NDA #204442
PROBUPHINE®
(buprenorphine HCl/ethylene vinyl acetate)
Implant CIII

Treatment of Opioid Dependence

Errata to Briefing Document and Risk Evaluation and Mitigation Strategy

for

January 12, 2016 Advisory Committee Meeting of the Psychopharmacologic Division of the U.S. Food and Drug Administration

Braeburn Pharmaceuticals, Inc. (US Commercial Licensee) 47 Hulfish Street, Suite 441
Princeton, NJ 08542
(609) 751-5375

Issued December 23, 2015

THIS DOCUMENT IS AVAILABLE FOR PUBLIC DISCLOSURE WITHOUT REDACTION
Erratum

In this erratum, revised text is shown in **bold + italics**, denoting new text, and **strikethrough**, denoting deleted text.

The Briefing Document and the Risk Evaluation and Mitigation Strategy (REMS) are updated to remove the requirement that Healthcare Providers Who Insert/Remove PROBUPHINE meet the criteria for a proceduralist.

**BRIEFING DOCUMENT**

**Executive Summary, Procedural Training Program, Section 1.4., Page 3 of Executive Summary.**

The Sponsor conducted a comprehensive evaluation of its training program for proper insertion and removal of Probuphine implants. This evaluation was undertaken to ensure that the favorable safety profile regarding insertion and removal procedures observed in clinical studies is continued post-approval. The evaluation showed that only HCPs who perform sterile procedures in their clinical practice (referred to herein as proceduralists) could reliably be certified to perform the insertion and removal procedures. As a result, the Sponsor determined that only certified proceduralists HCPs who have successfully completed the Sponsor’s training program and demonstrated procedural competency should be certified to perform the insertion and removal procedures. The training program was thereafter validated in a summative study that included a user group of 15 proceduralists HCPs comprised of physicians and other HCPs with experience in sterile procedures. This summative validation study showed that the training program successfully prepared proceduralists the HCPs to safely perform the insertion and removal procedures, including implanting Probuphine at the proper depth and distribution, palpation and removal.

**Validation of Procedural Training Program, Introduction, Section 6.1., Page 65 of Training Program.**

Over the course of developing Probuphine, it became clear that the insertion and removal of Probuphine required procedural expertise complemented by appropriate training and procedural competence.

To understand and build an appropriate training program that reflected this observation, the Sponsor engaged the National Center for Human Factors in Healthcare (NCHFH) within MedStar Health to design and conduct a thorough human factors validation of the initial Probuphine training program and associated instructional materials.

One consequence of this was a decision to differentiate between healthcare providers who were trained to and routinely perform procedures and those without procedural experience. As a result the Sponsor will limit insertion and removal to proceduralists trained to conduct the necessary minor surgical procedure. Physicians and midlevel providers are considered a proceduralist for these purposes if they had completed a medical residency or fellowship in a procedural specialty and currently practiced in that specialty, or if they had performed a sterile procedure with
suturing in the last 3 months. **require all HCPs who insert and remove Probuphine to complete a robust training program and demonstrate procedural proficiency.** In addition to educating HCPs on the proper insertion and removal procedures, the training program educates HCPs on the risks of (1) complications of, protrusion, expulsion, and nerve damage associated with the improper insertion and removal of Probuphine, and (2) accidental overdose, misuse, and abuse if an implant comes out or protrudes from the skin. **To be certified as an HCP who inserts and removes Probuphine, the HCP must demonstrate procedural competency under the observation of a master trainer.**

As described in Section 5, over time a number of SAE and AE problems decreased, which was likely attributable in part to improved procedural proficiency and training among healthcare providers performing the insertion and removal providers as well as improve instruments for the procedure.

The primary purpose of the human factors evaluation was to validate the Probuphine training program’s effectiveness in preparing end-users (proceduralists) to perform the Probuphine insertion and removal procedures. Consistent with FDA’s draft human factors guidance, a robust, iterative human factors program was designed and executed. The program included a series of expert human factors reviews and formative studies that informed the evolving training program and materials and ultimately resulted in a final program that was validated for incorporation into the proposed REMS.

The Sponsor initially intended to validate the effectiveness of the training program for training 2 different user groups: proceduralists and non-proceduralists. However, based on data obtained from several studies with the non-proceduralists, a decision was made to seek FDA approval only for the effectiveness of the training program to prepare proceduralists for safe use of the insertion and removal procedures. Consistent with FDA observations during the 2013 Advisory Committee, the Sponsor determined that non-proceduralists likely would require more hands-on training and require supervised procedures before being able to be qualified to perform the insertion and removal procedures safely and effectively.

The objective of the human factors evaluation and validation was to validate a training program.

The figure below Figure 19 presents a brief recap of the evolution of the human factors work since collaborating with MedStar Health’s National Center for Human Factors in Healthcare (NCHFH) in 2013. This thorough and iterative approach allowed for the inclusion of user feedback from previous clinical studies, human factors experts, healthcare training professionals, and a variety of intended users participating in numerous different studies across different healthcare settings. In the end, the training program was finalized that, in its current design, adequately prepares proceduralists to perform the Probuphine insertion and removal procedures. Figure 19 shows the evolution of the human factors program.

Validation of Procedural Training Program, Formative Human Factors Evaluation, IFU and Video Comprehension Study, Section 6.5.4., Page 74 of Training Program.

After the IFU and videos were complete, a comprehension study was designed and conducted with a representative group of 15 proceduralist and non-proceduralist participants HCPs. Comprehension studies can help to identify potential areas for labeling improvements.

Participants were asked to read specific sections of the IFU and watch parts of the videos. After each section, time was allowed for the moderator to ask knowledge comprehension questions. The most critical finding was that proper knowledge of operation and functionality of the Probuphine applicator was difficult to retain, likely due to the inability to become familiar with the actual applicator during a read-only session. The final recommendation was that the training materials should suffice as long as trainers were sure to spend an appropriate amount of time during the hands-on experience to carefully identify the components of the Probuphine
applicator, how the applicator functions, and best practices on how to use it to safely and properly insert Probuphine.

Validation of Procedural Training Program, Summative Human Factors Validation Study Section 6.6., Page 75 of Training Program.

A summative usability evaluation is the final validation intended to evaluate that all risks have been identified, adequately addressed and that the impacts of residual risks are minimized. A total of 15 proceduralists HCPs completed the training program and validation study. The study components included the training, the IFU, the quick guides and the kit components. There were 2 primary procedures of interest, each of which has several subtasks: 1) Insertion, and 2) Removal.

Validation of Procedural Training Program, Summative Human Factors Validation Study, Complete Training Program Validation, Section 6.6.6., Page 76 of Training Program.

Overall, participants performed well for their first unaided trials after participating in the full training program, and gave positive subjective feedback on the content and presentation of the training program. As previously described, participants performed exceptionally well across the board. Based on the validation study results, the training program successfully prepared participants to safely perform the insertion and removal procedures during their first unaided attempt. In addition, these findings indicate that the Probuphine training program prepares the intended users appropriately, and that these procedures are learnable and doable. It is expected that, as the proceduralist HCPs performed more procedures, their confidence and performance would will continue to improve as with any surgical procedure.

Validation of Procedural Training Program, Summary of the Human Factors Validation, Section 6.7., Page 76 of Training Program.

A robust summative human factors validation study was designed and executed, resulting in a comprehensive evaluation of all steps involved with the Probuphine insertion and removal procedures and the associated training components. The final Probuphine training program properly prepared the intended users with the knowledge and skills to safely complete both the insertion and removal of Probuphine while minimizing risk of harm to patient. No patterns of preventable error were evident that would suggest training program or instructional material changes are required at this time for deployment of this product for proceduralists HCPs. The Sponsor reviewed and modified its proposed REMS program consistent with the findings from the human factors evaluation and validation.

Benefit and Risk, Conclusion of Benefit Risk Assessment, Section 8.3., Page 84 of Risk Mitigation.

Probuphine has been studied in 7 clinical studies, including 3 randomized controlled studies. It has been demonstrated to be effective for maintaining patients stabilized on buprenorphine. It is expected to offer greater convenience, adherence, and ultimately satisfaction with treatment. It also has the potential to dramatically reduce the risks of abuse, misuse, diversion, and accidental pediatric exposure associated with daily-dosed buprenorphine.
The risks are in general those of buprenorphine and can be characterized as minor, transient and readily managed. The specific risks associated with the procedure have been minimized by improvements in the tools, techniques, and training, and certification of certified proceduralists HCPs for performance of the insertion and removal procedures.

Probuphine is well-suited to the needs of clinically stable patients. It brings opioid addiction treatment to the level of care and simplicity of treatment of other chronic disease states.

APPENDIX D: PROPOSED RISK MITIGATION AND EVALUATION STRATEGY (REMS)

I. GOALS

The goal of the Probuphine REMS is to mitigate the risk of complications of protrusion, expulsion, and nerve damage associated with the improper insertion and removal of Probuphine and the risks of accidental overdose, misuse and abuse if an implant comes out or protrudes from the skin by:

The goals of the PROBUPHINE REMS program are to mitigate (1) the risks of complications related to the PROBUPHINE insertion and removal procedures and (2) the risks of PROBUPHINE accidental overdose, misuse, and abuse by:

A. Educating, training, and certifying healthcare providers on appropriate patient selection for PROBUPHINE, proper PROBUPHINE insertion and removal procedures, and the risks of accidental overdose, misuse, and abuse associated with PROBUPHINE; and

B. Establishing a Closed Distribution System that ensures PROBUPHINE is distributed directly and exclusively to certified healthcare providers and that only certified healthcare providers perform the PROBUPHINE insertion and removal procedures.

A. Ensuring that prescribers are educated on the following:
   1. Risk of complications of protrusion, expulsion and nerve damage associated with improper insertion and removal of Probuphine
   2. Risks of accidental overdose, misuse, and abuse if the implant comes out or protrudes from the skin.

B. Ensuring that Probuphine is administered only to patients informed about the risks of complications of protrusion, expulsion, and nerve damage associated with improper insertion and removal, as well as, risks of accidental overdose, misuse, and abuse if any implant comes out or protrudes from the skin.

REMS Elements, Elements to Assure Safe Use, Section II, 2, e, Page 2.

e. Attest and agree on the PROBUPHINE REMS Program Healthcare Provider Who Inserts/Removes Enrollment Form to:

i. Be a U.S. state licensed healthcare provider who qualifies as a Proceduralist in one of the following ways:

   a) Completed a medical residency or fellowship in a Procedural Specialty and currently practices in that specialty at the time of
PROBUPHINE REMS Program enrollment. A Procedural Specialty is a practice that entails performing invasive procedures involving injection of local anesthetic and use of sterile technique, such as (but not limited to) anesthesia, surgery, obstetrics and gynecology, dermatology, emergency medicine, and critical care; or

b) At the time of the PROBUPHINE REMS Program enrollment, performed a sterile procedure in the last 3 months, which includes (but is not limited to) using aseptic technique, injecting local anesthetic, making skin incisions, placing sutures, and removing foreign objects; and

ii. Perform the PROBUPHINE insertion and removal procedures only in a clinical setting with proper equipment and materials to perform the implant insertion and removal procedures as set forth in the PROBUPHINE® Instructions for Use.

3.5.4 Non-Inferiority Design

The Sponsor considered the relative suitability of superiority and NI analysis before choosing NI. While it was agreed that a superiority study is usually preferable, there were 2 principal reasons why NI was chosen. First, in a population already stabilized on the standard of care, SL BPN, the number of responders would be expected to be high in both arms. As a practical matter, this leaves a very small margin for proving superiority. Evaluating clinically stable patients in an active comparator design presented inherent challenges in estimating the sample size required to demonstrate superiority.

Second, NI was deemed sufficient to establish what HCPs and patients most need to know about Probuphine, a different formulation of an established product: that transfer from SL BPN to Probuphine can safely and effectively maintain stability while allowing patients to access the benefits it provides in terms of convenience, reduction of stigma, and reduced risks of diversion, abuse, misuse and accidental pediatric exposure. A superiority design was thus viewed as unnecessary.

3.5.4 Non-Inferiority Design

Choice of Study Design: Non-inferiority vs Superiority

Four 80 mg Probuphine implants are expected to approximate the plasma concentrations of BPN observed following daily SL BPN doses of 8 mg or less. Previous clinical trials have demonstrated the efficacy of SL BPN doses of 8 mg/day or less for the maintenance treatment of opioid dependence (Johnson et al., 1992; Johnson et al., 1995; Ling et al., 1998). In addition, post-market studies have shown that clinicians are effectively treating many patients with maintenance BPN doses of 8 mg or less (Apelt et al., 2013; Mattick et al., 2008; Meade et al., 2010). The needs of patients who have been effectively maintained on relatively low SL BPN doses and require less frequent follow-up visits may be better met by Probuphine than by SL formulations. In addition, Probuphine provides an alternative dosage form that can reduce
diversion, misuse, abuse and accidental exposure. Therefore, the purpose of this study was to
demonstrate the maintenance of the safety and efficacy by an alternate delivery form of BPN,
Probuphine, in the continuing treatment of opioid dependence in clinically stabilized SL BPN
maintenance patients.

The design selected to meet the objectives of this study was a 24-week randomized double-
blind, double-dummy study with SL BPN as an active comparator in adult outpatients with
opioid dependence who are clinically stabilized on 8 mg or less of SL BPN. Probuphine was
compared to SL BPN using a non-inferiority analysis. Given the pharmacokinetic data
showing that four 80 mg Probuphine implants produce BPN plasma concentrations similar to
a daily SL BPN dose of 8 mg or less, the proposed design is consistent with other trials
evaluating the transfer of subjects to alternative dosage forms, where the overall plasma
concentrations have been demonstrated to be similar, such as the transfer from once-daily to
weekly dosing of anti-diabetics (Gastaldelli et al., 2013).

Braeburn and FDA agreed that conducting a superiority trial vs placebo would be considered
unethical in stable patients. Braeburn also considered the option of conducting a superiority
trial vs an active control (SL BPN) based on the hypothesis that improved compliance could
result in better outcomes for Probuphine vs SLBPN. Ultimately, the sponsor concluded that a
properly-powered superiority design would not be feasible, for the following reasons:

a) The patient population we intended to study was clinically stable and already judged to be
optimally treated. Due to this ceiling effect, it would be unlikely that these patients would
have room to improve significantly as a group. The only way for Probuphine to show
superiority over the SLBPN formulation would be for the patients who are randomized to
the SLBPN arm to deteriorate. While there was some potential for this to occur, it
appeared less likely since these patients would have been doing well on their SLBPN
therapy for some time.

b) While the potential for improved compliance to lead to improved efficacy outcomes did
exist, it was considered difficult to estimate the effect size for such an improvement.
Therefore, we could not reliably calculate the sample size for a superiority design where
overall event rates were expected to be very low (similar to stable epilepsy patients having
seizures) and where no data existed on the difference between the two treatment arms.

c) Any reasonable estimation of the potential difference would require a large study to
demonstrate superiority. For example, hypothesizing a 5% (85% vs. 90%) difference in
response rates, and a 90% power, would require a sample size of approximately 1800
patients.

In collaboration and agreement with the FDA, the sponsor then focused on the most
appropriate margin to test the hypothesis that patients could ultimately be transferred from the
SL form of buprenorphine to Probuphine effectively.
3.5.5 Non-Inferiority Margin

The NI margin to be used to plan a NI trial is generally based on how the chosen comparator for the trial performs against its active-control predicate, generally a null intervention. The NI margin is based on retaining a specified amount of the effect found for the chosen comparator relative to its predicate. Retaining 50% of this effect has been found to be acceptable in many situations. There are situations, however, where the chosen comparator performs so well against its predicate that 50% of effect retention leads to an NI margin that allows NI outcomes that cannot be regarded as clinically not inferior. In such cases the basis for identifying the NI margin must rely on clinical judgment. Buprenorphine maintenance treatment for clinically stable patients is just such a situation, even though FDA’s guidance on NI trial design suggested that certain factors might warrant selection of a less conservative NI margin.

A clinician survey suggested a performance outcome in the range of 10% to 25% when no intervention is being used for the types of patients that settle into stable maintenance treatment, the target population for this study. In other words, only 10-25% of stable patients would be expected to maintain their stability once buprenorphine treatment is stopped. This expectation is consistent with the available literature on patients on long-term treatment with buprenorphine and methadone.

We were also told by these clinicians that the expected performance outcome for the chosen comparator, SL BPN, would be approximately 75% or more. Assuming best case performance for no intervention (25%) and the projected worst case performance for the comparator (75%), the difference expected is conservatively estimated to be 75%—25% or 50 points: half of the patients maintaining stability. Using 50 points as the expected effect size and the 50% retention of effect rule the corresponding NI margin is computed as 25 points.

Thus, when the expected performance of the comparator is 75% then a 25-point margin would mean that 50% or less performance level for the experimental intervention would show it provides inferior treatment and outcomes greater than 50% would be consistent with it being a non-inferior treatment. However, given the high expectations for continuing treatment of clinically stable patients, clinicians would question these conclusions regarding this 50% experimental-performance threshold (compared to 75%) as allowing inferior clinical outcomes being defined as not inferior. To better reflect the higher bar for clinical relevance, further consultation led to a decision to use a more conservative 20 points as the NI margin, and FDA agreed to this proposed margin.

In order to compute trial size the following specifications were made: 1:1 randomization, 1-sided alpha = 0.025, power = 87.3%, control arm proportion = 0.75, margin = 0.20. These

1 FDA, Guidance for Industry: Non-Inferiority Clinical Trials (March 2010).
specifications give a requirement for 90 patients per arm (using Blackwelder’s formula for NI assessments).

Of interest are the critical outcomes for the trial, that is, those outcomes that lead to rejection of the inferiority hypothesis. The critical outcomes can be computed for a specification of the true control arm outcome; the critical outcomes are the experimental arm outcomes for which the inferiority hypothesis is rejected (leading to a conclusion of NI). Critical outcomes corresponding to a range of true control arm outcomes are computed from simulations of 10,000 trials for each control arm specification. The data are generated assuming no between arm difference. For each simulation (of 10,000 trials) the between arm critical difference is estimated, from which the critical outcome region is computed. For each simulation the percent of the 10,000 trials for which the inferiority hypothesis is rejected is also computed, and corresponds to an estimate of power for that specification of the true control arm outcome.

Table 2 provides a summary of critical outcomes and estimated power.

<table>
<thead>
<tr>
<th>Control Arm Outcome</th>
<th>Critical Between Arm Difference</th>
<th>Critical Outcomes</th>
<th>% Trials with Non-Inferiority</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.65</td>
<td>-0.087</td>
<td>≥ 0.583</td>
<td>77.6</td>
</tr>
<tr>
<td>0.70</td>
<td>-0.078</td>
<td>≥ 0.622</td>
<td>81.2</td>
</tr>
<tr>
<td>0.75</td>
<td>-0.089</td>
<td>≥ 0.664</td>
<td>86.2</td>
</tr>
<tr>
<td>0.80</td>
<td>-0.114</td>
<td>≥ 0.689</td>
<td>91.5</td>
</tr>
<tr>
<td>0.85</td>
<td>-0.122</td>
<td>≥ 0.728</td>
<td>97.0</td>
</tr>
<tr>
<td>0.90</td>
<td>-0.144</td>
<td>≥ 0.756</td>
<td>99.7</td>
</tr>
</tbody>
</table>

For example, the shaded row in Table 2 (control arm outcome is 0.75) corresponds to the trial as planned. When there is no between arm difference, the experimental arm outcomes with response estimate of 0.661 or higher (estimated between arm difference of 0.089 or greater) will lead to a NI conclusion. In terms of how the test is executed, the estimate of the lower limit of the 95% CI on the difference must be greater than 0.20 in order to reject the inferiority hypothesis.

3.5.5 Non-Inferiority Margin

The NI margin to be used to plan a NI trial is generally based on how the chosen comparator for the trial performs against its active-control, generally a null intervention. The NI margin is based on retaining a specified amount of the effect found to be acceptable in many situations. The 20 point margin is appropriate from a scientific validity perspective as well as meeting FDA guidance on NI trial design (FDA, 2010).

Although there was a paucity of literature on SL BPN treatment in long-term stabilized maintenance subjects, there was some empirical information available to lend credence to the 20 point non-inferiority margin. The sponsor relied on the available literature, supplemented
by a survey of addiction experts to estimate the expected effect size for the active control vs no
treatment (or placebo), based on which the non-inferiority margin could be proposed.

1) As summarized below, blinded taper studies following patients on longer-term BPN or
methadone treatment indicate rates of continued opioid abstinence following complete
withdrawal of only about 18% to 31%:
   a) A meta-analysis of tapered discontinuation following long-term methadone or BPN
treatment found an average abstinence rate of 33% (Korner & Waal, 2005). However,
because of the differences in methodology (single or double-blinding, naturalistic, etc.),
definitions of abstinence, treatments administered during MAT and durations of
follow-up, some studies are more relevant than others. In addition, this article didn’t
report on the baseline rates of percentage abstinence or urine toxicology results.
Breen et al., (2003) reported on a study of stable methadone patients (for at least 6
months) who were converted to BPN and then gradually tapered in blinded fashion to
0 mg BPN over an average duration of 11 weeks. At 1 month follow-up after complete
BPN discontinuation, subjects had 31% negative opioid samples (compared to about
73% negative at baseline, 89% negative during BPN induction, and 91% negative at
the beginning of BPN taper).
   b) One double-blind, double-dummy study in methadone users found 25% abstinence
overall during 1 month follow-up after discontinuation following gradual taper
regimens. Abstinence was 18% in the "rapid" withdrawal group (taper over 10 weeks)
(versus 100% negative urine opioid results for 4 weeks preceding study entry and 92%
negative for the 6 months prior to the study) (Senay, 1977), pointing to a 74% drug
treatment effect.
   c) Most of the studies used tapered discontinuation, but in terms of abrupt
discontinuation, one survey study in Australia reported that 15% of patients who
abruptly discontinued opioid maintenance therapy (BPN or methadone) were abstinent
for at least 3 months, while 26%-27% were abstinent after either self- or physician-
directed taper regimens (Winnstock et al., 2011).

2) To generate additional data to inform the proposal of an appropriate non-inferiority
margin, a survey of addiction experts was conducted by the Sponsor to estimate the
proportion of patients on a stable dose of 8 mg or less of SL BPN that would be expected to
maintain abstinence after discontinuation of the SL BPN. According to the addiction
experts, only a median 25% of clinically stabilized patients would not relapse to illicit
opioid use (i.e., maintain clinical stability) if these patients were taken off their stable dose
of 8 mg or less of SL BPN. These responses led the sponsor to estimate a 75% effect size.
The FDA Draft Guidance would thus suggest a 37.5 point margin to have been
appropriate (50% of the treatment effect). However, a 37.5 point margin, estimating a 50%
preservation of the effect size was considered less acceptable from a clinical standpoint.
The proposed 20 point margin would estimate preserving >70% of the effect and would be
considered clinically acceptable.

3) Finally, the FDA Draft Guidance notes that circumstances might support a less
conservative choice for the margin, including:
a) Pharmacologic properties of the test of the test drug that are very similar to those of the active control – Probuphine is an alternative dosage form of the same active entity with the expectation of similar overall plasma concentrations to the SL BPN arm;

b) Use of a persuasive biomarker - responder definition in this trial will include the standard and well-accepted objective urine toxicology results to confirm treatment success;

c) If the drug has been shown to be effective in closely related clinical settings – Probuphine has already been shown to be statistically superior relative to placebo in two trials in harder-to-treat populations;

d) If the test drug were shown to have some important advantage (e.g., on safety or on a secondary endpoint) – The safety issues associated with abuse of SL BPN are well-known; the Drug Abuse Warning Network confirmed an increasing trend to adverse medical outcomes associated with BPN abuse, i.e., a total of 21,483 emergency department visits related to abuse/misuse were reported in 2011 (DAWN, 2013). Probuphine has the potential to reduce misuse/abuse associated with SL BPN and have a significant positive public health impact, in addition to potentially increasing adherence.

Thus, the overall available data and circumstances associated with this study supported the use of a 20 point non-inferiority margin.

In order to compute trial size the following specifications were made: 1:1 randomization, 1-sided alpha = 0.025, power = 87.3%, control arm proportion = 0.75, margin = 0.20. These specifications give a requirement for 90 patients per arm (using Blackwelder’s formula for NI assessments).

In this study with 180 subjects, the observed data meeting the statistical 20 point margin would also be expected to meet scientific face-validity for the equivalency of the treatment arms. The table below summarizes the required Probuphine treatment response (% of responders) to satisfy non-inferiority based on the 20 point margin and sample size of 90 subjects per group at the two-sided 5% significance level for a range of observed SL BPN arm’s response rate. The table demonstrates that utilizing the 20 point margin to satisfy non-inferiority, the Probuphine treatment group’s proportion of responders should be at least similar to the rate for SL BPN.
<table>
<thead>
<tr>
<th>Observed SL BPN Arm’s Response Rate (% of Responders)</th>
<th>Required Minimum Observed Probuphine Arm’s Response Rate (% of Responders) to Satisfy Non-inferiority (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>81.1</td>
</tr>
<tr>
<td>84.4</td>
<td>76.7</td>
</tr>
<tr>
<td>80</td>
<td>73.3</td>
</tr>
<tr>
<td>74.4</td>
<td>67.8</td>
</tr>
<tr>
<td>70</td>
<td>64.4</td>
</tr>
<tr>
<td>64.4</td>
<td>58.9</td>
</tr>
<tr>
<td>60</td>
<td>54.4</td>
</tr>
</tbody>
</table>

### 3.6.1 Analysis Sets

Five datasets were pre-specified for this study:

- **Randomized Dataset**: all subjects who completed the Screening Phase and were randomized to a treatment arm;
- **Safety Dataset**: all subjects who received study medication. Analyses based on this population grouped subjects according to the treatment they actually received regardless of the treatment they were randomized to receive;
- **Intent-to-Treat (ITT) Dataset**: all randomized subjects who received study medication and provided some efficacy data. Primary efficacy analyses were based on the ITT dataset; *[NOTE: this definition is included in the final Statistical Analysis Plan and therefore, the final study report. The study Protocol had a different definition of ITT, which included all randomized subjects who have received at least one dose of study medication. The final definition of ITT dataset (modified ITT) included only subjects with at least one post-baseline assessment. Similar ITT approaches are standard and are recommended by many FDA guidance documents as often being more appropriate than unmodified ITT.]*
- **Completer’s Dataset**: all subjects who provided all required urine toxicology samples; and
- **Per Protocol (PP) Dataset**: all subjects who had no major protocol deviations.

### 4.4.2 Primary Efficacy Endpoint

Using the primary imputation penalty for missing urine testing results and adjusting urine toxicology results for subject self-report, the proportion of responders was 96.4% in the
Probuphine arm and 87.6% in the SL BPN arm. The 2-sided 95% CI (0.009, 0.167) of the proportion difference (Probuphine – SL BPN) was well above the pre-defined successful margin for NI.

Furthermore, after establishment of NI, superiority of Probuphine over SL BPN was tested. This is in keeping with the draft FDA Guidance on Non-Inferiority, which explains:

“When there is only one endpoint and one dose of the test treatment, a planned NI study can be tested for superiority without a need for Type 1 error alpha correction . . . One can also think of this as a two-stage analysis in which the showing of NI using a 95% confidence interval (invariably successful if the test drug is actually superior), is then followed sequentially by superiority testing.”

The same guidance comments on the meaningfulness of establishing superiority in an NI study:

“In some cases, a study planned as an NI study may show superiority to the active control. ICH E-9 and FDA policy has been that such a superiority finding arising in an NI study can be interpreted without adjustment for multiplicity. Showing superiority to an active control is very persuasive with respect to the effectiveness of the test drug, because demonstrating superiority to an active drug is much more difficult than showing superiority to placebo.”

The proportion of responders was statistically significantly higher in the Probuphine arm (P = 0.034). Thus, the number needed to treat (NNT) with Probuphine is 11.36 to achieve benefit beyond the standard of care, SL BPN.

Figure 6 shows the results of the primary analysis together with 2 pre-specified sensitivity analyses, both of which confirm NI (as described further in Section 4.4.3.)

Additionally, three subjects in the Probuphine arm were treated but excluded from the primary efficacy analysis. These patients were not included in the a priori-defined ITT dataset as outlined in the statistical analysis plan. However, a post-hoc analysis was conducted using the ITT definition from the study Protocol, which included all subjects who were randomized and were treated with study medication. Missing urine samples from these three subjects were imputed using the primary imputation method versus imputing these subjects as non-responders. This was deemed to be appropriated given that these subjects did not have any evidence of illicit opioid use for 90 days prior to enrollment and were expected to continue to remain stable.

In this analysis the responders were, 84 (96.6%) Probuphine and 78 (87.6%) SL BPN, 0.089 (0.011, 0.168), p= 0.029. We consider this statistical superiority result to further support the finding of non-inferiority, as demonstrated through primary and multiple sensitivity analyses.

4.4.6.1 Supplemental Use

During the NDA review cycle, FDA asked the Sponsor to provide additional information on how subjects who required supplemental SL BPN were evaluated with respect to treatment response. Because 27 of the 28 subjects who received at least 1 dose of supplemental SL BPN were also
treatment responders, a disproportionate distribution of subjects who received supplemental SL BPN could have had a significant effect on interpretation of study results if a judgment was made that individuals who received supplemental SL BPN should not be considered treatment successes. However, use of supplemental SL BPN did not have an effect on study outcome in an NI analysis of the primary endpoint because the use of supplemental SL BPN was distributed relatively evenly between the 2 treatment arms (17.9% in the Probuphine arm and 14.6% in the SL BPN arm).

The potential effect of supplemental SL BPN use on treatment response was evaluated in 3 post-hoc sensitivity analyses. The results of these sensitivity analyses are shown in Figure 9a. First, all 28 subjects who received supplemental were removed from the primary analysis, regardless of the total doses of SL BPN received. This analysis results in a small change in the proportion of responders in each treatment arm (a decrease of 0.7% in the Probuphine arm and a decrease of 0.8% in the SL BPN arm). However, NI was still demonstrated in this analysis.

Second, all 27 responders who received supplemental SL BPN were reclassified as non-responders. This analysis resulted in significant decreases in the proportion of responders in each treatment arm (decreases of 17.8% and 13.4% in the Probuphine and SL BPN arms, respectively). However, NI was still demonstrated in the analysis of responders in each treatment arm.

Third, the 8 subjects who received >100 total doses of 2 mg SL BPN during the 6-month study were reclassified as non-responders. (The threshold of 100 doses was meant to distinguish more significant from less significant use of supplemental SL BPN. This cut off point was recommended by our co-principle investigators.) This analysis resulted in modest decreases in the proportion of responders in each treatment arm (decreases of 7.2% and 2.2% in the Probuphine and SL BPN arms, respectively). However, NI was still demonstrated in the analysis of the proportions of responders in each treatment arm.

In each of these sensitivity analysis, the proportion of subjects who were responders remained numerically higher in the Probuphine arm than the SL BPN arm. These post-hoc analyses suggest that the treatment response for subjects who received Probuphine was clinically and statistically NI to those who were treated with SL BPN irrespective of whether or not subjects who required supplemental SL BPN are included in the responder analysis or reclassified as non-responders instead of responders. Thus, the primary efficacy endpoint was met in the pre-specified primary analysis as well as additional sensitivity analyses that controlled for the potential treatment effects of supplemental SL BPN.

Although supplemental use is not necessarily a marker of lack of response in the same way as illicit opioid use, we have included additional post-hoc analysis combining time to first illicit opioid use and time to first supplemental use. The results are shown in the Figure 9b below and demonstrate no difference between the two treatment arms, p=0.191.
**Figure 2a:**  **Post-hoc Sensitivity Analyses of Supplemental SL BPN Use (ITT Population) (PRO-814)**

<table>
<thead>
<tr>
<th>Sensitivity Analyses for Proportion of Responders</th>
<th>SL BPN n (%)</th>
<th>Probuphine n (%)</th>
<th>Proportion Difference (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>28 patients who received supplemental removed from the primary analysis, regardless of the total doses of SL BPN received</td>
<td>66/76 (86.8)</td>
<td>60/69 (95.7)</td>
<td>0.088 (-0.002, 0.178)</td>
<td>0.064</td>
</tr>
<tr>
<td>27 responders who received supplemental SL BPN reclassified as non-responders</td>
<td>66/89 (74.2)</td>
<td>66/87 (79.1)</td>
<td>0.044 (-0.082, 0.171)</td>
<td>0.495</td>
</tr>
<tr>
<td>8 patients who received &gt;100 total doses of SL BPN 2 mg tablets during the 6-month study reclassified as non-responders</td>
<td>76/89 (85.4)</td>
<td>75/84 (89.2)</td>
<td>0.039 (-0.060, 0.138)</td>
<td>0.442</td>
</tr>
</tbody>
</table>

**Figure 9b:**  **Post-hoc Sensitivity Analysis of Time to First Illicit Opioid Use or Supplemental SL BPN Use**
4.4.7 Missing Urine Samples

During the review cycle, FDA asked the Sponsor to scrutinize the potential effects on treatment outcome with different imputation methods for missing urine samples and reclassification of subjects who used supplemental SL BPN. Post-hoc analyses were therefore conducted imputing all missing urine samples as positive for illicit opioids. Moreover, primary efficacy was re-analyzed using this imputation method for both the Safety analysis set (Analysis 1) and ITT analysis set (Analysis 2). Additionally, post-hoc analyses were conducted classifying all subjects in the ITT who used supplemental SL BPN as non-responders (Analysis 3), as well as imputing all missing urine samples as positive and classifying all subjects in the safety population who received at least 1 supplemental SL BPN dose as non-responders (Analysis 4).

Figure 10 Figure 10a shows the results of these 4 sensitivity analyses. In all but 1 of these analyses (Analysis 4), the proportion of responders remained numerically higher in the Probuphine arm than the SL BPN arm. However, NI was demonstrated in all 4 sensitivity analyses.
Some individual opioid panel results were incomplete for several subjects. For example, some samples for norfentanyl were unable to be analyzed due to Matrix problems at the toxicology lab. These problems can occur when there are certain samples that may contain other compounds such as natural breakdown products within the urine or potential concomitant medications that could have interfered with the chromatography of the lab’s methods. This occurred repeatedly for the same subjects. It is therefore possible that individual subject diet, concomitant medication or specific subject internal chemistry could have contributed to the matrix problems for these subjects. However, attempts to tamper with the sample in order to conceal illicit opioid use cannot be ruled out at this time.

The oxymorphone results that were not reported on several subjects were due to “interfering substances”. When morphine is present, a morphine metabolite, morphine N-oxide, elutes at the same retention time as oxymorphone, thereby making it impossible to determine if a peak is actually oxymorphone or morphine metabolite. All of the samples that had missing
oxymorphone also had morphine present/detected in the urine sample and were reported as positive.

Opioid/creatinine ratio values were missing in several subjects where the creatinine values were not reported due to acceptance criteria not being met for creatinine stability. For this study, the sponsor (not the analysis lab) set the rules for sample acceptability for creatinine adjustment conservatively at 7 days for refrigerated (generally, creatinine would be stable for 30 days, when samples are refrigerated). Therefore, samples that were received later than 7 days were deemed out of stability and although creatinine values were analyzed, creatinine adjusted values were not reported.

Sensitivity analyses were conducted in which missing urine samples were imputed as positive and any missing opioid panel items resulted in the entire sample being imputed as positive. Additionally, missing urine samples and missing urine opioid panel items were imputed as positive AND subjects requiring supplemental use were imputed as non-responders. Figure 10b demonstrates that the results demonstrate non-inferiority of Probuphine to SL BPN, consistent with other sensitivity analyses and supports the robustness of the primary endpoint. Figure 10b represents the results of this analysis.

Figure 10b: Post-hoc Sensitivity Analysis of Missing Samples, Missing Panel Items, and Supplemental Use (PRO-814)