FDA Briefing Document

Psychopharmacologic Drugs Advisory Committee (PDAC) and Drug Safety and Risk Management (DSaRM) Advisory Committee Meeting

November 1, 2018

Topic: New Drug Application 210417

Buprenorphine and Samidorphan for the Adjunctive Treatment of Major Depressive Disorder

The attached package contains background information prepared by the Food and Drug Administration (FDA) for the panel members of the Advisory Committee. The FDA background package often contains assessments and/or conclusions and recommendations written by individual FDA reviewers. Such conclusions and recommendations do not necessarily represent the final position of the individual reviewers, nor do they necessarily represent the final position of the Review Division or Office. We have brought New Drug Application 210417, buprenorphine and samidorphan for the adjunctive treatment of major depressive disorder to this Advisory Committee to gain the Committee's insights and opinions. The background package may not include all issues relevant to the final regulatory recommendation and instead is intended to focus on issues identified by the Agency for discussion by the Advisory Committee. The FDA will not issue a final determination on the issues at hand until input from the advisory committee process has been considered and all reviews have been finalized. The final determination may be affected by issues not discussed at the advisory committee meeting.

Table of Contents

List of Tables

List of Figures

MEMORANDUM

DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

DATE: October 10, 2018

FROM: Mitchell V. Mathis, M.D.
Director
Division of Psychiatry Products, HFD-130

TO: Members of the Psychopharmacologic Drugs Advisory Committee (PDAC) and the Drug Safety and Risk Management (DSaRM) Advisory Committee

SUBJECT: November 1, 2018, Meeting of the PDAC and DSaRM for NDA 210417 buprenorphine and samidorphan sublingual tablet

1 Introduction

This single-day combined meeting of the Psychopharmacologic Drugs Advisory Committee (PDAC) and the Drug Safety and Risk Management (DSaRM) Advisory Committee will focus on a number of issues critical to the Center for Drug Evaluation and Research (CDER) in evaluating a pending New Drug Application (NDA) for a combination product containing buprenorphine (BUP) and samidorphan (SAM), referred to as ALKS 5461 during development and as BUP/SAM hereafter in this document. The combination product is intended to be used as an adjunct to established antidepressant pharmacotherapy in the treatment of major depressive disorder (MDD). BUP/SAM is an opioid combination product, and it is noteworthy that, if approved, BUP/SAM would be the first drug in an entirely new class of drugs for MDD. No other opioids have been formally evaluated for the treatment of depression.

The Applicant's development program includes numerous studies to evaluate the efficacy and safety of BUP/SAM; two studies, ALK5461-202 (Study 202) and ALK5461-207 (Study 207), are intended to provide substantial evidence of efficacy. The NDA raises a number of novel issues with respect to the demonstration of efficacy, and there are a number of safety issues that merit discussion.

Most of the studies in the development program utilized a novel sequential parallel comparison design (SPCD), originally proposed by Fava *et al* to reduce the high placebo response rates that

often plague studies of antidepressant drugs, and to increase statistical power[1]. Although SPCD has been discussed in clinical and statistical circles within CDER for several years, this is the first time that studies with a sequential parallel comparison design have been submitted to the Division of Psychiatry Products to provide evidence of efficacy for a new drug. The SPCD design includes two stages, which include separate randomizations and treatment comparisons in somewhat different populations. In Stage 1, patients who meet study entrance criteria are randomized to drug vs. placebo, with a skewed randomization ratio that places far more patients in the placebo group. Stage 2 includes only non-responders from the Stage 1 placebo group, and these patients are re-randomized to drug vs. placebo. Results from the two stages are merged to provide a single overall test of hypothesis. There are a number of unresolved statistical questions regarding the most appropriate method for analyzing the results of an SPCD study. We would like to discuss the SPCD design and its issues of interpretation with the PDAC and the DSaRM Advisory Committee.

Typically, studies for antidepressant therapies assess an efficacy endpoint at a specific time point following several weeks (6 to 12) of treatment, in part because the treatment effect may require several weeks to be manifested, and in part to provide evidence of durability of the treatment effect. Here the Applicant chose a different tactic, averaging the values from each patient over several weeks. Although the Division did not prospectively agree with the Applicant on this approach, it seems worthy of consideration. Averaging scores from multiple weeks can decrease the variability in mood disorder symptoms that tend to fluctuate over time; the trade-off is that scores at the final time point carry less weight, i.e., loss of efficacy at end-of-treatment is less important in the analysis.

The Montgomery Åsberg Depression Rating Scale is a 10-item diagnostic questionnaire used to measure the severity of depressive episodes in patients with mood disorders (MADRS-10). The Applicant used an abridged 6-item version of the MADRS-10 for the primary endpoint of one of the principal studies (Study 207). The Division had rendered advice explicitly against this plan, based on analyses of the MADRS-10 and MADRS-6 by both the Division and the Agency's Clinical Outcomes Assessment (COA) Staff. The COA Staff had concluded that the MADRS-6 could not replace the MADRS-10 for use as a primary endpoint because the abridged questionnaire excludes concepts that are relevant and important in MDD, specifically "reduced sleep," "reduced appetite," "concentration difficulties," and "suicidal thoughts."

Regarding safety, buprenorphine is an opioid. Although the Applicant has made several arguments that the other component of their drug, samidorphan, negates the μ-opioid properties of buprenorphine, this has not been conclusively proven. Although we agree with the Applicant that there are fewer opioid properties for this combination product than there likely would have been from the opioid alone, there remains some evidence of a mild opiate effect (including mild withdrawal effects) from the trials.

[1] Fava M, AE Evins, DJ Dorer, and DA Schoenfeld, 2003, The problem of the placebo response in clinical trials for psychiatric disorders: Culprits, possible remedies, and a novel study design approach. *Psychother Psychosom*, 72(3):115-127.

As part of the FDA's commitment to addressing the current opioid crisis in the US, the Agency has adopted a comprehensive approach to evaluating the risk-benefit balance of opioid-containing products that includes consideration of efficacy and safety in patients when used as directed, as well as potential risks to patients and others related to opioid misuse and abuse. To provide contextual information and help inform this broader consideration of the risk-benefit balance for this product, this background package includes an epidemiologic review (attached) covering the following topics: use, misuse, and abuse of currently marketed buprenorphine and buprenorphine-naloxone products, and the complex relationships between depression, pain, and substance use disorders.

We also note a number of difficulties in analyzing safety data in SPCD trials, and would appreciate some feedback from the Committees regarding potential approaches for display of the results.

The following are the draft points under consideration:

1. Has substantial evidence been presented by the applicant to support a claim of effectiveness for buprenorphine and samidorphan for the adjunctive treatment of major depressive disorder?

2. Has the applicant adequately characterized the safety profile of buprenorphine and samidorphan for the adjunctive treatment of major depressive disorder?

3. Related to the potential risks associated with the use, misuse, and abuse of BUP/SAM in post-market settings:
    a) Discuss any concerns you have about the risks of misuse, abuse, addiction, or overdose with BUP/SAM, in the intended patient population or in others who may access the drug.
    b) Discuss any concerns you have more generally about approving an opioid-containing product for the first time for the treatment of depression, and at a time of widespread opioid abuse, addiction, and overdose.
    c) Discuss whether you are concerned about the risk of opioid overdose if other opioids are used, misused, or abused concurrently with BUP/SAM, perhaps in higher doses intended to overcome BUP/SAM's μ-opioid antagonist effects, with potential additive opioid agonist effects that have not been fully characterized.
    d) What risk reduction strategies could be implemented to decrease risks associated with BUP/SAM use?

4. Do the available data support a favorable benefit-risk profile of buprenorphine and samidorphan to support approval?

5. What, if any, additional data are needed pre- or post-approval to address outstanding issues? Please be clear whether you believe these data should be required prior to approval.

2 Purpose

The purpose of this Advisory Committee meeting is to obtain input from the Committees on whether data provided by the applicant supports efficacy, safety, and a favorable benefit-risk profile of buprenorphine and samidorphan to support its approval for adjunctive treatment of major depressive disorder (MDD).

3 Adjunctive Treatment of Major Depressive Disorder

Major depressive disorder is a debilitating and chronic illness. A 2018 World Health Organization (WHO)[2] notes that depression is "the leading cause of disability worldwide, and is a major contributor to the overall global burden of disease." This disease is characterized by low mood, anhedonia, feelings of guilt and worthlessness, low energy, and other emotional and physical symptoms. In severe cases, MDD can result in suicide.

Partial response to pharmacologic treatment is common. In the National Institute of Mental Health (NIMH)-sponsored Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial[3], only 28% of patients achieved remission (defined in this study as a score of ≤ 7 on the 17-item Hamilton Depression Rating Scale ($HAM\text{-}D_{17}$)) during first-line treatment with a selective serotonin reuptake inhibitor (SSRI).

The $HAM\text{-}D_{17}$ is one of the two most commonly used measures in clinical trials to assess symptoms of depression. The other is the MADRS-10, described above, which was used in the BUP/SAM development program.

To date, only three drugs have received an indication for the adjunctive treatment of MDD—aripiprazole, quetiapine XR, and brexpiprazole. The approved products are atypical antipsychotic drugs, whereas BUP/SAM is an opioid. If approved, BUP/SAM would be the first drug in the opioid class of drugs for MDD, as there is no prior regulatory experience with opioids in the treatment of depression.

In addition to adjunctive treatment with one of the atypical antipsychotic drugs approved for this indication, other strategies commonly used by clinicians when patients have an inadequate response to antidepressant treatment include:

- maximizing antidepressant doses
- changing to a different antidepressant (different class or same class)

[2] World Health Organization. Fact Sheets: Depression, accessed October 9, 2018, http://www.who.int/en/news-room/fact-sheets/detail/depression.

[3] Trivedi MH, AJ Rush, SR Wisniewski, AA Nierenberg, D Warden, L Ritz, G Norquist, RH Howland, B Lebowitz, PJ McGrath, K Shores-Wilson, MM Biggs, GK Balasubramani, M Fava and STAR*D Study Team, 2006, Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: Implications for clinical practice, *Am J Psychiatry*, 163:28-40.

- augmentation of effect by combining two different approved antidepressants with different mechanisms of action
- adding or changing psychotherapy
- electroconvulsive therapy (ECT)
- off-label augmentation with lithium, thyroid hormone, anticonvulsants, or stimulants

4 Product Under Review

The product under review is a fixed-dose combination of buprenorphine (BUP), a μ-opioid receptor partial agonist and a κ-opioid receptor antagonist, and samidorphan (SAM), a μ-opioid receptor antagonist. The Applicant states that SAM is intended to reduce the risks of abuse and dependence with BUP. The product is formulated as a tablet for sublingual administration; it would be available in BUP/SAM 2 mg/2 mg, 1 mg/1 mg, and 0.5 mg/0.5 mg strengths (designated as 2/2, 1/1, and 0.5/0.5 throughout this document). The target dosage for most patients would be BUP/SAM 2 mg/2 mg per day.

Buprenorphine was approved in the United States as Subutex (NDA 020732; Indivior, Inc.) in 2002. This application is being reviewed under the 505(b)(2) regulatory pathway, relying in part on the Agency's findings of safety and effectiveness for Subutex. Samidorphan is a new molecular entity (NME), and is not currently marketed in the United States or elsewhere.

5 Regulatory Background

We present the salient regulatory background here in some detail so that the PDAC and DSaRM Advisory Committee can understand the genesis of some of the lack of alignment between the Applicant and FDA on this NDA.

The development program for BUP/SAM for the adjunctive treatment of MDD began with a pre-IND meeting on February 17, 2011. At the time, samidorphan was being investigated for the treatment of alcohol dependence and for the treatment of binge eating disorder. As a result, there was considerable human experience with samidorphan in clinical trials prior to the initiation of this development program.

In the pre-IND meeting briefing materials, the Applicant described their plans for an 8-week safety and tolerability study (ALK33BUP-201; Study 201) exploring different ratios of buprenorphine to samidorphan. They planned to use the information gleaned from Study 201 to inform the dosing strategy for what they called a "proof-of-efficacy" trial (ALK33BUP-202; Study 202), in adults with MDD and an inadequate response to antidepressant therapy. The Applicant also wished to discuss their plan to use a Sequential Parallel Comparison Design (SPCD) trial.

During the meeting, the Division and Applicant reached agreement that, to obtain an indication for adjunctive treatment of MDD, the clinical studies should be conducted in a population of

patients already on stable antidepressant treatment using an add-on design. Typically, for approval of a fixed-dose combination product such as BUP/SAM, an applicant must show that both components contribute to efficacy (i.e., drugs A and B need to be shown to be effective, and the combination (A+B) must be demonstrated to be more efficacious than both A used alone and B used alone). For two unapproved drugs, the combination rule is most efficiently fulfilled using a full factorial design (A vs. B vs. A+B vs. placebo). Problems with a full-factorial design to evaluate BUP/SAM were identified: the administration of buprenorphine alone in doses up to 8 mg per day in opioid-naïve patients could be dangerous. Thus, the Division advised the Applicant that a buprenorphine-only arm was not necessary; however, a samidorphan-only (plus SSRI/serotonin-norepinephrine reuptake inhibitor (SNRI)) arm was advised to ensure that efficacy of the buprenorphine-samidorphan combination was not due to samidorphan alone.

The Division voiced no objection to a SPCD in a proof-of-concept study, but strongly encouraged the Applicant to provide a detailed statistical analysis plan (SAP) and seek feedback prior to initiating the trial if they intended to use the study to support an efficacy claim. The Applicant clarified that Study 202 was intended to be used as a proof-of-concept study, and not to establish efficacy.

The Division also advised the Applicant that a fixed-dose study could help to determine the optimal dosing regimen and asked the Applicant to address concerns regarding how they would establish the optimal dose for each component of the combination. The Division noted that the Applicant would need to assess potential interactions between SSRIs, SNRIs, and buprenorphine, samidorphan, or the combination; they would also need to conduct a comprehensive abuse potential assessment.

The IND was submitted on April 8, 2011, with Study 202 as the opening trial. Early in the development program (April 30, 2013), the Applicant submitted a request for feedback regarding the nonclinical and clinical requirements needed to support an overall Controlled Substances Act (CSA) scheduling recommendation for BUP/SAM, and on the adequacy of the samidorphan data package for the Agency to support a petition with the Drug Enforcement Agency (DEA) to decontrol SAM. The Agency noted that FDA makes a recommendation for scheduling during the NDA review. The Agency also informed the Applicant that a label statement on the reduced drug liking and reduced abuse potential of BUP/SAM relative to BUP alone would need to be supported by an array of nonclinical and clinical studies, as well as post-marketing epidemiologic data.

In late 2013, the Applicant requested an End-of-Phase 2 meeting. In their meeting background package, the Applicant described their plan to conduct three phase 3 confirmatory efficacy studies (ALK5461-205, -206, and -207) and one long-term safety study (ALK5461-208). The Applicant was planning to use change from baseline in total MARDS score as the primary endpoint in all three of the planned phase 3 studies.

In preliminary meeting comments conveyed to the Applicant, the Division again expressed concerns about the planned SPCD analyses:

From a statistical perspective, although the proposed SPCD appears to be reasonable, there

has been no analytical proof for the validity of associated statistical analyses when there are missing data. In the cited Chen et al. paper, the type I error rates were estimated by simulation. Without theoretical proof, it is not guaranteed that the type I error rate will be controlled, especially in scenarios where there are extensive dropouts. Since statistical validity of the methods associated with this novel design is not yet clear when there are missing data, it will be a matter of review whether or not efficacy demonstration can primarily rely on this method. We note that you have pre-specified the MMRM approach as outlined by Chen et al. as the primary analysis and a few sensitivity analyses. To further assess the impact of missing data, you should propose sensitivity analyses that do not require the MAR (missing at random) assumption and provide details in the Statistical Analysis Plan.

After receiving the Division's preliminary responses, the Applicant withdrew the meeting request, noting that our preliminary responses sufficiently addressed their questions.

Following submission of the three phase 3 protocols (Studies 205, 206, and 207) and the long-term safety protocol (Study 208), the Agency advised the Applicant to monitor for respiratory depression throughout Study 208, and refine the time points for administration of the Clinical Opiate Withdrawal Scale (COWS). We also advised the Applicant to include monitoring for drug use (including prescription and over-the-counter medications as well as prohibited substances) and urine drug screening throughout all four proposed trials.

On May 12, 2015, the Applicant requested a meeting to discuss and reach agreement on the Statistical Analysis Plans (SAPs) for Studies 205 and 206. The Agency provided written responses to the Applicant's background package on July 24, 2015. In our advice, we noted:

We would like to reiterate that we haven't endorsed any analytical method for SPCD in a confirmatory trial setting. We are continuing making efforts in further understanding its pros and cons from a regulatory perspective. You are encouraged to collect efficacy data from both stages. However, we may determine the efficacy based on data from only Stage 1 if analysis associated with this novel design is still unsettled by the time of your NDA filing.

Because of the limited time available for review of submissions via a meeting category, we can only provide general guidance on the proposed questions. If there is any change in protocol or stand-alone SAP, we advise that you submit it (including tracked changes and/or a detailed list of changes) separately from a meeting package.

On September 26, 2016, the Applicant met with the Agency to share preliminary results from Studies 205 and 206. The Applicant acknowledged that neither study met its prespecified primary endpoint and inquired about any additional analyses that could be conducted. The Agency had no recommendations, but acknowledged that the additional analyses the Applicant already conducted could be informative for subsequent studies. For instance, the unadjusted *p*-value < 0.05 for the 2/2 dose based on exploratory analyses in Study 205 along with the numerical superiority of 2/2 vs. placebo suggested that this dose was more likely to be effective than the 0.5/0.5 dose.

The Applicant submitted an amendment to the SAP and protocol for Study 207 on September 19,

2016. The cover letter for this submission referenced the then-upcoming September 26 meeting; however, the revised SAP could not be adequately reviewed within that 7-day time frame and was not discussed during that meeting. The amendment to the analytical plan changed the primary efficacy endpoint from change from baseline to end-of-treatment on the MADRS-10 to three primary endpoints to be evaluated in a hierarchical fashion:

- Change in MADRS-6 using average of changes from baseline to Week 3 through the end of efficacy period (Week 5 for Stage 1; Week 6 for Stage 2)
- Change in MADRS-10 score using average of changes from baseline to Week 3 through the end of efficacy period (Week 5 for Stage 1; Week 6 for Stage 2)
- Change in MADRS-10 score from baseline to end of treatment (Week 5 for Stage 1; Week 6 for Stage 2)

These changes were ultimately discussed during a February 13, 2017, guidance meeting. In advance of the meeting, the Agency provided the following comments relative to the efficacy analyses:

1. *In general, we do not accept major changes, such as revising the primary efficacy measures, in the late stage of a clinical trial. It appears that the primary endpoint and duration of the efficacy period for Stage 2 were changed very late in the course of the study.*

2. *We have not previously accepted the MADRS-6 as a primary efficacy endpoint for a clinical trial. Before accepting this instrument as primary endpoint in a trial intended to support product registration, we would need data on the validity and reliability of the instrument, and clear documentation of how the biometric properties of the MADRS-6 compare to the MADRS-10. On face, we have concerns that the MADRS-6 omits diagnostically and clinically important aspects of depression.*

3. *We do not agree with the strategy of comparing the baseline MADRS-6 or MADRS-10 scores to the average of the scores from Week 3 to the end of the efficacy period. We note that the averaging of the change in MADRS-6 or MADRS-10 scores tends to obscure a possible dropoff in drug efficacy after the first few weeks of treatment. In Study 205, the change in MADRS-10 scores reached a peak at Week 3. In Study 207, the change in MADRS-6 and MADRS-10 scores both reached a peak at Week 4. It is important for us to know whether the drug has an effect that persists until the end of the study. We recommend using a single efficacy measure at the end of the study, and not an average over multiple time periods, as the primary efficacy endpoint.*

4. *With the protocol amendment for Study 207, the efficacy period in Stage 1 is now different in duration from the efficacy period in Stage 2. This adds some complexity to the comparison of data from the two SPCD stages. Please provide a rationale for the difference in duration of the efficacy periods.*

During the meeting, the Applicant presented a slide showing continued improvement in mean change in MADRS-6 and MADRS-10 scores in the ALKS 5461 2/2 group when this group was

followed from Week 0 in Stage 1 to Week 11 in Stage 2. We noted that, although the SPCD study design limits the conclusions that can be drawn with respect to drug efficacy when treatment response is compared across stages, the analysis did help to reduce the Division's concern about a possible loss of drug efficacy after Week 4. We also recommended that the Applicant submit a dossier for the MADRS-6, including reliability, validity, scoring instructions, rationale for item selection, and justification for its use in antidepressant efficacy trials.

The Applicant also proposed pooling Studies 205 and 207; the Agency stated that such pooled analyses could only be considered exploratory.

The Applicant submitted a request for preliminary Breakthrough Therapy Designation Request advice on March 3, 2017. The Division informed the Applicant that, because we had not yet determined whether the MADRS-6 was an acceptable endpoint, and because any statistical significance in the phase 3 study results depended on *post hoc* analyses, it would be difficult for us to grant Breakthrough Therapy Designation.

On April 24, 2017, the Applicant submitted the dossier on the MADRS-6. The Division consulted the Agency's Clinical Outcomes Assessment (COA) Staff to evaluate the submission. Consistent with the Division's original thinking, the COA Staff concluded that the MADRS-6 could not replace the MADRS-10 for use as a primary endpoint because it excludes concepts that are relevant and important in MDD. The MADRS-6 excludes "Reduced Sleep," "Reduced Appetite," "Concentration Difficulties," and "Suicidal Thoughts." Furthermore, results of a factor analysis on MADRS-10 suggested that the four items removed were highly associated, but not redundant with, the six items retained.

The Agency provided the Applicant with the analyses of the COA Staff and their conclusions during a July 24, 2017, pre-NDA meeting. The Division informed the Applicant that any analyses of MADRS-6 scores would be considered exploratory.

The final portion of the NDA was submitted January 31, 2018. The Agency initially refused to file the application; however, after the Applicant clarified the analyses intended to support their efficacy claim, we agreed to review the application.

6 Clinical Development of BUP/SAM for Adjunctive Treatment of Major Depressive Disorder

The Applicant provides a description of the relevant clinical trials in their background document. Their descriptions of the protocols and statistical analysis plans reflect the content of their submissions to the Agency. However, rather than presenting the study results in a conventional way (i.e., study-by-study), the Applicant chose to present the study results by endpoint. This approach tends to obfuscate the prospectively planned analyses for each study and, for this reason, we present the study-by-study results below, with greater emphasis on the studies the Applicant views as providing evidence of efficacy. Of note, in some cases, the Applicant's interpretations of the study results are not in alignment with the those of the Agency. The discussion below will note points of disagreement, when appropriate. For additional details of

the studies and their results, please see the Draft Statistical Review (attached).

Study 202

Study 202 is one of the two studies the Applicant describes as providing evidence of efficacy and as positive on the pre-specified primary endpoint. This was a phase 2, randomized, double-blind, placebo-controlled proof-of-concept study to evaluate the efficacy and safety of once-daily BUP/SAM—2/2 or 8/8—for the adjunctive treatment of MDD. Like all these studies, Study 202 enrolled subjects with MDD who had achieved an inadequate response to one or two adequate courses of treatment with a SSRI or a SNRI. The study employed an SPCD to reduce the influence of placebo response on the overall trial outcome. Both SPCD stages were 5 weeks in duration: a 4-week treatment period followed by a 1-week taper period.

In Stage 1, subjects were randomized to receive BUP/SAM 2/2, 8/8, or placebo in a 2:2:9 ratio. Following Stage 1, non-responders were defined as subjects with a baseline HAM-D_{17} score of 14 or worse and who had achieved less than 50% improvement from baseline to the End-of-Week 4. (A lower HAM-D_{17} score indicates less depression; possible HAM-D_{17} total score ranges from 0 to 52.) At the beginning of Stage 2, placebo non-responders from Stage 1 were re-randomized to receive BUP/SAM 2/2, 8/8, or placebo in a 1:1:1 ratio. Unlike the phase 3 studies, patients who received BUP/SAM in Stage 1 did not continue active treatment in Stage 2; rather, they were switched to placebo. Thus, no patient was treated with BUP/SAM for longer than 4 weeks.

The primary efficacy endpoint was the mean change from baseline to Week 4 in HAM-D_{17} total score, comparing both doses of BUP/SAM to placebo. The analysis included only subjects who had received at least 1 dose of study drug and had undergone at least 1 primary efficacy assessment (i.e., HAM-D_{17}) after study drug administration.

The analysis of mixed model repeated measure (MMRM) was used to evaluate the change from baseline to Week 4 (in each phase) in HAM-D_{17} total score. The model included the terms treatment, visit, treatment-by-visit interaction, and baseline HAM-D_{17} total score. The primary efficacy analyses from the two stages were combined using SPCD methodology to make inferences about the trial. The pre-specified weights for Stage 1 and Stage 2 were 0.6 and 0.4, respectively.

The Applicant's results were verified by FDA and are shown below:

Table 1: Study 202 – Change from Baseline in HAM-D_{17} Total Score at Week 4 (Stage 1)

	Placebo N=95	BUP/SAM 2/2 N=20	BUP/SAM 8/8 N=20
Baseline			
N	95	20	20
Mean (SD)	23.2 (4.2)	22.7 (4.2)	21.7 (3.3)
Change from Baseline at Week 4			
N	90	17	14
Lease squares mean (SE)	-7.1 (0.6)	-9.3 (1.5)	-6.6 (1.6)
Least-squares mean difference vs Placebo (SE)		-2.2 (1.6)	+0.5 (1.7)
p-value		0.168	0.787

SE = standard error; SD= Standard Deviation; Source: Applicant's Clinical Study Report

Table 2: Study 202 – Change from Baseline in HAM-D_{17} Total Score at Week 4 (Stage 2)

	Placebo N=20	BUP/SAM 2/2 N=23	BUP/SAM 8/8 N=22
Baseline			
N	20	23	22
Mean (SD)	17.3 (8.8)	16.1 (5.9)	19.0 (5.5)
Change from Baseline at Week 4			
N	20	18	18
Lease squares mean (SE)	-1.5 (1.1)	-5.2 (1.2)	-3.3 (1.1)
Least-squares mean difference vs Placebo (SE)		-3.7 (1.6)	-1.9 (1.6)
p-value		0.020	0.241

SE = standard error; SD= Standard Deviation; Source: Applicant's Clinical Study Report

Table 3: Study 202 – Stage 1 and Stage 2 Combined Inference Based on HAM-D_{17} Total Score

	BUP/SAM 2/2	BUP/SAM 8/8
Least-squares mean difference vs Placebo (SE)	-2.8 (1.2)	-0.5 (1.2)
p-value	0.014	0.699

SE = standard error; Source: Applicant's Clinical Study Report

Unadjusted for multiplicity, the treatment effect was statistically significant for the 2/2 treatment group compared to placebo in the combined stage analysis ($p = 0.014$), as well as in Stage 2 alone ($p = 0.020$), but not in Stage 1. The results for the higher 8/8 dose were not statistically significant for Stage 1, Stage 2, or overall.

The Applicant notes that this study was "designed and conducted with the same rigor as a pivotal trial" and that, based on the prespecified primary endpoint, the trial is positive. This study was considered a proof-of-concept trial throughout the development program, however, and there was no prospective plan to adjust for multiplicity.

Typically, for adequate and well-controlled studies, FDA accepts (at most) a 5% false positive rate, i.e., $\alpha = 0.05$. In Study 202, a single endpoint was tested at $\alpha = 0.05$; however, two dose combinations of BUP/SAM were evaluated in the study. With two independent active treatment groups tested at $\alpha = 0.05$, there is a multiplicity problem that inflates the Type-I error rate. Specifically, when success for either active treatment group by itself would lead to the conclusion of a drug effect, the chance of erroneously finding a treatment effect for at least one of the doses (a false positive finding) is ~10 percent [see FDA's 2017 Guidance for Industry: Multiple Endpoints in Clinical Trials for more details]:

$$1 - (1 - 0.05) \text{ X } (1 - 0.05) = 0.0975 \approx 10\%.$$

Although the statistical analytical plan could have included a prospectively planned adjustment to control the Type-1 error rate for the dual hypotheses, no statistical adjustments were planned. In the final SAP, submitted only 1 week before unblinding, the Applicant stated:

> "*The study will be considered positive if at least one dose group of ALKS 5461 is statistically superior to placebo. No multiplicity adjustment of 2 hypothesis tests will be made due to the exploratory nature of this study.*"

This statement was consistent with FDA's understanding, based on an agreement made during an early face-to-face meeting where the Applicant stated its intent that Study 202 was to be a "pure proof of concept study."

As noted in the attached Statistical Review and Evaluation:

> "…we noted that for the phase 3 studies included in this NDA, where multiple dose levels were investigated, the Applicant pre-specified a fixed sequence testing procedure starting from the high dose, not [a] Bonferroni procedure, to control the overall type I error rate. Applying such a fixed sequence testing procedure starting from the high dose 8/8 versus placebo would have rendered Study 202 to be an inconclusive study; that is, the trial would have failed on the 8/8 dose by this fixed sequence testing procedure and no further testing could be performed. Thus, this reviewer concludes that in Study 202, statistical significance for the 2/2 dose remains inconclusive."

<u>Importance of a single patient to the nominally positive result for the BUP/SAM 2/2 group</u>:

A potential financial conflict of interest was reported at a study site in Pennsylvania that had enrolled one subject. In Stage 1, the subject was randomized to BUP/SAM 2/2, and their HAM-D_{17} decreased from 39 at baseline to 3. With removal of this single subject from the analysis, the *p*-value for the study was no longer nominally statistically significant for the 2/2 dose group.

<u>Critique of Study 202 Results</u>:

In summary, Study 202 showed a nominally statistically significant treatment effect for the BUP/SAM 2/2-mg dose group. Results for the 8/8-mg dose were numerically similar to placebo. Because a drug's efficacy is typically related to exposure, the lack of even a positive trend for the 8/8-mg dose group casts doubt on the nominally positive findings for the 2/2-mg dose group. (In

practice, finding positive results with a lower dose of a drug despite negative results with a higher dose is extremely rare.) We also have concerns that the nominally positive result for the 2/2-mg dose group was critically dependent on the contribution of a single subject. Finally, because Study 202 lacked a prospective plan to control the Type-I error rate, we cannot view it as an adequate and well-controlled trial.

Study 207

Study 207 is the second study the Applicant views as providing evidence of efficacy. The Applicant and FDA agree that Study 207 was positive on its pre-specified 1° endpoint. This was a Phase 3, multinational, randomized, double-blind, placebo-controlled, SPCD trial to evaluate the efficacy and safety of once-daily BUP/SAM—1/1 mg or 2/2 mg—for adjunctive treatment of MDD. As in Study 202, Study 207 enrolled subjects who had not achieved an adequate response to one or two courses of an SSRI or SNRI. Stage 1 was 5 weeks and Stage 2 was 6 weeks in duration. Subjects assigned to both doses of BUP/SAM were to initiate therapy at a starting dose of 0.5/0.5 mg, increasing to the target dose over 1 week.

In Stage 1, subjects were randomized in a 2:2:9 ratio to BUP/SAM 1/1 mg, 2/2 mg, or placebo, respectively. In Stage 2, subjects who were originally randomized to placebo and had not responded (MADRS-10 score > 15 at Visit 7 and < 50% reduction in MADRS score between Visits 2 and 7) were re-randomized in a 1:1:1 ratio to BUP/SAM 1/1 mg, 2/2 mg, or placebo, respectively. All other subjects continued the treatment to which they were assigned in Stage 1.

The original primary efficacy endpoint, based on the protocol submitted in 2014, was the change in MADRS total score from baseline to Week 5 in both stages. Following a late change in the SAP (prior to database lock, when still blinded), the Applicant specified these primary endpoints in the following sequence – first for the 2/2 dose, and then for the 1/1 dose:

- Change in MADRS-6 using average of changes from baseline to Week 3 through the end of efficacy period (Week 5 for Stage 1; Week 6 for Stage 2)
- Change in MADRS-10 score using average of changes from baseline to Week 3 through the end of efficacy period (Week 5 for Stage 1; Week 6 for Stage 2)
- Change in MADRS-10 score from baseline to end of treatment (Week 5 for Stage 1; Week 6 for Stage 2)

A fixed sequence approach was used to control the Type-I error rate for the 6 comparisons. Each test was evaluated at 0.05 significance level; subsequent testing was to be stopped if a test was not statistically significant.

The MADRS-10 is a 10-item instrument with total score ranging from 0 to 60. A higher MADRS score indicates more severe depression. The 10 items are: sadness, apparent sadness, inner tension, reduced sleep, reduced appetite, concentration difficulties, lassitude, inability to feel, pessimistic thoughts, and suicidal thoughts.

The MADRS-6 is a 6-item subscale of the 10-item MADRS, with 4 omitted subscales: reduced sleep, reduced appetite, concentration difficulties, and suicidal thoughts. The total score ranges

from 0 to 36.

The Division conveyed concerns with 3 aspects of the planned efficacy analyses:

- The Division expressed caution with respect to the unequal durations of Stages 1 and 2 because of concern that this would complicate data analysis.
- The Division and the COA Staff did not accept the MADRS-6 as an efficacy endpoint, because the abridged questionnaire excluded concepts that are relevant and important in individuals with MDD.
- The Division disagreed with the Applicant's planned strategy to average the MADRS results over several weeks, and recommended use of the MADRS-10_{EOT}, as used in other antidepressant studies and as previously agreed.

Beyond the 6 primary endpoints, no key secondary endpoints were specified with a prospective plan to control the Type-I error rate.

Table 4: Study 207 – Stage 1 and Stage 2 Combined Results, 6 Endpoints

	Treatment Group	
	BUP/SAM 1/1	**BUP/SAM 2/2**
	Numbers of subjects	
BUP/SAM: Stage 1/Stage 2	62/62	63/63
Placebo: Stage 1/Stage 2	273/60	273/60
	MADRS-6$_{AVG}$ (change vs. placebo)	
Least squares mean difference (SE)	-0.6 (0.62)	-1.5 (0.62)
95% CI	(-1.8, 0.6)	(-2.7, -0.3)
p-value	0.329	0.018
	MADRS-10$_{AVG}$ (change vs. placebo)	
Least squares mean difference (SE)	-0.9 (0.85)	-1.9 (0.86)
95% CI	(-2.6, 0.7)	(-3.6, -0.2)
p-value	0.277	0.026
	MADRS-10, end-of-treatment (change vs. placebo)	
Least squares mean difference (SE)	-1.3 (0.95)	-1.7 (0.96)
95% CI	(-3.2, 0.5)	(-3.6, 0.2)
p-value	0.165	0.076

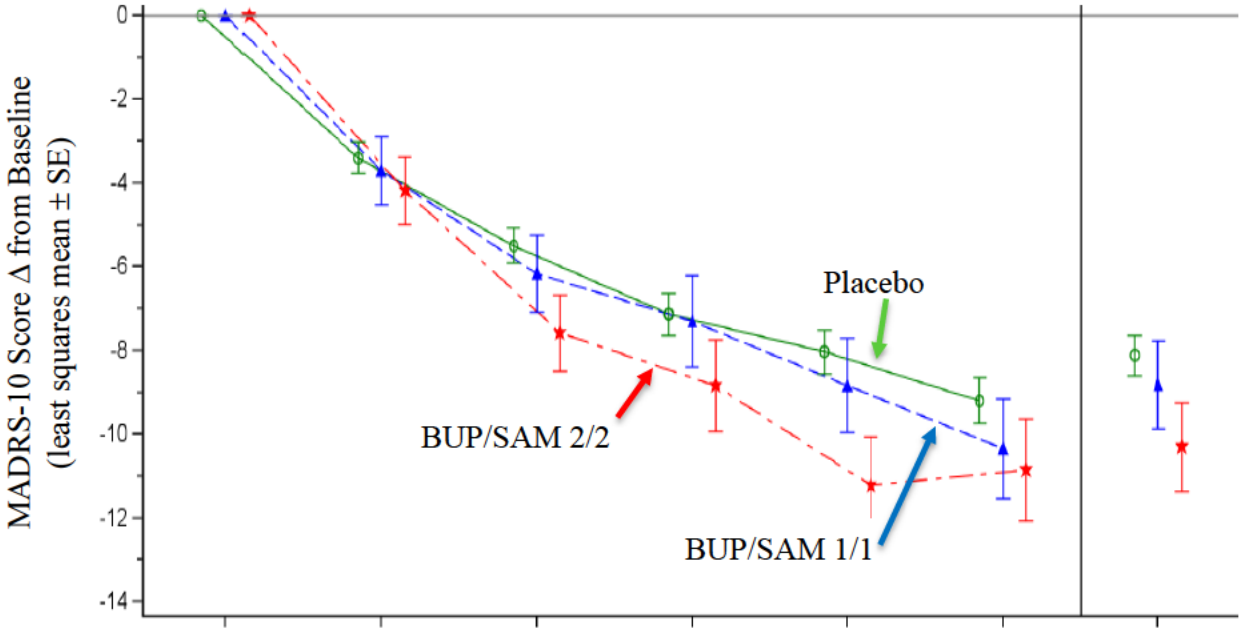
CI = confidence interval; AVG = average; SE = standard error; Source: Table 18 of Applicant's Clinical Study Report

The treatment effect for the 2/2 mg dose was statistically significant for the MADRS-6$_{AVG}$ and the MADRS-10$_{AVG}$, but not statistically significant for the MADRS-10 at End-of-Treatment, the endpoint typically used as the basis for approval of drugs for MDD. (The negative results for MADRS-10 End-of-Treatment terminated the testing sequence.) Given that both the MADRS-6$_{AVG}$ and the MADRS-10$_{AVG}$ showed statistically significant treatment effects, designation of MADRS-6$_{AVG}$ as the initial primary endpoint was not consequential. Importantly, for the 2/2 mg group, the mean placebo-subtracted treatment effect was <2 points for both the MADRS-6$_{AVG}$ and the MADRS-10$_{AVG}$ on Stages 1 and 2 combined, a treatment effect that seems modest at best.

Results for the 1/1 mg doses were not nominally statistically significant; therefore, results would have been negative irrespective of the testing sequence.

Results of Study 207 for MADRS-10 are shown graphically in Figure 1 (Stage 1) and Figure 2 (Stage 2).

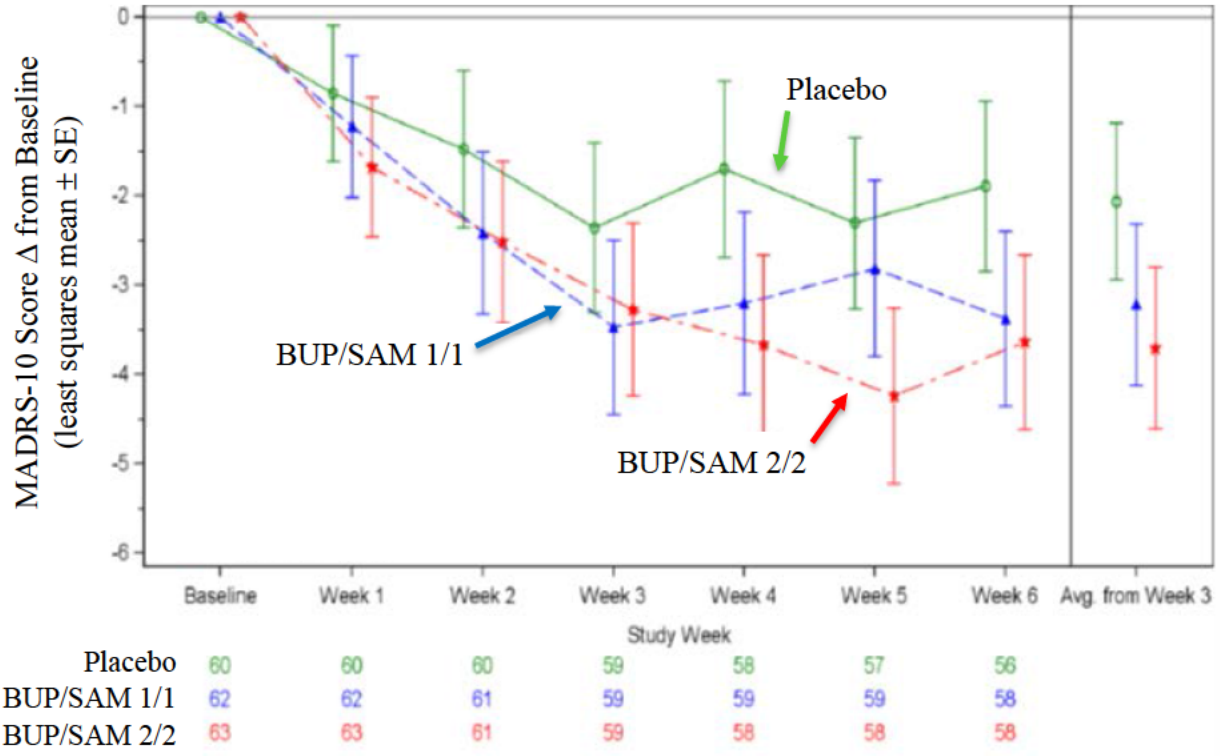
Figure 1: Study 207 – Change from Baseline in MADRS-10 Score by Visit in Stage 1



Placebo	273	273	270	265	263	257
BUP/SAM 1/1	62	62	59	56	56	55
BUP/SAM 2/2	63	63	59	53	49	49

Baseline for Stage 1 is the randomization baseline at Visit 2; SE = standard error; Source: Applicant's Clinical Study Report

Figure 2: Study 207 – Change from Baseline in MADRS-10 Score by Visit in Stage 2



Baseline for Stage 2 is Visit 7; SE = standard error; Source: Applicant's Clinical Study Report

General improvement is evident in depression scores in all treatment groups in Stage 1 (up to Week 4, Figure 1) and Stage 2 (up to Week 5, Figure 2). In the final week of both Stage 1 and Stage 2, however, the MADRS-10 tended to worsen in the BUP/SAM 2/2 group. The figures show graphically why the results for MADRS-10 were not statistically significant at the End-of-Treatment.

The Applicant believes that fluctuations in symptoms over time, the subjective nature of assessments, week-to-week variability of treatment effect, and other factors support the strategy of averaging assessments over multiple weeks. They state that a "more complete estimate of the difference between the various arms of Study 207 was achieved by utilizing data from multiple time points over treatment." We note, however, that the issues cited by the Applicant affect all antidepressant trials and, to date, all marketed antidepressants (monotherapy or adjunctive) have been approved based on improvement in symptoms from baseline to End-of-Treatment. Using a similar approach ($MADRS\text{-}10_{EOT}$) for Study 207, the results are not statistically significant.

Study 205

The Applicant also cites Study 205 as supportive evidence of effectiveness; however, they acknowledge that the trial was not statistically significant on its prespecified primary endpoint. Prior to the endpoint modification in Study 207 noted above, Studies 205 and 207 were designed similarly. Both were multicenter, randomized, double-blind, placebo-controlled trials evaluating

the efficacy and safety of BUP/SAM for the adjunctive treatment of MDD. Both employed SPCD to reduce the influence of placebo responders on the efficacy results. Study 205 explored the effect of 2/2 and 0.5/0.5 (rather than 1/1 as in Study 207), and the primary endpoint was change from baseline in MADRS-10 at Week 5 for each stage (although the total treatment period was 11 weeks: 5 weeks in Stage 1 and 6 weeks in Stage 2). The least-squared (LS) mean difference for BUP/SAM 2/2 treatment group relative to placebo was -1.8 on the Week 5 MADRS-10 (p=0.109).

The Applicant points out that, although MADRS-10 at Week 5 was not statistically significant, improvement in depressive symptoms on BUP/SAM was nominally significantly better than placebo on MADRS-10_{EOT} (end of treatment: Week 5 for Stage 1; Week 6 for Stage 2) (LS mean difference vs placebo = -2.5; p=0.025) based on a *post hoc* analysis. Of the 60 patients who received BUP/SAM 2/2 in Stage 1, 52 continued BUP/SAM 2/2 in Stage 2, and received a total of 11 weeks of treatment. Among the placebo non-responders, 56 patients were randomized to BUP/SAM 2/2 in Stage 2. Thus, the end-of-treatment analysis included 52 patients with up to 11 weeks of treatment and 56 patients with 6 weeks of treatment. In Study 207, the similar end-of-treatment analysis included 48 patients with 11 weeks of treatment and 63 patients with 6 weeks of treatment.

Given that the end-of-treatment analysis was nominally positive in the study with a greater proportion of patients with longer duration or treatment (48% of Study 205 vs. 43% of Study 207), one might posit that it simply takes longer than 5 weeks to observe a treatment effect, and that the greater proportion of patients with longer duration of treatment helped to shift the results of this trial from negative to positive in the exploratory analysis. In the absence of a study with a later prespecified primary endpoint, however, we cannot assume this is the case. As designed, Study 205 only demonstrates a lack of a significant treatment effect at 5 weeks. The Agency views this as a negative study that is not capable of providing supportive evidence of effectiveness.

The Applicant also presents a meta-analysis including results from Studies 202, 205, and 207 as additional supportive evidence of effectiveness. However, as noted in our statistical review (attached), this meta-analysis is exploratory in nature and cannot substitute for substantial evidence of effectiveness as would be provided by separate adequate and well-controlled trials.

Study 206

Study 206 had a different design from the three studies described above. This was a phase 3 multicenter, randomized, placebo-controlled trial to evaluate the safety and efficacy of BUP/SAM 2/2 mg when administered to adults with MDD and who had failed one to two antidepressants. Rather than utilizing SPCD, this study employed a simple placebo lead-in period. Patients with more severe symptoms (HAM-$D_{17} \geq 20$; Group 1) received 4 weeks of placebo before the placebo non-responders were randomized to BUP/SAM 2/2 or continued placebo for 6 weeks; those with less severe symptoms (HAM-D_{17} = 18 or 19; Group 2) received 10 weeks of double-blind treatment. Both groups entered a 2-week safety follow-up period at the end of the double-blind treatment period. The efficacy period was defined as the 6 weeks of randomized, double-blind treatment from randomization baseline (Visit 6 for Group 1) to the

sixth week of randomized treatment period (Visit 12 for Group 1). The primary efficacy endpoint was the change in MADRS-10 score from randomization baseline to the end of the efficacy period. At the end of the 6-week efficacy period, the LS mean change in the MADRS-10 score was -4.6 for placebo and -4.8 for BUP/SAM 2/2 (p=0.78). Both the Applicant and FDA agree the trial was negative.

The Applicant attributes the lack of separation between drug and placebo to a "substantial" placebo response. For reference, the change from baseline to endpoint on the MADRS-10 in the 6-week adjunctive treatment trials for the three atypical antipsychotic drugs approved for this indication ranged from -5.2 to -11.2 in the placebo arms. Across those studies, treatment arms that were statistically superior to placebo demonstrated placebo-subtracted differences on MADRS-10 (change from baseline) in the range of -2.7 to -3.2.

7 Sequential Parallel Comparison Design

The rate and magnitude of placebo response in antidepressant trials has been increasing over time[4]. The SPCD is intended to mitigate the impact of placebo responders on the outcome of clinical trials[5]. Essentially, an SPCD trial is conducted in two stages. In Stage 1, subjects are randomized to receive either drug or placebo, with a greater proportion in the placebo group than in the active treatment arm. After a predetermined period of time, treatment response is evaluated. Those in the placebo group who have nonetheless improved to a prespecified degree are considered placebo responders and continue to receive placebo. The placebo non-responders are randomized to either active treatment or continued placebo in Stage 2. The results of the two stages are combined in a weighted analysis to determine the overall treatment effect and an overall test of hypothesis (p-value).

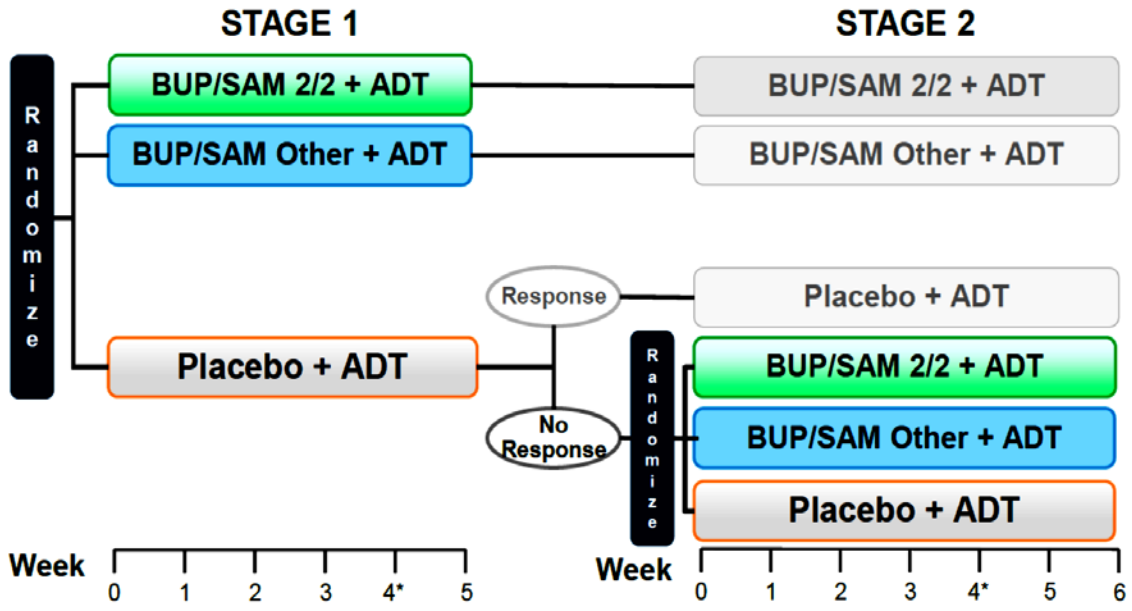
There are a number of unresolved questions regarding the clinical interpretability and the statistical analyses of SPCD data. As a result, the Agency has not endorsed the use of SPCD in a confirmatory trial setting. The key clinical challenges involve unequal durations between the two stages, and interpretation of the weighted treatment effect from the two stages where Stage 1 consists of a typical population and Stage 2 an enriched population. In calculating the weighted treatment effect, the weight assigned to each placebo non-responder is indeed larger than to other patients because placebo non-responders' data from both stages are included in the analysis. The key statistical challenges include whether there is a bias in estimating the weighted treatment effect because the validity of analyses of SPCD data hinges on certain crucial assumptions. The presence of dropouts in the placebo group in Stage 1 adds complexity to clinical interpretability and to the statistical analysis. Dropouts in the placebo group in Stage 1 may represent a different population from the completers in the placebo group and the dropouts (even if they are non-responders) cannot be re-randomized in Stage 2. The impact of the dropouts on Type-I error

[4] Khin NA, YF Chen, Y Yang, P Yang, and TP Laughren, 2011, Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications, J Clin Psychiatry, 72(4) 464-472.
[5] See footnote 1.

control in statistical analysis of SPCD data is also under research. In the context of this application, however, the dropout rates were not substantial; also, it appears that the weights assigned to the different trial stages were sensible. Study 202 fixed the Stage 1 weight at 0.6; Study 207 weighted Stage 1 and Stage 2 equally (0.5). Of note, the results remain consistent when an alternative statistical inference procedure based on bootstrap sampling is used.

The Applicant's briefing document provides an illustration of the application of SPCD in the BUP/SAM development program.

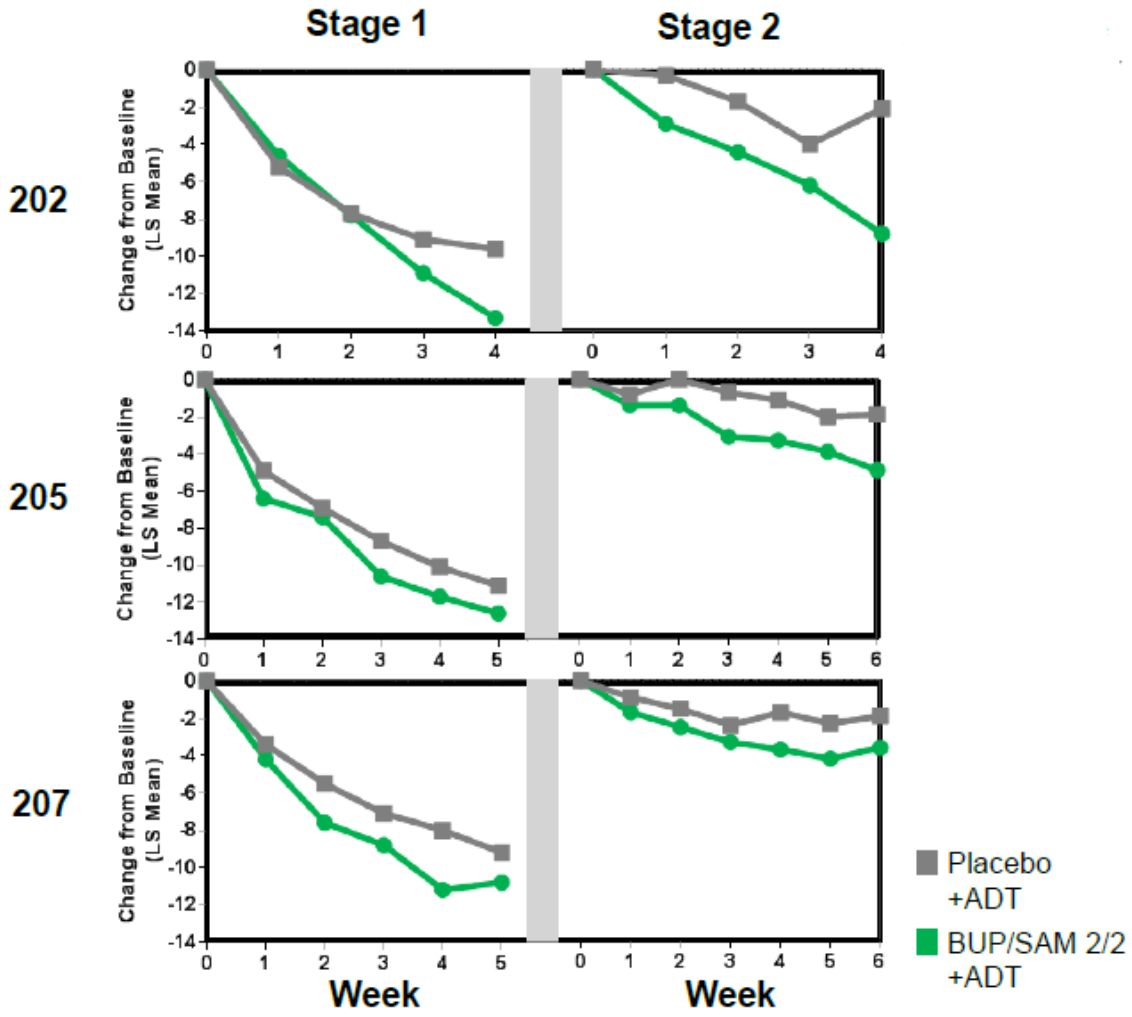
Figure 3: Studies 202, 205, and 207 Design (SPCD)



Abbreviations: ADT=antidepressant therapy; BUP=buprenorphine; SAM=samidorphan.
* In Study 202 the treatment period used for efficacy evaluation was 4 weeks in duration.
Note: Patients randomized to placebo and active treatment (represented by the colored boxes) were included in the efficacy analysis for each stage.
Note: For Study 202, patients who received BUP/SAM in Stage 1 were switched to placebo in Stage 2.
Source: Applicant briefing document, Figure 4, page 18.

This study design is intended to reduce the impact of placebo responders on the trial result because one would expect an amplification of the drug-placebo difference once the placebo responders have been removed from consideration. Thus, if the active treatment does not separate from placebo despite the trial design, there are two possible conclusions: either the placebo response in the trial is unusually large and persistent, or the active treatment is not superior to placebo.

Figure 4: BUP/SAM Reduction vs Placebo in MADRS-10 Scores Across SPCD Studies



Source: Adapted from Applicant briefing document, Figure 16, page 35

Based on the plots above, it appears that there was considerably less improvement in the placebo group in Stage 2 relative to Stage 1. This would suggest that the SPCD design was working as intended. However, the expected amplification of the drug-placebo difference in Stage 2 is not observed. Thus, it is difficult to conclude that an unusually high placebo response rate is to blame for the inability to demonstrate a treatment difference on a traditionally acceptable endpoint in prespecified analysis with appropriate multiplicity controls.

8 Safety

SPCD trials present a unique challenge in terms of how best to analyze and display the safety findings. (Just as it is difficult to determine the size of an overall treatment effect across Stage 1 and 2 of an SPCD trial, it is difficult to determine the effect size for safety issues.) In more

traditionally designed parallel group trials, drug and placebo are administered to approximately equal number of patients beginning at baseline and running through the same duration. In SPCD, one can compare drug- and placebo-treated patients in Stage 1 in the usual manner, despite the over-representation of placebo-treated patients. In Stage 2, the drug group is comprised of patients receiving drug for the first time and those who are continuing treatment from Stage 1, complicating safety analyses.

Prior to the pre-NDA meeting, the Applicant proposed the following pooling strategy for analyzing safety data in the four placebo-controlled trials. The Agency agreed that the pooling strategy was reasonable.

Table 5: Pooling of BUP/SAM MDD Placebo-Controlled Trials for Safety Analyses

Pooling Nomenclature	Description	Stage(s) of Studies to Be Included	Parameters
Stage 1 Pooling	Treatment group comparisons as randomized in Stage 1 (202, 205, 206 [i.e. Group 2], and 207)	Stage 1 (5 weeks)	Numbers enrolled, numbers exposed, demographics, baseline characteristics, descriptive statistics for dose, duration, compliance, study disposition, TEAEs (including subgroup analyses and AESIs[a]), descriptive summary and outlier analyses for VS, laboratory analyses, urinalysis, ECGs, and C-SSRS
Stage 2 Pooling	Treatment group comparisons for placebo non-responders defined in Stage 1 (202, 205, 207) who are re-randomized in Stage 2 and Group 1 (206) placebo run-in non-responders randomized.	Stage 2 (5 or 6 weeks)	
Continuous Treatment Pooling	Treatment group comparisons with subjects who were not re-randomized (or were re-randomized to the same treatment) presented as randomized in Stage 1 (205, 206 (group 2), and 207)	Uses all available safety data from both stages (treatment durations of 10 to 11 weeks PBO controlled	
Post-discontinuation Pooling	Treatment group comparisons for subjects who completed or had an early termination visit, based on last treatment in study (205, 206, and 207)	Any	PDEAEs, AESIs[a] of withdrawal, COWS

Abbreviations: AESI=adverse events of special interest; COWS=Clinical Opioid Withdrawal Scale; ECG=electrocardiogram; TEAE=treatment emergent adverse event; PBO=placebo; PDEAE= post-discontinuation emergent adverse events; VS=vital signs

Source: Applicant's pre-NDA briefing package, Table 16, page 61

Although pooling of safety results in this manner can provide an overarching view of the drug's safety, the question of how to analyze these data and present them in a product label can be challenging. Typically, safety results are presented on both individual trial results and pooled analyses. Pooled analyses include greater numbers of patients that may allow identification of rare or serious events that might not be observed in a single clinical trial. In some cases, however, results are best presented trial-by-trial, particularly if there are studies with disparate

lengths of drug exposure, dissimilar dosage groups, or different randomization ratios. In particular, pooling studies with different randomization ratios and different sample sizes can yield deceptive results, e.g., a Simpson's paradox. Recognizing the limitations and pitfalls of this pooling approach, we will nevertheless focus on the pooled safety analyses for purposes of review by the PDAC and the DSaRM Advisory Committee.

The overall safety database for BUP/SAM includes data from 18 completed studies and one ongoing long-term safety study. These include four completed short-term placebo-controlled studies in subjects with MDD, one completed short-term dose titration study in subjects with MDD, one ongoing long-term (52-week) open-label safety study in subjects with MDD, and 13 completed phase 1 studies (seven studies in healthy volunteers and six studies in special populations, including one study in MDD subjects).

A total of 2165 subjects have been exposed to BUP/SAM across these 19 studies, with 1860 subjects receiving the therapeutic dose of 2/2, 1531 of whom had MDD. To characterize adverse events (AEs) that may be rare or late-occurring, the International Conference on Harmonization (ICH) E1[6] guideline recommends that 300-600 patients be exposed to drug for at least 6 months and 100 patients exposed for at least 1 year. In this development program, 947 subjects received BUP/SAM for at least 6 months and 743 subjects for at least 1 year.

In the pooled analysis of Studies 202, 205, 206, and 207, the most common AEs (≥5% on BUP/SAM 2/2 and at least twice the rate of placebo) were nausea, dizziness, constipation, vomiting, fatigue, somnolence, and sedation.

In addition to the routine safety concerns evaluated in any drug development program, we are also cognizant of the fact that this product is an opioid. Thus, we must evaluate the risk of misuse, abuse, and addiction. The Division of Epidemiology II has reviewed available information about the misuse and abuse of currently marketed buprenorphine products, and their review is attached. The abuse liability assessments for BUP/SAM were reviewed by our Controlled Substances Staff, and their review is attached.

9 Risk Management

The FDA has identified safety concerns associated with the use of BUP/SAM warranting consideration of a Risk Evaluation and Mitigation Strategy (REMS). In the following sections, we discuss the Applicant's proposed REMS, the approved REMS for buprenorphine-containing products approved for pain and opioid dependence, and considerations for risk mitigation for BUP/SAM for the treatment of MDD.

[6] International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), 1994, The extent of population exposure to assess clinical safety for drugs intended for long-term treatment of non-life-threatening conditions E1, accessed October 10, 2018, https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E1/Step4/E1_Guideline.pdf..

9.1 FDA Safety Concerns Associated with BUP/SAM

There are safety concerns that are associated with the class of opioids as well as risks that may be different or of greater concern in the proposed population for BUP/SAM. The risks and the concerns are provided below.

BUP/SAM contains buprenorphine (partial μ-opiate receptor agonist and a partial ƙ-opiate receptor antagonist) and samidorphan (full μ-opiate receptor antagonist that is metabolized into two full μ-opiate receptor agonists). If approved, BUP/SAM will be a chronic medication and the first opioid approved for use in MDD. It is unknown whether chronic exposure to BUP/SAM will contribute to an increased risk of opiate dependence, misuse, and abuse, particularly in patients with a previous history of opioid use disorder. There is a potential concern that buprenorphine can easily be chemically separated from samidorphan by individuals wishing to misuse or abuse the buprenorphine component.

Mental illness, including MDD, is associated with non-medical use of prescription opioids and opioid use disorder. Studies suggest that more than 50% of patients with depression have co-morbid pain conditions.[7] Individuals with MDD who fail to improve with treatment have a higher baseline risk of substance use disorder and polypharmacy. The likelihood that BUP/SAM could be co-administered with other controlled substances such as benzodiazepines is also concerning. The majority of MDD treated in the U.S. is treated by primary care clinicians specializing in family and internal medicine.[8] It is important for prescribers to understand that BUP/SAM is an opioid.

BUP/SAM could be prescribed as an adjunctive treatment to women who are pregnant and are already taking antidepressants. The prevalence of moderate to severe major depressive disorder is 9.3% in women aged 18 to 39.[9] Women who are pregnant and are treated with BUP/SAM may be putting their unborn infants at risk for development of Neonatal Opiate Withdrawal Syndrome (NOWS); this was not studied in the clinical program for BUP/SAM.

The uncharacterized pharmacodynamics of the samidorphan component of BUP/SAM is also a concern. Since samidorphan is a full μ-opiate receptor antagonist that is metabolized into two full μ-opiate receptor agonists, there may be a potential impact of BUP/SAM on patients who require opioid analgesics due to the antagonist effects of buprenorphine and samidorphan combined, especially in emergency situations where patients need analgesia. It is unknown whether patients treated with BUP/SAM will require higher doses to achieve analgesia. In addition, patient nonadherence with antidepressant therapy may lead to variable levels of μ-antagonism and potentially increase the risk of respiratory depression/overdose in the setting of opioid use.

[7] FDA Office of Surveillance and Epidemiology Review (OSE)/Office of Pharmacovigilance and Epidemiology (OPE) Epidemiology Review ALKS 5461 AC Background Document

[8] Mark, TL, KR Levit, and JA Buck, 2009. Psychotropic drug prescriptions by medical specialty, Psychiatr Serv, 60(9):1167

[9] Pratt LA, and DJ Brody, 2014, Depression in the U.S. household population, 2009–2012, NCHS data brief, no 172. Hyattsville, MD: National Center for Health Statistics.

9.2 Risk Evaluation and Mitigation Strategy (REMS) Background

Section 505-1 of the Food, Drug, and Cosmetic Act (FDCA), added to the law by the Food Drug Administration Amendments Act of 2007 (FDAAA) authorizes the FDA to require pharmaceutical manufacturers to develop and comply with a REMS for a drug if FDA determines that a REMS is necessary to ensure that the benefits of the drug outweigh the risks. A REMS is a required risk management plan that uses risk minimization strategies beyond the professional labeling. The elements of a REMS can include: a Medication Guide or patient package insert (PPI), a communication plan to healthcare providers, elements to assure safe use, and an implementation system. FDAAA also requires that all REMS approved for drugs or biologics under New Drug Applications (NDA) and Biologics License Applications (BLA) have a timetable for submission of assessments of the REMS. These assessments are prepared by the sponsor and reviewed by FDA.

A communication plan consists of FDA approved materials used to aid a sponsor's implementation of the REMS and/or inform healthcare providers about serious risk(s) of an approved product. This can include, for example, "Dear Healthcare Professional" letters, collaboration with professional societies, and education pieces (such as letters, drug fact sheets) to inform prescribers of the risks and the safe use practices for the drug.

Elements to assure safe use (ETASU) can include one or more of the following requirements:

- Healthcare providers who prescribe the drug have training or experience or special certifications
- Pharmacies, practitioners, or healthcare settings that dispense the drug are specially certified.
- The drug may be dispensed only in certain healthcare settings
- The drug may be dispensed to patients with evidence of safe-use conditions that each patient must be subject to monitoring
- Patients must be enrolled in a registry

Because ETASU can impose significant burdens on the healthcare system and reduce patient access to treatment, ETASU are required only if FDA determines that the product could be approved only if, or would be withdrawn unless, ETASU are required to mitigate a specific serious risk listed in the labeling. Accordingly, the statute [FDCA 505-1(f)(2)] specifies that ETASU:

- Must be commensurate with specific serious risk(s) listed in the labeling.
- Cannot be unduly burdensome on patient access to the drug.
- To minimize the burden on the healthcare delivery system, must, to the extent practicable, conform with REMS elements for other drugs with similar serious risks and be designed for compatibility with established distribution, procurement, and dispensing systems for drugs.

9.3 Applicant's Risk Minimization Plan

The Applicant proposes a REMS that includes prescriber education through letters, the Prescribing Information, a Healthcare Provider Brochure, as well as an Appropriate Use Checklist. The Applicant also proposes a Medication Guide for patients. Their proposal has the following goal:

The goal of the BUP/SAM REMS is to mitigate the risk of misuse and accidental exposure by:

- Informing healthcare providers and patients of the risks associated with misuse of BUP/SAM.

Informing healthcare providers and patients of the risks associated with accidental exposure of BUP/SAM to children and others for whom it was not prescribed.

9.4 Approved REMS for buprenorphine-containing products approved for pain and opioid dependence

Table 5 provides a list of the approved REMS for buprenorphine-containing products.

Table 6: REMS for Buprenorphine-Containing Products			
REMS Name	**Buprenorphine Indication**	**Risks mitigated in the REMS**	**REMS Summary**
Opioid Analgesics REMS (OA REMS)	Pain	adverse outcomes of addiction, unintentional overdose, and death resulting from inappropriate prescribing, abuse, and misuse	ETASU Healthcare provider training based on FDA Blueprint Includes the Patient Counseling Guide and Medication Guide for patient education
Buprenorphine* Transmucosal Products for Opioid Dependence REMS (BTOD REMS) Suboxone/Subutex REMS	Treatment of opioid dependence	accidental overdose, misuse, and abuse	ETASU Provider and pharmacy education Includes the Medication Guide for patients and the Appropriate Use Checklist as a tool for prescribers
Sublocade REMS*	Depot Injection for treatment of opioid dependence	serious harm or death that could result from intravenous self-administration	ETASU Healthcare setting and pharmacy certification
Probuphine REMS*	Implant for treatment of opioid dependence	complications of migration, protrusion, expulsion and nerve damage associated with the insertion and removal of Probuphine, accidental overdose, misuse and abuse	ETASU Healthcare provider training and certification and pharmacy certification and Medication Guide for patient education
* In order to prescribe or dispense buprenorphine for opioid use disorder, prescribers must apply for a Drug Addiction Treatment Act of 2000 (DATA 2000) waiver through the Substance Abuse and Mental Health Services Administration (SAMHSA), which includes complete eight hours of required training.			

The Sublocade REMS and Propuphine REMS are designed to mitigate risks specific to those formulations. The Opioid Analgesics REMS and BTOD REMS are required to mitigate the known opioid-related risks of addiction, accidental overdose, misuse and abuse. These known opioid-related risks are relevant to BUP/SAM.

9.5 Risk Mitigation Considerations for BUP/SAM

A variety of strategies are used to minimize risks associated with drugs and therapeutic biologics. These strategies minimize risks in several ways. They can communicate specific risk information, as well as information about the safe use of the product. In addition, they can provide guidance and encourage, remind, or support adherence to certain prescribing, dispensing, or monitoring requirements, and/or limit use of a product to only the most appropriate patients where the benefits outweigh the risks. The FDA is still evaluating how best to mitigate the risks associated with BUP/SAM to help ensure the its safe use.

All of the buprenorphine-containing products are approved with a REMS. Both the Suboxone/Subutex and the Buprenorphine Transmucosal Products for Opioid Dependence (BTOD) REMS are required to mitigate the adverse outcomes of accidental overdose, misuse, and abuse. The REMS programs for these products used to treat opioid dependence include provider and pharmacy educational materials and an appropriate use checklist. In addition, there is a required training for these prescribers through the Substance Abuse and Mental Health Services Administration (SAMHSA) which manages the Drug Addiction Treatment Act of 2000 (DATA 2000). Physicians who apply and hold waivers can prescribe and/or dispense buprenorphine products for medication-assisted treatment. SAMHSA sets eligibility and certification requirements for the DATA 2000 waiver which includes mandatory training.

The buprenorphine-containing products approved for the treatment of pain are part of the Opioid Analgesic REMS, which requires manufacturers to make training available to healthcare providers involved in the management of patients with pain. The training is focused in part on pain management, but also covers identifying risk factors for abuse and addiction as well as how to counsel patients and their families on the safe use of opioids and fundamentals on addiction medicine.

The Agency is concerned that BUP/SAM carries the same risks of accidental overdose, misuse, and abuse seen with other opioid products. In addition, we are evaluating whether there are other safety concerns that may be unique in the indicated population, such as co-prescribing with other controlled substances, neonatal exposure, as well as the need for higher opioid doses to achieve analgesia in emergencies. Overall, these risks are not well characterized in the clinical development program.

We are seeking the advice of the PDAC and the DSaRM Advisory Committee on how best to mitigate the risks associated with BUP/SAM and the necessary components of a REMS.

10 References

Fava M, AE Evins, DJ Dorer, and DA Schoenfeld, 2003, The problem of the placebo response in clinical trials for psychiatric disorders: Culprits, possible remedies, and a novel study design approach. *Psychother Psychosom*, 72(3):115-127.

FDA Office of Surveillance and Epidemiology Review (OSE)/Office of Pharmacovigilance and Epidemiology (OPE) Epidemiology Review ALKS 5461 AC Background Document

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), 1994, The extent of population exposure to assess clinical safety for drugs intended for long-term treatment of non-life-threatening conditions E1, accessed October 10, 2018, https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E1/Step4/E1_Guideline.pdf.

Khin NA, YF Chen, Y Yang, P Yang, and TP Laughren, 2011, Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications, *J Clin Psychiatry*, 72(4) 464-472.

Mark, TL, KR Levit, and JA Buck, 2009. Psychotropic drug prescriptions by medical specialty, *Psychiatr Serv*, 60(9):1167

Pratt LA, and DJ Brody, 2014, Depression in the U.S. household population, 2009–2012, NCHS data brief, no 172. Hyattsville, MD: National Center for Health Statistics.

Trivedi MH, AJ Rush, SR Wisniewski, AA Nierenberg, D Warden, L Ritz, G Norquist, RH Howland, B Lebowitz, PJ McGrath, K Shores-Wilson, MM Biggs, GK Balasubramani, M Fava and STAR*D Study Team, 2006, Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: Implications for clinical practice, *Am J Psychiatry*, 163(1): 28-40.

World Health Organization. Fact Sheets: Depression, accessed October 9, 2018, http://www.who.int/en/news-room/fact-sheets/detail/depression.

11 Attachments

Statistical review
CSS review
Epidemiology review

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	210417
Drug Name:	buprenorphine and samidorphan (BUP/SAM)
Indication:	Adjunctive treatment for major depressive disorder (MDD)
Applicant:	Alkermes, Inc.
Date(s):	Submission Date: 1/31/2018 PDUFA Date: 1/31/2019
Review Priority:	Standard
Biometrics Division:	Division of Biometrics I
Statistical Reviewer:	Semhar Ogbagaber, Ph.D.
Concurring Reviewers:	Peiling Yang, Ph.D., HM James Hung, Ph.D.
Medical Division:	Division of Psychiatry Products
Clinical Team:	Daniel J Lee, M.D., Javier Muniz, M.D.
Project Manager:	Nam Chun, Pharm.D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

This NDA included 4 studies (Studies 202, 205, 206 and 207), which investigated efficacy of BUP/SAM (buprenorphine and samidorphan) for adjunctive treatment of major depressive disorder (MDD). BUP/SAM is submitted as an NME (new molecular entity) which combines equal proportions of buprenorphine and samidorphan. For notational simplicity, the drug combination, 2 mg buprenorphine and 2 mg samidorphan daily, is represented as 2/2. Other possible dose combinations are defined similarly. This combination product was evaluated in patients who exhibit residual symptoms after being treated with a standard anti-depressant medication for MDD. In these trials, patients continued their primary anti-depressant medication and received either ALKS 5461 or placebo.

Based on data from the two studies (Studies 202 and 207), the Applicant has sought a labeling claim to describe BUP/SAM's effect on MDD compared to placebo. This submission is unique as it features utilization of sequential parallel comparison design (SPCD). SPCD is a special two-stage design constructed to address high placebo response in psychiatry trials. The design allows filtering placebo responders out from the first stage in order to demonstrate the effect of a drug under investigation. SPCD combines data from the two stages: all incoming subjects in the first stage and placebo non-responders in the second stage.

Among the 4 trials, one is a phase 2 SPCD trial (202) and three are phase 3 trials (205, 206, 207). Studies 205 and 207 used SPCD and Study 206 used the conventional randomized, double-blind with placebo lead-in component in design. Currently, there is one on-going phase 3 trial (217). Studies 205 and 206 were negative trials. This NDA review will focus on studies 202 and 207, which were presented as positive according to the Applicant. However, at the Type B pre-IND meeting (February 17, 2011), the Applicant clarified that study 202 was conducted for proof of concept. This statistical review was conducted with this background information in mind.

Study 202 was only a proof of concept trial and was not intended to support efficacy. Meeting minutes from February 25, 2011 pre-IND attests the clear intention of the Applicant. The ensued discussion clearly stated, "*The applicant also clarified that study ALKS33BUP-202 is not intended to be a "proof of efficacy" study as stated in their meeting package. This will instead be a proof of concept study to examine the validity of their hypothesis in a preliminary fashion*."

Study 202 was an exploratory trial but the Applicant claimed to have shown evidence of efficacy for the lower dose of BUP/SAM without accounting for multiplicity adjustment. Two doses of BUP/SAM (2/2, 8/8) were investigated in this study. The primary efficacy endpoint was the change from baseline to week 4 on the 17-item Hamilton Rating Scale for Depression (HAM-D) total score.

Study 207 was an SPCD, phase 3 trial which investigated 2 doses of BUP/SAM (1/1 and 2/2) versus placebo. The 3 primary efficacy endpoints, which were amended at a late stage of the trial, were change from baseline for MADRS-6_{AVG} (averaged MADRS-6 score of change from baseline to week 3 up to end of each stage), MADRS-10_{AVG} (averaged MADRS-10

score of change from baseline to week 3 up to end of each stage) and MADRS-10_{EOT} (change from baseline to end of each stage).

Protocol Amendment #3 of study 207, dated September 15, 2016, introduced changes to the initially planned primary endpoint and its analyses. The original primary efficacy endpoint from study 207 protocol submitted in 2014 was the change in MADRS total score from baseline to week 5 in each stage (despite 6 weeks treatment duration in stage 2). The Protocol Amendment states that the primary efficacy endpoints will be:

[1] Change in MADRS-6 using the average of changes from baseline to Week 3 through the end of each stage;

[2] Change in MADRS-10 total score using average of changes from baseline to Week 3 through the end of each stage;

[3] Change in MADRS-10 total score from baseline to end of each stage.

In the original protocol, the efficacy endpoint was defined at week 5 in each stage despite 6 weeks treatment duration in stage 2. But, in Protocol Amendment #3, the efficacy endpoint was defined as the end of each stage, resulting in different durations for efficacy comparison between the two stages. The purpose for treatment period of 6 weeks in Stage 2 was to mask both investigators and subjects to the timing of the primary efficacy endpoint, which would occur at Week 5.

After evaluation of Applicant's primary endpoint, the Clinical Outcomes Assessment (COA) team at FDA determined that they did not agree with the use of the MADRS-6 in Study 207 as the primary efficacy endpoint due to "the omission of clinically important symptoms that are included in the MADRS-10" (Meeting Minutes, 07/24/2017, DARRTS Ref ID: 4136401; COA Consult Review, 06/06/2017).

2 INTRODUCTION

2.1 Overview

Studies 202 and 207 were submitted in support of a new NDA 210417, which evaluates the effect of BUP/SAM for treatment of adjunctive major depressive disorder (MDD). Study 207 was international in scope (US, Canada, Germany) while study 202 was conducted in the US.

SPCD was first proposed by Fava et al (2003)[1]. The rationale behind the use of SPCD is to reduce excessive placebo response in an attempt to reduce the number of subjects required

[1] Fava M, Evins A, E, Dorer D, J, Schoenfeld D, A, The Problem of the Placebo Response in Clinical Trials for Psychiatric Disorders: Culprits, Possible Remedies, and a Novel Study Design Approach. *Psychother Psychosom* 2003;72:115-127.

without compromising the statistical power for efficacy demonstration. Psychiatric trials are highly linked with rising placebo response and declining trend of treatment effect over years. This has led to a substantial number of negative trials in MDD because of small treatment differences between drug and placebo, in which case it is often difficult to demonstrate a statistical significance without a large enough sample size. A phase 2 study 202 and phase 3 study 207 by Alkermes utilized two-stage SPCD to evaluate efficacy of BUP/SAM. Two other studies (205, 206) that also investigated BUP/SAM were negative.

The Applicant used MADRS-6_{AVG} (averaged MADRS-6 score of change from baseline to week 3 up to the end of each stage (week 5 for stage 1; week 6 for stage 2) as a primary efficacy endpoint for adjunctive treatment of MDD in only one study, 207, that showed statistically significant separation between the 2/2 dose group and placebo. Regardless of analysis results, MADRS-6 was found to be unacceptable by the clinical team. The Applicant also submitted exploratory study 202, where there were two dose groups to compare with placebo. Although the nominal p-value from the comparison between placebo and dose group 2/2 on the HAM-D_{17} primary efficacy measure was less than 0.05, the overall type I error rate may be inflated because a multiple testing procedure was not pre-specified. Emphasis in this review is put on the results of studies 202 and 207.

Upon FDA's request, the Applicant sent a comprehensive evaluation report which compared the psychometric assessments of MADRS-6 and MADRS-10 on March 28, 2017. However, the Clinical Outcomes Assessment team deemed MADRS-6 as unacceptable and concluded:

- *MADRS-6 may not be sensitive in distinguishing between severe and less severe MDD patients owing to the removal of symptoms that are often the most disabling and the most important in clinical decision making regarding patients with MDD: "Reduced Sleep," "Reduced Appetite," "Concentration Difficulties," and "Suicidal Thoughts." These are the symptoms that a clinician often will use to make decisions such as whether to treat with medication or with verbal or psychosocial therapies, and whether to treat on an outpatient or inpatient basis. All four symptoms are included in the DSM-5 diagnostic criteria of MDD, and are understood by clinicians as clinically important characteristics of the illness.*

- *Results of factor analysis on MADRS-10 suggested that the four items removed were highly associated but not redundant with the six items retained. This finding suggested the four items provide information that is not captured by the MADRS-6.*

2.2 Data Sources

The Applicant's submitted data, SAS program listings for the two studies and clinical study reports are available in the following directories of the CDER' electronic document room (EDR):
Study 202: \\Cdsesub1\evsprod\NDA210417\0014\m5\datasets\alk5461-202\analysis\adam\datasets, \\Cdsesub1\evsprod\NDA210417\0014\m5\53-clin-stud-rep\533-rep-human-pk-stud\5332-patient-pk-init-tol-stud-rep\alk5461-202
Study 207: \\Cdsesub1\evsprod\NDA210417\0014\m5\datasets\alk5461-207\analysis\adam\datasets, \\Cdsesub1\evsprod\NDA210417\0014\m5\53-clin-stud-rep\535-rep-effic-safety-stud\adjunct-treatment-mdd\5351-stud-rep-contr\alk5461-207

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The reviewer found the quality and integrity of the submitted data satisfactory and acceptable for the review analysis.

3.2 Evaluation of Efficacy

The primary objective of these studies was to provide evidence of efficacy of ALKS-5461 as an adjunctive treatment for MDD in adult patients whose response to SSRI or SNRI was not adequate.

3.2.1 Study Design and Endpoints

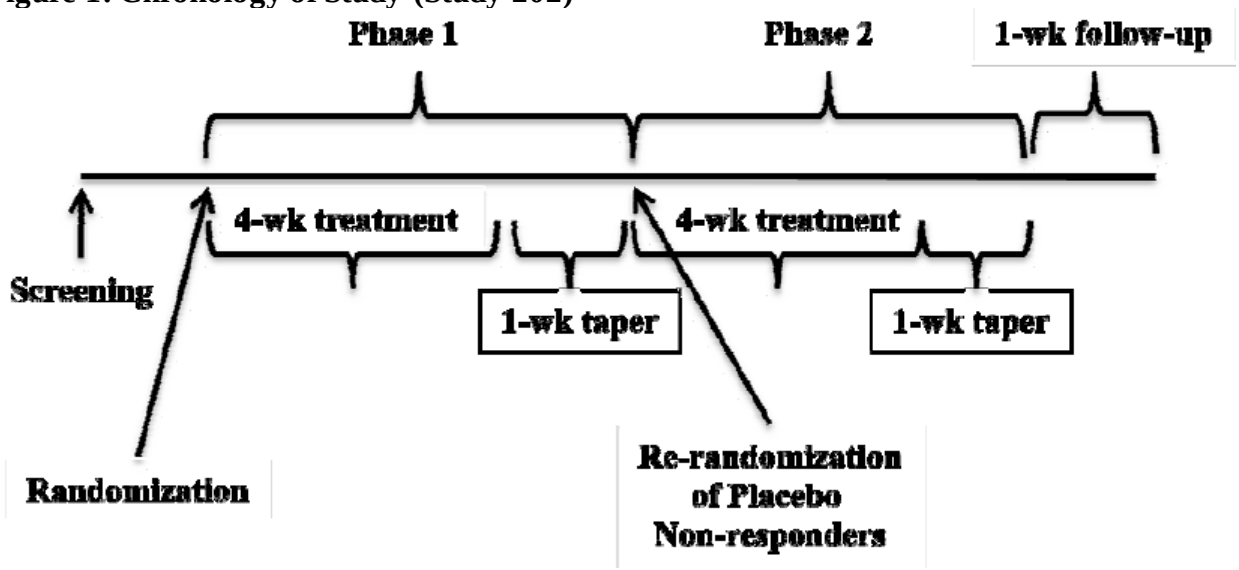
3.2.1.1 Study 202

Study Design
This was a Phase 2, multicenter, randomized, double-blind, placebo controlled study, which utilized the sequential parallel comparison design (SPCD) to evaluate ALKS-5461. ALKS-5461 is an equal combination of buprenorphine (BUP) and samidorphan. It was investigated in subjects with MDD who have had an inadequate response to one to two adequate courses of treatment with a SSRI (selective serotonin reuptake inhibitor) or a SNRI (serotonin-norepinephrine reuptake inhibitor). Two doses, 2 mg/2 mg and 8 mg/8 mg, in equal proportions of buprenorphine and samidorphan, were studied in two treatment phases of SPCD. Each SPCD phase lasted 5 weeks, which comprised a 4-week treatment period followed by a 1-week taper period (Figure 1).

In phase 1 at Visit 2, subjects were randomized to receive ALKS-5461 2/2, ALKS-5461 8/8, or placebo in a 2:2:9 ratio. At the beginning of stage 2 (Visit 7), placebo non-responders from phase 1 were re-randomized to receive either ALKS-5461 2/2, ALKS-5461 8/8, or placebo in a 1:1:1 ratio. Subjects were categorized as placebo non-responders at the start of phase 2

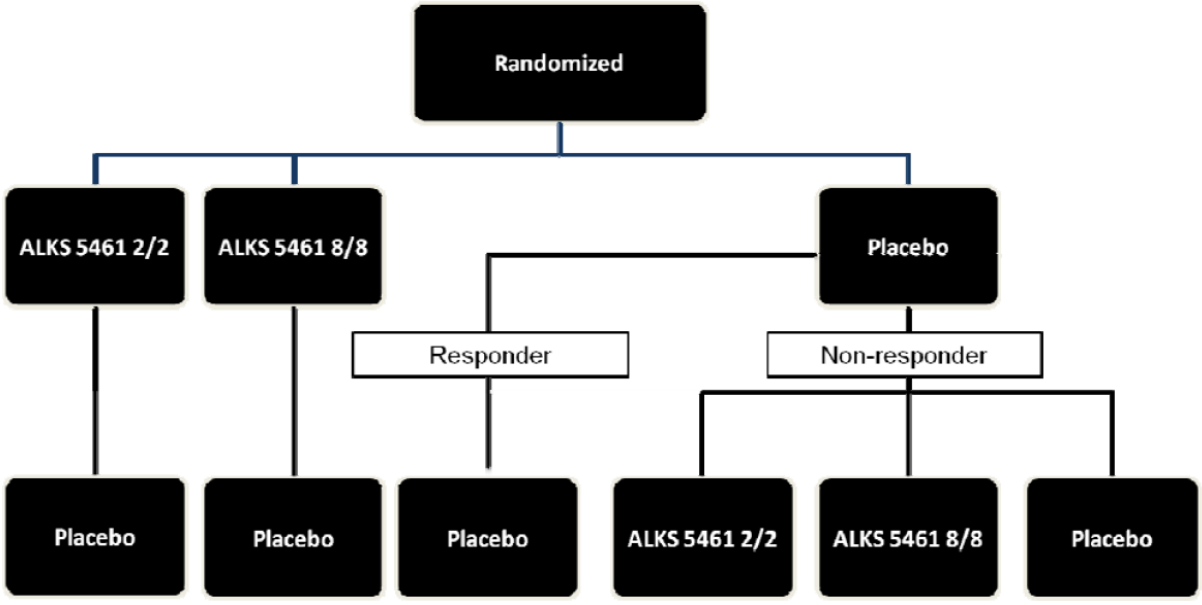
treatment phase. A placebo non-responder was defined as a subject who met the criteria: *a HAM-D_{17} total score > 14 at phase 1 baseline and a <50% reduction from baseline to end of week 4 in the HAM-D_{17} total score.*

Figure 1: Chronology of Study (Study 202)



Source: Figure 1 of Applicant's Clinical Study Report (Page 26)

Figure 2: Overall Study Schema (Study 202)



Source: Figure 2 of Applicant's Clinical Study Report (Page 28)

After the original protocol was issued on 3 November 2011, there were four amendments, issued on 27 March 2012, 12 June 2012, 28 September 2012 and 20 November 2012, respectively.

Reviewer's Note: In a pre-IND meeting with the Applicant on February 17, 2011, the Applicant clarified that SPCD methodology in study 202 would be utilized as purely a "proof of concept" study and not "proof of efficacy" for confirmatory purposes (02/25/2011, DARRTS Ref ID: 2910466). The Statistical Analysis Plan (SAP) was submitted on March 21, 2013, which was approximately near the date of database unblinding on March 28, 2013.

3.2.1.2 Study 207

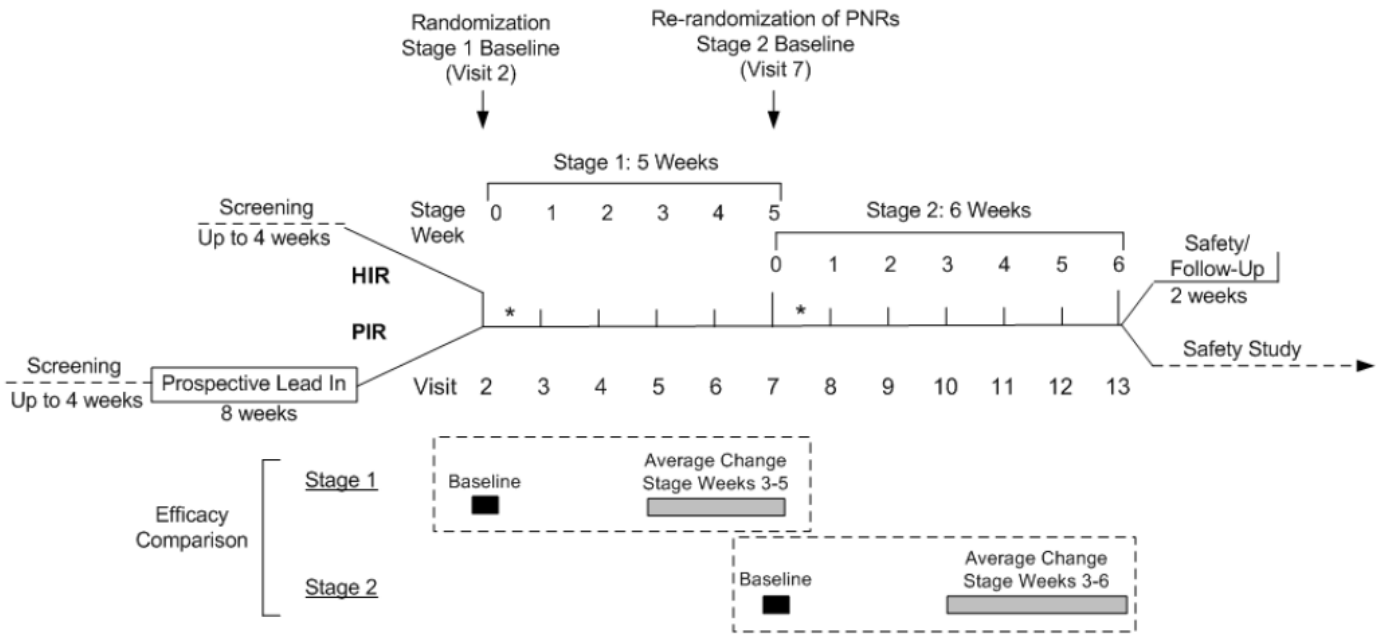
Study Design

This Phase 3 multicenter, randomized, double-blind study evaluated the efficacy and safety of once-daily adjunctive sublingual ALKS 5461 treatment (ALKS 5461 1/1; and ALKS 5461 2/2) in male and female subjects (18-70 years, inclusive) with MDD. Eligible subjects must have shown no improvement or inadequate response to anti-depressant therapy (ADT) during their current major depressive episode.

Study 207 included up to 4-week screening period, an 8-week, prospective lead-in period (to evaluate eligible participants whether they qualify as inadequate ADT responders), an 11-week (stage 1: 5 weeks, stage 2: 6 weeks), double-blind treatment period, and a 2-week safety follow-up period. All potentially eligible screened subjects were classified as either historical inadequate responders (HIR) or as prospective inadequate responders (PIR) based on their Visit 1 17-item Hamilton Rating Scale for Depression (HAM-D) total score; number of prior inadequate ADT responses, and current ADT (if any) response. Subjects who are identified as PIR are evaluated at the end of the 8-week period; those who qualify as inadequate responders proceed to Visit 2 in stage 1. Subjects could opt to enter a 52-week extension trial (ALK5461-208) at the conclusion of the double-blind treatment period.

At Visit 7, subjects in the placebo group were categorized as placebo responders or placebo non-responders according to the non-response criteria: *MADRS score > 15 at Visit 7 and < 50% reduction in MADRS score from Visit 2 to Visit 7*.

Figure 3: Overall Study Schema (Study 207)



Abbreviations: HIR=historical inadequate responder; PIR=prospective inadequate responder; PNR=placebo non-responder
*=3-day dose titration for ALKS 5461 1/1 subjects; 1-week dose titration for ALKS 5461 2/2 subjects

Source: Figure 6.a of Applicant's Clinical Study Report (Page 24)

There were three amendments after the original protocol was issued on 13 January 2014. The protocol was amended on 1 April 2014, 13 November 2014, 15 September 2016, and respectively.

Reviewer's Note: In a Type C meeting with the Applicant on 02/13/2017, FDA did not endorse MADRS-6 as a primary endpoint as well as the strategy of comparing dose groups based on averaged MADRS-6 or averaged MADRS-10 over several weeks (03/15/2017, DARRTS Ref ID: 4107096). The COA team at the FDA communicated that it did not agree with MADRS-6 as a primary efficacy endpoint because it omits important symptoms contained in MADRS-10 (SDN 61, Review Completion Date: May 18, 2017, DARRTS Ref ID: 4107096). The Applicant specified to compare the change from baseline in MADRS-6 or MADRS-10 score averaged from Week 3 to the end of each stage (week 5 for stage 1 and week 6 for stage 2).

3.2.2 Statistical Methodologies

Before describing the pre-specified statistical methodologies, it is worthwhile to introduce the special design called SPCD.

Sequential Parallel Comparison Design (SPCD)

Unlike the conventional parallel group design which has one treatment period, in SPCD evidence is based on efficacy data from two treatment stages. The SPCD methodology entails merging of analyses from the two stages and inference is made based on a linear weighted test statistic from stage 1 and stage 2. The weighted statistic is $\frac{w\widehat{\theta}^{(1)} + (1-w)\widehat{\theta}^{(2)}}{\sqrt{w^2 \operatorname{var}\left(\widehat{\theta}^{(1)}\right) + (1-w)^2 \operatorname{var}\left(\widehat{\theta}^{(2)}\right)}}$, where $\widehat{\theta}^{(1)}$ and $\widehat{\theta}^{(2)}$ correspond to point estimates based on ANCOVA (Analysis of Covariance) or MMRM (Mixed Model with Repeated Measures) analysis, w is a pre-specified weight for stage 1 data. The weighted test statistic, which combines estimated treatment effects from stage 1 and stage 2, has been considered for purposes of inference. It is used to test the null hypothesis, $H_O : \theta^{(1)} = \theta^{(2)} = 0$. The p-value is computed based on the weighted test statistic and the normality assumption under the null hypothesis.

Chen et al. (2011)[2] proposed re-randomization of stage 1 placebo non-responders at the beginning of stage 2. In this paper, the point estimates in the weighted test statistic are based on ordinary least squares (OLS) estimators. The validity of the test statistic could be affected by the weight term as well as missing data. Therefore, clinically meaningful and justifiable weights should be pre-specified in the SAP.

[2] Chen YF, Yang Y, Hung HJ, Wang SJ. Evaluation of performance of some enrichment designs dealing with high placebo response in psychiatric clinical trials. *Contemporary Clinical Trials* 2011; **32**:592–604.

Crucial assumptions stated in this paper (page 4) to proving zero covariance between stage 1 and stage 2 treatment effects, $\widehat{\theta}^{(1)}$ and $\widehat{\theta}^{(2)}$, are:

a. Constant correlation between the continuous endpoint measures in stage 1 and stage 2 for placebo non-responders entering stage 2;
b. Constant variance of the continuous endpoint measures for all patients in each stage.

Whether these assumptions hold can only be explored after data has been collected and analyzed. A key assumption is that the treatment arms have equal variances in both stages.

Adding to the complexity of inference in SPCD is how missing data impacts the weighted test statistic. There has been an on-going extensive research in academia, industry and regulatory agencies regarding SPCD in the presence of missing outcomes, so this area is still evolving. As noted in the statistical review of Study 207 protocol (February 26, 2014), FDA conveyed its concerns with regard to type I error and missing data:

1. *Without theoretical proof, it is not guaranteed that the type I error rate will be controlled, especially in scenarios where there are extreme dropouts.*
2. *There is no analytical proof for validity of associated statistical analysis when there are missing data.*

The Applicant's estimation and inference approach for studies 202 and 207 are based on this paper.

Throughout the NDA review, we use the terms 'stage' and 'phase' interchangeably as prefix to describe the two treatment stages in SPCD.

3.2.2.1 Study 202

The primary analysis for the primary efficacy endpoint was carried out on the full analysis sets (FAS) from Phase 1 and Phase 2.

Phase 1 FAS: consists of all randomized subjects who received at least 1 dose of study drug (ALKS-5461 or placebo) and had at least 1 primary efficacy assessment (i.e., HAM-D_{17}) after administration of study drug in Phase 1.

Phase 2 FAS: consists of all placebo non-responders from Phase 1 who were re-randomized and received at least 1 dose of study drug in Phase 2 and had at least 1 primary efficacy assessment (i.e., HAM-D_{17}) after administration of study drug in Phase 2.

Efficacy Endpoints and Analyses Methods (Primary and Key Secondary Efficacy)

The primary efficacy comparisons were the mean change from baseline to Week 4 in HAM-D_{17} total score between each of the two doses of ALKS-5461 and placebo. Lower score of HAM-D_{17} is associated with improved depression. The HAM-D_{17} total score ranges from 0 to 52.

There were no key secondary endpoints but the reviewer analyzed the Applicant's secondary endpoint: change from baseline in MADRS-10 total score (4-week treatment period in each phase). The MADRS-10 endpoint was one of the primary endpoints specified in study 207.

The analysis of mixed model repeated measure (MMRM) with AR(1) covariance structure was used to evaluate the change from baseline to week 4 (in each phase) in HAM-D_{17} total score. The model included terms: treatment, visit, treatment-by-visit interaction, and baseline HAM-D_{17} total score. The primary efficacy analyses from the two phases were combined using SPCD methodology to make inferences about the trial. The pre-specified weights for phase 1 and phase 2 were 0.6 and 0.4, respectively.

Multiple Testing Procedure

Due to the exploratory nature of Study 202, no statistical adjustments were planned to test multiple hypotheses (SAP, Version 1.0, March 21, 2013, Page 18). However, on Page 20 of the SAP, the Applicant states that "*The study will be considered positive if at least one dose group of ALKS 5461 is statistically superior to placebo. No multiplicity adjustment of 2 hypothesis tests will be made due to the exploratory nature of this study.*" (SAP, Version 1.0, March 21, 2013, Page 20). This statement is consistent with the agreement during face to face Type B meeting (February 17, 2011), in which the Applicant intended this study to be a "pure proof of concept study".

Reviewer's Note: Following an initial refuse-to-file (RTF) letter issued to Alkermes on March 30, 2018, there were two informal tele-conferences between FDA and the Applicant. Since the RTF contained some factual errors, it was later rescinded by the division of psychiatry products (DPP) and review of the Applicant's NDA submission resumed immediately afterwards. For the meeting, the Applicant explored data from Study 202 using a post-hoc Bonferroni procedure to adjust for multiple testing. This attempt could be construed as "significant p-value" fishing expedition. Cherry-picking convenient but unspecified multiplicity adjustment methods after the trial results are known to often make the statistical interpretability of trial results highly questionable. Furthermore, we noted that for the phase 3 studies included in this NDA, where multiple dose levels were investigated, the Applicant pre-specified a fixed sequence testing procedure, not Bonferroni procedure, to control the overall type I error rate. Applying such a fixed sequence testing procedure starting from the high dose 8/8 versus placebo would have rendered Study 202 to be an inconclusive study; that is, the trial would have failed on the 8/8 dose by this fixed sequence testing procedure and no further testing could be performed.

Sample Size Calculation

Assume the mean (standard deviation) of change from baseline in HAM-D_{17} for the placebo and the investigational drug arm in stage 1: -7 (7) and -10.5 (7.5); stage 2: -3 (5.3) and -7 (6.5), respectively. Based on simulations, a total of 130 subjects in Stage 1 (in 9:2:2 ratio, where 9 represents placebo) would be sufficient to provide at least 85% power based on a 5% (two-sided) significance level on a global test that includes two phases of SPCD.

Sensitivity Analyses

To evaluate robustness of the main results from MMRM analysis, the Applicant conducted the following sensitivity analyses:

1. An ANCOVA (LOCF, OC) model which included treatment group and baseline HAM-D_{17} total score.
2. Same as primary analysis using MMRM via weights 0.7 and 0.3 for Stage 1 and Stage 2, respectively.

3.2.2.2 Study 207

The SPCD was also used in Study 207. The primary efficacy analysis was based on two FAS populations, namely, stage 1 FAS (FAS1) and stage 2 FAS (FAS2).

Stage 1 FAS: subjects in stage 1 safety population* who had at least 1 post-baseline assessment of MADRS-10 score in stage 1.

Stage 2 FAS: subjects in stage 2 safety population** who had at least 1 post-baseline assessment of MADRS-10 score in stage 2.

*Stage 1 Safety Population: all randomized subjects who received at least one dose of study drug (double blind placebo or ALKS-5461) during Stage 1.

**Stage 2 Safety Population: stage 1 placebo non-responders entering stage 2 and who received at least one dose of study drug (double blind placebo or ALKS-5461) during stage 2.

Efficacy Endpoints and Analyses Methods (Primary and Key Secondary Efficacy)

The primary efficacy endpoints are

1. change in MADRS-6 using average of changes from baseline to Week 3 through the end of each stage (Week 5 for stage 1, Week 6 for stage 2).

2. change in MADRS-10 total score using average of changes from baseline to Week 3 through the end of each stage.

3. change in MADRS-10 total score from baseline to end of each stage.

MADRS-10: is a 10-item instrument and total score ranges from 0 to 60. Higher MADRS score indicates more severe depression.

MADRS-6: is a 6-item subscale of the 10-item MADRS and total score ranges from 0 to 36. The 6 items are: 1=sadness, 2=apparent sadness, 3=inner tension, 7=lassitude, 8=inability to feel, 9=pessimistic thoughts.

No key secondary endpoints were specified.

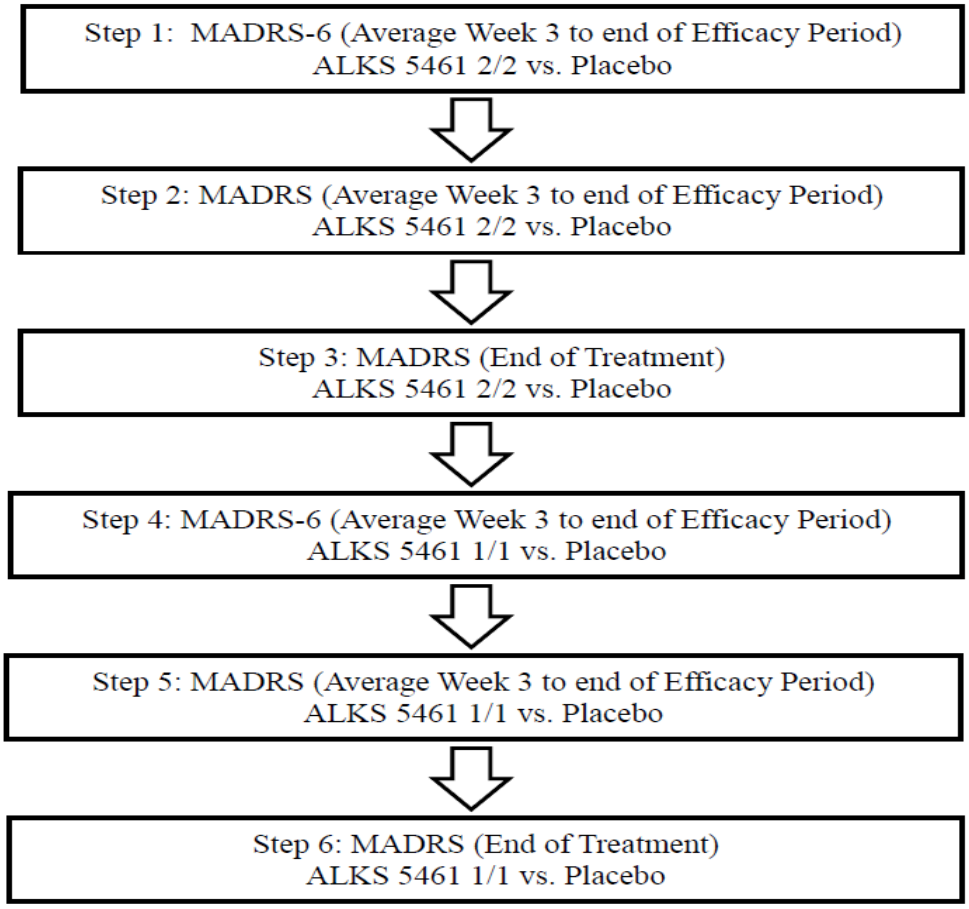
The MMRM analysis with unstructured (UN) covariance matrix (to model within subject variability) was used to evaluate the change from baseline in the primary efficacy outcomes in each stage. The Kenward-Roger approximation was used to adjust the denominator degrees of freedom. The model included terms: treatment group, visit, treatment-by-visit interaction, baseline, and baseline-by-visit interaction.

Efficacy conclusions (comparing ALKS 5461 1/1 vs. Placebo; ALKS 5461 2/2 vs. Placebo) were based on p-value, which was computed based on the weighted test statistic and the normality assumption under the null hypothesis. Treatment effects from stage 1 and stage 2 were combined by allocating equal weights of 0.5 to each stage.

Multiple Testing Procedure

A fixed sequence approach was used to adjust for multiple comparisons. Tests at each pre-specified order were evaluated at 0.05 significance level. Subsequent testing stopped if one of the hypotheses was found not to be significant. Figure 4 displays the testing order of the endpoints.

Figure 4: Fixed Sequence Approach for Multiple Comparisons (Study 207)



Source: Applicant's SAP (page 19)

Sample Size Calculation

Sample size was computed using simulations based on estimates from previous ALKS 5461 trials. Assumptions were made on repeated measures quantities such as mean at each visit, variance, correlations, etc. Using a fixed stage 1 randomization ratio (2:2:9) to ALKS 5461 1/1, ALKS 5461 2/2 and placebo, assuming the baseline mean (standard deviation) of MADRS total score equal to 30 (9.3), unstructured correlation matrix for within subject variability (values ranging from 0.1 to 0.96 depending on visit pair and treatment groups), stage 1/stage 2 discontinuation rates of 10%/5% and 20%/10% for the placebo and the investigational drug arms, a total of 350 subjects would provide approximately 90% power to detect a treatment difference of 3 points in stage 1 and 5 points in stage 2 (between the investigational drug and placebo in MADRS-10 change from baseline at Week 5) using MMRM least-squares means, Wald test derived from analysis and a 2-sided hypothesis test with a significance level at 0.05.

Sensitivity Analyses

1. Delta-adjusted pattern-mixture model (PMM), which incorporates the clinical assumption that subjects who discontinue at a given time point would have, on average, their unobserved efficacy score worse by some amount, δ, compared with the observed efficacy score of subjects on the same treatment arm who continue to the next time point.

2. ANCOVA model using LOCF. In each stage, the ANCOVA model included: change from baseline to the average of assessment from week 3 to end of each stage as a dependent variable; treatment group factor and baseline value as a covariate.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

3.2.3.1 Study 202

There were 31 sites within the United States. First subject visit was on December 29, 2011 and last visit on March 14, 2013.

As displayed in Table 1, one hundred and forty-two (142) subjects (62.9%) were randomized (99 in placebo, 22 in ALKS-5461 2/2 and 21 in in ALKS-5461 8/8).

Phase 1 FAS population (n = 135; placebo (95), ALKS 5461 2/2 (20), ALKS 5461 8/8 (20)) was the primary efficacy analysis set in stage 1 treatment period.

Phase 2 FAS population (n = 65; placebo (20), ALKS 5461 2/2 (23), ALKS 5461 8/8 (22)) was the primary efficacy analysis set in stage 2 treatment period.

Table 1 shows that 121 (85.82%) subjects completed phase 1 (90 (91.84%) in placebo; 18 (75%) in ALKS 5461 2/2; 13 (68.42%) in ALKS 5461 8/8). The most common reasons for study discontinuation were adverse event (7.09%), lost to follow-up (3.55%), and withdrawal by subject (2.13%). Completion rate in phase 2 is summarized in Table 2 by treatment sequence to

which they were randomized sequentially: 89.83% subjects completed phase 2 in which 100% in Placebo, 78.26% in ALKS 5461 2/2, 81.82% in ALKS 5461 8/8.

Table 1: Summary of Analysis Population and Disposition in Phase 1 (Study 202; All Randomized Subjects)

		Phase 1 Treatment Group [1]		
Category	All	Placebo	ALKS 5461 2/2	ALKS 5461 8/8
Subjects Randomized, N	142	99	22	21
Subjects Randomized but not Dosed, N (%) [2]	1 (0.70)	1 (1.01)	0	0
Subjects in the Safety Population, N	141	98	24	19
Subjects in the FAS Population, N (%) [2]	135 (95.07)	95 (95.96)	20 (90.91)	20 (95.24)
Subjects in the PK Population, N (%) [3]	34 (24.11)	0	19 (79.17)	15 (78.95)
Subjects who Discontinued during Phase 1, N (%) [3]	20 (14.18)	8 (8.16)	6 (25.00)	6 (31.58)
Adverse Event	10 (7.09)	1 (1.02)	4 (16.67)	5 (26.32)
Lost to Follow-up	5 (3.55)	2 (2.04)	2 (8.33)	1 (5.26)
Withdrawal by Subject	3 (2.13)	3 (3.06)	0	0
Non-compliance with Study Drug	1 (0.71)	1 (1.02)	0	0
Physician Decision	1 (0.71)	1 (1.02)	0	0
Death	0	0	0	0
Lack of Efficacy	0	0	0	0
Other	0	0	0	0
Pregnancy	0	0	0	0
Protocol Violation	0	0	0	0
Study Termination by Sponsor	0	0	0	0
Subjects who Completed Phase 1, N (%)[3, 4]	121 (85.82)	90 (91.84)	18 (75.00)	13 (68.42)
Subjects who Entered Phase 2	118 (83.69)	88 (89.80)	17 (70.83)	13 (68.42)
Subjects who did not Enter Phase 2	3 (2.13)	2 (2.04)	1 (4.17)	0
Lost to Follow-up	1 (0.71)	1 (1.02)	0	0
Physician Decision	1 (0.71)	1 (1.02)	0	0
Withdrawal by Subject	1 (0.71)	0	1 (4.17)	0

FAS = Full analysis set.

[1] Subjects are summarized by the planned treatment for the categories of randomized, randomized but not dosed, and FAS population; and are summarized by the actual treatment for other categories. Subjects (b) (6) and (b) (6) were randomized to ALKS 5461 8/8 but received ALKS 5461 2/2 during Phase 1, so their randomized treatment and actual treatment are different.

[2] Percentages are based on the randomized subjects in the treatment group.

[3] Percentages are based on the subjects in the safety population in the treatment group.

[4] Subjects who completed taper period during Phase 1 are considered completing Phase 1.

Source: Table 14.1.1.1 of Applicant's Clinical Study Report (Page 15-16)

Table 2: Summary of Analysis Population and Disposition in Phase 2 (Study 202; Phase 2 FAS)

Category	All	Phase 1-Phase 2 Treatment Sequence [1]		
		PBO(NR)-PBO	PBO(NR)-ALKS 5461 2/2	PBO(NR)-ALKS 5461 8/8
Subjects Entering Phase 2, N	118	20	23	22
Subjects Entering Phase 2 but not Dosed, N	0	0	0	0
Subjects in the Safety Population, N	118	20	23	22
Subjects in the FAS Population, N (%) [2]	65 (55.08)	20 (100)	23 (100)	22 (100)
Subjects in the PK Population, N (%) [3]	30 (25.42)	0	15 (65.22)	15 (68.18)
Subjects who Discontinued during Phase 2, N (%) [3]	12 (10.17)	0	5 (21.74)	4 (18.18)
Adverse Event	9 (7.63)	0	5 (21.74)	4 (18.18)
Withdrawal by Subject	2 (1.69)	0	0	0
Lost to Follow-up	1 (0.85)	0	0	0
Death	0	0	0	0
Lack of Efficacy	0	0	0	0
Non-compliance with Study Drug	0	0	0	0
Physician Decision	0	0	0	0
Protocol Violation	0	0	0	0

Category	All	Phase 1-Phase 2 Treatment Sequence [1]		
		PBO(NR)-PBO	PBO(NR)-ALKS 5461 2/2	PBO(NR)-ALKS 5461 8/8
Subjects who Discontinued during Phase 2 (continued)				
Other	0	0	0	0
Pregnancy	0	0	0	0
Study Termination by Sponsor	0	0	0	0
Subjects who Completed Phase 2, N (%) [3, 4]	106 (89.83)	20 (100)	18 (78.26)	18 (81.82)
Subjects who Completed Study	105 (88.98)	20 (100)	18 (78.26)	18 (81.82)
Subject who did not Complete Study	1 (0.85)	0	0	0
Withdrawal by Subject	1 (0.85)	0	0	0

(See next page for footnotes and table source.)

PBO = Placebo; PBO(NR) = Placebo non-responder; PBO(R)= Placebo responder; FAS = Full analysis set.

A placebo non-responder is defined as a subject who has a HAM-D-17 total score >14 at baseline and has a <50% reduction in HAM-D-17 total score from baseline to the end the 4 week treatment period in Phase 1.

[1] Subjects are summarized by the planned treatment sequence for the categories of subjects entering Phase 2, subjects entering Phase 2 but not dosed, and FAS population; and are summarized by the actual treatment sequence for other categories. Subject (b) (6) was randomized to ALKS 5461 8/8 but received ALKS 5461 2/2 during Phase 1, so the subject's randomized treatment sequence and actual treatment sequence are different.

[2] Percentages are based on the subjects entering Phase 2 in the treatment group.

[3] Percentages are based on the subjects in the safety population in the treatment group.

[4] Subjects who completed taper period during Phase 2 are considered completing Phase 2.

Source: Table 14.1.1.2 of Applicant's Clinical Study Report (Page 17-18)

Table 3 refers to dropout rates across treatment periods in stages 1 and 2. The dropout rate is the highest (30%) in ALKS 5461 8/8 in stage 1; 21.7% in ALKS 5461 2/2 in stage 2. The dropout rate in the placebo group was the lowest (5.3% in stage 1 and zero dropouts in stage 2).

Table 3: Dropout Rates across Weeks for Each Stage (Study 202)

Stage 1			
	Placebo (%)	ALKS 5461 2/2 (%)	ALKS 5461 8/8 (%)
Week 1	1.05	0	0
Week 2	3.2	15	20
Week 3	2.1	15	25
Week 4	5.3	15	30
Stage 2			
Week 1	0	0	0
Week 2	0	17.4	4.5
Week 3	0	21.7	9.1
Week 4	0	21.7	18.2

Source: Reviewer's Result

Summary of baseline characteristics in the safety set for all subjects in phase 1 are presented in Table 4. The average age of patients was 46.3 years, ranging from 20 to 65 years. Overall, the scores and measurements at baseline were similar across the treatment groups with respect to age, gender, ethnicity, duration of current major depressive episode, and lifetime number of major depressive episodes. There was a noticeable higher proportion of the white in the placebo group, but the noticeable difference may be due to small sample sizes in the other two treatment groups. Many of the participants were females (68.09%), Caucasians (Whites including Hispanic) (72.34%) followed by Black or African American (26.95%). Majority of the patients were of non-Hispanic and non-Latino origin (92.20%).

Phase 2 FAS consists of only placebo non-responders from phase 1. Table 5 summarizes baseline characteristics across treatment groups for Phase 2 FAS by treatment sequence to which they were randomized sequentially. There were no systematic observed differences across treatment groups with respect to age and gender. The proportion of whites was larger in placebo group and ALKS 2/2 than ALKS 8/8. The mean duration of current major depressive episodes (months) and spread was smaller in ALKS 8/8 treatment group; the mean life time number of major depressive episodes and standard deviation is relatively smaller in ALKS 2/2 treatment group.

The stage 1 and stage 2 data don't show evidence of difference in baseline characteristics of

subjects between the two stages.

Table 4: Summary of Demographics and Baseline Characteristics (Study 202; Phase 1 Safety Population)

Variable	All (N=141)	Phase 1 Treatment Group: Placebo (N=98)	ALKS 5461 2/2 (N=24)	ALKS 5461 8/8 (N=19)
Age (years)				
n	141	98	24	19
Mean (SD)	46.3 (11.1)	46.6 (11.0)	45.2 (10.9)	45.8 (11.9)
Median	48.0	48.5	47.5	47.0
Q1 - Q3	40.0 - 54.0	39.0 - 55.0	39.5 - 52.0	41.0 - 56.0
Min - Max	20 - 65	20 - 65	23 - 65	20 - 61
Gender, n (%)				
Female	96 (68.09)	68 (69.39)	17 (70.83)	11 (57.89)
Male	45 (31.91)	30 (30.61)	7 (29.17)	8 (42.11)
Primary Race, n (%)				
White	102 (72.34)	74 (75.51)	15 (62.50)	13 (68.42)
Black or African American	38 (26.95)	23 (23.47)	9 (37.50)	6 (31.58)
Native Hawaiian or other Pacific Islander	1 (0.71)	1 (1.02)	0	0
American Indian or Alaska Native	0	0	0	0
Asian	0	0	0	0
Ethnicity, n (%)				
Not Hispanic or Latino	130 (92.20)	89 (90.82)	23 (95.83)	18 (94.74)
Hispanic or Latino	11 (7.80)	9 (9.18)	1 (4.17)	1 (5.26)
Duration of current major depressive episode (months)				
n	141	98	24	19
Mean (SD)	9.2 (7.6)	9.1 (8.2)	9.7 (6.1)	8.7 (6.1)
Median	7.0	7.0	8.5	6.0
Q1 - Q3	4.0 - 11.0	4.0 - 11.0	5.0 - 14.5	4.0 - 12.0
Min - Max	2 - 63	2 - 63	2 - 24	2 - 23
Lifetime number of major depressive episodes				
n	138	96	23	19
Mean (SD)	6.2 (5.9)	6.2 (5.8)	6.5 (5.7)	6.2 (7.4)
Median	4.0	4.0	4.0	4.0
Q1 - Q3	3.0 - 7.0	3.0 - 6.5	3.0 - 8.0	2.0 - 6.0
Min - Max	1 - 32	1 - 30	1 - 20	1 - 32

Height (cm)				
n	141	98	24	19
Mean (SD)	167.67 (8.86)	167.14 (9.17)	167.76 (6.34)	170.24 (9.85)
Median	167.00	166.35	167.80	171.50
Q1 - Q3	160.00 - 173.50	160.00 - 173.20	164.40 - 171.75	161.00 - 176.00
Min - Max	148.8 - 196.5	148.8 - 196.5	157.0 - 182.9	155.0 - 188.0
Weight (kg)				
n	141	98	24	19
Mean (SD)	83.07 (18.34)	80.86 (18.19)	87.28 (18.42)	89.11 (17.65)
Median	82.10	79.20	86.10	92.50
Q1 - Q3	70.80 - 94.60	68.50 - 89.20	74.85 - 97.60	77.90 - 100.70
Min - Max	44.2 - 130.0	44.2 - 122.5	53.1 - 130.0	55.6 - 121.2
Body Mass Index (kg/m^2)				
n	141	98	24	19
Mean (SD)	29.44 (5.56)	28.82 (5.47)	30.84 (5.23)	30.86 (6.11)
Median	29.43	27.67	32.00	32.77
Q1 - Q3	25.13 - 33.75	24.84 - 32.56	26.50 - 34.30	26.48 - 34.39
Min - Max	18.6 - 39.9	18.6 - 39.9	19.8 - 39.2	19.1 - 39.5

Source: Table 14.1.2.1 of Applicant's Clinical Study Report (Section 14 Tables-and-Figures, Page 21-24)

Table 5: Summary of Demographics and Baseline Characteristics (Study 202; Phase 2 FAS)

Variables	All (N=118)	Phase 1-Phase 2 Treatment Sequence: PBO(NR)-PBO (N=20)	PBO(NR)-ALKS 5461 2/2 (N=23)	PBO(NR)-ALKS 5461 8/8 (N=22)
Age (years)				
n	118	20	23	22
Mean (SD)	46.2 (11.3)	45.5 (12.7)	47.5 (11.1)	46.9 (10.2)
Median	48.0	46.0	51.0	47.5
Q1 - Q3	39.0 - 55.0	36.0 - 56.0	42.0 - 57.0	39.0 - 55.0
Min - Max	20 - 65	22 - 65	27 - 63	26 - 63
Gender, n (%)				
Female	76 (64.41)	15 (75.00)	14 (60.87)	12 (54.55)
Male	42 (35.59)	5 (25.00)	9 (39.13)	10 (45.45)
Primary Race, n (%)				
White	89 (75.42)	19 (95.00)	18 (78.26)	14 (63.64)
Black or African American	28 (23.73)	1 (5.00)	5 (21.74)	7 (31.82)
Native Hawaiian or other Pacific Islander	1 (0.85)	0	0	1 (4.55)
American Indian or Alaska Native	0	0	0	0
Asian	0	0	0	0
Ethnicity, n (%)				
Not Hispanic or Latino	108 (91.53)	19 (95.00)	22 (95.65)	20 (90.91)
Hispanic or Latino	10 (8.47)	1 (5.00)	1 (4.35)	2 (9.09)
Duration of current major depressive episode (months)				
n	118	20	23	22
Mean (SD)	9.4 (7.9)	10.0 (8.0)	9.2 (6.8)	8.3 (3.6)
Median	7.5	9.0	8.0	7.5
Q1 - Q3	5.0 - 11.0	4.5 - 11.5	4.0 - 11.0	6.0 - 11.0
Min - Max	2 - 63	3 - 37	2 - 32	3 - 19
Lifetime number of major depressive episodes				
n	115	20	22	21
Mean (SD)	6.1 (5.7)	6.2 (6.3)	4.3 (2.0)	7.5 (7.4)
Median	4.0	3.5	4.0	4.0
Q1 - Q3	3.0 - 6.0	3.0 - 5.5	3.0 - 5.0	3.0 - 9.0
Min - Max	1 - 32	2 - 21	1 - 10	1 - 25

Variables	All (N=118)	Phase 1-Phase 2 Treatment Sequence		
		PBO(NR)-PBO (N=20)	PBO(NR)-ALKS 5461 2/2 (N=23)	PBO(NR)-ALKS 5461 8/8 (N=22)
Height (cm)				
n	118	20	23	22
Mean (SD)	167.81 (9.24)	168.56 (10.41)	168.19 (9.70)	168.77 (8.52)
Median	166.65	166.60	166.00	171.00
Q1 - Q3	160.00 - 175.00	159.00 - 175.10	161.00 - 175.00	162.50 - 175.00
Min - Max	148.8 - 196.5	154.2 - 188.0	153.0 - 196.5	154.0 - 184.0
Weight (kg)				
n	118	20	23	22
Mean (SD)	83.82 (18.95)	84.67 (19.84)	88.36 (19.82)	81.81 (18.40)
Median	82.85	85.35	85.30	79.20
Q1 - Q3	70.00 - 96.40	70.00 - 103.75	75.70 - 103.50	72.00 - 89.20
Min - Max	44.2 - 130.0	56.2 - 116.0	53.2 - 122.5	44.2 - 118.2
Body Mass Index (kg/m^2)				
n	118	20	23	22
Mean (SD)	29.63 (5.64)	29.59 (5.40)	31.07 (5.67)	28.62 (5.66)
Median	29.67	29.84	31.26	27.11
Q1 - Q3	25.27 - 34.06	24.44 - 34.73	27.06 - 37.19	24.84 - 31.73
Min - Max	18.6 - 39.9	20.8 - 37.4	20.7 - 39.9	18.6 - 39.6

PBO = Placebo; PBO(NR) = Placebo non-responder; PBO(R)= Placebo responder
SD = standard deviation; Q1 = first quartile; Q3 = third quartile; Min = minimum; Max = maximum
[1] A subject can be on more than one antidepressant therapy.

Source: Table 14.1.2.2 of Applicant's Clinical Study Report (Section 14 Tables-and-Figures, Page 25-32)

Baseline psychiatric efficacy measurements are also summarized for phase 1 and phase 2 treatment periods. Table 6 below shows the mean (standard deviation) of HAM-D_{17} total score, MADRS total score, CGI-S at baseline in the phase 1 and phase 2 primary analysis sets. The baseline measurements were fairly balanced across treatment groups in phase 1, and in phase 2 they were not considerably different across treatment groups. However, these baseline psychiatric efficacy measurements appeared to be lower in the phase 2 analysis set than in phase 1.

Table 6: Baseline Psychiatric Efficacy Measurements (Study 202; Phase 1 and Phase 2 FAS)

Psychiatric Efficacy Endpoints	Population	Placebo Mean (SD)	ALKS-5461 (2/2) Mean (SD)	ALKS-5461 (8/8) Mean (SD)
HAM-D_{17}	Phase 1	23.2 (4.2)	22.7 (4.2)	21.7 (3.3)
	Phase 2	17.3 (8.8)	16.1 (5.9)	19.0 (5.5)
MADRS-10	Phase 1	31.0 (5.6)	31.1 (5.6)	28.1 (4.1)
	Phase 2	23.6 (12.5)	21.6 (9.0)	26.2 (7.4)
CGI-S	Phase 1	4.4 (0.5)	4.5 (0.5)	4.6 (0.6)
	Phase 2	3.6 (1.2)	3.5 (1.1)	4.2 (0.8)

Source: Phase 1: Table 14.1.3.1 of Applicant's Clinical Study Report (Section 14 Tables-and-Figures, Page 33-34); Phase 2: Table 14.1.3.2 of Applicant's Clinical Study Report (Section 14 Tables-and-Figures, Page 36-37)

3.2.3.2 Study 207

The study was conducted at 57 sites but randomized subjects were from only 46 sites (37 sites in the United States in which 2 sites were in Puerto Rico, 2 sites in Canada, and 7 sites in Germany). First subject visit was on July 24, 2014 and last visit on September 27, 2016.

As summarized in Table 7, the efficacy analyses populations includes 398 subjects in FAS1 who were randomized to Placebo (273), ALKS-5461 1/1 (62), ALKS-5461 2/2 (63), and 185 subjects in FAS2 who were randomized to Placebo (60), ALKS-5461 1/1 (62), ALKS-5461 2/2 (63).

Table 8 summarizes dropout rates across treatment weeks in the FAS population. The rates correspond to the ones based on the randomized set.

Table 7: Data Sets Analyzed (Study 202)

Category	Treatment Groups			
	PBO	ALKS 5461 1/1	ALKS 5461 2/2	All
Efficacy Populations				
Randomized Population, N	281	63	63	407
Stage 1 Full Analysis Set, N[a]	273	62	63	398
Stage 2 Full Analysis Set, N[b]	60	62	63	185
Safety Populations				
Stage 1 Safety Population, N[c]	280	63	63	406
Stage 2 Safety Population 1, N[d]	62	62	63	187
Continued Treatment Population, N[e]	131	56	48	235

Abbreviations: PBO=placebo

[a] Subjects in the Stage 1 Safety Population who had at least one postbaseline assessment of MADRS-10 score in Stage 1.

[b] Subjects in the Stage 2 Safety Population 1 who had at least one postbaseline assessment of MADRS-10 in Stage 2

[c] All randomized subjects who received at least one dose of study drug (double-blind placebo or ALKS 5461) during Stage 1.

[d] Stage 1 placebo non-responders entering Stage 2 and who received at least one dose of study drug (double-blind placebo or ALKS 5461) during Stage 2.

[e] Subjects who were in the Stage 2 Safety Population 1 or Stage 2 Safety Population 2 and received the same treatment for both stages.

Source: Table 9 of Applicant's Clinical Study Report (Page 87)

Table 8: Dropout Rates Across Weeks for Each Stage (Study 207; FAS)

Stage 1			
	Placebo (%)	ALKS 5461 1/1 (%)	ALKS 5461 2/2 (%)
Week 1	0	0	0
Week 2	1.1	4.8	6.3
Week 3	2.9	9.7	15.9
Week 4	3.7	9.7	22.2
Week 5	5.9	11.3	22.2
Stage 2			
Week 1	0	0	0
Week 2	0	1.6	3.2
Week 3	1.7	4.8	6.3
Week 4	3.3	4.8	7.9
Week 5	5.0	4.8	7.9
Week 6	6.7	6.4	7.9

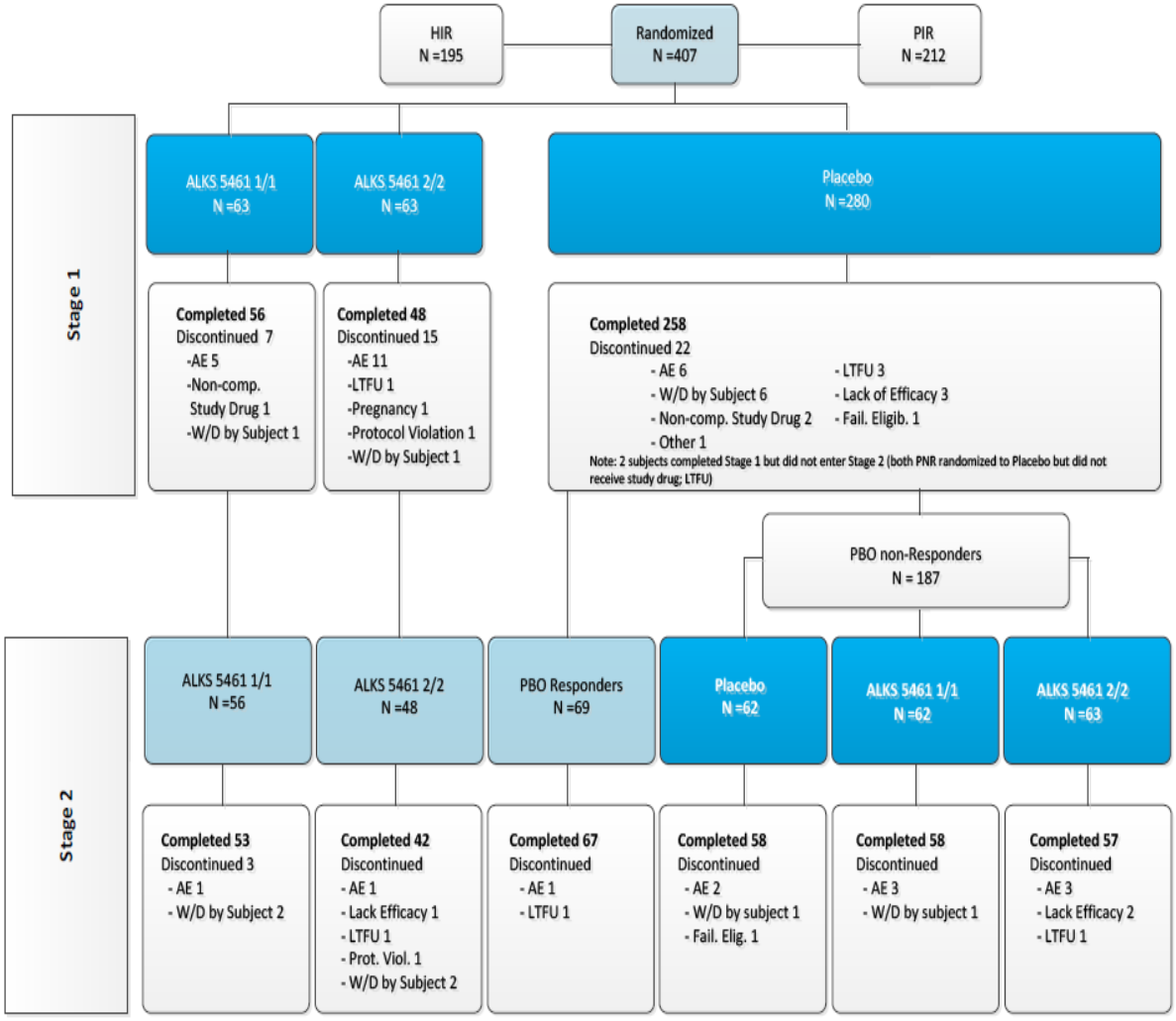
Source: Reviewer's Result

Figure 5 shows the number of completers as well as numbers of discontinued subjects, including discontinuation reasons, in each stage. Table 9 displays the rate of discontinuation in stage 1 for all randomized subjects in the 3 treatment groups: placebo (7.9%), ALKS 5461 1/1 (11.1%), and ALKS 5461 2/2 (23.8%), which was similar across treatment groups. The frequent discontinuation reasons included adverse event (placebo: 2.1%, ALKS 1/1: 7.9%, ALKS 2/2: 17.5%), withdrawal by subject (placebo: 2.1%, ALKS 1/1: 1.6%, ALKS 2/2: 1.6%), loss to follow-up (placebo: 1.1%, ALKS 1/1: 0%, ALKS 2/2: 01.6%) and lack of efficacy (placebo: 1.1%, ALKS 1/1: 0%, ALKS 2/2: 0%). In stage 2, the dropout rates for the all randomized set were: placebo (6.5%), ALKS 5461 1/1 (6.5%), ALKS 5461 2/2 (9.5%). Adverse event was the main reason for discontinuation in the second stage (placebo: 3.2%, ALKS 1/1: 4.8%, ALKS 2/2: 4.8%).

Baseline demographic characteristics, psychiatric history for stage 1 and stage 2 populations are summarized in Table 10 and Table 11. In stage 1, the average age of patients was 45.2 years, ranging from 18 to 69 years. Overall, treatment groups were balanced with respect to age, sex, race, weight and BMI. Majority of participants were white (74.1%), Non-Hispanic or Latino origin (83.5%), and females (68.2%). In stage 2, the average age was 45.0 years, majority of subjects were white (73.3%) and females (68.4%).

Taking the smaller sample size in stage 2 into account, similar baseline patterns were also observed as noted in Table 11.

Figure 5: Subject Disposition (Study 207; Stage 1 and Stage 2)



Abbreviations: AE=adverse event; Fail. Elig=failure to meet eligibility requirement; HIR=Historical Inadequate Responder; PBO=placebo; PIR=Prospective Inadequate Responder; LTFU=lost to follow-up; Non-comp=Non-compliance; W/D=withdrawal

Note: Non-compliance with study drug was defined as a lack of adherence to study drug regimen.

Note: In stage 1, one subject was randomized to placebo but never received study drug due to failure to meet randomization criteria and therefore is not included in figure.

Source: Figure 5 (Cross Reference: Table 14.1.1.1, 14.1.1.2, 14.1.1.3) of Applicant's Clinical Study Report (Page 82)

Table 9: Disposition of Randomized Subjects in Stage 1 (Study 207; All Randomized Subjects)

Category	Stage 1 Treatment Group		
	PBO	ALKS 5461 1/1	ALKS 5461 2/2
Subjects Randomized, N	281[a]	63	63
Subjects in the Stage 1 Safety Population, N	280	63	63
Subjects who Completed Stage 1 Treatment Period[b], n (%)	258 (92.1)	56 (88.9)	48 (76.2)
Subjects who Discontinued the Study During Stage 1, n (%)	22 (7.9)	7 (11.1)	15 (23.8)
Reason for Discontinuation, n (%)			
Adverse Event	6 (2.1)	5 (7.9)	11 (17.5)
Withdrawal by Subject	6 (2.1)	1 (1.6)	1 (1.6)
Lack of Efficacy	3 (1.1)	0	0
Lost to Follow-up	3 (1.1)	0	1 (1.6)
Non-compliance with Study Drug	2 (0.7)	1 (1.6)	0
Other	1 (0.4)	0	0
Pregnancy	0	0	1 (1.6)
Protocol Violation	0	0	1 (1.6)
Failure to Meet Eligibility Criteria	1 (0.4)	0	0
Subjects who Entered Stage 2, n (%)	256[c] (91.4)	56 (88.9)	48 (76.2)

PBO=Placebo.

[a]Subject (b) (6) was randomized to placebo but never received study drug due to failure to meet randomized criteria.
[b] Subjects who completed the Stage 1 Treatment Period are those who had assessments at Visit 7 (Day 36)
[c] Two subjects (b) (6) and (b) (6)) who completed Stage 1 did not enter Stage 2. They were both randomized to placebo for Stage 2 and both were lost to follow-up

Source: Table 6 of Applicant's Clinical Study Report (Page 83)

Table 10: Summary of Demographics and Baseline Characteristics (Study 207; Stage 1 Safety Population)

Variable Statistics	Stage 1 Treatment Group: Placebo (N=280)	ALKS 5461 1/1 (N=63)	ALKS 5461 2/2 (N=63)	All Subjects (N=406)
Age (years)				
n	280	63	63	406
Mean (SD)	45.7 (12.87)	45.1 (11.46)	42.9 (14.48)	45.2 (12.93)
Median	47.0	47.0	43.0	47.0
Min - Max	18 - 68	19 - 66	18 - 69	18 - 69
Gender, n (%)				
Female	193 (68.9)	42 (66.7)	42 (66.7)	277 (68.2)
Male	87 (31.1)	21 (33.3)	21 (33.3)	129 (31.8)
Primary Race, n (%)				
White	207 (73.9)	44 (69.8)	50 (79.4)	301 (74.1)
Black or African American	67 (23.9)	17 (27.0)	11 (17.5)	95 (23.4)
Asian	5 (1.8)	2 (3.2)	2 (3.2)	9 (2.2)
Native Hawaiian or Other Pacific Islander	1 (0.4)	0	0	1 (0.2)
American Indian or Alaska Native	0	0	0	0
Ethnicity, n (%)				
Not Hispanic or Latino	232 (82.9)	53 (84.1)	54 (85.7)	339 (83.5)
Hispanic or Latino	48 (17.1)	10 (15.9)	9 (14.3)	67 (16.5)
Region, n (%)				
US	229 (81.8)	54 (85.7)	52 (82.5)	335 (82.5)
Non-US	51 (18.2)	9 (14.3)	11 (17.5)	71 (17.5)
Subject Type, n (%)				
PIR	144 (51.4)	32 (50.8)	35 (55.6)	211 (52.0)
HIR	136 (48.6)	31 (49.2)	28 (44.4)	195 (48.0)
Duration of Current MDE (months)				
n	280	63	63	406
Mean (SD)	9.0 (5.50)	9.4 (5.19)	9.0 (5.30)	9.1 (5.41)
Median	8.0	9.0	7.0	8.0
Min - Max	2 - 24	2 - 22	3 - 24	2 - 24

Lifetime Number of MDEs[1], n (%)				
1	25 (8.9)	3 (4.8)	3 (4.8)	31 (7.6)
2	48 (17.1)	13 (20.6)	14 (22.2)	75 (18.5)
3-4	119 (42.5)	22 (34.9)	23 (36.5)	164 (40.4)
>4	88 (31.4)	25 (39.7)	23 (36.5)	136 (33.5)
Lifetime Number of Antidepressants at Screening, n (%)				
0	35 (12.5)	7 (11.1)	7 (11.1)	49 (12.1)
1	67 (23.9)	14 (22.2)	14 (22.2)	95 (23.4)
2	81 (28.9)	17 (27.0)	21 (33.3)	119 (29.3)
>2	97 (34.6)	25 (39.7)	21 (33.3)	143 (35.2)
Lifetime Number of Antidepressants at Randomization[2], n (%)				
1	61 (21.8)	13 (20.6)	10 (15.9)	84 (20.7)
2	88 (31.4)	19 (30.2)	24 (38.1)	131 (32.3)
>2	131 (46.8)	31 (49.2)	29 (46.0)	191 (47.0)
Number of Inadequate Responses to ADT in the Current MDE at Randomization[2], n (%)				
0	1 (0.4)	0	0	1 (0.2)
1	247 (88.2)	57 (90.5)	56 (88.9)	360 (88.7)
2	32 (11.4)	6 (9.5)	7 (11.1)	45 (11.1)
Total Lifetime ADTs Failed at Screening[3], n (%)				
0	69 (24.6)	18 (28.6)	16 (25.4)	103 (25.4)
1	113 (40.4)	22 (34.9)	26 (41.3)	161 (39.7)
>=2	98 (35.0)	23 (36.5)	21 (33.3)	142 (35.0)
Total Lifetime ADTs Failed at Randomization[2], n (%)				
1	143 (51.1)	32 (50.8)	33 (52.4)	208 (51.2)
>=2	137 (48.9)	31 (49.2)	30 (47.6)	198 (48.8)
ADT Class for Current MDE, n (%)				
SSRI	174 (62.1)	32 (50.8)	36 (57.1)	242 (59.6)
SNRI	76 (27.1)	22 (34.9)	20 (31.7)	118 (29.1)
Other	30 (10.7)	9 (14.3)	7 (11.1)	46 (11.3)

Height (cm)				
n	280	63	63	406
Mean (SD)	168.9 (9.2)	168.6 (10.5)	169.9 (10.5)	169.0 (9.6)
Median	168.0	170.0	170.5	168.2
Min - Max	144.8 - 193.0	150.0 - 198.0	139.7 - 194.3	139.7 - 198.0
Weight (kg)				
n	280	63	63	406
Mean (SD)	83.5 (19.3)	85.6 (22.1)	82.9 (18.7)	83.8 (19.6)
Median	81.7	84.3	82.8	81.8
Min - Max	40.7 - 141.1	47.8 - 140.8	46.1 - 121.0	40.7 - 141.1
Body Mass Index (kg/m^2)				
n	280	63	63	406
Mean (SD)	29.2 (5.7)	29.9 (6.0)	28.7 (5.7)	29.2 (5.8)
Median	28.6	28.9	28.2	28.7
Min - Max	17.8 - 41.7	18.1 - 40.0	18.6 - 40.0	17.8 - 41.7

Source: Summary of Demographics of Applicant's CSR (Table 14.1.3.1.1, Page 205-209)

Table 11: Summary of Demographics and Baseline Characteristics (Study 207; Stage 2 Safety Population)

Summary of Demographics and Baseline Characteristics - Stage 2 Safety Population 1

Variable Statistics	Stage 2 Treatment Group			
	Placebo (N=62)	ALKS 5461 1/1 (N=62)	ALKS 5461 2/2 (N=63)	All Subjects (N=187)
Age (years)				
n	62	62	63	187
Mean (SD)	44.0 (12.65)	44.1 (12.36)	46.9 (13.84)	45.0 (12.97)
Median	46.0	45.0	48.0	46.0
Min - Max	20 - 68	19 - 68	18 - 67	18 - 68
Gender, n (%)				
Female	46 (74.2)	40 (64.5)	42 (66.7)	128 (68.4)
Male	16 (25.8)	22 (35.5)	21 (33.3)	59 (31.6)
Primary Race, n (%)				
White	42 (67.7)	49 (79.0)	46 (73.0)	137 (73.3)
Black or African American	18 (29.0)	11 (17.7)	16 (25.4)	45 (24.1)
Asian	2 (3.2)	2 (3.2)	1 (1.6)	5 (2.7)
American Indian or Alaska Native	0	0	0	0
Native Hawaiian or Other Pacific Islander	0	0	0	0
Ethnicity, n (%)				
Not Hispanic or Latino	48 (77.4)	51 (82.3)	56 (88.9)	155 (82.9)
Hispanic or Latino	14 (22.6)	11 (17.7)	7 (11.1)	32 (17.1)
Region, n (%)				
US	53 (85.5)	55 (88.7)	52 (82.5)	160 (85.6)
Non-US	9 (14.5)	7 (11.3)	11 (17.5)	27 (14.4)
Subject Type, n (%)				
PIR	27 (43.5)	40 (64.5)	33 (52.4)	100 (53.5)
HIR	35 (56.5)	22 (35.5)	30 (47.6)	87 (46.5)
Duration of Current MDE (months)				
n	62	62	63	187
Mean (SD)	10.0 (5.35)	9.3 (5.69)	8.7 (4.96)	9.3 (5.34)
Median	8.0	8.0	8.0	8.0
Min - Max	2 - 24	2 - 24	2 - 22	2 - 24

Number of Inadequate Responses to ADT in the Current MDE at Randomization[2], n (%)				
0	0	0	1 (1.6)	1 (0.5)
1	53 (85.5)	54 (87.1)	55 (87.3)	162 (86.6)
2	9 (14.5)	8 (12.9)	7 (11.1)	24 (12.8)
Total Lifetime ADTs Failed at Screening[3], n (%)				
0	12 (19.4)	17 (27.4)	19 (30.2)	48 (25.7)
1	29 (46.8)	20 (32.3)	22 (34.9)	71 (38.0)
>=2	21 (33.9)	25 (40.3)	22 (34.9)	68 (36.4)
Total Lifetime ADTs Failed at Randomization[2], n (%)				
1	30 (48.4)	29 (46.8)	33 (52.4)	92 (49.2)
>=2	32 (51.6)	33 (53.2)	30 (47.6)	95 (50.8)
ADT Class for Current MDE, n (%)				
SSRI	31 (50.0)	43 (69.4)	40 (63.5)	114 (61.0)
SNRI	23 (37.1)	11 (17.7)	17 (27.0)	51 (27.3)
Other	8 (12.9)	8 (12.9)	6 (9.5)	22 (11.8)
Lifetime Number of MDEs[1], n (%)				
1	4 (6.5)	4 (6.5)	9 (14.3)	17 (9.1)
2	12 (19.4)	12 (19.4)	9 (14.3)	33 (17.6)
3-4	24 (38.7)	28 (45.2)	27 (42.9)	79 (42.2)
>4	22 (35.5)	18 (29.0)	18 (28.6)	58 (31.0)
Lifetime Number of Antidepressants at Screening, n (%)				
0	7 (11.3)	9 (14.5)	7 (11.1)	23 (12.3)
1	18 (29.0)	13 (21.0)	16 (25.4)	47 (25.1)
2	15 (24.2)	19 (30.6)	18 (28.6)	52 (27.8)
>2	22 (35.5)	21 (33.9)	22 (34.9)	65 (34.8)
Lifetime Number of Antidepressants at Randomization[2], n (%)				
1	13 (21.0)	14 (22.6)	12 (19.0)	39 (20.9)
2	22 (35.5)	17 (27.4)	23 (36.5)	62 (33.2)
>2	27 (43.5)	31 (50.0)	28 (44.4)	86 (46.0)
Height (cm)				
n	62	62	63	187
Mean (SD)	168.5 (7.2)	169.5 (10.0)	169.2 (9.7)	169.0 (9.0)
Median	167.6	170.1	168.0	168.0
Min - Max	154.9 - 185.4	144.8 - 190.0	149.0 - 190.5	144.8 - 190.5
Weight (kg)				
n	62	62	63	187
Mean (SD)	81.9 (18.0)	82.7 (21.0)	85.8 (20.0)	83.5 (19.7)
Median	78.9	81.8	85.0	81.8
Min - Max	53.4 - 129.7	40.7 - 121.6	53.8 - 141.1	40.7 - 141.1
Body Mass Index (kg/m^2)				
n	62	62	63	187
Mean (SD)	28.8 (5.7)	28.5 (5.8)	29.8 (5.6)	29.1 (5.7)
Median	26.9	27.8	30.1	28.4
Min - Max	19.8 - 40.9	17.8 - 39.0	19.5 - 40.6	17.8 - 40.9

Source: Summary of Demographics of Applicant's CSR (Table 14.1.3.1.2, Page 210-214)

Table 12: Summary of Psychiatric History - Stage 1 and Stage 2 Treatment Periods

Variable Statistics	Stage 1 Treatment Group (Stage 1 Safety Population)			Stage 2 Treatment Group (Stage 2 Safety Population 1)		
	PBO (N=280)	ALKS 5461 1/1 (N=63)	ALKS 5461 2/2 (N=63)	PBO (N=62)	ALKS 5461 1/1 (N=62)	ALKS 5461 2/2 (N=63)
Duration of current MDE (months)						
Mean (SD)	9.0 (5.50)	9.4 (5.19)	9.0 (5.30)	10.0 (5.35)	9.3 (5.69)	8.7 (4.96)
Median	8.0	9.0	7.0	8.0	8.0	8.0
Min - Max	2 - 24	2 - 22	3 - 24	2 - 24	2 - 24	2 - 22
Lifetime number of MDEs[a], n (%)						
1	25 (8.9)	3 (4.8)	3 (4.8)	4 (6.5)	4 (6.5)	9 (14.3)
2	48 (17.1)	13 (20.6)	14 (22.2)	12 (19.4)	12 (19.4)	9 (14.3)
3-4	119 (42.5)	22 (34.9)	23 (36.5)	24 (38.7)	28 (45.2)	27 (42.9)
>4	88 (31.4)	25 (39.7)	23 (36.5)	22 (35.5)	18 (29.0)	18 (28.6)
Lifetime number of antidepressants at randomization[b], n (%)						
1	61 (21.8)	13 (20.6)	10 (15.9)	13 (21.0)	14 (22.6)	12 (19.0)
2	88 (31.4)	19 (30.2)	24 (38.1)	22 (35.5)	17 (27.4)	23 (36.5)
>2	131 (46.8)	31 (49.2)	29 (46.0)	27 (43.5)	31 (50.0)	28 (44.4)
Number of inadequate responses to ADT in the current MDE at randomization[b], n (%)						
0	1 (0.4)	0	0	0	0	1 (1.6)
1	247 (88.2)	57 (90.5)	56 (88.9)	53 (85.5)	54 (87.1)	55 (87.3)
≥2	32 (11.4)	6 (9.5)	7 (11.1)	9 (14.5)	8 (12.9)	7 (11.1)
Total lifetime ADTs failed at randomization[b], n (%)						
1	143 (51.1)	32 (50.8)	33 (52.4)	30 (48.4)	29 (46.8)	33 (52.4)
≥2	137 (48.9)	31 (49.2)	30 (47.6)	32 (51.6)	33 (53.2)	30 (47.6)
ADT class for current MDE, n (%), n (%)						
SSRI	174 (62.1)	32 (50.8)	36 (57.1)	31 (50.0)	43 (69.4)	40 (63.5)
SNRI	76 (27.1)	22 (34.9)	20 (31.7)	23 (37.1)	11 (17.7)	17 (27.0)
Other	30 (10.7)	9 (14.3)	7 (11.1)	8 (12.9)	8 (12.9)	6 (9.5)

Abbreviations: ADT=antidepressant therapy; Max=maximum; MDE=major depressive episode; Min=minimum; PBO=placebo; SD=standard deviation; SNRI=serotonin-norepinephrine reuptake inhibitor; SSRI=selective serotonin reuptake inhibitor

[a] Lifetime number of MDEs includes the current episode

[b] The value of the variables 'The number of inadequate responses to ADT in the current MDE at randomization,' 'Lifetime Number of Antidepressants at Randomization' and 'Total Lifetime ADTs Failed at Randomization' is the value collected at screening visit and the value collected at screening visit + 1 for HIR and PIR subjects, respectively

Note: "Other" ADT class included bupropion

Source: Table 11 of Applicant's Clinical Study Report (Page 90)

3.2.4 Results and Conclusions

3.2.4.1 Study 202

Primary Endpoint

The reviewer confirmed the Applicant's numerical findings (Table 13 through Table 15). ALKS-5461 2/2 dose appeared to have a larger improvement from baseline in HAM-D_{17} total score than placebo. Combining results from the two stages, the least squares mean treatment difference between ALKS-5461 2/2 and placebo from baseline to week 4 in HAM-D_{17} total score was -2.8 (p = 0.014, unadjusted for multiple testing). As mentioned above, this study was designed as a proof-of-concept study and there was no pre-specified plan for multiple testing. Later the Applicant explored data using a post-hoc Bonferroni procedure to adjust for multiple testing. In this reviewer's view, this attempt could be construed as "significant p-value" fishing expedition. Cherry-picking convenient but unspecified multiplicity adjustment methods after the trial results are known often make the statistical interpretability of trial results highly questionable. Furthermore, we noted that for the phase 3 studies included in this NDA, where multiple dose levels were investigated, the Applicant pre-specified a fixed sequence testing procedure starting from the high dose, not Bonferroni procedure, to control the overall type I error rate. Applying such a fixed sequence testing procedure starting from the high dose 8/8 versus placebo would have rendered Study 202 to be an inconclusive study; that is, the trial would have failed on the 8/8 dose by this fixed sequence testing procedure and no further testing could be performed. Thus, this reviewer concludes that in Study 202, statistical significance for the 2/2 dose remains inconclusive.

Table 13: Change from Baseline in HAM-D_{17} Total Score At Week 4 (Study 202; Stage 1 FAS; MMRM)

	Placebo N=95	ALKS 5461 2/2 N=20	ALKS 5461 8/8 N=20
Baseline			
N	95	20	20
Mean (SD)	23.2 (4.2)	22.7 (4.2)	21.7 (3.3)
Change from Baseline at Week 4			
N	90	17	14
LS Mean (SE)	-7.1 (0.6)	-9.3 (1.5)	-6.6 (1.6)
LSMD vs Placebo (SE)		-2.2 (1.6)	0.5 (1.7)
p-value		0.168	0.787

LS = least-squares; LSMD = least-squares mean difference; MMRM = mixed model repeated measures; SE = standard error; SD= Standard Deviation; FAS = Full Analysis Set
Source: Table 11 of Applicant's Clinical Study Report (Page 65)

Table 14: Change from Baseline in HAM-D_{17} Total Score At Week 4 (Study 202; Stage 2 FAS; MMRM)

	Placebo N=20	ALKS 5461 2/2 N=23	ALKS 5461 8/8 N=22
Baseline			
N	20	23	22
Mean (SD)	17.3 (8.8)	16.1 (5.9)	19.0 (5.5)
Change from Baseline at Week 4			
N	20	18	18
LS Mean (SE)	-1.5 (1.1)	-5.2 (1.2)	-3.3 (1.1)
LSMD vs Placebo (SE)		-3.7 (1.6)	-1.9 (1.6)
p-value		0.020	0.241

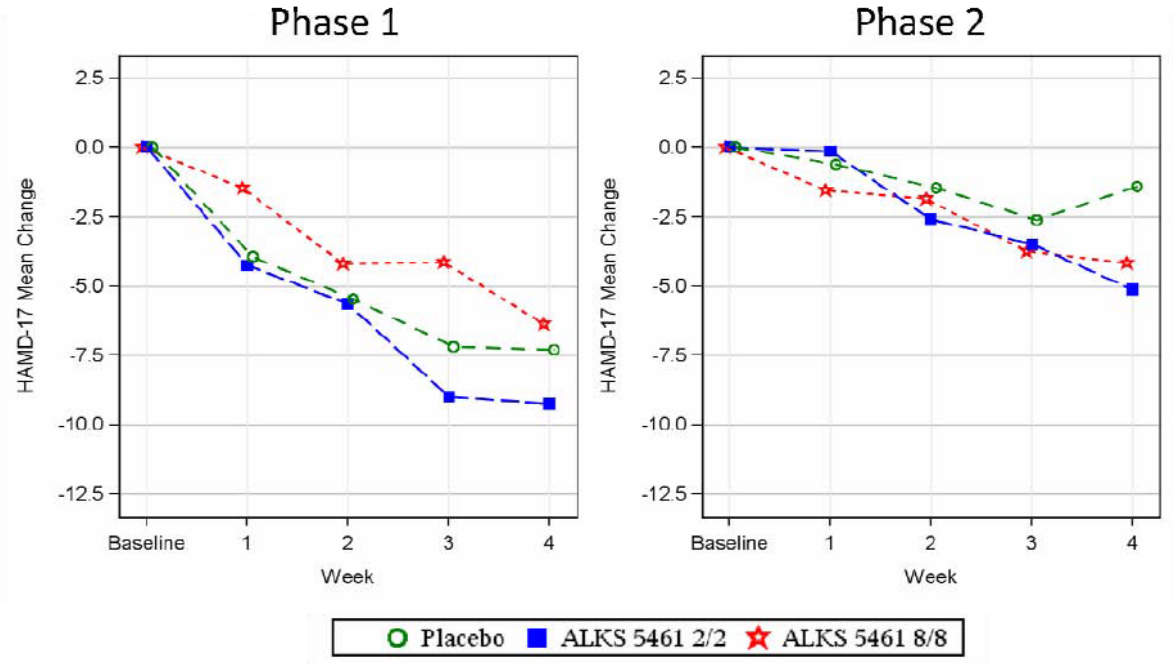
LS = least-squares; LSMD = least-squares mean difference; MMRM = mixed model repeated measures; SE = standard error; SD= Standard Deviation; FAS = Full Analysis Set
Source: Table 11 of Applicant's Clinical Study Report (Page 65)

Table 15: Stage 1 and Stage 2 Combined Inference Based on HAM-D_{17} Total Score (Study 202)

Phase 1 and Phase 2 Combined (SPCD)	ALKS 5461 2/2	ALKS 5461 8/8
Phase 1 Weight = 0.6; Phase 2 Weight = 0.4		
LSMD vs Placebo (SE)	-2.8 (1.2)	-0.5 (1.2)
p-value	0.014	0.699

Source: Table 11 of Applicant's Clinical Study Report (Page 65)
LSMD = least-squares mean difference; SE = standard error

Figure 6: Change from Baseline in HAM-D_{17} Total Score by Visit (Study 202; FAS)



Source: Figure 3 (Cross Reference Tables 14.2.1.1.1 and 14.2.1.1.2) of Applicant's Clinical Study Report (Page 66)

Sensitivity analyses using MMRM with SPCD weights of 0.7, 0.3 for stage 1 and stage 2 yielded the same conclusions as the primary efficacy analysis results.

Secondary Exploratory Efficacy Endpoint: MADRS-10 Total Score

The Applicant analyzed MADRS-10 total score as an exploratory endpoint. The least squares mean change in MADRS-10 total score from phase 1 baseline to Week 4 was -9.6 (0.9), -13.3 (2.2), and -11.3 (2.3) for the placebo, ALKS 5461 2/2, and ALKS 5461 8/8, respectively (Table 16). In phase 2, the least squares estimates were -2.1 (1.6), -8.8 (1.7), and -4.7 (1.7) for the placebo, ALKS 5461 2/2, and ALKS 5461 8/8, respectively (Table 17). Dose 2/2 statistically significantly separated from placebo in phase 2 (nominal $p = 0.005$) if unadjusted for multiple testing (Table 17). When data from the two phases were combined, the results were consistent with those from phase 2 ($p = 0.004$, unadjusted for multiple testing) (Table 18).

Table 16: Change from Baseline in MADRS Total Score At Week 4 (Study 202; Stage 1 FAS; MMRM)

Stage 1	Placebo N=95	ALKS 5461 2/2 N=20	ALKS 5461 8/8 N=20
Baseline			
N	95	20	20
Mean (SD)	31.0 (5.6)	31.1 (5.6)	28.1 (4.1)
Change from Baseline at Week 4			
N	90	17	14
LS Mean (SE)	-9.6 (0.9)	-13.3 (2.2)	-11.3 (2.3)
LSMD vs Placebo (SE)		-3.7 (2.4)	-1.8 (2.5)
p-value		0.119	0.483

LS = least-squares; LSMD = least-squares mean difference; MMRM = mixed model repeated measures; SE = standard error; SD= Standard Deviation; FAS = Full Analysis Set
Source: Table 14.2.3.3.1 of Applicant's Clinical Study Report (Page 342)

Table 17: Change from Baseline in MADRS Total Score At Week 4 (Study 202; Stage 2 FAS; MMRM)

Stage 2	Placebo N=20	ALKS 5461 2/2 N=23	ALKS 5461 8/8 N=22
Baseline			
N	20	23	22
Mean (SD)	23.6 (12.5)	21.6 (9.0)	26.2 (7.4)
Change from Baseline at Week 4			
N	20	18	18
LS Mean (SE)	-2.1 (1.6)	-8.8 (1.7)	-4.7 (1.7)
LSMD vs Placebo (SE)		-6.7 (2.3)	-2.6 (2.3)
p-value		0.005	0.260

LS = least-squares; LSMD = least-squares mean difference; MMRM = mixed model repeated measures; SE = standard error; SD= Standard Deviation; FAS = Full Analysis Set
Source: Table 14.2.3.3.2 of Applicant's Clinical Study Report (Page 344)

Table 18: Stage 1 and Stage 2 Combined Inference Based on MADRS Total Score (Study 202)

Phase 1 and Phase 2 Combined (SPCD)	ALKS 5461 2/2	ALKS 5461 8/8
Phase 1 Weight = 0.6; Phase 2 Weight = 0.4		
LSMD vs Placebo (SE)	-4.9 (1.7)	-2.1 (1.8)
p-value	0.004	0.233

Source: Table 14.2.3.3.3 of Applicant's Clinical Study Report (Page 345)
LSMD = least-squares mean difference; SE = standard error

Reviewer's Note 1: Table 19 and Figure 7 are displayed to explore if the assumption of homogeneous variances of HAM-D_{17} total scores between treatment arms in each stage holds. The descriptive statistics suggest that this assumption might not be far-fetched.

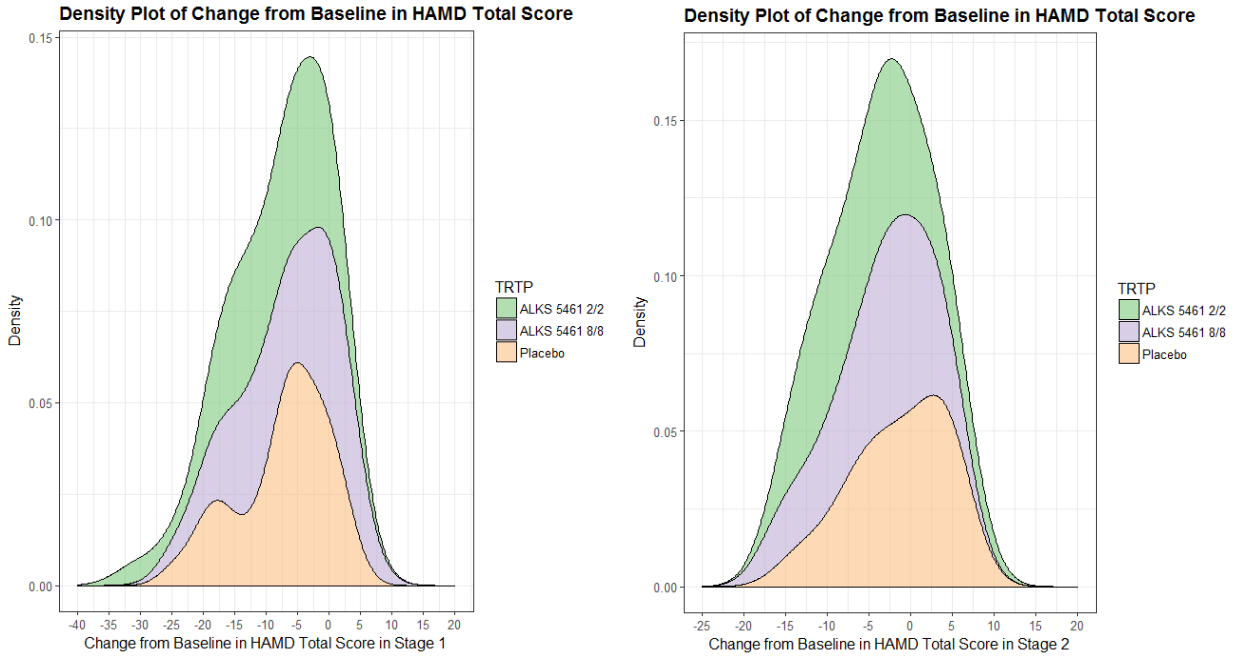
Table 19: Descriptive Statistics (Observed Cases): HAM-D_{17} Total Score at Baseline and Week 4 (Study 202)

Stage 1	Placebo N=95	ALKS 5461 2/2 N=20	ALKS 5461 8/8 N=20
Baseline			
N	95	20	20
Mean (SD)	23.2 (4.2)	22.7 (4.2)	21.7 (3.3)
Week 4			
N	90	17	14
Mean (SD)	15.8 (7.3)	13.0 (6.3)	15.4 (6.6)
Change from Baseline to Week 4			
Mean (SD)	-7.3 (7.3)	-9.2 (8.1)	-6.36 (7.9)

Stage 2	Placebo N=20	ALKS 5461 2/2 N=23	ALKS 5461 8/8 N=22
Baseline			
N	20	23	22
Mean (SD)	17.3 (8.8)	16.1 (5.9)	19.0 (5.5)
Week 4			
N	20	18	18
Mean (SD)	15.9 (7.8)	11.1 (7.3)	15.1 (6.6)
Change from Baseline to Week 4			
Mean (SD)	-1.4 (5.8)	-5.1 (6.0)	-4.2 (5.9)

N: number of patients; SD: standard deviation
Source: Reviewer's Result

Figure 7: Density Plots of Change from Baseline in HAM-D_{17} Total Score by Stage (Study 202)



Note: The total area under the density plot of the three treatment arms sums to 1.
Source: Reviewer's Result

Table 20 displays the correlation matrices for stage 1 and stage 2, respectively, where stage 2 was for placebo non-responders from stage 1. These are correlations that would be obtained if each treatment group were to be analyzed separately. Note that there are two dose groups in the treatment arm (ALKS 5461 2/2, ALKS 5461 8/8). The diagonal entries display the within-group variances at each of the visits and off-diagonal entries are within-group correlations between any two visits. The primary efficacy endpoint was HAM-D_{17}. For each dose group and placebo, the cells in the tables are populated with the correlation coefficient and the sample size used to calculate the correlation. For example, in stage 1, the correlation between the baseline and HAM-D_{17} total score at Week 4 are: 0.30, -0.16, -0.08 for placebo, ALKS 5461 2/2, and ALKS-5461 8/8 groups, respectively. Likewise, the correlation between baseline of stage 2 (that is, end of stage 1 measurement) and stage 2 measurement at Week 4 are 0.77, 0.62, 0.52 for the Placebo, ALKS 5461 2/2, and ALKS 5461 8/8 groups, respectively. Based on Table 20, there appeared to be some differences in between-visit correlations between the treatment arms. These differences need to be accounted for in statistical analysis.

Table 20: Pearson Correlation Coefficients Between Visit-wise HAM-D_{17} Total Scores Within Each Stage (Study 202)

Placebo

	Stage 1					**Stage 2**				
	Baseline	Week 1	Week 2	Week 3	Week 4	Baseline	Week 1	Week 2	Week 3	Week 4
Baseline	1.00 95	0.36 94	0.32 92	0.33 93	0.30 90	1.00 20	0.73 20	0.84 20	0.84 20	0.77 20
Week 1	0.36 94	1.00 94	0.64 91	0.66 92	0.60 89	0.73 20	1.00 20	0.69 20	0.64 20	0.63 20
Week 2	0.32 92	0.64 91	1.00 92	0.71 91	0.78 89	0.84 20	0.69 20	1.00 20	0.94 20	0.90 20
Week 3	0.33 93	0.66 92	0.71 91	1.00 93	0.77 90	0.83 20	0.64 20	0.94 20	1.00 20	0.87 20
Week 4	0.30 90	0.60 89	0.78 89	0.77 90	1.00 90	0.77 20	0.63 20	0.90 20	0.87 20	1.00 20

Top number in cell: correlation; bottom number in cell: sample size used to calculate the correlation.

ALKS 5461 2/2

	Stage 1					**Stage 2**				
	Baseline	Week 1	Week 2	Week 3	Week 4	Baseline	Week 1	Week 2	Week 3	Week 4
Baseline	1.00 20	0.46 20	0.14 17	-0.30 17	-0.16 17	1.00 23	0.56 23	0.76 19	0.57 18	0.62 18
Week 1	0.46 20	1.00 20	0.62 17	0.43 17	0.48 17	0.56 23	1.00 23	0.45 19	0.30 18	0.44 18
Week 2	0.14 17	0.62 17	1.00 17	0.68 17	0.58 17	0.76 19	0.45 19	1.00 19	0.84 18	0.84 18
Week 3	-0.30 17	0.43 17	0.68 17	1.00 17	0.76 17	0.57 18	0.30 18	0.84 18	1.00 18	0.93 18
Week 4	-0.16 17	0.48 17	0.58 17	0.76 17	1.00 17	0.62 18	0.44 18	0.84 18	0.93 18	1.00 18

ALKS 5461 8/8

	Stage 1					Stage 2				
	Baseline	Week 1	Week 2	Week 3	Week 4	Baseline	Week 1	Week 2	Week 3	Week 4
Baseline	1.00	0.49	0.06	0.23	-0.08	1.00	0.55	0.49	0.62	0.52
	20	20	16	15	14	22	22	21	20	18
Week 1	0.49	1.00	0.53	0.42	0.37	0.55	1.00	0.76	0.73	0.74
	20	20	16	15	14	22	22	21	20	18
Week 2	0.06	0.54	1.00	0.71	0.79	0.49	0.76	1.00	0.78	0.68
	16	16	16	15	14	21	21	21	20	18
Week 3	0.24	0.42	0.71	1.00	0.65	0.62	0.73	0.78	1.00	0.83
	15	15	15	15	14	20	20	20	20	18
Week 4	-0.08	0.37	0.79	0.65	1.00	0.52	0.74	0.68	0.83	1.00
	14	14	14	14	14	18	18	18	18	18

Source: Reviewer's Result

Reviewer's Note 2: Accurate inference of the fixed effects in MMRM analysis depends on specification of reasonable covariance structure for the longitudinal data. Study 202 assumed a first-order autoregressive variance-covariance structure, i.e., AR(1), which is a strong assumption that rarely holds. Re-analyzing the dataset using unstructured variance-covariance matrix, which is typically used in statistical analysis for psychiatric trials, does not conclude statistical significance of ALKS 5461 2/2 ($p = 0.057$, unadjusted for multiple testing). Based on Pearson's correlation coefficients and the AIC criterion, AR(1) variance-covariance structure does not seem to be the best assumption as compared to the unstructured variance-covariance. See Table 21 below.

Table 21: AR(1) versus UN Covariance Structure (Study 202)

Stage 1	AIC	BIC	-2Log-Lik
AR(1)	3020.2	3026.6	3016.2
UN	2974.5	3003.6	2954.5
Stage 2			
AR(1)	1328.0	1332.4	1324.0
UN	1315.8	1337.5	1295.8

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; Log-Lik: log-likelihood
Source: Reviewer's Results

Reviewer's Note 3: In Study 202, the clinical reviewer discovered potential financial conflict of interest by investigators at sites 124 and 128. As such, this reviewer examined efficacy results using an ad-hoc sensitivity analysis by removing each site. Site 124 had only one subject, who was on the treatment sequence "ALKS 5461 2/2 – Placebo". Site 128 had two subjects and both were on the treatment sequence "Placebo – Placebo". Statistical significance is inconclusive in both dose groups when site 124 is removed (The magnitude of treatment effect of ALKS 5461 2/2 versus placebo dropped from 2.8 to 2.2 and the magnitude of treatment effect of ALKS 5461 8/8 versus placebo dropped from 0.5 to 0.4). Removal of site 128 has little impact.

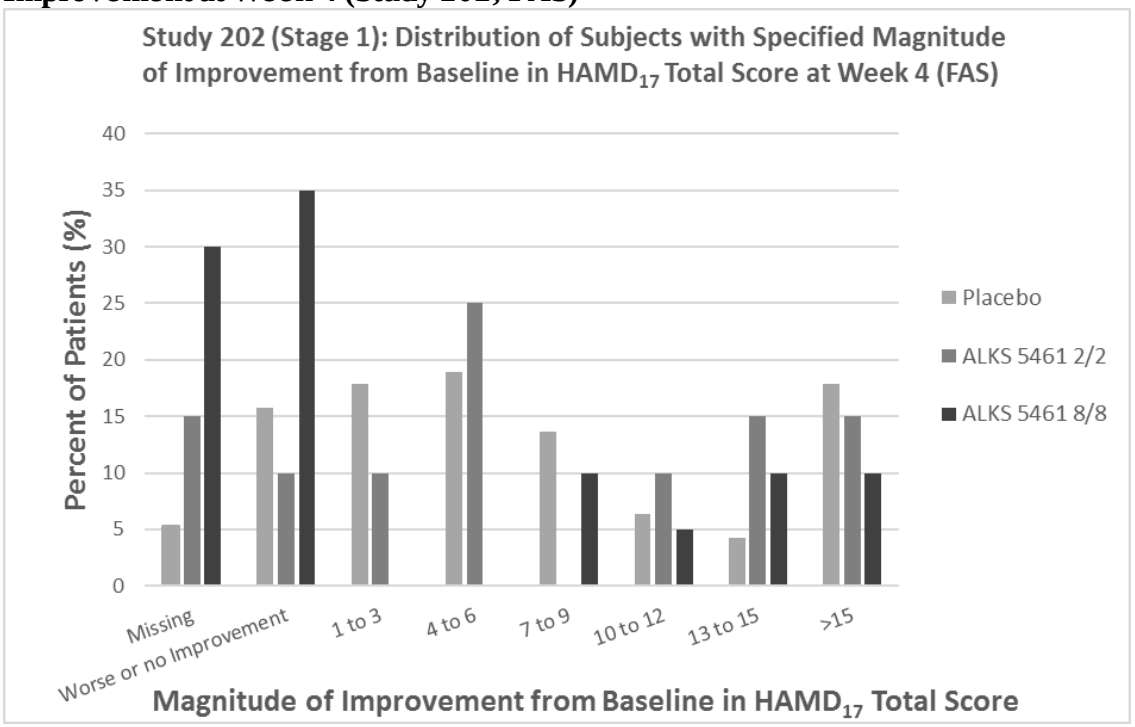
Reviewer's Note 4: This reviewer has included two figures to visualize the distribution of change in HAM-D_{17} total score and percent of improved subjects (Figure 8, Figure 9) for study 202. Distribution of improved subjects is categorized in a 2-unit bin; subjects who didn't show improvement and missing are located in the left corner bins.

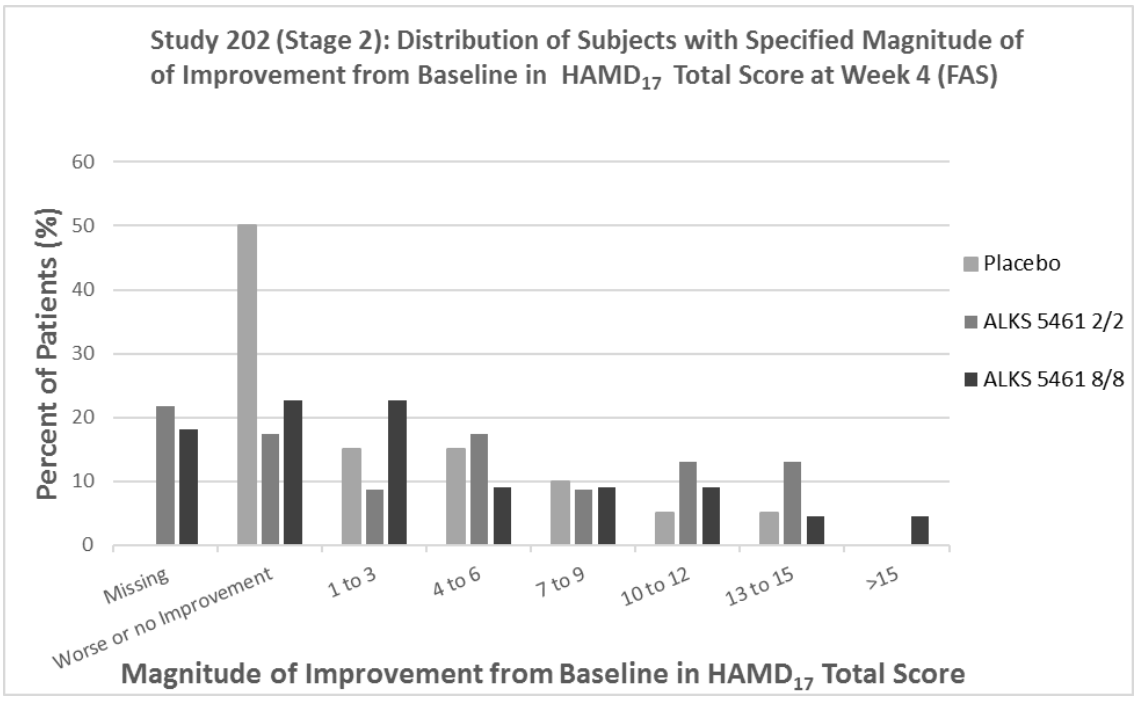
The graphs in Figure 8 show distributions of change from baseline in HAM-D_{17} total score at week 4 in stage 1 and stage 2. In stage 1, in the dose group ALKS 5461 8/8, there were no subjects with magnitudes of improvement '1-3' and '4-6', and smaller proportions of subjects with larger magnitudes of improvement. This could be the main factor for the small effect of ALKS 5461 8/8 noted in the primary efficacy analysis of stage 1.

In stage 2, except in bin category '1-3' (where the proportion was higher for the ALKS 5461 8/8 group), the dose group ALKS 5461 2/2 generally shows larger proportions of subjects with improvement compared to placebo.

The graphs in
Figure 9 were the cumulative distributions of specific magnitude of improvement in HAM-D_{17} total score in stage 1 and stage 2. In stage 1, there were more proportions of subjects in placebo and ALKS 5461 2/2 with improved symptoms of MDD (at categories ≥ 1, ≥ 3, ≥ 5, and ≥ 7). High placebo response was expected in stage 1. This was compatible with the primary efficacy results. In stage 2, the dose group ALKS 5464 2/2 showed consistent improvement at each response cutoff. For example, approximately 57% of subjects improved at least by 3 points.

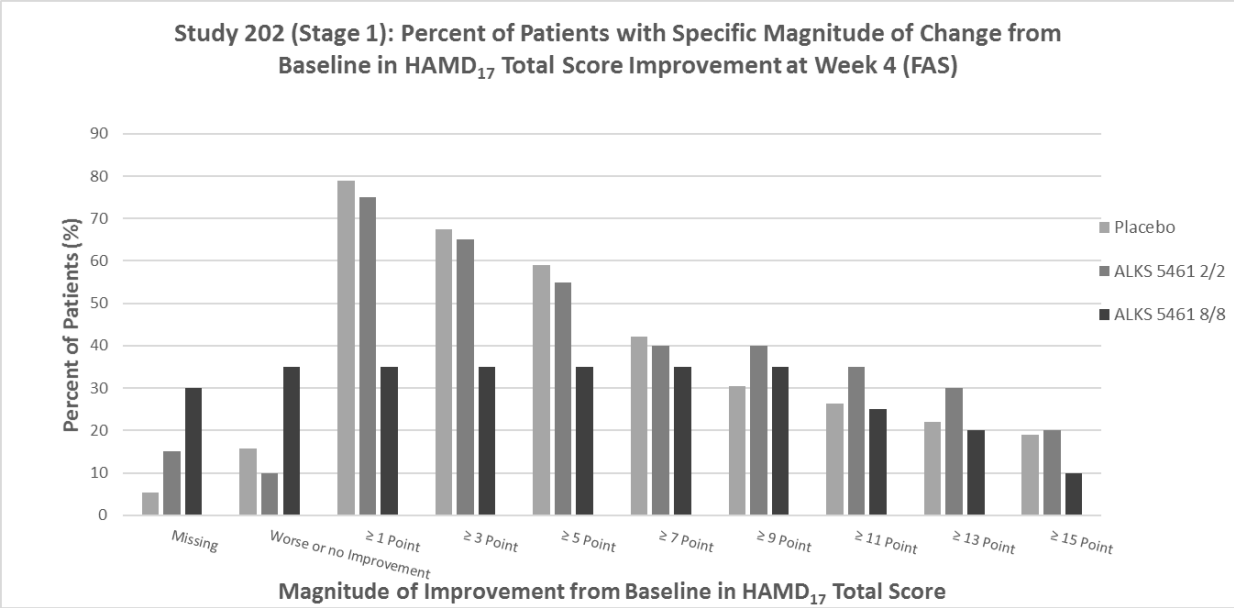
Figure 8: Percent of Patients with Specified Magnitude of HAM-D_{17} Total Score Improvement at Week 4 (Study 202; FAS)

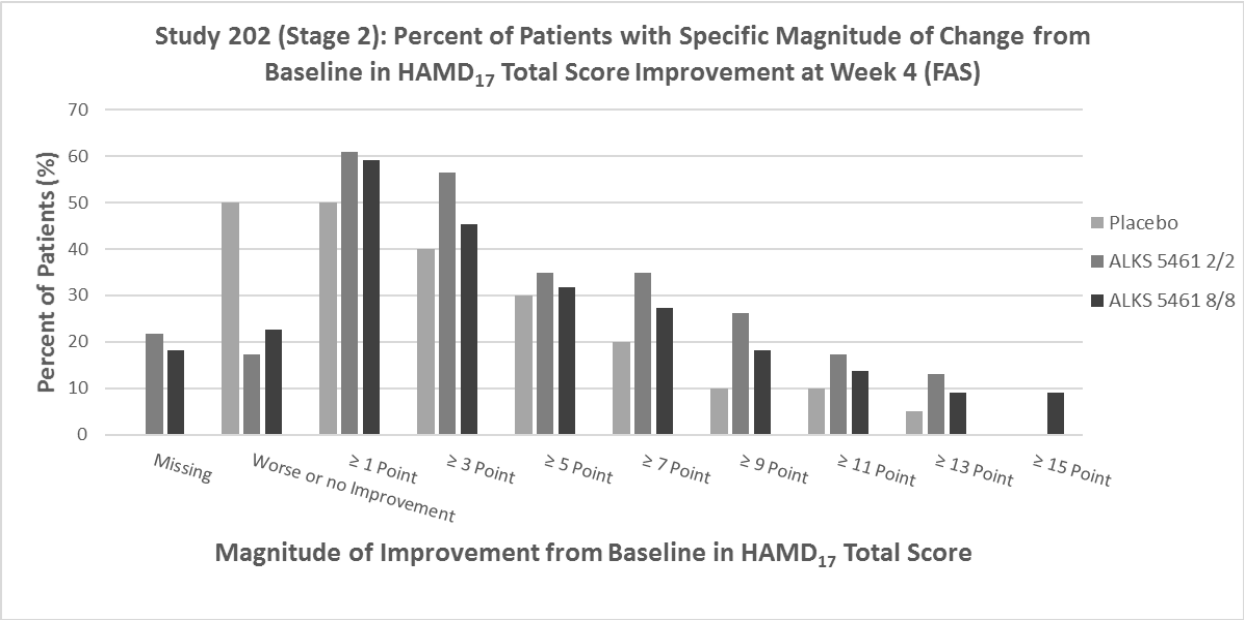




Source: Reviewer's Result

Figure 9: Percent Response (HAM-D_{17} Total Score) at Week 4 by Various Response Thresholds (Study 202; Stage 1 FAS)





Source: Reviewer's Result

3.2.4.2 Study 207

Primary Endpoint

The reviewer confirmed the Applicant's efficacy findings (Table 22 through Table 24). ALKS-5461 2/2 dose showed a statistically significantly larger improvement than placebo on the endpoints: MADRS-6 and MADRS-10 total scores averaged from Week 3 to the end of each stage (Week 5 for stage 1 and Week 6 for stage 2). The least squares mean treatment differences [LSMD (SE)] in MADRS-6_{AVG} and MADRS-10_{AVG} were -1.5 (0.62) and -1.9 (0.86), respectively, with unadjusted p-values 0.018 and 0.026 respectively.

Since there was no agreement on the Applicant defined endpoint, MADRS-6, between the Applicant and FDA, the focus will center on analysis based the MADRS-10 endpoint.

Note on MADRS-10_{AVG}: The Applicant specified average change from baseline across multiple time points (week 3 to week 5 in stage 1; week 3 to week 6 in stage 2) in MADRS-10. The rationale behind the averaging scheme is to provide a more precise estimate of the treatment effect. That is, averaging point estimates reduces the variability (noise) because of increasing sample size to triple or quadruple.

Table 22: Change from Baseline in MADRS-6, MADRS-10 Total Score at Week 5 (Study 207; Stage 1 FAS; MMRM)

Stage 1	Placebo N=273	ALKS 5461 1/1 N=62	ALKS 5461 2/2 N=63
MADRS-6			
Baseline N Mean (SD)	 273 21.4 (3.71)	 62 21.2 (3.58)	 63 21.3 (4.06)
Change from Baseline at Week 5 N LS Mean (SE) LSMD vs Placebo (SE) p-value	 257 -6.3 (0.48)	 55 -6.9 (0.864) -0.6 (0.95) 0.551	 49 --7.1 (0.89) -0.8 (0.97) 0.426
Average from End of Treatment: Change from Baseline at Week 5 LS Mean (SE) LSMD vs Placebo (SE) p-value	 -5.6 (0.34)	 -6.0 (0.74) -0.5 (0.81) 0.565	 -6.8 (0.75) -1.3 (0.83) 0.123
MADRS-10			
Baseline N Mean (SD)	 273 31.7 (5.64)	 62 31.8 (5.27)	 63 31.8 (5.64)
Change from Baseline at Week 5 N LS Mean (SE) LSMD vs Placebo (SE) p-value	 257 -9.2 (0.55)	 55 -10.3 (1.19) -1.1 (1.31) 0.382	 49 -10.8 (1.22) -1.6 (1.34) 0.220
Average from End of Treatment: Change from Baseline at Week 6 LS Mean (SE) LSMD vs Placebo (SE) p-value	 -8.1 (0.48)	 -8.8 (1.05) -0.7 (1.15) 0.541	 -10.3 (1.06) -2.2 (1.17) 0.063

LS = least-squares; LSMD = least-squares mean difference; MMRM = mixed model repeated measures; SE = standard error; SD= Standard Deviation; FAS = Full Analysis Set
Source: Table 14.2, 14.2.1.1.1, 14.2.1.1.2 of Applicant's Clinical Study Report (Page 307-308, 747-748)

Table 23: Change from Baseline in MADRS-6, MADRS-10 Total Score At Week 6 (Study 207; Stage 2 FAS; MMRM)

Stage 2	Placebo N=60	ALKS 5461 1/1 N=62	ALKS 5461 2/2 N=63
MADRS-6			
Baseline N Mean (SD)	 60 17. (5.18)	 62 18.5 (4.66)	 63 17.9 (4.66)
Change from Baseline at Week 6 N LS Mean (SE) LSMD vs Placebo (SE) p-value	 56 -1.4 (0.71)	 58 -2.3 (0.73) -1.0 (1.02) 0.331	 58 -3.2 (0.73) -1.9 (1.01) 0.068
Average from End of Treatment: Change from Baseline at Week 6 LS Mean (SE) LSMD vs Placebo (SE) p-value	 -1.5 (0.65)	 -2.2 (0.67) -0.7 (0.93) 0.430	 -3.2 (0.67) -1.7 (0.93) 0.075
MADRS-10			
Baseline N Mean (SD)	 60 26.3 (7.58)	 62 27.7 (7.15)	 63 26.0 (6.45)
Change from Baseline at Week 6 N LS Mean (SE) LSMD vs Placebo (SE) p-value	 56 -1.9 (0.96)	 58 -3.4 (0.98) -1.5 (1.37) 0.281	 58 -3.6 (0.98) -1.7 (1.37) 0.203
Average from End of Treatment: Change from Baseline at Week 6 LS Mean (SE) LSMD vs Placebo (SE) p-value	 -2.1 (0.88)	 -3.2 (0.91) -1.2 (1.26) 0.362	 -3.7 (0.90) -1.6 (1.25) 0.192

Source: Table 14.2.1.1.1, 14.2.1.1.2 of Applicant's Clinical Study Report (Page 316-317, 756-757)

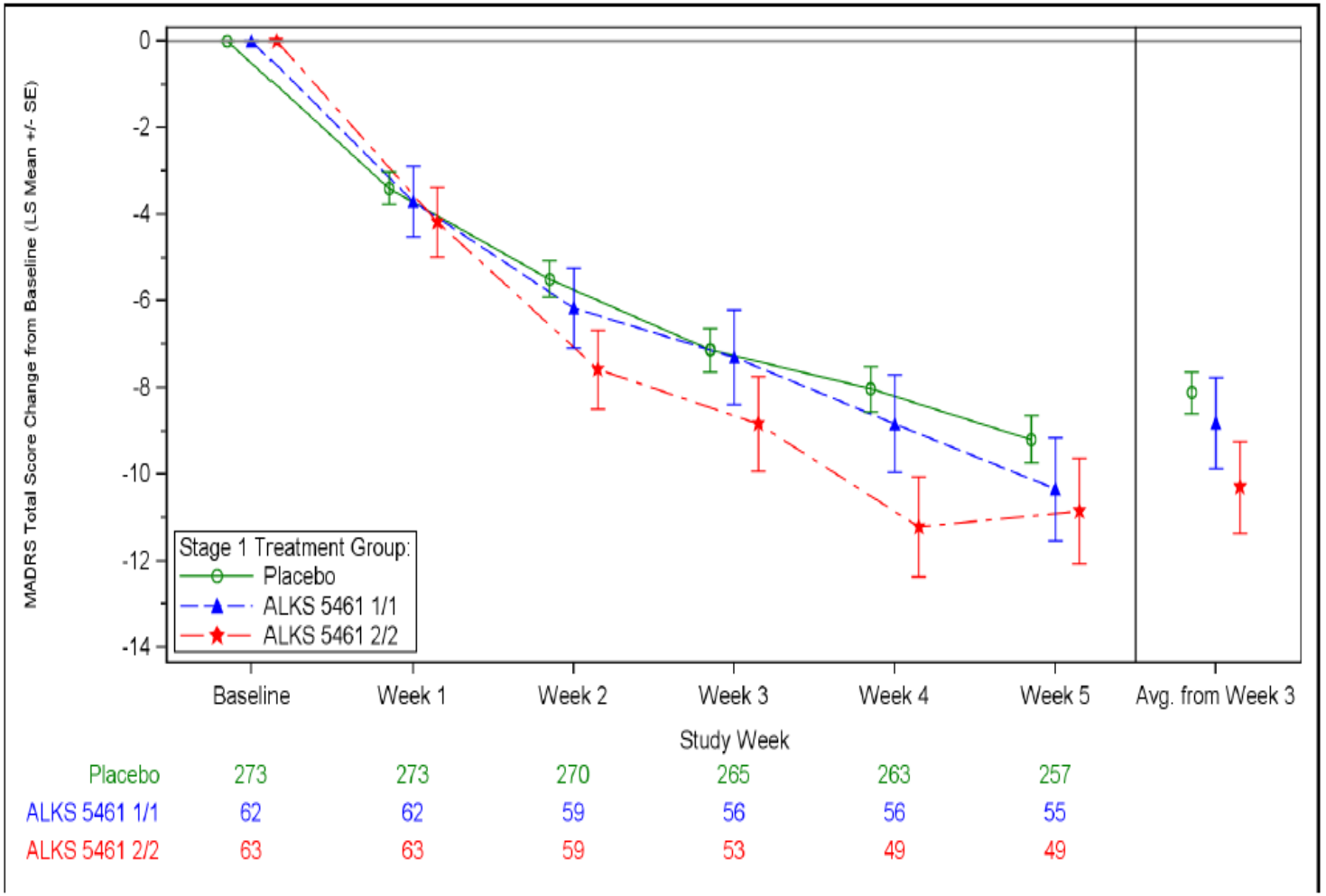
Table 24: Primary Efficacy Endpoints: Weighted Analysis of Change from Baseline in MADRS-6_{AVG}, MADRS-10_{AVG}, MADRS-10_{EOT} (Study 207; FAS)

Primary Efficacy Endpoint Statistics	Treatment Group	
	ALKS 5461 1/1 vs Placebo	ALKS 5461 2/2 vs Placebo
Number of Subjects per Treatment Group and Stage		
ALKS 5461, Stage 1, n/Stage 2, n	62/62	63/63
Placebo, Stage 1, n/Stage 2, n	273/60	273/60
MADRS-6_{AVG}		
LSMD (SE)	-0.6 (0.62)	-1.5 (0.62)
95% CI	(-1.8, 0.6)	(-2.7, -0.3)
P-value	0.329	0.018
MADRS-10_{AVG}		
LSMD (SE)	-0.9 (0.85)	-1.9 (0.86)
95% CI	(-2.6, 0.7)	(-3.6, -0.2)
P-value	0.277	0.026
MADRS-10_{EOT}		
LSMD (SE)	-1.3 (0.95)	-1.7 (0.96)
95% CI	(-3.2, 0.5)	(-3.6, 0.2)
P-value	0.165	0.076

CI = confidence interval; EOT = end of treatment; AVG = average; LSMD = least squares mean difference; MMRM = mixed model repeated measures; SE = standard error.
Source: Table 18 of Applicant's Clinical Study Report (Page 99)

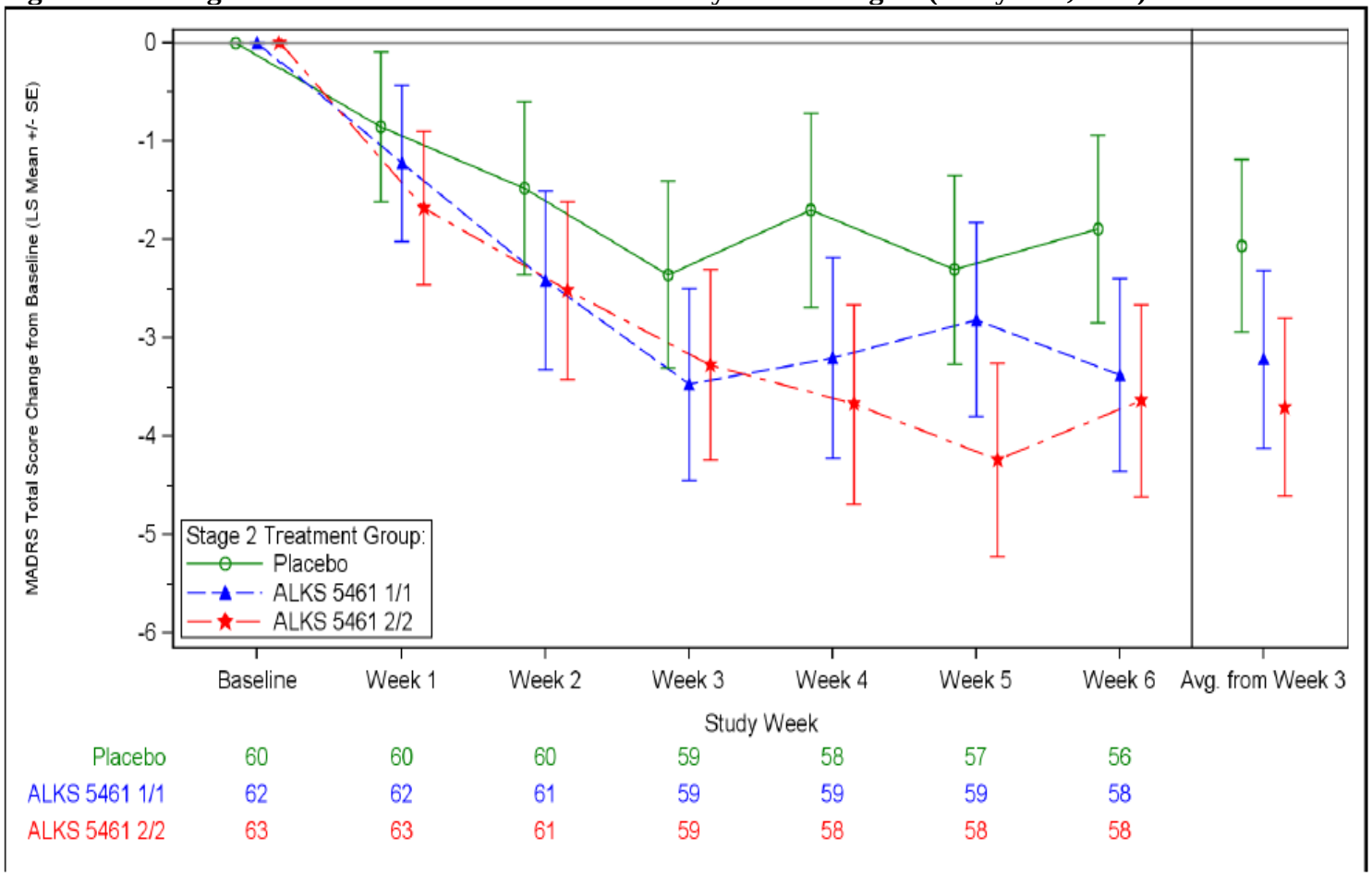
There was generally visual improvement in depression scores in ALKS 5461 2/2 in stage 1 (up to week 4) and stage 2 (up to week 5) as depicted in Figure 10 and Figure 11. In the last week of stage 1, the improvement leveled off while there was an uptick at week 6 in stage 2. The forest plot and bar plot in Figure 12 and Figure 13 summarize trial results that were numerically confirmed in Table 16 through Table 18.

Figure 10: Change from Baseline in MADRS-10 Score by Visit in Stage 1 (Study 207; FAS)



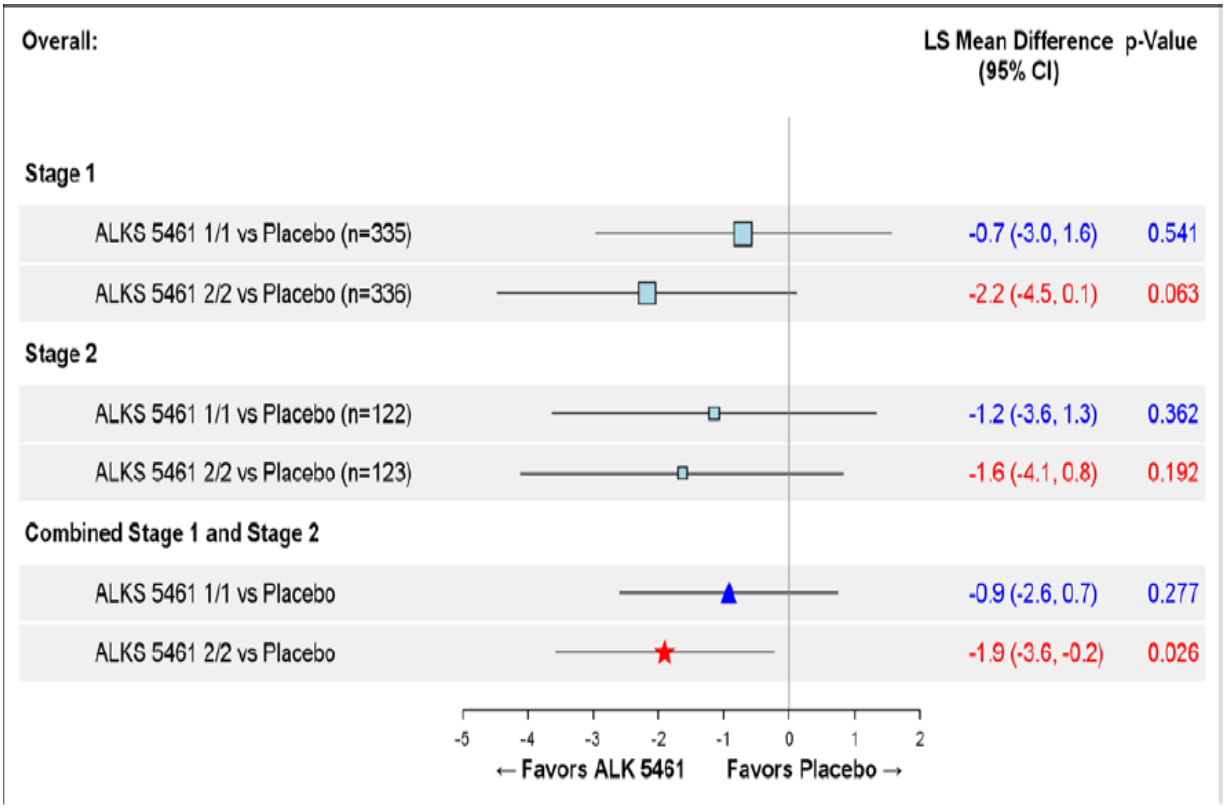
LS mean = least squares mean; SE = standard error
Note: Baseline is the randomization baseline at Visit 2
Source: Table 4.2.2.1.1 of Applicant's Clinical Study Report (Page 1779)

Figure 11: Change from Baseline in MADRS-10 Score by Visit in Stage 2 (Study 207; FAS)



LS mean = least squares mean; SE = standard error
Note: Baseline is the stage 2 baseline at Visit 7
Source: Table 4.2.2.1.2 of Applicant's Clinical Study Report (Page 1780)

Figure 12: Change from Baseline to End of Each Stage (Week 5 for Stage 5, Week 6 for Stage 2) in MADRS-10_{AVG} Score (Study 207)

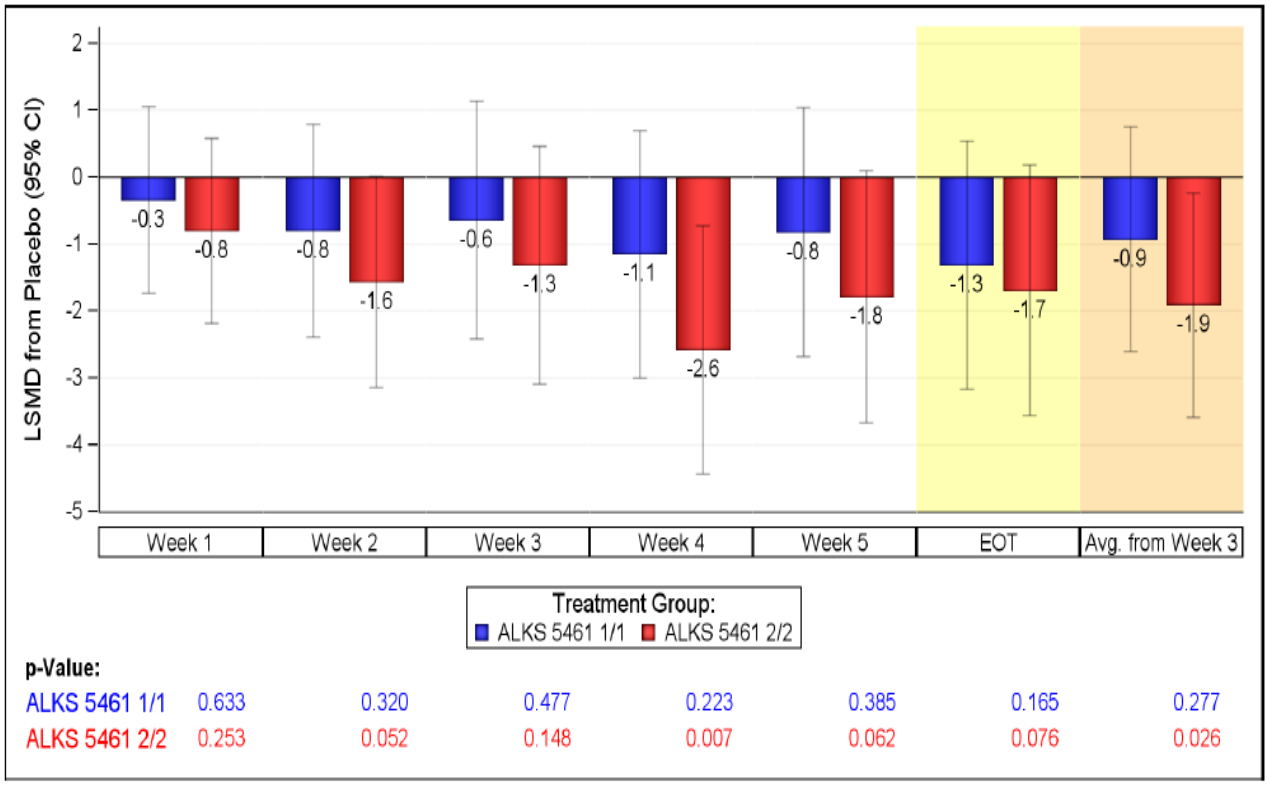


Abbreviations: CI=confidence interval; LS mean=least squares mean; MADRS=Montgomery-Åsberg Depression Rating Scale; MMRM=mixed-effects model for repeated measures

Note: The number of subjects (n) listed in Stage 1 and Stage 2 includes both the placebo group and ALKS treatment group subject counts. Analyses based on MMRM.

Source: Figure 11 (Cross Reference Figure 14.2.2.3) of Applicant's Clinical Study Report (Page 105)

Figure 13: Least Squares Mean Difference (95% CI) Between ALKS 5461 and Placebo in Change from Baseline in MADRS-10 Total Score by Visit Combined Across Stage 1 and Stage 2 (Study 207; MMRM with Equal Weights)



Note: CI = confidence interval; EOT = end of treatment; MMRM = mixed effects model for repeated measures.

Source: Figure 14.2.2.1.3 of Applicant's Clinical Study Report (Page 1781)

Sensitivity Analyses

The delta adjusted PMM was conducted as a tipping point analysis to identify when statistical significance is lost as the shift parameter changes progressively. Increasing the magnitude of the shift parameter* was associated with shrinking of treatment effect and loss of statistical significance. Statistical significance was lost when the (stage 1, stage 2) shift parameter was set at (2, 1) as compared to (0, 0) for the primary efficacy analysis.

**Shift parameter: a clinical assumption that subjects who discontinue at a given time point would have, on average, their unobserved efficacy score worsen by some amount delta compared with the observed efficacy score of subjects on the same treatment arm who continue to the next time point.* (Clinical Study Report, Page 106)

Reviewer's Note 5: Table 25 and Figure 14 explored the constant variance assumption of the continuous endpoint MADRS-10 in the three treatment groups: Placebo, ALKS 5461 1/1, ALKS 5461 2/2. The assumptions seem to be reasonable, though there were still some differences in standard deviation (SD) among the treatment groups.

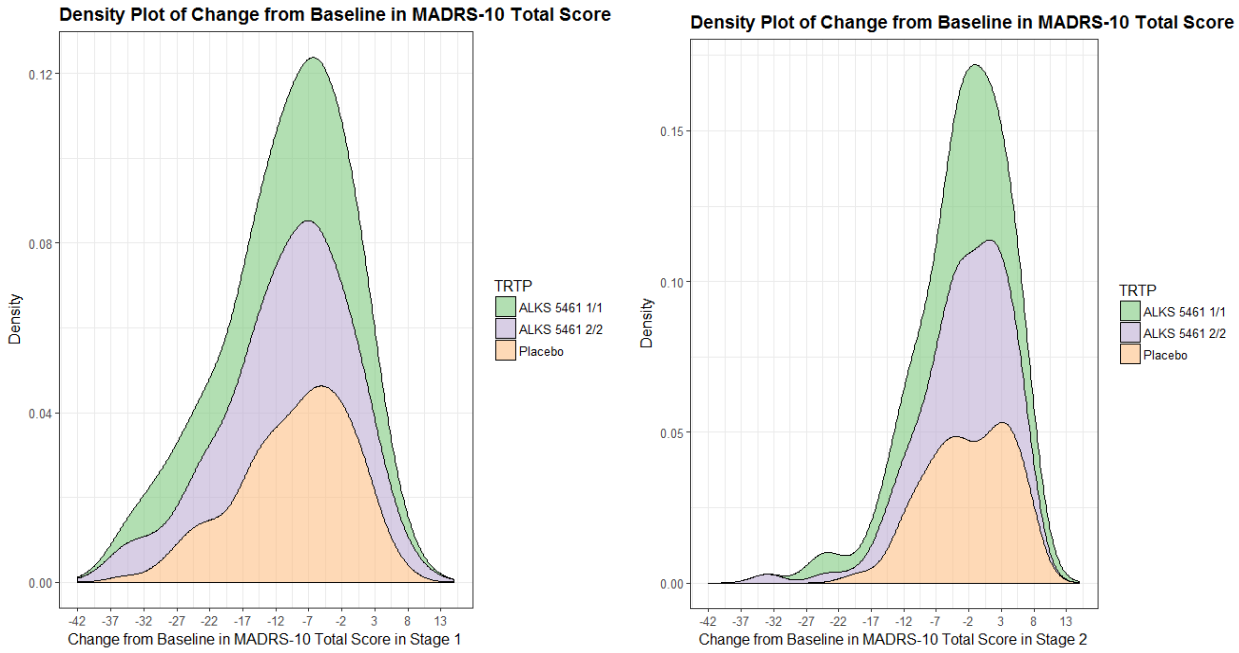
Table 25: Descriptive Statistics (Observed Cases): MADRS-10 Total Score at Baseline and End of Each Stage (Study 207)

Stage 1	Placebo N=273	ALKS 5461 1/1 N=62	ALKS 5461 2/2 N=63
Baseline			
N	273	62	63
Mean (SD)	31.7 (5.6)	31.8 (5.3)	31.8 (5.6)
Week 5			
N	257	55	49
Mean (SD)	22.4 (9.6)	20.8 (10.6)	20.5 (9.9)
Change from baseline to week 5			
Mean (SD)	-9.1 (8.56)	-10.9 (9.99)	-11.3 (10.08)

Stage 2	Placebo N=60	ALKS 5461 1/1 N=62	ALKS 5461 2/2 N=63
Baseline			
N	60	62	63
Mean (SD)	26.3 (7.6)	27.7 (7.2)	26.0 (6.5)
Week 6			
N	56	58	58
Mean (SD)	24.1 (8.73)	24.0 (10.30)	22.8 (9.0)
Change from baseline to week 6			
Mean (SD)	-2.1 (6.34)	-3.6 (7.64)	-3.2 (7.43)

Source: reviewer's results

Figure 14: Density Plots of Change from Baseline in MADRS-10 Total Score by Stage (Study 207)



Source: Reviewer's Result

Similar to study 202 (Reviewer's Note 1), as seen in **Table 26**, for study 207 the correlations between the baseline and end of treatment MADRS-10 total score in stage 1 are: 0.47, 0.35, 0.24 for placebo, ALKS 5461 1/1, and ALKS 5461 2/2 groups, respectively. In addition, the MADRS-10 total score correlations between stage 2 baseline (equivalent to end of stage 1) measurement) and end of stage 2 measurement, which appear similar across treatment groups, are 0.69, 0.67, 0.58 for placebo, ALKS 5461 1/1, and ALKS 5461 2/2 groups, respectively. There appear some differences in the correlations between the treatment arms.

Table 26: Pearson Correlation Coefficients between Visit-wise MADRS-10 Total Scores within Each Stage (Study 207)

Placebo

	Stage 1						Stage 2						
	Baseline	Week 1	Week 2	Week 3	Week 4	Week 5	Baseline	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Baseline	1.00 273	0.61 273	0.52 270	0.48 265	0.48 263	0.47 257	1.00 60	0.81 60	0.73 60	0.74 59	0.73 58	0.67 57	0.69 56
Week 1	0.61 273	1.00 273	0.78 270	0.73 265	0.68 263	0.65 257	0.81 60	1.00 60	0.79 60	0.82 59	0.82 58	0.78 57	0.76 56
Week 2	0.52 270	0.78 270	1.00 270	0.81 265	0.72 263	0.72 257	0.73 60	0.79 60	1.00 60	0.87 59	0.76 58	0.72 57	0.72 56
Week 3	0.48 265	0.73 265	0.81 265	1.00 265	0.84 263	0.81 257	0.74 59	0.82 59	0.87 59	1.00 59	0.88 58	0.82 57	0.80 56
Week 4	0.48 263	0.68 263	0.72 263	0.84 263	1.00 263	0.82 257	0.73 58	0.82 58	0.76 58	0.88 58	1.00 58	0.87 57	0.81 56
Week 5	0.47 257	0.65 257	0.72 257	0.81 257	0.82 257	1.00 257	0.67 57	0.78 57	0.72 57	0.82 57	0.87 57	1.00 57	0.91 56
Week 6	NA	NA	NA	NA	NA	NA	0.69 56	0.76 56	0.72 56	0.80 56	0.81 56	0.91 56	1.00 56

Top number in cell: correlation; bottom number in cell: sample size used to calculate the correlation.
NA: not available because there were only 5 weeks in stage 1.

ALKS 5461 1/1

	Stage 1						Stage 2						
	Baseline	Week 1	Week 2	Week 3	Week 4	Week 5	Baseline	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Baseline	1.00 62	0.42 62	0.57 59	0.42 56	0.46 56	0.35 55	1.00 62	0.71 62	0.62 61	0.66 59	0.58 59	0.65 59	0.67 58
Week 1	0.42 62	1.00 62	0.64 59	0.38 56	0.28 56	0.36 55	0.71 62	1.00 62	0.78 61	0.81 59	0.75 59	0.75 59	0.77 58
Week 2	0.57 59	0.64 59	1.00 59	0.80 56	0.77 56	0.72 55	0.62 61	0.78 61	1.00 61	0.89 59	0.83 59	0.79 59	0.85 58
Week 3	0.42 56	0.381 56	0.81 56	1.00 56	0.87 56	0.87 55	0.66 59	0.81 59	0.89 59	1.00 59	0.86 59	0.85 59	0.91 58
Week 4	0.46 56	0.28 56	0.77 56	0.87 56	1.00 56	0.85 55	0.58 59	0.75 59	0.83 59	0.86 59	1.00 59	0.84 59	0.87 58
Week 5	0.35 55	0.36 55	0.72 55	0.83 55	0.85 55	1.00 55	0.65 59	0.75 59	0.79 59	0.85 59	0.84 59	1.00 59	0.90 58
Week 6	NA	NA	NA	NA	NA	NA	0.67 58	0.77 58	0.85 58	0.91 58	0.87 58	0.90 58	1.00 58

ALKS 5461 2/2

	Stage 1						Stage 2						
	Baseline	Week 1	Week 2	Week 3	Week 4	Week 5	Baseline	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Baseline	1.00 63	0.52 63	0.55 59	0.29 53	0.42 49	0.24 49	1.00 63	0.58 63	0.57 61	0.57 59	0.58 58	0.52 58	0.58 58
Week 1	0.52 63	1.00 63	0.72 59	0.69 53	0.69 49	0.57 49	0.58 63	1.00 63	0.78 61	0.76 59	0.76 58	0.76 58	0.76 58
Week 2	0.55 59	0.72 59	1.00 59	0.82 53	0.80 49	0.70 49	0.57 61	0.78 61	1.00 61	0.89 59	0.78 58	0.80 58	0.85 58
Week 3	0.29 53	0.69 53	0.82 53	1.00 53	0.81 49	0.71 49	0.57 59	0.76 59	0.89 59	1.00 59	0.82 58	0.83 58	0.87 58
Week 4	0.42 49	0.69 49	0.80 49	0.81 49	1.00 49	0.82 49	0.58 58	0.76 58	0.78 58	0.82 58	1.00 58	0.88 58	0.86 58
Week 5	0.24 49	0.57 49	0.70 49	0.71 49	0.82 49	1.00 49	0.52 58	0.76 58	0.80 58	0.83 58	0.88 58	1.00 58	0.92 58
Week 6	NA	NA	NA	NA	NA	NA	0.58 58	0.76 58	0.85 58	0.87 58	0.86 58	0.92 58	1.00 58

Source: Reviewer's Result

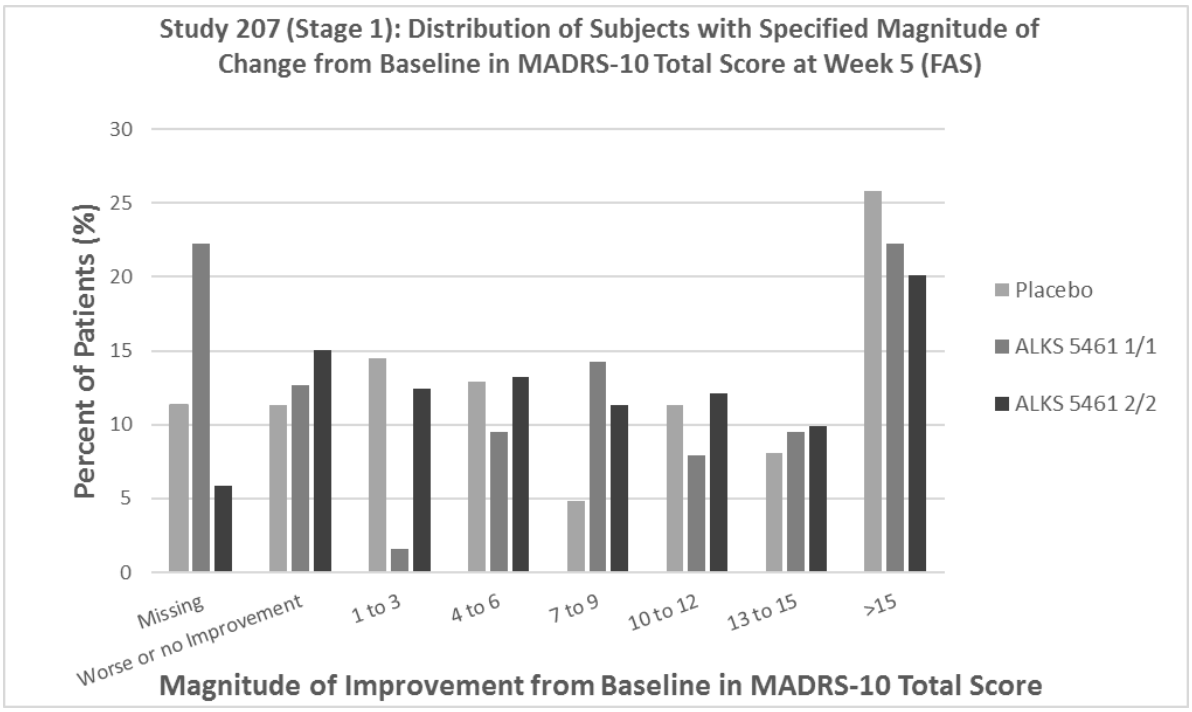
Reviewer's Note 6: This reviewer has included two figures to visualize the distribution of change in MADRS-10 total score and percent of improved subjects (Figure 15, Figure 16**).**

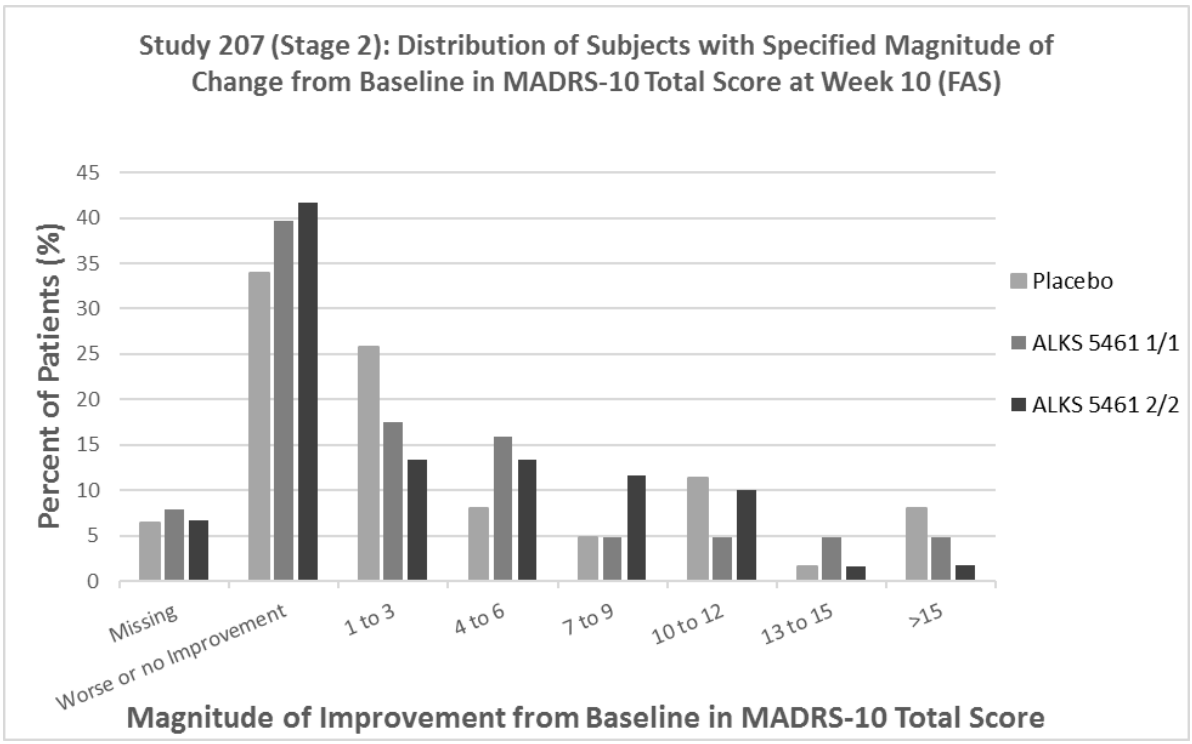
Figure 15 captured the distribution of change from baseline in MADRS-10 total score. In stage 1, with increasing magnitude of change (going from no improvement bins to the right of the plot), there was generally an increase in the proportion of improved subjects in the placebo and the ALKS 5461 2/2 groups. It is interesting to note that the placebo group had considerable proportion of subjects with 15 or more points of improvement.

In stage 2, a large proportion of subjects in placebo showed improvement in the range of 1 to 3 points change in MADRS-10 total score. The figure also reveals a large proportion of subjects who did not have an improvement or had worsened in each treatment group.

The empirical distribution in Figure 16 suggested that neither dose group showed considerable separation from placebo across each response cutoff, which agreed with the primary efficacy result.

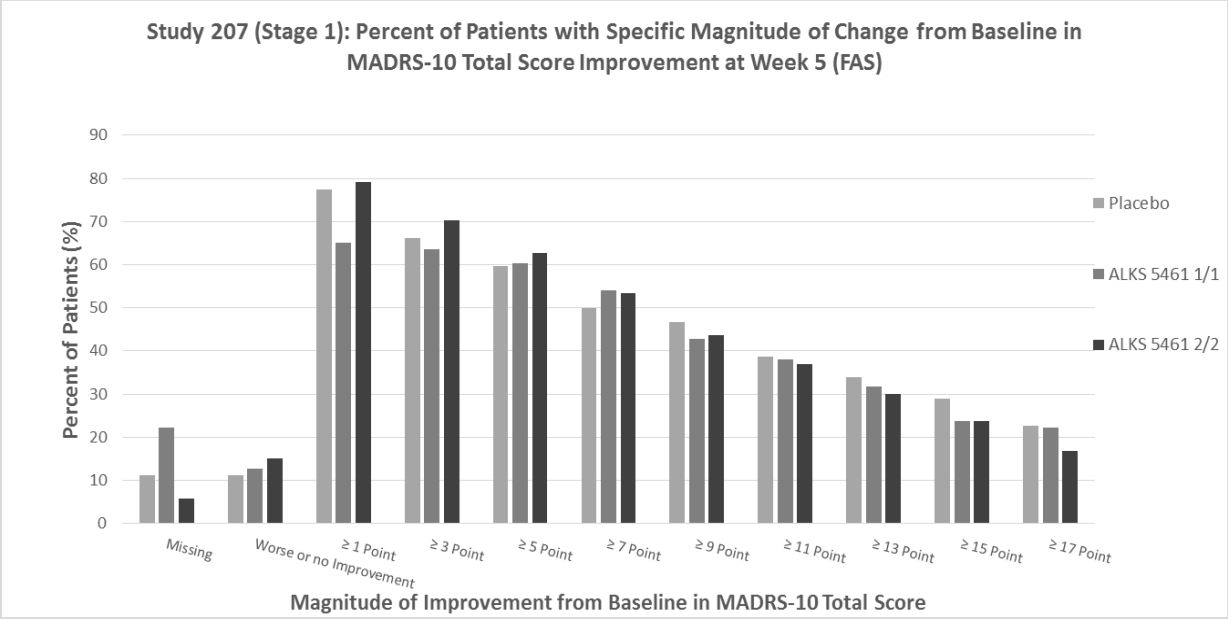
Figure 15: Percent of Patients with Specified Magnitude of MADRS-10 Total Score Improvement in Each Stage (Study 207; FAS)

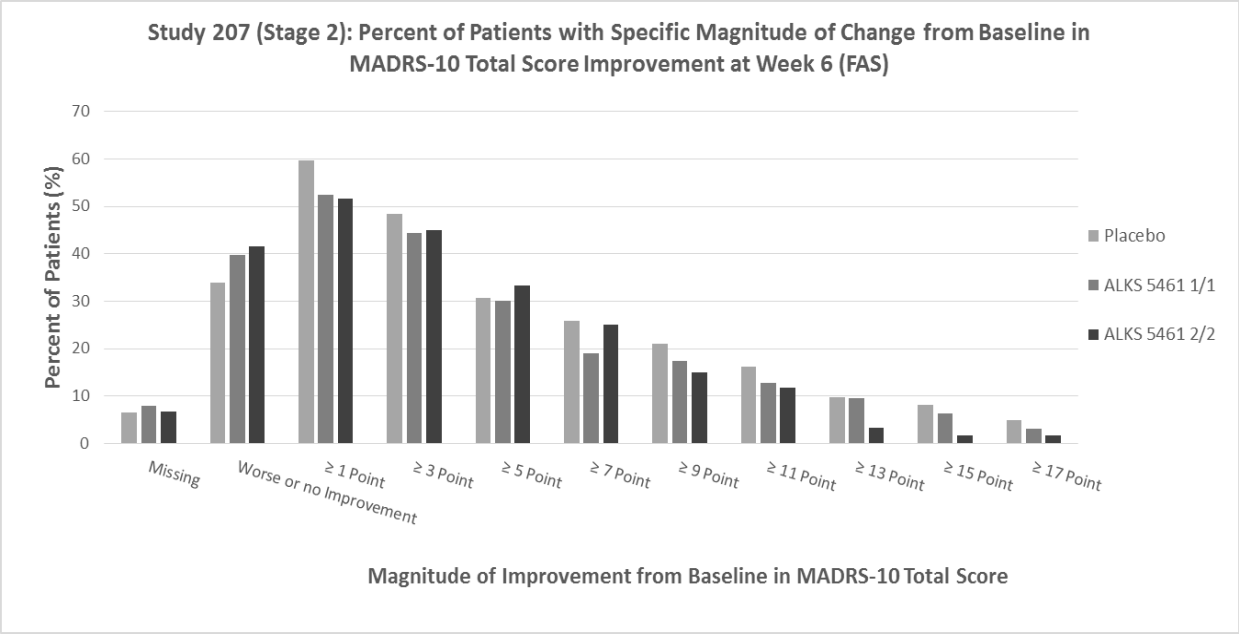




Source: Reviewer's Result

Figure 16: Percent Response (MADRS-10 Total Score) for Each Treatment Group in Each Stage by Response Threshold (Study 207)



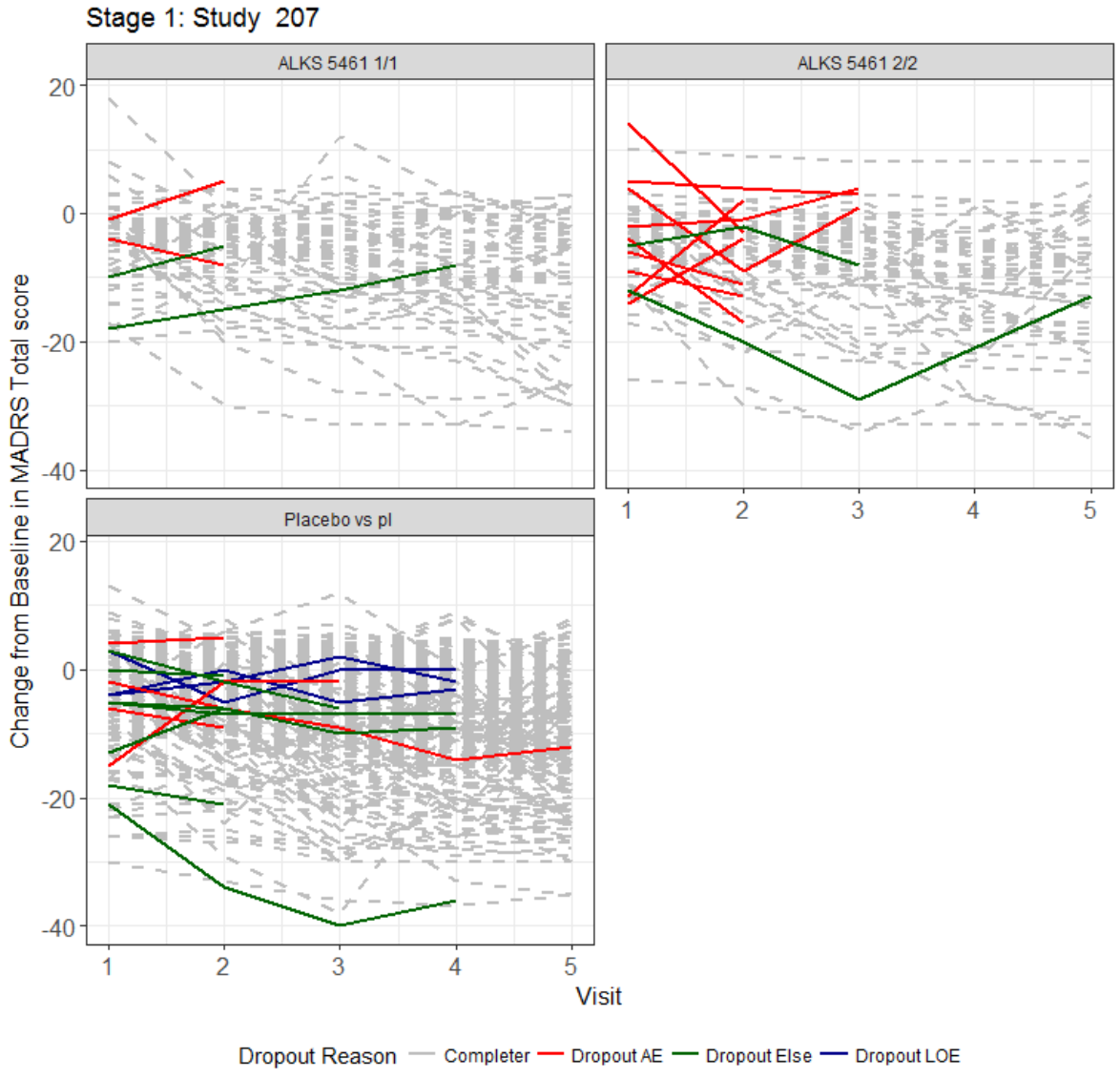


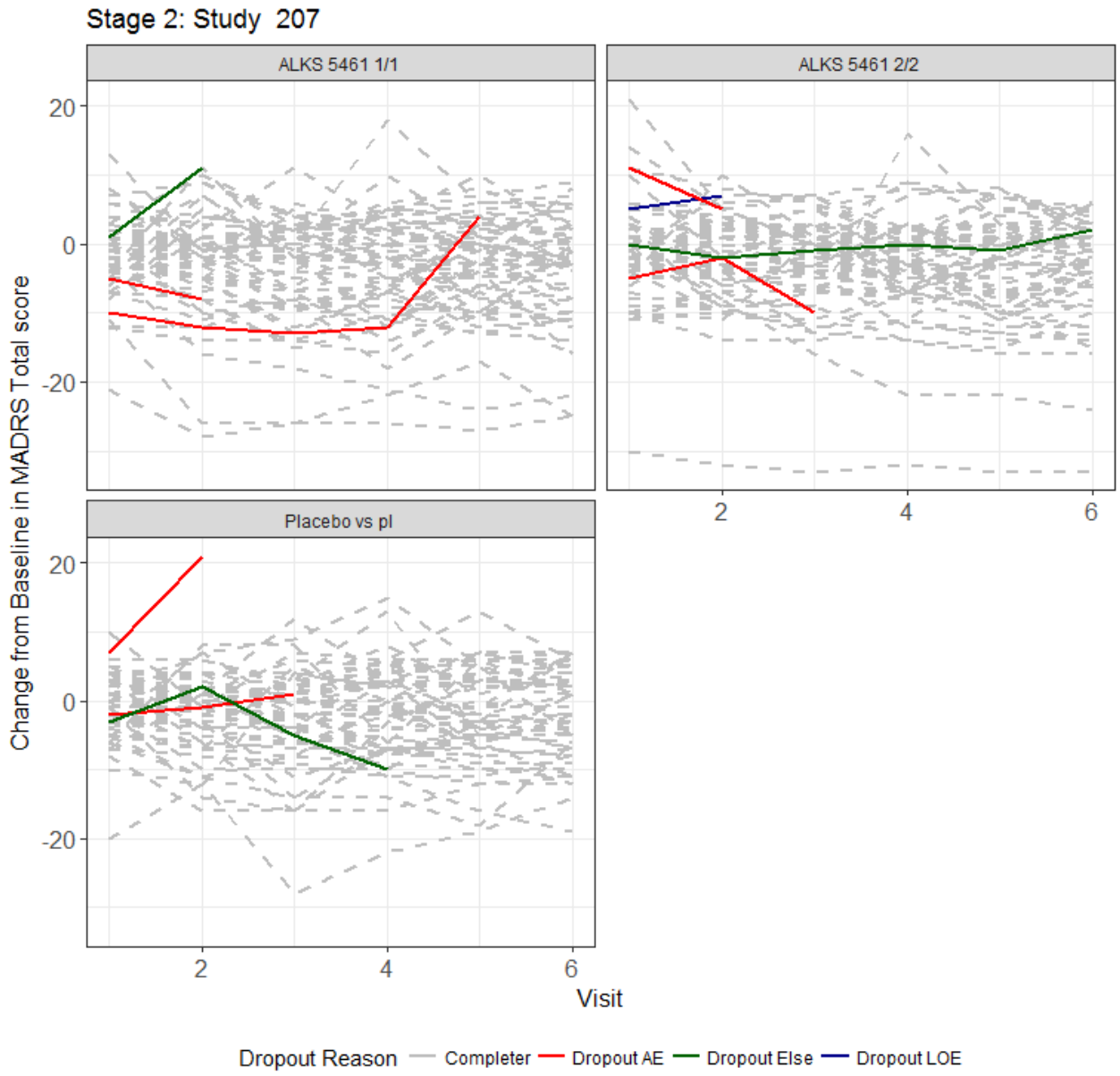
Source: Reviewer's Result

Reviewer's Note 7: Missing Data Profiles

The MMRM analysis assumes *missing at random* (MAR) as the underlying missing data mechanism. To visually explore missingness in study 207, the missing data patterns were explored according to dropout reasons: adverse events, lack of efficacy and combined reasons in stage 1 and stage 2 in Figure 17. In each treatment group, by and large, there was no noticeable signal suggesting that for almost all dropouts, before dropping out, their MADRS-10 scores patterns were very different than the completers. Thus, there was no discernible evidence that MAR assumption is totally unreasonable.

Figure 17: Missing Data Profiles Based on MADRS-10 Total Score in Each Stage (Study 207)





Source: Reviewer's Result

Alternative Statistical Review: Bootstrap Plan

Following several rounds of discussion within the statistics review team for this application, this reviewer implemented a bootstrap methodology in SPCD as an alternative to the existing analytic methodology. The plan assesses if the assumption of unequal variance-covariance structure across treatment groups could be invoked and would yield a different conclusion. If two treatment groups can have unequal variances, then the variance of the weighted estimator is not yet developed, e.g., when $\hat{\theta}_1$ and $\hat{\theta}_2$ are MMRM estimators (even if the MAR assumption holds). To account for potential differences in variances and covariances, we conducted a bootstrap inference plan.

Based on the bootstrap procedure (B=10,000 replicates), the bootstrap-based empirical 95% CIs yield a consistent conclusion (Table 27) as compared to the Applicant's normal-based 95% CIs (Table 28). The density plots of bootstrap treatment effects suggest a typical normal distribution (Figure 18). The treatment effect estimates in dose 2/2 based on the bootstrap sampling were slightly bigger than the Applicant's estimates (possibly due to repeated use of same patients). The standard errors were around 1.2 times (0.86 vs 1.01) those of Applicant's (assuming no difference in variances and covariances between treatment arms).

Table 27: Stage 1 and Stage 2 Weighted Bootstrap (B=10000) (Study 207)

Phase 1 and Phase 2 Combined (SPCD)	ALKS 5461 1/1	ALKS 5461 2/2
Change from Baseline: MADRS-10_{EOT}		
Phase 1 Weight = 0.5; Phase 2 Weight = 0.5		
LSMD vs Placebo (SE)	-1.4 (0.98)	-1.9 (1.1)
95% CI	(-3.3, 0.5)	(-4.2, 0.2)
Empirical 95% CI	(-3.4, 0.5)	(-4.2, 0.2)
P-value	0.188	0.080
Average Change from Week 3 to End of Treatment: MADRS-10_{AVG}		
Phase 1 Weight = 0.5; Phase 2 Weight = 0.5		
LSMD vs Placebo (SE)	-1.1 (0.86)	-2.2 (1.01)
95% CI	(-2.8, 0.6)	(-4.2, -0.2)
Empirical 95% CI	(-2.8, 0.6)	(-4.2, -0.3)
P-value	0.201	0.028

B: Number of bootstrap replicates; SE: standard error; CI; confidence interval.
Source: reviewer's results.

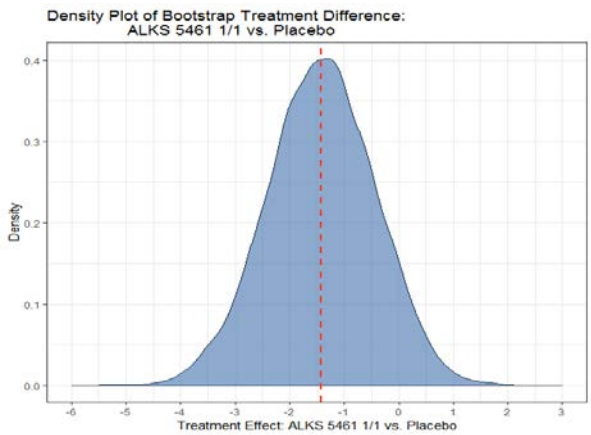
Table 28: Stage 1 and Stage 2 Combined Inference (Study 207)

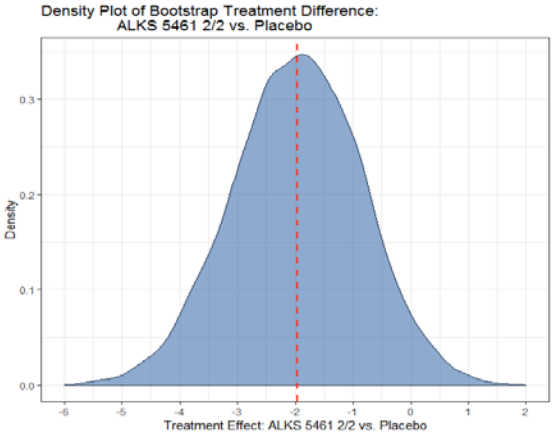
Phase 1 and Phase 2 Combined (SPCD)	ALKS 5461 1/1	ALKS 5461 2/2
Change from Baseline: MADRS-10_{EOT}		
Phase 1 Weight = 0.5; Phase 2 Weight = 0.5		
LSMD vs Placebo (SE)	-1.3 (0.95)	-1.7 (0.96)
95% CI	(-3.2, 0.5)	(-3.6, 0.2)
P-value	0.165	0.076
Average Change from Week 3 to End of Treatment: **MADRS-10_{AVG}**		
Phase 1 Weight = 0.5; Phase 2 Weight = 0.5		
LSMD vs Placebo (SE)	-0.9 (0.85)	-1.9 (0.86)
95% CI	(-2.6, 0.7)	(-3.6, -0.2)
P-value	0.277	0.026

Source: Applicant's Results

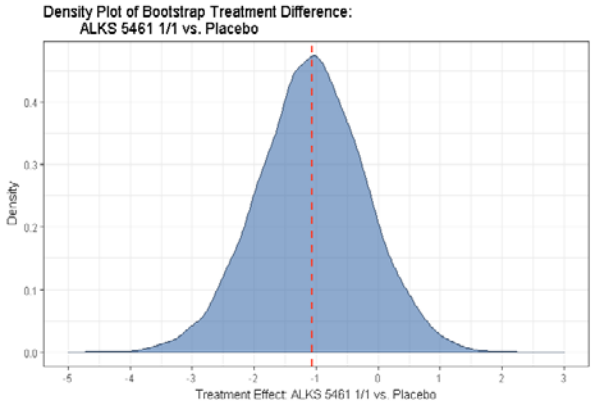
Figure 18: Density Plots of Bootstrap Treatment Difference (Study 207)

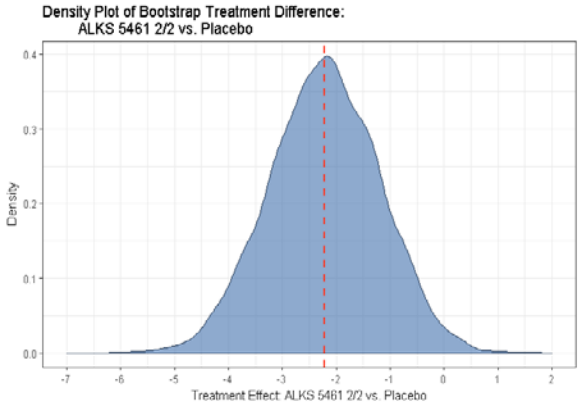
Change from Baseline to End of Each Stage (Week 5 for Stage 1; Week 6 for Stage 2): $MADRS\text{-}10_{EOT}$





Average from Week 3 to End of Each Stage (Week 5 for Stage 1; Week 6 for Stage 2): $MADRS\text{-}10_{AVG}$





Source: reviewer's results.

3.3 Evaluation of Safety

Safety evaluation was not conducted in this review.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

This section contains the reviewer's results of subgroup analyses. For exploratory purposes, the primary efficacy models were used to investigate treatment effect in the subgroups.

4.1 Gender, Race, Age, and Geographic Region

This reviewer conducted exploratory subgroup analysis by race, gender according to the respective primary efficacy analyses models specified. Subgroup analysis by age is not explored because most patients were less than 65 years old.

In study 202, whites had larger reduction numerically in HAM-D_{17} total scores than non-whites in ALKS 5461 2/2 compared to placebo. Similar trend is observed in which the overall treatment effect is numerically better in males than females in ALKS 5461 2/2.

In study 207, non-whites and males showed numerically improved treatment effect in ALKS 5461 2/2 compared to placebo.

Study 202

Table 29: Subgroup Analysis by Race: HAM-D_{17} Total Score (Study 202; FAS, MMRM)
(a) Stage 1

	Baseline		LS◊ Mean Change from Baseline			LS◊ Mean Difference from Placebo	
Treatment Group	N	Mean (SD*)	N	Mean	SE#	Mean	SE#
White							
Placebo	72	22.64 (4.04)	68	-6.71	0.69	--	--
ALKS 5461 2/2	13	20.54 (3.02)	12	-9.31	1.68	-2.61	1.83
ALKS 5461 8/8	14	20.79 (2.94)	9	-7.91	1.86	-1.20	1.99
Non-White							
Placebo	23	24.78 (4.27)	32	-8.08	1.47	--	--
ALKS 5461 2/2	7	26.57 (3.36)	5	-9.79	3.10	-1.71	3.43
ALKS 5461 8/8	6	23.67 (3.50)	5	-4.59	3.10	3.50	3.43

(b) Stage 2

	Baseline		LS◊ Mean Change from Baseline			LS◊ Mean Difference from Placebo	
Treatment Group	N	Mean (SD*)	N	Mean	SE#	Mean	SE#
White							
Placebo	19	16.89 (8.92)	19	-1.39	1.18	--	--
ALKS 5461 2/2	18	15.39 (4.55)	14	-5.67	1.37	-4.27	1.81
ALKS 5461 8/8	14	19.29 (5.45)	11	-3.74	1.51	-2.35	1.92
Non-White							
Placebo	1	24.0 (.)	1	-1.77	3.94	--	--
ALKS 5461 2/2	5	18.80 (9.63)	4	-4.07	1.95	-2.30	4.41
ALKS 5461 8/8	8	18.37 (5.88)	7	-2.67	1.47	-0.90	4.21

(c) Combined Stages

	LS◊ Mean Difference from Placebo	
Treatment Group	Mean	SE#
White		
ALKS 5461 2/2	-3.3	1.3
ALKS 5461 8/8	-1.7	1.4
Non-White		
ALKS 5461 2/2	-1.9	2.7
ALKS 5461 8/8	1.7	2.7

*Standard Deviation, #Standard Error, ◊Least Squares
Source: Reviewer's Results

Table 30: Subgroup Analysis by Gender: HAM-D_{17} Total Score (Study 202; FAS, MMRM)

(a) Stage 1

	Baseline		LS◊ Mean Change from Baseline			LS◊ Mean Difference from Placebo	
Treatment Group	N	Mean (SD*)	N	Mean	SE#	Mean	SE#
Male							
Placebo	30	23.2 (4.2)	29	-5.4	1.1	--	--
ALKS 5461 2/2	6	25.8 (3.4)	6	-11.6	2.5	-6.3	2.7
ALKS 5461 8/8	8	21.8 (3.5)	7	-3.4	2.3	2.0	2.5
Female							
Placebo	65	23.2 (4.2)	61	-7.9	0.8	--	--
ALKS 5461 2/2	14	21.3 (3.9)	11	-8.3	1.8	-0.4	2.0
ALKS 5461 8/8	12	21.6 (3.4)	7	-9.6	2.2	-1.7	2.3

(b) Stage 2

	Baseline		LS◊ Mean Change from Baseline			LS◊ Mean Difference from Placebo	
Treatment Group	N	Mean (SD*)	N	Mean	SE#	Mean	SE#
Male							
Placebo	5	20.8 (4.3)	5	-4.2	2.2	--	--
ALKS 5461 2/2	9	16.2 (7.2)	7	-7.0	1.9	-2.7	3.0
ALKS 5461 8/8	10	21.2 (6.0)	9	-4.4	1.7	-0.2	2.8
Female							
Placebo	15	16.1 (9.7)	15	-0.4	1.3	--	--
ALKS 5461 2/2	14	16.1 (5.2)	11	-4.2	1.5	-3.8	2.0
ALKS 5461 8/8	12	17.1 (4.4)	9	-2.5	1.6	-2.1	2.0

(c) Combined Stages

	LS◊ Mean Difference from Placebo	
Treatment Group	Mean	SE#
Male		
ALKS 5461 2/2	-4.9	2.0
ALKS 5461 8/8	1.1	1.9
Female		
ALKS 5461 2/2	-1.8	1.4
ALKS 5461 8/8	-1.9	1.6

*Standard Deviation, #Standard Error, ◊Least Squares

Source: Reviewer's Results

Study 207

Table 31: Subgroup Analysis by Race: MADRS-10 Total Score (Study 207; FAS, MMRM)
(a) Stage 1

	Baseline		LS◊ Mean Change from Baseline			LS◊ Mean Difference from Placebo	
Treatment Group	N	Mean (SD*)	N	Mean	SE#	Mean	SE#
White							
Placebo	202	31.3 (5.35)	191	-9.4	0.63	--	--
ALKS 5461 1/1	44	31.9 (5.06)	38	-12.3	1.39	-2.9	1.53
ALKS 5461 2/2	50	31.3 (5.46)	37	-10.8	1.37	-1.4	1.50
Non-White							
Placebo	71	33.0 (6.26)	66	-8.6	1.09	--	--
ALKS 5461 1/1	18	31.3 (5.88)	17	-5.7	2.16	2.8	2.42
ALKS 5461 2/2	13	33.6 (6.20)	12	-11.1	2.57	-2.5	2.79

(b) Stage 2

	Baseline		LS◊ Mean Change from Baseline			LS◊ Mean Difference from Placebo	
Treatment Group	N	Mean (SD*)	N	Mean	SE#	Mean	SE#
White							
Placebo	42	24.9 (5.98)	40	-2.3	1.13	--	--
ALKS 5461 1/1	49	27.3 (6.96)	45	-3.5	1.09	-1.2	1.57
ALKS 5461 2/2	46	25.0 (6.18)	41	-4.1	1.13	-1.7	1.59
Non-White							
Placebo	18	29.6 (9.86)	16	-0.6	1.83	--	--
ALKS 5461 1/1	13	29.2 (7.95)	13	-2.4	2.09	-1.9	2.76
ALKS 5461 2/2	17	28.5 (6.67)	17	-2.6	1.83	-2.0	2.58

(c) Combined Stages

	LS◊ Mean Difference from Placebo	
Treatment Group	Mean	SE#
White		
ALKS 5461 1/1	-2.0	1.1
ALKS 5461 2/2	-1.6	1.09
Non-White		
ALKS 5461 1/1	0.5	1.84
ALKS 5461 2/2	-2.3	1.90

*Standard Deviation, #Standard Error, ◊Least Squares
Source: Reviewer's Results

Table 32: Subgroup Analysis by Gender: MADRS-10 Total Score (Study 207; FAS, MMRM)

(a) Stage 1

	Baseline		LS◊ Mean Change from Baseline			LS◊ Mean Difference from Placebo	
Treatment Group	N	Mean (SD*)	N	Mean	SE#	Mean	SE#
Male							
Placebo	84	31.0 (6.05)	80	-7.7	0.98	--	--
ALKS 5461 1/1	20	33.0 (3.49)	19	-12.6	2.08	-4.9	2.30
ALKS 5461 2/2	21	30.1 (4.87)	17	-11.8	2.07	-4.1	2.29
Female							
Placebo	189	32.0 (5.43)	177	-9.9	0.66	--	--
ALKS 5461 1/1	42	31.2 (5.88)	36	-9.1	1.46	0.7	1.6
ALKS 5461 2/2	42	32.6 (5.87)	32	-10.4	1.52	-0.5	1.66

(b) Stage 2

	Baseline		LS◊ Mean Change from Baseline			LS◊ Mean Difference from Placebo	
Treatment Group	N	Mean (SD*)	N	Mean	SE#	Mean	SE#
Male							
Placebo	15	29.3 (9.16)	14	-2.5	1.89	--	--
ALKS 5461 1/1	22	27.6 (7.27)	22	-3.9	1.60	-1.3	2.47
ALKS 5461 2/2	21	26.4 (6.58)	19	-2.7	1.65	-0.1	2.52
Female							
Placebo	45	25.3 (6.81)	42	-1.7	1.11	--	--
ALKS 5461 1/1	40	27.8 (7.18)	36	-3.1	1.24	-1.4	1.67
ALKS 5461 2/2	42	25.7 (6.45)	39	-4.1	1.17	-2.4	1.61

(c) Combined Stages

	LS◊ Mean Difference from Placebo	
Treatment Group	Mean	SE#
Male		
ALKS 5461 1/1	-3.1	1.69
ALKS 5461 2/2	-2.1	1.70
Female		
ALKS 5461 1/1	-0.3	1.16
ALKS 5461 2/2	-1.5	1.16

*Standard Deviation, #Standard Error, ◊Least Squares

Source: Reviewer's Results

4.2 Other Special/Subgroup Populations: U.S. versus Non-US

Study 202 was conducted in the US; study 207 was conducted in US (37 sites) and non-US regions (11 sites). Table 33 below presents subgroup analysis by region for study 207.

In Study 207, the US population makes up the majority of subjects (82.5%). Numerically, the least squares treatment difference between placebo and ALKS 5461 2/2 is larger in the US.

Table 33: Subgroup Analysis by Region: MADRS-10 Total Score (Study 207; FAS, MMRM)

(a) Stage 1

	Baseline		LS◊ Mean Change from Baseline			LS◊ Mean Difference from Placebo	
Treatment Group	N	Mean (SD*)	N	Mean	SE#	Mean	SE#
US							
Placebo	222	32.0 (5.57)	211	-8.9	0.59	--	--
ALKS 5461 1/1	53	32.3 (5.16)	46	-10.8	1.25	-1.9	1.38
ALKS 5461 2/2	52	31.9 (5.61)	40	-9.9	1.35	-1.1	1.47
Non-US							
Placebo	51	30.2 (5.76)	46	-12.0	1.25	--	--
ALKS 5461 1/1	9	28.6 (4.98)	9	-9.4	2.89	2.6	3.14
ALKS 5461 2/2	11	31.2 (6.03)	9	-13.4	2.75	-1.5	3.02

(b) Stage 2

	Baseline		LS◊ Mean Change from Baseline			LS◊ Mean Difference from Placebo	
Treatment Group	N	Mean (SD*)	N	Mean	SE#	Mean	SE#
US							
Placebo	51	26.8 (7.91)	47	-1.5	1.04	--	--
ALKS 5461 1/1	55	28.2 (6.99)	52	-3.8	1.01	-2.3	1.44
ALKS 5461 2/2	52	26.8 (6.23)	49	-5.1	1.17	-3.7	1.56
Non-US							
Placebo	9	23.2 (4.55)	9	-3.6	2.38	--	--
ALKS 5461 1/1	7	24.3 (8.08)	6	-1.0	2.84	2.6	3.69
ALKS 5461 2/2	11	22.0 (6.26)	9	-1.7	2.33	1.9	3.29

(c) Combined Stages

Region	Treatment Group	LS◊ Mean Difference from Placebo	
		Mean	SE#
US	**ALKS 5461 1/1**	-2.1	1.00
	ALKS 5461 2/2	-2.4	1.07
Non-US	**ALKS 5461 1/1**	2.6	2.42
	ALKS 5461 2/2	0.2	2.23

*Standard Deviation, #Standard Error, ◊Least Squares

Source: Reviewer's Results

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

A 25 year FDA data on trials for major depression (1983-2008) revealed that the placebo response rate has increased and about 50% of the trials were negative[3]. During the same time period, the treatment effect (relative to placebo) has generally been decreasing.

The FDA encourages innovative designs, such as the SPCD, which could potentially alleviate problems associated with high placebo response in psychiatry trials. But the Division of Psychiatry Products (DPP) has not endorsed regulatory acceptance of this novel design under confirmatory trial settings, primarily because of the challenges to analyze the data collected from it. It also requires more understanding in spectrum of the pros and cons resulting from the use of this novel design.

Studies 202 and 207 utilized SPCD, where data from two stages were first evaluated separately, followed by combining the two estimates with a prespecified weight, to produce evidence in a form of p-value or confidence interval. This design is intended to enrich the primary analysis population, which would help reduce the high placebo response. The key statistical challenges are whether a fixed weighted combination of the two MMRM estimates of the two stages is statistically valid in terms of type I error control and correct estimation of the treatment effect.

The estimated treatment effects from each stage in studies 202 and 207 were extracted from MMRM analysis method. Study 202 had stage 1 weight fixed at 0.6, but weights were equally allocated in study 207. However, the variance of the weighted estimator is not yet developed in the context of MMRM.

To overcome this caveat, the statistical review team conducted an alternative statistical inference procedure based on bootstrap sampling. In this application, the bootstrap-based inference yields a consistent conclusion on MADRS-10 and MADRS-10_{AVG}, as compared to the Applicant's inference based on the assumption of equal variance-covariance matrix in the treatment arms.

The meta-analysis submitted by the Applicant is deemed exploratory in nature and cannot be a substitute for substantial evidence in the NDA submission. Some of the concerns with Applicant's meta-analysis are:

1. Combinability of SPCD and non-SPCD studies using meta-analysis does not appear to be sound due to incompatibility of populations in stage 1 and stage 2.
2. The rationale behind combining effects based on endpoints that were not specified in the SAP/protocol (such as MADRS-10_{AVG} in studies 202, 205, 206) is questionable.

[3] Ni A. Khin, Yeh-Fong Chen, Yang Yang, Peiling Yang and Thomas P. Laughren. Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications. *Journal of Psychiatry* 2011; 72 (4): 464-472.

3. There appears to be a large amount of heterogeneity between trials/studies, casting doubt of combining studies with the fixed effect model.

5.2 Collective Evidence

Study 202 was considered "proof of concept", not "proof of efficacy". Fixed doses ALKS 5461 2/2 and ALKS 5461 8/8 were explored without a prespecified plan to adjust for multiplicity when comparing them against placebo. Review of the statistical evidence in this NDA is solely based on the phase 3 study 207.

The two studies didn't show consistent differences between baseline populations characteristics in stage 1 and stage 2. However, it remains theoretically possible that under SPCD subjects in stage 2 have different baseline characteristics than the population in stage 1.

5.3 Conclusions and Recommendations

Evidence based on phase 3 study 207, ALKS 5461 2/2 dose seems to have a statistically significantly larger treatment effect relative to placebo on averaged MADRS-10 total score over several visits, when data from both stages are combined. Potential advantage of a primary endpoint averaged over visits is to improve statistical precision of treatment effect estimation resulting from multiplying the sample size. Whether the primary endpoint, MADRS-10_{AVG}, is acceptable, interpretable and can sensibly assess improvement over treatment resistant depression remains a clinical call. Typically, psychiatric trials have in the past evaluated improvement of symptoms using change from baseline in MADRS-10 or HAM-D_{17} total score to a fixed time point (instead of average over visits).

Results derived from the study 202 are considered exploratory or learning, without a pre-specified multiplicity adjustment procedure to control the overall type I error rate, which undermines statistical interpretability of the trial results. As such, the results generate a hypothesis of the particular 2/2 dose for test in Study 207.

To summarize, we defer to the clinical team regarding the following critical issues identified in the NDA review:

1. clinical relevance of the averaged MADRS-10 total score over a number of visits as a primary endpoint;
2. clinical relevance of combining estimated treatment effects over unequal durations between the two stages (5 weeks in stage 1 and 6 weeks in stage 2);
3. interpretability of results from SPCD: different patient populations (not only different baseline scores, placebo non-responders who stay through the end of stage 1 may be intrinsically different from the placebo dropouts or placebo responders.

6 APPENDICES

6.1 APPENDIX A Placebo Responders Entering Stage 2 (Study 202)

Table 34: Summary of Placebo Responder Population in Stage 2 (Study 202)

Category	Phase 1-Phase 2 Treatment Sequence [1]		
	PBO(R)-PBO	ALKS 5461 2/2-PBO	ALKS 5461 8/8-PBO
Subjects Entering Phase 2, N	23	16	14
Subjects Entering Phase 2 but not Dosed, N	0	0	0
Subjects in the Safety Population, N	23	17	13
Subjects in the FAS Population, N (%) [2]	0	0	0
Subjects in the PK Population, N (%) [3]	0	0	0
Subjects who Discontinued during Phase 2, N (%) [3]	2 (8.70)	1 (5.88)	0
Adverse Event	0	0	0
Withdrawal by Subject	1 (4.35)	1 (5.88)	0
Lost to Follow-up	1 (4.35)	0	0
Death	0	0	0
Lack of Efficacy	0	0	0
Non-compliance with Study Drug	0	0	0
Physician Decision	0	0	0
Protocol Violation	0	0	0

Category	Phase 1-Phase 2 Treatment Sequence [1]		
	PBO(R)-PBO	ALKS 5461 2/2-PBO	ALKS 5461 8/8-PBO
Subjects who Discontinued during Phase 2 (continued)			
Other	0	0	0
Pregnancy	0	0	0
Study Termination by Sponsor	0	0	0
Subjects who Completed Phase 2, N (%) [3, 4]	21 (91.30)	16 (94.12)	13 (100)
Subjects who Completed Study	21 (91.30)	16 (94.12)	12 (92.31)
Subject who did not Complete Study	0	0	1 (7.69)
Withdrawal by Subject	0	0	1 (7.69)

Variables	Phase 1-Phase 2 Treatment Sequence		
	PBO(R)-PBO (N=23)	ALKS 5461 2/2-PBO (N=17)	ALKS 5461 8/8-PBO (N=13)
Age (years)			
n	23	17	13
Mean (SD)	46.3 (12.1)	44.6 (11.7)	45.8 (11.0)
Median	49.0	47.0	47.0
Q1 - Q3	34.0 - 55.0	39.0 - 52.0	41.0 - 54.0
Min - Max	20 - 64	23 - 65	26 - 60
Gender, n (%)			
Female	18 (78.26)	11 (64.71)	6 (46.15)
Male	5 (21.74)	6 (35.29)	7 (53.85)
Primary Race, n (%)			
White	17 (73.91)	13 (76.47)	8 (61.54)
Black or African American	6 (26.09)	4 (23.53)	5 (38.46)
Native Hawaiian or other Pacific Islander	0	0	0
American Indian or Alaska Native	0	0	0
Asian	0	0	0

Variables	Phase 1-Phase 2 Treatment Sequence		
	PBO(R)-PBO (N=23)	ALKS 5461 2/2-PBO (N=17)	ALKS 5461 8/8-PBO (N=13)
Ethnicity, n (%)			
Not Hispanic or Latino	19 (82.61)	16 (94.12)	12 (92.31)
Hispanic or Latino	4 (17.39)	1 (5.88)	1 (7.69)
Duration of current major depressive episode (months)			
n	23	17	13
Mean (SD)	10.8 (12.8)	9.7 (5.7)	7.6 (6.5)
Median	6.0	8.0	5.0
Q1 - Q3	4.0 - 12.0	5.0 - 15.0	4.0 - 10.0
Min - Max	2 - 63	3 - 21	2 - 23
Lifetime number of major depressive episodes			
n	23	16	13
Mean (SD)	6.0 (4.4)	6.2 (5.6)	7.0 (8.2)
Median	4.0	4.0	5.0
Q1 - Q3	3.0 - 7.0	3.0 - 8.5	3.0 - 6.0
Min - Max	1 - 20	1 - 20	2 - 32

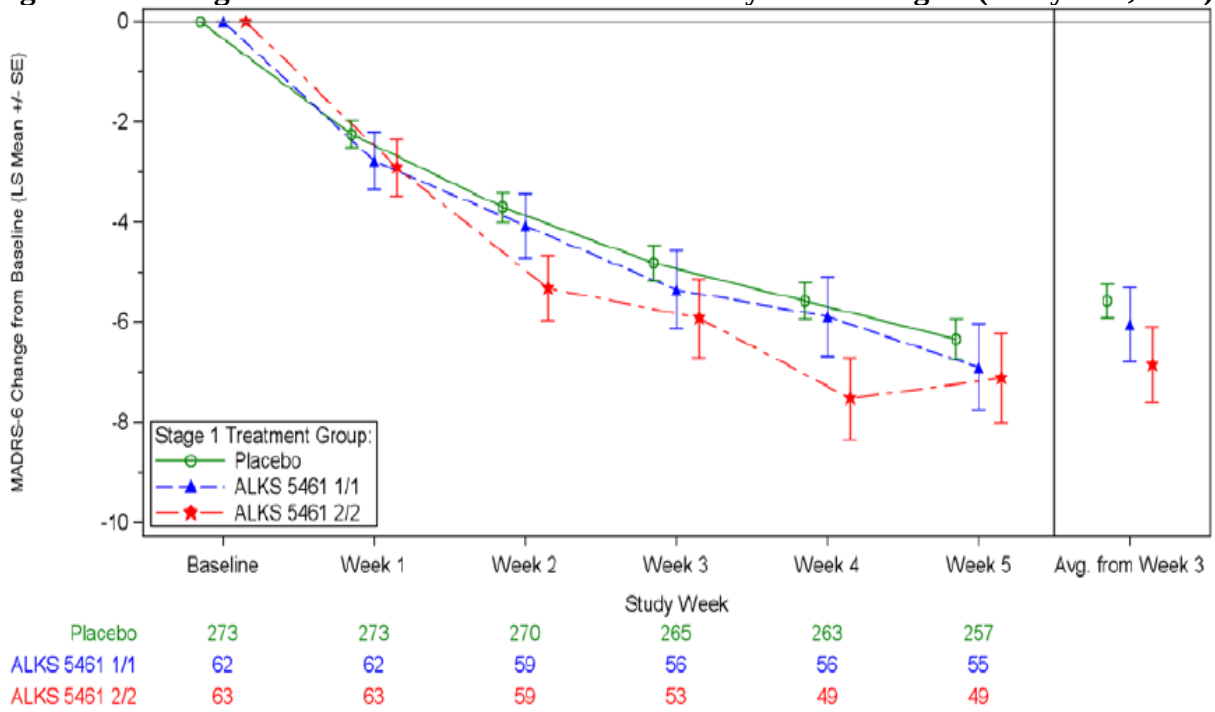
Variables	Phase 1-Phase 2 Treatment Sequence		
	PBO(R)-PBO (N=23)	ALKS 5461 2/2-PBO (N=17)	ALKS 5461 8/8-PBO (N=13)
Height (cm)			
n	23	17	13
Mean (SD)	164.17 (8.28)	167.74 (7.19)	170.88 (11.36)
Median	163.00	169.70	175.00
Q1 - Q3	158.80 - 167.60	163.80 - 172.00	160.00 - 176.00
Min - Max	148.8 - 185.0	157.0 - 182.9	155.0 - 188.0
Weight (kg)			
n	23	17	13
Mean (SD)	70.96 (12.95)	87.26 (19.25)	96.10 (14.85)
Median	73.30	83.20	95.50
Q1 - Q3	62.20 - 82.10	78.30 - 98.90	91.30 - 105.60
Min - Max	46.0 - 98.8	53.1 - 130.0	61.4 - 121.2
Body Mass Index (kg/m^2)			
n	23	17	13
Mean (SD)	26.34 (4.66)	30.80 (5.30)	33.14 (5.45)
Median	25.71	32.16	34.29
Q1 - Q3	22.36 - 29.43	26.70 - 33.57	31.86 - 36.04
Min - Max	19.4 - 36.5	19.8 - 38.9	19.1 - 39.5

Source: Table 14.1.1.2 of Applicant's Clinical Study Report (Page 17-18)

6.2 APPENDIX B MADRS-6_{AVG} Results (Study 207)

Figure 19 and Figure 20 display stage 1 and stage 2 plots of visit-wide change, as well as average change over time in Applicant's chosen endpoint, change from baseline in MADRS-6. In both stages, ALKS 5461 2/2 generally showed more separation from placebo than ALKS 5461 1/1. Figure 21 neatly summarizes the combined stage results in a forest plot.

Figure 19: Change from Baseline in MADRS-6 Score by Visit in Stage 1 (Study 207; FAS)



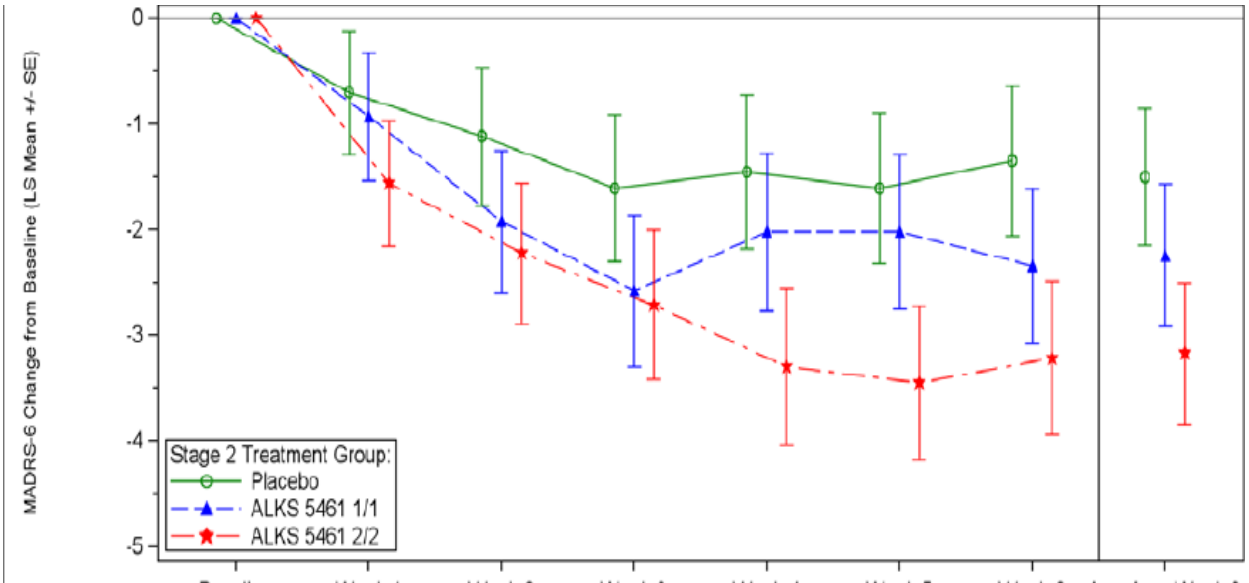
Placebo	273	273	270	265	263	257
ALKS 5461 1/1	62	62	59	56	56	55
ALKS 5461 2/2	63	63	59	53	49	49

Abbreviations: Avg=average; LS mean=least squares mean; MADRS=Montgomery-Åsberg Depression Rating Scale; SE=standard error

Notes: Avg from Week 3 (displayed in the line plot) includes data from Weeks 3, 4 and 5. Baseline is the randomization baseline at Visit 2. Subject counts in each treatment group per time point are provided in the Figure (eg, 273, 62, and 63).

Source: Figure 8 (Cross Reference Figure 14.2.1.1.1) of Applicant's Clinical Study Report (Page 102)

Figure 20: Change from Baseline in MADRS-6 Score by Visit in Stage 2 (Study 207; FAS)



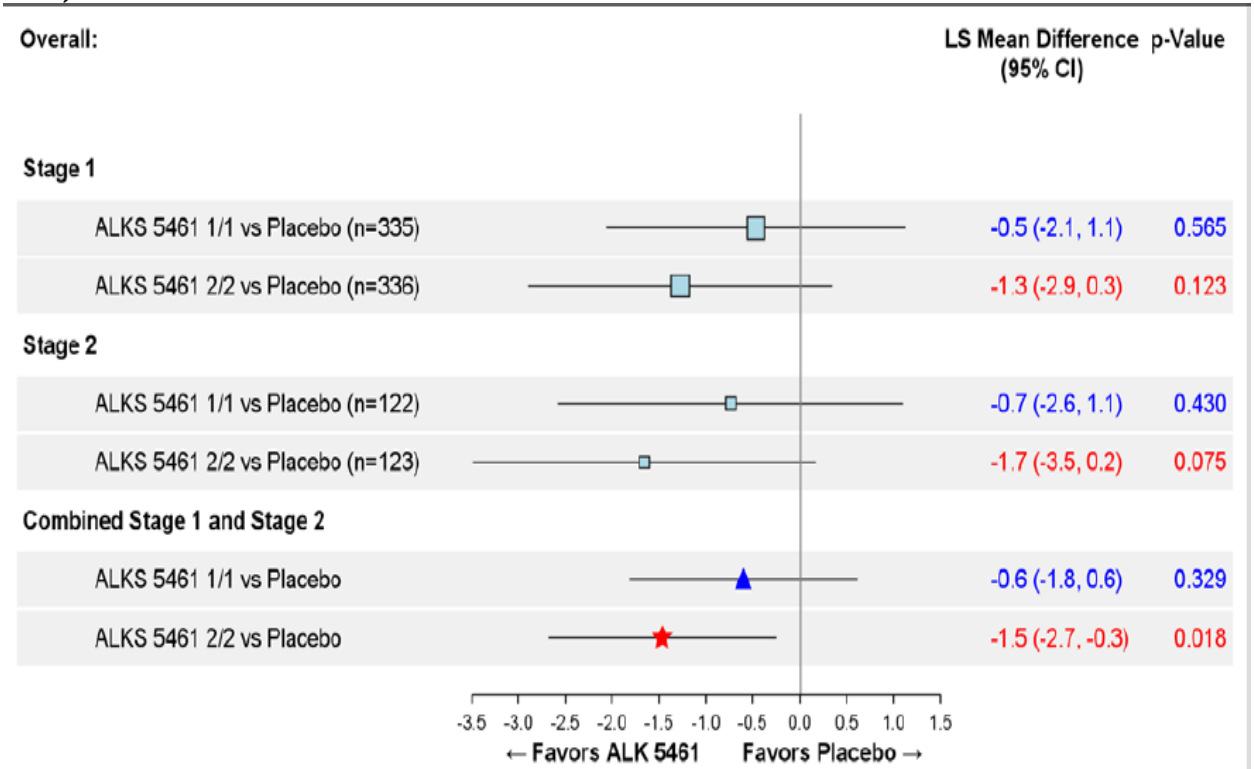
Placebo	60	60	60	59	58	57	56
ALKS 5461 1/1	62	62	61	59	59	59	58
ALKS 5461 2/2	63	63	61	59	58	58	58

Abbreviations: Avg=average; LS mean=least squares mean; MADRS=Montgomery-Åsberg Depression Rating Scale; SE=standard error

Notes: Avg from Week 3 (displayed in the line plot) includes data from Weeks 3, 4, 5 and 6.Baseline is the randomization baseline at Visit 2. Subject counts in each treatment group per time point are provided in the Figure (eg, 60, 62, and 63).

Source: Figure 9 (Cross Reference Figure 14.2.1.1.2) of Applicant's Clinical Study Report (Page 103)

Figure 21: Change from Baseline to End of Each Stage in MADRS-6_{AVG} Score (Study 207; FAS)

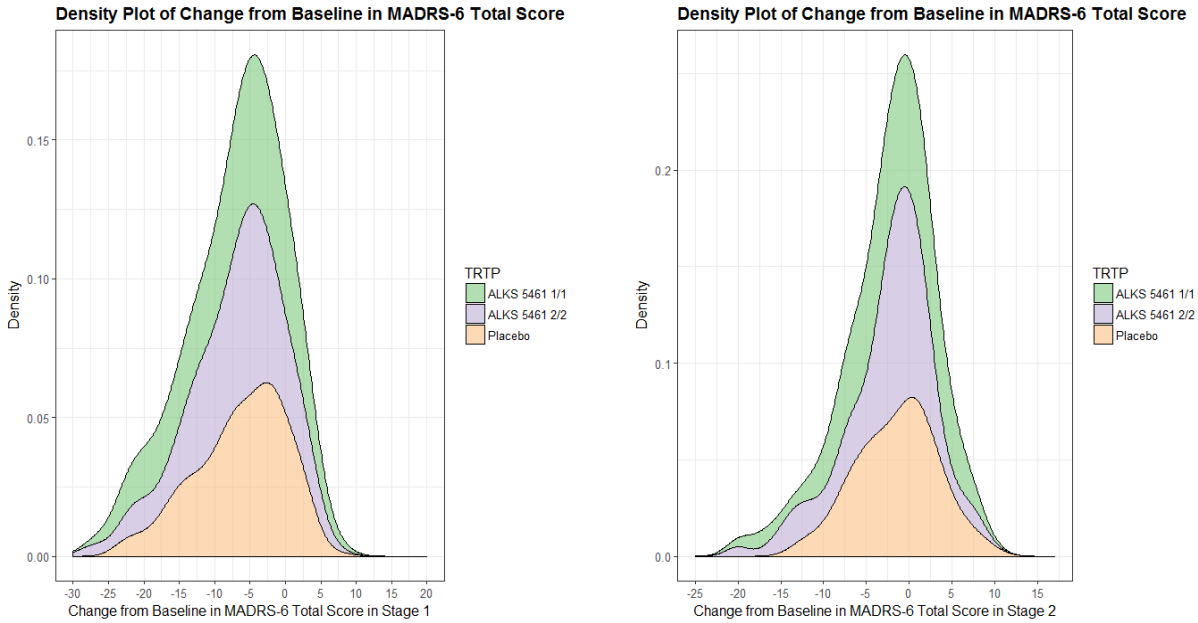


Abbreviations: CI=confidence interval; LS mean=least squares mean; MADRS=Montgomery-Åsberg Depression Rating Scale; MMRM=mixed-effects model for repeated measures

Note: The number of subjects (n) listed in Stage 1 and Stage 2 includes both the placebo group and ALKS treatment group subject counts. Analyses based on MMRM.

Source: Figure 10 (Cross Reference Figure 14.2.1.3) of Applicant's Clinical Study Report (Page 104)

Figure 22: Density Plots of Change from Baseline in MADRS-6_{AVG} Total Score by Stage (Study 207)



Source: Reviewer's results.

M E M O R A N D U M

Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research

Date: September 25, 2018

To: Mitch Mathis, MD, Director
Division of Psychiatry Products

Through: Dominic Chiapperino, PhD, Director
Silvia Calderon, PhD, Team leader
Martin Rusinowitz, MD, Team Leader
Controlled Substance Staff

From: Edward Hawkins, PhD, Pharmacologist
Katherine Bonson, PhD, Pharmacologist
Controlled Substance Staff

Subject: OPEN SESSION BACKGROUND DOCUMENT for buprenorphine and samidorphan sublingual tablet (NDA 210417). Prepared for the FDA Joint Meeting of the Psychopharmacologic Drugs Advisory Committee (PDAC) and the Drug Safety and Risk Management Advisory Committee (DSARM), November 1, 2018

I. BACKGROUND

BUP/SAM is a fixed-dose combination product containing buprenorphine hydrochloride and samidorphan maleate formulated as tablets for sublingual administration. The Sponsor claims that "efficacy of [BUP/SAM] is mediated through modulation of the opioid system." It is indicated for the adjunctive treatment of major depressive disorder (MDD) at a daily recommended dose of 2 mg/2 mg (buprenorphine/samidorphan). The proposed use of a combination product containing buprenorphine and samidorphan as an adjuvant in the treatment of MDD constitutes a novel indication for buprenorphine and for an opioid-containing product in general.

Buprenorphine is a mu opioid receptor partial agonist and a kappa opioid receptor antagonist. It has two FDA approved indications: 1) treatment of pain (as a single entity formulation), and 2) use in medication-assisted treatment (MAT) of opioid addiction (as a single entity formulation or in combination with naloxone). All buprenorphine-containing drug products are Schedule III under the Controlled Substances Act (CSA), as is buprenorphine substance[1]. This CSA

[1] Buprenorphine was placed in Schedule III of the CSA on November 7, 2002, (67 FR 62354)

placement was based on abuse potential assessments conducted when the drug products were under FDA review prior to NDA approval.

Buprenorphine has low oral bioavailability, which has led to its abuse primarily via the sublingual (SL), intranasal (IN), intramuscular (IM), or intravenous (IV) routes of administration. Buprenorphine is considered to have a lower abuse potential compared to opioids that are full mu opioid receptor agonists (Jasinski et al., 1978, Walsh et al., 1995, Comer et al., 2008), primarily because of its mu partial agonism, high affinity for the mu opioid receptor, and low dissociation rate from the mu opioid receptor. However, administration of buprenorphine through SL, IV, IM and IN routes produces positive subjective effects that can be comparable to the effects of full opioid agonists (Walsh et al., 1994, Walsh et al., 1995, Comer and Collins, 2002, Comer et al., 2002, Comer et al., 2005, Comer et al., 2010, Jones et al., 2015).

Human abuse potential studies conducted with buprenorphine demonstrate that its reinforcing properties, which are considered predictive of the abuse potential of a drug, can vary substantially depending on the route of administration and the state of opioid dependence in the individual taking the drug. These studies demonstrate that when buprenorphine is administered to *non-dependent opioid subjects* or to *subjects maintained on buprenorphine*, buprenorphine produces a subjective effect profile similar to that of an opioid agonist; whereas when taken by *opioid dependent* individuals maintained on a full mu-opioid agonist, the partial opioid agonist properties of buprenorphine mediate a decrease in the reinforcing effects of the drug (Jasinski et al., 1978, Walsh et al., 1994, Walsh et al., 1995, Comer et al., 2002, Comer et al., 2005, Comer et al., 2010)

Samidorphan is a new molecular entity that acts as a mu opioid receptor antagonist. It is currently controlled in Schedule II of the CSA because it is derived from thebaine-derived naltrexone[2].

The assessment of the abuse potential of a drug is based upon the comprehensive evaluation of data from chemistry, abuse-related animal behavioral studies, pharmacokinetic studies, human abuse potential (HAP) studies, abuse-related adverse events from clinical studies, and epidemiological data, if available.

The abuse potential of buprenorphine has been well characterized, and the next sections summarize the studies conducted by the Sponsor to characterize the abuse potential of samidorphan, and of the combination of buprenorphine and samidorphan as BUP/SAM.

---

[2] As defined in Title 21 CFR § 1308.12 (b) (1) (xvii), derivatives of the thebaine-derived naltrexone are not excluded from Schedule II controls under the CSA

II. MAIN FINDINGS OF THE ABUSE POTENTIAL EVALUATION

BUP/SAM is a drug product that contains buprenorphine and samidorphan. Since buprenorphine is already a Schedule III drug (either as a substance or when formulated in a drug product with or without naltrexone), an abuse potential assessment for buprenorphine alone was not necessary.

However, an abuse assessment was required for samidorphan alone to determine if its current placement as a Schedule II substance was scientifically justified, since it is purported by the Sponsor to be a mu opioid antagonist. Similarly, since the Sponsor asserts that the addition of samidorphan to buprenorphine has an abuse potential less than that of buprenorphine alone, an abuse assessment was required for BUP/SAM.

Abuse Potential of BUP/SAM

CSS has evaluated the abuse-related studies conducted with BUP/SAM and concludes that human abuse potential studies and analysis of abuse-related adverse events from clinical trials indicate that the drug product has an abuse potential that is lower than that of buprenorphine alone (see below). However, the full characterization of the abuse potential of the product should also take under consideration the chemistry of the product (e.g., in vitro extraction studies), and epidemiology data regarding the abuse and misuse of buprenorphine containing products currently on the market.

- An *in vitro* functional study (Study AP-5461-01) shows that the maximal stimulation of [^{35}S]GTPγS at mu opioid receptors produced by buprenorphine (60%) can be reduced by half (to 32%) when samidorphan is added at a 1:1 ratio. This shows that samidorphan is acting as a mu opioid antagonist.
- A HAP study (Study ALK5461-212) shows that in nondependent individuals with a history of opioid abuse:
    - BUP/SAM at therapeutic (2 mg/2 mg) and supratherapeutic (8 mg/8 mg, and 16 mg/16 mg) doses produces scores on positive subjective measures that are within the acceptable placebo range (Drug Liking and Overall Drug Liking) or are slightly outside the acceptable placebo range (Take Drug Again, High, and Good Drug Effects), but are all statistically significantly lower than those produced by buprenorphine alone at comparable doses (8 and 16 mg).
    - Although BUP/SAM produced some euphoria-related adverse events (19-21%), they occurred much less frequently than the adverse events produced by buprenorphine alone (48-67%)
    - The subjective responses and adverse events that were reported after administration of BUP/SAM in this study appear to be primarily the result of the effects of buprenorphine, since two HAP studies conducted with samidorphan alone (see below) did not produce drug effects outside of the acceptable placebo range. However, samidorphan did not reduce all abuse-related subjective responses or euphoria-related adverse effects to levels similar to those of placebo. This shows that BUP/SAM retains some clinically significant abuse potential.

- An assessment of abuse-related adverse events (AEs) from seven Phase 1 pharmacokinetic studies in healthy volunteers (Studies 33-008, 33 BUP-101, A108, A109, A110, A111, and 212) and 4 Phase 2 studies in MDD subjects (Studies 202, 205, 206, and 207) show that:
    - BUP/SAM produces fewer abuse-related AEs when compared to the same dose of buprenorphine alone. This shows that samidorphan is acting as a mu opioid antagonist.
    - BUP/SAM produced a similar incidence of euphoria-related AEs in single and multiple dose studies in healthy volunteers. This suggests that there are no clinically meaningful changes in abuse potential with continuing exposure to BUP/SAM.
    - BUP/SAM produced a higher incidence of euphoria-related AEs in studies with healthy volunteers compared to studies with MDD subjects. This is a common pattern in clinical studies, where individuals with a disease state do not report euphoria-related events as frequently as healthy individuals.

Abuse Potential of Samidorphan

CSS also evaluated the abuse-related studies conducted with samidorphan alone and concludes that it is a mu opioid antagonist with negligible abuse potential, based on the following:

- *In vitro* receptor binding and functional studies show that samidorphan is a mu opioid receptor antagonist that is approximately four-fold more potent than the mu and kappa opioid antagonist, naltrexone (Study 702-03231).
- Metabolism studies show that samidorphan produces two major metabolites, RDC-9986 (through *N*-dealkylation) and RDC-1066 (through *cis*-*N*-oxidation) (Study AM-0313-07).
- *In vitro* binding and functional activity studies show that RDC-9986 is an agonist at mu, kappa, and delta opioid receptors, while RDC-1066 is an agonist at the mu opioid receptor (Studies 702-03234 and 702-07219).
- In animals, RDC-9986 produces analgesic effects in a model of inflammatory pain (Study 702-07224), confirming that it has opioid agonist effects.
- In human pharmacokinetic studies, acute administration of BUP/SAM (2 mg/2 mg) produces similar Tmax values for both samidorphan and RDC-9986, although exposure levels for the metabolite are markedly lower than the parent. However, accumulation of RDC-9986 occurs after 7 days of samidorphan administration, such that RDC-9986 doubles and nearly equals exposure to the parent. The RDC-1066 metabolite does not appear to accumulate appreciably over 7 days (Study ALK5461-A111).
- Although these data show that the samidorphan metabolites are mu opioid agonists that accumulate with chronic samidorphan administration, the contribution of samidorphan's major metabolites to the overall pharmacodynamic effects of BUP/SAM is unknown.
- A general behavioral study in rats (Study AT-0313-11) shows that samidorphan produces some depressant-like effects such as reduced responsivity during handling, reduced locomotor activity, and reduced rearings.

- Analgesic studies in mice show that samidorphan does not reduce painful responses (Study 702-03368). However, it does attenuate the analgesic effects of the mu opioid receptor agonist, morphine (Study 702-03371), showing that samidorphan is acting as a mu opioid antagonist.
- Studies in beagle dogs and in monkeys (Studies 00081 and 01675) demonstrate that intramuscular samidorphan reversed the cardiac and respiratory depressant effects elicited by a constant intravenous infusion of fentanyl. This shows that samidorphan is acting as a mu opioid antagonist.
- A drug discrimination study (Study AP-0313-06) in rats shows that samidorphan did not generalize to the morphine discriminative cue. This means that animals did not recognize the effects of samidorphan as morphine-like.
- A self-administration study in rats shows that samidorphan produces some self-administration that is greater than saline but comparable to the mu opioid antagonist, naltrexone, and significantly lower than the mu opioid agonist, heroin (Study AP-0313-11).
- Two HAP studies conducted in nondependent individuals with a history of opioid abuse show that samidorphan alone did not produce positive subjective responses or euphoria-related AEs that were outside of the acceptable placebo range (Studies ALK33-012 and ALK33-B109). These data show that samidorphan does not produce rewarding effects in humans.

III. REFERENCES

Comer SD, Collins ED (2002) Self-administration of intravenous buprenorphine and the buprenorphine/naloxone combination by recently detoxified heroin abusers. J Pharmacol Exp Ther 303:695-703.

Comer SD, Collins ED, Fischman MW (2002) Intravenous buprenorphine self-administration by detoxified heroin abusers. J Pharmacol Exp Ther 301:266-276.

Comer SD, Sullivan MA, Vosburg SK, Manubay J, Amass L, Cooper ZD, Saccone P, Kleber HD (2010) Abuse liability of intravenous buprenorphine/naloxone and buprenorphine alone in buprenorphine-maintained intravenous heroin abusers. Addiction 105:709-718.

Comer SD, Sullivan MA, Walker EA (2005) Comparison of intravenous buprenorphine and methadone self-administration by recently detoxified heroin-dependent individuals. J Pharmacol Exp Ther 315:1320-1330.

Comer SD, Sullivan MA, Whittington RA, Vosburg SK, Kowalczyk WJ (2008) Abuse liability of prescription opioids compared to heroin in morphine-maintained heroin abusers. Neuropsychopharmacology 33:1179-1191.

Jasinski DR, Pevnick JS, Griffith JD (1978) Human pharmacology and abuse potential of the analgesic buprenorphine: a potential agent for treating narcotic addiction. Arch Gen Psychiatry 35:501-516.

Jones JD, Sullivan MA, Vosburg SK, Manubay JM, Mogali S, Metz V, Comer SD (2015) Abuse potential of intranasal buprenorphine versus buprenorphine/naloxone in buprenorphine-maintained heroin users. Addict Biol 20:784-798.

Walsh SL, Preston KL, Bigelow GE, Stitzer ML (1995) Acute administration of buprenorphine in humans: partial agonist and blockade effects. J Pharmacol Exp Ther 274:361-372.

Walsh SL, Preston KL, Stitzer ML, Cone EJ, Bigelow GE (1994) Clinical pharmacology of buprenorphine: ceiling effects at high doses. Clin Pharmacol Ther 55:569-580.

Department of Health and Human Services
Public Health Service
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Surveillance and Epidemiology Review (OSE)
Office of Pharmacovigilance and Epidemiology (OPE)

Epidemiology Review

Date:	October 5, 2018
Reviewers:	Jana McAninch, MD, MPH, MS, Senior Medical Epidemiologist Celeste Mallama, PhD, MPH, Epidemiology Reviewer Sara Karami, PhD, Epidemiology Data Analyst *Division of Epidemiology II*
Team Leader:	Tamra Meyer, PhD, MPH, Epidemiology Team Lead *Division of Epidemiology II*
Associate Office Director:	Judy Staffa, PhD, RPh, Associate Office Director for Public Health Initiatives *Office of Surveillance and Epidemiology*
Drug Name:	Buprenorphine/Samidorphan (ALKS 5461)
Subject:	Epidemiology Review of Use, Misuse and Abuse of Buprenorphine-Containing Products, and the Associations Between Depression, Pain, and Opioid Use Disorder

LIST OF ABBREVIATIONS:

AAPCC: American Association of Poison Control Centers

BUP-ONLY: Single-ingredient buprenorphine

BUP-NAL: Buprenorphine/naloxone combination

CDC: Centers for Disease Control and Prevention

CNS: Central Nervous System

DEPI: Division of Epidemiology II

ED: Emergency Department

FDA: Food and Drug Administration

IV: Intravenous

MAT: Medication Assisted Treatment

MG: Milligrams

NASEM: National Academies of Sciences, Engineering, and Medicine

NEISS-CADES: National Electronic Injury Surveillance System – Cooperative Adverse Drug Event Surveillance

NPDS: National Poison Data System

NSDUH: National Survey on Drug Use and Health

PCC: Poison Control Center

SUD: Substance Use Disorder

US: United States

WHO: World Health Organization

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Buprenorphine/samidorphan (BUP/SAM) is a fixed dose combination product, containing 2 mg of buprenorphine and 2 mg of samidorphan. Buprenorphine is a partial μ-opiate receptor agonist and a partial ƙ-opiate receptor antagonist, and samidorphan is a full μ-opiate receptor antagonist. BUP/SAM's proposed indication is as an adjunctive treatment for major depressive disorder, a novel indication for an opioid-containing product. As part of the US Food and Drug Administration's (FDA) commitment to addressing the current opioid crisis in the US, the Agency has adopted a comprehensive approach to evaluating the risk-benefit balance of opioid-containing products that includes consideration of efficacy and safety in patients when used as directed as well as potential risks to patients and others related to opioid misuse and abuse. The purpose of this epidemiologic review is to help inform this broader consideration of the risk-benefit balance for BUP/SAM by addressing the following topics: use, misuse, and abuse of currently marketed buprenorphine and buprenorphine-naloxone (bup-nal) products, and the complex relationships between depression, pain, and substance use disorders. The presentation of real-world buprenorphine abuse patterns is not intended as an indicator of expected abuse levels or patterns for BUP/SAM, were it to be approved and marketed. The goal is rather to provide contextual information for discussion of the risk-benefit balance of the product under review.

In 2017, approximately 13.5 million prescriptions were dispensed for buprenorphine-containing oral solid film and tablet products from US outpatient retail pharmacies. The large majority of these prescriptions were for bup-nal combination products indicated for the treatment of opioid dependence, as part of evidence-based medication assisted treatment (MAT) of opioid use disorder. National survey data show that among people reporting any past-year use of various opioids, buprenorphine has the second highest proportion reporting of misuse[a] of the drug. When examining overall prevalence of misuse in the population, however, buprenorphine is less frequently misused than more commonly prescribed opioids like hydrocodone and oxycodone. In 2016-2017, there were an estimated 26,360 emergency department visits in the US involving abuse of buprenorphine-containing products, the majority involving the more widely prescribed bup-nal combination products. More than 40% of abuse cases presenting to emergency departments involving single-ingredient buprenorphine (bup-only) and bup-nal products documented abuse through the injection route. Internal analyses of poison control center call data and review of published literature also indicate that abuse of both single-ingredient buprenorphine and bup-nal combination products is common and frequently occurs through non-oral routes, primarily injection. Comparing buprenorphine abuse patterns to those of other opioids is challenging due to the high-risk nature of the populations in which buprenorphine is used as part of MAT for opioid use disorder. Attempts to estimate the risk of addiction associated with buprenorphine is also complicated by its use predominantly in individuals with pre-existing opioid use disorders. Nonetheless, the observed levels of buprenorphine misuse and abuse, and the substantial proportion of bup-nal abuse cases involving the injection route provide real-world evidence of the drug's abuse potential and demonstrate the ability of motivated individuals to circumvent the intended abuse-deterrent properties of buprenorphine-antagonist combination products. These data also raise the concern about potential unknown harms associated with parenteral abuse of samidorphan.

---

[a] Data are from the National Survey on Drug Use and Health, which uses the following definition for misuse: *"use in any way not directed by a doctor, including use without a prescription of one's own; use in greater amounts, more often, or longer than told."*

There is substantial overlap between mental illness and substance use disorders, and complex relationships exist between depression, pain, and opioid use disorder. Pain conditions are highly prevalent in patients with depression, and individuals with depression are at elevated risk for opioid misuse, opioid use disorder, and opioid overdose. An evaluation of the risk-benefit balance of a buprenorphine-samidorphan combination product in real-world settings should consider the risk profile of the indicated patient population, as well as potential risks associated with concomitant medical and nonmedical use of opioids and other central nervous system (CNS) depressants.

1 INTRODUCTION

Buprenorphine/samidorphan (BUP/SAM) is a fixed dose combination product, containing 2 mg of buprenorphine and 2 mg of samidorphan. Buprenorphine is a partial μ-opiate receptor agonist and a partial ƙ-opiate receptor antagonist, and samidorphan is a full μ-opiate receptor antagonist. BUP/SAM's proposed indication is as an adjunctive treatment for major depressive disorder, a novel indication for an opioid-containing product. The vast majority of opioid products marketed in the US are indicated for the treatment of moderate to severe pain, and the increase in prescribing of opioid analgesics since the 1990s has contributed to the current crisis of opioid addiction and overdose, which is responsible for more than 350,000 deaths in the US since 1999.[1] The US Food and Drug Administration (FDA) has made addressing this public health crisis a top priority, while also recognizing the need for appropriate access to prescription opioids in patients with a legitimate medical need.[2]

At the request of FDA, in July 2017, the National Academies of Sciences, Engineering, and Medicine (NASEM) issued the report *Pain Management and the Opioid Epidemic: Balancing Societal and Individual Benefits and Risks of Prescription Opioid Use*.[3] The NASEM committee's charge was to help the FDA develop a framework for opioid review, approval, and monitoring that balances individual patient needs with the broader public health consequences of opioid misuse. The report suggests that such an approach to evaluation of risk-benefit should include assessing evidence of a product's potential for diversion and misuse and predicted risks to family members and society.

Buprenorphine is a Schedule III drug under the Controlled Substances Act and is marketed in the US in products indicated for treatment of pain as well as products indicated for treatment of opioid dependence as part of evidence-based medication assisted treatment (MAT) of opioid use disorders. Several buprenorphine MAT products are formulated in combination with naloxone, a μ-opiate receptor antagonist with low oral bioavailability intended to reduce the risk of abuse via non-oral routes, particularly injection. The μ-opiate receptor antagonist, samidorphan, is a new molecular entity. As such, our understanding of its potential to reduce misuse, abuse and related risks when combined with buprenorphine is based on premarket studies currently under review by the agency. These studies are described elsewhere in the briefing document.

The purpose of this epidemiologic review memo is to provide information that may help inform the evaluation of the risk-benefit balance of BUP/SAM for patients when used as directed while also considering potential harms to patients or others arising from misuse and abuse. This review addresses the following topics: use, misuse, and abuse of currently marketed buprenorphine-containing products, and the complex relationships between depression, pain, and substance use disorders. In the US, buprenorphine-containing products are used primarily for the treatment of opioid use disorders. Therefore, the population to which these drugs are prescribed has a very different risk profile from the intended population for BUP/SAM. Furthermore, no marketed buprenorphine product contains samidorphan, which is intended to reduce the μ-opiate effects associated with risk of misuse, abuse, dependence, addiction, and overdose. Therefore, the review of real-world buprenorphine abuse patterns is not intended as a prediction of expected abuse of

BUP/SAM, were it to be approved and marketed. The goal is rather to provide contextual information for discussion of the risk-benefit balance of the product under review.

2 REVIEW METHODS AND MATERIALS

2.1 DEFINITIONS

Terminology varies considerably in this area, depending on data sources being used. Unless otherwise specified, this review will use the following definitions:

> **Misuse:** the intentional therapeutic use of a drug product in an inappropriate way and specifically excludes the definition of abuse
>
> **Abuse:** the intentional, non-therapeutic use of a drug product or substance, even once, to achieve a desirable psychological or physiological effect
>
> **Opioid Use Disorder:** problematic pattern of opioid use ranging from mild to severe and characterized by impaired control of use, impaired social functioning, risky use, tolerance, and withdrawal
>
> **Opioid Addiction**: a chronic disease of brain reward, motivation, memory, and related circuitry; generally corresponds to moderate to severe opioid use disorder

2.2 DRUG UTILIZATION

This review presents data from a recent FDA analysis of IQVIA™ drug utilization data describing nationally projected estimates of the number of dispensed prescriptions in the outpatient retail setting for buprenorphine-containing products.[4]

2.3 NATIONAL SURVEY OF DRUG USE AND HEALTH (NSDUH)

NSDUH is an annual national survey sponsored by the Substance Abuse and Mental Health Services Administration *(*SAMHSA) designed to provide nationally representative estimates of illicit as well as prescription drug misuse in the general US population.[5] NSDUH uses a multistage probability sample design to provide representative state and country-level estimates for non-institutionalized residents of the US who are aged 12 years and above. Population subgroups not covered by the survey include individuals residing within institutional facilities (e.g., jails, nursing homes), as well as those without a permanent address (e.g., homeless individuals). The survey does sample people from noninstitutional group quarters like shelters, halfway houses, and college dormitories. The survey is conducted in a face-to-face manner, and during the year 2017, the interview response rate of 50% included 68,032 completed interviews. In a survey redesign implemented in 2015, NSDUH began to include more detailed data on use and misuse of specific prescription opioid analgesic subtypes.

We extracted estimates of past-year use and misuse of prescription opioids, overall and by subtype, from published detailed results tables from the most recent survey year (i.e., 2017).[6] NSDUH defines misuse of a drug as the following: *"use in any way not directed by a doctor, including use without a prescription of one's own; use in greater amounts, more often, or longer than told."*

2.4 NATIONAL ELECTRONIC INJURY SURVEILLANCE SYSTEM -- COOPERATIVE ADVERSE DRUG EVENT SURVEILLANCE (NEISS-CADES)

Cases and national estimates of the number of emergency department (ED) visits for drug-related adverse events were based on data from the NEISS-CADES project, a national stratified probability sample of approximately 60 hospitals with a minimum of six beds and a 24-hour ED in the US and its territories. The NEISS-CADES project, which has been described in detail elsewhere, is a joint effort of the Centers for Disease Control and Prevention (CDC), the US Consumer Product Safety Commission, and the FDA.[7-10] In brief, trained coders located at each participating hospital review clinical records of every ED visit to identify clinician-diagnosed drug related adverse events, to report up to four medications implicated in each adverse event, and to record narrative descriptions of the incident. Each NEISS-CADES case is assigned a sample weight derived from the inverse probability of selection, adjusted for nonresponse and post-stratified to adjust for the number of annual hospital ED visits.

NEISS-CADES has historically focused exclusively on ED visits due to use of medications for a therapeutic indication, or unintended medication exposures by young children. However, in 2016 NEISS-CADES surveillance activities were expanded to represent the full spectrum of pharmaceutical-related harm, encompassing ED visits resulting from abuse, self-harm, drugs used for unknown intent, and assault, in addition to therapeutic adverse drug events.

In NEISS-CADES, an abuse case is defined as a clinician diagnosis of abuse (for current ED visit) or documentation of recreational use (e.g., "to get high", "at a party", "crushing and snorting", "bought off street"). In addition to documented cases of abuse, we examined cases of overdose involving buprenorphine-containing products where the intent (i.e., misuse/abuse versus therapeutic use) was unclear. These cases include scenarios where, for example, the patient was altered or unresponsive and may not have been able or willing to provide a history or details of drug use.

Analyses of 2016-2017 NEISS-CADES data were conducted and provided to FDA by the CDC Division of Healthcare Quality Promotion. Cases were identified by searching for cases with the text string "BUPRENORPHINE" in any of the generic drug fields (generic1-4). National estimates of ED visits were calculated using the SURVEYMEANS procedure in SAS version 9.3 (SAS Institute, Inc, Cary, NC) to account for weighting and complex sample design. NEISS-CADES estimates based on <20 cases or total estimates <1200 for the study period are considered statistically unreliable and are not shown. Similarly, estimates with a coefficient of variation >30% may be statistically unreliable and are noted.

2.5 NATIONAL POISON DATA SYSTEM (NPDS)

Poison control centers (PCCs) are publicly available, confidential resources that supply exposure management resources and poison related public and professional education, with an emphasis on accurate data collection and coding. Information on case characteristics such as age, sex, intent, and substance(s) involved are recorded for all exposure-related calls. Follow-up calls are used to monitor case progress and consequent medical outcome.[11] NPDS, maintained by the American Association of Poison Control Centers (AAPCC), captures data on calls to US PCCs on a near real-time basis. Currently, AAPCC's 55 PCCs serve the entire US population, including all 50 states and US territories.

DEPI searched NPDS for intentional exposure calls coded as abuse involving oral solid (film and tablet) single-ingredient buprenorphine (bup-only) and bup-nal combination products from 2013-2017. Abuse cases were stratified by total (involving any number of products) and single-substance (involving only one product) exposure cases, and single-substance cases were further analyzed by route of exposure. NPDS defines abuse as "an exposure resulting from the intentional improper or incorrect use of a substance where the victim was likely attempting to gain a high, euphoric effect or some other psychotropic effect."

2.6 REVIEW OF PUBLISHED LITERATURE

Abuse of Buprenorphine Products in the Postmarket Setting

To provide information on patterns of abuse for marketed buprenorphine products formulated with and without a μ-opiate receptor antagonist, we updated a 2013 FDA epidemiologic review of the abuse of bup-only and bup-nal products.[12] We searched PubMed and EMBASE for relevant studies published from January 2013 to July 2018. The search was limited to these years because prior years were covered in the previous FDA literature review. We limited the search to reports of human subjects that were written in English. The following search strings were used.

PubMed:

((((((((((((buprenorphine abuse[Title/Abstract]) OR buprenorphine-naloxone abuse[Title/Abstract]) OR suboxone abuse[Title/Abstract]) OR buprenorphine diversion[Title/Abstract]) OR buprenorphine-naloxone diversion[Title/Abstract]) OR suboxone diversion[Title/Abstract]) OR illicit buprenorphine[Title/Abstract]) OR illicit buprenorphine-naloxone[Title/Abstract]) OR illicit suboxone[Title/Abstract]) OR buprenorphine inject*[Title/Abstract]) OR buprenorphine-naloxone inject*[Title/Abstract]) OR suboxone inject*[Title/Abstract]) OR buprenorphine inhal*[Title/Abstract]) OR buprenorphine-naloxone inhal*[Title/Abstract]) OR suboxone inhal*[Title/Abstract]

EMBASE:

('buprenorphine':ti,ab OR 'buprenorphine naloxone':ti,ab OR 'suboxone':ti,ab) AND ('abuse':ti,ab OR 'diversion':ti,ab) AND [humans]/lim AND [english]/lim AND [2013-2018]/py AND ('clinical article'/de OR 'clinical trial'/de OR 'controlled clinical trial'/de OR 'controlled study'/de OR 'cross-sectional study'/de OR 'drug surveillance program'/de OR 'major clinical study'/de OR 'observational study'/de OR 'retrospective study'/de) AND ('addiction'/de OR 'drug dependence'/de OR 'drug overdose'/de OR 'opiate addiction'/de OR 'withdrawal syndrome'/de)

Association between Depression and Opioid Use Disorder

To better understand the population for which BUP/SAM is intended, and the potential risks associated with its use in real-world settings, we conducted a general review of the epidemiologic literature and publicly-available population data on the associations between depression and opioid use disorder. The initial search included various combinations of the following terms: "risk," "depression," "mental," and "opioid use disorder." The search was restricted to human studies published in the English language. Additional relevant articles, including several discussing associations between depression, pain, and substance use disorders more broadly, were identified through reference lists of recent review articles and other publications identified in our initial search. This was not a systematic review of all the published literature on these topics.

3 REVIEW RESULTS

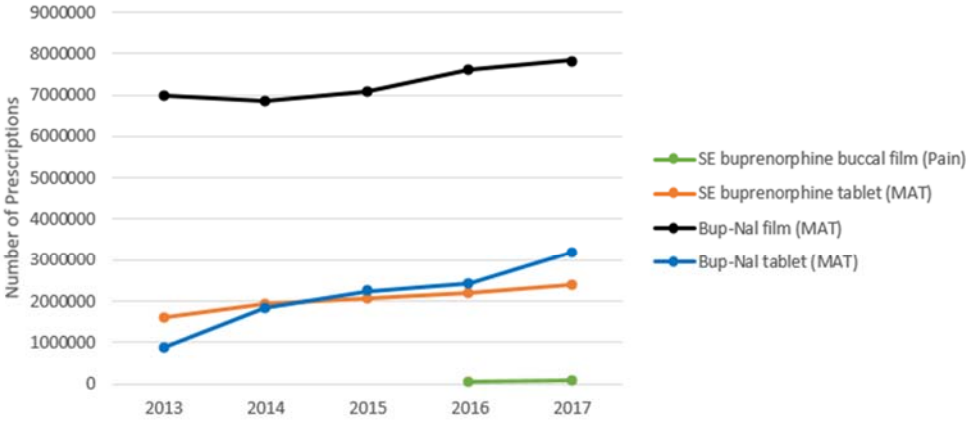
3.1 USE, MISUSE, AND ABUSE OF CURRENTLY MARKETED BUPRENORPHINE-CONTAINING PRODUCTS

3.1.1 Drug Utilization Data

As shown in **Figure 1**, national projections for the number of prescriptions dispensed from outpatient retail settings demonstrate that the vast majority of the market for oral solid buprenorphine-containing products consists of products indicated for opioid dependence as part of MAT as opposed to pain, and that dispensings of bup-nal combination products greatly

outnumber bup-only products. Buprenorphine transdermal patches, injectable, and implant products are not included in these data.

Figure 1. Nationally Estimated Number of Prescriptions for Buprenorphine-Containing Oral Solid Formulations Dispensed from US Outpatient Retail Pharmacies, 2013-2017



Source: IQVIA, National Prescription Audit™ (NPA). January 2013-December 2017. Data extracted March 2018.

Bup-Nal, buprenorphine-naloxone; MAT, medication-assisted treatment; SE, single entity
Source: Generated by DEPI reviewer from Table 3.2.1, FDA Briefing Document: Joint Meeting of Anesthetic and Analgesic Drug Products Advisory Committee and Drug Safety and Risk Management Advisory Committee, May 22, 2018. Available from https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/AnestheticAndAnalgesicDrugProductsAdvisoryCommittee/UCM608101.pdf, accessed 9/6/2018.
.

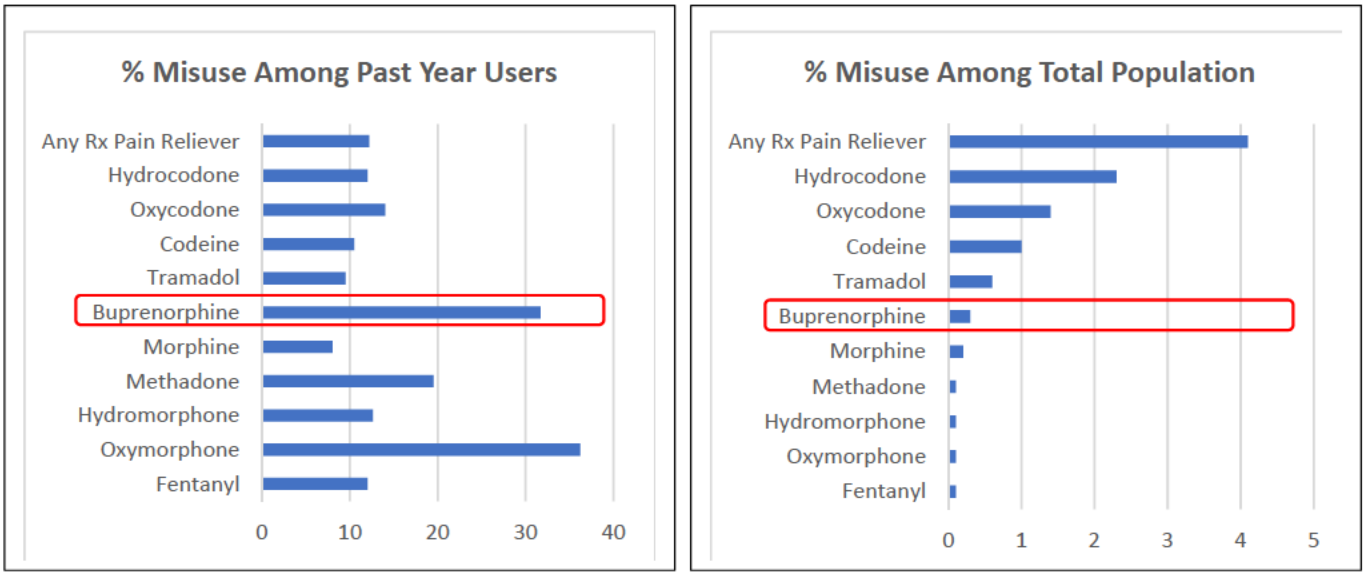
3.1.2 Misuse and Abuse

National Survey Data—NSDUH

Figure 2 shows the estimated past-year misuse of prescription pain relievers among non-institutionalized US residents aged 12 years and older, based on 2017 NSDUH data. Overall, slightly more than 4% of the survey population reported misusing prescription opioids in the past year. In this survey, misuse is defined as any use other than as directed by a healthcare provider.

Looking only at those reporting any past-year use of each opioid (including use as directed), buprenorphine has the second highest proportion reporting of misuse of the drug. When examining overall prevalence of misuse in the population, buprenorphine is less frequently misused than more commonly prescribed opioids like hydrocodone and oxycodone.

Figure 2. Past-year Misuse* of Prescription Pain Relievers in the US, Ages 12 Years and Older, NSDUH 2017



NSDUH: National Survey on Drug Use and Health; Rx, Prescription
*NSDUH defines misuse as any use other than as directed by a healthcare provider (i.e., misuse and abuse)
Source: Generated by DEPI reviewer from Center for Behavioral Health Statistics and Quality. (2018). *2017 National Survey on Drug Use and Health: Detailed Tables*. Substance Abuse and Mental Health Services Administration, Rockville, MD.

Emergency Department Visit Data—NEISS-CADES

Table 2 displays the number of cases and national estimates of total ED visits in the US for buprenorphine-related adverse events, based on 2016-2017 NEISS-CADES data. During this 2-year period, there were an estimated 48,924 ED visits where buprenorphine was an implicated drug, including 26,360 (54%) where the record specifically indicated abuse of buprenorphine. Of these estimated abuse-related visits, 21,329 involved abuse of bup-nal combination products. Approximately 45% of buprenorphine abuse cases were documented as involving the injection route, and this percentage was similar for bup-only and bup-nal abuse cases. In only 40% of the abuse cases was a buprenorphine product the only substance implicated. Co-implicated substances were most commonly alcohol or illicit drugs.

In addition to the abuse cases, there were an estimated 8,930 (95% CI 3,891-13,968) ED visits nationally in 2016-2017 for overdoses involving buprenorphine-containing products, where the intent was unknown. Most of these visits involved additional pharmaceutical drugs (76.6%) or concurrent alcohol or illicit drug use (55.4%).

Table 2. Emergency Department Visits for Buprenorphine-related Adverse Events, NEISS-CADES 2016-2017

	Any buprenorphine-containing product		Single-ingredient buprenorphine product*		Buprenorphine-naloxone combination product*	
	Cases	National estimate (95% CI)	Cases	National estimate (95% CI)	Cases	National estimate (95% CI)
Total cases implicating drug	574	48,924 (24,120-73,648)	109	**	469	39,209 (20,014-58,405)
Abuse cases (N)	**287**	**26,360 (10,542-42,178)**	**55**	**	**236**	**21,329 (9,166-33,492)**
Oral (N,%)	102	26.8% (13.7%-39.9%)	9	*	93	30.4% (15.5%-45.3%)
Nasal (N,%)	13	*	3	*	10	*
Injection (N,%)	112	45.4% (31.5%-59.2%)	24	44.7% (23.5%-65.8%)	90	45.6% (31.1%-60.1%)
Unknown/other route (N,%)	60	23.4% (14.5%-32.2%)	19	*	43	20.0% (14.1%-25.9%)
Only drug implicated, no concurrent alcohol or illicit substance use (N,%)	101	40.0% (30.2%-49.8%)	15	*	87	42.4% (32.0%-52.9%)
Multiple drugs implicated (N,%)						
With prescription opioid(s) other than buprenorphine	21	7.8% (10.5%-17.4%)	7	*	14	*
With other buprenorphine-containing product	4	*	4	*	4	*
With benzodiazepine(s)	78	25.1% (16.8%-33.5%)	18	*	61	22.5% (14.6%-30.3%)
With any pharmaceutical drugs***	117	39.1% (29.5%-48.7%)	29	53.1% (31.4%-74.7%)	89	34.6% (26.3%-42.9%)
With concurrent alcohol or Illicit drug use	139	44.8% (35.0%-54.6%)	28	50.3% (37.5%-63.0%)	113	43.5% (33.1%-53.9%)
Overdose cases with unknown intent (N)	**108**	**8,930 (3,891-13,968)**	**23**	***	**85**	**6,692 (3,047-10,337)**
Only drug implicated, no concurrent alcohol or illicit substance use (N,%)	10	*	1	*	9	*
Multiple drugs implicated (N,%)						
With prescription opioid(s) other than buprenorphine	8	*	1	*	7	*
With other buprenorphine-containing product	0	*	0	*	0	*
With benzodiazepine(s)	59	58.1% (50.3%-66.0%)	15	*	44	55.8% (46.5%-65.2%)
With any pharmaceutical drugs***	77	76.6% (67.4%-85.7%)	19	*	58	74.6% (64.5%-84.7%)
With concurrent alcohol or Illicit drug use	60	55.4% (46.8%-64.0%)	8	*	52	62.2% (54.4%-69.9%)

* indicates unstable estimates due to <20 cases
** indicates unstable estimates due to Coefficient of Variation >0.33
*** including drugs detected only by laboratory testing or identified only by drug class

Poison Control Center Data—NPDS

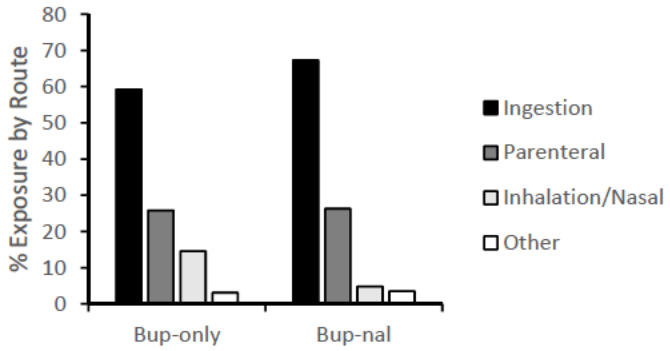
From 2013-2017, there were 652 abuse-related PCC calls in the US involving bup-only oral film and tablets and 2,581 involving bup-nal oral film and tablets (**Table 3**). Of these, there were 260 single-substance abuse-related calls involving bup-only and 1,124 involving bup-nal combination products. Of the single-substance abuse-related calls for bup-only, 25.8% were parenteral

exposures, and 14.6% were inhalation/nasal exposure. Of the single-substance abuse-related calls for bup-nal combination products, 26.3% were parenteral exposure, and 4.8% were inhalation/nasal (**Figure 3**).

Table 3. Abuse-related PCC calls for single-ingredient buprenorphine (bup-only) and buprenorphine-naloxone (bup-nal) oral film and tablet products, NPDS 2013-2017

	Bup-only	Bup-nal
Total exposures	652	2581
Single-substance exposures	260	1124

Figure 3. Percent of single-substance abuse-related PCC calls involving single-ingredient buprenorphine (bup-only) and buprenorphine-naloxone (bup-nal) oral film and tablet products, by route of exposure, NPDS 2013-2017*



* A single-substance exposure in NPDS may be associated with more than one route of exposure.

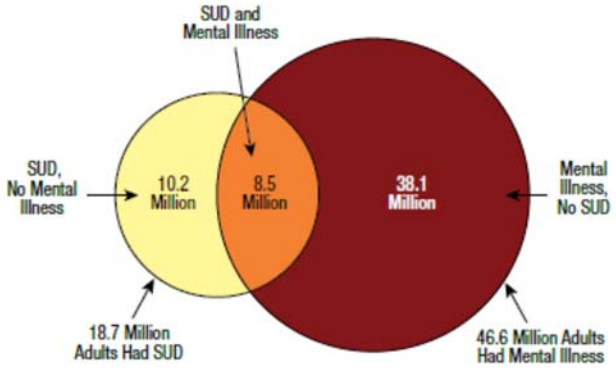
Review of Published Literature

A 2013 FDA review assessed abuse of bup-only compared to bup-nal combination products [12] and found that although some observational studies suggested the possibility of lower rates of intravenous (IV) abuse of bup-nal as compared to bup-only among opioid dependent individuals who were not primarily dependent on buprenorphine, the existing body of literature did not support the broad claim that bup-nal combinations were less abused and/or diverted than bup-only. Since this review, published studies from the US demonstrate the ongoing occurrence of diversion, misuse, and abuse of both bup-only and bup-nal combination products.[13-15] Although some studies suggest that naloxone may deter non-oral or injection abuse to some degree,[13,15] the data are conflicting, and abuse patterns vary by pharmaceutical dosage form (e.g., tablet, film, patch) and study population. In addition, the studies show that injection of both bup-only and bup-nal combination products is not uncommon among those abusing buprenorphine. In a sample of individuals entering substance use disorder treatment, Cicero et al. (2014) reported that 34.4% of those reporting past-month buprenorphine use "to get high" indicated that they had injected it in the month prior to treatment. Of those respondents who indicated past month injection of buprenorphine, 43.6% injected bup-nal tablets and 32.1% injected bup-nal oral film. When asked about circumventing barriers of the bup-nal combination by injection, participants reported a number of simple and easy methods to separate buprenorphine from naloxone, resulting in what they believed to be 'pure buprenorphine' for injection.[13] Motivations reported by respondents for buprenorphine use outside of a treatment program fell into two main categories: (1) use of buprenorphine for the express purpose of treating/preventing withdrawal sickness and (2) as a substitute to get high when other, more preferred drugs were unavailable.[13] Respondents also reported using buprenorphine to help treat anxiety, depression, or other psychological symptoms.[13]

3.2 ASSOCIATIONS BETWEEN DEPRESSION, PAIN, AND SUBSTANCE USE DISORDERS

The most recent year of NSDUH data indicate that there is considerable overlap between self-reported mental illness and substance use disorders (**Figure 4**).

Figure 4. Past Year Substance Use Disorder (SUD) and Mental Illness among Adults Aged 18 or Older, in Millions, NSDUH 2017



Source: Figure 54, Substance Abuse and Mental Health Services Administration. (2018). Key substance use and mental health indicators in the United States: Results from the 2016 National Survey on Drug Use and Health (HHS Publication No. SMA 17-5044, NSDUH Series H-52). Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. Retrieved from https://www. samhsa.gov/data/, 9/16/2018.

Individuals with depression and nonmedical use of opioids share a number of sociodemographic and behavioral characteristics, [16,17] and multiple population-based studies indicate that mood disorders, and specifically persistent depression and major depressive disorder, are associated with nonmedical use of prescription opioids.[16-19] NSDUH survey results show that opioid analgesics are commonly used in a manner other than as directed by a healthcare provider, often in attempts to manage pain, stress, anxiety, depressed mood, or insomnia, as well as sometimes to achieve a euphoric effect, or "high."[6] Cross-sectional studies have demonstrated associations between mood disorders and prescription opioid use disorder,[20-22] and longitudinal population studies indicate that individuals with baseline mental health diagnoses, including depression, are at increased risk of subsequently developing substance use disorders, including prescription opioid use disorder.[17,23,24]

Pain is highly prevalent in populations with depression. Studies suggest that more than 50% of patients with depression have co-morbid pain conditions and that the converse is also true; depression is also extremely common in patients with chronic pain.[25,26] A World Health Organization (WHO) prospective study of chronic pain in primary care found that "persistent pain at baseline predicted the onset of a psychological disorder with the same strength that a baseline psychological disorder predicted the onset of persistent pain."[26,27] Furthermore, the coexistence of depression and pain makes each condition more difficult to treat and results in worse outcomes than when either condition is present alone .[26,28,29] Chronic pain patients with depression have been found to be more likely than those without depression to progress to higher-dose long-term opioid use,[26,30,31] and use of psychotropic medications has also been associated with opioid misuse.[21,32] In studies examining chronic pain populations specifically, mood disorders, and unipolar major depression in particular, have been identified as risk factors for opioid misuse, although this is not a consistent finding across all studies.[33-36] Among patients receiving opioids, depression and other psychiatric disorders (other than substance use disorders) are also independent risk factors for fatal and non-fatal opioid overdose.[37-41]

4 DISCUSSION

4.1 USE, MISUSE, AND ABUSE OF CURRENTLY MARKETED BUPRENORPHINE-CONTAINING PRODUCTS

Buprenorphine is a partial opioid agonist recognized as having potential for abuse and addiction, and as such it is a scheduled drug under the Controlled Substances Act. The epidemiologic data indicate that buprenorphine products are indeed misused and abused, sometimes resulting in adverse health consequences, reflected in the many ED visits and PCC calls involving buprenorphine abuse. However, comparing the levels of abuse and related adverse outcomes associated with buprenorphine to other opioids is very difficult, because the majority of marketed buprenorphine products are dispensed as part of evidence-based treatment for opioid use disorder. Therefore, compared to other prescription opioid analgesics, buprenorphine is disproportionately dispensed to those with existing opioid use disorder at elevated risk for abusing opioids—including through injection or other non-oral routes—or diverting it within social networks of individuals with opioid use disorders. Furthermore, any estimation of the risk of addiction associated with buprenorphine is also complicated by its use predominantly in individuals with pre-existing opioid use disorders.

Both bup-only and bup-nal products are abused, with absolute numbers of ED and PCC cases being higher for the more widely dispensed bup-nal combination drugs. It remains unclear to what extent naloxone actually deters abuse through non-oral routes, but clearly, injection of bup-nal combination products continues to occur, as documented through ED visits, PCC calls, and self-report of individuals entering treatment for substance use disorders. Again, non-oral opioid use is prevalent in individuals with advanced opioid use disorders, and it is this population that is mostly likely to be exposed to existing buprenorphine MAT products. Therefore, observing relatively prevalent abuse of buprenorphine through injection and other non-oral routes is also not surprising and may largely reflect the drug's availability in these high-risk populations. Furthermore, samidorphan and naloxone cannot be equated with regard to their ability to deter misuse and abuse when combined with buprenorphine. Although both are μ-receptor antagonists, they are different molecular entities and differ in important ways, including their relative bioavailability through different routes, and their differing pharmacological properties are expected to have distinct impacts on abuse liability. More detailed discussion of the clinical and pharmacologic properties of samidorphan is found in other sections of the background package. Nonetheless, the high proportion of bup-nal abuse cases involving the injection route demonstrates the ability of motivated individuals to circumvent the intended abuse-deterrent properties of buprenorphine-antagonist combination products, and it also raises the concern about potential unknown harms associated with parenteral abuse of samidorphan.

4.2 ASSOCIATIONS BETWEEN DEPRESSION, PAIN, AND SUBSTANCE USE DISORDERS

Strong and complex associations exist between depression, pain, and substance use disorders. As stated in Sullivan 2018,[26] depression, chronic pain, and substance use disorders share many sociodemographic and behavioral risk factors, and the conditions themselves can interact with one another as cause, consequence, and promotor. The data suggest that a high proportion of individuals being treated for major depression either have a pain condition or will develop one in the future, and that both conditions are more difficult to manage when they occur together. Misuse of prescription opioids remains a prevalent problem in the general population, and the data also suggest that individuals with depression may be at elevated risk of misusing opioids, for example taking a family member's opioid medication, taking a higher dose or for a longer duration than recommended, or using it concomitantly with another central nervous system (CNS) depressant in an attempt to self-manage pain, anxiety, or insomnia. Among chronic pain

patients, individuals with depression and other psychiatric disorders are also at elevated risk of opioid overdose.

These complex interrelationships and the real-world use and misuse of opioid analgesics in patients with depression should be considered when evaluating the overall balance of risks and benefits of BUP/SAM. This is a novel opioid combination product with pharmacological properties that differ from opioid products with which patients and prescribers are likely to be familiar. Consider the following hypothetical scenario of a patient taking BUP/SAM as adjunct therapy for major depression who develops back pain. The patient might be prescribed an opioid analgesic by a primary care or emergency department provider, or might even obtain opioids from a friend or family member. If BUP/SAM reduces the analgesic efficacy of the opioid, the patient may increase the dose in an attempt to achieve adequate pain control. However, if the μ-opioid antagonist effect diminishes or the patient misses doses or discontinues BUP/SAM while taking this high dose of opioid, he or she could be at increased risk for respiratory depression and overdose from the now unopposed full opioid agonist, particularly if also using benzodiazepines or other CNS depressants. This hypothetical example illustrates the potentially complex drug interactions that could occur in the intended patient population, which may have an elevated probability of both use and misuse of opioid analgesics.

5 CONCLUSIONS

A comprehensive evaluation of the risk-benefit balance of an opioid-containing product includes consideration of potential harms related to misuse and abuse by the patient or others. Abuse of both bup-only and bup-nal combination products is common and occurs through both oral and non-oral routes, including injection; however, comparison of buprenorphine abuse patterns with those of other opioids is challenging due to the high-risk nature of the populations in which buprenorphine is used as part of evidence-based treatment for opioid use disorder. There is substantial overlap between mental illness and substance use disorders, and complex relationships exist between depression, pain, and opioid use disorder. Pain conditions are highly prevalent in patients with depression, and individuals with depression are at elevated risk for opioid misuse, opioid use disorder, and overdose. An evaluation of the risk-benefit balance of a buprenorphine-samidorphan combination product should consider the risk profile of the indicated patient population in real world settings, as well as potential risks associated with concomitant medical and nonmedical use of opioids and other CNS depressants.

6 APPENDIX

NPDS Search Parameters- Oral Solid Buprenorphine Products	
Products included in query	*Bup-only* Belbuca, Subutex, Buprenorphine sublingual generics *Bup-nal* Bunavail, Suboxone, Zubsolv, Buprenorphine-Naloxone generics
Month/year of query	7/2018
Call Type	Exposure
Date range for query	1/1/2013-12/31/2017
Case status	Closed
Species	Human
Exposure Reason	Intentional - Abuse

7 REFERENCES

1. Centers for Disease Control and Prevention. Opioid Overdose, Opioid Basics, Understanding the Epidemic. https://www.cdc.gov/drugoverdose/epidemic/index.html. Accessed 9/19/2018, 2018.
2. Statement by FDA Commissioner Scott Gottlieb, M.D., on balancing access to appropriate treatment for patients with chronic and end-of-life pain with need to take steps to stem misuse and abuse of opioids. 2018; https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm612779.htm.
3. National Academies of Sciences, Engineering and Medicine. Pain Management and the Opioid Epidemic: Balancing Societal and Individual Benefits and Risks of Prescription Opioid Use. In: Phillips JK, Ford MA, Bonnie RJ, eds. Washington (DC): National Academies Press (US). Copyright 2017 by the National Academy of Sciences. All rights reserved.; 2017.
4. FDA Briefing Document: Joint Meeting of Anesthetic and Analgesic Drug Products Advisory Committee and Drug Safety and Risk Management Advisory Committee, May 22, 2018. https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/AnestheticAndAnalgesicDrugProductsAdvisoryCommittee/UCM608101.pdf.
5. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: Results from the 2017 National Survey on Drug Use and Health (HHS Publication No. SMA 18-5068, NSDUH series H-53). 2018; https://www.samhsa.gov/data/.
6. Substance Abuse and Mental Health Services Administration. 2017 National Survey on Drug Use and Health: Detailed Tables. Center for Behavioral Health Statistics and Quality, Rockville, MD
7. Shehab N, Lovegrove MC, Geller AI, Rose KO, Weidle NJ, Budnitz DS. US Emergency Department Visits for Outpatient Adverse Drug Events, 2013-2014. *Jama.* Nov 22 2016;316(20):2115-2125.
8. Jhung MA, Budnitz DS, Mendelsohn AB, Weidenbach KN, Nelson TD, Pollock DA. Evaluation and overview of the National Electronic Injury Surveillance System-Cooperative Adverse Drug Event Surveillance Project (NEISS-CADES). *Medical care.* Oct 2007;45(10 Supl 2):S96-102.
9. Budnitz DS, Pollock DA, Weidenbach KN, Mendelsohn AB, Schroeder TJ, Annest JL. National surveillance of emergency department visits for outpatient adverse drug events. *Jama.* Oct 18 2006;296(15):1858-1866.
10. Schroeder TJ AK. National Electronic Injury Suveillance System (NEISS) sample design and implementation from 1997 to present. . In: Commission UCPS, ed. Washington, DC2001.
11. Gummin DD, Mowry JB, Spyker DA, Brooks DE, Fraser MO, Banner W. 2016 Annual Report of the American Association of Poison Control Centers' National Poison Data System (NPDS): 34th Annual Report. *Clinical toxicology (Philadelphia, Pa.).* Dec 2017;55(10):1072-1252.

12. McAninch J SA. Epidemiology Review: Review of the literature and other data sources comparing abuse of buprenorphine-only and buprenorphine-naloxone combination products. Uploaded to DARRTS December 23, 2013.

13. Cicero TJ, Ellis MS, Surratt HL, Kurtz SP. Factors contributing to the rise of buprenorphine misuse: 2008-2013. *Drug and alcohol dependence.* Sep 1 2014;142:98-104.

14. Butler SF, Black RA, Severtson SG, Dart RC, Green JL. Understanding abuse of buprenorphine/naloxone film versus tablet products using data from ASI-MV(R) substance use disorder treatment centers and RADARS(R) System Poison Centers. *Journal of substance abuse treatment.* Jan 2018;84:42-49.

15. Lavonas EJ, Severtson SG, Martinez EM, et al. Abuse and diversion of buprenorphine sublingual tablets and film. *Journal of substance abuse treatment.* Jul 2014;47(1):27-34.

16. Becker WC, Sullivan LE, Tetrault JM, Desai RA, Fiellin DA. Non-medical use, abuse and dependence on prescription opioids among U.S. adults: psychiatric, medical and substance use correlates. *Drug and alcohol dependence.* Apr 1 2008;94(1-3):38-47.

17. Martins SS, Fenton MC, Keyes KM, Blanco C, Zhu H, Storr CL. Mood and anxiety disorders and their association with non-medical prescription opioid use and prescription opioid-use disorder: longitudinal evidence from the National Epidemiologic Study on Alcohol and Related Conditions. *Psychological medicine.* Jun 2012;42(6):1261-1272.

18. Saha TD, Kerridge BT, Goldstein RB, et al. Nonmedical Prescription Opioid Use and DSM-5 Nonmedical Prescription Opioid Use Disorder in the United States. *The Journal of clinical psychiatry.* Jun 2016;77(6):772-780.

19. Goldner EM, Lusted A, Roerecke M, Rehm J, Fischer B. Prevalence of Axis-1 psychiatric (with focus on depression and anxiety) disorder and symptomatology among non-medical prescription opioid users in substance use treatment: systematic review and meta-analyses. *Addictive behaviors.* Mar 2014;39(3):520-531.

20. Blanco C, Wall MM, Okuda M, Wang S, Iza M, Olfson M. Pain as a Predictor of Opioid Use Disorder in a Nationally Representative Sample. *The American journal of psychiatry.* Dec 1 2016;173(12):1189-1195.

21. Boscarino JA, Hoffman SN, Han JJ. Opioid-use disorder among patients on long-term opioid therapy: impact of final DSM-5 diagnostic criteria on prevalence and correlates. *Substance abuse and rehabilitation.* 2015;6:83-91.

22. Schepis TS, Hakes JK. Age of initiation, psychopathology, and other substance use are associated with time to use disorder diagnosis in persons using opioids nonmedically. *Substance abuse.* Oct-Dec 2017;38(4):407-413.

23. Katz C, El-Gabalawy R, Keyes KM, Martins SS, Sareen J. Risk factors for incident nonmedical prescription opioid use and abuse and dependence: results from a longitudinal nationally representative sample. *Drug and alcohol dependence.* Sep 1 2013;132(1-2):107-113.

24. Edlund MJ, Steffick D, Hudson T, Harris KM, Sullivan M. Risk factors for clinically recognized opioid abuse and dependence among veterans using opioids for chronic non-cancer pain. *Pain.* Jun 2007;129(3):355-362.

25. Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: a literature review. *Archives of internal medicine.* Nov 10 2003;163(20):2433-2445.

26. Sullivan MD. Depression Effects on Long-term Prescription Opioid Use, Abuse, and Addiction. *The Clinical journal of pain.* Sep 2018;34(9):878-884.

27. Gureje O, Simon GE, Von Korff M. A cross-national study of the course of persistent pain in primary care. *Pain.* May 2001;92(1-2):195-200.

28. Surah A, Baranidharan G, Morley S. Chronic pain and depression. *Continuing Education in Anaesthesia Critical Care & Pain.* 2014/04/01/ 2014;14(2):85-89.

29. Ohayon MM, Schatzberg AF. Using chronic pain to predict depressive morbidity in the general population. *Archives of general psychiatry.* Jan 2003;60(1):39-47.

30. Halbert BT, Davis RB, Wee CC. Disproportionate longer-term opioid use among U.S. adults with mood disorders. *Pain.* Nov 2016;157(11):2452-2457.

31. Braden JB, Sullivan MD, Ray GT, et al. Trends in long-term opioid therapy for noncancer pain among persons with a history of depression. *General hospital psychiatry.* Nov-Dec 2009;31(6):564-570.

32. Chou R, Turner JA, Devine EB, et al. The effectiveness and risks of long-term opioid therapy for chronic pain: a systematic review for a National Institutes of Health Pathways to Prevention Workshop. *Annals of internal medicine.* Feb 17 2015;162(4):276-286.

33. Voon P, Karamouzian M, Kerr T. Chronic pain and opioid misuse: a review of reviews. *Substance abuse treatment, prevention, and policy.* Aug 15 2017;12(1):36.

34. Turk DC, Swanson KS, Gatchel RJ. Predicting opioid misuse by chronic pain patients: a systematic review and literature synthesis. *The Clinical journal of pain.* Jul-Aug 2008;24(6):497-508.

35. Friedman R, Li V, Mehrotra D. Treating pain patients at risk: evaluation of a screening tool in opioid-treated pain patients with and without addiction. *Pain medicine (Malden, Mass.).* Jun 2003;4(2):182-185.

36. Ives TJ, Chelminski PR, Hammett-Stabler CA, et al. Predictors of opioid misuse in patients with chronic pain: a prospective cohort study. *BMC health services research.* Apr 4 2006;6:46.

37. Bohnert AS, Valenstein M, Bair MJ, et al. Association between opioid prescribing patterns and opioid overdose-related deaths. *Jama.* Apr 6 2011;305(13):1315-1321.

38. Dunn KM, Saunders KW, Rutter CM, et al. Overdose and prescribed opioids: Associations among chronic non-cancer pain patients. *Annals of internal medicine.* 2010;152(2):85-92.

39. Nadpara PA, Joyce AR, Murrelle EL, et al. Risk Factors for Serious Prescription Opioid-Induced Respiratory Depression or Overdose: Comparison of Commercially Insured and Veterans Health Affairs Populations. *Pain medicine (Malden, Mass.).* Jan 1 2018;19(1):79-96.

40. Turner BJ, Liang Y. Drug Overdose in a Retrospective Cohort with Non-Cancer Pain Treated with Opioids, Antidepressants, and/or Sedative-Hypnotics: Interactions with Mental Health Disorders. *Journal of general internal medicine.* Aug 2015;30(8):1081-1096.

41. Yang Z, Wilsey B, Bohm M, et al. Defining risk of prescription opioid overdose: pharmacy shopping and overlapping prescriptions among long-term opioid users in medicaid. *The journal of pain : official journal of the American Pain Society.* May 2015;16(5):445-453.