PhUSE 2016

Paper SP01

Standard Methods for Analysis and Reporting of VAS or NRS Derived Pain Relief Response Scores

James R. Johnson, Flexion Therapeutics, Inc., Cary, NC USA

ABSTRACT

The design of clinical trials to assess the Pain Relief (PR) properties of analgesic drugs often utilize subject reported visual analog scales (VAS) or numerical rating scale (NRS) instruments to assess Pain Intensity (PI) over a defined sampling interval. The decision to use VAS or NRS should consider the pain type (acute or chronic), temporal aspects of reporting PI, study population, PK of the study medication, and clinically relevant differences from placebo.

Missing PI data must be accounted for with imputation techniques such as Last Observation Carried Forward, Baseline Observation Carried Forward, Worst Observation Carried Forward, or a PI score obtained prior to administration of rescue medication. Common efficacy endpoints derived from VAS or NRS are the PI Difference (PID), Time weighted Sum PID (SPID), and Area under the PI curve (AUE). Analysis of these endpoints are completed using linear mixed models for repeated measures (MMRM) or analysis of variance (ANOVA).

Methods for implementation of the VAS and NRS scales to address the predictive qualities for treatment discrimination with these instruments and endpoints are presented. Methods for the implementation of MMRM and ANOVA models as well a graphical display standards for PR data and endpoints are also proposed as standards for pain data obtained using VAS or NRS.

INTRODUCTION

Pain is a subjective experience composed of two complementary features (i) a localized sensation afflicting a particular part of the body, and (ii) an unpleasant quality of varying degrees of severity commonly that is associated with behaviors or treatments directed at relieving or terminating the pain experience.[1] The degree to which a subject experiences pain is highly variable and associated with the temporal and physical aspects of the source of the pain, thus making any real objective assessment of pain very unreliable. Valid, reliable, and reproducible subjective patient reported assessments of pain intensity are essential for both clinical trials and effective pain management strategies.[2,3] These patient reported PI and PR assessments are the basis for determining the analgesic efficacy and is the subject of regulatory guidance for the development and approval of drugs and biologics for analgesic indications.[3,4,5,6]

Pain can be classified based on physiology (e.g. nociceptive, neuropathic, inflammatory), intensity (e.g. mild-moderate-severe, 0-10 NRS), temporal characteristics (e.g. acute: <3 months duration, Chronic: >3 months duration), type of tissue affected (e.g. skin, muscle, viscera, joints, tendons, bones), and syndrome (e.g. cancer, fibromyalgia, osteoarthritis, migraine, others).[2] Analgesic treatments for pain management take into consideration each of these biological and temporal factors.

Clinical trials designed to test an analgesic treatments impact on pain relief (PR) will examine PI related to specific diseases and disease syndromes. Key design features for clinical trials that assess PI and PR include selecting the patient reported instrument for recording pain. Two main, validated, instruments are currently utilized in practice: Numerical Rating Scale (NRS) or Visual Analog Scale (VAS). Efficacy endpoints derived from VAS or NRS are the PI Difference (PID), Time weighted Sum PID (SPID), and Area under the PI curve (AUE). Analysis of these endpoints are completed using linear mixed models for repeated measures (MMRM) or analysis of variance (ANOVA).

Clinical trials for analgesic interventions to provide symptomatic PR from either acute (e.g. soft tissue injury, dental molar extractions, post-operative pain) or chronic (e.g. osteoarthritis, fibromyalgia) pain conditions have different properties and often will have different PI sampling schemas. Acute pain trials are typically conducted over a short temporal period generally less than 3 months. Chronic pain trials are typically conducted over longer temporal periods of 12 weeks or longer (some up to a year or longer). The rates of treatment discontinuation and lost to follow-up are much greater in chronic pain trials compared to acute pain trials. The reasons for treatment discontinuation or administration of rescue medication in analgesic trials are different between the treatment and the control groups, especially placebo-controlled trials. Discontinuation in the placebo treatment groups have higher rates associated from inadequate pain relief, whereas discontinuation in the experimental treatment groups have higher

discontinuation rates associated with poor tolerability and adverse events. Clinical trial designs must account for the missing PI data associated with discontinuation from treatments or lost to follow-up.

All of these features of clinical trials designed to assess PR for an analgesic are key features of the study design and analysis planning. In follow on sections we present the instruments used to collect PI, and introduction to how missing data are managed for analysis, methods for computing efficacy endpoints, models for the analysis of efficacy endpoints, and a proposal for the standard presentation of PI and PR data from analgesic clinical trials.

PAIN INTENSITY COLLECTION INSTRUMENTS

The choice of implementing a VAS or NRS instrument to discriminate PI in a clinical trial is not a trivial decision, as the sensitivity and precision of each instrument must be considered for evaluating pain, and computing these common efficacy endpoints. The decision to use VAS or NRS pain scales should consider the type of pain setting (acute or chronic), temporal aspects of pain reporting (length of time and frequency), targeted disease study population, pharmacokinetics of the medication under study, and the clinically relevant separation of treatments from placebo or an active control. Another key consideration is the implementation of either NRS or VAS: In-clinic settings allow for close supervision of instrument administration, ensuring proper data collection, whereas out-patient settings provide little or no supervision with instrument administration. The VAS instrument is best administered in an in-clinic setting, whereas the NRS instrument can be used in both settings effectively.

NUMERICAL PAIN INTENSITY SCALE (NRS)

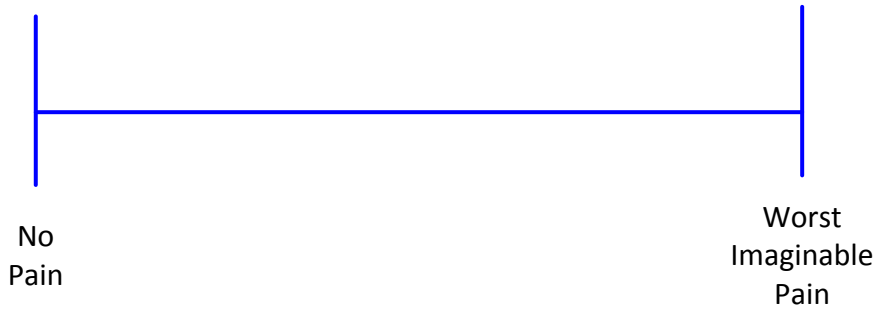
The NRS is used to collect PI scores at each selected time interval and is defined on an 11 point scale 0 to 10, in increments of 1, where 0 is No Pain and 10 is Worst Imaginable pain. Often times the NRS instrument is implemented in clinical trials as presented in Figure 1, where a subject selects a single number that best describes their current pain.

Figure 1 Numerical Rating Scale (NRS) Instrument

VISUAL ANALOG PAIN SCALE (VAS)

The VAS instrument is used to collect PI at each selected time interval and is defined on a sliding scale from 0 to 100 mm (3.94 in), where the left anchor at 0 mm is No Pain, and the 100 mm anchor is Worst Imaginable pain. Often times the VAS is implemented in clinical trials as presented in Figure 2, where the subject selects a point on the scale that best describes their current pain. The subject marks the scale and PI is the measured distance from the No Pain anchor (0) to the first point where the subject marks the scale. Scores range from 0 to 100 mm. Most measurements are rounded to the nearest millimeter (mm).

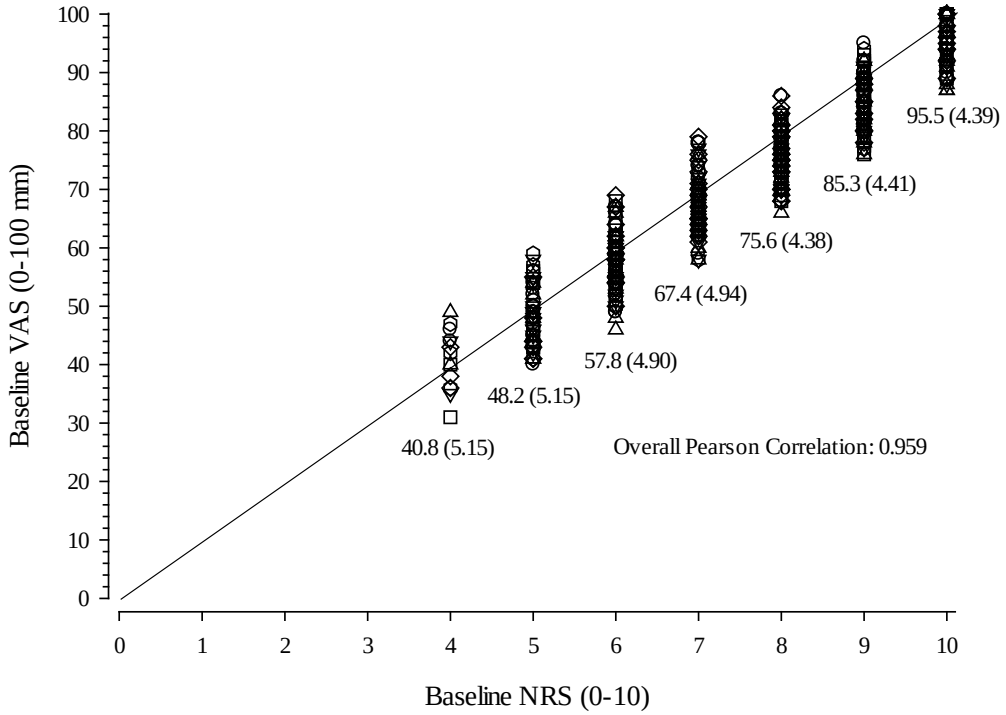
Figure 2 Visual Analog Pain Intensity Scale (VAS)



The specificity of the NRS and VAS instruments in a clinical trial is an important consideration when selecting which

instrument to utilize in a trial. The majority of clinical trials will utilize a baseline assessment of PI either as a single assessment or assessments over a period of time to establish a baseline PI score, with protocol qualifying requirements with a minimum PI intensity (e.g. NRS ≥ 4 or VAS ≥ 40 mm). The difference in baseline reporting of PI with the NRS and VAS instruments, in the presence of a protocol mandated qualifying baseline score, can have profound effects in detecting a difference between treatments, and is a point to consider in the design of an analgesic clinical trial. Consider that VAS has a higher degree of measurement precision and lower correlation with the NRS than the NRS instrument at pain levels associated with the middle of the NRS scale (i.e. NRS 3-7) (Figure 3).

Figure 3 Matched Baseline NRS and VAS scores from two studies with qualifying PI of VAS ≥ 40 mm or NRS ≥ 4 (N=563)



The choice of implementing the NRS or VAS must be carefully considered. The main advantages of the NRS are that the instrument is simple, with a limited number of choices (0 to 10) and when anchored with "No Pain" and "Worst Imaginable Pain" provide a reference from which to choose a PI score at any given time point. The NRS can be used effectively across all types of acute and chronic disorders. A disadvantage of the NRS may be that the level of specificity of PI is limited to a discrete 11-point scale, in studies where greater specificity in PI scoring is required. The VAS is also simple to implement and score (by measurement in mm), and provides many points from which to choose on the 0-100 mm scale, which increases the level of precision and specificity in PI reporting. A disadvantage of the VAS is that a subject requires more concentration and coordination to respond to reporting PI at a given time point, which can be more difficult for post-operative pain studies or studies with neurological disorders.

MISSING PAIN INTENSITY DATA

The collection of PI data with the VAS or NRS instruments over a period will almost always result in the presence of missing PI scores at one or more time points over the sampling interval. The statistical management of missing data is an important topic that is critical to the planned analyses of clinical trial data.[7-14] Methods for the imputation of missing PI data has been studied and a large number of methodological papers have discussed this topic.[14-18] Missing PI data offer several unique challenges to consider with respect to how the PI data resulted in missing: Consider (a) subjects who have an adverse event and drop from the study, (b) subjects do not record PI at one or more time points, (c) subjects are lost to follow-up. How missing PI data are handled for the specific reason PI data are missing for the computation of efficacy endpoints must be fully specified in the SAP. The most common imputation methods are Last Observation carried Forward (LOCF), Worst Observation Carried Forward (WOCF), and Baseline Observation Carried Forward (BOCF). It should be noted that each of these imputation methods has their own statistical analysis challenges with respect to biasing results.[11,12] It is also worth noting that while LOCF analysis is simple, and easy to implement, it's use is associated with the introduction of biased results and should probably be avoided as a primary method for imputation of missing PI data. By way of example consider an acute pain study where the assessment of PI is over an 8 hour sampling interval following single dose administration. How missing PI

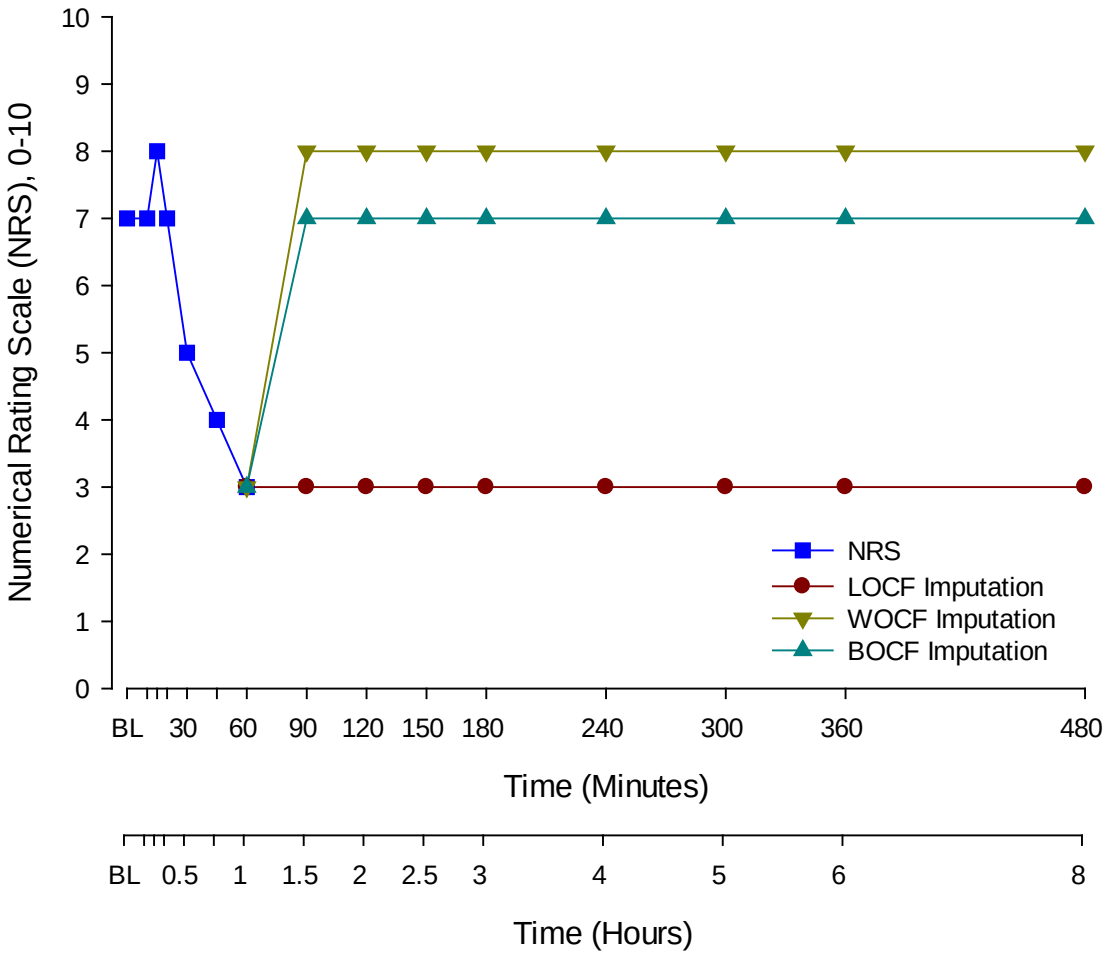
data may be implemented with each of these imputation strategies is presented in Table 1. The use of these three implementation strategies is useful for sensitivity analysis.

Table 1 Implementation of LOCF, WOCF, and BOCF

Time (min)	0 (BL)	10	15	20	30	45	60	90	120	150	180	240	300	360	480
PI (0-10)	7	7	8	7	5	4	3	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missing
LOCF	7	7	8	7	5	4	3	3	3	3	3	3	3	3	3
WOCF	7	7	8	7	5	4	3	8	8	8	8	8	8	8	8
BOCF	7	7	8	7	5	4	3	7	7	7	7	7	7	7	7

The impact that the implementation of LOCF, WOCF or BOCF on the computation of the efficacy endpoints can be seen in the overall PI curve for a subject (Figure 4).

Figure 4 Individual Subject Pain Intensity Profile with LOCF, WOCF and BOCF



ADJUSTMENT OF PAIN INTENSITY DATA FOR RESCUE MEDICATION USAGE

The design of clinical trials for analgesia most often will include the use of rescue medication when adequate PR is not achieved by a subject in a trial, especially in placebo controlled trials. A common practice in these study designs is that when rescue medication is requested by the subject prior to receiving the rescue medication a “pre-rescue PI score” is obtained. This score is then carried forward in the planned analyses, and the subject continues to record their PI over the rest of the sampling interval, or a pre-specified window of sampling time that is based on the

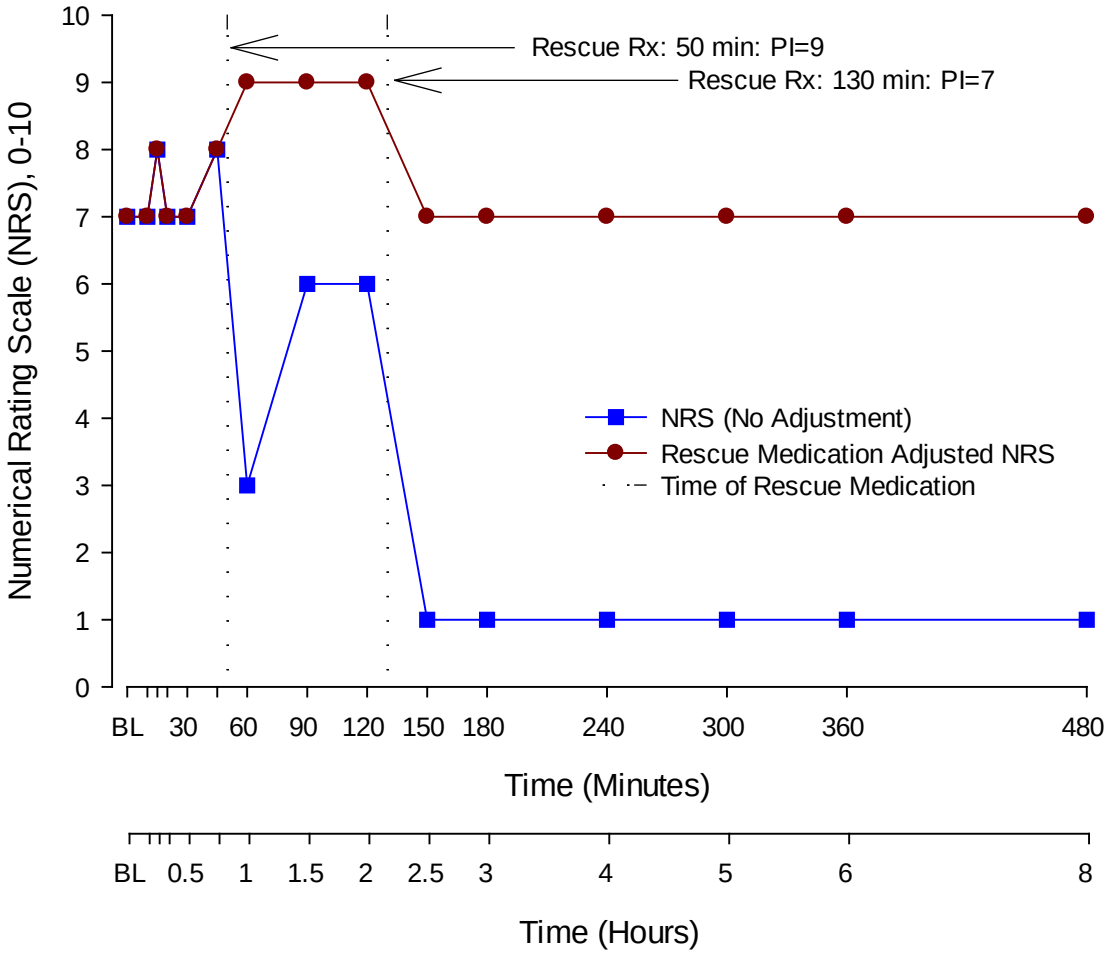
pharmacokinetics of the rescue medication. Pre-rescue medication PI scores are carried forward in the analysis record. The impact of multiple uses of rescue medication and multiple pre-rescue medication PI scores are included in the analysis records (Table 2).

Table 2 Implementation of Rescue Medication PI scores over a sampling interval.

Time (min)	0 (BL)	10	15	20	30	45	60	90	120	150	180	240	300	360	480
PI (0-10)	7	7	8	7	7	8	3	6	6	1	1	1	1	1	1
Rescue Time						50			130						
Pre-Rescue PI						9			7						
Rescue Adjusted PI	7	7	8	7	7	8	9	9	9	7	7	7	7	7	7

The impact that the implementation of Rescue medication adjusted PI scores on the computation of the efficacy endpoints can be seen in the overall PI curve for a subject (Figure 5).

Figure 5 Individual Subject Pain Intensity Profile with Rescue Medication PI Score Adjustment



Each of these imputation methods for handling missing data can be adequately implemented in ADaM data sets with an appropriate DTYPE variable identifier for the PI values (AVAL) substituted for missing or rescue medication implemented imputation. It should be noted that while we do not present the underlying CDISC Compliant ADaM analysis data set specifications, the use of ADaM standard is assumed for analysis data.

The imputation of missing PI data is not restricted to these methods. One technique, not presented herein, is implementation of a multiple imputation technique where instead of substituting a single PI value for each missing value a multiple imputation procedure replaces each missing value with a set of plausible values that represent the uncertainty about the right value to impute.

EFFICACY ENDPOINTS

Three common efficacy endpoints are often derived for assessing PR and include the Pain Intensity Difference (PID), Time weighted Sum pain Intensity Differences (SPID), and Sum Pain Intensity Differences Area Calculation (AUE). Each of these derived endpoints is specific to analgesic assessment in either an acute or chronic pain setting. Choosing which endpoint is most appropriate for the analgesic intervention is a function of defining a clinically important difference to be observed from either placebo or active control, as well as the regulatory requirements for the analgesic medication being investigated.[5,6,20]

Calculation of the efficacy endpoints is on an individual subject profile over a defined sampling interval. The choice of the time unit (e.g. minutes, hours, days, weeks) should be driven by the temporal aspects of the sampling regimen. Typically time is expressed in minutes for acute pain settings where the sampling interval is less than 1 day (24 hours, 1440 minutes). Hours are used in the computations where multiple days are sampled, up to approximately one month's sampling duration. Days or weeks are assigned as the time unit for long term chronic pain assessments generally exceeding one month's duration. It is critically important to define the unit of time from which efficacy endpoints will be computed in the protocol and statistical analysis plan. The equations for each endpoint are presented as follows:

Pain Intensity Difference (PID): Two methods for computing a PID are based on the desire for expressing the direction of the change from baseline, and expressed as follows:

$$(1)\ PID_t = PI_{baseline} - PI_t \quad \text{or} \quad (2)\ PID_t = PI_t - PI_{baseline}$$

Time Weighted Sum Pain Intensity Differences (SPID):

$$(3)\ SPID_{t_i - t_{i+n}} = \sum_{t_i}^{t_{i+n}} (PID_i) * (t_{i+1} - t_i)$$

Where $t_{(0)}$ = baseline (Pre-dose), t_i is the scheduled assessment time, $PID_{(i)}$ is the pain intensity difference score calculated at each post-baseline time point.

Sum Pain Intensity Differences Area Calculation (Area Under the Effect (AUE)):

$$(4)\ AUE_{t_i - t_{i+n}} = \sum_{t_i}^{t_{i+n}} ((PID_i + PID_{i+1})/2) * (t_{i+1} - t_i)$$

Where PID_i = the pain intensity difference score at t_i, and PID_{i+1} = the pain intensity difference score at t_{i+1} from the adjacent time point.

The increment of time interval associated with the efficacy endpoints needs to be specified in the protocol and the SAP and should be linked to the pharmacokinetics and expected duration of PR of the analgesic under investigation. For instance SPID endpoints will be calculated over the sampling interval defined, and may also be computed for partial-time intervals over the planned sampling interval that are linked to the maximum concentration (C_{max}) from the PK profile of the medication.

EFFICACY ENDPOINT EXAMPLE:

By way of example consider a post-operative dental pain clinical trial with two test analgesics (Treatments A and B) and Placebo. PI is assessed over an 8 hour sampling interval following single dose oral study medication administration, where PI is collected with the NRS instrument, and adjustment of PI is implemented with rescue medication PI scores (Figure 5). Derivation of efficacy endpoints are described for two sample subjects, and include partial-time intervals are presented in Table 3. Note that in the example VAS PI assessments can be used interchangeably for all calculations.

Table 3 Efficacy Endpoint Computations

(a) Efficacy Endpoints without Rescue Medication PI Score Adjustment (Subject 1)

	BL	10 Mins.	15 Mins.	20 Mins.	30 Mins.	45 Mins.	1 Hr.	1.5 Hrs.	2 Hrs.	2.5 Hrs.	3 Hrs.	4 Hrs.	5 Hrs.	6 Hrs.	8 Hrs.
Time (minutes)	0	10	15	20	30	45	60	90	120	150	180	240	300	360	480
TD min (t_{i+1} - t_i)		10	5	5	10	15	15	30	30	30	30	60	60	60	120
Pain Intensity (0-10)	7	7	8	7	7	8	3	6	6	1	1	1	1	1	1
PID (t_i-BL)	0	0	1	0	0	1	-4	-1	-1	-6	-6	-6	-6	-6	-6
PIDi*(t_{i+1}-t_i)		0	5	0	0	15	-60	-30	-30	-180	-180	-360	-360	-360	-720
(PID_{i+1}+PID_i)/2		0	0.5	0.5	0	0.5	-1.5	-2.5	-1	-3.5	-6	-6	-6	-6	-6
[(PID_i+PID_{i+1})/2]*TD min		0.0	2.5	2.5	0.0	7.5	-22.5	-75.0	-30.0	-105.0	-180.0	-360.0	-360.0	-360.0	-720.0

Efficacy Endpoint	Value
Time Weighted $SPID_{0-180}$	-460
Time Weighted $SPID_{0-360}$	-1540
Time Weighted $SPID_{0-480}$	-2260

Efficacy Endpoint	Value
AUE_{0-180}	-400
AUE_{0-360}	-1480
AUE_{0-480}	-2200

(b) Efficacy Endpoints with Rescue Medication PI Score Adjustment (Subject 1)

	BL	10 Mins.	15 Mins.	20 Mins.	30 Mins.	45 Mins.	1 Hr.	1.5 Hrs.	2 Hrs.	2.5 Hrs.	3 Hrs.	4 Hrs.	5 Hrs.	6 Hrs.	8 Hrs.
Time (minutes)	0	10	15	20	30	45	60	90	120	150	180	240	300	360	480
TD min (t_{i+1} - t_i)		10	5	5	10	15	15	30	30	30	30	60	60	60	120
Rescue Adjusted Pain Intensity (0-10)[1]	7	7	8	7	7	8	9	9	9	7	7	7	7	7	7
PID (t_i-BL)	0	0	1	0	0	1	2	2	2	0	0	0	0	0	0
PIDi*(t_{i+1}-t_i)		0	5	0	0	15	30	60	60	0	0	0	0	0	0
(PID_{i+1}+PID_i)/2		0	0.5	0.5	0	0.5	1.5	2	2	1	0	0	0	0	0
[(PID_i+PID_{i+1})/2]*TD min		0.0	2.5	2.5	0.0	7.5	22.5	60.0	60.0	30.0	0.0	0.0	0.0	0.0	0.0

Efficacy Endpoint	Value
Time Weighted $SPID_{0-180}$	170
Time Weighted $SPID_{0-360}$	170
Time Weighted $SPID_{0-480}$	170

Efficacy Endpoint	Value
AUE_{0-180}	185
AUE_{0-360}	185
AUE_{0-480}	185

[1]: Subject requests Rescue Medication at 50 minutes post treatment with NRS=9 and again at 130 minutes post treatment with NRS=7.

(a) Efficacy Endpoints without Rescue Medication PI Score Adjustment (Subject 2)

	BL	10 Mins.	15 Mins.	20 Mins.	30 Mins.	45 Mins.	1 Hr.	1.5 Hrs.	2 Hrs.	2.5 Hrs.	3 Hrs.	4 Hrs.	5 Hrs.	6 Hrs.	8 Hrs.
Time (minutes)	0	10	15	20	30	45	60	90	120	150	180	240	300	360	480
TD min (t_{i+1} - t_i)		10	5	5	10	15	15	30	30	30	30	60	60	60	120
Pain Intensity (0-10)	9	9	7	6	5	2	2	2	1	1	1	3	4	4	4
PID (t_i-BL)	0	0	-2	-3	-4	-7	-7	-7	-8	-8	-8	-6	-5	-5	-5
PIDi*(t_{i+1}-t_i)		0	-10	-15	-40	-105	-105	-210	-240	-240	-240	-360	-300	-300	-600
(PID_{i+1}+PID_i)/2		0	-1	-2.5	-3.5	-5.5	-7	-7	-7.5	-8	-8	-7	-5.5	-5	-5
[(PID_i+PID_{i+1})/2]*TD min		0.0	-5.0	-12.5	-35.0	-82.5	-105.0	-210.0	-225.0	-240.0	-240.0	-420.0	-330.0	-300.0	-600.0

Efficacy Endpoint	Value
Time Weighted $SPID_{0-180}$	-1205
Time Weighted $SPID_{0-360}$	-2165
Time Weighted $SPID_{0-480}$	-2765

Efficacy Endpoint	Value
AUE_{0-180}	-1155
AUE_{0-360}	-2205
AUE_{0-480}	-2805

(b) Efficacy Endpoints with Rescue Medication PI Score Adjustment (Subject 2)

	BL	10 Mins.	15 Mins.	20 Mins.	30 Mins.	45 Mins.	1 Hr.	1.5 Hrs.	2 Hrs.	2.5 Hrs.	3 Hrs.	4 Hrs.	5 Hrs.	6 Hrs.	8 Hrs.
Time (minutes)	0	10	15	20	30	45	60	90	120	150	180	240	300	360	480
TD min (t_{i+1} - t_i)		10	5	5	10	15	15	30	30	30	30	60	60	60	120
Rescue Adjusted Pain Intensity (0-10)[1]	9	9	7	6	5	2	2	2	1	1	1	3	5	5	5
PID (t_i-BL)	0	0	-2	-3	-4	-7	-7	-7	-8	-8	-8	-6	-4	-4	-4
PIDi*(t_{i+1}-t_i)		0	-10	-15	-40	-105	-105	-210	-240	-240	-240	-360	-240	-240	-480
(PID_{i+1}+PID_i)/2		0	-1	-2.5	-3.5	-5.5	-7	-7	-7.5	-8	-8	-7	-5	-4	-4
[(PID_i+PID_{i+1})/2]*TD min		0.0	-5.0	-12.5	-35.0	-82.5	-105.0	-210.0	-225.0	-240.0	-240.0	-420.0	-300.0	-240.0	-480.0

Efficacy Endpoint	Value
Time Weighted $SPID_{0-180}$	-1205
Time Weighted $SPID_{0-360}$	-2045
Time Weighted $SPID_{0-480}$	-2525

Efficacy Endpoint	Value
AUE_{0-180}	-1155
AUE_{0-360}	-2115
AUE_{0-480}	-2595

[1]: Subject requests Rescue Medication at 260 minutes post treatment with NRS=5.

PAIN INTENSITY DIFFERENCE (PID) ENDPOINT

The efficacy endpoint PID is computed for each time interval sampled, and adjustment for rescue medication PI score is included for the primary computation. PID (change from baseline) information is analyzed using a longitudinal mixed model for repeated measures (MMRM) with fixed effects for treatment, time, treatment-by-time interaction, study site and baseline pain score. Subject will be the random effect. Treatment differences from control (Placebo) are estimated from the least squares means from the analysis model along with 95% confidence intervals and associated 2-sided p-values. Typically PID data are completed on observed cases (no imputation for missing data) as the primary analysis and then on the same imputation methods applied for the calculation of the SPID and AUE endpoints. In studies where the adjustment of PI information is applied in the presence of rescue medication then a second analysis often times completed is on the PI data without rescue medication adjustment for observed cases.

The basic model for the analysis of the PID change information implemented with SAS® Software[1] PROC MIXED is as follows:

```
   proc mixed data=<ADNRS> method=reml;
     where avisit) ^= 'baseline' and paramcd="PIDRA";
     class usubjid trt01p avisitn;
     model chg = base trt01p avisitn trt01p*avisitn / DDFM=KR;
     repeated avisitn / type = un subject=usubjid;
     lsmeans trt01p / diff=control ('Placebo') cl;
     lsmeans trt01p*avisitn / pdiff cl;
     ods output lsmeans = lsmeans
                 diffs = diffs
                Tests3 = test;
   run;
quit;
```

Descriptive summaries of PID at the individual time points, along with the model output for test of differences in the Least Square Means (LSM) are proposed as the standard display, along with a graphical representation of the LSM change from baseline. (See Table 4)

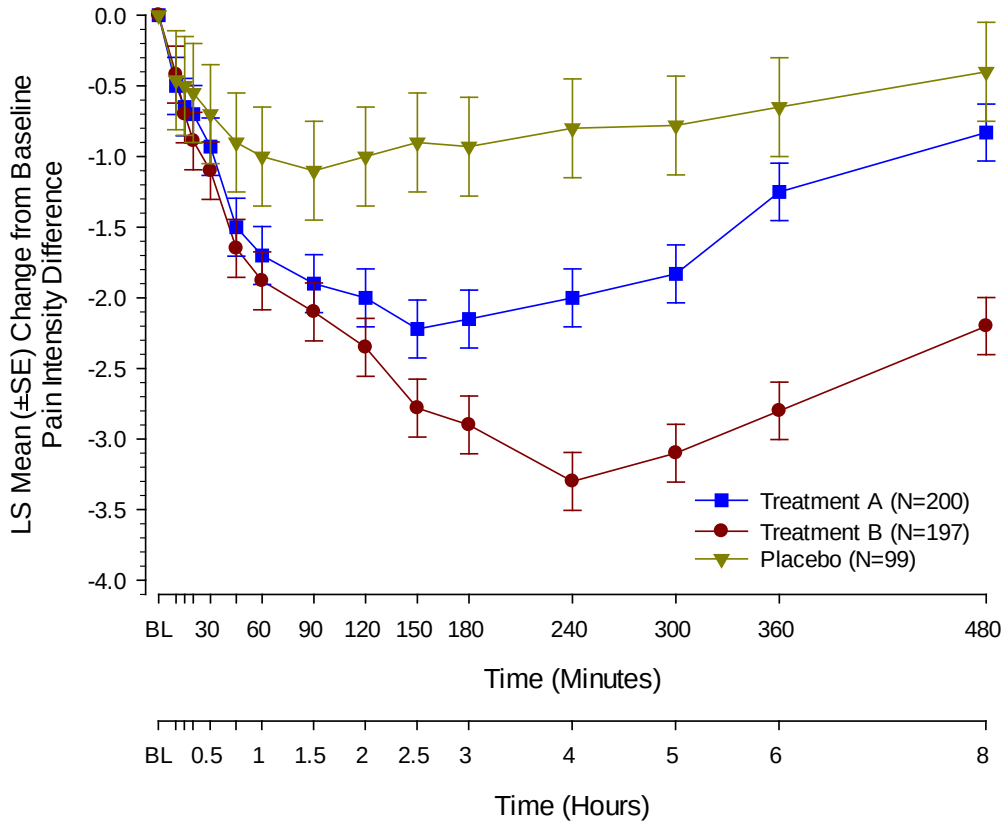
GRAPHICAL DISPLAY OF PID

As shown in equations (1) and (2) PID is computed for each time interval sampled, and the direction can be either positive or negative. It is recommended that if the intent is PR (Pain Relief) is the reduction in PI then the PID values are negative for reduction. It is acknowledged that some sponsors prefer to display PID as a positive number and complete the analysis with this direction. In either case the computation of PID needs to be declared in both the protocol and the SAP. A recommend standard for the display of PID change over time is presented as an example for our case study in Figure 6.

In addition to the LSM derived PID display over the sampling interval standard displays for the individual subject PI scores with all associated imputation or rescue medication adjusted PI scores overlaid on By Subject plots, as presented in Figure 4 and Figure 5.

[1] The analyses completed for this paper were generated using SAS® software. Copyright 2015, SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Figure 6 **Standard Display of Pain Intensity Difference**



TIME WEIGHTED SUM PAIN INTENSITY DIFFERENCES (SPID) AND AREA UNDER THE EFFECT (AUE) ENDPOINTS

The efficacy endpoints that are derived in Table 3 (Efficacy Endpoints) are computed over the entire time interval that are sampled. SPID and AUE information are analyzed using a linear model (ANOVA) with fixed effects for treatment, and baseline pain score. Treatment differences from control (Placebo) are estimated from the least squares means from the analysis model along with 95% confidence intervals and associated 2-sided p-values. The basic ANOVA model implemented with SAS Software for the analysis of these endpoints is as follows:

```
proc mixed data=<ADEFF>;
  where paramcd = "<endpoint code>;
  class trt01p;
  model aval = trt01p base / ddfm=kr;
  lsmeans trt01p / pdiff cl;
  estimate 'Treatment A v Placebo' trt01p -1  0 1 / cl alpha=0.05;
  estimate 'Treatment B v Placebo' trt01p  0 -1 1 / cl alpha=0.05;
  ods output lsmeans = lsmeans
              diffs = diffs
             Tests3 = test
          Estimates = est;
run;
quit;
```

In these analyses it is recommended that the primary analysis be completed on the endpoint with the appropriate imputation method for missing PI data for the calculation. It is also recommended that the sensitivity analyses be completed and displayed in the same manner. The protocol and the SAP must fully specify the imputation methods to be applied for missing data or rescue adjusted PI data. In our example analyses would have been displayed for the endpoints computed with and without rescue medication PI score adjustment in Table 3 (SPID0-180, SPID0-360, SPID0-360, etc.).

Table 4 Standard Table Display for Pain Intensity Difference (PID) from Baseline with comparisons to Placebo

	Treatment A (N=xxx)		Treatment B (N=xxx)		Placebo (N=xxx)	
Time Point / Statistic	Observed	Change from Baseline	Observed	Change from Baseline	Observed	Change from Baseline
Baseline						
n	xxx		xxx		xxx	
Mean (SD)	xxx.x (xx.xx)		xxx.x (xx.xx)		xxx.x (xx.xx)	
Median	xxx.x		xxx.x		xxx.x	
Min, Max	xxx, xxx		xxx, xxx		xxx, xxx	
<Time Point>	xxx	xxx	xxx	xxx	xxx	xxx
n	xxx.x (xx.xx)	xxx.x (xx.xx)	xxx.x (xx.xx)	xxx.x (xx.xx)	xxx.x (xx.xx)	xxx.x (xx.xx)
Mean (SD)	xxx.x	xxx.x	xxx.x	xxx.x	xxx.x	xxx.x
Min, Max	xxx, xxx	xxx, xxx	xxx, xxx	xxx, xxx xxx.x	xxx, xxx	xxx, xxx xxx.x
LSM (SE)		xxx.x (xx.xx)		(xx.xx)		(xx.xx)
95% CI		xxx.x, xxx.x		xxx.x, xxx.x		xxx.x, xxx.x
LSM Difference from Placebo		xx.xx		xx.xx		
95% CI		xx.xx, xx.xx		xx.xx, xx.xx		
p-value, 2-sided		0.xxxx		0.xxxx		

Note: LSM (SE), mean difference from placebo, CI and p-values from mixed model, modeling Pain Intensity Difference from baseline with fixed effects of Treatment, time point, treatment by time interaction, and model covariates of baseline PI score and <covariates>

<Other Footnotes>

Programming Note:

- Display one time point per page for clarity of analysis
- Additional descriptive statistics may include %CV or Interquartile ranges if needed. Insert on separate lines

Table 5 Standard Table Display for Sum Pain Intensity Difference (SPID) and Area Under the Effect endpoints with comparisons to Placebo

Endpoint / Statistic	Treatment A (N=xxx)	Treatment B (N=xxx)	Placebo (N=xxx)
<Endpoint>	xxx	xxx	xxx
n	xxx.x (xx.xx)	xxx.x (xx.xx)	xxx.x (xx.xx)
Mean (SD)	xxx.x	xxx.x	xxx.x
Min, Max	xxx, xxx	xxx, xxx	xxx, xxx
LSM (SE)	xxx.x (xx.xx)	xxx.x (xx.xx)	xxx.x (xx.xx)
95% CI	xxx.x, xxx.x	xxx.x, xxx.x	xxx.x, xxx.x
LSM Difference from Placebo	xx.xx	xx.xx	
95% CI	xx.xx, xx.xx	xx.xx, xx.xx	
p-value, 2-sided	0.xxxx	0.xxxx	

Note: LSM (SE), mean difference from placebo, CI and p-values from linear model (ANOVA), modeling <Efficacy Endpoint> Difference from baseline with fixed effects of Treatment, and model covariates of baseline PI score and <covariates>

<Other Footnotes>

Programming Note:

- Display efficacy endpoint per page for clarity of analysis
- Additional descriptive statistics may include %CV or Interquartile ranges if needed. Insert on separate lines.

CONCLUSION AND RECOMMENDATION

The purpose of this paper is to provide a first proposal on the analysis and reporting (displaying, summarizing, and analysis) of Pain Intensity information collected using either an NRS or VAS instrument. The development of analgesic drugs to provide PR will often utilize one of these two instruments to collect serial pain intensity data over a defined period of time (acute or chronic). The methods for handling missing PI data has been the topic of many peer reviewed articles and appropriate methods are defined in the literature. Imputation methods for missing PI data, as well as implementation for use of rescue medication adjusted PI scores must be accounted for in the computation of efficacy endpoints in the protocol and SAP. Efficacy endpoints of PID, SIPD, and AUE are well characterized and accepted endpoints both in the literature and by regulatory authorities for evaluation of the analgesic benefits of investigational medications. Clear descriptions of the primary efficacy endpoint (PID, SPID or AUE), along with the primary analysis model, and sensitivity analyses for the endpoint must also be defined in the protocol and SAP.

We have presented a first proposal for a standardized analysis approach and display of Pain Intensity data collected with NRS or VAS instruments. In this paper we have not provided detailed programming code other than the models to be implemented with SAS® Software using PROC MIXED. It is recommended that standard programming code (e.g. SAS® Software or R) be developed utilizing standardized SDTM and ADaM data architecture to provide standardized input to the statistical models and made publically available in the PhUSE CSS Standard Scripts Repository.

Ideally further development (possibly as a PhUSE CSS sponsored White Paper) on the development of standard Tables, Figures, and Listings (TLF) and associated analysis approaches would enhance standardization of the collection and reporting of pain intensity information and pain related efficacy endpoints through standard data architecture with SDTM and ADaM models for regulatory compliant submissions. Another standards development recommendation would be to ensure that information on the PhUSE Wiki for the Pain Therapeutic Areas is comprehensive and updated with the latest most relevant information on standards for pain endpoint development in clinical trials (PhUSE Wiki: Pain).

In this paper we have only touched on some of the endpoints available when implementing the NRS or VAS for the collection of PI data. Other endpoints such as TOTPAR (Total Pain Relief) using a five-point categorical PR scale (0-None, 1=Slight, 2=Moderate, 3=Good, and 4=Complete) have similar properties and should also be included in the analysis of PR data in standard recommendations. Other disease specific PI or PR instruments (e.g. WOMAC A for OA pain or the McGill Pain Questionnaire) are tools that have well published standard approaches for the analysis of data from these instruments. These disease specific instruments would also benefit from a standards recommendation for SDTM and ADaM dataset architecture for regulatory compliant submissions, as well as statistical model TLF recommended standards. Other derived endpoints such as Time to Onset of PR and PR responder analyses would also benefit from published standard approaches.

REFERENCES

1. Institute of Medicine (US) Committee on Pain, Disability, and Chronic Illness Behavior; Osterweis M, Kleinman A, Mechanic D, editors. Pain and Disability: Clinical, Behavioral, and Public Policy Perspectives. Washington (DC): National Academies Press (US); 1987. 7, The Anatomy and Physiology of Pain. Available from: http://www.ncbi.nlm.nih.gov/books/NBK219252/
2. Summers, S., 2000. Evidence-based practice part 1: Pain definitions, pathophysiologic mechanisms, and theories. Journal of Perianesthesia Nursing, 15(5), pp.357-365. http://www.ncbi.nlm.nih.gov/pubmed/11811340
3. Macrae, W.A. and Davies, H.T.O., 1999. Epidemiology of pain. Core topics in pain, p.99. http://ayurvedavignan.in/freeEbooks/Core%20Topics%20in%20Pain.pdf#page=114
4. Breivik, H. Borchgrevink, P.C., Allen, A.M. Rosseland, L.A., Romundstad, L., Breivik Hals, E.K., Kvarstein, G., and Stubhaug, A. 2008. Assessment of pain. British Journal of Anesthesia 101(1): 17-24. http://bja.oxfordjournals.org/content/101/1/17.abstract
5. FDA Guidance for Industry. February 2014. Analgesic Indications: Developing Drug and Biological Products. USHHS/FDA/CDER http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm384691.pdf
6. European Medicines Agency, 17 December 2015. Guideline on the clinical development of medicinal products intended for the treatment of pain. EMA/CHMP/970057, Corr. 1 http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/12/WC500199242.pdf
7. LaVange, L. 2015. Missing Data Issues in Regulatory Clinical Trials. JPMA Symposium. February 12, 2015. http://www.jpma.or.jp/medicine/shinyaku/tiken/symposium/pdf/20150213/20150213_3.pdf.
8. ICH E9. 1998. Statistical Principles for Clinical Trials. CPMP/ICH/363/96. http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/statistical-principles-for-clinical-trials.html.
9. Little RJ, D'Agostino R, Cohen M, et al. The prevention and treatment of missing data in clinical trials. N Engl J Med, 367(14): 1355-1360. http://www.nejm.org/doi/full/10.1056/NEJMsr1203730#t=article

10. O'Neill RT and Temple R. 2012. The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. Clin. Pharmacol Ther, 91: 550-554. http://www.ncbi.nlm.nih.gov/pubmed/22318615
11. European Medicines Agency. 2011. Guideline on Missing data in Confirmatory Clinical Trials. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf
12. National Research Council. 2010. The Prevention and Treatment of Missing Data in Clinical Trials. Panel on handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The national Academies Press. http://www.ncbi.nlm.nih.gov/books/NBK209904/pdf/Bookshelf_NBK209904.pdf
13. Allison, PD. 2001. Missing Data. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-136. Thousand Oaks, CA: Sage.
14. Little, RJ and Rubin DB. 2002. Statistical Analysis with Missing Data, 2nd Edition. Wiley Series in Probability and Statistics. John Wiley and Sons. New York, NY.
15. Yang CY. 2000. Multiple Imputation for Missing Data: Concepts and New Development. P267-25. http://www.ats.ucla.edu/stat/sas/library/multipleimputation.pdf
16. Altman DG. 2009. Missing outcomes in randomized trials: addressing the dilemma. Open Medicine. 2009:3(2). http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2765768/
17. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. 2006. Review: A gentle introduction to imputation of missing values. Journal of Clinical Epidemiology 2006:59:1087-1091. http://www.ncbi.nlm.nih.gov/pubmed/16980149
18. Siddiqui O, Hung HMJ, O'Neill R. 2009. MMRM vs. LOCF: A comprehensive comparison based on simulation study and 25 NDA datasets. Journal of Biopharmaceutical Statistics. 2009:19:227-246. http://www.ncbi.nlm.nih.gov/pubmed/19212876
19. Molnar FJ, Man-Song-Hing M, Hutton B, Fergusson D. 2009. Have last-observation-carried-forward analyses caused us to favor more toxic dementia therapies over less toxic alternatives? A systematic review. Open Medicine 2009:3(2):31-50. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2765769/
20. Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL. 2000. Defining the clinically important difference in pain outcome measures. Pain. 2000:88;287-294. http://www.ncbi.nlm.nih.gov/pubmed/11068116

ACKNOWLEDGMENTS

I wish to acknowledge Dr's Mark Jaros and Phil Shale from Summit Analytical who have been colleagues on many different acute and chronic pain clinical trials. Ms. Teresa Curto and Mr. Angelo Tinazzi from Cytel, two highly skilled statistical professionals, who I have had the privilege of working with on a chronic pain relief development program and have pressure tested the proposed standards in this paper through our collaborative work on this program. Lastly, my colleagues at Flexion Therapeutics, Inc. who have graciously provided a venue and time to continue my work on statistical and statistical programming methods for analgesic clinical trials.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

James R. Johnson, PhD ("Jim")
Flexion Therapeutics, Inc.
104 Lutterworth Court
Cary, North Carolina 27519-8682 USA
Work Phone: +1 (919) 244-4756
Email: jrjphd@gmail.com or jjohnson@flexiontherapeutics.com