# CENTER FOR DRUG EVALUATION AND RESEARCH

***APPLICATION NUMBER:***

**22-348**

# 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

## AMENDMENT

| | |
|---|---|
| **NDA:** | 22-348/000 |
| **Drug Name:** | CALDOLOR (ibuprofen) Injection |
| **Indication(s):** | Management of mild to severe pain, reduction of fever |
| **Applicant:** | Cumberland Pharmaceuticals Inc. |
| **Date(s):** | Letter 12/11/2008, PDUFA 6/11/2009 |

| | |
|---|---|
| **Biometrics Division:** | Division of Biometrics II |
| **Statistical Reviewer:** | Jonathan D. Norton, Ph.D. |
| **Concurring Reviewers:** | Dionne Price, Ph.D. |

| | |
|---|---|
| **Medical Division:** | Division of Anesthesia, Analgesia, and Rheumatology Products |
| **Clinical Team:** | Christina Fang, M.D. |
| | Ellen Fields, M.D. |
| **Project Manager:** | Kathleen Davies |

**Keywords:** clinical studies, NDA review, protocol violations

## 1 Background

On May 8, 2009, I submitted my original statistical review for Caldolor. My findings were that the submitted studies showed adequate evidence of efficacy for fever, but that no conclusions could be reached from the two pain studies. In the review I expressed concerns with the conduct, analysis, and robustness of results of the pain studies. On May 17, Dionne Price submitted a secondary review which focused on those studies. Dr. Price concurred with my assessment that Study CPI-CL-008A did not show statistical significance, but concluded that Study CPI-CL-008B showed "evidence that Caldolor 800 mg decreases the amount of morphine needed following surgery."

On May 19, the Division of Scientific Investigations (DSI) submitted the Clinical Inspection Summary for this NDA. This review is in response to the DSI summary, and will focus mainly on the pain studies.

## 2 Review

DSI inspected four sites, the principal investigators of which were Dr. Peter Morris, Dr. John Promes, Dr. Henry Frazer, and Dr. Lamar Snow. The first two sites participated in a fever study, CPI-CL-004. The Summary does not report any irregularities from Dr. Morris' site. At Dr. Promes' site, a patient with a closed head injury was enrolled in violation of protocol. DSI concluded that study conduct was adequate at both of these sites and that their data could be used in support of registration.

Dr. Henry Frazer participated in Study CPI-CL-008A, which I will refer to as Study 8A. The Clinical Inspection Summary does not list any irregularities, "based on communications with the FDA Investigator". The Establishment Inspection Report had not been reviewed at the time the Summary was submitted, so the conclusions were subject to change.

Dr. Lamar Snow participated in both pain studies. The Summary noted several regulatory violations. For four subjects, there were discrepancies between the source documents and the submitted data listings for the primary endpoint, 24-hour morphine usage. One of the subjects was in enrolled in Study 8A, the other three in Study 8B (CPI-CL-008B). Also, the treatment assignments of two subjects in enrolled in Study 8A were unblinded to the study staff. Another violation was that two ineligible patients were enrolled in Study 8B. Of particular concern is the following statement in regard to these subjects in the Summary, "The timings of the morphine doses for subjects 5056 and 8060 on protocol 008B ***appear to be altered*** [emphasis added] in the copies of hospital records maintained at the study site compared with source documents maintained at the hospital." As a general comment, the

Summary notes that "Dr. Snow, the clinical investigator (CI) did not appear to have adequate oversight of the study."

DSI's overall conclusion for Dr. Snow's site was as follows,

> The reported morphine totals for the 4 subjects noted above could not be verified by the source documents, and one AE of nausea and vomiting was not reported to the sponsor. Given the inability to verify the primary endpoint data for the 4 subjects noted above, the review division may wish to consider excluding from analyses the efficacy data from these subjects. Data generated by this site for the other subjects appear acceptable in support of the pending application.

Following a request from Dr. Ellen Fields, I replicated several of my analyses of the primary endpoint with all subjects from this site excluded. See Table 1 below, which is comparable to Table 16 in my original review.

**Table 1: Analyses of Primary Endpoint, without Snow**

| Imputation | Transform | Model | p-value for Arm |
|---|---|---|---|
| Applicant Method 1 | None | Planned stepwise | .115 |
| Applicant Method 1 | None | Reduced stepwise | .0036 |
| Applicant Method 1 | Log(x + 1) | Reduced stepwise | .0002 |
| Applicant Method 1 | Square root | Reduced stepwise | .0006 |
| Applicant Method 1 | Rank | Reduced stepwise | .0003 |
| Conservative | Log(x + 1) | Reduced stepwise | .0003 |

The results from these analyses are qualitatively similar to those from the full ITT set (as shown in Table 16). The 800 mg dose of Caldolor is found to be significantly better than placebo for all analyses conducted except one fitting the "planned stepwise" model to the untransformed primary endpoint. Note that what I call the "planned stepwise" model was the result of my attempt to follow the vague instructions in the Applicant's prospective Statistical Analysis Plan. I consider the "Reduced stepwise" model to be more appropriate, albeit not prospectively chosen.

In addition to the results shown in Table 1, I also ran the same analyses with only five subjects excluded from the Snow site. These included the three subjects noted earlier who showed a discrepancy on the primary endpoint, as well as the two subjects who were found to be ineligible. The "planned stepwise" model showed a marginally significant difference ($p = .06$) between Caldolor 800 mg and placebo, and the other analyses showed the difference to be highly significant ($p < .001$).

## 3 Conclusions and Recommendations

In my original review I discounted the strength of evidence from the pain studies because I was concerned about their quality. The findings from DSI are consistent with my concerns, as the data from Dr. Snow's site were compromised by multiple regulatory violations. Admittedly, my sensitivity analyses showed that excluding the data from this site does not change the previous findings of statistical significance under several models. However, this result does not alleviate my concerns about the overall quality of the pain studies. My original determination, that these studies were "inadequate to reach any conclusion about efficacy" for pain, remains unchanged.

**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
6/2/2009 08:41:50 AM
BIOMETRICS

Dionne Price
6/2/2009 04:51:18 PM
BIOMETRICS
I concur that the results of the analyses excluding Dr. Snow's site are consistent with those including the site.

DEPARTMENT OF HEALTH AND HUMAN SERVICES
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH
OFFICE OF BIOSTATISTICS

# Statistical Review and Evaluation

## SECONDARY REVIEW

| | |
|---|---|
| NDA: | 22-348 |
| Name of drug: | Caldolor (ibuprofen) Injection |
| Applicant: | Cumberland Pharmaceuticals Inc. |
| Indication: | Reduction of fever; Treatment of mild to severe pain |
| Dates: | Cumberland Pharmaceuticals Inc. |
| Review Priority: | Priority |
| Biometrics Division: | Division of Biometrics II |
| Statistical reviewer: | Jonathan Norton, Ph.D. |
| Secondary reviewer: | Dionne Price, Ph.D. |
| Medical Division: | Division of Anesthesia, Analgesia, and Rheumatology Products |
| Clinical Team: | Christina Fang, M.D. |
| | Ellen Fields, M.D. |
| Project Manager: | Kathleen Davies |

Keywords: clinical studies, NDA review

# 1 BACKGROUND

Cumberland Pharmaceuticals proposes Caldolor (ibuprofen) Injection for the reduction of fever and management of mild to severe pain. The primary statistical review of the new drug application (NDA) was conducted by Dr. Jonathan Norton. Dr. Norton's assessment of the efficacy of Caldolor focused on four studies, two studies per proposed indication. Dr. Norton concluded that the clinical studies demonstrated the efficacy of Caldolor 400 mg in the reduction of fever. He found the clinical studies submitted in support of the latter indication to be inadequate to reach any conclusion.

Caldolor was developed under IND 62,605. The Applicant submitted a special protocol for Study CPI-CL-008 dated September 13, 2004. At the time of protocol submission, the study included two parts, Part A and Part B. An agreement was not reached and comments were conveyed to Cumberland. The statistical comments are of relevance to the NDA review and included the following:

> The intent-to-treat population should include all randomized patients with at least one dose of study medication. The intent-to-treat population should be the primary population that the efficacy analyses are based on.
>
> When analyzing the primary endpoints, important baseline characteristics used as randomization strata, i.e., age and weight, can be included in the ANOVA model to reduce variation and provide more precise results.
>
> The sponsor had listed multiple secondary endpoints. Secondary endpoints should be clinically relevant. If the secondary endpoints are intended for labeling, an adjustment for multiplicity is needed.
>
> In reference to your proposal on the management of missing morphine use data, we disagree with imputing the treatment group mean for patients discontinuing early or with more conservative missing time points. We suggest a more conservative approach such as imputing the mean placebo response for either the active treatment or placebo group or cross imputing the treatment group means. We also note your intention to submit a detailed Statistical Analysis Plan and we will further review the method of imputing missing data in that submission.

A Guidance meeting was held July 15, 2005. Cumberland inquired about the adequacy of the Statistical Analysis Plans (SAP) for Study CPI-CL-004, (b) (4) and Study CPI-CL-008. For Study CPI-CL-008, the Division recommended that Cumberland adjust the analysis to reflect variables used to stratify the randomization. On May 29, 2008, a pre-NDA was held. A question posed in the meeting package pertained to the non-normality of the data.

The question and response were as follows:

**Cumberland Question:**

Reference is made to pivotal studies CPI-CL-008A and CPI-CL-008B. Both studies met their primary endpoint demonstrating a statistically significant reduction in morphine use in post-surgical patients receiving Ibuprofen Injection compared to those receiving placebo. The data from these studies are not normally distributed. Appropriate statistical testing for normality and analyses (log transformation as specified in the statistical analysis plan (attached in Appendix A)) will be presented in the final study report. Further, an additional transformation, the Box-Cox transformation, was applied for robustness of analysis. Does the agency agree that including the log transformation and Box- Cox transformation would be adequate to provide in the NDA?

**FDA Response:**

Yes, provide the results from all analyses in the NDA.

## 2 REVIEW

I am in agreement with Dr. Norton's assessment of the studies evaluating Caldolor for the reduction of fever. Consequently, my review will focus on the studies submitted in support of the proposed (b) (4) pain indication.

### 2.1 STUDY CPI-CL-008A

Study CPI-CL-008A was a randomized, double-blind, placebo-controlled study conducted in postoperative adult patients. Details of the study design are provided in Dr. Norton's review. The primary efficacy variable was total morphine usage in the 24 hours following surgery. In the clinical study report and statistical analysis plan, the Applicant stated that analysis of variance (ANOVA) and covariance procedures would be used and that the primary model would include terms for treatment and center. The Applicant allowed for the inclusion of additional terms in the model for "sensitivity analysis and robustness." In addition, the Applicant stated that the assumptions underlying the analysis of variance and covariance would be evaluated and that the data would be transformed if warranted based on a violation of the assumptions. Lastly, the Applicant's analysis plan allowed for the possibility that analog nonparametric procedures would be used to corroborate the results.

Analyses were conducted on the intent-to-treat population including all randomized patients receiving one dose of treatment. The Applicant employed Dunnett's test to address multiplicity concerns. Missing data was imputed using the mean of the assessments immediately prior to and after the missing value for patients having no more than two consecutive missing assessments. The mean morphine use in the placebo group

(for the corresponding time point) was imputed for patients having more than two consecutive missing assessments. Missing data was replaced by the mean morphine use from the placebo group for each time point for patients dropping out of the study. As a sensitivity analysis, the Applicant also utilized the mean morphine use among the placebo group for all missing data. The strategies for handling missing data were consistent with the advice provided by the Agency.

The results of the analyses conducted on the original (i.e. untransformed) and transformed data are provided in Table 1.

**Table 1: Applicant's results for original and rank transformed data**

| | Placebo (n=134) | Caldolor 400 mg (n=134) | Caldolor 800 mg (n=138) |
|---|---|---|---|
| Total Morphine Use (mg) | | | |
| LS Mean (SE) | 49 (4) | 46 (4) | 44 (3) |
| LS Mean difference from placebo | | -3 | -5 |
| 95% CI of difference | | (-10,5) | (-13,3) |
| Pairwise p-value[1] | | 0.667 | 0.237 |
| Rank Transformed Morphine Use (mg) | | | |
| LS Mean (SE) | 223 (14) | 209 (14) | 191 (13) |
| LS Mean difference from placebo | | -14 | -32 |
| 95% CI of difference | | (-44,16) | (-62,-3) |
| Pairwise p-value[1] | | 0.458 | 0.030 |

Sources: Clinical Study Report Table 21 and Table 14.2.1.1.1
[1] Analyses based on an ANOVA model with factors for age, weight, center, and treatment

Dr. Norton appropriately expressed concern that the Applicant's primary results were based on an ANOVA model of the rank-transformed data that included effects for age, weight, center, and treatment although the study report and SAP suggested the primary analysis on the original data would employ an ANOVA model with effects for center and treatment only. To address the concern, Dr. Norton conducted an analysis of the data employing a model with effects for center and treatment. His analysis was conducted with and without data transformations. A statistically significant difference between Caldolor 800 mg and placebo was evident for the transformed data only.

**Table 2: Dr. Norton's analysis based on the SAP**

| | Placebo (n=134) | Caldolor 400 mg (n=134) | Caldolor 800 mg (n=138) |
|---|---|---|---|
| Total Morphine Use (mg) | | | |
| LS Mean (SE) | 50 (4) | 48 (4) | 45 (3) |
| LS Mean difference from placebo | | -2 | -5 |
| Pairwise p-value[1] | | 0.720 | 0.230 |

Source: Dr. Norton's statistical review, Table 13
[1] Analysis based on ANOVA model with treatment and center as factors

I agree that the Applicant did not specify the analysis in a concise manner. However in my opinion, the choice of the model was not completely arbitrary. Cumberland received previous advice from the Agency suggesting that randomization factors (i.e. age and weight) be included in the model. For additional clarity, an information request (dated May 15, 2009) was sent to the Applicant. The request and response are provided in the Appendix of this review. The Applicant acknowledged that age and weight were included in the model as "secondary covariates". According to the Applicant, a decision was made to retain all randomization factors. The Applicant further stated that inclusion of the factors did not change the inference. While the response confirmed the seemingly post-hoc nature of the selection of the covariates to be included in the final model, I agree with the Applicant in the assertion that inclusion of age and weight covariates did not appear to change the overall conclusions as evidenced by a comparison of the Applicant's results to those of Dr. Norton.

In the Clinical Study Report (CSR), the Applicant presented the results for the rank-transformed data in the main body of the report and presented results from the original data in *Section 14: Tables, Figures, and Graphs Referred To But Not Included In The Text.* The rank transformation approach applied by the Applicant was analogous to a nonparametric procedure and evaluated the difference in mean ranks between the treatment groups. The analysis methods applied by the Applicant yielded conflicting results as indicated in Table 1. According to the Applicant, the ANOVA model conducted on the original data, the log transformed data, and the Box-Cox transformed data violated the underlying assumption of normality. Thus, a nonparametric procedure was used and deemed most appropriate (by the Applicant).

In general, nonparametric approaches are useful statistical tools that require less restrictive assumptions compared to their parametric counterparts. I did not necessarily agree that the nonparametric procedure was the most appropriate analytical approach since an ANOVA model performed on the original data should be robust with respect to moderate departures from the basic assumptions. Specifically, the normality assumption can be relaxed provided the sample size is large and/or the departures from normality are not extreme. For this reason, I expected the results from the parametric analysis and the nonparametric analysis to yield consistent results. While the presence of outliers or unequal variances could have produced inconsistent findings, it was difficult to determine the cause of the conflicting results. I based my conclusions on the pre-specified analysis defined as an analysis of variance on the original data. The analysis plan specified that an analog nonparametric procedure might be used to "corroborate" the results. In this study, the nonparametric approach performed did not "corroborate" the results from the primary analysis as intended by the Applicant.

Dr. Norton stated that the imputation strategies were not clearly documented, and he was not convinced that the Applicant used the pre-specified strategies for handling missing data. His analyses were conducted based on his interpretation of the pre-specified strategies. While Dr. Norton's course of action was appropriate given his unease, I was

not overly concerned about the handling of missing data since the number of discontinuations appeared small and balanced across treatment groups. Moreover, the conclusions remain the same regardless of the strategy utilized. My lack of concern regarding missing data does not negate the seriousness of Dr. Norton's assertion that the Applicant did not use the planned imputation strategy.

After consideration of Dr. Norton's analysis of the data (as specified in the protocol), the Applicant's analysis of the original data incorporating the randomization factors, and Dr. Norton's concerns, I conclude that the Applicant has failed to demonstrate a statistically significant difference in the total morphine use between arms.

### 2.2 STUDY CPI-CL-008B

Study CPI-CL-008B was a randomized, double-blind, placebo-controlled study conducted in postoperative adult patients. Details of the study design are provided in Dr. Norton's review. Similar to Study CPI-CL-008A, the primary efficacy variable was total morphine usage in the 24 hours following surgery. The analyses specified in the final CSR dated October 22, 2008 varied from the analyses specified in the SAP dated January 10, 2008.

The analyses specified in the CSR were similar to those specified in Study CPI-CL-008A. However, the primary analysis outlined in the SAP included a model selection procedure. Such a strategy is used to determine the particular variables that should be in the final model. The initial model included factors for center, age, weight, treatment, and all possible two-way and three-way interactions. Both the analysis specified in the SAP as well as the CSR included an evaluation of the assumptions underlying the analysis of variance and covariance and stated that the data would be transformed if warranted based on a violation of the assumptions. The Applicant stated, "...additional techniques may be applied to investigate the robustness of any conclusions from the model derived using the methodology described above. Among the methods that may be attempted are log and square root transformations; non-parametric methods, such as those suggested by Conover and Inman (1981), may also be attempted."

As a result of the discrepancy between the CSR and the SAP, an information request was sent (May 13, 2009) for clarification. In response, the Applicant clarified that the language in the CSR was "not representative of the analyses performed and presented in the CSR." According to the Applicant, the SAP was followed and the model selection procedure was used to derive the final model. The information request and response are provided in the Appendix of this review. The Applicant additionally provided documentation which more clearly demonstrated the model selection procedure that was utilized. The Applicant's final model included factors for treatment, age, weight, and center. The results are provided in Table 3.

**Table 3: Applicant's results for original and Box-Cox transformed data**

| | Placebo (n=153) | Caldolor 800 mg (n=166) |
|---|---|---|
| Total Morphine Use | | |
| LS Mean (SE) | 57 (2) | 49 (2) |
| LS Mean difference from placebo | | -8 |
| 95% CI of difference | | (-13,-3) |
| Pairwise p-value[1] | | <0.001 |
| Box-Cox Transformed Morphine Use | | |
| LS Mean (SE) | 14 (.4) | 12 (.4) |
| LS Mean difference from placebo | | -2 |
| 95% CI of difference | | (-2,-1) |
| Pairwise p-value[1] | | <0.001 |

Source: Clinical Study Report Table 13 and Table 14.2.1.1.3
[1] Analysis based on ANOVA model with effects for age, weight, center, and treatment.

One of Dr. Norton's primary concerns was the lack of a concise analysis plan. Thus, Dr. Norton conducted an analysis based on his understanding of the SAP (including the handling of missing data). Dr. Norton's preferred final model included terms for age, treatment, and center. The results are shown in Table 4 and are nearly identical to those of the Applicant.

**Table 4: Table 2: Dr. Norton's analysis based on the SAP**

| | Placebo (n=153) | Caldolor 800 mg (n=166) |
|---|---|---|
| Total Morphine Use | | |
| LS Mean (SE) | 57 (2) | 49 (2) |
| LS Mean difference from placebo | | -8 |
| 95% CI of difference | | (-13,-3) |
| Pairwise p-value[1] | | <0.001 |

Source: Adapted from Dr. Norton's statistical review, Table 16
[1] Analysis based on ANOVA model with treatment and center as factors.

Dr. Norton conducted several additional analyses applying various transformations to the data. His rationale for performing the analyses was to diminish the multiplicity concern arising from several different analyses. All of the analyses produced significant results. Specifically, he stated the following:

> As noted earlier, the Applicant raised a multiplicity problem in the SAP by describing a host of potential analyses that might have been performed. One way to address this problem is adequately cover the space of possible analyses tried, then that would suggest that the positive finding reported by the Applicant is not an artifact of post-hoc choices.

When evaluating the evidence, I did not focus on the results produced by transforming the data. As previously stated, the Applicant deemed the data to be non-normal and consequently transformed the data. However, the normality assumption underlying an analysis of variance can be relaxed provided the sample size is large and/or the departure from normality is not extreme. In Study CPI-CL-008B, the sample size was not small, I was not convinced that the departures from normality were extreme, and I was concerned that the transformed data would not easily be clinically interpretable; therefore, an analysis of the untransformed data is preferable in my opinion. Although I did not focus on the analyses on the transformed data, the results from those analyses (performed by the Applicant and Dr. Norton) were similar to those from the original data.

## 3 CONCLUSIONS AND RECOMMENDATIONS

Dr. Norton had several concerns with the studies including high rates of protocol deviations and/or violations and the lack of concise pre-specified analyses. The proportion of study participants with protocol deviations and/or violations was of greater concern in Study CPI-CL-008A where the number of deviations was large. However, the deviations and violations were fairly balanced across all treatment arms in both studies and therefore may not have influenced the treatment effect. Additionally, all randomized patients were included in the efficacy analysis despite protocol deviations and/or violations. The large number of protocol deviations could potentially suggest poor study conduct; however, the study conduct should be assessed collectively among the Division of Scientific Investigation, the statistical team, and the clinical team.

A pre-specified statistical analysis is a germane feature of a clinical trial. Pre-specification assists in negating multiplicity concerns arising from post-hoc selection of endpoints and/or analyses. The Applicant pre-specified the analyses, but they were found to be lacking in sufficient detail. I garnered some clarity regarding the analyses from the Applicant's responses to the information requests. However, the requests also demonstrated that the Applicant was negligent in reviewing the CSR for accuracy and providing needed details within the CSR.

While I find the Applicant's submission to be flawed, I am able to glean sufficient information to draw conclusions. I conclude that there is evidence that Caldolor 800 mg decreases the amount of morphine needed following surgery; however, the evidence is derived from a single study. The review team will consider the totality of the data presented in the NDA to weigh the risks and benefits of Caldolor for the proposed indication.

# 4 APPENDIX

## 4.1 CONTENT OF INFORMATION REQUEST AND RESPONSE FOR STUDY CPI-CL-008A

***FDA Request:***

On page 41 of the Clinical Study Report for Study CPI-CL-008A, you state the following:

> The primary endpoint for the study was reduction in the requirement for morphine use in the 24 hours following surgery as measured by total morphine usage compared to placebo. Analysis of variance and covariance procedures were to be used to compare the reduction in the requirement for morphine use in the 24 hours following surgery among the treatment groups. Dunnett's test was to be used as a multiple comparison test to compare active dose groups with the placebo group at an overall alpha level of 0.05. Comparison of morphine use among active doses of IVIb was to be made using an alpha level of 0.10 to declare significance. In the primary model, center was to be introduced as a covariate. Center-by-Treatment interaction was to be examined to evaluate the consistency of results among centers for the primary efficacy endpoint, morphine requirements post-surgery. Type of surgery, weight, gender and other covariates identified through the demographic, background and baseline analysis were to be introduced as secondary covariates **for sensitivity analysis and robustness.**

Based on the aforementioned, it appears that the primary model only included factors for treatment and center, and other covariates were to be introduced for sensitivity analyses and robustness. However in your NDA, the primary model appears to include factors for treatment, center, weight, and age. Explain this apparent discrepancy, and indicate where the results for the primary model are located within your NDA.

***Cumberland Response:***

The final model for CPI-CL-008A includes factors for treatment, center, age group, and weight group. Consistent with the SAP plan, factors for age group and weight group were introduced into the model as secondary covariates. Since this was a center based stratified randomization, stratified within center by weight group and age group and at least one of these additional factors was significant in the model, the decision was made to keep all randomization factors in the final model for CPI-CL-008A. Please note that the inclusion of these factors does not change the inference. The inference from a model with factors for treatment, center, age group, and weight group is the same as the inference from a model with treatment and center. The results of the final model for CPI-CL-008A with factors for treatment, center, age group, and weight group is what was presented in the NDA and can be found in CPI-CL-008A table 14.2.1.1.4.

### 4.2 CONTENT OF INFORMATION REQUEST AND RESPONSE FOR STUDY CPI-CL-008B

***FDA Requests:***

On pages 41- 42 of the Final Clinical Study Report (CSR) for Study CPI-CL-008B, the analysis was described as follows:

> The primary endpoint for the study was reduction in the requirement for morphine use in the 24 hours following surgery as measured by total morphine usage compared to placebo. Analysis of variance and covariance procedures were used to compare the reduction in the requirement for morphine use in the 24 hours following surgery among the treatment groups. Dunnett's test was used as a multiple comparison test to compare active dose groups with the placebo group at an overall alpha level of 0.05. In the primary model, center was introduced as a covariate. Center-by-Treatment interaction was examined to evaluate the consistency of results among centers for the primary efficacy endpoint, morphine requirements postsurgery. Type of surgery, weight, gender and other covariates identified through the demographic, background and baseline analysis were introduced as secondary covariates for sensitivity analysis and robustness.
>
> The assumptions of analysis of variance and covariance were examined. If the primary endpoint was not normally distributed or had very large variance, a log (natural log or log to the base 10) transformation was applied such as log (Dose+1) to stabilize the variance and the analysis would be presented as the primary analysis of efficacy. Other assumptions of the analysis of variance such as homogeneity of variance, independence of errors, and homogeneity of slopes (for covariance analysis) were also examined. Results may have been corroborated using analog non-parametric procedures.

On page 14 of the Statistical Analysis Plan for Study CPI-CL-008B (dated January 10, 2008), the analysis of the primary variable was described as follows:

> Analysis of variance will be implemented with SAS® PROC GLM; tests of significance will be done with Type 3 sums of squares. The initial model will contain factors for center, age group, weight group, treatment group and all two-way and three-way interactions of the four factors. Three-way interactions will be removed from the model if their corresponding p-values are greater than 0.20 and model will be rerun. Two-way interactions will be removed if their corresponding p-values are greater than 0.20 and model will be rerun. As a final step, if there is no interaction between centers and the other factors and the p-value associated with centers is greater than 0.20, then centers may be removed from the model. The test for treatment differences will be done when the simplest model has been obtained. The Type 3 least squares treatment means will be reported.
>
> The residuals of the simplest model will be assessed for normality. If the residuals violate model assumptions, then additional techniques may be applied to investigate the robustness of any conclusions from the model derived using the methodology described above. Among the methods that may be attempted are log and square root transformations; nonparametric methods, such as those suggested by Conover and Inman (1981), may also be attempted.

The Statistical Analysis Plan suggests that a model selection approach was used to derive the final model. However, the analysis outlined in the clinical study report does not suggest the use of a model selection approach. Clarify this apparent discrepancy. If a model selection procedure was used, provide the location within your NDA where the results from the initial model as well as subsequent models are provided. Currently, it is not apparent that the final model used to produce Table 13 in the CSR was the model resulting from the model selection procedure.

***Cumberland Response:***

The language on page 41-42 of the Clinical Study Report (CSR) for study CPI-CL-008B was obtained from the study protocol (dated 08Dec06, pages 32-33) and is not representative of the analyses performed and presented in the CSR. The analyses in the CSR are according to the statistical analysis plan (SAP), dated 10 Jan 2008 which was completed and submitted to the Agency after generation of the protocol.

Amended pages 41-42 of the CSR are included in this response.

Further, a model selection approach was used to derive the final model. Supporting documentation concerning the model selection procedure is included in this response. The model generated from the model selection procedure (consistent with the SAP dated 10Jan08) is the final model used to produce Table 13 in the CSR.

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

/s/
---
Dionne Price
5/17/2009 04:13:13 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

## CLINICAL STUDIES

**NDA/Serial Number:** 22-348/000

**Drug Name:** CALDOLOR (ibuprofen) Injection

**Indication(s):** Management of mild to severe pain, reduction of fever

**Applicant:** Cumberland Pharmaceuticals Inc.

**Date(s):** Letter 12/11/2008, PDUFA 6/11/2009

**Review Priority:** Priority

**Biometrics Division:** Division of Biometrics II

**Statistical Reviewer:** Jonathan D. Norton, Ph.D.

**Concurring Reviewers:** Dionne Price, Ph.D.

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

**Clinical Team:** Christina Fang, M.D.
Ellen Fields, M.D.

**Project Manager:** Kathleen Davies

**Keywords:** clinical studies, Cochran-Mantel-Haenszel, data imputation, intent-to-treat, multiple comparisons, NDA review, protocol violations, 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 ..... 5**

2.1 OVERVIEW ..... 5
2.2 DATA SOURCES ..... 6

**3. STATISTICAL EVALUATION ..... 6**

3.1 EVALUATION OF EFFICACY ..... 6
3.2 EVALUATION OF SAFETY ..... 25

**4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS ..... 26**

4.1 GENDER, RACE AND AGE ..... 26
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS ..... 30

**5. SUMMARY AND CONCLUSIONS ..... 31**

5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE ..... 31
5.2 CONCLUSIONS AND RECOMMENDATIONS ..... 31
5.3 REVIEW OF THE PROPOSED LABEL ..... 32

**APPENDIX I: REFERENCES ..... 36**

**APPENDIX II: INFORMATION REQUESTS ..... 37**

# 1. EXECUTIVE SUMMARY

## 1.1 Conclusions and Recommendations

The Applicant seeks to have an injectable ibuprofen product with the proposed tradename Caldolor approved for the following indications: management of mild to moderate pain in adults, and reduction of fever in adults (b) (4) . The proposed doses are 400 mg (4 mL vial) and 800 mg (8 mL vial). The application is a 505(b)(2) and cites three reference listed drugs: Children's Motrin suspension, Advil Liqui-Gels capsules, and Motrin tablets. Ibuprofen is a non-steroidal ant-inflammatory drug (NSAID) with a long history as an approved active moiety.

After considering the clinical evidence, I find that the 400 mg dose of Caldolor is effective for fever in adults. In regard to the pain indication, I find the clinical studies to be inadequate to reach any conclusion about efficacy. This judgment was reached by considering both the quality of the studies and the strength of the results. It is not meant to rule out a determination from other evidence, e.g., a pharmacokinetic "bridge", that the product is effective for pain. (b) (4)

## 1.2 Brief Overview of Clinical Studies

The Applicant submitted four Phase 3 trials. Study CPI-CL-004 was a double-blind, placebo-controlled study of the safety, efficacy, and pharmacokinetics of Caldolor for fever. The study was conducted at nine centers in the United States, Thailand, and Australia. The 123 adult subjects were randomized into the following treatment arms in roughly equal numbers: placebo, 100 mg ibuprofen, 200 mg ibuprofen, and 400 mg ibuprofen. Patients qualified for the study by having a temperature of at least 38.3° C (101° F). After randomization, subjects were given intravenous investigational drug (ibuprofen or placebo) over a period of 30 minutes every four hours. They were given six doses in total, over a treatment period of 24 hours. Subjects who had a high fever after two hours were considered treatment failures, and the investigators could provide rescue treatment. The primary efficacy endpoint was the proportion of subjects with a temperature less than 38.3° C at four hours.

Study CPI-CL-006 was a double-blind, placebo-controlled trial for the efficacy and safety of Caldolor in febrile adults. Sixty hospitalized patients were enrolled at a single center in Thailand, and assigned in equal numbers to either placebo or 400 mg intravenous ibuprofen. In order to qualify a patient had to have a diagnosis of uncomplicated falciparum malaria and a temperature greater than 100.4° F (38.0° C). The patients were given study medication every six hours for the first three days, then every six hours "as needed" for an additional two days. The protocol allowed for rescue treatment for highly febrile subjects, but they were considered treatment failures. The primary efficacy endpoint was the 24-hour area under the curve (AUC) for temperature.

Study CPI-CL-008A was a double-blind, placebo-controlled trial of intravenous ibuprofen for pain in adults. The study was conducted on 406 patients at 16 sites in the U.S., South Africa, and Australia. Patients were randomized to the following treatment arms in roughly equal numbers:

placebo, ibuprofen 400 mg, and ibuprofen 800 mg. The study enrolled patients who were scheduled for single-site elective surgery and were expected to need post-operative intravenous morphine for more than 24 hours. The first study medication was given at initiation of skin closure, then additional doses were given every six hours, for a total of eight scheduled doses. In addition to study medication, patients had access to morphine either upon request or via patient controlled analgesia. The primary efficacy endpoint was total morphine usage for 24 hours after surgery.

Study CPI-CL-008B was similar in design to study CPI-CL-008A. It was conducted on 335 female subjects at 10 sites in the U.S. These subjects were randomized to the either placebo or ibuprofen 800 mg. In order to be included, subjects had to be scheduled for abdominal hysterectomy surgery with anticipated need for at least 24 hours of post-operative morphine. The first dose of study medication was administered at the initiation of skin closure, and then every six hours. The primary efficacy endpoint was again 24-hour morphine use.

### 1.3 Statistical Issues and Findings

The submitted trials had several features that lead to concern. In Study 6, the values of the primary endpoint submitted in the SAS files differed from the values used in the Applicant's own analyses. The Applicant explained that two different methods were used to compute the primary endpoint from the raw data. The two pain studies had high rates of reported protocol deviations. In Study 8A, the Applicant reports 269 "major protocol deviations" among 406 subjects. (Some subjects had more than one deviation.) In Study 8B, 63 subjects had a "protocol violation", which in this study was more severe than a "protocol deviation". In both pain studies the prospective analysis plans were vague, making it difficult to determine which analytical decisions were made after looking at the data. In addition to the foregoing, there are a number of errors and inconsistencies in the NDA which collectively raise concern.

In order to reach a conclusion about efficacy, I weighed my concerns about the Applicant's trials against the strength of the results. The two fever studies showed large treatment effects that are unlikely to be attributable to unintentional bias arising from the conduct and analysis of the studies. (***Bias*** usually refers to the potential for a statistical procedure to systematically over- or underestimate the quantity that it is intended to estimate. In this case I include the data generating mechanism, i.e., the clinical trials, as a source of bias.) In both studies, a subject receiving the 400 mg dose of ibuprofen was more than twice as likely to become afebrile than a subject receiving placebo treatment. In contrast, the results from the two pain studies were less "robust" in a couple of senses of the word. First of all, the statistical significance of the results depend on the methods of analysis, e.g., whether the data from Study 8A are transformed. Secondly, the size of the analgesic effect in the studies was modest, on the order of a 10-15% reduction in concomitant morphine use. Although this magnitude of effect may be in line with what is expected from an NSAID, it does imply that an equally modest source of unintentional bias could confound the results.

Taking into account the stated concerns about the Applicant's trials and the strength of the results, I find sufficient evidence of efficacy for the indication of fever in adults. Making the same assessment for fever, I find the studies inadequate to reach a conclusion.

# 2. INTRODUCTION

## 2.1 Overview

The Applicant seeks to have an injectable ibuprofen product with the proposed tradename Caldolor approved for the following indications: management of mild to moderate pain in adults, and reduction of fever in adults (b) (4). The proposed doses are 400 mg (4 mL vial) and 800 mg (8 mL vial). The application is a 505(b)(2) and cites three reference listed drugs: Children's Motrin suspension, Advil Liqui-Gels capsules, and Motrin tablets. Ibuprofen is a non-steroidal ant-inflammatory drug (NSAID) with a long history as an approved active moiety. It was first approved by FDA in 1974, and since that time it been marketed by a number of manufacturers in both prescription and over-the-counter products.

Table 1 lists the Phase 3 studies that were submitted by the Applicant. The Clinical Studies section of the proposed label includes (b) (4). I did not review this study, however, for a couple of reasons. First of all, the primary endpoint was mortality. Also, the study was conducted from 14 to 20 years ago, using clinical trial material from a different manufacturer.

**Table 1: Phase III Clinical Trials**

| Study ID and Indication | Study Design | No. Patients (ITT) and Centers |
|---|---|---|
| CPI-CL-004, Fever | Randomized, placebo-controlled study of hospitalized patients<br><br>Arms: Placebo, 100 mg, 200 mg, 400 mg IVIb | 120, 9 centers |
| CPI-CL-006, Fever | Randomized, placebo-controlled study of febrile hospitalized patients with malaria<br><br>Arms: Placebo, 400 mg IVIb | 60, 1 center |
| CPI-CL-008A, Pain | Randomized, placebo-controlled study of surgical patients<br><br>Arms: Placebo, 400 mg, 800 mg IVIb | 406, 16 centers |
| CPI-CL-008B, Pain | Randomized, placebo-controlled study of hysterectomy patients<br><br>Arms: Placebo, 800 mg IVIb | 319, 10 centers |

Note: IVIb = Intravenous ibuprofen, ITT = intent-to-treat set

The Division of Anti-Inflammatory, Analgesic and Ophthalmic Drug Products (DAIOP) held a Type A meeting with the Applicant on 9/29/2004. DAIOP agreed with the Applicant's proposal to stratify the randomization by age and weight, and stated that these factors should be included in the analysis of variance (ANOVA) model. DAIOP also stated that the intent-to-treat (ITT) population should include all randomized patients who received a dose of study medication.

Following a reorganization within FDA, this product fell within the purview of the Division of Anesthesia, Analgesia and Rheumatology Products (DAARP). DAARP held a teleconference with the Applicant on 7/16/2006. In regard to studies CPI-CL-004 and CPI-CL-008, DAARP recommended that the analysis be adjusted for the variables used in stratification. The Applicant confirmed that the interim analysis in study CPI-CL-008 would be blinded and could only result in an increase in sample size.

DAARP also had a pre-NDA meeting with the Applicant on 5/29/2008. At this meeting, DAARP requested both raw and analysis-ready datasets from each study. (b) (4) [redacted]. The Applicant noted that the primary endpoint data from study CPI-CL-008 was not normally distributed, and proposed to include the results using both Box-Cox and log transformations. DAARP requested that they submit the results of all analyses.

DAARP also submitted a number of information requests to the Applicant after NDA submission. Table 25 in Appendix I summarizes those requests which were related to the statistical review, as well the Applicant's response.

### 2.2 Data Sources

The electronic version of this NDA can be found at \\cdsesub1\EVSPROD\NDA022348.

## 3. STATISTICAL EVALUATION

### 3.1 Evaluation of Efficacy

***Fever – CPI-CL-004***

**Study Design and Endpoints**

Study CPI-CL-004, which I will refer to as Study 4, was a double-blind, placebo-controlled study of the safety, efficacy, and pharmacokinetics of Caldolor for fever in adult subjects. The study was conducted at nine centers in the United States, Thailand, and Australia. The 120 subjects in the ITT set were randomized to the following treatment arms: placebo (28 patients), 100 mg ibuprofen (31 patients), 200 mg ibuprofen (30 patients), and 400 mg ibuprofen (31 patients). (There were a total of 123 randomized patients.)

Patients qualified for the study by having a fever with a temperature of at least 38.3° C (101° F). The fever had to have begun within the previous seven days. After randomization, subjects were given intravenous investigational drug (ibuprofen or placebo) over a period of 30 minutes every four hours. They were given six doses in total, over a treatment period of 24 hours. The temperature was to be recorded immediately before the first dose, and then at Hours 1, 2, 4, 6, 8, 12, 18, and 24 after that dose. Patients who had a temperature of 39.4 ° C (103° F) or greater after a minimum of two hours were considered treatment failures, and the investigators were allowed (but not required) to give them rescue treatment. The total duration of the study was 120 hours.

The primary efficacy endpoint was the proportion of subjects with a temperature less than 38.3° C at four hours. Pre-specified secondary endpoints included the proportion of treatment failures, the time to afebrility (temperature < 38.3° C), the area under the temperature curve at 4 and 24 hours, and change from baseline in temperature at 4 and 24 hours. The ITT population was defined as all patients with a baseline assessment and at least one post-baseline temperature reading.

**Patient Disposition, Demographic and Baseline Characteristics**

Of the 123 subjects who were randomized, 120 received study treatment. Three subjects were discontinued prior to treatment: one due to "poor venous access" and two because their temperature dropped below 101° F before receiving their first dose. All 120 treated subjects were included in the ITT set. Of the 120, 108 (90%) received all six doses of study medication. Eleven subjects were withdrawn from the study for reasons shown in Table 2. The twelfth subject (of the 120) missed one dose "due to nursing error" and was not withdrawn. Table 3 shows baseline characteristics of the subjects.

**Table 2: Applicant's Reasons for Discontinuation** (Source: Table 6 in Clinical Study Report)

| Reason CTM Discontinued | Rand Number | Treatment | # Doses Received |
|---|---|---|---|
| Adverse Event | 4065<br>4072<br>4057 | 100 mg IVIb<br>100 mg IVIb<br>Placebo | 5<br>1<br>4 |
| Excluded Concomitant Medication Required | 4066<br>4080 | Placebo<br>100 mg IVIb | 1<br>3 |
| Treatment Failure | 4060<br>5051<br>6005 | 400 mg IVIb<br>Placebo<br>Placebo | 3<br>1<br>1 |
| Inadequate IV Access | 9001 | 100 mg IVIb | 5 |
| Bleeding Precaution | 1001<br>5055 | 400 mg IVIb<br>Placebo | 1<br>5 |

Notes: Rand Number = Subject ID, CTM = Clinical Trial Material.

**Table 3: Demographics and Baseline Values**

| Variable | | Treatment | | | | | |
|---|---|---|---|---|---|---|---|
| | | 100 mg Ibu | 200 mg Ibu | 400 mg Ibu | Placebo | All Ibu | All Subjects |
| | N | 31 | 30 | 31 | 28 | 92 | 120 |
| Age | Mean | 40.1 | 34.5 | 39.2 | 37.0 | 38.0 | 37.8 |
| | Median | 35.0 | 28.0 | 36.0 | 31.0 | 34.5 | 33.5 |
| | SD | 19.0 | 15.0 | 17.2 | 19.1 | 17.1 | 17.5 |
| | Min | 18 | 18 | 17 | 18 | 17 | 17 |
| | Max | 83 | 68 | 73 | 89 | 83 | 89 |
| Height (cm) | Mean | 170.3 | 170.1 | 168.2 | 168.8 | 169.6 | 169.4 |
| | Median | 169.0 | 168.0 | 166.4 | 165.0 | 168.0 | 168.0 |
| | SD | 10.3 | 10.5 | 10.1 | 11.8 | 10.2 | 10.6 |
| | Min | 152.0 | 147.0 | 148.0 | 153.0 | 147.0 | 147.0 |
| | Max | 188.0 | 190.5 | 188.0 | 193.0 | 190.5 | 193.0 |
| Weight (kg) | Mean | 77.9 | 80.3 | 72.5 | 78.4 | 76.9 | 77.2 |
| | Median | 69.5 | 69.8 | 67.1 | 68.2 | 69.3 | 69.2 |
| | SD | 32.6 | 29.2 | 23.4 | 26.6 | 28.5 | 28.0 |
| | Min | 42.0 | 41.0 | 42.5 | 46.0 | 41.0 | 41.0 |
| | Max | 191.4 | 150.0 | 129.0 | 129.9 | 191.4 | 191.4 |
| Base Temp. (Celsius) | Mean | 39.1 | 39.1 | 39.2 | 38.9 | 39.1 | 39.1 |
| | SD | 0.6 | 0.7 | 0.7 | 0.5 | 0.7 | 0.6 |
| Gender | Female | 8 | 8 | 9 | 7 | 25 | 32 |
| | Male | 23 | 22 | 22 | 21 | 67 | 88 |
| Race | Asian | 10 | 10 | 10 | 10 | 30 | 40 |
| | Black | 5 | 4 | 1 | 3 | 10 | 13 |
| | White | 13 | 15 | 16 | 14 | 44 | 58 |
| | Other | 3 | 1 | 4 | 1 | 8 | 9 |

Note: SD = standard deviation.

**Protocol Deviations**

Thirteen patients were excluded from the Applicant's efficacy evaluable population due to major protocol deviations. In nine of these cases, I was able to determine that the major deviation occurred after the primary endpoint was assessed. The other four cases were as follows: one patient in the 100 mg arm received corticosteroids, one in the placebo arm received Xigris, and two patients in the placebo arm received rescue therapy due to treatment failure. The Applicant notes that not all "deviations" actually violated the protocol.

In addition to these deviations which the Applicant deemed major, I found additional deviations which appeared to have the potential to affect the efficacy analysis. Two patients, one in the 100 mg arm and one in the 200 mg arm, received corticosteroids during the treatment period.

The Applicant reported that four subjects were misclassified in regard to severity of illness at time of randomization, and hence assigned to the wrong strata. For randomization, 55 subjects were deemed critically ill and 65 were deemed non-critically ill. After "review of the criteria for critically ill", 53 were considered critically ill and 67 non-critically ill.

## Statistical Methods

Randomization to treatment was stratified by center and whether the subject was critically ill. The statistical analysis plan (SAP) defines ***critically ill*** it as "in the hospital requiring mechanical ventilation for respiratory failure, pressor support for hypotension, or both". A block randomization scheme was used, with a block size of 4.

The Applicant describes the primary efficacy analysis as follows on page 18 of the SAP, "For each of the [doses of ibuprofen], comparison against placebo will be evaluated using the [Cochran-Mantel-Haenszel] procedure adjusted for center by using exact Chi-squared test statistic, to be calculated using SAS Proc Freq." The SAP also states (p. 16) "Dunnett's test will be used to account for multiple comparisons between the active doses and the placebo." There are a number of methods that might be called "Dunnett's test", and they are used to adjust for multiple comparisons when comparing multiple treatments to a control. The usual setting for a "Dunnett" method is a continuous outcome, and it appears that the Applicant performed multiplicity corrections for the continuous secondary outcomes (as opposed to the dichotomous primary). According to the SAP, missing temperatures were to be imputed using last observation carried forward.

For my analysis, I conducted stratified Cochran-Mantel-Haenszel (CMH) tests and obtained both asymptotic and exact p-values. Asymptotic p-values are mathematical approximations based on a large sample size; exact p-values are valid even for small samples. What constitutes a "large" enough sample varies, so the exact p-values should be more reliable in this case. Unless otherwise specified, my analyses are based on the strata assignments used for randomization.

## Results and Conclusions

Table 4 shows the results for the primary endpoint, afebrility (temperature less than 38.3° C) at four hours. The 200 and 400 mg doses of ibuprofen are superior to placebo, with marginal significance, depending on statistical method, for the 100 mg dose. The analysis is based on the original strata for randomization, in order to ensure a valid statistical test. Table 5 shows the results broken down by corrected stratum, i.e., the severity subgroup to which the subject actually belonged. While the non-critically-ill subgroup had a high placebo response rate, there was an apparent effect of 400 mg ibuprofen treatment in both strata.

**Table 4: Analysis of Primary Endpoint, ITT Set**

| Treatment | N | Num. Afebrile (%) | Asymptotic p | Exact p | Applicant p |
|---|---|---|---|---|---|
| Placebo | 28 | 9 (32%) | N.A. | N.A. | N.A. |
| 100 mg | 31 | 19 (61%) | .0328 | .0581 | .0264 |
| 200 mg | 30 | 21 (70%) | .0079 | .0147 | .0043 |
| 400 mg | 31 | 24 (77%) | .0006 | .0011 | .0005 |

**Table 5: Analysis of Primary Endpoint by Corrected Stratum, ITT Set**

| Stratum | Treatment | N | Num. Afebrile (%) | Exact p |
|---|---|---|---|---|
| Critically Ill | Placebo | 13 | 1 (8%) | N.A. |
| | 100 mg | 14 | 9 (64%) | .0064 |
| | 200 mg | 12 | 6 (50%) | .0533 |
| | 400 mg | 14 | 8 (57%) | .0175 |
| Non-Critically Ill | Placebo | 15 | 8 (53%) | N.A. |
| | 100 mg | 17 | 10 (59%) | 1.0000 |
| | 200 mg | 18 | 15 (83%) | .1264 |
| | 400 mg | 17 | 16 (94%) | .0220 |

The prior analyses were based on last-observation-carried-forward (LOCF) imputation. There was only one subject (4068), who did not have a temperature recorded at hour 4. A conservative approach (i.e., favoring the null hypothesis of no treatment effect) is to classify this subject as febrile. This subject was in the 400 mg arm, and when this conservative imputation is used the 400 mg treatment is still superior to placebo with exact p = .0024. Another subject, 6005, was classified as a treatment failure at Hour 2.5 but was later classified as afebrile at four hours. The study report does not indicate whether this subject received rescue treatment. This subject was in the placebo arm, however, so the effect of classifying him as afebrile is conservative.

In the foregoing analyses I did not perform any corrections for the multiple comparisons between active doses and placebo. It does not appear that the Applicant performed any multiplicity correction on the primary endpoint, either. Applying a simple Bonferroni correction however, i.e., tripling the p-values, the 400 mg dose is still found to be superior to placebo (exact p = .0033).

*Fever – CPI-CL-006*

**Study Design and Endpoints**

Study CPI-CL-006, which I will refer to as Study 6, was a double-blind, placebo-controlled trial for the efficacy and safety of Caldolor in febrile adults. Sixty hospitalized patients were enrolled at a single center in Thailand. They were assigned in equal numbers to either placebo or 400 mg intravenous ibuprofen.

In order to qualify a patient had to have a diagnosis of uncomplicated falciparum malaria and a temperature greater than 100.4° F (38.0° C). The patients were given study medication every six hours for the first three days, then every six hours "as needed" for an additional two days. Patients were considered treatment failures if they had a temperature greater than 106.0° F (41.1° C) after a minimum of two hours after the first dose of study medication, or if it was greater than 103.0° F (39.4° C) two or more hours after a subsequent dose. These patients were then to be given rescue treatment. The patients were observed for a total of 21 days.

The primary efficacy endpoint was the area under the curve (AUC) above 98.6 F (37.0° C) for 24 hours. Secondary endpoints included the AUC for 4 and 72 hours, the AUC for the period 24-72 hours, the proportion of treatment failures, and the parasite clearance time. The intent-to-treat (ITT) set was defined as all patients with a baseline assessment and at least one post-baseline temperature reading.

**Patient Disposition, Demographic and Baseline Characteristics**

A total of 60 patients enrolled in the study and none withdrew in the first 24 hours of the study. No patients were excluded from the intent-to-treat set.

Demographic and baseline characteristics of the patients are shown in Table 6.

**Table 6: Demographics and Baseline Temperature**

| | Ibuprofen<br>Mean (SD) | Placebo<br>Mean (SD) |
|---|---|---|
| Sex – Female | 6 | 6 |
| Sex – Male | 24 | 24 |
| Race – Asian | 30 | 30 |
| Age | 32.4 (9.4) | 27.8 (9.4) |
| Weight in kg: | 55.4 (5.5) | 53.2 (8.6) |
| Baseline Temperature in °C | 38.7 (.3) | 38.8 (.5) |

**Protocol Deviations**

The Applicant reported four protocol deviations, two of which involved multiple subjects. The data for the study were supposed to be double-entered, but the data were single-entered for all subjects. The Applicant states that the data were double-checked, however. Also, for 38 subjects the clinical laboratory testing was performed at screening, but not on days one through five as specified in the protocol. There was one subject in the placebo group who received rescue treatment despite not having a high enough temperature to qualify as a treatment failure, and another subject in the placebo group who was a treatment failure but did not receive rescue treatment.

**Statistical Methods**

A block randomization scheme was used with a block size of six and no stratification. The primary efficacy analysis on the AUC endpoint was an analysis of covariance with treatment and baseline temperature as factors. The baseline temperature was defined as "the mean of the qualifying temperatures and the temperature immediately before dosing." The treatment-by-baseline-temperature interaction was to be investigated. The AUC was to be calculated using the linear trapezoidal rule. Missing values were to be imputed using linear interpolation for intermediate points and LOCF otherwise.

The analysis data sets that the Applicant provided in the NDA submission showed different values for the primary endpoint than the data listings in the Clinical Study Report (CSR). Note that the primary endpoint is ***derived***, so this discrepancy does not imply a problem with the raw data. In a letter of 3/20/2009, the Applicant explained as follows:

> Study CPI-CL-006 was conducted utilizing Microsoft Excel as the database for raw data collection and analysis. The AUC presented in the clinical study report and in Table 2.5 is calculated by using the time of initial study drug administration as time 0 temperature collection time and the actual temperature collection time for the remaining values.
>
> The SAS dataset was then converted to transport version 5 and [Study Data Tabulation Model] formatting. This SAS dataset does not provide the results presented in the clinical study report because the AUC calculation utilizes the actual temperature collection times for all timepoints. This calculation was utilized for analyses included in the integrated summary of efficacy and in the dataset provided for study CPI-CL-006.
>
> The analyses in the CSR for CPI-CL-006 resulted in a mean difference between IVIB and placebo of 8.95 °C x hour. The calculation of AUC was later revised to use the actual time of the first temperature measurement to be consistent with the definition outlined in the SAP. This analysis resulted in a mean difference between IVIB and placebo of 8.9 °C x hour. The different calculation methods yielded similar results (< 1% difference in mean values) and provided consistent statistical conclusions.

The analysis presented herein was based on the SAS data sets provided in the NDA. For comparison with Study 4, I included as an additional endpoint the primary efficacy

analysis from that study (afebrility). My analysis of this endpoint used a Pearson chi-square test.

**Results and Conclusions**

Table 7 shows the results for the primary and other efficacy endpoints. I did not examine the effects of the Applicant's imputation method. Since the subjects were hospitalized, however, it is unlikely that the results were strongly affected by missing temperature readings.

**Table 7: Efficacy Endpoints, ITT Set**

| Endpoint | Statistic | Placebo (N=30) | Ibuprofen 400 mg (N=30) |
|---|---|---|---|
| AUC – 24 hours | Mean (SD) | 16.94 (11.69) | 7.99 (7.96) |
| | p-value vs. placebo | -- | .002 |
| AUC – 4 hours | Mean (SD) | 5.30 (2.94) | 2.56 (1.52) |
| | p-value vs. placebo | -- | < .0001 |
| Afebrile (< 38.3° C) at 4 Hours | Proportion (%) | 12/30 (40%) | 29/30 (97%) |
| | p-value vs. placebo | -- | .0006 |

***Pain – CPI-CL-008, Part A***

**Study Design and Endpoints**

Study CPI-CL-008A, or Study 8A, was a double-blind, placebo-controlled trial of intravenous ibuprofen for pain in adults. The study was conducted on 406 patients at 16 sites in the U.S., South Africa, and Australia. Patients were randomized to the following treatment arms: placebo (134 patients), ibuprofen 400 mg (134), and ibuprofen 800 mg (138).

The study enrolled patients who were scheduled for single-site elective surgery and were expected to need post-operative intravenous morphine for more than 24 hours. The first study medication was given at initiation of skin closure. Additional doses were given every six hours, for a total of eight scheduled doses. In addition to study medication, patients had access to morphine either upon request or via patient controlled analgesia. No other treatments were allowed for pain. Patients could continue to receive study medication for as long as five days, but after 24 hours they could be discontinued when they were able tolerate oral medication.

The primary efficacy endpoint was total morphine usage for 24 hours after surgery. Morphine use was to be assessed every three hours. Secondary efficacy endpoints included pain intensity on an 11-point visual analog scale (VAS) at several time points, assessed both at rest and with movement. Another secondary endpoint was treatment failure, which was defined as a patient "requiring" additional pain medications. The intent-to-treat analysis set was all patients who were randomized and received at least one dose of study medication.

## Patient Disposition, Demographic and Baseline Characteristics

The Applicant made the useful distinction between ending study medication and withdrawing from the study. Of 406 patients randomized into the study, all discontinued treatment medication prior to day 5. Table 8 shows the reasons for discontinuation and distribution of the number of doses. As the table shows, for most patients the reason for discontinuation was ability to tolerate oral treatment. It can also be seen that in every arm at least 90% of patients received the fourth dose of medication, which preceded the 24-hour efficacy evaluation.
Table 9, in contrast, shows the reason for and times of study withdrawal. Both tables were provided by the Applicant.

**Table 8: Applicant's Disposition of Patients, Study Medication**
(Source: Table 14.1.4 in Clinical Study Report)

| | Placebo (N=134) | 400 mg (N=134) | 800 mg (N=138) |
|---|---|---|---|
| Did Patient Receive Full Course of Study Medication? | | | |
| Yes | 0 | 0 | 0 |
| No | 134 (100%) | 134 (100%) | 138 (100%) |
| Reason Patient Did not Receive Full Course of Study Medication | | | |
| Resolution Of Pain | 9 ( 7%) | 10 ( 7%) | 12 ( 9%) |
| Ability to tolerate pain medication by mouth | 91 ( 68%) | 89 ( 66%) | 101 ( 73%) |
| No intravenous access | 3 ( 2%) | 1 ( <1%) | 2 ( 1%) |
| Physician request for safety reasons | 1 ( <1%) | 0 | 1 ( <1%) |
| Withdrawal at patient's request | 4 ( 3%) | 5 ( 4%) | 0 |
| Discontinued secondary to adverse event | 9 ( 7%) | 8 ( 6%) | 6 ( 4%) |
| Treatment Failure | 11 ( 8%) | 7 ( 5%) | 4 ( 3%) |
| Other | 6 ( 4%) | 14 ( 10%) | 12 ( 9%) |
| Patients at Each Scheduled Post-Surgery Dosing Time | | | |
| Study Hour 0 | 134 (100%) | 134 (100%) | 138 (100%) |
| Study Hour 6 | 128 ( 96%) | 123 ( 92%) | 130 ( 94%) |
| Study Hour 12 | 128 ( 96%) | 120 ( 90%) | 130 ( 94%) |
| Study Hour 18 | 126 ( 94%) | 120 ( 90%) | 129 ( 93%) |
| Study Hour 24 | 98 ( 73%) | 105 ( 78%) | 108 ( 78%) |
| Study Hour 30 | 45 ( 34%) | 57 ( 43%) | 59 ( 43%) |
| Study Hour 36 | 37 ( 28%) | 50 ( 37%) | 47 ( 34%) |
| Study Hour 42 | 31 ( 23%) | 38 ( 28%) | 41 ( 30%) |
| Patients who required at least one dose of CTM more than 42 hours after surgery | 1 ( <1%) | 4 ( 3%) | 12 ( 9%) |

**Table 9: Applicant's Disposition of Patients, Study Participation**
(Source: Table 14.1.2.1 in Clinical Study Report)

| | Placebo (N=134) | 400 mg IVIb (N=134) | 800 mg IVIb (N=138) |
|---|---|---|---|
| Patients Who | | | |
| Completed Study | 115 ( 86%) | 115 ( 86%) | 127 ( 92%) |
| Prematurely Discontinued Study | 19 ( 14%) | 19 ( 14%) | 11 ( 8%) |
| Withdrawn by | | | |
| Self | 7 ( 5%) | 8 ( 6%) | 0 |
| Physician | 4 ( 3%) | 1 ( <1%) | 2 ( 1%) |
| Principal Investigator | 8 ( 6%) | 10 ( 7%) | 9 ( 7%) |
| Reason for Withdrawal from Study | | | |
| Adverse Event | 6 ( 4%) | 9 ( 7%) | 6 ( 4%) |
| Intercurrent Illness | 1 ( <1%) | 0 | 0 |
| Non-Compliance with Protocol | 0 | 2 ( 1%) | 1 ( <1%) |
| Other | 12 ( 9%) | 8 ( 6%) | 4 ( 3%) |
| Completed all aspects of study without deviation from the Protocol? | | | |
| Yes | 29 ( 22%) | 17 ( 13%) | 27 ( 20%) |
| No | 105 ( 78%) | 117 ( 87%) | 111 ( 80%) |
| Patients at Study Visit [1] | | | |
| Screening/Baseline (-72 to 0 hours) | 134 (100%) | 134 (100%) | 138 (100%) |
| Day 1 (>0-24 hours) | 134 (100%) | 134 (100%) | 138 (100%) |
| Day 2 (>24-48 hours) | 127 ( 95%) | 119 ( 89%) | 131 ( 95%) |
| Day 3 (>48-72 hours) | 49 ( 37%) | 45 ( 34%) | 54 ( 39%) |
| Day 4 (>72-96 hours) | 26 ( 19%) | 19 ( 14%) | 29 ( 21%) |
| Day 5 (>96-120 hours) | 15 ( 11%) | 14 ( 10%) | 19 ( 14%) |
| Post-Treatment Period Day 1 (>120-144 hours) | 11 ( 8%) | 11 ( 8%) | 14 ( 10%) |
| Post-Treatment Period Day 2 (>144-168 hours) or Discharge | 128 ( 96%) | 126 ( 94%) | 130 ( 94%) |
| Day 14 Telephone Contact | 130 ( 97%) | 132 ( 99%) | 138 (100%) |

Table 10, which was produced by the Applicant, shows the demographics and stratification factors for the study. I was able to reproduce the table from the data listings provided by the Applicant, with a minor exception: based on the data provided, the labels on the "Hispanic" and "Other" rows should be switched.

**Table 10: Applicant's Subject Demographics and Stratification Factors**
(Source: Table 17 in Clinical Study Report)

| | | Placebo n=134 | 400 mg IVIb n=134 | 800 mg IVIb n=138 | Total n=406 |
|---|---|---|---|---|---|
| Age | | | | | |
| | Mean (SD) | 45 (10.9) | 45 (12.5) | 46 (11.9) | 45 (11.8) |
| Age Category | | | | | |
| | ≤45 years old | 77 (57%) | 80 (60%) | 76 (55%) | 233 (57%) |
| | >45 years old | 57 (43%) | 54 (40%) | 62 (45%) | 173 (43%) |
| Gender | | | | | |
| | Male | 25 (19%) | 35 (26%) | 27 (20%) | 87 (21%) |
| | Female | 109 (81%) | 99 (74%) | 111 (80%) | 319 (79%) |
| Race | | | | | |
| | Caucasian | 118 (88%) | 112 (84%) | 118 (86%) | 348 (86%) |
| | Black | 14 (10%) | 16 (12%) | 15 (11%) | 45 (11%) |
| | Hispanic | 0 | 1 (<1%) | 0 | 1 (<1%) |
| | Asian | 2 (1%) | 2 (1%) | 3 (2%) | 7 (2%) |
| | Other | 0 | 3 (2%) | 2 (1%) | 5 (1%) |
| Height (cm) | | | | | |
| | Mean (SD) | 167.3 (9.00) | 167.7 (9.22) | 166.7 (9.34) | 167.2 (9.17) |
| Weight (kg) | | | | | |
| | Mean (SD) | 83.4 ( 18.15) | 83.0 (18.16) | 84.9 (20.83) | 83.8 (19.08) |
| Weight Category | | | | | |
| | ≤75 kg | 51 (38%) | 50 (37%) | 52 (38%) | 153 (38%) |
| | >75 kg | 83 (62%) | 84 (63%) | 86 (62%) | 253 (62%) |
| Age & Weight Stratum | | | | | |
| | Age ≤45 yrs; Weight ≤75 kgs | 29 (22%) | 31 (23%) | 31 (22%) | 91 (22%) |
| | Age ≤45 yrs; Weight >75 kgs | 48 (36%) | 49 (37%) | 45 (33%) | 142 (35%) |
| | Age >45 yrs; Weight ≤75 kgs | 22 (16%) | 19 (14%) | 21 (15%) | 62 (15%) |
| | Age >45 yrs; Weight >75 kgs | 35 (26%) | 35 (26%) | 41 (30%) | 111 (27%) |
| | | | | | |
| Geographical Region | | | | | |
| | USA | 42 (31%) | 41 (31%) | 43 (31%) | 126 (31%) |
| | Australia | 41 (31%) | 41 (31%) | 43 (31%) | 125 (31%) |
| | Republic of South Africa | 51 (38%) | 52 (39%) | 52 (38%) | 155 (38%) |

## Protocol Deviations

The Applicant reported a total of 269 major protocol deviations, as seen in Table 11.

**Table 11: Applicant's Major Protocol Deviations** (Source: Table 13 in Clinical Study Report)

| | Treatment Group Assignment | | | Total |
|---|---|---|---|---|
| | 400 mg | 800 mg | Placebo mg | |
| Consenting | 3 | 6 | 5 | 14 |
| CTM administration | 48 | 46 | 49 | 143 |
| Exclusion criteria | 6 | 5 | 6 | 17 |
| Received restricted concomitant medications | 28 | 37 | 23 | 88 |
| Randomization | 3 | 1 | 3 | 7 |
| Total | 88 | 95 | 86 | 269 |

The Applicant reports that the deviations in clinical trial material (CTM) administration were "primarily" related to timing. A dose was considered within compliance if it was administered within 60 minutes of the scheduled time. There were 64 patients who received at least one of the first four doses (relevant to the primary endpoint) out-of-window. The effect of restricted concomitant medications on the efficacy analysis is evaluated in Results and Conclusions.

**Statistical Methods**

Randomization was stratified by dichotomous age (<= 45, > 45) and weight (<= 75 kg, > 75 kg). The original plan, as represented in the study protocol and the first four amendments, was for central randomization. Due to "logistical issues", the Applicant decided that randomization would be by-site for sites outside of Australia, with the four sites in Australia continuing to use central randomization.

The Applicant describes the primary efficacy analysis somewhat differently, and vaguely, in different documents. The SAP for Studies 8A and 8B (p. 20) states:

> Analysis of variance and covariance procedures will be used to compare the reduction in the requirement for morphine use in the 24 hours following surgery among the treatment groups... In the primary model, center will be introduced as a covariate. Center-by-Treatment interaction will be examined to evaluate the consistency of results among centers for the primary efficacy endpoint, morphine requirements post-surgery. Type of surgery, gender and other covariates identified through the demographic, background and baseline analysis will be introduced as secondary covariates for sensitivity analysis and robustness. The primary efficacy variable total morphine requirements to treat pain in the 24 hours following surgery will also be analyzed using an analysis of covariance model with adjustment age and weight as continous [sic] variables as well as stratification variables (age: <=45 years; >45 years; and weight: <=70Kgs; >70Kgs).
>
> The assumptions of analysis of variance and covariance will be examined. If the primary endpoint is not normally distributed or has very large variance, a log (natural log or log to the base 10) transformation will be applied such as log (Dose+1) to stabilize the variance and the analysis will be presented as the primary analysis of efficacy. Other assumptions

> of the analysis of variance such as homogeniety of variance, independence of errors, and homogeneity of slopes (for covariance analysis) will also be examined. Results may be corroborated using analog non-parametric procedures.

The quoted paragraphs allow for several possibilities, but I infer that the primary analysis was intended to be an analysis of variance on the untransformed data with treatment and center as factors. A log transformation is to be used for non-normal data or "very large variance".

The Applicant reported a number of analyses in the study report, and it is difficult to determine which is intended to be definitive. Based in its prominence in the Applicant's report, I take Table 12 ("Table 21" in their report) to show their primary analysis. As the notes state, the Applicant fit an ANOVA model with four factors to *rank*-transformed data. Despite this apparent shift by the Applicant, I consider the primary efficacy analysis to be the one described in the SAP. Under the standard frequentist theory of statistical testing, an analysis chosen post-hoc (which I must presume is the case) may not yield a valid p-value. I did reproduce the results reported in Table 12, however.

The SAP states that when morphine use is missing for one or two time points, the mean of the previous and subsequent non-missing observations will be imputed for those time points. When more than two consecutive time points are missing, the placebo group mean for each time point is to be imputed. The placebo group mean is also to be used when "there is not an available timepoint prior to or after the missing data." The SAP further stipulates that if a patient drops out in the first 24 hours then the placebo group mean will be imputed at missing time points. As a sensitivity analysis, the Applicant also proposed a "more conservative" method in which the placebo mean would be used for all missing time points. The same imputation methods were described in the clinical study report, but it is not clear whether they were actually used. (See next section for further discussion.)

In regard to multiplicity, the SAP states, "Dunnett's test will be used to account for multiple comparisons between the active doses and the placebo." Although the Applicant does not give a reference, I was able to determine from the submitted SAS code that they used the Dennett-Hsu method (Hsu, 1992). Dunnett-Hsu is a version of Dunnett's test which addresses a technical problem in computing multiplicity-adjusted probabilities and confidence intervals in the general setting of the test, i.e., comparing multiple treatment groups to a control group.

The original sample size for the study was 225 patients. A planned interim analysis was performed after "approximately" 168 patients (75%) had completed the study. It was performed by an independent statistician who was not given the treatment assignments. The purpose of the analysis was to re-estimate the sample size based on the standard deviation. The independent statistician recommended increasing the sample to 360 subjects (120 per group).

**Results and Conclusions**

Table 12 shows the results of what I take to be the Applicant's primary efficacy analysis. I was able to reproduce the values in the table. The p-values were adjusted for multiplicity using Dunnett-Hsu.

**Table 12: Applicant's Post-Hoc Efficacy Analysis, ITT Set**
(Source: Table 21 in Clinical Study Report)

| | Intent-to-Treat | | |
|---|---|---|---|
| | Placebo (n=134) | 400mg IVIb (n=134) | 800mg IVIb (n=138) |
| **Morphine Requirement (mg)** | | | |
| Mean (SD) | 48.9 (27.7) | 46.3 (29.4) | 43.8 (33.7) |
| Median | 45.3 | 44.0 | 35.5 |
| Min, Max | 0.0, 144.0 | 3.0, 198.25 | 0.0, 221.3 |
| **Transformed Morphine Requirement (mg)** | | | |
| LSMeans (SE)[1] | 223.0 (13.8) | 208.5 (13.6) | 190.6 (13.1) |
| **Pairwise Comparison to Placebo** | | | |
| p-value[2] | --- | 0.458 | 0.030 |

[1] Data are transformed using the rank transformation. LS means are adjusted for age group, weight group, randomization center, and treatment group.
[2]The analysis is based on a linear 4-way ANOVA model with fixed effects for age group, weight group, randomization center, and treatment group.

I was not able to find the results of the planned primary efficacy analysis in the Clinical Study Report, but I performed it with the results shown in Table 13. Since the data are in fact skewed (i.e., have an asymmetrical probability distribution), the results are given for both untransformed and log-transformed (log(X+ 1)) data. Regardless of the transformation, the 400 mg dose is not statistically different from placebo. The 800 mg is statistically better than placebo only after the data are log-transformed. Based on the least-squares means for the untransformed data, the 800 mg dose reduced morphine use by 11% compare to placebo.

**Table 13: Planned Efficacy Analysis, ITT Set, Applicant's Imputation**

| Transform | Statistic | Placebo | 400 mg Ibu | 800 mg Ibu |
|---|---|---|---|---|
| None | LS Mean (SE) | 50.1 (3.6) | 47.7 (3.6) | 44.8 (3.4) |
| | Adj. p-value | --- | .72 | .23 |
| Log(x + 1) | LS Mean (SE) | 3.78 (.09) | 3.73 (.09) | 3.53 (.09) |
| | Adj. p-value | --- | .84 | .014 |

Note: P-value is for comparison of dose to placebo, using Dunnett-Hsu adjustment. LS = least squares; SE = standard error.

The analysis data set provided by the Applicant includes two different versions of the 24-hour morphine use variable that are purported to be based on different imputation methods. The imputation methods are not clearly documented. Based on an examination of the SAS code the Applicant submitted, it does not appear that the original imputation plan was followed. The two methods used result in quite similar values for the primary endpoint, having a Pearson correlation of .99986. In attempting to reproduce the Applicant's results from their primary endpoint, I used the version which they labeled "Imp. Meth. 1".

Since it appears that the Applicant did not use the planned imputation, I also performed the "conservative" analysis described in the SAP, imputing the mean from the placebo group at each missing time point. For the untransformed data, this analysis results in a finding that neither the 400 mg dose nor the 800 mg dose are superior to placebo, with adjusted p-values of .98 and .66, respectively. After a log transform, the 800 mg dose might be deemed marginally superior to placebo (adj. p = .068) but the 400 mg dose is not (p = .99).

Concerned about the use of concomitant medications, we requested that the Applicant send a list of subjects who received restricted analgesics and NSAIDs during the first 24 hours of treatment with study medication. The listing that they sent in response did not precisely convey this information. (See Table 25 in Appendix I.) Table 16.2.9.2 in the clinical study report, however, does list the subjects who received restricted analgesics and NSAIDs at some point during the study, along with the day (but not the *time*) it was received. This served as the basis for a list of 33 subjects who were *potentially* exposed in the relevant time period, and they were predominantly in the active treatment groups. To be precise, I selected subjects who received the restricted medications of interest either on the calendar day when double-blind treatment began or the next day. In order to adjust for the effect of (suspected) restricted analgesia use, I added a "dummy variable" (1 if restricted analgesia use, 0 otherwise) to the model for log morphine use. This variable was not statistically significant (p = .14), and in fact the observed trend was for subjects on the list to use *more* morphine. Although this analysis was based on incomplete data, it does provide some reassurance that concomitant analgesia did not inflate the apparent effect of ibuprofen.

The secondary pain endpoints are particularly important in this study, as they can be used to eliminate the possibility that subjects who used less morphine were simply tolerating a higher level of pain. The SAP did not specify a single endpoint to summarize the VAS, but the 24 hour area AUC is a natural analog to the 24 hour morphine use. The Applicant reports that the both the 400 mg dose and the 800 mg dose significantly lowered the "with movement" pain AUC compared to placebo, with p = .033 and p = .004, respectively. My analysis of their derived variables yielded similar results.

***Pain – CPI-CL-008, Part B***

**Study Design and Endpoints**

Study CPI-CL-008B was similar in design in Study 8A. It was conducted on 335 female subjects at 10 sites in the U.S. These subjects were randomized to the either placebo (161) or ibuprofen 800 mg (174). In order to be included, subjects had to be scheduled for abdominal hysterectomy surgery with anticipated need for at least 24 hours of post-operative morphine. The first dose of study medication was administered at the initiation of skin closure, then every six hours for 48 hours (eight doses), and then continued as needed for a total of five days. The only rescue medication allowed was morphine. After 24 hours, study medication could be discontinued due to resolution of pain, "no IV access", or discharge from hospital.

The primary efficacy endpoint was total morphine use in the first 24 hours after initiation of study medication. Secondary efficacy endpoints included pain intensity on a VAS at several time points, treatment failure (i.e., requirement of additional rescue medication besides morphine), time to first analgesia for breakthrough pain, and time to first analgesia for nocturnal awakening due to pain. The ITT population included all patients who were randomized and received at least one dose of study medication. The efficacy evaluable population is defined, somewhat confusingly, as, "Patients included in the ITT who have no major protocol violations with regard to inclusion or exclusion criteria or study conduct and have all primary efficacy assessments... specifically, patients included in the ITT who received the first four doses of CTM administered within 60 minutes of the scheduled administration time." (See page 5 of SAP).

**Patient Disposition, Demographic and Baseline Characteristics**

On page 46 of the CSR, the Applicant states, "A total of 319 patients were randomized into two treatment groups at 10 clinical centers: 800 mg IVIb 166; Placebo 153. All 319 patients received at least one dose of CTM." In fact, the submitted data show that ***335*** patients had assigned treatments and were hence, by definition, randomized. There were 319 patients who were ***randomized and received at least one dose of CTM***, hence were in the ITT set. All 319 patients discontinued study medication before day 5, mainly due to discontinuation of intravenous access. However, all but 37 patients (12%) completed the study. The breakdown of withdrawals by treatment was 22 (14%) in the placebo arm and 15 (9%) in the active arm.

Table 14 summarizes the baseline characteristics of the subjects. It was provided by the Applicant, but I was able to reproduce the contents.

**Protocol Deviations**

The Applicant made a distinction in this study between protocol violations and protocol deviations, the former being more severe. Protocol violations included the following: inclusion criteria not met, exclusion criterion met, receiving restricted medication within 24 hours of starting the study medication, or CTM administered outside of a 60 minute window. The Applicant reported protocol violations for 36 patients in the placebo arm and 32 patients in the ibuprofen arm. More specifically, 63 subjects took restricted medications within 24 hours of the beginning of study treatment (before or after): 33 in the placebo arm, 30 in the active arm.

**Table 14: Applicant's Subject Demographics and Stratification Factors**
(Source: Table 10 in CSR.)

| | | Placebo n=153 | IVIb n=166 | Total n=319 |
|---|---|---|---|---|
| Age | | | | |
| | Mean (SD) | 42 (7.2) | 42 (7.0) | 42 (7.1) |
| Age Category | | | | |
| | ≤45 years old | 104 (68%) | 110 (66%) | 214 (67%) |
| | 45-70 years old | 49 (32%) | 56 (34%) | 105 (33%) |
| Gender | | | | |
| | Male | 0 | 0 | 0 |
| | Female | 153 (100%) | 166 (100%) | 319 (100%) |
| Race | | | | |
| | Caucasian | 52 (34%) | 71 (43%) | 123 (39%) |
| | Black | 91 (59%) | 84 (51%) | 175 (55%) |
| | Hispanic | 8 (5%) | 11 (7%) | 19 (6%) |
| | Asian | 0 | 0 | 0 |
| | Other | 2 (1%) | 0 | 2 (<1%) |
| Height (cm) | | | | |
| | Mean (SD) | 163.0 (7.56) | 163.8 (8.22) | 163.4 (7.91) |
| Weight (kg) | | | | |
| | Mean (SD) | 83.0 (19.07) | 83.9 (18.67) | 83.5 (18.83) |
| Weight Category | | | | |
| | ≤75 kg | 57 (37%) | 61 (37%) | 118 (37%) |
| | >75 kg | 96 (63%) | 105 (63%) | 201 (63%) |
| Age & Weight Stratum | | | | |
| | Age ≤45 yrs; Weight ≤75 kgs | 41 (27%) | 43 (26%) | 84 (26%) |
| | Age ≤45 yrs; Weight >75 kgs | 63 (41%) | 67 (40%) | 130 (41%) |
| | Age 45-70 yrs; Weight ≤75 kgs | 16 (10%) | 18 (11%) | 34 (11%) |
| | Age 45-70 yrs; Weight >75 kgs | 33 (22%) | 38 (23%) | 71 (22%) |

**Statistical Methods**

In addition to the SAP described in the Statistical Methods section for Study 8A, there was an amendment specific to Study 8B. This amendment describes the primary efficacy analysis as follows:

> Analysis of variance will be implemented with SAS® PROC GLM; tests of significance will be done with Type 3 sums of squares. The initial model will contain factors for center, age group, weight group, treatment group and all two-way and three-way interactions of the four factors. Three-way interactions will be removed from the model if their corresponding p-values are greater than 0.20 and model will be rerun. Two-way interactions will be removed if their corresponding p-values are greater than 0.20 and model will be rerun. As a final step, if there is no interaction between centers and the other factors and the p-value associated with centers is greater than 0.20, then centers may be removed from the model. The test for treatment differences will be done when the

> simplest model has been obtained. The Type 3 least squares treatment means will be reported.
>
> The residuals of the simplest model will be assessed for normality. If the residuals violate model assumptions, then additional techniques may be applied to investigate the robustness of any conclusions from the model derived using the methodology described above. Among the methods that may be attempted are log and square root transformations; non-parametric methods, such as those suggested by Conover and Inman (1981), may also be attempted.

The analysis plan described above is vague. The algorithm for stepwise covariate selection described in the first paragraph does not specify whether interaction terms will be dropped if they exceed the .2 boundary for inclusion at a later step of the algorithm, nor whether they will be dropped if the lower order terms they are associated with are dropped. The term "simplest model" is not defined; arguably, the simplest linear model only includes an intercept term. The second paragraph is equally nebulous, as the Applicant leaves itself of option of addressing model violations with a variety of methods or of simply ignoring the violations. By introducing the potential to make choices after looking at the data, the Applicant has raised a clear multiplicity problem. As a consequence, positive results can only be trusted if they are significant under a variety of possible analyses that fall within the scope of the SAP.

The case report form for Study 8B collects morphine usage data differently than the form for Study 8A. In addition to recording usage since the last assessment, the form keeps a running total of morphine usage. Also, it does not list fixed time points for collection as in Study 8A. For these reasons, the amended SAP proposes a different imputation method for the primary endpoint of 24-hour morphine usage. The new imputation method is described in a confusing manner (p.13 of SAP), however, and does not seem to reflect the actual changes on the form. For this reason, I did not attempt to reproduce their computation of the primary endpoint, but rather made my own computations from the longitudinal data using the "conservative" approach of imputing the mean from the placebo arm for each patient whose 24 hour morphine use was missing.

**Results and Conclusions**

As with Study 8A, the Applicant reported a number of analyses of the primary endpoint, and it is unclear which analysis is intended to be taken as definitive.

Table 15 shows the analyses that are featured most prominently in the report. I was able to reproduce the descriptive statistics for the untransformed data, but could not match their statistics for the transformed data in the efficacy evaluable population. I was also able to confirm the bound on the p-values (< .001). The Applicant's analysis differs from the SAP in using a Box-Cox transformation, rather than the proposed log and square-root transformations. (Admittedly, the log and square root are special cases of the Box-Cox transformation.) Also, it appears that they did not use the planned stepwise method to choose factors for inclusion in the model.

**Table 15: Applicant's Summary of the Primary Endpoint**
(Source: Table 13 in Clinical Study Report)

| | | Intent-to-Treat | | Efficacy Evaluable | |
|---|---|---|---|---|---|
| | | Placebo (n=153) | IVIb (n=166) | Placebo (n=137) | IVIb (n=150) |
| Morphine Requirement (mg) | | | | | |
| | Mean (SD) | 55.9 (20.6) | 47.3 (25.6) | 54.3 (20.5) | 45.5 (24.6) |
| | Median | 54.0 | 43.5 | 53.5 | 42.0 |
| | Min, Max | 14.5, 114.0 | 4.0, 143.3 | 14.5, 114.0 | 4.0, 143.3 |
| Transformed Morphine Requirement (mg) | | | | | |
| | Mean (SD) | 13.4 (3.1) | 11.8 (4.0) | 12.1 (2.7) | 10.7 (3.5) |
| | LSMeans (SE)[1] | 13.6 (0.4) | 12.1 (0.4) | 12.2 (0.4) | 10.8 (0.3) |
| | Median | 13.4 | 11.8 | 12.3 | 10.7 |
| | Min, Max | 5.8, 20.6 | 2.0, 23.5 | 5.5, 18.7 | 2.0, 21.2 |
| Pairwise Comparison to Placebo | | | | | |
| | p-value[2] | | <0.001 | | <0.001 |

[1] Data are transformed using the Box Cox transformation. LS means are adjusted for age group, weight group, randomization center, and treatment group.
[2]The analysis is based on a linear 4-way ANOVA model with fixed effects for age group, weight group, randomization center, and treatment group.

In addition to the results shown in Table 15, the Applicant fit log-and rank-transformed versions of the primary endpoint to the four-factor model. (See Table 27 on pages 68-69 of clinical study report). They report that ibuprofen was superior to placebo ($p < .05$) in both cases. I was able to reproduce these results.

The results of my own analyses are summarized in Table 16. Although the analysis described in the amended SAP for this study is vague, I first attempted to follow it as literally as possible in order to approach the ideal of a pre-specified analysis. Following their instructions for dropping terms from the model, I obtained a model with the following factors: Arm, Center, Age, Weight, Arm-by-Center-by-Weight interaction, and Arm-by-Center-by-Age interaction. Fitting this model to their morphine endpoint, I obtained a p-value of .064 for the effect of treatment Arm.

Applying my own judgment, I find the model selected by the planned stepwise procedure to contain extraneous terms. The weight term is far from significant ($p = .25$), as is the Arm-by-Center-by-Age interaction ($p = .33$). I also chose to drop Arm-by-Center-by-Weight interaction on the basis that the main effect of Weight is no longer in the model. (The Weight effect is also insignificant when only the main effects are included in the model.) The resulting reduced model contains only the main effects of Arm, Center, and Age, and the p-value for treatment Arm is .0009. Returning to the planned analysis, the residuals are now "assessed for normality", and they show evidence of skewness (in other words, the histogram is not symmetrical like the classic "bell curve"). There are many possible ways to address this issue.

**Table 16: Analyses of Primary Endpoint**

| Imputation | Transform | Model | p-value for Arm |
|---|---|---|---|
| Applicant Method 1 | None | Planned stepwise | .064 |
| Applicant Method 1 | None | Reduced stepwise | .0009 |
| Applicant Method 1 | Log(x + 1) | Reduced stepwise | < .0001 |
| Applicant Method 1 | Square root | Reduced stepwise | .0001 |
| Applicant Method 1 | Rank | Reduced stepwise | < .0001 |
| Conservative | Log(x + 1) | Reduced stepwise | <. .0001 |

As noted earlier, the Applicant raised a multiplicity problem in the SAP by describing a host of potential analyses that ***might have been performed.*** One way to address this problem is adequately cover the space of possible analyses described by the SAP. If a significant (or nearly so) result were found on every analysis tried, then that would suggest that the positive finding reported by the Applicant is not an artifact of post-hoc choices.

As remedies for non-normality, the SAP proposes transforming the endpoint using log and square root functions. Using either of these methods, the p-value for the treatment effect is .0001 or lower. The SAP also refers to the paper by Conover and Iman (1981) as a source of potential analysis methods for non-normal data. The key point of the paper is that traditional parametric methods can be made nonparametric by substituting ranks for the original observations. The obvious application of the principle in this case is to perform an ANOVA analysis on ranks, which yields a p-value <.0001.

Additional analyses were suggested by ambiguities in the data. As noted earlier, I was not able to reproduce the Applicant's imputation method. Using a putatively conservative imputation for missing values (impute placebo means) and fitting the reduced model to the log morphine use, the resulting p-value for treatment is again highly significant (p <. 0001). As a final sensitivity analysis, I created a "dummy variable" (as in Study 8A) for subjects who were listed as having violated protocol by having restricted medications within 24 hours. When this variable was added to the model, the treatment effect remained highly significant (p < .0001).

In regard to pain, the Applicant reported a significant reduction in the 24-hour AUC for the "with movement" VAS score with active treatment (p = .009). Using their derived endpoint, I replicated this finding of a reduction in pain. As with Study 8A, it should be noted that the AUC was not a pre-specified endpoint.

### 3.2 Evaluation of Safety

The safety profile of this product was reviewed by Christina Fang, M.D.

# 4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

## 4.1 Gender, Race and Age

The 60 patients in Study 6 were all non-elderly people of Asian race, and 80% were male. Due to this homogeneity, I did not perform any subgroup analyses for this study.

Tables 17 and 18 summarize the primary endpoints of Studies 4 and 8A by gender. Due to the overall sample size and the gender imbalance in these studies, I did not perform any tests. All subjects in Study 8B were women.

**Table 17: Study 4, Afebrility by Gender**

| Gender | Arm | N Obs | Perc |
|---|---|---|---|
| Female | 100mg IVIb | 8 | 38% |
| | 200mg IVIb | 8 | 50% |
| | 400mg IVIb | 9 | 67% |
| | Placebo | 7 | 43% |
| Male | 100mg IVIb | 23 | 70% |
| | 200mg IVIb | 22 | 77% |
| | 400mg IVIb | 22 | 82% |
| | Placebo | 21 | 29% |

**Table 18: Study 8A, Morphine Use by Gender**

| Gender | Arm | N Obs | Mean | Std Dev |
|---|---|---|---|---|
| Female | 400mg IVIb | 99 | 44.64 | 27.56 |
| | 800mg IVIb | 111 | 40.72 | 33.13 |
| | Placebo | 109 | 47.85 | 25.91 |
| Male | 400mg IVIb | 35 | 50.83 | 34.21 |
| | 800mg IVIb | 27 | 56.57 | 33.55 |
| | Placebo | 25 | 53.24 | 34.83 |

Tables 19-21 show the primary endpoints by race. Due to the sizes of the overall sample and the subgroups, Studies 4 and 8A did not allow for meaningful inferential analysis. In Study 8B, however, there are sufficient numbers to compare Black and Caucasian patients. Adding race and

an interaction effect to the reduced model for the log-transformed data, it is seen that black patients used significantly less morphine than white patients (p = .0001) and this effect did not interact with treatment (p = .76). These findings are illustrated in Figure 1.

**Table 19: Study 4, Afebrility by Race**

| Race | Arm | N Obs | Perc |
|---|---|---|---|
| Black or African American | 100mg IVIb | 5 | 60% |
| | 200mg IVIb | 4 | 75% |
| | 400mg IVIb | 1 | 100% |
| | Placebo | 3 | 67% |
| Caucasian or White | 100mg IVIb | 13 | 62% |
| | 200mg IVIb | 15 | 73% |
| | 400mg IVIb | 16 | 69% |
| | Placebo | 14 | 14% |
| Other | 100mg IVIb | 13 | 62% |
| | 200mg IVIb | 11 | 64% |
| | 400mg IVIb | 14 | 86% |
| | Placebo | 11 | 45% |

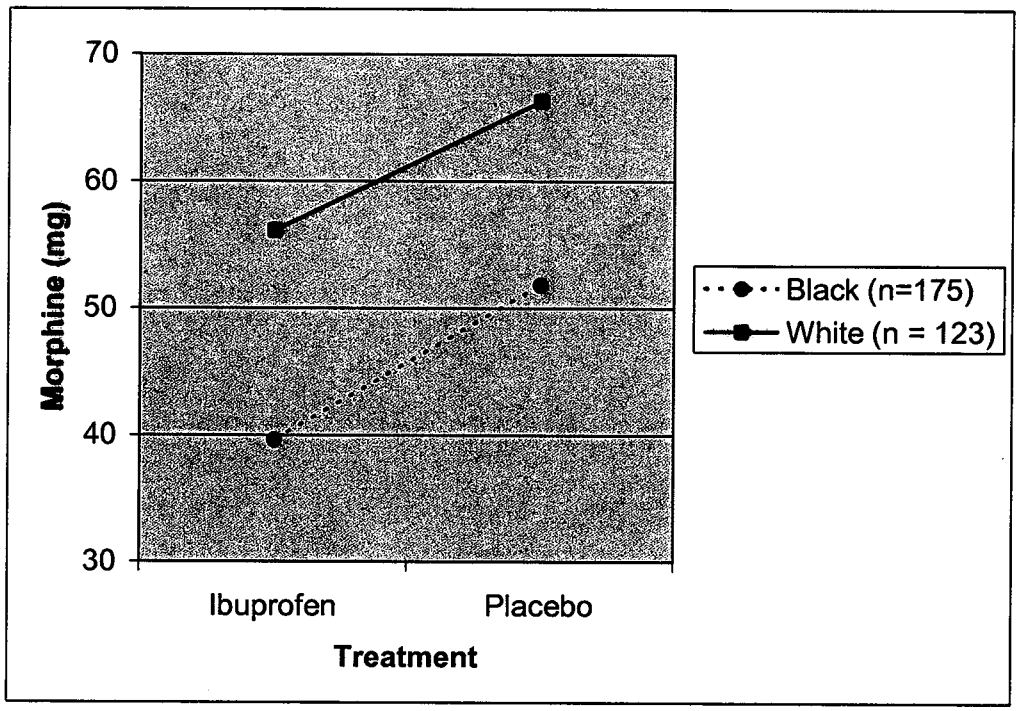
**Table 20: Study 8A, Morphine Use by Race**

| Race | Arm | N Obs | Mean | Std Dev |
|---|---|---|---|---|
| Asian | 400mg IVIb | 2 | 32.5 | 26.2 |
| | 800mg IVIb | 3 | 23.0 | 26.3 |
| | Placebo | 2 | 79.6 | 37.3 |
| Black or African American | 400mg IVIb | 16 | 45.8 | 42.9 |
| | 800mg IVIb | 15 | 34.7 | 22.2 |
| | Placebo | 14 | 42.0 | 18.5 |
| Caucasian or White | 400mg IVIb | 112 | 45.8 | 27.2 |
| | 800mg IVIb | 118 | 45.1 | 34.9 |
| | Placebo | 118 | 49.1 | 28.3 |
| Other | 400mg IVIb | 4 | 66.7 | 29.9 |
| | 800mg IVIb | 2 | 66.3 | 35.2 |

**Table 21: Study 8B, Morphine Use by Race**

| Race | Arm | N Obs | Mean | Std Dev |
|---|---|---|---|---|
| Black | 800mg IVIb | 85 | 39.6 | 20.9 |
| | Placebo | 92 | 51.8 | 18.6 |
| Caucasian | 800mg IVIb | 74 | 56.1 | 27.9 |
| | Placebo | 55 | 66.3 | 20.1 |
| Hispanic | 800mg IVIb | 11 | 48.7 | 25.2 |
| | Placebo | 8 | 33.6 | 14.9 |
| Other | Placebo | 2 | 67.4 | 0.7 |

**Figure 1: Study 8B, Morphine Use by Race**




Tables 22-24 show the primary endpoints by age group (<= 45 year, > 45 years). In Study 8A, there was a main effect of age (p = .0002) but no interaction with treatment (p = .79). Similarly, Study 8B showed an effect of age (p = .019) but no interaction (p = .93). Both studies showed younger patients using more morphine.

**Table 22: Study 4, Afebrility by Age**

| Age | Arm | N Obs | Perc |
|---|---|---|---|
| 45 and Under | 100mg IVIb | 19 | 58% |
| | 200mg IVIb | 20 | 70% |
| | 400mg IVIb | 19 | 84% |
| | Placebo | 22 | 36% |
| Over 45 | 100mg IVIb | 12 | 67% |
| | 200mg IVIb | 10 | 70% |
| | 400mg IVIb | 12 | 67% |
| | Placebo | 6 | 17% |

**Table 23: Study 8A, Morphine Use by Age**

| Age | Arm | N Obs | Mean | Std Dev |
|---|---|---|---|---|
| <=45 yrs | 400mg IVIb | 80 | 51.4 | 31.5 |
| | 800mg IVIb | 76 | 45.1 | 30.8 |
| | Placebo | 77 | 52.3 | 29.7 |
| > 45 yrs | 400mg IVIb | 54 | 38.7 | 24.5 |
| | 800mg IVIb | 62 | 42.3 | 37.2 |
| | Placebo | 57 | 44.2 | 24.3 |

**Table 24: Study 8B, Morphine Use by Age**

| Age | Arm | N Obs | Mean | Std Dev |
|---|---|---|---|---|
| <=45 yrs | 800mg IVIb | 115 | 49.7 | 25.9 |
| | Placebo | 109 | 59.2 | 21.2 |
| > 45 yrs | 800mg IVIb | 59 | 42.4 | 24.4 |
| | Placebo | 52 | 49.1 | 17.7 |

### 4.2 Other Special/Subgroup Populations

In Study 4, the primary endpoint was analyzed by severity as shown in Table 5. No other subgroup analyses were performed.

# 5. SUMMARY AND CONCLUSIONS

## 5.1 Statistical Issues and Collective Evidence

The submitted trials had several features that lead to concern. In Study 6, the values of the primary endpoint submitted in the SAS files differed from the values used in the Applicant's own analyses. The Applicant explained that two different methods were used to compute the primary endpoint from the raw data. The two pain studies had high rates of reported protocol deviations. In Study 8A, the Applicant reports 269 "major protocol deviations" among 406 subjects. (Some subjects had more than one deviation.) In Study 8B, 63 subjects had a "protocol violation", which in this study was more severe than a "protocol deviation". In both pain studies the prospective analysis plans were vague, making it difficult to determine which analytical decisions were made after looking at the data. In addition to the foregoing, there are a number of errors and inconsistencies in the NDA which may individually be deemed minor but collectively raise concern. For example, in both Studies 4 and 8B the Applicant gave multiple definitions for the efficacy-evaluable population. Also, the Clinical Study Report for Study 8B misstates the number of subjects randomized into the study.

In order to reach a conclusion about efficacy, I weighed my concerns about the Applicant's trials against the strength of the results. The two fever studies showed large treatment effects that are unlikely to be attributable to unintentional bias arising from the conduct and analysis of the studies. (***Bias*** usually refers to the potential for a statistical procedure to systematically over- or underestimate the quantity that it is intended to estimate. In this case I include the data generating mechanism, i.e., the clinical trials, as a source of bias.) In both studies, a subject receiving the 400 mg dose of ibuprofen was more than twice as likely to become afebrile than a subject receiving placebo treatment. In contrast, the results from the two pain studies were less "robust" in a couple of senses of the word. First of all, the statistical significance of the results depend on the methods of analysis, e.g., whether the data from Study 8A are transformed. The Applicant aggravated this problem by not thoroughly specifying the analysis in advance. Secondly, the size of the analgesic effect in the studies was modest, on the order of a 10-15% reduction in concomitant morphine use. Although this magnitude of effect may be in line with what is expected from an NSAID, it does imply that an equally modest source of unintentional bias could confound the results.

Taking into account the stated concerns about the Applicant's trials and the strength of the results, I find sufficient evidence of efficacy for the indication of fever in adults. Making the same assessment for fever, I find the studies inadequate to reach a conclusion.

## 5.2 Conclusions and Recommendations

After considering the clinical evidence, I find that the 400 mg dose of Caldolor is effective for fever in adults. In regard to the pain indication, I find the clinical studies to be inadequate to reach any conclusion about efficacy. This judgment was reached by considering both the quality of the studies and the strength of the results. It is not meant to rule out a determination from other evidence, e.g., a pharmacokinetic "bridge", that the product is effective for pain. (b) (4)

■

## 5.3 Review of the Proposed Label

***14.1 Analgesia (Pain)***

*The effect of Caldolor on post-operative pain was studied in two multi-center randomized, double-blind studies.*

(b) (4) *In a* (b) (4) *study, 406 patients (87 men, 319 women) who had undergone an elective abdominal or orthopedic surgery were randomized to receive Caldolor 400 mg, Caldolor 800 mg, or placebo administered every 6 hours (started intra-operatively), and morphine, as needed.* (b) (4)

The statistical analysis plan states that the ITT population is the primary analysis population. Therefore the degree of morphine sparing should be calculated from this population, which showed a 22% reduction in median morphine use. Also, it may be more appropriate to use the mean, which is consistent with the primary analysis.

I recommend removal of (b) (4) from the clinical studies section.

(b) (4)

The AUC endpoint was not pre-specified, and should be excluded, barring a strong clinical argument to the contrary. In regard to the VAS at different timepoints, I did not find a significant difference at hours 36, 42, 45, and 48. If this information is deemed clinically useful and included, then the comparisons should end at hour 24.

(b) (4)

These were not pre-specified endpoints, and should be excluded.

(b) (4) *In a* (b) (4) *study, 319 female patients who had undergone an elective abdominal hysterectomy were randomized to receive Caldolor 800 mg or placebo administered every 6 hours (started intra-operatively), and morphine, as needed.*

As noted earlier, 335 patients were randomized. The 319 patients referred to were randomized ***and treated.***

(b) (4)

(b) (4)

As with the first study, the primary analysis population was ITT, which showed a 19% reduction in median morphine use.

(b) (4)

.

As with the first study, the AUC endpoint was not pre-specified and should be excluded. The second part of the sentence(b) (4) is accurate, based on the Applicant's imputation.

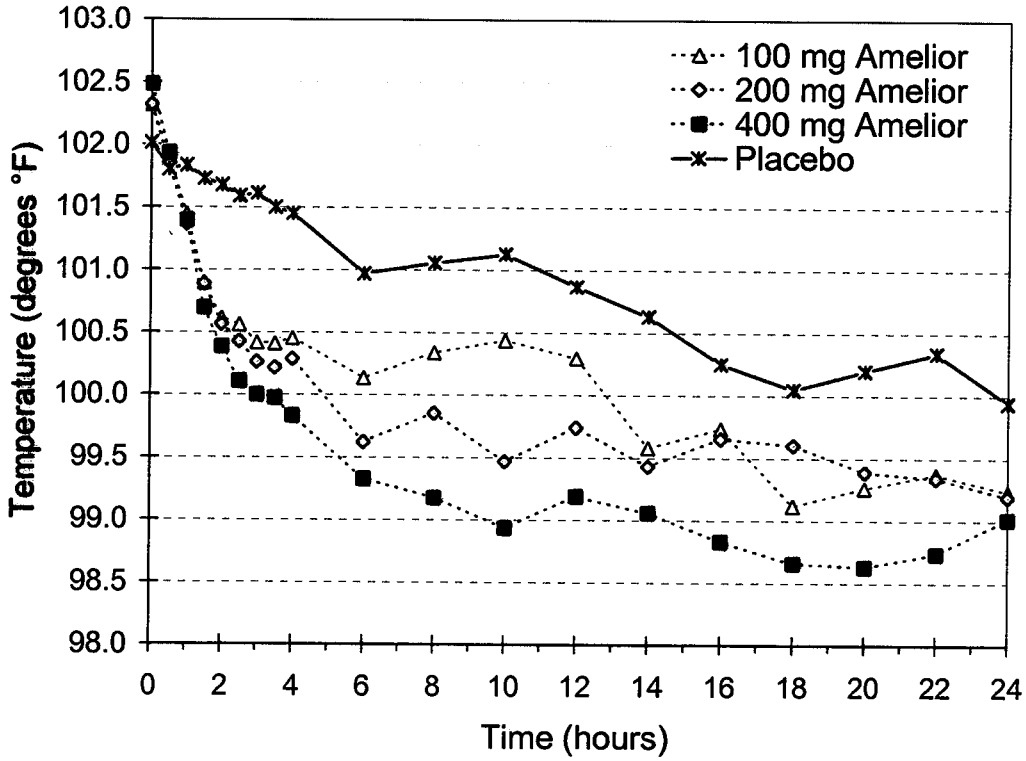
(b) (4)

This endpoint is proposed in the SAP, but in the context of reducing opiate-related side effects. Barring a strong clinical argument to the contrary, it should be excluded.

### *14.2 Antipyretic (Fever)*

*The effect of Caldolor on fever was studied in two randomized, double-blind studies.* (b) (4) *In a multi-center study comparing Caldolor to placebo, 120 patients (88 men, 32 women)*(b) (4) *who had temperatures of 101°F or greater were randomized to Caldolor at 400 mg, 200 mg, and 100 mg or placebo, administered every 4 hours for 24 hours. All Caldolor dose levels resulted in a statistically greater number of patients with a reduced temperature (<101°F) after 4 hours, compared to placebo (placebo 32%* (b) (4) (b) (4) *65%*(b) (4) (b) (4) *73%*(b) (4) *and* (b) (4) *77%* (b) (4) *The dose response is shown in the figure below.*

My own results for the primary endpoint were as follows: placebo, 9/28; 100 mg, 19/31; 200 mg, 21/30; 400 mg, 24/31. (The discrepancy may be due to my using the Celsius-based cutoff for fever.) I agree that all three active doses were superior to placebo, (b) (4)
(b) (4)




The statistics on this figure were not pre-specified in the statistical analysis plan. Unless it provides vital clinical information, it should not be included. I did not verify the accuracy.

(b) (4)

As my Table 5 shows, the highest response rate among critically ill patients was actually in the 100 mg arm. There is a similar statement in the Dosage and Administration section which is also unsupported by the submitted studies.

(b) (4) *In a single center study comparing Caldolor to placebo, 60 patients (48 men, 12 women) with uncomplicated P. falciparum malaria having temperatures ≥100.4°F were randomized to Caldolor 400 mg or placebo, administered every 6 hours for 72 hours of treatment.*

According to the clinical study report (p. 6), qualifying patients had "fevers ***greater than*** 38.0° C (100.4° F)" (emphasis added).

(b) (4) *reduction of fever within the first 24 hours of treatment, measured as the area above the temperature 98.6°F (37.0°C) vs. time curve* (b) (4) (b) (4)

(b) (4)

(b) (4)

(b) (4)

(b) (4)

(b) (4)

## APPENDIX I: References

Conover, WJ and Iman, RL (1981), Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics, The American Statistician, 35:6895-6900.

Hsu, J. C. (1992), The Factor Analytic Approach to Simultaneous Inference in the General Linear Model, Journal of Computational and Graphical Statistics, 1:151 - 168.

# APPENDIX II: Information Requests

**Table 25: Selected Information Requests**

| Request | Response (Date) |
|---|---|
| Analysis datasets and statistical analysis plans (SAP) for fever studies. | Analysis datasets provided. SAP provided for study 4, but no prospective SAP was created for study 6. (1/13/2009) |
| Code used for analysis in all four pivotal studies, as well as code used to make analysis data set from tabulation data. | Some SAS code provided. For studies 4 and 6, however, the Applicant's analysis was partly done in Microsoft Excel. The analysis data for the four studies were not derived from the tabulation data, but from "raw" data sets which were not submitted. (1/30/2009) |
| Describe randomization for study 6, and give block sizes for other studies. | Satisfactory response (2/20/2009) |
| Information on subjects randomized to incorrect stratum. | Satisfactory response (3/11/2009) |
| Explain two different set of values for primary endpoint of study 6. | Explanation of different methods for calculating endpoint. (3/20/2009) |
| List all subjects who received restricted analgesics during first 24 hours of treatment with study medication. | Provided list of all subjects who received restricted study medication, regardless of timing or type of medication. (3/20/2009) |
| Provide case report forms for certain subjects who were excluded from the ITT population for study 8B. | Satisfactory response (4/8/2009) |

**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
5/8/2009 01:30:29 PM
BIOMETRICS

Dionne Price
5/8/2009 04:42:49 PM
BIOMETRICS
Please see the secondary review for my perspective on the application