CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

022544Orig1s000

STATISTICAL REVIEW(S)

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	NDA 22-544
Drug Name:	Gralise (gabapentin) extended release tablet
Indication(s):	Management of postherpetic neuralgia (PHN)
Applicant:	Abbott Products, Inc.
Date(s):	Submitted: March 30, 2010
	PDUFA: January 30, 2011
Review Priority:	Standard
Biometrics Division:	Division of Biometrics II
Statistical Reviewer:	Yongman Kim, Ph.D.
Concurring Reviewers:	Dionne Price, Ph.D.
Medical Division:	Division of Anesthesia and Analgesia Products
Clinical Team:	Timothy Jiang, M.D.
	Ellen Fields, M.D.
Project Manager:	Allison Meyer

Keywords: Clinical studies, missing data, sensitivity analyses

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

Study 81-0062 was conducted to evaluate the safety and efficacy of gabapentin extended release tablets in patients with postherpetic neuralgia. The study demonstrated a statistically significant difference between gabapentin extended release (G-ER) and placebo in terms of the change in average daily pain scores from baseline to end of treatment. The efficacy was further supported by a cumulative evaluation of the proportion of patients achieving various degrees of improvement.

1.2 Brief Overview of Clinical Studies

Study 81-0062 was an 11-week, double-blind, placebo-controlled, international, multi-center trial investigating the safety and efficacy of once daily G-ER in patients with postherpetic neuralgia (PHN). In the study, 452 patients were randomized to G-ER (n 221) or placebo (n 231). The primary efficacy variable was the change in average daily pain score from baseline to Week 11. Secondary efficacy measures included Patient Global Impression of Change (PGIC), Clinical Global Impression of Change (CGIC), and average daily sleep interference (ADSI) score. The primary analysis used an analysis of covariance (ANCOVA) model with the baseline observation carried forward (BOCF) strategy to impute missing data due to dropout.

1.3 Statistical Issues and Findings

There were no statistical issues. During the Type C meeting (December 19, 2005), concerns regarding the use of a last observation carried forward (LOCF) imputation strategy were conveyed. Consequently, the Applicant used a baseline observation carried forward (BOCF) strategy to impute missing pain scores due to dropout. The primary analysis with BOCF showed a statistically significant difference between G-ER and placebo. My continuous responder analysis using a van der Waerden test also showed a statistically significant separation between responder curves. During the End-of Phase 2 meeting (March 2, 2006), the Division requested clarification regarding incorporation of rescue use into the analysis. The Applicant did not incorporate rescue use in the analysis. However, my analysis treating patients who took rescue medications as treatment failures produced results consistent with the primary analysis. Subgroup analyses by country (US, Russia, or Argentina) suggested that the overall treatment effect was mainly driven by the US.

In order to adjust for multiple secondary endpoints PGIC, CGIC, and ADSI, a sequential testing procedure was employed. Based on the sequential testing procedure, significance was not achieved for these secondary endpoints.

2. INTRODUCTION

2.1 Overview

2.1.1 Drug class and regulatory history

Gabapentin Extended Release (G-ER) tablets 300 mg and 600 mg are an extended release formulation containing gabapentin intended for once-daily administration of 1800 mg for the management of Postherpetic Neuralgia (PHN) in adults. The NDA was submitted under 505(b)(2) of the Food, Drug and Cosmetic Act. To support the application, Abbott is relying on the previous findings of safety and efficacy of Neurontin® which was approved for PHN indication in 2002.

At the Type C meeting on December 19, 2005, concerns regarding the use of a LOCF imputation strategy were conveyed and Abbott proposed a baseline observation carried forward methodology. Also the Division encouraged a continuous responder analysis in which dropouts are considered non-responders.

At the End of Phase 2 meeting on April 6, 2006, it was agreed that a single adequate and well-controlled positive trial would be sufficient to support an indication of treatment of pain due to PHN. At the meeting, the Division asked for clarification on how the use of rescue would be incorporated into the statistical analysis.

Study 81-0062 was submitted as an adequate and well controlled trial.

2.1.2 Proposed Indication

The proposed indication is for the treatment of pain due to postherpetic neuralgia (PHN).

2.2 Data Sources

NDA 22-544 was submitted on March 30, 2010. Data are located in the electronic document room (EDR) of the Center for Drug Evaluation and Research. The electronic SAS data sets were also provided in the EDR using the following path:

\\CDSESUB1\EVSPROD\NDA022544\0000

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy of Study 81-0062

3.1.1 Study Design and Endpoints

Study 81-0062 was an 11-week (1 week baseline period and 10 week efficacy treatment period), double-blind, placebo-controlled, international, multi-center trial investigating the safety and efficacy of once daily G-ER in PHN patients. In the study, 452 eligible patients were randomized in a 1:1 ratio to G-ER (n 221) or placebo (n 231) at 89 centers, 57 in the United States, 24 in Russia, and 8 in Argentina. Randomization was stratified by center, and patients were required to have an average daily pain score of at least 4 to be randomized.

The primary efficacy endpoint was the change in average daily pain score from baseline week to the final week of the efficacy treatment period, evaluated from the daily pain diary. The endpoint was measured on the 11-point numerical rating scale, ranging from 0 (no pain) to 10 (worst possible pain).

The secondary endpoints (b) (4) were:

- Proportion of patients who were categorized as "very much" or "much improved" in Patient Global Impression of Change (PGIC)
- Proportion of patients who were categorized as "very much" or "much improved" in Clinical Global Impression of Change (CGIC)
- Mean change in average daily sleep interference (ADSI) score from baseline week to the final week of efficacy treatment period, evaluated from the daily sleep diary assessed on an 11-point numerical rating scale, ranging from 0 (pain does not interfere with sleep) to 10 (pain completely interferes with sleep).

3.1.2 Disposition and Demographics

Approximately 17% of the patients discontinued before the end of study (Table 1). However, approximately the same proportion of patients discontinued in both the G-ER and placebo groups. Sixteen percent of G-ER patients discontinued while 17% of placebo patients discontinued. As expected, the majority of the G-ER dropouts were due to adverse events, and the majority of placebo dropouts were due to lack of efficacy. Eight percent of G-ER patients discontinued due to adverse events and 6% of placebo patients discontinued due to lack of efficacy. Two subjects were excluded from the intent-to-treat (ITT) population which included all randomized patients receiving study medication. One subject randomized to the G-ER group had no baseline data recorded and one subject randomized to the placebo group was previously randomized to the G-ER group.

Table 1 Subject Disposition: 81-0062

	Number of Patients	
	G-ER	**Placebo**
Randomized	221 (100%)	231 (100%)
ITT	220	230
Completed	185 (84%)	192 (83%)
Reasons for dropout		
AE	19 (8%)	8 (3%)
Lack of efficacy	7 (3%)	12 (6%)
Protocol violation	2 (1%)	2 (1%)
Withdrawal of consent	4 (2%)	9 (4%)
Other reason	4 (2%)	8 (3%)

Note: See tables 10-1 and 10-2 of the 81-0062 Clinical Study Report.

Patient demographics are presented by treatment groups in the appendix (Table 12). There were no noticeable imbalances between treatment groups with respect to demographic variables of age, race, and sex. The average age of study population was about 66 years and 63% of the population was female and 88% was Caucasian.

Table 12 also shows baseline values for the efficacy variable of average daily pain score by treatment groups. Distributions of the efficacy variable at baseline were comparable between treatment groups.

3.1.3 Statistical Methodologies

The primary analysis used an analysis of covariance (ANCOVA) model including terms for treatment and center as factors and baseline pain score as a covariate. Missing data due to dropouts were imputed employing the baseline observation carried forward (BOCF) strategy in the primary analysis.

As an additional analysis, I also conducted a continuous responder analysis treating dropouts as non-responders. To compare the responder curves, I conducted a van der Waerden test.

To assess the impact of rescue medication usage on the primary analysis, I conducted the same analysis of covariance as in the primary analysis treating all patients who took rescue medications as treatment failures.

To assess differences in the primary and secondary outcome variables among countries, subgroup analyses were conducted by country.

The primary analysis was conducted on the intent-to-treat (ITT) population defined as all patients who were randomized and received the dispensed study medication.

For the analysis of the average daily sleep interference, an analysis of covariance model including terms for treatment and center as factors and baseline score as a covariate was used. For the analysis of Patient Global Impression of Change and Clinical Global Impression of Change, a chi-square test was used.

In order to adjust for multiple testing on the secondary endpoints, a hierarchical test procedure was performed in the following order:

1. PGIC or CGIC (tested at 2.5% level)
2. ADSI (tested at 5% level).

Using this strategy, the analyses of PGIC and CGIC were conducted simultaneously at an alpha level of 0.025 for each variable. If either test was statistically significant, the analysis of the ADSI was conducted at an alpha level of 0.05.

3.1.4 Results and Conclusions

Study 81-0062 demonstrated a statistically significant difference between G-ER and placebo using the pre-specified primary analysis (Table 2).

Table 2 Applicant's Primary Efficacy Analysis: 81-0062 (ITT)

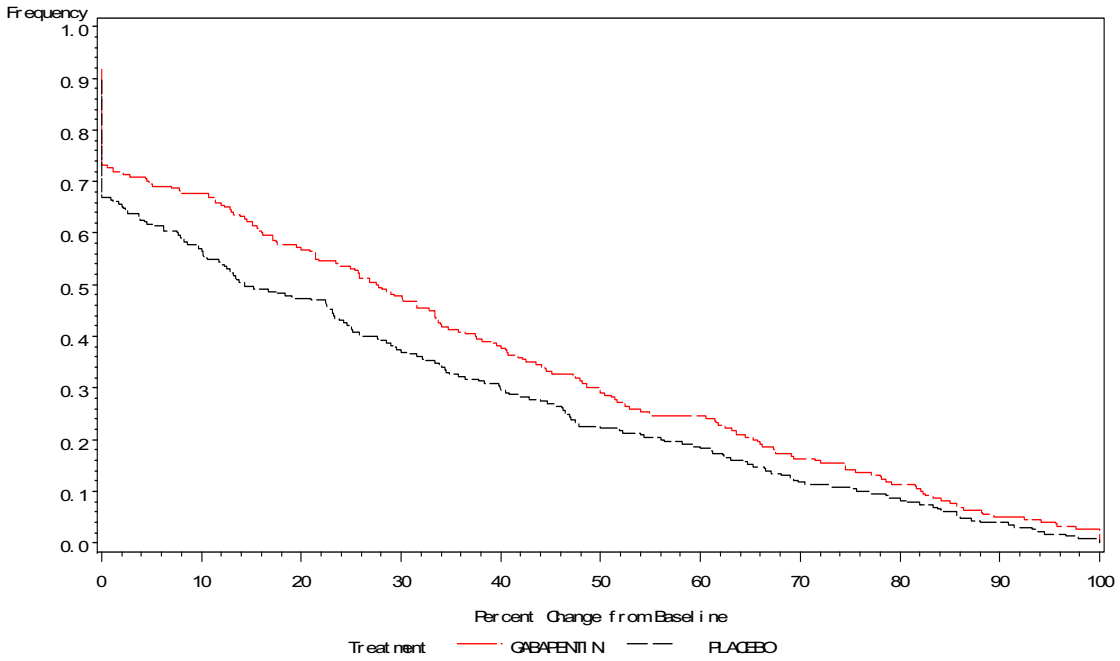
Average Daily Pain	**G-ER (N 220)**	**Placebo (N 230)**	**P-value**
LS Mean Change (SE) from Baseline to Week 11*	-2.1 (0.17)	-1.6 (0.16)	**0.013**
Difference from Placebo (SE)	-0.5 (0.20)		
(95% CI)	(-0.9, -0.1)		

*P value calculated from ANCOVA model with terms for treatment and center, and baseline pain score as a covariate.

The continuous responder curves between G-ER and placebo appear to be separated and the van der Waerden test resulted in statistical significance (Figure 1). The applicant also

generated continuous responder curves which generally matched my curves. However, they did not provide results of statistical comparison.

Figure 1 Reviewer's Continuous Responder Analysis on Primary Efficacy Variable: 81-0062 (ITT)



Note: P-value of **0.039** is generated by van der Waerden test.

As a sensitivity analysis regarding use of rescue medication, I conducted an ANCOVA with same model as in the primary analysis after treating patients who took rescue medication as treatment failures. The results also gave a statistically significant difference between G-ER and placebo (Table 3). Approximately 5% (24/450) of population took rescue medications and approximately 0.4% (2/450) took rescue medications during the last week of the efficacy treatment period. The applicant did not conduct an analysis based on rescue medication since there was little use of rescue medication for pain during the last 7 days of the efficacy treatment period.

Table 3 Reviewer's Sensitivity Analysis: 81-0062 (ITT)

Average Daily Pain	G-ER (N 220)	Placebo (N 230)	P-value
LS Mean Change (SE) from Baseline to Week 11*	-2.0 (0.17)	-1.6 (0.16)	**0.030**
Difference from Placebo (SE)	-0.4 (0.19)		
(95% CI)	(-0.8, -0.0)		

Note: Patients who took rescue medications were treated as treatment failures.
*P value calculated from ANCOVA model with terms for treatment and center, and baseline score as a covariate.

The secondary efficacy analyses on PGIC and CGIC failed to demonstrate statistically significant differences at 2.5% level (Tables 4 5). Therefore, the sequential testing procedure stopped after the tests. However, I present the analysis result for ADSI for descriptive purposes only (Table 6).

Table 4 Applicant's Analysis of Secondary Efficacy Variables: 81-0062 (ITT)

Patient Global Impression of Change	G-ER (N 220)	Placebo (N 230)
Very Much or Much Improved at Endpoint	94 (43%)	77 (34%)
Difference from Placebo	9%	
p-value vs. Placebo*	0.043	

*P value calculated from chi square test. Not significant at 2.5% level.

Table 5 Applicant's Analysis of Secondary Efficacy Variables: 81-0062 (ITT)

Clinical Global Impression of Change	**G-ER (N 220)**	**Placebo (N 230)**
Very Much or Much Improved at Endpoint	97 (44%)	78 (34%)
Difference from Placebo	10%	
p-value vs. Placebo*	0.027	

*P value calculated from chi square test. Not significant at 2.5% level.

Table 6 Applicant's Analysis of Secondary Efficacy Variables: 81-0062 (ITT)

Average Daily Sleep Interference	**G-ER (N 220)**	**Placebo (N 230)**	**P-value**
LS Mean Change (SE) from Baseline to Week 11*	-2.3 (0.16)	-1.6 (0.15)	0.030
Difference from Placebo (SE)	-0.7 (0.18)		
(95% CI)	(-1.1, -0.4)		

*P value calculated from ANCOVA model with terms for treatment and center, and baseline score as a covariate.

3.2 Evaluation of Safety

The evaluation of safety was conducted by the clinical reviewer, Timothy Jiang, M.D.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

I explored the heterogeneity of the treatment effect across age, race, and sex by inclusion of interaction terms in the ANCOVA model. In the analyses of primary efficacy variables, there were no statistically significant interactions between treatment and age group ('<65 yr.' or '≥65 yr.') or sex. There was a statistically significant interaction (p 0.021) between treatment and race group ('Caucasian' or 'Other'). The interaction suggested that G-ER is better than placebo in Caucasians, but G-ER is worse than placebo in others. However, since the majority of the patients were Caucasian (approximately 90%), the negative effect shown in 'Other' race group (approximately 10%) does not appear concerning. I also conducted subgroup analyses by the above demographic variables (Tables 7).

Table 7 Subgroup Analyses on Primary Efficacy Endpoint: 81-0062 (ITT)

LS Mean Change (SE) from Baseline to Week 11 in average daily pain*	**G-ER (N 220)**	**Placebo (N 230)**
White	-2.2 (0.18)	-1.6 (0.17)
Non-White	-1.3 (0.63)	-2.2 (0.67)
Age <65	-2.2 (0.29)	-2.0 (0.28)
Age > 65	-1.9 (0.19)	-1.3 (0.19)
Female	-2.4 (0.20)	-1.7 (0.19)
Male	-1.5 (0.26)	-1.4 (0.27)

*LSMeans calculated from ANCOVA/BOCF model with terms for treatment and center, and baseline pain score as a covariate.

Subgroup analyses by country (US, Russia, or Argentina) showed that the primary analysis with patients from the US sites resulted in statistical significance and the analysis with patients from the non-US sites did not give statistical significance. Therefore, efficacy in the primary analysis appeared to come from the US region only. Furthermore, patients in Argentina showed a negative effect of G-ER compared to placebo. However, considering the small number of patients from Argentina, the negative effect is not convincing and therefore not concerning (Tables 8-10).

Table 8 Reviewer's Subgroup Analysis: 81-0062 (ITT US)

Average Daily Pain	**G-ER (N 126)**	**Placebo (N 131)**	**P-value**
LS Mean Change (SE) from Baseline to Week 11*	-2.5 (0.24)	-1.7 (0.24)	**0.006**
Difference from Placebo (SE)	-0.8 (0.28)		
(95% CI)	(-1.3, -0.2)		

*P value calculated from ANCOVA model with terms for treatment and center, and baseline score as a covariate.

Table 9 Reviewer's Subgroup Analysis: 81-0062 (ITT Russia)

Average Daily Pain	**G-ER (N 81)**	**Placebo (N 80)**	**P-value**
LS Mean Change (SE) from Baseline to Week 11*	-1.6 (0.21)	-1.2 (0.16)	0.109
Difference from Placebo (SE)	-0.4 (0.27)		
(95% CI)	(-1.0, 0.1)		

*P value calculated from ANCOVA model with terms for treatment and center, and baseline score as a covariate.

Table 10 Reviewer's Subgroup Analysis: 81-0062 (ITT Argentina)

Average Daily Pain	**G-ER (N 13)**	**Placebo (N 19)**	**P-value**
LS Mean Change (SE) from Baseline to Week 11*	-1.7 (0.83)	-2.3 (0.64)	0.455
Difference from Placebo (SE)	0.6 (0.84)		
(95% CI)	(-1.1, 2.4)		

*P value calculated from ANCOVA model with terms for treatment and center, and baseline score as a covariate.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

5.1.1 Statistical Issues

There were no statistical issues. During the Type C meeting (December 19, 2005), concerns regarding the use of a LOCF imputation strategy were conveyed. Consequently, the Applicant used a baseline observation carried forward (BOCF) strategy to impute missing pain scores due to dropout. The primary analysis with BOCF showed a statistically significant difference between G-ER and placebo. My continuous responder analysis using a van der Waerden test also showed a statistically significant separation between responder curves. During the End-of Phase 2 meeting (March 2, 2006), the Division requested clarification regarding incorporation of rescue use into the analysis. The Applicant did not incorporate rescue use in the analysis. However, my analysis treating patients who took rescue medications as treatment failures produced results consistent with the primary analysis. Subgroup analyses by country (US, Russia or Argentina) suggested that the overall treatment effect was mainly driven by the US.

In order to adjust for multiple secondary endpoints PGIC, CGIC, and ADSI, a sequential testing procedure was employed. Based on the sequential testing procedure, significance was not achieved for these secondary endpoints.

5.1.2 Collective Evidence

In reviewing the evidence from the applicant's primary analyses as well as my additional analyses, I conclude that the data from Study 81-0062 provide evidence of the efficacy of G-ER for treating PHN.

5.2 Conclusions and Recommendations

The Applicant proposes G-ER for the management of postherpetic neuralgia. Based on my review of Study 81-0062, the Applicant has demonstrated that PHN patients receiving G-ER had a reduction in pain compared to patients receiving placebo.

5.3 Review of Clinical Studies of Proposed Label

The following portion of the Clinical Study section from the proposed label includes the applicant's results of data analyses from the study in the current submission. It is

consistent with the study report, and I agree with the presentation in general. (b) (4) [redacted] Lastly in Figure 2, the Y-axis should extend to 100.

I highlighted the relevant sentences below in yellow.

> The efficacy of TRADENAME for the management of postherpetic neuralgia was established in a double-blind, placebo-controlled, multicenter study. This study enrolled patients between the age of 21 to 89 with neuralgia persisting for at least 6 months following healing of herpes zoster rash and a minimum baseline pain intensity score of at least 4 on an 11-point numerical pain rating scale ranging from 0 (no pain) to 10 (worst possible pain).
>
> This 11-week study compared TRADENAME 1800 mg once daily with placebo. A total of 221 and 231 patients were treated with TRADENAME or placebo, respectively. The study treatment including titration for all patients comprised a 10-week treatment period followed by 1-week of dose tapering. Double-blind treatment began with titration starting at 300 mg/day and titrated up to a total daily dose of 1800 mg over 2 weeks, followed by 8 weeks fixed dosing at 1800 mg once daily, and then 1 week of dose tapering. During the 8-week stable dosing period, patients took 3 active or placebo tablets each night with the evening meal. During baseline and treatment, patients recorded their pain in a daily diary using an 11-point numeric pain rating scale. (b) (4) [redacted]. The mean baseline pain score was 6.6 and 6.5 for TRADENAME and placebo-treated patients, respectively.
>
> (b) (4) [redacted]

(b) (4)

Treatment with TRADENAME statistically significantly improved the endpoint mean pain score (b) (4) from baseline. For various degrees of improvement in pain from baseline to study endpoint, (b) (4) shows the fraction of patients achieving that degree of improvement. The figure is cumulative, so that patients whose change from baseline is, for example, 50%, are also included at every level of improvement below 50%. Patients who did not complete the study were assigned 0% improvement. (b) (4)

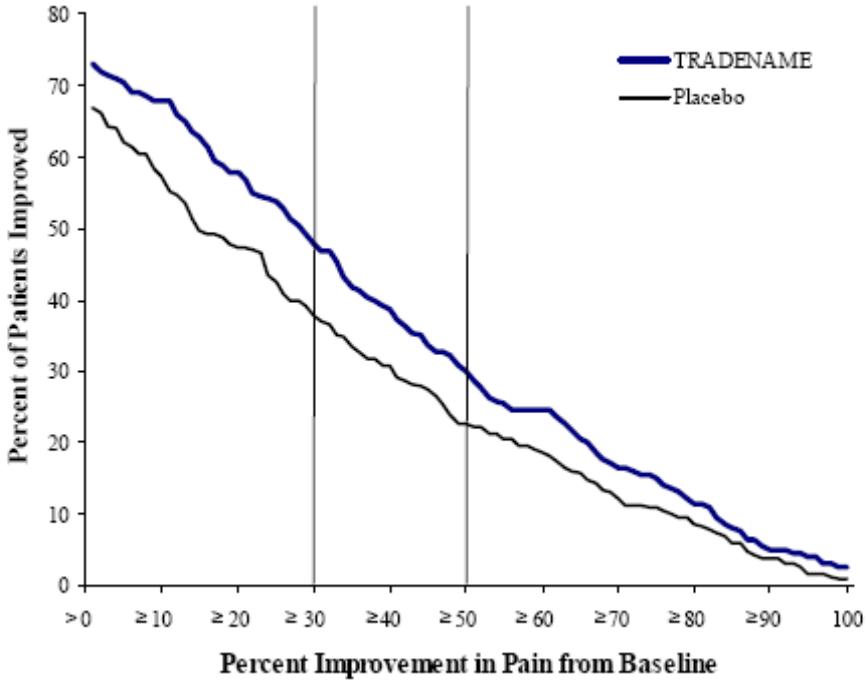


Figure 2. Percent of Patients Achieving Various Levels of Pain Relief

APPENDIX

Table 11 Patient Demographic and Baseline Characteristics: 81-0062

	G-ER (n 221)	**Placebo (n 231)**
Gender n (%)		
Female	135 (61%)	148 (64%)
Male	86 (39%)	83 (36%)
Race n (%)		
Caucasian	197 (89%)	205 (88%)
Black	10 (4%)	6 (3%)
Asian	1 (1%)	2 (1%)
Other	13 (6%)	18 (8%)
Age (years)		
Mean (SD)	65 (13.3)	66 (11.1)
Average Daily Pain		
Mean (SD)	6.6 (1.4)	6.5 (1.4)

SIGNATURES/DISTRIBUTION LIST

Primary Statistical Reviewer: Yongman Kim, Ph.D.
Mathematical Statistician
Date: December 3, 2010

Concurring Reviewer: Dionne Price, Ph.D.
Team Leader

cc:

DAARP/Allison Meyer
DAARP/Timothy Jiang, M.D.
DAARP/Ellen Fields, M.D.
HFD-715/Yongman Kim, Ph.D.
HFD-715/Dionne Price, Ph.D.
HFD-715/Thomas Permutt, Ph.D.
HFD-700/Edward Nevius, Ph.D.
HFD-700/Ram Tiwari, Ph.D.
HFD-700/Lillian Patrician

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

/s/

YONGMAN KIM
12/06/2010

DIONNE L PRICE
12/06/2010
concur