo). It is plausible that this is because the higher the dose, the more dropouts for toxicity among patients who were given good scores by LOCF. My continuous responder curves did not show separation between tramadol 300 mg dose group and placebo and van der Waerden test did not show a statistically significant separation. My continuous responder curves showed separation between tramadol 100 mg dose group and placebo. However, van der Waerden test did not show a statistically significant separation. (See Tables 7 – 11 and Figure 3 below.)

Table 5 Sponsor Analysis of WOMAC Pain (with site term not included in the model): Study MDT3-003 FAS with LOCF

Percent Change from Baseline to Endpoint (Week 12)				
	TRA100 mg (n=103)	TRA200 mg (n=107)	TRA300 mg (n=105)	PBO (n=224)
LSMean Percent Change (SE)	42% (5%)	43% (5%)	46% (5%)	32% (3%)
Diff. from PBO (95% CI)	10% (-2%, 21%)	11% (-0%, 22%)	14% (3%, 24%)	
p-value vs. placebo*	(0.0933)	0.0504	**0.0165**	
Holm's adjusted level	0.05	0.025 (=0.05/2)	0.0167 (=0.05/3)	

LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.
*p-values are compared with adjusted significance levels for multiplicity based on Holm's sequentially rejective method. The sequential testing procedure stops prior to calculating p-values in the parenthesis.

Table 6 Reviewer Analysis of WOMAC Pain (with site term included in the model): Study MDT3-003 FAS with LOCF

Percent Change from Baseline to Endpoint (Week 12)				
	TRA100 mg (n=103)	TRA200 mg (n=107)	TRA300 mg (n=105)	PBO (n=224)
LSMean Percent Change (SE)	41% (5%)	41% (5%)	44% (5%)	31% (4%)
Diff. from PBO (95% CI)	10% (-1%, 21%)	10% (-1%, 21%)	13% (2%, 24%)	
p-value vs. placebo*	(0.0710)	(0.0626)	0.0227	
Holm's adjusted level	0.05	0.025 (=0.05/2)	0.0167 (=0.05/3)	

LSMeans and p-values calculated from ANCOVA model: Y = trt + site + baseline.
*p-values are compared with adjusted significance levels for multiplicity based on Holm's sequentially rejective method. The sequential testing procedure stops prior to calculating p-values in the parenthesis.

Table 7 Reviewer Analysis of WOMAC Pain (with site term not included in the model): Study MDT3-003 FAS with LOCF

Absolute Change from Baseline to Endpoint (Week 12)				
	TRA100 mg (n=103)	TRA200 mg (n=107)	TRA300 mg (n=105)	PBO (n=224)
LSMean Change (SE)	126 (14)	130 (13)	136 (13)	98 (9)
Diff. from PBO (95% CI)	28 (4, 60)	32 (1, 63)	37 (6, 69)	
p-value vs. placebo*	(0.0830)	(0.0443)	0.0196	
Holm's adjusted level	0.05	0.025 (=0.05/2)	0.0167 (=0.05/3)	

LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.
*p-values are compared with adjusted significance levels for multiplicity based on Holm's sequentially rejective method. The sequential testing procedure stops prior to calculating p-values in the parenthesis.

Table 8 Reviewer Analysis of WOMAC Pain (with site term not included in the model): Study MDT3-003 FAS with BOCF

Percent Change from Baseline to Endpoint (Week 12)				
	TRA100 mg (n=103)	TRA200 mg (n=107)	TRA300 mg (n=105)	PBO (n=224)
LSMean Percent Change (SE)	36% (4%)	32% (4%)	31% (4%)	29% (3%)
Diff. from PBO (95% CI)	7% (-3%, 17%)	3% (-6%, 13%)	2% (-8%, 12%)	
p-value vs. placebo*	0.1682	(0.4843)	(0.7064)	
Holm's adjusted level	0.0167 (=0.05/3)	0.025 (=0.05/2)	0.05	

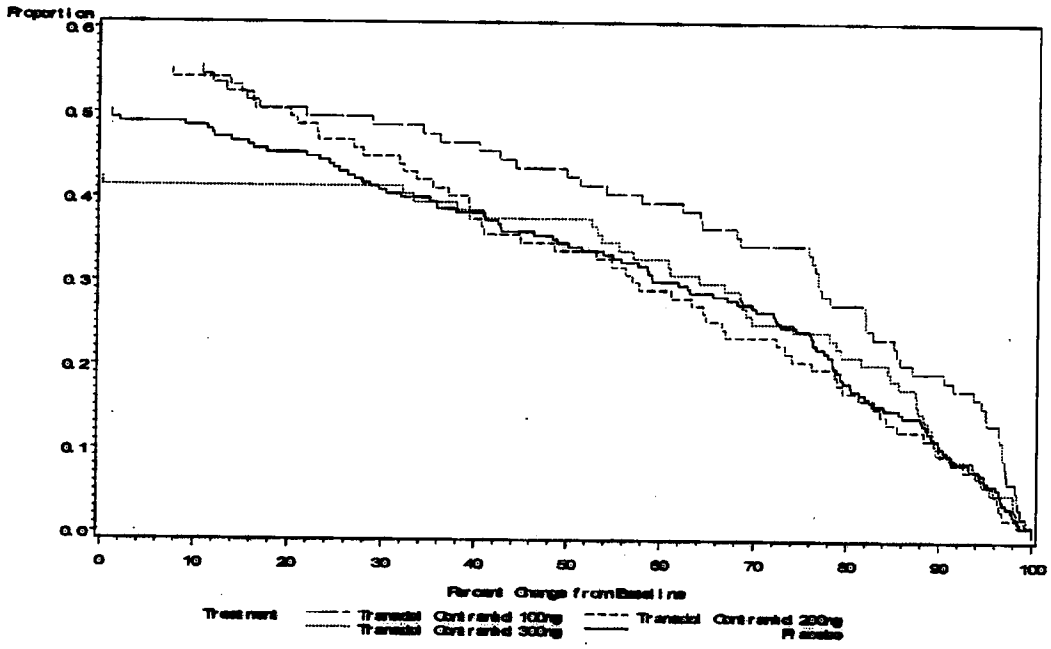
LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.
*p-values are compared with adjusted significance levels for multiplicity based on Holm's sequentially rejective method. The sequential testing procedure stops prior to calculating p-values in the parenthesis.

Table 9 Reviewer Analysis of WOMAC Pain (with site term not included in the model): Study MDT3-003 ITT with LOCF

Percent Change from Baseline to Endpoint (Week 12)				
	TRA100 mg (n=106)	TRA200 mg (n=111)	TRA300 mg (n=108)	PBO (n=227)
LSMean Percent Change (SE)	39% (5%)	42% (4%)	44% (5%)	32% (3%)
Diff. from PBO (95% CI)	7% (-3%, 18%)	10% (-1%, 20%)	12% (2%, 23%)	
p-value vs. placebo*	(0.1781)	(0.0696)	0.0253	
Holm's adjusted level	0.05	0.025 (=0.05/2)	0.0167 (=0.05/3)	

LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.
*p-values are compared with adjusted significance levels for multiplicity based on Holm's sequentially rejective method. The sequential testing procedure stops prior to calculating p-values in the parenthesis.

Figure 2 Continuous Responder Analysis: Study MDT3-003 FAS



Absolute Change from Baseline to Endpoint (Week 12)				
	TRA100 mg (n=103)	**TRA200 mg (n=107)**	**TRA300 mg (n=105)**	**PBO (n=224)**
p-value vs. placebo*	0.3940	0.6451	0.5927	

*p-values calculated from van der Waerden Nonparametric test.

Study MDT3-005:

Data from the study showed statistically significant difference of tramadol Contramid OAD when compared to placebo in primary endpoint, Pain Intensity Numeric Rating Scale at Week 12 in patients with OA of the knee. The statistically significant difference was shown by the sponsor in analysis on FAS with LOCF using ANCOVA model with terms for treatment and baseline pain as covariate (4.3 for tramadol vs. 4.8 for placebo; p = 0.0157). My analysis with center term in the ANCOVA model also showed statistically significant difference between tramadol dose group and placebo (4.1 for tramadol vs. 4.6 for placebo; p = 0.0254).

However, my BOCF analysis on FAS population did not show statistically significant difference between tramadol dose group and placebo (4.8 for tramadol vs. 5.0 for placebo; p =0.2134). Although my continuous responder curves showed separation between tramadol dose group and placebo, van der Waerden test did not show a statistically significant separation. (See Tables 12 – 15 and Figure 4 below.)

Table 10 Sponsor Analysis of Pain Intensity Score (with site term not included in the model): Study MDT3-005 FAS with LOCF

Pain Intensity Score at Endpoint (Week 12)		
	TRA CONT (n=431)	**PBO (n=214)**
Baseline Mean (SD)	7.2 (1.6)	7.2 (1.6)
Endpoint LSMean (SE)	4.3 (0.1)	4.8 (0.2)
Difference vs PBO (95% CI)	-0.5 (-0.9, -0.1)	
p-value vs. placebo	**0.0157**	

LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.

Table 11 Reviewer Analysis of Pain Intensity Score (with site term included in the model): Study MDT3-005 FAS with LOCF

Pain Intensity Score at Endpoint (Week 12)		
	TRA CONT (n=431)	PBO (n=214)
Baseline Mean (SD)	7.2 (1.6)	7.2 (1.6)
Endpoint LSMean (SE)	4.1 (0.2)	4.6 (0.2)
Difference vs PBO (95% CI)	-0. 5 (-0.9, -0.1)	
p-value vs. placebo	0.0254	

LSMeans and p-values calculated from ANCOVA model: Y = trt + site + baseline.

Table 12 Reviewer Analysis of Pain Intensity Score (with site term not included in the model): Study MDT3-005 FAS with BOCF

Pain Intensity Score at Endpoint (Week 12)		
	TRA CONT (n=431)	PBO (n=214)
Baseline Mean (SD)	7.2 (1.6)	7.2 (1.6)
Endpoint LSMean (SE)	4.8 (.1)	5.0 (.2)
Difference vs PBO (95% CI)	-0.2 (-0.7, 0.1)	
p-value vs. placebo	0.2134	

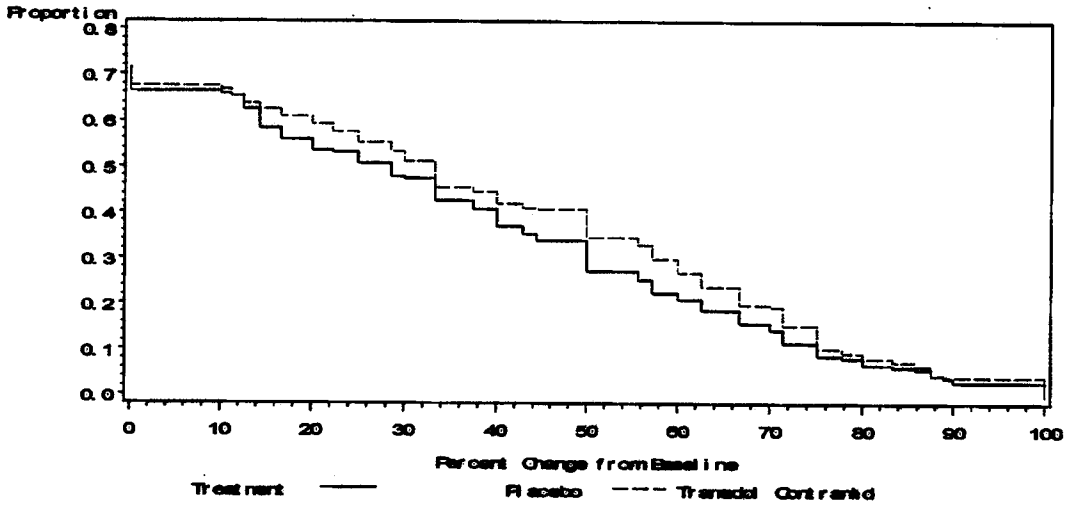
LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.

Table 13 Sponsor Analysis of WOMAC Pain (with site term not included in the model): Study MDT3-002 FAS with LOCF

Percent Change from Baseline to Endpoint (Week 12)		
	TRA CONT (n=431)	**PBO (n=226)**
LSMean Percent Change (SE)	37% (38%)	31% (40%)
Diff. from PBO (95% CI)	6% (-0%, 12%)	
p-value vs. placebo	0.0584	

LSMeans and p-values calculated from ANCOVA model: Y = trt + baseline.

Figure 3 Continuous Responder Analysis: Study MDT3-005 FAS



Absolute Change from Baseline to Endpoint (Week 12)		
	TRA CONT (n=431)	**PBO (n=226)**
p-value vs. placebo*	0.3466	

*p-values calculated from van der Waerden Nonparametric test.

3.2 Evaluation of Safety

Safety analyses were done by Clinical reviewer, Jin Chen, M.D.

No statistical problems or issues were found.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

In subgroup analyses for the studies MDT3-002, MDT3-003, and MDT3-005, there were no statistically significant interactions between treatment and age, sex, or race in the WOMAC Pain Subscale Score and Pain Intensity Numeric Rating Scale.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

5.1.1 Statistical Issues

For the efficacy analysis, the sponsor based their inferences on full analysis set (FAS) data with LOCF for a statistically significant difference on WOMAC Pain Subscale in studies MDT3-002 and MDT3-003 and on Pain Intensity Numerical Rating Scale in study MDT3-005 comparing tramadol Contramid OAD with placebo. Sponsor's FAS population was defined as all randomized patients who received at least one dose of study medication and at least one post baseline pain assessment. However, since ITT population defined as all randomized patients is more appropriate as the primary analysis set, I conducted the same analysis as the sponsor based on ITT population as a sensitivity analysis.

The sponsor proposed LOCF in order to impute 'missing data' due to dropouts in their primary analysis. However, FDA requested a sensitivity analysis using alternative imputation methods. I construe the agreement between the sponsor and FDA that sensitivity analyses should support a successful LOCF analysis for a study to be a success. I think that non-existence of the outcome variable at target endpoint does not necessarily mean 'missing data' and a method assigning bad scores to 'missing' outcomes due to dropout is desirable. In this sense, LOCF is not a good way of handling dropouts because it could assign good scores and BOCF is a more acceptable method because it usually assigns bad scores.

The sponsor provided post hoc sensitivity analyses assessing their conclusion with respect to imputation methods for missing data due to dropout. The sponsor conducted BOCF analysis, Time-Weighted Average analysis, and Repeated Measures ANOVA. I also conducted a continuous responder analysis, in which a continuous responder curve for each group is generated by changing responder criterion from 0% to 100% improvement from baseline to Week 12 and the curves are compared. Along with the continuous responder analysis, I used van der Waerden normal score test to compare tramadol groups and placebo. Since the test is also post hoc, it is subject to multiplicity. The sponsor also conducted a responder analysis with 30% response criterion. However, the analysis was problematic because they used LOCF for dropouts before determining the response.

The sponsor prespecified ANCOVA model with terms for treatment, center, and baseline pain score as a covariate as the primary analysis. However, they dropped the center term from the model with the reason that many centers had too few subjects. I conducted an analysis including the center term in the model as a sensitivity analysis.

The sponsor proposed Holm's sequentially rejective method to adjust for multiple comparisons in the studies MDT3-002 and MDT3-003, which is acceptable.

5.1.2 Collective Evidence

Put together, the data from the three studies - MDT3-002, MDT3-003, and MDT3-005 – failed to provide a substantial evidence of analgesic efficacy of tramadol Contramid OAD. Studies MDT3-003 and MDT3-005 showed statistically significant difference between tramadol Contramid OAD and placebo in analyses specified in the protocol. However, the statistically significant difference was not strong enough to retain the significance in more conservative analysis such as BOCF analysis and continuous responder analysis where dropouts are treated as non-improvement or non-responder, respectively.

5.2 Conclusions and Recommendations

The study MDT3-002 with knee OA patients failed to show a statistically significant difference on WOMAC Pain Subscale Score, percent change from baseline to Week 12, as primary outcome variable at all dose levels (100 mg, 200 mg, and 300 mg) of tramadol Contramid OAD compared to placebo in the LOCF analysis on the FAS.

The study MDT3-003 with knee OA patients showed a statistically significant difference on WOMAC Pain Subscale Score, percent change from baseline to Week 12, at dose of 300 mg of tramadol Contramid OAD compared to placebo in the LOCF analysis on the FAS.

The study MDT3-005 with knee OA patients showed a statistically significant difference on Pain Intensity Numerical Rating Scale at Week 12 at dose of 200 mg or 300 mg of tramadol Contramid OAD compared to placebo in the LOCF analysis on the FAS. However, the statistically significant difference shown in the studies MDT3-003 and MDT3-005 was sensitive to dropout handling methods and to choice of analysis set. The statistically significant difference was not shown in the BOCF analysis, in the continuous responder analysis, or in the ITT LOCF analysis in both studies.

Overall, although the evidence of efficacy was replicated in two well controlled studies MDT3-003 and MDT3-005, the submitted data of studies with high dropout rates failed to provide substantial evidence supporting pain indication of tramadol Contramid once a day formulation because the efficacy shown might be driven by imputation of missing data due to dropouts.

5.3 Review of Clinical Studies of Proposed Label

Following is the text portion in the Clinical Study section from 'PROPOSED LABELING TEXT' regarding results from the three efficacy studies:

3 Page(s) Withheld

_____ Trade Secret / Confidential (b4)

✓ Draft Labeling (b4)

_____ Draft Labeling (b5)

_____ Deliberative Process (b5)

b(4)

I found that the clinical study section of the proposed label is not consistent with my review of the study reports.

APPENDIX

Table 14 Patient Disposition by Treatment Group

Study MDT3-002:

	TRA100 mg	TRA200 mg	TRA300 mg	PLACEBO	TOTAL
Randomized	110	113	115	227	565
FAS	109	110	113	226	558
PP	64	56	48	130	298
Completers	64 (58%)	60 (53%)	54 (47%)	144 (63%)	322 (57%)
Dropouts	46 (42%)	53 (47%)	61 (53%)	83 (37%)	243 (43%)
AE	21 (19%)	19 (17%)	41 (36%)	10 (4%)	91 (16%)
LOE	17 (15%)	15 (13%)	13 (11%)	52 (23%)	97 (17%)
Other	8	19	7	21	45

Study MDT3-003:

	TRA100 mg	TRA200 mg	TRA300 mg	PLACEBO	TOTAL
Randomized	106	111	108	227	552
FAS	103	107	105	224	539
PP	61	63	49	126	229
Completers	62 (58%)	65 (59%)	50 (46%)	134 (59%)	311 (56%)
Dropouts	44 (42%)	46 (41%)	58 (54%)	93 (41%)	241 (44%)
AE	13 (12%)	20 (18%)	35 (32%)	17 (7%)	85 (15%)
LOE	21 (20%)	11 (10%)	11 (10%)	47 (21%)	90 (16%)
Other	10	15	12	29	66

Study MDT3-005:

	TRA CONT	PLACEBO	TOTAL
OL Enrolled			**1028**
DB Randomized (ITT)	**432**	**214**	**646**
FAS	**431**	**214**	**645**
PP	**378**	**192**	**570**
Completers	**326 (75%)**	**165 (77%)**	**491 (76%)**
Dropouts	**106 (25%)**	**49 (23%)**	**155 (24%)**
AE	**44 (10%)**	**11 (5%)**	**55 (9%)**
LOE	**34 (8%)**	**22 (10%)**	**56 (9%)**
Other	**28**	**16**	**44**

Table 15 Patient Demographics and Baseline Efficacy Variable (FAS Subjects)

Study MDT3-002:

	TREATMENT				
	TRA100 MG (N=109)	TRA200 MG (N=110)	TRA300 MG (N=113)	PLACEBO (N=226)	TOTAL (n=558)
Gender n (%)					
Male	45 (41%)	47 (43%)	43 (38%)	87 (38%)	222 (40%)
Female	64 (59%)	63 (57%)	70 (62%)	139 (62%)	336 (60%)
Race n (%)					
Asian	3 (3%)	1 (.9%)	2 (2%)	3 (1%)	9 (2%)
Black	11 (10%)	9 (8%)	11 (10%)	31 (14%)	62 (11%)
Caucasian	83 (76%)	87 (79%)	92 (81%)	176 (78%)	438 (78%)
Hispanic	11 (10%)	13 (12%)	8 (7%)	15 (7%)	47 (8%)
Other	1 (.9%)	-	-	1 (.4%)	2 (.4%)
Age (years)					
Mean ± SD	60 ± 9	60 ± 8	61 ± 10	61 ± 10	60 ± 9
Median	61	59	62	62	62
Range	40 - 75	40 - 74	40 - 76	41 - 80	40 - 80
BMI (kg/m^2)					
Mean ± SD	30.5 ± 4.8	30.6 ± 4.3	30.4 ± 4.5	31.2 ± 4.7	30.8 ± 4.6
Median	31.2	30.2	30.9	32.0	31.2
Range	20.1 – 39.8	18.4 – 38.7	19.8 – 41.8	19.6 – 44.4	18.4 – 44.4
WOMAC Pain Score					
Mean ± SD	299.6 ± 81.4	310.5 ± 96.5	308.7 ± 89.2	302.4 ± 85.9	
Median	294.0	316.5	311.0	304.5	
Range	162 - 485	114 - 495	145 - 495	145 - 494	

Study MDT3-003:

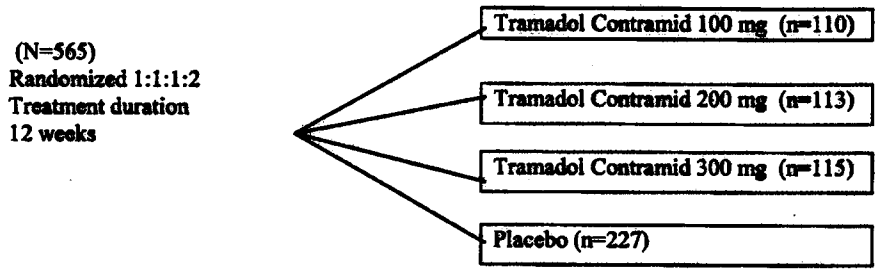
	TREATMENT				TOTAL (n=539)
	TRA100 MG (N=103)	TRA200 MG (N=107)	TRA300 MG (N=105)	PLACEBO (N=224)	
Gender n (%)					
Male	41 (40%)	43 (40%)	36 (34%)	86 (38%)	206 (38%)
Female	62 (60%)	64 (60%)	69 (66%)	138 (62%)	333 (62%)
Race n (%)					
Asian	2 (2%)	1 (.9%)	2 (2%)	-	5 (.9%)
Black	10 (10%)	6 (6%)	12 (11%)	25 (11%)	53 (10%)
Caucasian	72 (70%)	83 (78%)	73 (70%)	159 (71%)	387 (72%)
Hispanic	18 (17%)	16 (15%)	17 (16%)	35 (16%)	86 (16%)
Other	1 (1%)	1 (.9%)	1 (1%)	5 (2%)	8 (1.1%)
Age (years)					
Mean ± SD	63 ± 8	61 ± 9	60 ± 9	61 ± 10	61 ± 9
Median	63	62	61	62	62
Range	40 - 76	39 - 75	39 - 75	40 - 82	39 - 82
BMI (kg/m^2)					
Mean ± SD	30.7 ± 4.5	30.2 ± 4.6	31.0 ± 4.0	30.7 ± 4.6	30.6 ± 4.4
Median	30.5	30.0	31.0	30.5	30.6
Range	21.4 – 39.9	19.5 – 38.1	22.0 – 43.2	19.1 – 46.2	19.1 – 46.2
WOMAC Pain Score					
Mean ± SD	287.8 ± 78.8	283.8 ± 81.7	314.4 ± 97.1	300.7 ± 88.8	
Median	286.0	278.0	308.0	295.0	
Range	161 - 480	149 - 458	63 - 497	94- 495	

Study MDT3-005:

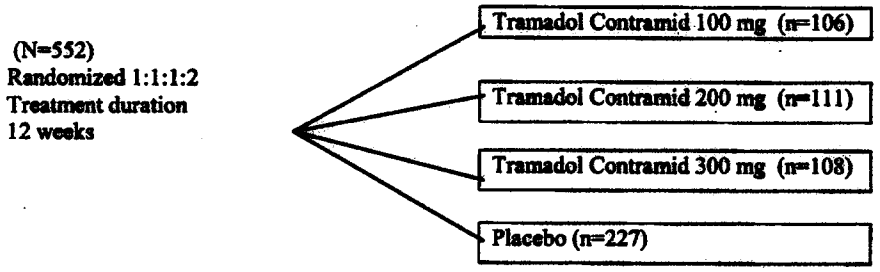
	TREATMENT		
	TRA CONT (N=431)	PLACEBO (N=214)	TOTAL (n=645)
Gender n (%)			
Male	157 (36%)	81 (38%)	238 (37%)
Female	274 (64%)	133 (62%)	407 (63%)
Race n (%)			
Asian	1 (.2%)	1 (.5%)	2 (.3%)
Black	21 (5%)	12 (6%)	33 (5%)
Caucasian	379 (88%)	185 (86%)	564 (87%)
Hispanic	28 (6%)	15 (7%)	43 (7%)
Other	2 (.5%)	1 (.5%)	3 (.5%)
Age (years)			
Mean ± SD	62 ± 9	62 ± 9	62 ± 9
Median	63	63	63
Range	41 - 80	41 - 79	41 - 80
BMI (kg/m^2)			
Mean ± SD	29.7 ± 4.0	29.5 ± 4.3	29.6 ± 4.1
Median	29.9	30.1	30.0
Range	19.5 – 37.6	19.5 – 37.4	19.5 – 37.6
Pain Intensity Score			
Mean ± SD	7.2 ± 1.6	7.2 ± 1.6	
Median	7.0	7.0	
Range	3 - 10	4 - 10	
WOMAC Pain Score			
Mean ± SD	11.2 ± 3.5	11.1 ± 3.2	
Median	11.0	11.0	
Range	2 - 20	3 - 20	

Figure 4 Schematic of Study Design

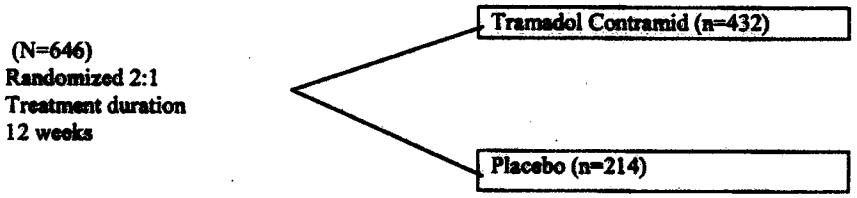
Study MDT3-002:



Study MDT3-003:



Study MDT3-005:



SIGNATURES/DISTRIBUTION LIST

Primary Statistical Reviewer:	Yongman Kim, Ph.D. Mathematical Statistician
Date:	August 1, 2006
Concurring Reviewer:	Thomas Permutt, Ph.D. Acting Division Director

cc:

DAARP/Paul Balcer
DAARP/Jin Chen, M.D.
HFD-715/Yongman Kim, Ph.D.
HFD-715/Thomas Permutt, Ph.D.
HFD-715/Edward Nevius, Ph.D.
HFD-715/Steve Wilson, Ph.D.
HFD-700/Lillian Patrician
HFD-700/Robert O'Neil, Ph.D.

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

/s/

Yongman Kim
8/1/2006 12:04:06 PM
BIOMETRICS

Thomas Permutt
8/1/2006 12:47:03 PM
BIOMETRICS
concur