Title: Effect of Duloxetine 30/ 60 mg Once Daily versus Placebo in Adolescents with Juvenile Primary Fibromyalgia Syndrome.

NCT ID: NCT01237587

Document Approval Date: 27-Jun-2017

1. Effect of Duloxetine 30/60 mg Once Daily versus Placebo in Adolescents with Juvenile Primary Fibromyalgia Syndrome

Duloxetine hydrochloride (LY248686)

Phase IIIb, multicenter, randomized, double-blind, placebo-controlled, 13-week clinical trial to assess the efficacy and safety of duloxetine 30/60 mg QD compared to placebo in adolescents with Juvenile Primary Fibromyalgia Syndrome. Continued efficacy and safety of duloxetine 30/60 mg QD will be assessed in a 26-week, open-label treatment extension period.

Eli Lilly and Company
Protocol F1J-MC-HMGW
Phase IIIb

Statistical Analysis Plan Version 3 electronically signed and approved by Lilly on date provided below.

Statistical Analysis Plan Version 2 electronically signed and approved by Lilly on date 11Dec2014
Statistical Analysis Plan Version 1 electronically signed and approved by Lilly: 10Feb2011

Approval Date: 27-Jun-2017 GMT

2. Table of Contents

Section **Page**

3. Revision History

SAP version 1 was approved on 10Feb2011, and SAP version 2 was approved on 11Dec2014 both before unblinding. The following changes were made to the SAP since the version 1:

- In Section 5.3, added patients could be re-screened based on revised protocol (b).
- The section of 'Determination of Sample Size' was moved to Section 5.2 and updated to reflect more accurate assumptions around dropout rate.
- More details of patient population, baseline and post-baseline definitions for all efficacy and safety analyses are summarized in Table HMGW.6.1.
- The details of definition of treatment group for Study Period II/III were added in Section 6.1.2, and analyses for Study Period II/III were added in Section 6.1.13 and 6.1.14.
- Removed the potential DLX_NOIP treatment group for SP IV in Section 6.1.2.
- In Section 6.1.4, a selection model and a placebo multiple imputation (pMI) approach were added to handle dropouts or missing data.
- The details of definition of baseline concurrent/ongoing therapies were added in Section 6.1.8.
- The details of definition of significant protocol deviations were added in Section 6.1.9 and summarized in Table HMGW.6.2.
- More details of definition of last dose date for Study Period II and Study Period III were added in Section 6.1.10.
- More details of the test of maintenance of effect were added in Section 6.1.13.2 and Section 4.2.
- Added Section 6.1.13.3 for sensitivity analysis and updated the sensitivity analysis employed.
- More details of definition of treatment-emergent adverse events (TEAE) for Study Period II and Study Period III were added in Section 6.1.14.1.
- The details of definition of normal values at baseline were added for treatment emergent and potentially clinically significant analyses in vital signs and ECGs in Section 6.1.14.3 and 6.1.14.4 respectively.
- Analyses for non-suicidal self-injurious behavior were added in Section 6.1.14.5.
- Adjusted paragraph order in the safety Section 6.1.14.
- The details of American College of Rheumatology (ACR) criteria for fibromyalgia (Wolfe et al. 1990) were added in Section 6.3.2.

- The section of 'Changes in the Conduct of the Study or Planned Analyses' was removed since it will be summarized in the clinical study report for this study.
- Update the Interim Analysis in Section 6.2.
- Removed Appendix 1 Details of Sensitivity Analysis and Appendix 2 Implementation of Placebo Multiple Imputation.

4. Study Objectives

4.1. Primary Objective

The primary objective of this study is to evaluate the efficacy of duloxetine 30/60 mg once daily (QD) compared with placebo on the reduction of average pain severity as measured by the Brief Pain Inventory (BPI) – Modified Short Form: Adolescent Version 24-hour average pain severity rating (for simplicity, it is referred to as the BPI average pain severity hereafter) during a 13-week, double-blind treatment phase in adolescents (aged 13 to 17 years) with Juvenile Primary Fibromyalgia Syndrome (JPFS), as defined by Yunus and Masi (Yunus and Masi 1985).

4.2. Secondary Objectives

The secondary objectives of the study are as follows:

- To evaluate the efficacy of duloxetine 30/60 mg QD in the treatment of adolescents with JPFS during a 13-week, double-blind treatment phase, based on the improvement of the following measures:
    - Brief Pain Inventory (BPI) Modified Short Form: Adolescent Version severity (worst pain, least pain, pain right now) and interference.
    - Response to treatment, as defined by a 30% and 50% reduction in the BPI average pain severity.
    - Pediatric Pain Questionnaire (PPQ) pain right now, worst pain and average pain items.
    - Clinical Global Impression of Severity for Overall Illness (CGI-Severity: Overall Illness) scale.
    - Clinical Global Impression of Severity for Mental Illness (CGI-Severity: Mental Illness) scale.
    - Functional Disability Inventory - child version scale (FDI-child).
    - Functional Disability Inventory - parent version scale (FDI-parent).
    - Children's Depression Inventory (CDI).
    - Multidimensional Anxiety Scale for Children (MASC).
- To evaluate the safety and tolerability of duloxetine 30/60 mg QD in the treatment of adolescents with JPFS during a 13-week, double-blind treatment phase.
- To evaluate the efficacy and safety of duloxetine 30/60 mg QD during a 26-week, open-label, extension treatment phase, as assessed by the following:
    - Maintenance effect of duloxetine 30/60 mg QD during the open-label extension treatment phase. Maintenance effect will be assessed using the BPI average pain severity in only the duloxetine treated acute phase responders (defined as those patients with ≥30% pain reduction from baseline on the BPI average pain severity measure at the last non-missing assessment in Study Period II).
    - Effect of duloxetine 30/60 mg QD during an extension treatment phase as measured by the following: BPI, PPQ, CGI-Severity: Overall Illness, CGI-Severity: Mental Illness, FDI-child, FDI-parent, CDI, and MASC.

- Safety of duloxetine 30/60 mg QD during an extension treatment phase.

4.3. Exploratory Objectives

Exploratory objectives of the study are as follows:

- To evaluate the correlation between BPI pain severity items and PPQ pain items.
- To evaluate the appropriateness of the ACR FM criteria (Wolfe et al. 1990) in an adolescent population.

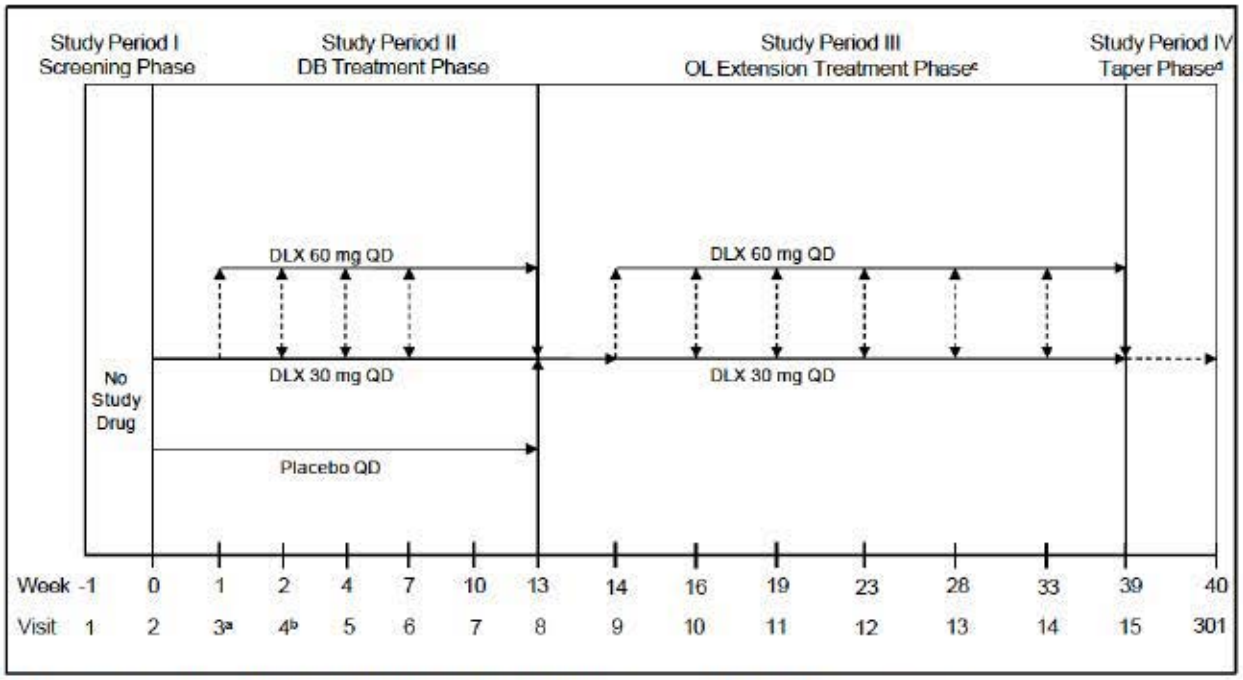
5. Study Elements

This section contains the summary of study design and the method of treatment assignment from Study F1J-MC-HMGW protocol.

5.1. Summary of Study Design

Study F1J-MC-HMGW (HMGW) is a Phase 3b, multicenter, randomized clinical trial of duloxetine versus placebo for the treatment of JPFS in adolescents. The study design includes 4 study periods. Following Study Period I (screening phase), eligible patients enter Study Period II, the acute 13-week parallel, randomized, double-blind, placebo-controlled treatment period of duloxetine 30/60 mg QD. Upon completion of the 13-week double-blind period, patients enter Study Period III, a 26-week, open-label extension phase of duloxetine 30/60 mg QD. A 1-week drug taper period (Study Period IV) is required for all patients who discontinue for any reason between Visit 4 to Visit 8 during Study Period II, or who complete/discontinue Study Period III on duloxetine 60 mg QD.

Figure HMGW.5.1 illustrates the study design.



Abbreviations: DB = double-blind; DLX = duloxetine; OL = open-label; QD = once-daily

[a] After Visit 3, patients increased to 60 mg duloxetine can have their dose decreased 1 time, at a scheduled or unscheduled Visit. Once 60 mg dose is decreased to 30 mg, it cannot be increased again during Study Period II.

[b] At Visits 4 through 8, patients randomized to placebo or DLX 30 mg who discontinue prior to entering the open-label extension treatment period (Study Period III) will receive placebo in the drug taper period.

[c] At Visits 9 through 14 in Study Period III, dose increases are permitted at scheduled visits (to a maximum dose of 60 mg QD) and dose decreases are permitted at scheduled or unscheduled visits (to a minimum dose of 30 mg QD).

[d] Patients who complete/discontinue Study Period III on DLX 30 mg do not need to enter the drug taper period.

Figure HMGW.5.1 Illustration of study design for Protocol F1J-MC-HMGW.

Study Period I is a screening period of approximately 1 week in duration, but not more than 30 days, during which patients will be screened for study eligibility.

Study Period II is a 13-week, acute treatment period. During Study Period II, at Visit 2, patients who meet all study criteria for enrollment will be randomly assigned in a 1:1 ratio to duloxetine or placebo by a computer-generated random sequence using an interactive voice response system (IVRS). All patients will receive investigational product (IP) beginning at Visit 2 and will be instructed to start taking their IP the morning after Visit 2.

Patients randomly assigned to placebo will remain on placebo throughout Study Period II.

Patients randomly assigned to the duloxetine treatment group will remain on duloxetine throughout Study Period II. Duloxetine patients will initially take a 30 mg QD dose of duloxetine for 1 week. Starting at Visit 3 through Visit 6, the dose can be increased from 30 mg QD to 60 mg QD. By Visit 4, comorbid MDD and/or GAD patients receiving a 30-mg QD dose should be increased to a 60-mg QD dose, unless the patient is experiencing significant intolerance to the 30-mg dose. Beginning at Visit 7, no dose increases can be made. Dose –escalation criteria for all patients is based on investigator's clinical judgment. Dose increase can only occur at scheduled study visits. If necessary due to intolerability, for those patients on a 60-mg QD dose, a one-time dose decrease to 30 mg may occur at a scheduled or unscheduled visit. Once the 60-mg dose is decreased to 30 mg, it cannot be increased again during Study Period II. If the patient has already had their 1 dose decrease, and at a subsequent time cannot tolerate the investigational product dose well enough to remain compliant, the patient should be discontinued.

Study Period III is a 26-week, open-label extension treatment phase. Patients randomized either to the duloxetine or placebo treatment group during Study Period II will enter Study Period III on a 30-mg dose of duloxetine for 1 week. The duloxetine dose may be increased to 60 mg QD only at a scheduled Visit starting at Visit 9 (Week 14). As in Study Period II, additional dose-escalation consideration should be given to all patients with comorbid MDD and/or GAD at study entry. At Visit 9, the comorbid MDD and/or GAD patients should have their 30-mg dose increased to 60 mg, unless the patient is experiencing significant intolerance to the 30-mg dose. From Visit 9 to Visit 14, dose modifications are allowed. Dose increases (up to 60 mg QD) are only permitted at scheduled visits. Dose decreases (down to 30 mg QD) are permitted at either scheduled or unscheduled visits. Dose adjustments are to be based on the investigator's clinical judgment of treatment response and tolerability at the patient's current dose.

Study Period IV is a 1-week taper phase to minimize the occurrence of discontinuation-emergent adverse events (AEs). Patients who complete Study Period III or discontinue after Visit 9 of Study Period III while on duloxetine 60 mg QD, will receive duloxetine 30 mg QD for 1 week. The 1-week taper is not required for patients in Study Period III who complete or discontinue early on a 30-mg QD dose of duloxetine.

Patients who discontinue between Visits 4 to 8 of Study Period II should enter into the 1-week drug taper period. However, tapering of IP should be based on the investigator's determination of safety for his or her patient. Patients on duloxetine 60 mg QD will receive 30 mg QD for

1 week. Patients on duloxetine 30 mg QD will receive placebo for 1 week. Patients on placebo will receive placebo for 1 week. The 1-week taper is implemented in a blinded manner using the IVRS. The taper period is not required for patients who discontinue at Visit 3 and therefore have not received 2 weeks of treatment.

5.2. Determination of Sample Size

Approximately 184 patients will be randomized in a 1:1 ratio to the duloxetine and placebo treatment groups. This study is to assess the difference of the mean change in BPI average pain severity from baseline to the last time point (Visit 8, Week 13) in Study Period II between treatment groups. Assuming that there will be some missing post-baseline data, and about 81% completion rate at acute phase by Visit 8 or Week 13, this sample size will provide at least 80% power to detect the treatment difference of 1.0 point with α=.05. All the parameters used in the sample size calculations were based on the MMRM analysis of data from 3-month, placebo-controlled, acute treatment period in 2 adult FM studies: Study F1J-MC-HMCA and Study F1J-MC-HMCJ.

5.3. Method of Assignment to Treatment

After the informed consent form (ICF) is signed and dated, a patient is considered to be "entered" into the study and will be assigned a patient number. Patients who meet all criteria for enrollment will be randomized to double-blind treatment at Visit 2. Assignment to treatment groups will be determined by a computer generated random sequence using an IVRS. The IVRS will be used to assign investigational product packages containing double-blind investigational product to each patient during Study Period II, and open-label investigational product during Study Period III. Site personnel will confirm that they have located the correct investigational product packages by entering a confirmation number found on the investigational product packages into the IVRS.

Individuals who do not meet certain criteria for participation in this study (screen failure) may be re-screened. Individuals may be re-screened one time. The interval between screening and re-screening should be at least 3 weeks. When re-screening is performed, the individual must sign a new ICF and will be assigned a new identification number.

To achieve between-group comparability for site factor, the randomization will be stratified by site.

6. A Priori Statistical Methods

6.1. Statistical and Analytical Plans

The protocol for this study was approved on 25 October 2010. The protocol amendment (a) was approved on 13 July 2011. The protocol amendment (b) was approved on 24 April 2013. The version 2 of SAP will supersede the statistical plans described in the protocol amendment (b). Section 6.1 addresses the statistical analyses planned before unblinding.

6.1.1. Analysis Populations

All data presentations will be based the intent-to-treat (ITT) population, which includes all randomized patients.

6.1.2. General Considerations

Efficacy and safety analyses will be done on an intent-to-treat (ITT) basis unless otherwise specified. An intent-to-treat analysis is an analysis of data by the groups to which patients are assigned by random allocation, even if the patient does not take the assigned treatment, does not receive the correct treatment, or otherwise does not follow the protocol.

All efficacy and safety analyses of continuous measures will include randomized patients with both a baseline and at least one post-baseline value for the variable being analyzed. For efficacy and continuous safety variables, baseline is defined as the last measurement taken at, or prior to the visit where the study period begins; endpoint is defined as the last non-missing measurement for the study period of interest. Table HMGW 6.1 provides baseline and post-baseline definitions for all variables.

Statistical analyses will be carried out for Study Period II, III, and IV (SP II-IV). All tests of hypotheses will be evaluated based on a two-sided significance level of 0.05 unless otherwise specified. No adjustments for multiple comparisons will be made.

For SP II, a patient's treatment group will be determined by the treatment to which the patient is randomized: either duloxetine (DLX) or placebo (PLA).

For SP III, the following treatment groups will be determined by the randomized treatment at SP II and extension phase treatment at SP III:

> PLA/DLX: patients who are randomized to placebo at SP II and who enter SP III and take duloxetine.
>
> DLX/DLX: patients who are randomized to duloxetine at SP II and who enter SP III and take duloxetine.

No statistical comparisons will be conducted to compare the two treatment groups PLA/DLX and DLX/DLX during SP III unless otherwise noted. For continuous measures, when appropriate, within treatment testing of change from baseline for selected safety and efficacy measures will be conducted for SP III.

For SP II/III, one treatment group will be determined by the randomized treatment at SP II:

> DLX: patients who are randomized to duloxetine at SPII, including patients who may have discontinued early at SPII or who enter SPIII.

For SP IV, a patient's treatment group assignment will be based on the last treatment received before entering Study Period IV and the treatment received during SP IV: PLA_PLA, DLX30_PLA, or DLX60_DLX30. SP IV is not required for patients who discontinue the study during SP III on the duloxetine dose of 30 mg (Visit 9-15), however it would be allowed based on investigator decision.

Table HMGW.6.1 below presents baseline and post-baseline period definitions for each study period and for each safety and efficacy analysis.

Table HMGW.6.1 Patient Population with Baseline and Post-baseline Definitions by Study Period and Type of Analysis

Study Period / Analysis	Patient Population	Baseline Definition	Post-baseline Definition
Acute Phase			
Efficacy	Patients with a baseline and at least one post-baseline observation	Last of Visits 1-2	All Visits 3–8
TEAEs	All randomized patients	All Visits 1–2	All Visits 2.01–8
SAE and DCAE	All randomized patients	NA	All Visits 2.01–8
C-SSRS categorical analyses	All randomized patients	Lifetime baseline: All Visits 1–2 Lead-in baseline: Visit 2	All Visits 2.01–8
Treatment Emergent Abnormal Labs	Patients with normal laboratory values at all nonmissing baseline visits (with respect to direction being analyzed) and who have at least one post-baseline observation	All of Visits 1-2	All Visits 2.01–8
Treatment emergent low blood pressure	Patients with normal values at all nonmissing baseline visits (with respect to direction being analyzed) and who have at least one post-baseline observation. See Table HMGW.6.4 for normal limits of blood pressure.	Minimum value from Visits 1-2	All Visits 2.01–8
PCS in vital signs, weight, sustained blood pressure analyses	Vital signs: Patients with normal values at all nonmissing baseline visits (with respect to direction being analyzed) and who have at least one post-baseline observation. See Table HMGW.6.4 for normal limits of vital signs. Weight:	Low: Minimum value from Visits 1-2High: Maximum value from Visits 1-2	All Visits 2.01–8

	Patients with non-missing baseline and at least one post-baseline value.		
Treatment-emergent changes in ECG intervals and heart rate	Patients with normal values at all nonmissing baseline visits (with respect to direction being analyzed) and who have at least one post-baseline observation. See Table HMGW.6.8 for normal limits of ECG.	Low: Minimum value from Visits 1–2 High: Maximum value from Visits 1–2	All Visits 2.01–8
Continuous Safety measures	Patients with a baseline and at least one post-baseline observation	Last of Visits 1-2	All Visits 3–8
Extension Phase			
Efficacy	Patients with a baseline and at least one post-baseline observation	Last of Visits 3-8	All Visits 9-15
TEAEs	All randomized patients entering extension phase	Maximum severity between 7<=Visit<=8	All Visits 8.01-15
SAE and DCAE	All randomized patients entering extension phase	NA	All Visits 8.01-15
C-SSRS categorical analyses	All randomized patients entering extension phase.	Lifetime baseline: All Visits 1–8 Lead-in baseline: 7<=Visit<=8	All Visits 8.01-15
Treatment Emergent Abnormal Labs	Patients with normal laboratory values at all nonmissing baseline visits (with respect to direction being analyzed) and who have at least one post-baseline observation	All Visits 2.01-8	All Visits 8.01-15
Treatment emergent low blood pressure	Patients with normal values at all nonmissing baseline visits (with respect to direction being analyzed) and who have at least one post-baseline observation. See Table HMGW.6.4 for normal limits of blood pressure.	Minimum value from Visits 2.01-8	All Visits 8.01-15
PCS in vital signs, weight, sustained blood pressure analyses	Vital signs: Patients with normal values at all nonmissing baseline visits (with respect to direction being analyzed) and who have at least one post-baseline observation. See Table HMGW.6.4 for normal limits of vital signs. Weight: Patients with non-missing baseline and at least one post-baseline value.	Low: Minimum value from Visits 2.01–8 High: Maximum value from Visits 2.01–8	All Visits 8.01-15
Treatment-emergent changes in ECG intervals and heart rate	Patients with normal values at all nonmissing baseline visits (with	Low: Minimum value	All Visits 8.01–15

	respect to direction being analyzed) and who have at least one post-baseline observation. See Table HMGW.6.8 for normal limits of ECG.	from Visits 2.01–8 High: Maximum value from Visits 2.01–8	
Continuous Safety measures	Patients with a baseline and at least one post-baseline observation	Last of Visits 3-8 (for vital sign analysis, another baseline is: Last of Visits 1-2)	All Visits 9-15
Acute and Extension Phase Combined			
Efficacy	Patients with a baseline and at least one post-baseline observation	Last of Visits 1-2	All Visits 3-15
Treatment emergent low blood pressure	Patients with normal values at all nonmissing baseline visits (with respect to direction being analyzed) and who have at least one post-baseline observation. See Table HMGW.6.4 for normal limits of blood pressure.	Minimum value from Visits 1-2	All Visits 2.01-15
PCS in vital signs, weight, sustained blood pressure analyses	Vital signs: Patients with normal values at all nonmissing baseline visits (with respect to direction being analyzed) and who have at least one post-baseline observation. See Table HMGW.6.4 for normal limits of vital signs. Weight: Patients with non-missing baseline and at least one post-baseline value.	Low: Minimum value from Visits 1–2 High: Maximum value from Visits 1–2	All Visits 2.01-15
Treatment-emergent changes in ECG intervals and heart rate	Patients with normal values at all nonmissing baseline visits (with respect to direction being analyzed) and who have at least one post-baseline observation. See Table HMGW.6.8 for normal limits of ECG.	Low: Minimum value from Visits 1–2 High: Maximum value from Visits 1–2	All Visits 2.01-15
Continuous Safety measures	Patients with a baseline and at least one post-baseline observation	Last of Visits 1-2	All Visits 3-15
Taper Phase			
Discontinuation emergent AEs	All randomized patients entering taper phase	Maximum severity at last two visits before entering taper	Visit 301
SAE and DCAE	All randomized patients entering	NA	Visit 301

	taper phase		
C-SSRS categorical analyses	All randomized patients entering taper phase	Lifetime baseline: All visits before entering taper phase (Visits 1-15) Lead-in baseline: Last visit before entering taper phase	Visit 301
Continuous measure in vital signs and weight	All randomized patients entering taper phase	Last scheduled visit before entering taper phase	Visit 301

Note: Visit 2.01 indicates the first unscheduled visit occurring after Visit 2 and prior to Visit 3.
Visit 8.01 indicates the first unscheduled visit occurring after Visit 8 and prior to Visit 9.

Abbreviation: AE = adverse event; C-SSRS = Columbia-Suicide Severity Rating Scale; DCAE = adverse event leading to discontinuation; NA = not applicable; PCS = potentially clinically significant; SAE = serious adverse event; TEAE = treatment-emergent adverse event.

A repeated measures analysis refers to a restricted maximum likelihood (REML)-based, mixed-effects repeated measures (MMRM) analysis using all the longitudinal observations at each post-baseline visit. The model for this analysis will include the fixed categorical effect of treatment, pooled investigative site, visit, and treatment-by-visit interaction, as well as the continuous, fixed covariate of baseline value of the variable being analyzed and baseline value of the variable being analyzed-by-visit interaction. An unstructured covariance structure will be used to model the within-patient errors. And Kenwood-Roger approximation (Kenward and Roger, 1997) will be used to estimate denominator degrees of freedom. If the default Newton-Raphson algorithm used by SAS PROC MIXED does not converge, then the Fisher scoring algorithm will be applied.

However, if the model with unstructured covariance structure still fails to converge, the sandwich estimator (Diggle, Liang, and Zeger, 1994; Lu and Mehrotra, 2009) will be used to estimate the standard errors of the fixed effects parameters and the model will be fitted using covariance structures in the following order until convergence is met:

heterogeneous toeplitz	type = toeph
heterogeneous autoregressive (1st order)	type = arh(1)
heterogeneous compound symmetric	type = cs(h)
toeplitz	type = toep
autoregressive (1st order)	type = ar(1)
compound symmetric	type = cs

When the sandwich estimator is used, the Kenward-Roger approximation for denominator degrees of freedom cannot be used. Instead, the denominator degrees of freedom will be partitioned into between-subject and within-subject portions (denoted by DDFM= BETWITHIN in the MODEL statement).

For repeated measures analyses of efficacy and safety variables that are not collected at each post-baseline visit, data may exist at visits where the variable was not scheduled to be collected due to early discontinuation visits. Data collected at scheduled visits and early discontinuation visits will be used in the analysis. In these situations, the data from the early discontinuation visit will be carried forward to the next regularly scheduled collection visit for the repeated measures analysis. Significance tests will be based on least-squares means (LSMean) and Type III sum-of-squares, using a two-sided α=0.05. Analyses will be implemented using SAS® PROC MIXED.

The repeated measures analysis for categorical variables will use a categorical, pseudo-likelihood-based repeated measures (MMRM-CAT) approach. The model will include the fixed, categorical effect of treatment, pooled investigative site, visit, and treatment-by-visit interaction, as well as the continuous, fixed covariate of baseline value of the variable being analyzed and baseline value of the variable being analyzed-by-visit interaction. An unstructured covariance structure will be used to model the within-patient errors. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. If the model does not converge with the default fitting algorithm used by PROC GLIMMIX, the Fishers scoring algorithm will be utilized by the SCORING option in SAS. If the model still fails to converge, the model will be fitted using covariance matrices in the following order specified by a decreasing number of covariance parameters until convergence is met: heterogeneous toeplitz, heterogeneous autoregressive (1st order), heterogeneous compound symmetric, toeplitz, autoregressive (1st order),compound symmetry. For models where the unstructured covariance matrix is not utilized, the sandwich estimator (Diggle, Liang, and Zeger 1994) will be used to estimate the standard errors of the fixed effects parameters. The sandwich estimator is utilized by the EMPIRICAL option in SAS. When the sandwich estimator is utilized, the Kenward-Roger approximation for denominator degrees of freedom cannot be used. Instead, the denominator degrees of freedom will be partitioned into between-subject and within-subject portions by the DDFM=BETWITHIN option in SAS. This analysis will be implemented using SAS® PROC GLIMMIX.

Unless otherwise specified, when an analysis of variance (ANOVA) model is used to analyze a continuous efficacy variable, the model will contain the main effects of treatment and pooled investigative site. The significance of treatment-by-pooled investigative site interaction will be evaluated in a separate model, when appropriate. Similar logic is applied to an analysis of covariance (ANCOVA) model, which in general, refers to the ANOVA model with baseline values added as covariates. Type III sum-of-squares for the LSMean will be used for the statistical comparison of main effects using ANOVA or ANCOVA. Statistical inference for ANOVA or ANCOVA interaction terms will be based on type II sum-of-squares for the LSMean.

The same MMRM, ANOVA and ANCOVA models excluding treatment effect will be used to present data for Study Period III, where appropriate.

Unless otherwise specified, for all analyses using the last-observation-carried forward (LOCF) approach in Study Period II, endpoint is defined as the last nonmissing observation obtained from Visit 3 through Visit 8; for the analyses using LOCF in Study Period III, endpoint is defined as the last nonmissing observation from Visit 9 through Visit 15.

In the analyses of some efficacy variables for Study Period II, the baseline-observation-carried-forward (BOCF) approach is considered. The BOCF endpoint is defined as follows:

> For randomized patients who complete the treatment period (that is, complete the scheduled Visit 8) and have last nonmissing value at Visit 8, the BOCF endpoint is defined as the last nonmissing observation, for randomized patients who discontinue early (that is, do not complete the scheduled Visit 8) but with at least one non-missing baseline value, the BOCF endpoint is defined as the baseline value.

Categorical comparisons between treatment groups will be performed using Cochran-Mantel-Haenszel (CMH) test controlling for pooled investigative site and Fisher's exact test for Study Period II, when appropriate. For the categorical data, the number and percentage of patients for the categorical variables will be presented for Study Period III.

Where appropriate, variables will be summarized descriptively (frequency and percent will be summarized for categorical variables; number of patients with a non-missing observation, mean, SD, median, minimum, and maximum will be summarized for continuous variables) by study visit.

For those data summaries accompanied by individual patient data listings, data will be sorted by investigator ID, patient ID, and visit or event number if it is available.

Changes made to the data analysis methods will not necessarily require a protocol amendment and will be described in an updated SAP and reflected in the final report.

Statistical analysis of this study will be the responsibility of Eli Lilly and Company or designee. SAS® software will be used to perform most or all statistical analyses.

6.1.3. Adjustments for Covariates

In general, when a repeated measures analysis is used, the model will include the fixed categorical effect of treatment, pooled investigative site, visit, treatment-by-visit interaction, and the continuous, fixed covariates of the baseline value of the variable being analyzed and baseline value of the variable being analyzed-by-visit interaction. The baseline value of the variable being analyzed and the baseline value of the variable being analyzed-by-visit interaction are included to account for the differing influence over time of the baseline score on the post-baseline scores.

When an analysis of covariance (ANCOVA) model is used to analyze a continuous variable, the model will contain the main effects of treatment, pooled investigative site, and appropriate baseline values included as covariates.

Data presentations for Study Period III will be based on the same MMRM and ANCOVA models excluding treatment effect, where appropriate.

6.1.4. Handling of Dropouts or Missing Data

Missing efficacy and safety data can occur for multiple reasons, including missed patient visits (for example, patients discontinuing from the study early) and scales or measures with missing item scores.

When change from baseline is assessed, only patients with a baseline and at least one post-baseline measurement will be included in the analyses.

For the repeated measures analyses, the model parameters are simultaneously estimated using restricted likelihood estimation incorporating all of the observed data.

Unless otherwise specified, when a total score or an average score is calculated from individual items, it will be considered missing if any of the individual items are missing.

For the assessment of efficacy, handling of missing data depends on the missing data mechanism assumptions. There are three missing data mechanisms: 1) missing completely at random (MCAR), 2) missing at random (MAR) and 3) missing not at random (MNAR). Three primary statistical approaches to handling missing data will be utilized: mixed model repeated measures (MMRM) analyses, ANCOVA/ANOVA using last observation carried forward (LOCF) change from baseline to endpoint, and ANCOVA/ANOVA using baseline observation carried forward (BOCF) change from baseline to endpoint. The LOCF and BOCF imputation methods in general make MNAR assumption, while the repeated measures analysis assumes MAR. A sensitivity analysis will be performed to assess the impact of missing data assumptions on the primary analysis conclusions (see section 6.1.13.3 for details).

6.1.5. Multicenter Studies

All investigative sites with fewer than 2 patients randomized to each treatment group (each patient with non-missing change from baseline in BPI average pain severity rating score) will be pooled together within each country and considered a single site for analyses. If this results in a "site" still having fewer than 2 patients randomized to each treatment, these sites will be pooled together with the next smallest site in that country. If there are no other sites in that country, then these sites would be pooled with the next smallest site in the whole study. All analyses will use pooled investigative sites. The actual investigative site numbers will be included in the listings.

6.1.6. Multiple Comparisons/Multiplicity

The primary efficacy analysis will be the contrast between duloxetine and placebo in the BPI average pain score change from baseline to endpoint (week 13) in Study Period II from a repeated measures analysis. No adjustments will be made for multiple comparisons.

6.1.7. Patient Disposition

The number and percentage of randomized patients who complete the study or discontinue early will be tabulated for all treatment groups for Study Period II, Study Period III, and Study Period

IV both overall and by visit. Reasons for discontinuation presented for Study Period II will be compared between treatment groups using Fisher's exact test. Reasons for discontinuation presented for Study Period III and Study Period IV will be summarized by treatment group.

Patient disposition data from all study periods will be listed.

6.1.8. Patient Characteristics

Patient allocation by investigator, grouped by country, will be summarized for Study Period II for all ITT patients.

Patient allocation by investigator will also be listed for all ITT patients.

The following patient characteristics will be summarized by treatment group for all ITT patients for Study Period II, and for those ITT patients entering the Study Period III.

- Demographic information: age, gender, ethnicity, race origin, country, height, weight, and BMI
- Baseline comorbid diagnoses: MDD, GAD, and ADD
- Baseline severity of illness measured by: BPI severity and interference ratings, PPQ pain ratings, CGI-Severity: Mental Illness, CGI-Severity: Overall Illness, FDI-child, and FDI-parent, CDI, and MASC
- Medical history: duration of fibromyalgia (years); onset of fibromyalgia (age); presence of fibromyalgia in family history (yes/no)
- Pre-existing conditions
- Historical alcohol and tobacco consumption
- Baseline concurrent/ongoing therapies for FM symptoms: 1) NSAIDS, 2) Physical therapy, and 3) Psychotherapy
- Historical illnesses
- Onset of menses (female patients only): incidence prior to or post study entry among female patients; age of onset of first menses, defined as (date of first menses)-(date of birth)

Baseline concurrent/ongoing therapies are those therapies which are started, ongoing or stopped at baseline (visit 1 to 2). Baseline concurrent/ongoing therapies include psychotherapy, physical therapy, and NSAIDs.

Comparisons between treatment groups will be performed using Fisher's exact test for categorical data and ANOVA with treatment and pooled investigative site as independent variables in the model for continuous data for Study Period II. Patient characteristics will be summarized for Study Period III without any statistical comparison between treatment groups.

A listing of patient demographic characteristics will be presented.

The pre-existing conditions table will present the number and percent of patients with pre-existing conditions by preferred term in decreasing frequency within MedDRA system organ class. Pre-existing conditions for ITT patients in Study Period II and all ITT patients entering

Study Period III are events reported at baseline (visit 1 to 2) regardless if they are resolved or ongoing at the time of randomization.

The baseline historical illness table will summarize all significant illnesses reported at visit 1 by MedDRA preferred term and treatment group for all ITT patients for Study Period II and for those ITT patients entering Study Period III.

6.1.9. Significant Protocol Deviations

Significant protocol deviations will be listed for all Study Periods. The significant protocol deviations are defined in Table HMGW.6.2.

Table HMGW.6.2 Definition of Significant Protocol Deviations

Category of Protocol Deviation	Details	Methods of Checking
Improper administration of ICF	ICF not signed prior to initiation of protocol procedures	For all patients, if ICF (including assent form if applicable) date is after Visit 1 date or Visit 1 lab date, then result is a protocol deviation.
Inclusion/exclusion criteria	Randomized patient <13 years old at study entry	For randomized patient, if age<13 at ICF date, then result is a protocol deviation [criterion 1]
	Randomized patient >=18 years old at study entry	For randomized patient, if age >=18 at Visit 1 date, then result is protocol deviation [criterion 1]
	Randomized patient does not meet criteria for primary JPFS as defined by Yunus and Masi	For randomized patient, if JPFS criteria are not met at Visit 1, then result is protocol deviation. [criterion 2]
	Randomized patient does not have a score of ≥4 on BPI average pain severity at Visit 1 and Visit 2	For randomized patient, if BPI item #3 <4 at Visit 1 or Visit 2, then result is protocol deviation. [criterion 3]
	Randomized female patient tested positive for pregnancy	For randomized female patients, if pregnancy test is positive any time prior to Visit 2, then result is protocol deviation.

		[criterion 4]
	Randomized patient had current diagnosis of <*XXX*> at study entry	For randomized patients, use Pre-existing conditions and study adverse event module. If there exists a current diagnosis of rheumatoid arthritis, arthritis, lupus, ankylosing spondylitis, psoriasis, then result is protocol deviation. [criterion 17]
	Randomized patient had current diagnosis or diagnosis within 1 year of <*XXX*> at study entry	For randomized patients, use Pre-existing conditions and study adverse event module. If there exists a current diagnosis or diagnosis within 365 days (1 year) prior to visit 1 of bipolar disorder, psychotic depression, schizophrenia or other psychotic disorder, anorexia, bulimia, obsessive compulsive disorder, posttraumatic stress disorder, panic disorder, pervasive development disorder, then result is protocol deviation. [criterion 18]
	Randomized patient had previous diagnosis of <*XXX*> at study entry	For randomized patients, use historical illness module. If there exists a previous diagnosis of bipolar disorder, psychotic depression, schizophrenia, schizoaffective disorder or other psychotic disorder, substance abuse or dependence (excluding caffeine and nicotine) (in the past 183 days (6 months) prior to Visit1), seizure disorder excluding febrile seizures, then result is protocol deviation. [criterion 20, 22, 28]
	Baseline weight < 20 kg for randomized patient	For randomized patients, if weight < 20 kg at Visits 1 or 2, then result is protocol deviation. [criterion 26]
	Randomized patient treated with stimulant or other ADHD medication: <*medication name*> during screening.	For randomized patients, if patient has continued use of stimulant (eg, methylphenidate, amphetamine, lisdexamfetamine) or atomoxetine at Visit 1 or therapy end date after ICF date, then result is protocol deviation. [criterion 19,

		29]
	Randomized patient treated with MAOI : *<medication name>* <14 days of visit 1	For randomized patients, if DOV for visit 1 minus therapy end date for MAOI (e.g. isocarboxazid, phenelzine, selegilin, tranylcypromine) <14 days, then result is protocol deviation. [criterion 31]
	Randomized patient treated with fluoxetine <30 days of completion of visit 1	For randomized patients, if DOV for visit 1 minus therapy end date for fluoxetine <30 days, then result is protocol deviation. [criterion 30]
	Randomized patient treated with duloxetine < 6 months prior to visit 1	For randomized patients, if DOV for visit 1 minus therapy end date for duloxetine < 183 days (6 months), then result is protocol deviation. [criterion 14]
	Positive UDS prior to randomization (Visit 2).	For randomized patients if prior to visit 2 a patient has a positive UDS result and a repeated UDS not done or last repeat UDS is positive or UDS never collected, then result is protocol deviation. [criterion 23]
	Have uncontrolled narrow-angle glaucoma or acute liver injury at baseline	For randomized patients, use pre-existing/AE module. [criterion 33,34]
	Randomized patient treated with MAOI during study	For randomized patients, taking MAOI anytime > or = visit 2, then result is protocol deviation. [criterion 31]
Study Conduct – Concomitant Medications	Randomized patient reported prohibited concomitant medication use: *<concomitant medication name>*	For randomized patients, if any excluded medication (eg, SSRI, SNRI, stimulant, antipsychotic, benzodiazepine, anticonvulsant) is reported after Visit 2, then result is protocol deviation. Specify which concomitant medication in the protocol deviation details. Provide one record per patient per concomitant medication.
	Randomized patient who had consecutive use of NSAID less	For randomized patients in Study Period II, if patients had consecutive use of NSAID <

	than 1 month immediately prior to visit 2 reported *NSAID* use: <*NSAID name*> exceeded 3 consecutive or exceeded 20 cumulative days during Study Period II	30 days (1 month) immediately prior to visit 2, and reported NSAID use exceeds 3 consecutive days or exceeds 20 cumulative days after visit 2 in SP II, then result is protocol deviation. Specify which NSAID in the protocol deviation details.
	Randomized patient who had consecutive use of NSAID less than one month immediately prior to visit 8 reported NSAID use : <*NSAID name*> exceeded 3 days or exceeded 40 cumulative days during Study Period III	For randomized patients in Study Period III, if patients had consecutive use of NSAID < 30 days (1 month) immediately prior to visit 8, and reported NSAID use exceeds 3 consecutive days or exceeds 40 cumulative days after visit 8 in SP III, then result is protocol deviation. Specify which NSAID in the protocol deviation details.
	Randomized patient reported opiate use : <*opiate name*> exceeded 3 consecutive or exceeded 20 cumulative days during Study Period II	For randomized patients in study Period II, opiate use exceeds 3 consecutive days or exceeds 20 cumulative days then result is protocol deviation. Specify which opiate in the protocol deviation details.
	Randomized patient reported opiate use : <*opiate name*> exceeded 3 days/month or exceeded 40 cumulative days during Study Period III	For randomized patients in study Period III, opiate use exceeds 3 days/month or exceeds 40 cumulative days then result is protocol deviation. Specify which opiate in the protocol deviation details.
Study Conduct - Compliance	Randomized patient took <80% or >120% of study drug for >= 2 visit intervals	For randomized patients, if <80% or >120% of study drug is taken >= 2 visits (consecutive or non-consecutive), then result is protocol deviation.

Study Conduct – Key measurements not collected	Missing key efficacy measurement: <*BPI average pain score, CGI-S overall illness*>	For randomized patients, if BPI average pain severity or CGI-S overall illness is missing at any scheduled visit, then result is protocol deviation. Provide one record per patient per missing measurement per visit.
	Missing safety measurement: <*CSSRS, Blood Pressure, Pulse, Height, Weight, Labs, ECGs*>	For randomized patients, if C-SSRS, blood pressure, pulse, height, weight, are missing at any scheduled baseline or post-baseline visits at or before the early discontinuation visit, then result is protocol deviation. Provide one record per patient per missing measurement per visit. For ALT lab test, if any scheduled or unscheduled baseline labs are missing or if a patient discontinued at or after visit 6 (the first scheduled post-baseline visit for labs is visit 6), and without any scheduled or unscheduled post-baseline labs data, then it is a protocol deviation. For ECG QTcF measure, if any scheduled or unscheduled baseline ECGs are missing or if a patient discontinued at or after visit 8 (the first scheduled post-baseline visit is visit 8), and without any post-baseline data, then it is a protocol deviation.
Visit interval outside specified limits	Visit interval length exceeds (number of days study drug dispensed + 2 days) (excluding taper phase)	For randomized patients, compare actual visit interval length to number of days study drug dispensed at each visit. Allowable visit interval lengths (resulting in no more than 2 days off drug): Visit 2,3, 8: 11 days Visit 4, 9: 20 days Visit 5,6,7, 10: 29 days Visit 11: 38 days Visit 12,13: 47 days

		Visit 14: 56 days

6.1.10. Study Drug Exposure

Duration of study drug exposure in days during the acute and extension periods will be summarized by treatment and overall for all ITT patients in Study Period II and for all ITT patients entering Study Period III.

Total patient-years of exposure will also be displayed by treatment group and overall.

Duration of study drug exposure will be defined as the difference between the date of last dose of study drug in the respective study period and the date of first dose of study drug in the respective study period.

The first dose date of study drug for Study Period II is the first dose date collected on the eCRF. The first dose date for Study Period III will be one day after the visit 8 date.

The date collected on the eCRF module "date of last dose" is the last dose date of study drug for the whole study period. If the "date of last dose" collected in the eCRF is missing, then the last dose date will be imputed as the minimum of two dates: 1) the date determined by the number of days dispensed dose allowed; 2) the last non-missing disposition date.

The last dose date of study drug for Study Period II will be determined as following:

- For patients who did not enter the Study Period III or taper phase, the last dose date is the date collected in the eCRF module "date of last dose";
- For all other patients, the last dose date will be imputed as the minimum of two dates during SP II as described above.

The last dose date of study drug for Study Period III will be determined as following:

- For patients who did not enter the taper phase, the last dose date is the date collected in the eCRF module "date of last dose";
- For all other patients, the last dose date will be imputed as the minimum of two dates during SP III as described above.

Comparisons between treatment groups for duration of study drug exposure will be performed for Study Period II using an ANOVA with treatment and pooled investigative site in the model. Patient exposure durations will also be summarized categorically and assessed for treatment differences using Fisher's exact test for Study Period II.

Number and percentages of patients with doses of DLX30 and DLX60 will be tabulated at each visit for all patients randomized to duloxetine during Study Period II and all ITT patients entering Study Period III.

Number and percentages of patients with modal dose of DLX30 and DLX60 will be summarized for all patients randomized to duloxetine during Study Period II and all ITT patients entering

Study Period III. The last prescribed dose at LOCF endpoint for Study Period II and III will be also summarized using this method.

The modal dose is defined as the dose that a patient was prescribed for the most number of days during the study period. If there is a tie in the greatest number of days on a dose for a patient, the modal dose will be set to the highest dose in the tie.

The mean daily dose is calculated as the mean of all doses that a patient was prescribed during the study period. The mean of all doses is calculated as sum of dose multiplied by the number of days on that dose divided by total number of days during the study period. The mean daily dose will be summarized for all patients randomized to duloxetine at Study Period II and all ITT patients entering Study Period III.

Dose adjustment by visit will be summarized as well.

Dose reduction due to adverse events will be listed.

6.1.11. Treatment Compliance

The percentage of patients deemed compliant will be summarized by visit and overall for Study Period II, and for all patients entering Study Period III. A patient will be considered to be compliant with study drug for each interval if they take between 80% and 120% of study drug capsules prescribed for that interval. Overall compliance is defined as having been compliant at all nonmissing visits within each study period. Comparisons between treatment groups will be performed for Study Period II both overall and at each visit using Fisher's exact test. Descriptive statistics will be used to present the two treatment groups for the Study Period III results and will include only those patients who entered the study period.

A listing of study drug compliance will be presented for all study periods.

6.1.12. Previous and Concomitant Therapy

Previous therapies for fibromyalgia are those therapies for the treatment of fibromyalgia that started and stopped prior to or at baseline (Visit 2). Concomitant therapies include psychotherapy, physical therapy, and concomitant medication. Concomitant therapies for Study Period II are those which started prior to visit 2 or after visit 2 and are taken during SP II. Concomitant therapies for Study Period III are those that started prior to visit 8 or after visit 8 and are taken during SP III.

Previous therapies for fibromyalgia will be summarized by treatment group for all ITT patients, displaying the number and percentage of patients by preferred term for Study Period II and for all ITT patients entering Study Period III. Concomitant medications will be summarized in the same way for Study Period II and Study Period III, respectively. The tables will be sorted in decreasing frequency of preferred terms. The denominator used for calculating the percentages will be the total number of patients included in the ITT population for each treatment group for Study Period II, and will be the total number of patients included in the ITT patients entering Study Period III for each treatment group for Study Period III. Concomitant psychotherapy and

concomitant physical therapy will also be summarized by treatment group for Study Period II and Study Period III separately. Comparison between treatment groups will be performed using Fisher's exact test for Study Period II.

A listing of concomitant therapy will be presented for SP II and SP III.

6.1.13. Efficacy Analyses

6.1.13.1. Primary Outcome and Methodology

The primary objective of this study is to evaluate the efficacy of duloxetine 30/60 mg once daily compared with placebo on the reduction of average pain severity as measured by the BPI average pain severity rating during the Study Period II in adolescents with JPFS.

The primary efficacy analysis will be the contrast between duloxetine and placebo at the last visit in Study Period II (Visit 8, Week 13) from a MMRM analysis on change from baseline in the BPI average pain score as described in Sections 6.1.2 and 6.1.3.

6.1.13.2. Additional Secondary Analyses

Table HMGW.6.3 presents the secondary efficacy measures and derived variables that will be evaluated for Study Period II. Baseline and endpoint definitions are described above in Table HMGW.6.1.

Maintenance of effect of duloxetine 30/60 mg QD during Study Period III will be assessed using the BPI average pain severity in only the duloxetine treated acute phase responders (defined as those patients with ≥30% pain reduction from baseline on the BPI average pain severity measure at the last non-missing assessment in Study Period II). The T-statistics will be used to evaluate the maintenance of effect of duloxetine 30/60 mg QD during Study Period III. If the upper bound of the one-sided 97.5% CI of the change from baseline to endpoint for patients in the extension treatment phase who responded to duloxetine 30/60 mg QD in Study Period II (acute phase duloxetine responders) is less than or equal to the pre-specified margin of 1.5, then the treatment effect of duloxetine was maintained during the extension treatment phase. The margin of 1.5 points on the BPI average pain scale was extrapolated from data and results of prior adult duloxetine clinical studies, where a mean pain reduction of 3 to 4 points were observed for patients who met pre-defined response criteria after 2-3 months duloxetine therapy. Therefore, an increase of 1.5 indicates that the average pain at end of extension phase would still be 1.5 to 2 points below the pain severity prior to initiating acute treatment, which would be considered as clinical meaningful.

Table HMGW.6.3 Secondary Efficacy Analysis

Efficacy Variable	Derivation and Details	Analysis
1. Area under the curve of pain relief (AUC) a. BPI average pain severity	The relief score at a visit is defined as baseline score minus the BPI average pain score at the particular visit. The	Variable 1.a will be analyzed by the ANCOVA model described in Section

	area under the curve of relief (AUC) is the sum of each trapezoidal area circumscribed by the sides of relief scores at two consecutive non-missing visits and the side of days between the two visits.	6.1.2 and 6.1.3 (using baseline BPI average pain severity as covariate) for Study Period II.
2. Change from baseline to LOCF endpoint: a. BPI pain severity items: worst pain, least pain, average pain, and pain right now b. BPI-Interference items: general activity, mood, walking normally, normal work, relations with others, sleep, enjoyment of life, and school work (new item). c. BPI Average Interference score d. PPQ average pain, pain right now, and worst pain rating e. CGI-S: Mental Illness f. CGI-S: Overall Illness g. FDI-child h. FDI-parent i. CDI: total score j. MASC: total score, factor scores	a-b. eCRF data. c. Average of the non-missing ratings of the 8 individual items. d-f. eCRF data. g-h. FDI total score is the sum of 15 individual item scores. i. CDI total score is the sum of 27 individual item scores. j. MASC total score is the sum of 39 individual item scores ; the 4 factor scores are 1) physical symptoms, 2) social anxiety, 3) harm avoidance and 4) separation anxiety	The Variable 2.a to 2.j will be analyzed by the ANCOVA models as described in Section 6.1.2 and 6.1.3 for Study Period II.
3. Change from baseline to BOCF endpoint: a. BPI average pain severity b. PPQ average pain rating	a-b. eCRF data.	The Variable 3.a to 3.b will be analyzed by the ANCOVA models as described in Section 6.1.2 and 6.1.3 for Study Period II.
4. Change from baseline to each post-baseline visit: a. BPI pain severity items: worst pain, least pain, and pain right now b. BPI-Interference items: general activity, mood, walking normally, normal work, relations with others, sleep, enjoyment of life, and school work (new item). c. BPI Average Interference score d. CGI-Severity: Overall Illness e. CGI-Severity: Mental Illness	a-b. eCRF data. c. Average of the non-missing ratings of the 8 individual items. d-e. eCRF data.	Variables 4.a to 4.e will be analyzed by a repeated measures analysis as described in Section 6.1.2 and 6.1.3 for Study Period II.
5. Categorical variable: a. 30% Response rate (LOCF) b. 30% Response rate (BOCF) c. 50% Response rate (LOCF) d. 50% Response rate (BOCF) e. 30% Response rate (MMRM-	a-b. Response: at least 30% reduction from baseline to endpoint (LOCF or BOCF) for BPI average pain score. c-d. Response: at least 50% reduction from baseline to endpoint (LOCF or BOCF) for BPI average pain score.	For Variables 5.a to 5.d, and 5.g, proportions will be summarized by treatment group and will be analyzed by Fisher's exact test and CMH test controlling for

CAT) f. 50% Response rate (MMRM-CAT) g. Sustained response rate h. Cumulative distribution of BPI average pain score reduction (BOCF)	e. Response at each post-baseline visit: at least 30% reduction from baseline. f. Response at each post-baseline visit: at least 50% reduction from baseline. g. Sustained response: at least 30% reduction from baseline to endpoint; with a 30% reduction from baseline at an earlier visit than the last visit, and remains at least 20% reduction from baseline in every visit in between, if there are any intervening visits (based on BPI average pain score). The sustained response analyses will include patients with both a baseline and at least two post-baseline values for BPI average pain score. h. The percentage of patients who have reached each threshold of BPI average pain reduction from baseline to BOCF endpoint (from >0% to 100% with a 10% increase) will be calculated. Discontinued patients will be considered as "no change".	pooled investigative site for Study Period II. For Variable 5.e and 5.f, the categorical, pseudo-likelihood-based repeated measures (MMRM-CAT) analysis will be used. See section 6.1.2 for details. For Variable 5.h, the treatment group difference in the empirical cumulative distribution of the percentage pain reduction will be evaluated using Van der Waerden test for Study Period II.
6. Time to event variable: a. Time to first 30% reduction in BPI average pain score b. Time to first 50% reduction in BPI average pain score c. Time to sustained response	a. For the patients with a 30% reduction at a visit in the treatment phase in SP II, time = days from the date of the visit that the earliest 30% reduction is observed to the date of the first day of SP II. The date of the first day of SP II is the randomization date. b. For the patients with a 50% reduction at a visit in the treatment phase in SP II, time = days from the date of the visit that the earliest 50% reduction is observed to the date of the first day of SP II. c. For the sustained responders defined above in SP II, time = the days from the date of the visit which is the earlier visit from which the sustained response is observed to the date of the first day of SP II.	For Variables 6.a to 6.c, the Kaplan-Meier survival curves of time to event will be calculated by treatment group for Study Period II. In the calculation, patients who do not have the event will be considered as right-censored observation. The comparison of the survival curves between treatment groups will be conducted by a log-rank test and stratified log-rank test controlling for pooled investigator.

Abbreviations: ANCOVA = analysis of covariance; BOCF = baseline observation carried forward; BPI = Brief Pain Inventory-Adolescent Version; CDI = Children's Depression Inventory; CGI-S = Clinical Global Impressions of Severity; eCRF = electronic case report form; FDI = Functional Disability Inventory; LOCF = last observation carried forward; MASC = Multidimensional Anxiety Scale for Children; PPQ = Pediatric Pain Questionnaire.

The treatment-by-investigator interaction will be tested using an ANCOVA model. When the interaction is statistically significant, the nature of the interaction will be investigated.

Descriptive statistics will be used to summarize all variables presented in HMGW.6.3 except variables 3a-b, 5(b, d, h) and 6a-c by treatment group for all ITT patients entering Study Period III. Descriptive statistics will be used to summarize variables 1a, 2a-f, and 4a-e for Study Period II/III. For continuous measures, when appropriate, within treatment testing of change from baseline will be conducted for SP III and SP II/III.

For efficacy measures, data from all visits will be listed.

6.1.13.3. Sensitivity Analysis

To assess the possible impact of missing data on the primary efficacy analysis conclusions, a sensitivity analysis will be performed leveraging a delta based approach. The approach for this analysis is to vary the assumptions of missing data for the primary analysis in a systematic way. Basically, the method will be to predict the missing outcomes and then add a value (Δ) to the predictions in the duloxetine treatment group consistent with the sensitivity approach suggested in Permutt (2015). This procedure will be repeated multiple times for different values of Δ using the following steps:

1) Predict the missing outcomes for each treatment via multiple imputation based on observed primary endpoint and baseline values. Such imputation will be carried out using a Markov Chain Monte Carlo method with a Jeffreys prior via SAS® PROC MI. Thirty (30) such imputations will be created.
2) Add the corresponding Δ value to the imputed duloxetine treatment values.
3) Conduct the primary analysis separately for each of the 30 imputations.
4) Combine the results of these analyses using Rubin's combining rules, as implemented in SAS® PROC MI ANALYZE.

The above steps will be repeated multiple times for different values of Δ with Δ ranging from (0, twice the observed treatment effect seen in the primary analysis). For example, if the overall mean change from baseline for placebo is -3.6 and the maximum overall treatment difference is -1.5, then Δ would range from (0,7.2).

6.1.14. Safety Analyses

The safety and tolerability of treatment with duloxetine will be assessed by summarizing the following:

- Treatment-emergent adverse events (TEAEs)

- Serious Adverse Events (SAEs)
- Vital signs
- Height
- Weight
- Rates and reasons for early discontinuation
- Laboratory measurements and electrocardiograms (ECG)
- Suicide risk and suicide-related events (behavior and/or ideation) as assessed by the Columbia-Suicide Severity Rating Scale (C-SSRS)

For Study Period II, statistical comparisons will be conducted to compare treatment groups. For Study Period III, descriptive statistics only will be presented for each treatment group.

The baseline periods for all safety data are as described for safety variables in Table HMGW.6.1 unless otherwise specified.

6.1.14.1. Adverse Events

All adverse events (AEs) will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). In this thesaurus, each verbatim term is mapped to a "preferred" MedDRA term, which is then mapped to a system organ class.

An adverse event is treatment emergent if it is newly occurring or worsened in severity during post-baseline compared with baseline. The evaluations of adverse events will include separate summaries for Study Periods II, III, and IV of the number and percentage of patients with treatment-emergent adverse events (TEAEs), serious adverse events (SAEs) including deaths, and discontinuations due to AEs. Also, the TEAEs for patients who are initially randomized to duloxetine will be summarized for SP II/III. The TEAEs will be summarized by severity and MedDRA preferred term. The TEAEs will be summarized by system organ class and MedDRA preferred term. Adverse events will be summarized by MedDRA preferred term.

Any event having an onset date within the post-baseline range of its Study Period is a TEAE. For each event, the severity level of an event at first report and the change in severity level is recorded according to the patient's or physician's perceived severity of the event (mild, moderate, or severe). In addition, a change in severity to more severe than that recorded during the baseline period can be recorded during the study. For each event, the maximum severity during visits for baseline will be used as the baseline severity. If the maximum severity during post-baseline visits for the study period is greater than the baseline severity, the event is considered to be treatment-emergent.

For Study Period II, an adverse event is treatment emergent if the event onset date falls after visit 2 and prior to visit date of visit 8. An event which was ongoing during the baseline period becomes treatment emergent if it worsens in severity after visit 2 and prior to visit date of visit 8, compared with the maximum severity prior to visit 2. For Study Period III, unless stated otherwise, an adverse event is treatment emergent if the onset date falls after visit 8 and prior to visit date of visit 15. An event which was ongoing during the earlier study periods becomes treatment emergent if it worsens in severity after visit 8 and prior to visit date of visit 15,

compared with the maximum severity between visit 7 and 8 during Study Period II. For Study Period II, all SAEs from after visit 2 to visit 8 will be summarized. For Study Period III, all SAEs from after visit 8 to visit 15 will be summarized. For Study Period IV, all SAEs during the taper phase will be summarized.

Discontinuation emergent AEs will be summarized for patients entering Study Period IV from Study Period II and for patients entering Study Period IV from Study Period III. For these analyses, the baseline period will be the last two visits during the previous study period. Any AEs with an onset or worsening severity on or after the start of Study Period IV, compared to the maximum severity during the baseline period, will be considered discontinuation emergent for the taper phase.

Listings for AEs and pre-existing conditions, SAEs, discontinuations due to AEs or death will be presented.

The Kaplan-Meier survival curve for time to onset of the most common TEAEs during SP II/III will be calculated for patients who are initially randomized to duloxetine. In the calculation, patients who do not have the event will be considered as right-censored observation. The same Kaplan-Meier survival curve will be calculated for time to discontinuation due to adverse event.

6.1.14.2. Clinical Laboratory Evaluation

The incidence rates of patients with treatment-emergent abnormal, high or low laboratory values at endpoint (LOCF) will be summarized for SP II, SP II and SP II/II, and compared between treatment groups for SP II. Listings of abnormal lab values and treatment-emergent abnormal lab values will also be presented.

A "treatment-emergent abnormal value at endpoint" is defined as a change from normal at all visits during baseline to abnormal at endpoint. A "treatment-emergent high value at endpoint" is defined as a change from a value less than or equal to the high limit at all visits during baseline to a value greater than the high limit at endpoint. A "treatment-emergent low value at endpoint" is defined as a change from a value greater than or equal to the low limit at all visits during baseline to a value less than the low limit at endpoint. High and low reference ranges for adolescent patients will be used to determine treatment emergent abnormal, high, low values. These reference ranges will be provided by a central laboratory. The same analysis will be done for "treatment-emergent abnormal value at anytime", "treatment-emergent high value at any time", and "treatment-emergent low value at any time".

Treatment emergent liver function abnormalities at any time post-baseline for patients with normal ALT values (ALT <1 time upper limit normal) at last non-missing baseline will be summarized for SP II, SP III and SP II/III, and compared between treatment groups for SP II for the following categories:

ALT >= 3 times upper limit normal

ALT>=5 times upper limit normal

ALT>=10 times upper limit normal

Hy's Rule, ALT >= 3 times upper limit normal and total bilirubin >=2 times upper limit normal.

A listing of patients with ALT >= 3 times upper limit normal at any time during post-baseline will be presented. And, a listing of all laboratory values will be presented.

For the continuous laboratory analytes, change from baseline to endpoint will be assessed using an ANOVA model (see Section 6.1.2 and 6.1.3) for Study Period II, and will be summarized for Study Period III and Study Period II/III. Rank transformed data for lab analytes will be used for the laboratory analysis. Within treatment testing of change from baseline will be based on a Wilcoxon Signed-Rank Test.

6.1.14.3. Vital Signs and Other Physical Findings

In this study, the vital signs (blood pressure and pulse) are collected in triplicate at every visit, with readings at least 3 minutes apart. At each visit, the mean of this triplicate will be used as the blood pressure or pulse value for that visit. For categorical analysis, the lowest value during baseline (mean value of triplicates from each visit) will be used to determine treatment-emergent or potentially clinically significant low value; the highest value during baseline (mean value of triplicates from each visit) will be used to determine potentially clinically significant high values.

The incidence of patients with treatment-emergent low blood pressure at endpoint (LOCF) and at any time will be summarized for SP II, SP III, and SP II/III, and compared between treatment groups for SP II. Table HMGW.6.5 provides the criteria for treatment emergent low blood pressure for the relevant age categories. The baselines and post-baseline of blood pressure for Study Period II, Study Period III and Study Period II/III are as described in Table HMGW.6.1. Patients with low blood pressure at baseline will be excluded from the analysis, and only patients with normal baseline values with respect to the direction will be included in the analysis. The normal limits of vital signs at baseline are as described in Table HMGW.6.4.

Table HMGW.6.4 Normal Limits of Vital Signs at Baseline

Parameter	Age (Years)	Normal Limits for treatment emergent Low	Normal Limits for treatment emergent or PCS high
Pulse (bpm)	Adolescent (13-17)	>= 50	<= 120
Diastolic BP (mm Hg)	Adolescent (13-17)	> 50	Value <= 95th percentile (based on age, gender, and height)
Systolic BP (mm Hg)	Adolescent (13-17)	> 90	Value <= 95th percentile (based on age, gender, and height)

Table HMGW.6.5 Criteria to Identify Patients with Treatment-Emergent Abnormalities in Vital Signs.

Parameter	Age (Years)	Low Value	High Value
Diastolic BP (mm Hg)	Adolescent (13-17)	≤ 50 and decrease of ≥ 10	NA
Systolic BP (mm Hg)	Adolescent (13-17)	≤ 90 and decrease of ≥ 20	NA

Table HMGW.6.6 Criteria to Identify Patients with Potentially Clinically Significant Abnormalities in Vital Signs.

Parameter	Low Value	High Value
Pulse (bpm) Adolescents (13-17)	<50 and a decrease of ≥ 15	>120 and increase of ≥ 15
Weight	Decrease of at least 3.5% from baseline low value	N/A
Diastolic BP (mm Hg)	N/A	Value > 95th percentile* (based on age, gender, and height) AND Increase ≥ 5 mm Hg from baseline
Systolic BP (mm Hg)	N/A	Value > 95th percentile* (based on age, gender, and height) AND Increase ≥ 5 mm Hg from baseline

N/A = Not applicable

*National High Blood Pressure Education Program Working Group. The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents. *Pediatrics* 2004 August;114(2 Suppl 4th Report):555-76.

The incidence of patients meeting criteria for potentially clinically significant (PCS) values in vital signs and weight at endpoint (LOCF) and at any time will be summarized for SP II, SP III and SP II/III, and compared between treatment groups for SP II. Patients with abnormal vital signs at baseline with respect to direction being analyzed will be excluded from the analysis. Table HMGW 6.4 displays the normal limits of vital signs at baseline. In addition, the number of patients with abnormal blood pressure values at baseline will be summarized. Potentially

clinically significant changes in vital signs will be based on criteria shown in Table HMGW.6.6. To calculate the percentiles of blood pressure value for each age/gender/height percentile combination, the following steps will be followed:

1. Height percentiles for each age and gender combination will be calculated first based on the most recent CDC growth charts. The SAS code from CDC website (http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm) will be utilized for the height percentile calculation.

   a. First, L (power in the Box-Cox transformation), M (Mean) and S (Coefficient of variation) will be extracted from 2000 CDC growth charts for United States for height.
   b. Z-score of height with corresponding age in months will be calculated using the following formula:

$$Zscore = \frac{(\frac{Height}{M})^{L} - 1}{L * S}$$

   c. Percentile of height will be the probability that an observation from the standard normal distribution is less than or equal to Z-score.

2. The blood pressure percentile for each age, gender and height percentile combination will be calculated as follows (National Heart, Lung, and Blood Institute's National High Blood Pressure Education Program Working Group on High Blood Pressure in Children and Adolescents (2004)).

   a. Compute the expected blood pressure values (μ) for kids of age y years and height Z-score(Zht) given by

$$\mu = \alpha + \sum_{j=1}^{4} \beta_j (y-10)^j + \sum_{k=1}^{4} \gamma_j (Zht)^k$$

   Where $\alpha, \beta_1, \ldots\ldots \beta_4 and \gamma_1, \ldots\ldots \gamma_4$ are given in the 3rd column of appendix table B-1 of the fourth report of the diagnosis, evaluation and Treatment of High Blood Pressure in Children and Adolescents (2004).

   b. Z-score of blood pressure is given by :

$$Zbp = \frac{(x-\mu)}{\sigma}$$

   Where σ is given in the 3rd column of appendix table B-1 of the fourth report of the diagnosis, evaluation and Treatment of High Blood Pressure in Children and Adolescents (2004).

   c. Percentile of blood pressure will be the probability that an observation from the standard normal distribution is less than or equal to Z-score.

The incidence of patients with sustained elevated blood pressure will be summarized for SP II, SP III, and SP II/III, and compared between treatment groups for SP II. Blood pressure has sustained elevation if patient's blood pressure value meets PCS criteria at 3 consecutive post-baseline visits. PCS criteria are outlined at Table HMGW 6.6.

All vital signs data, patients with sustained blood pressure elevation, and patients with PCS vital signs will be presented in separate listings.

Change from baseline in vital signs (blood pressure and pulse), weight, and height data will be analyzed using an ANCOVA model with baseline values as covariates (see Sections 6.1.2 and 6.1.3). Raw values will be used for these analyses unless normality assumptions appear to be violated in which case the data will be rank transformed. These analyses will include summaries of baseline values, endpoint (LOCF) values, and change from baseline to endpoint (LOCF) by treatment group for SP II, SP III and SP II/III separately. Change from baseline in vital signs (blood pressure and pulse) and weight will also be summarized for SP IV. For SP IV, change from baseline in vital signs and weight data will be analyzed using a simple linear regression model with change from baseline as the response variable and the baseline value as the independent variable.

Change from baseline in vital signs (blood pressure and pulse) and weight will also be analyzed using an MMRM model (see Section 6.1.2) for SP II, SPIII and SP II/III separately. Within treatment testing of change from baseline will be conducted.

6.1.14.4. Electrocardiograms

In this study, 12-lead ECGs will be obtained in triplicate at defined study visits (Visits 1, 8, 12, and 15/discontinuation), with each measurement done approximately 1 minute apart.

The ECGs recordings will be electronically transmitted to a centralized ECG vendor who will complete the ECG overread. All three of the ECG recordings will be overread by the vendor. For patient management purposes during the study, the vendor will send the study site a full overread/interpretation for only one of the triplicate ECG recordings (the first readable) obtained at each visit. For data analysis purpose, the vendor overread of QT, PR, RR and QRS measurements (the non-missing values) from all 3 ECG recordings will be averaged, and these averaged values will represent the ECG results for each patient/visit. The QTc result for each ECG recording will be determined using the HR, RR and QT intervals and then the average QTc will be calculated.

The incidence of patients meeting criteria for treatment-emergent abnormal values at endpoint (LOCF) and at any time will be summarized for SP II, SP III and SP II/III, and compared between treatment groups for SP II. The abnormality criteria are presented in Table HMGW.6.7. In those analyses, only patients with normal values at all baseline visits with respect to direction being analyzed will be included. The normal limits at baseline are described in Table HMGW 6.8. The Fridericia's corrected QT interval (QTcF) (msec) will be calculated as $QT/RR^{1/3}$. The Bazette's corrected QT interval (QTcB) (msec) will be calculated as $QT/RR^{1/2}$.

In addition to categorical analyses of the QTc intervals based on absolute values, it is customary (and expected by regulatory bodies, particularly the Committee for Proprietary medicinal Products [CPMP]) that the proportion of subjects with increases from baseline above certain thresholds also be analyzed. These thresholds are based on estimates of normal variance (versus drug-induced changes) in QTc intervals. These thresholds are:

>=30 msec

>=60 msec

>=75 msec

A listing of patients with treatment-emergent abnormal values using the criteria in Table HMGW.6.7 will be presented.

Table HMGW. 6.7 Criteria to Identify Patients with Treatment-Emergent Abnormalities in ECG.

Parameter	Low Value	High Value
Heart Rate (bpm) Adolescents (13-17)	<50 bpm	>110 bpm
PR Interval	N/A	>220 msec
QRS Interval	N/A	>120msec
QTc Bazett – Female	N/A	> = 470 msec
QTc Bazett – Male	N/A	> = 450 msec
QTc Bazett – Male or Female	N/A	> 40 msec increase from baseline
QTc Fridericia– Female	N/A	> = 470 msec
QTc Fridericia – Male	N/A	> = 450 msec
QTc Fridericia – Male or Female	N/A	> 40 msec increase from baseline

N/A = Not applicable

Table HMGW.6.8 Normal Limits at Baseline for ECG Analyses

Parameter	Normal Limits for treatment emergent or PCS Low	Normal Limits for treatment emergent or PCS High

Heart Rate (bpm) Adolescents (13-17)	>=50 bpm	<=110 bpm
PR Interval	N/A	<=220 msec
QRS Interval	N/A	<=120msec
QTc Bazett – Female QTc Bazett – Male	N/A N/A	< 470 msec < 450 msec
QTc Fridericia– Female QTc Fridericia – Male	N/A N/A	< 470 msec < 450 msec
PCS QTc Bazett and QTc Fridericia	N/A	<=500 msec

The incidence of patients with PCS high QTc Bazett and QTc Fridericia at any time and at endpoint, defined as value greater than 500 msec, will be summarized for SP II, SP III and SP II/III. In these analyses, only patients with a value less than or equal to 500 msec at all baseline visits will be included.

Change from baseline in ECG data will be analyzed using an ANCOVA model with baseline values as covariates for SP II, SP III and SP II/III. Raw values will be used for these analyses unless normality assumptions appear to be violated in which case the data will be rank transformed. These analyses will include summaries of baseline values, endpoint (LOCF) values, and change from baseline to endpoint (LOCF) by treatment group.

6.1.14.5. Columbia Suicidal-Severity Rating Scale

Suicide-related thoughts and behaviors, based on the Columbia-Suicide Severity Rating Scale (C-SSRS), will be summarized by treatment group for Study Period II, Study Period III, and Study Period IV. In particular, for each of the following suicide-related events, the number and percent of patients with the event will be enumerated by treatment group: completed suicide, nonfatal suicide attempt, interrupted attempt, aborted attempt, preparatory acts or behavior, active suicidal ideation with specific plan and intent, active suicidal ideation with some intent to act without specific plan, active suicidal ideation with any methods (no plan) without intent to act, nonspecific active suicidal thoughts, and wish to be dead. In addition, the number and percent of patients who experienced at least one of various composite measures will be presented. These include suicidal behavior (completed suicide, non-fatal suicidal attempts, interrupted attempts, aborted attempts, and preparatory acts or behavior), suicidal act (completed suicide, non-fatal suicidal attempts), suicidal ideation [active suicidal ideation with specific plan and intent, active suicidal ideation with some intent to act without specific plan, active suicidal

ideation with any methods (no plan) without intent to act, non-specific active suicidal thoughts, and wish to be dead]. Table HMGW.6.9 is a shell of the format that will be followed for Study Period II. Similar tables but without treatment comparison will be created for Study Period III and Study Period IV.

Specifically, the following outcomes are C-SSRS categories and have binary responses (yes/no). The categories have been re-ordered from the actual scale in an increasing order of severity from 1 to 10 to facilitate the definitions of the comparative endpoints.

- Category 1 – Wish to be Dead
- Category 2 – Non-specific Active Suicidal Thoughts
- Category 3 – Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act
- Category 4 – Active Suicidal Ideation with Some Intent to Act, without Specific Plan
- Category 5 – Active Suicidal Ideation with Specific Plan and Intent
- Category 6 – Preparatory Acts or Behavior
- Category 7 – Aborted Attempt
- Category 8 – Interrupted Attempt
- Category 9 – Actual Attempt (non-fatal)
- Category 10 – Completed Suicide

The following outcomes are numerical scores derived from the C-SSRS categories. The scores are created at each assessment for each patient.

- **Suicidal Ideation Score:** The maximum suicidal ideation category (1-5 on the C-SSRS) present at the assessment. Assign a score of 0 if no ideation is present.
- **Suicidal Behavior Score:** The maximum suicidal behavior category (6-10 on the C-SSRS) present at the assessment. Assign a score of 0 if no behavior is present.
- **Suicidal Act Score**: The maximum suicidal act category (9-10 on the C-SSRS) present at the assessment. Assign a score of 0 if no act is present.
- **Suicidal Ideation or Behavior Score:** The maximum suicidal ideation or behavior category (1-10 on the C-SSRS) present at the assessment. Assign a score of 0 if no ideation or behavior is present.

Composite endpoints based on the above categories are defined below.

- **Suicidal ideation:** A "yes" answer at any time during treatment to any one of the five suicidal ideation questions (Categories 1-5) on the C-SSRS.
- **Suicidal behavior:** A "yes" answer at any time during treatment to any one of the five suicidal behavior questions (Categories 6-10) on the C-SSRS.
- **Suicidal act:** A "yes" answer at any time during treatment to any one of the two suicidal act questions (Categories 9-10) on the C-SSRS.
- **Suicidal ideation or behavior:** A "yes" answer at any time during treatment to any one of the ten suicidal ideation and behavior questions (Categories 1-10) on the C-SSRS.

In addition, non-suicidal self-injurious behavior will be summarized and compared between treatment groups as shown in Table HMGW.6.9.

Also, the number and percent of patients with at least one of the following composite measures will be presented by treatment group for Study Period II and Study Period III separately: treatment-emergent suicidal ideation, treatment-emergent suicidal behavior, emergence of serious suicidal ideation, improvement in suicidal ideation at endpoint, treatment-emergent suicidal behavior or ideation, and treatment-emergent non-suicidal self-injurious behavior. Taper emergent suicide-related events in suicidal ideation or behavior, and taper emergent non-suicidal self-injurious behavior during taper phase will also be summarized. Table HMGW.6.10 is a shell of the format that will be followed. Treatment-emergent events will be assessed in 2 ways: compared to lifetime baseline and compared to lead-in baseline. The definitions of lifetime and lead-in baseline for Study Period II are as following and also described in Table HMGW.6.1. For details of baseline definitions for Study Period III and study Period IV, see table HMGW.6.1.

Lifetime Baseline = suicide information collected on the Columbia-Suicide Severity Rating Scale (C-SSRS) Baseline form plus information collected on the C-SSRS "Since Last Visit" form during any lead-in period prior to randomization. Lifetime baseline is used to capture all the prior history.

Lead-in Baseline = information collected on the C-SSRS "Since Last Visit" form at all visits prior to randomization. Lead-in baseline is used to capture only recent history

Treatment emergent/taper emergent suicidal ideation (category 1-5) compared to lifetime (or lead-in) baseline is defined as an increase in maximum suicidal ideation over lifetime (or lead-in) baseline during the study period, or any ideation during the study period, if there was none at lifetime (or lead-in) baseline. Treatment emergent/taper emergent suicidal behavior (category 6-10) compared to lifetime (or lead-in) baseline is defined as an increase in maximum suicidal behavior over lifetime (or lead-in) baseline during the study period, or any behavior during the study period, if there was none at lifetime (or lead-in) baseline. Emergence of serious suicidal ideation compared to lifetime (or lead-in) baseline is defined as an increase in the maximum suicidal ideation score to 4 or 5 during study period from no suicidal ideation (scores of 0) at lifetime (or lead-in) baseline. Improvement in suicidal ideation at endpoint compared to baseline is defined as a decrease in suicidal ideation score at endpoint (the last nonmissing measurement during study period) from the lead-in baseline. Treatment-emergent suicidal ideation or behavior compared to lifetime (lead-in) baseline is defined as an increase in maximum suicidal ideation or behavior over lifetime (or lead-in) baseline during the study period, or any ideation or behavior during the study period, if there was none at lifetime (or lead-in) baseline.

If the number of patients with post-baseline suicide-related events is greater than or equal to 4, then shift tables will be also be used to summarize the data. Table HMGW.6.11 and Table HMGW.6.12 are shells of format that show the number and percentage of patients with shifts between the most severe event category at baseline and the most severe event category post-baseline by treatment group, for Study Period II and lifetime baseline. Similar tables will be created for Study Period II and lead-in baseline, and Study Period III with both baselines if needed.

Subjects who discontinued from the study with no post-baseline C-SSRS value will be considered unevaluable for analyses of suicidality. Only evaluable subjects will be considered in the analyses.

Fisher's exact test will be used for treatment comparison for Study Period II. For each event, p-values will only be displayed if at least 4 events occurred in at least one treatment group.

A listing of patients with suicidal ideation or behavior will be displayed. For patients with suicidal ideation or behavior at any time, data from all visits are displayed. See the listing HMGW.6.13.

Table HMGW.6.9 Number of Patients with Suicide-Related Events Based on the C-SSRS During Treatment; All Randomized Patients

Events during treatment	Duloxetine N=xx n (%)	Placebo N=xx n (%)	p-values[a] (to compare percentages)
Suicidal Ideation (1-5)	x (%)	x (%)	0.xxx
1 – Wish to be dead	x (%)	x (%)	
2 – Non-specific active suicidal thoughts	x (%)	x (%)	
3 – Active suicidal ideation with any methods (not plan) without intent to act	x (%)	x (%)	
4 – Active suicidal ideation with some intent to act, without specific plan	x (%)	x (%)	
5 – Active suicidal ideation with specific plan and intent	x (%)	x (%)	
Suicidal Behavior (6-10)	x (%)	x (%)	0.xxx
6 – Preparatory acts or behavior	x (%)	x (%)	
7 – Aborted attempt	x (%)	x (%)	
8 – Interrupted attempt	x (%)	x (%)	
Suicidal Act (9-10)			
9 – Non-fatal suicide attempt	x (%)	x (%)	
10 – Completed suicide	x (%)	x (%)	
Suicidal Ideation or Behavior (1-10)	x (%)	x (%)	0.xxx

Non-suicidal Self-injurious	x (%)	x (%)	0.xxx

[a] p-values are from Fisher's exact test.
Notes: N = number of enrolled patients with at least one post-baseline C-SSRS assessment. In this table, n and (%) refer to the number and percent of patients who experience the event at least once during treatment. For the composite endpoint of suicidal ideation (1-5), n and (%) refer to the number and percent of patients who experience any one of the five suicidal ideation events at least once during treatment. For the composite endpoint of suicidal behavior (6-10), n and (%) refer to the number and percent of patients who experience any one of the five suicidal behavior events at least once during treatment. For the composite endpoint of suicidal ideation or behavior (1-10), n and (%) refer to the number and percent of patients who experience any one of the ten suicidal ideation or behavior events at least once during treatment.

Table HMGW.6.10 Number of Patients with Suicide-Related Treatment-Emergent Events Based on the C-SSRS During Treatment; All Randomized Patients

	Duloxetine		Placebo		
Treatment-emergent (TE) Events	**N**	**n (%)**	**N**	**n (%)**	**p-values[a]**
TE Suicidal ideation (1-5) compared to lifetime baseline[b]	xx	x (%)	xx	x (%)	0.xxx
TE Suicidal ideation (1-5) compared to lead-in baseline[b]	xx	x (%)	xx	x (%)	0.xxx
Emergence of serious suicidal ideation (4-5) compared to lifetime baseline[c]	xx	x (%)	xx	x (%)	0.xxx
Emergence of serious suicidal ideation (4-5) compared to lead-in baseline[c]	xx	x (%)	xx	x (%)	0.xxx
Improvement in suicidal ideation at endpoint compared with lead-in baseline[d]	xx	x (%)	xx	x (%)	0.xxx
TE Suicidal behavior (6-10) compared to lifetime baseline[e]	xx	x (%)	xx	x (%)	0.xxx
TE Suicidal behavior (6-10) compared to lead-in baseline[e]	xx	x (%)	xx	x (%)	0.xxx
TE Suicidal ideation or behavior compared to lifetime baseline[f]	xx	x (%)	xx	x (%)	0.xxx

TE Suicidal ideation or behavior compared to lead-in baseline[f]	xx	x (%)	xx	x (%)	0.xxx
TE Non-suicidal self-injurious compared to lifetime baseline[g]	xx	x (%)	xx	x (%)	0.xxx
TE Non-suicidal self-injurious compared to lead-in baseline[g]	xx	x (%)	xx	x (%)	0.xxx

[a] p-values are from Fisher's exact test.
[b] N=Number of randomized patients with at least one post-baseline suicidal ideation score and whose maximum C-SSRS suicidal ideation score during the comparison period is non-missing and <5.
[c] N=Number of randomized patients with at least one post-baseline suicidal ideation score and whose maximum C-SSRS suicidal ideation score during the comparison period is 0.
[d] N=Number of randomized patients with at least one post-baseline suicidal ideation score and whose suicidal ideation score is non-missing and >0 just prior to treatment.
[e] N=Number of randomized patients with baseline and at least one post-baseline C-SSRS suicidal behavior score.
[f] N=Number of randomized patients with baseline and at least one post-baseline C-SSRS suicidal ideation or behavior score.
[g] N=number of randomized patients without non-suicidal self-injurious behavior at baseline and with at least one post-baseline non-suicidal self-injurious behavior score.

Notes: For the composite endpoint of suicidal ideation (1-5), n and (%) refer to the number and percent of patients who experience treatment-emergent suicidal ideation during treatment. For the composite endpoint of suicidal behavior (6-10), n and (%) refer to the number and percent of patients who experience treatment-emergent suicidal behavior during treatment.

Table HMGW.6.11 Shift-table to Demonstrate Changes in C-SSRS Categories from Baseline During Treatment; All Randomized Patients

Treatment	Maximum Baseline Category	Maximum During Treatment		
		No suicidal ideation or behavior n (%)	Suicidal ideation n (%)	Suicidal behavior n (%)
Duloxetine (N=xxx)	No suicidal ideation or behavior	x (%)	x (%)	x (%)
	Suicidal Ideation	x (%)	x (%)	x (%)
	Suicidal Behavior	x (%)	x (%)	x (%)
Placebo (N=xxx)	No suicidal ideation or behavior	x (%)	x (%)	x (%)

	Suicidal Ideation	x (%)	x (%)	x (%)
	Suicidal Behavior	x (%)	x (%)	x (%)

Notes: N = number of patients with a baseline and post-baseline C-SSRS assessment, n = number of patients in category. % = 100*n/N.

Baseline refers to the screening period; Maximum refers to the category associated with the maximum C-SSRS suicidal ideation or behavior score during treatment (0 = least severe, 10 = most severe) where 0=No Suicidal Ideation or Behavior, 1=Wish to be Dead, 2=Non-specific Active Suicidal Thoughts, 3=Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act, 4=Active Suicidal Ideation with Some Intent to Act, without Specific Plan, 5=Active Suicidal Ideation with Specific Plan and Intent, 6=Preparatory Acts or Behavior, 7=Aborted Attempt, 8= Interrupted Attempt, 9=Actual Attempt (non-fatal), 10=Completed Suicide.

Table HMGW.6.12 Shift-table to Demonstrate Changes in C-SSRS Suicidal Ideation Scores from Baseline During Treatment; All Randomized Patients

Treatment	Maximum Baseline Score	Maximum Suicidal Ideation Score During Treatment					
		0 n (%)	1 n (%)	2 n (%)	3 n (%)	4 n (%)	5 n (%)
Duloxetine (N=xxx)	0	x (%)	x (%)	x (%)	x (%)	x (%)	x (%)
	1	x (%)	x (%)	x (%)	x (%)	x (%)	x (%)
	2	x (%)	x (%)	x (%)	x (%)	x (%)	x (%)
	3	x (%)	x (%)	x (%)	x (%)	x (%)	x (%)
	4	x (%)	x (%)	x (%)	x (%)	x (%)	x (%)
	5	x (%)	x (%)	x (%)	x (%)	x (%)	x (%)
Placebo (N=xxx)	0	x (%)	x (%)	x (%)	x (%)	x (%)	x (%)
	1	x (%)	x (%)	x (%)	x (%)	x (%)	x (%)
	2	x (%)	x (%)	x (%)	x (%)	x (%)	x (%)
	3	x (%)	x (%)	x (%)	x (%)	x (%)	x (%)
	4	x (%)	x (%)	x (%)	x (%)	x (%)	x (%)
	5	x (%)	x (%)	x (%)	x (%)	x (%)	x (%)

Notes: N = number of patients with a baseline and post-baseline C-SSRS suicidal ideation score, n = number of patients in category, % = 100*n/N.

Baseline refers to screening period; Maximum refers to the maximum C-SSRS suicidal ideation score during treatment (0 = least severe, 5 = most severe) where 0=No Suicidal Ideation, 1=Wish to be Dead, 2=Non-specific Active Suicidal Thoughts, 3=Active

Suicidal Ideation with Any Methods (Not Plan) without Intent to Act, 4=Active Suicidal Ideation with Some Intent to Act, without Specific Plan, and 5=Active Suicidal Ideation with Specific Plan and Intent.

Listing HMGW.6.13 Listing of C-SSRS Suicidal Ideation and Behavior Data[a] and Non-suicidal Self-injurious Data; All Randomized Patients; Study Period II ,III and IV.

				Suicidal Ideation					Suicidal Behavior					Non-suicidal Self-injurious
Investigator ID	Patient ID	Treatment	Visit	1	2	3	4	5	6	7	8	9	10	
	xxxx			Yes	Yes	Yes	No	Yes	No	No	No	No	No	

Note: Only patients with suicidal ideation or behavior are displayed. For patients with suicidal ideation or behavior at any time, data from all visits are displayed.

[a] Key: 1=Wish to be Dead, 2=Non-specific Active Suicidal Thoughts, 3=Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act, 4=Active Suicidal Ideation with Some Intent to Act, without Specific Plan, 5=Active Suicidal Ideation with Specific Plan and Intent, 6=Preparatory Acts or Behavior, 7=Aborted Attempt, 8= Interrupted Attempt, 9=Actual Attempt (non-fatal), 10=Completed Suicide.

6.1.15. Subgroup Analyses

Some of the baseline severity of illness will be summarized and analyzed by comorbid diagnosis subgroup, which includes: CDI by MDD and non-MDD, CGI-Severity: Mental Illness by MDD/GAD and non-MDD/GAD, and MASC by GAD and non-GAD.

Suicide-related ideation and behavior, treatment-emergent suicidal ideation and behavior will also be analyzed by the following subgroups: patients with or without comorbid MDD; >= or < the baseline median CDI score for the study. The Fisher's exact test will be used to overall compare the proportion of suicide-related events between the treatment groups. A logistic

regression model with factors of treatment, subgroup and treatment-by-subgroup will be performed to compare the proportion accounting for the subgroup effect.

For the BPI average pain severity, a subgroup analysis will be conducted for Study Period II. Table HMGW.6.14 lists the subgroup analysis variables by which the analyses performed.

To analyze a specific subgroup's impact, change from baseline to endpoint (LOCF and BOCF) in BPI average pain will be analyzed using an ANCOVA model with all the terms described generally in Section 6.1.2 and 6.1.3 with additional terms of the subgroup and the subgroup-by-treatment interaction. The primary statistical test will be for the treatment-by-subgroup interaction, which will be tested at the significance level of 0.05.

Furthermore, treatment group differences will be evaluated within each category of a subgroup regardless of the significance level of the treatment-by-subgroup interaction. For the subgroup of Race Origin, all the categories that have <10% of the patients in the study will be combined in the analysis. For the subgroup of NSAIDs use, the NSAIDs use is based on concomitant medication took during study and medical review.

Subgroup analysis for other efficacy and safety variables will be conducted as deemed appropriate and necessary.

Table HMGW.6.14 Definition of Subgroup Variables

Subgroup Variable	Categories
1. Gender	1. Female or Male
2. Race Origin	2. White American Indian or Alaska Native Asian Black or African American Native Hawaiian or Other Pacific Islander Multiple
3. Baseline Pain Severity	3. ≤6 vs. >6 on BPI average pain severity at baseline
4. Family history of Fibromyalgia	4. Yes/No
5. Onset age of Fibromyalgia (years)	5. ≤12 vs. >12
6. Comorbid MDD	6. Yes/No
7. Comorbid GAD	7. Yes/No
8. Country	8. Country name
9. NSAID use (ongoing, concurrent regimen)	9. Yes/No

Abbreviations: BPI = Brief Pain Inventory; MDD = major depressive disorder; GAD = generalized anxiety disorder; vs. = versus, NSAID = non-steroidal anti-inflammatory drugs.

6.2. Interim Analysis

Due to low patients enrollment throughout this study, Lilly provided FDA with a document (Sequence No. 0036) to request release from the post marketing requirement (PMR) in April 2014. FDA requested Lilly to explore the possibility of conducting an interim analysis of the efficacy data in their July 2014 written response. Lilly proposed a futility analysis which was

subsequently requested to conduct from FDA. The appropriate Lilly regulatory scientists were consulted and determined that the unplanned interim analysis plan will be documented in the SAP, and the protocol was not amended.

The interim analysis would focus on the primary efficacy analysis of duloxetine 30/60 mg once daily (QD) compared with placebo on the reduction of average pain severity as measured by the BPI average pain severity rating during the Study Period II in adolescents (aged 13 to 17 years) with JPFS. The purpose of the interim analysis is to evaluate the primary efficacy endpoint to stop the study early for futility using pre-specified decision rules, and it will not require adjustment of α. Sites will remain blinded to any information produced from the interim analysis. A statistical analysis center (SAC), which would be external to the study team in order to maintain data integrity, will be unblinded and conducting the analyses in the restricted access folder, https://sddchippewa.sas.com/webdav/lillyce/prd/ly248686/f1j_mc_hmgw/ac_unblinded1. The study team will remain blinded to the interim analysis data if the study continues. The SAC members will not be in contact with study site personnel.

The randomization cut-off date for the interim analysis will be the date when approximately 150 patients are randomized, which is approximately 71% of planned enrollment. The data lock for the interim analysis would occur approximately 3 months later to ensure those randomized patients will subsequently either complete or discontinue the double-blind treatment phase. The patient population for the interim analysis will be all patients randomized on or before the randomization cut-off date.

The futility interim analysis will be based on the calculation of conditional power (CP) . Conditional power represents the probability that the ongoing trial will result in statistically significant difference between duloxetine 30/60 mg QD and placebo at end of the trial, based on the data available at the time of interim analysis. The calculation of conditional power follows that proposed in Lan and Wittes (1988) and DeMets (2006). If Z(t) represents the test statistic at interim, then the conditional power for expected test statistic θ (the assumed treatment effect) at end of trial could be calculated as following:

$$CP = 1 - \Phi\left\{\frac{Z_{\alpha/2} - Z(t)\sqrt{t} - \theta(1-t)}{\sqrt{1-t}}\right\}$$

Where Φ is the cumulative distribution function of the standard normal distribution, α is the type I error for two-sided test, and t is the information fraction which defined as proportion of planned patients randomized at interim analysis. If assuming data at end of the trial will follow the trend observed at interim analysis, the calculation of conditional power could be simplified as following:

$$CP = 1 - \Phi\left\{\frac{Z_{\alpha/2} - Z(t)/\sqrt{t}}{\sqrt{1-t}}\right\}$$

The test statistic Z(t) will be calculated based on estimators from primary efficacy analysis. The primary efficacy analysis will be the contrast between duloxetine 30/60 mg QD and placebo at the last visit during double-bind treatment phase (Visit 8, Week 13) from a mixed model repeated measures analysis (MMRM) on change from baseline in BPI average pain severity. The model will include the fixed categorical effect of treatment, pooled investigate site, visit, and treatment-by-visit interaction, as well as the continuous, fixed covariate of baseline value of BPI average pain severity and baseline value of BPI average pain severity-by-visit interaction. The test statistic at interim will be approximately calculated as $\widehat{Z(t)} = \hat{\delta}/se(\hat{\delta})$, where $\hat{\delta}$ is the estimator of the contrast between duloxetine 30/60 mg QD and placebo at Week 13, and $se(\hat{\delta})$ is the corresponding standard error.

The SAC will only communicate one of the two possible decisions to Lilly:

- Conditional Power (CP) <0.6: Trial termination for futility
- Conditional Power (CP) >=0.6: Trial may continue

with the pre-determined Lilly individuals (medical director, regulatory scientist, statistician, clinical trial manager, chief operating officer, and clinical project management advisor) in order to decide further study implementation. The outline of the plan will be based on the document that Lilly submitted to FDA on September 2, 2014, containing Lilly responses to the FDA guidance provided in the written responses dated July 11, 2014 titled "Response to FDA's Information Request Regarding Release from Post Marketing Requirement for Adolescents with Fibromyalgia: Study F1J-MC-HMGW" (Sequence No. 0037).

The SAC will inform the FDA project manager of the primary efficacy results from a mixed model repeated measures analysis (MMRM) on change from baseline in BPI average pain severity as well as the conditional power. This will be sent by a password protected encrypted file with a cover letter.

When the CP meets the criterion of futility stopping rule, and the study is terminated early, the sensitivity analyses to address the impact of missing data on primary efficacy analysis may not be performed.

6.3. Exploratory Objective Analyses

6.3.1. Correlation between Pediatric Pain Questionnaire and Brief Pain Inventory

The correlation between PPQ and BPI for each item (average pain rating, pain right now rating, worst pain rating) will be assessed. The correlation for all records where both PPQ and BPI data are non-missing at visits across the entire study for all ITT will be assessed by using Pearson's correlation coefficient.

6.3.2. Appropriateness of ACR criteria

Patients meet the American College of Rheumatology (ACR) criteria for fibromyalgia (Wolfe et al. 1990) if the following criteria are met:

- History of widespread pain >= 3 months (pain in the left and right side of the body, and pain above and below the waist, and axial skeletal pain)
- Tenderness in 11 or more of 18 specific points on digital palpation

The number and frequency of randomized patients who meet the ACR criteria for fibromyalgia will be summarized. McNemar test will be used to assess the agreement between ACR criteria and JPFS criteria defined by Yunus and Masi (Yunus and Masi, 1985). Also, the baseline

demographic characteristics and baseline illness characteristic will be summarized by treatment group for the patients who meet JPFS criteria but not ACR criteria.

7. References

30 May 2000. National Center for Health Statistics. http://www.cdc.gov/growthcharts/percentile_data_files.htm, then STATAGE [XLS - 125 KB].

30 May 2000. National Center for Health Statistics. http://www.cdc.gov/growthcharts/percentile_data_files.htm, then WTAGE [XLS - 125 KB].

Birhanu, T., Lipkovich, I., Molenberghs, G. and Mallinckrodt, C.H. (2014) A Multiple-Imputation-Based Approach to Sensitivity Analyses and Effectiveness Assessments in Longitudinal Clinical Trials. *Journal of Biopharm. Stat.*, 24:2, 211-228.

Demets DL. Futility approaches to interim monitoring by data monitoring committees. *Clinical Trials.* 2006;3(6):522-529.

Diggle, P.J., Liang, K.Y., and Zeger, S.L. (1994), Analysis of Longitudinal Data, Oxford: Clarendon Press.

Diggle, P.J. and Kenward, M.G. (1994), Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics* 43: 49-93.

Lan KK and Wittes J. The B-value: A tool for monitoring data. *Biometrics.* 1998;44(2):579-585.

Lu, K. and Mehrotra, D. Specification of covariance structure in longitudinal data analysis for randomized clinical trials, *Statist. Med.* 2010, 29: 474—488

Kenward, M.G. and Roger, J.H. Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood, *Biometrics* 1997, 53: 983 - 997.

National High Blood Pressure Education Program Working Group. The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents. *Pediatrics* 2004 August;114(2 Suppl 4th Report):555-76.

Permutt T. 2015. Sensitivity analysis for missing data in regulatory submissions. *Stat Med.* 35:2876-2879.

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons.

Yunus MB, Masi AT. Juvenile primary fibromyalgia syndrome. A clinical study of thirty-three patients and matched normal controls. Arthritis Rheum. 1985; 28 (2): 138-145.

Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, Tugwell P, Campbell SM, Abeles M, Clark P, Fam AG, Farber SJ, Fiechtner JJ, Franklin CM, Gatter RA, Hamaty D, Lessard J, Lichtbroun AS, Masi AT, McCain GA, Reynolds WJ, Romano TJ, Russell IJ, Sheon RP. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia. Report of the Multicenter Criteria Committee. Arthritis Rheum, 1990; 33 (2): 160-172.

Verbeke , G., Molenberghs, G. 2000. Linear Mixed Models for Longitudinal Data. New York: Springer-Verlag.