Graduate thesis

NTNU
Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Mental Health

Mina Kim
Mari Olsen Melhus

A systematic review on the efficacy of buprenorphine as treatment for major depressive disorder in adult patients

Graduate thesis in Medicine
Supervisor: Terje Torgersen
Co-supervisor: Sindre Andre Pedersen
December 2023

NTNU
Norwegian University of
Science and Technology

Mina Kim
Mari Olsen Melhus

A systematic review on the efficacy of buprenorphine as treatment for major depressive disorder in adult patients

Graduate thesis in Medicine
Supervisor: Terje Torgersen
Co-supervisor: Sindre Andre Pedersen
December 2023

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Mental Health

NTNU
Norwegian University of
Science and Technology

Acknowledgements

We would like to thank our main supervisor Terje Torgersen, whose insight and advice helped us form this thesis, and our co-supervisor Sindre Pedersen, for his editorial feedback and for sharing his knowledge with us when planning our search strategy. Also, thank you to the both of you for your cordiality and for answering our never-ending questions.

We would also like to express our gratitude to our family and friends for their great support.

Table of Contents

Acknowledgements **1**

Figures **3**

Tables **3**

Abstract **4**

Sammendrag **5**

Background **6**

Objectives 7

Methods **7**

Eligibility criteria 7

Search strategy 8

Selection and data collection process 8

Method for risk of bias assessment 9

Method for certainty assessment 10

Synthesis method 10

Results **11**

Study selection 11

Study details 12

Risk of bias assessment 14

GRADE assessment 18

Effect of intervention 21

Primary outcome 21

Secondary outcomes 28

Discussion **29**

Conclusion **35**

References **35**

Appendix 1 – search strategy **41**

Figures

Figure 1 PRISMA flow diagram for study selection 12
Figure 2 Risk of Bias assessment 15
Figure 3 Risk of Bias summary plot 15

Tables

Table 1 GRADE assessment 19
Table 2 Results table 21
Table 3 Effect direction plot 27
Table 4 Summary of findings 27

Abstract

Selective serotonin reuptake inhibitors (SSRIs) are currently the most utilized pharmacological treatment for depression. It is most often prescribed to those who have moderate to severe depression, but only around half of the patients experience tangible improvement in their symptoms. As such, there is a demand for alternative medications. Recently, there has been a renewed interest in exploring the potential antidepressant properties of buprenorphine, an opioid analgesic. The aim of this systematic review was to assess the efficacy of buprenorphine in treating major depression in adult patients. Secondary outcomes examined were opioid abuse, addiction, adverse events and suicidal ideation when using buprenorphine. We searched MEDLINE, Embase, Web of Science, Cochrane Central Register of Controlled Trials (CENTRAL) and ClinicalTrials.gov. Included studies were randomized controlled trials of buprenorphine vs. placebo in adult patients with confirmed major depressive disorder (MDD), excluding dementia, chronic pain and substance use disorders. Included studies were evaluated for risk of bias and certainty of evidence, and the primary outcome was evaluated through standardized, validated scales of depression. Seven studies with a total of 1 623 participants were included. Five out of seven studies showed a reduction in depression scores with the changes from baseline being overall larger for the intervention groups than the placebo groups. However, the difference between them was minimal. Certainty of evidence in the studies was deemed moderate based on GRADE. Five studies were deemed to have low risk of bias, one had some concerns while another had high risk of bias. Lastly, we found no clear evidence of abuse or addiction, nor changes in suicidal ideation. Overall, our findings suggest that buprenorphine might reduce depressive symptoms in adults with major depression, but the difference from placebo is minimal. More research is needed to further assess the efficacy, but also long-term and post-treatment effects.

Sammendrag

Selektive serotoninreopptakshemmere (SSRI) er den mest brukte medisinske behandlingen for depresjon. Det forskrives oftest til de med moderat til alvorlig depresjon, men forbedringen i symptomer er ikke merkbar for mer enn omtrent halvparten. Etterspørselen etter alternative medikamenter har i løpet av de siste årene ført til en fornyet interesse for de potensielle antidepressive egenskapene til opioid-analgetikumet buprenorfin. Hensikten med vår systematiske oversikt var å vurdere effekten av buprenorfin som behandling for alvorlig depresjon for voksne pasienter. Sekundære utfall som ble utforsket var opioidmisbruk og -avhengighet, skadevirkninger og selvmordstanker ved bruk av buprenorfin. Vi utførte et litteratursøk via MEDLINE, Embase, Web of Science, Cochrane Central Register of Controlled Trials (CENTRAL) and ClinicalTrials.gov. Inkluderte studier var randomiserte kontrollerte studier som evaluerte buprenorfin sammenlignet med placebo som behandling for alvorlig depresjon hos voksne uten demens, kronisk smerte eller rusmiddellidelser. Risiko for bias og tillitt til dokumentasjon (certainty of evidence) ble vurdert for de inkluderte studiene, og primærutfall ble vurdert ved bruk av standardiserte, validerte skalaer for depresjon. Vi inkluderte sju studier med totalt 1623 deltakere. For fem av sju studier var endringene fra utgangsverdier generelt større for intervensjonsgruppene enn for placebo. Forskjellen mellom buprenorfin og placebo var dog minimal. Vi vurderte tilliten til dokumentasjonen som moderat basert på GRADE. Risiko for bias var lav for fem studier, noe bekymringsverdig for én og høy for én. Vi fant ingen klar evidens for opioidmisbruk eller -avhengighet, eller endringer i selvmordstanker. Funnene fra denne systematiske oversikten tyder på at buprenorfin kan redusere depressive symptomer for voksne med alvorlig depresjon, men at forskjellen fra placebo er minimal. Mer forskning er nødvendig for å vurdere effekten på depresjon nærmere, men også for å vurdere langsiktig- og etterbehandlingseffekt.

Background

Nearly 980 million people suffer from depressive disorders worldwide (1), with 5% of adults experiencing a depressive disorder at any given time (2). Clinical depression is a disorder that can be divided into mild, moderate and severe depression depending on the number and intensity of experienced symptoms (3, 4). Correspondingly, it can also have different expressions depending on the manifested symptoms. Core symptoms are feelings of sadness, anhedonia, and energy depletion, while additional symptoms include suicidal ideation, sleep disruption and low self-worth. The International Statistical Classification of Diseases (ICD-11) and the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) agree that to satisfy the criteria for a major depressive disorder (MDD), the patient should have many or most of these listed symptoms to a marked degree, greatly affecting important areas in the patient's function (3, 4).

Presently, the first line treatment for MDD is selective serotonin (5-HT) re-uptake inhibitors (SSRIs) or other newer antidepressants (5, 6). The use of SSRIs is based on the monoamine hypothesis presented in the 1960s which suggests a connection between depression and lower levels of monoamines (7, 8). Essentially, the monoamines serotonin and noradrenaline are involved in the regulation of mood and drive, and as a premise, increasing their level at the synapse may produce an antidepressant effect. However, SSRIs are only effective among 40-50% of patients with MDD (9). Although the difference in efficacy between SSRI and placebo is statistically significant, (10, 11), the difference in actual effect size is relatively small (10). Furthermore, approximately 63% of patients encounter at least one side-effect during their treatment (12), and abrupt discontinuation can lead to withdrawal syndrome with both somatic and psychological symptoms including nausea, dizziness, anxiety and flu-like symptoms, among others (13, 14). Thus, new treatment options are needed, and alternative medications are being assessed, such as the opioid analgesic buprenorphine (15, 16).

The endogenous opioid system consists of the three classic families of opioid receptors called μ (MOR), δ (DOR) and κ (KOR) and the non-opioid receptor nociceptin (NOP) and their endogenous ligands (17, 18). This system is located throughout both the peripheral and central nervous system (19). Consequently, the endogenous opioid system impacts many different physiological processes such as respiration, pain processing and stress regulation (20), but it also directly affects mood and other key factors involved in the pathophysiology of depression (21, 22). For instance, the results of animal studies have shown that MORs influence reward processing, motivational behavior and social functioning (19, 23-26). Furthermore, deficiency of DORs has a pro-depressive and anxiogenic effect, indicating its possible impact on mood with DOR agonists potentially leading to improvements in mood disorders (27). DORs also mediate reward processing, but in a different manner than MORs (19). In contrast, KORs are a part

of the anti-reward system, and activation of KORs increases anhedonic and dysphoric-like states and aversive and depressive behavior (28). Lastly, recent studies suggests that NOPs also contribute to mood alteration with NOP-blockers eliciting an antidepressant-like effect (29). Mood regulation and many of the forementioned behavioral traits are often impaired in depressive disorders, and emerging research suggests that the endogenous opioid system is dysregulated in patients with MDD (21, 22) and in suicidal patients (30).

As a result, opioids may present as a potential treatment for depression, and currently, buprenorphine is an opioid analgesic being studied for this purpose (15, 16). Buprenorphine is a semisynthetic opioid which was first developed in the 1970s (31). It is a partial MOR agonist and KOR antagonist (32). Due to its partial MOR agonistic effects, buprenorphine is most commonly used as a treatment for opioid use disorder (OUD) to reduce opioid withdrawal symptoms and cravings (33), and like other opioids, the medication can give typical side effects like sedation, constipation and respiration depression, although it is less likely compared to a full agonist (33). Buprenorphine has also displayed antidepressive effects as first seen in a study by Emrich et al. in the early 80s (34). The medication has been shown to reduce depressive symptoms for opioid-naïve patients in later open-label trials as well (15, 16). The partial KOR antagonism is what is hypothesized to give buprenorphine its antidepressive effect (35), and during the last decade, buprenorphine has resurfaced as a potential treatment for MDD in randomized controlled trials (RCTs) (15, 16).

Objectives

The aim of this systematic review is to summarize the evidence and evaluate the efficacy of buprenorphine as treatment for MDD in adult (>18 years) patients.

Methods

Eligibility criteria

Eligible studies had to fulfill the following requirements for inclusion: 1) have RCT as study design, 2) contain a study population of human participants with clinically diagnosed MDD, 3) utilize buprenorphine as primary intervention, 4) compare intervention to a control group of placebo or standard care, 5) evaluate depression as an outcome through validated, standardized measurements of depression and lastly, 6) be published in English. On the other hand, study populations with known co-occurring disorders and conditions such as substance use disorders and chronic pain conditions were excluded since

it would be difficult to assess whether the intervention had a direct impact on depression or indirectly through the alleviation of comorbidity symptoms. People with established dementia were excluded because of potential ethical issues with informed consent and the plausible challenges with the evaluation process as a result of cognitive impairment.

Search strategy

A medical research librarian was consulted during the development of the search strategy. Thus, five databases were searched: MEDLINE (via PubMed), Embase (via Elsevier), Web of Science, The Cochrane Central Register of Controlled Trials (CENTRAL) and the register ClinicalTrials.gov. The main concepts were depression, opioids, treatment and randomized controlled trials. Words related to each concept were combined with the Boolean operator "OR" to cover synonyms and related terms. Both free-text words and MeSH/Emtree terms, when applicable, were used. Finally, the concepts were combined with "AND" to identify the records that covered all the specified concepts. Specific search filters developed by Cochrane and optimized for PubMed (36) and Embase (37) were utilized to restrict the studies to randomized controlled trials. A modified version of the Embase-filter (37) was used for Web of Science by removing subject headings from the search filter. No RCT-filter was applied to the search in the last two databases, CENTRAL and ClinicalTrials.gov, since they are mainly restricted to RCTs (see Appendix 1 for a detailed description of the search strategy used in the different databases). The searches were last updated 25th of august 2023.

Selection and data collection process

Two review authors screened title and abstract independently for potentially relevant studies and the same process was done for the full text screening. Disagreements on inclusion of studies were resolved through further discussion, and the main supervisor provided a third-party opinion if a conclusion was not reached between the reviewers. Missing results from eligible studies were sought for retrieval by direct contact with the study investigators through e-mail. Each of the review authors extracted relevant information from the included studies using a standardized form, and this was subsequently double-checked and corrected by the other. Any disparity in opinion of data interpretation was settled through discussion.

Data items

The primary outcome was change in depression scores as measured on any validated, standardized scale for instance HAM-D17 (38) and MADRS (39). Most of the results were collected at the end of each

treatment period. Data was extracted from the last evaluations and/or the most representative values at end of treatment (EOT) were collected in cases where several measurements from different time points were reported in the same study. The reason was to keep consistency in the chosen values and assess the effect when used over the longest available time period. Even so, other effect estimates were also included in a liberal manner since we were not planning on pooling them together into a single estimate. Hence the inclusion of other values at additional time points if these were prespecified and properly justified by the study investigators as clinically meaningful. Moreover, adjusted measures of effect were selected over unadjusted ones whenever possible. There were no restrictions on types of effect measures collected though most were reported as mean difference from baseline to EOT or placebo-adjusted mean change. We also included the associated standard deviations, standard errors, p-values and/or confidence intervals when provided. Results from different intervention dosages and scores from multiple scales were also gathered.

Additionally, the following secondary outcomes were briefly assessed:

- Suicidal ideation/behavior as assessed with any standardized scoring instrument e.g. Columbia-Suicide Severity Rating Scale (C-SSRS).
- Opioid withdrawal as evaluated with Clinical Opiate Withdrawal Scale (COWS) when available.
- Adverse events presented as a general overview.

Originally, these outcome domains were also included in the review process and in-depth analysis. However, due to insufficient reporting of the results, only a narrative summary was done on these outcomes.

Other characteristics sought for data collection were author name, year of publication, study design and locations, funding sources, clinicaltrials.gov ID, number of participants, study and treatment durations, intervention dose and allocation, type of depression rating instrument, threshold in depression score for participant inclusion and additional characteristics needed for risk of bias assessment (see segment about “Risk of bias assessment” in the results section for further details).

Method for risk of bias assessment

The included studies were assessed for risk of bias with Cochrane’s Risk-of-Bias tool (RoB 2) which appraises five standardized domains: the randomization process, deviations from intended treatment, missing outcome data, measurement of outcomes and selection of the reported results (40). It provides a framework for evaluating different features of a study like the design, management during trial and

reporting of results. A series of questions is presented where the options for answering are “yes”, “probably yes”, “probably no”, “no” and “no information”. An algorithm then suggests either low risk of bias, some concerns or high risk of bias in each domain and an overall judgement of the study. In this review, risk of bias evaluation was done independently by the two review authors then compared and discussed to reach consensus.

Method for certainty assessment

As for gauging the certainty of evidence, an adjusted version of the Grading of Recommendations Development, Assessment and Evaluation (GRADE) approach was chosen (41). A slightly modified variant was adopted since we did not perform a meta-analysis in this review which made it difficult to follow the original GRADE-method rigidly. The modified version (41) still encapsulates the same concepts and domains as the original (42), but evaluates the certainty of evidence in a more narrative and slightly broader sense due to the lack of a single effect estimate. Akin to the original (42), the modified version (41) examines methodological limitations of the studies (risk of bias), indirectness, imprecision, inconsistency and likelihood of publication bias to appraise the quality of evidence for each outcome. Additionally, other elements such as large effects in the studies, a clear dose-response relation and recognizing opposing plausible confounders/biases can increase the certainty in the findings. Each outcome is then summarized into high, moderate, low or very low certainty based on the assessed domains and how confident we are that the estimated effect is close to the true effect.

Synthesis method

In this review, vote counting based on direction of effect was the chosen data synthesis method. This approach (43) relies on effect direction as the standardized metric, and all the effect estimates are dichotomized by their numerical values into whether they indicate either benefit or harm, disregarding other parameters like statistical significance and effect size. The number of effects in favor of the intervention is then compared to the number in opposition within each outcome domain. These proportions can then be used in a sign test (43) to assess whether there is any indication of true effect in the results by rejecting the null hypothesis of no difference in number of positive and negative results. Though this test is only applicable when there is a clear direction of effect in the results as mixed results are discarded.

Data presentation

All the results were described narratively with accompanying figures and tables summarizing the information. Extracted data were displayed in a summary table containing an overview of study design, treatment duration, dose and participant allocation, depression rating scale, reported numerical results and lastly, a short conclusion of the findings. The results were further simplified and visualized in an effect direction plot (44) with the primary studies in order of overall study quality/risk of bias. Arrows were plotted in to represent general direction of effect as positive, negative or mixed/conflicting, and the sizes of the arrows differentiate smaller from larger studies.

Heterogeneity and sensitivity analysis

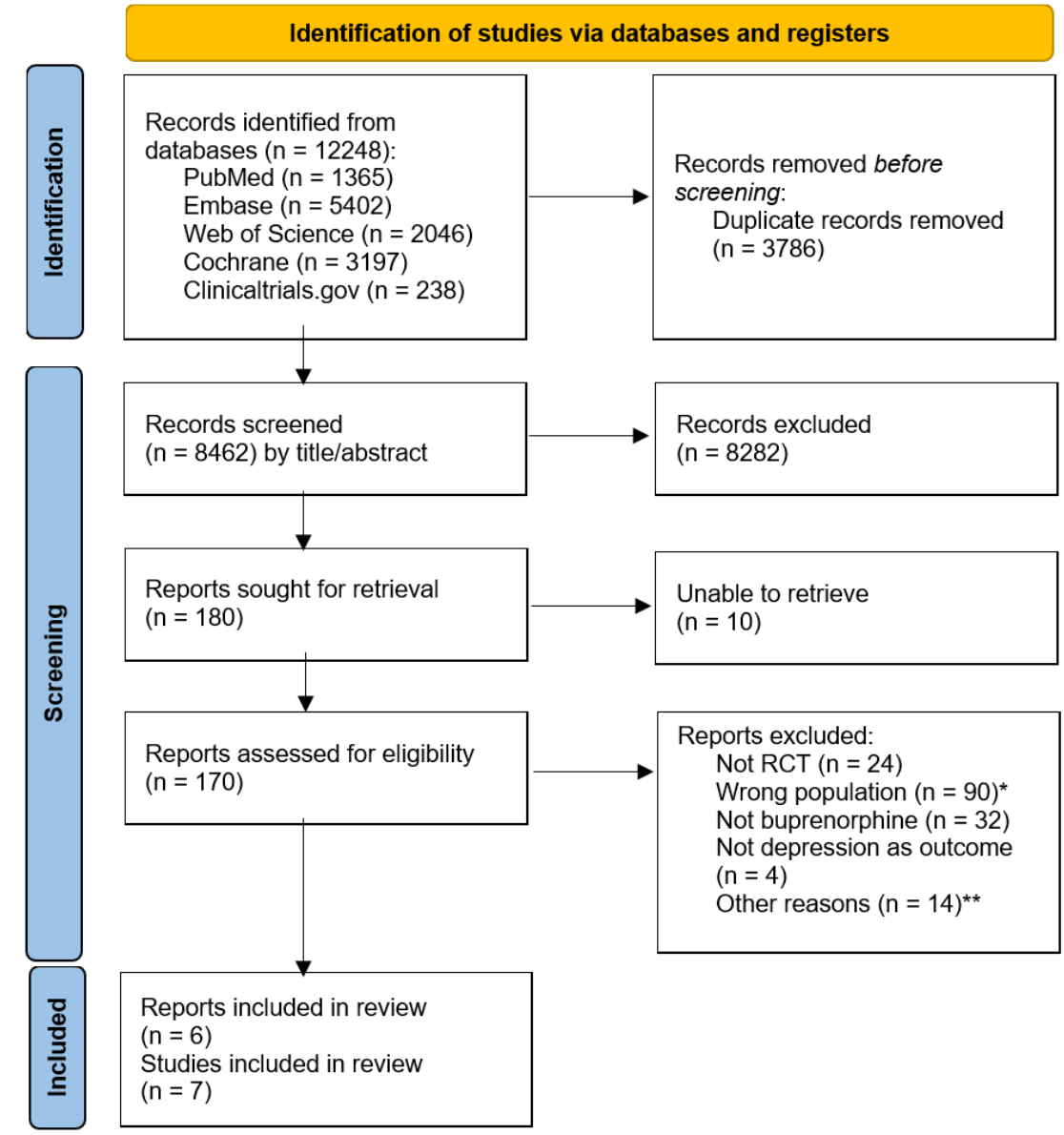
Sensitivity analysis and a formal appraisal of heterogeneity through statistical test, subgroup-analysis or meta-regression were omitted because of the lack of a meta-analysis. Without a meta-analysis, these procedures were deemed negligible since the effect estimates were not combined and the statistics were not transformed in this review. The intention of the review was to examine the general trend of current literature. Thus, only an informal, surface-level of heterogeneity was explored narratively using the summary table. Differences in dosages, study sizes, durations and other apparent distinctions between the studies were compared to assess the possible impact on effect sizes as one would in a normal review.

Results

Study selection

Our database search yielded a total of 12 248 records, where 3 786 were removed as duplicates and an additional 8 282 records were excluded in the title/abstract screening. We were only able to retrieve 170 out of 180 remaining reports for full-text review due to unavailability, restricted access and no response from respective study investigators. Of the retrieved reports, 90 did not match our population criteria: 48 had participants with some form of substance use disorder, 24 were animal studies, 8 encompassed participants with chronic pain and 1 assessed if analgesics reduced depression in demented patients. An additional 8 did not have participants with MDD, and 1 included a mixed population where many psychiatric diagnoses were included without separate analysis for participants with MDD. Another 32 studies used other treatments than opioid analgesics, and 24 were not RCT studies. Additional 4 studies did not assess depression as an outcome, 1 study was not published in English, and we excluded another 13 as duplicates.

In total, 164 reports were excluded during the full-text review, and as noted in the PRISMA flow diagram (Figure 1) (45), we were left with six reports, presenting a total of seven studies. The discrepancy between the number of studies and reports is due to two of the studies being presented in the same publication.



*Wrong population; substance use disorder (n=48), animal study (n=24), not MDD (n=8), chronic pain (n=8), dementia (n=1) or mixed population (n=1)
**Other reasons; additional duplicates (n=13), wrong language (n=1)

Figure 1 PRISMA flow diagram for study selection

Study details

The seven double-blind randomized control studies that met our criteria are presented in Table 2 and had a total population of N=1 623 (N=1 538 for buprenorphine/samidorphan, N=85 for buprenorphine). Study

designs included two parallel RCTs by Ehrich et al. and Lee et al. (46, 47), another parallel RCT with a placebo run-in period by Zajecka et al. (48) and four sequential parallel comparison design (SPCD) RCTs of which one is presented in an unpublicized trial (49) and one is by Fava et al. in 2016 (50) and the remaining two by Fava et al. in 2020 (51).

In SPCD, stage 1 analysis is conducted on the whole study population while for stage 2, placebo non-responders from stage 1 are re-randomized into each treatment group and this smaller sample size is further analyzed. By comparing the intervention groups to placebo non-responders in stage 2, this study method lowers the potential placebo response in analysis, and thus increases the potential difference between experimental treatment and placebo (50-52). The SPCD and placebo run-in design are similar in that both locate placebo non-responders for re-randomization; however in the run-in design, no participants receive study drug before non-responders are found, while for SPCD, the first stage is designed as a parallel RCT in itself (48, 52, 53).

Ehrich et al. (46), Fava et al. 2016 (50), Zajecka et al. (48), Fava et al. 2020 (51) and the unpublished study NCT03188185 (49) were all sponsored by Alkermes, Inc. While the latter's study protocol explicitly states that "Alkermes will remain blinded throughout the interim analysis", the others do not clarify the company's degree of involvement in their publication. Furthermore, Lee et al. (47) received a grant from Indivior Inc., specifying that the company was not involved in any part of the study.

Location and population

Zajecka et al. (48) included participants from USA and Bulgaria, and the FORWARD-4 and FORWARD-5 trials by Fava et al. 2020 (51) were conducted in USA, Canada and Australia, and USA, Canada, Germany, and Puerto Rico respectively. The unpublished trial (49) included participants from USA, Australia, and Puerto Rico. The remaining three studies were conducted in USA (46, 47, 50).

Demographics and/or clinical characteristics were presented for all studies, and major depressive disorder (MDD) defined by DSM-IV-TR, DSM-IV or DSM-V (4, 54, 55) was an inclusion criterion in all studies with a current depressive episode required. Excluding Lee et al. (47) and the unpublished trial (49), all studies defined a lower and/or upper duration cutoff for the current depressive episode, ranging from a minimum of 8 weeks to a maximum of 24 months. Five studies included minimum baseline depression scores ranging from 16 to 18 in HAM-D (48, 50), or 15 in MADRS (47, 51). Mean age was between 43.0 years and 65.6 years. Four studies (48, 49, 51) included patients between 18-70 years, and two (46, 50) included ages 18-65. Lee et al. (47) were the only ones to solely include adults over 50 years of age in their study.

Intervention

Buprenorphine in dosages ranging from 0.2mg/day to 8mg/day was administered sublingually, either alone (47) or with samidorphan in a 1:1 ratio (46, 48-51) or 8:1 ratio (46). Samidorphan is a MOR antagonist and partial KOR and DOR agonist that was added to block buprenorphine's agonistic effects on MOR, and thus prevent abuse and addiction (56, 57). All treatment groups received adjunctive antidepressant treatment (ADT) for the duration of the studies, and in six studies (46, 48-51), excluding Lee et al. (47), participants were required to have been treated with ADT with "inadequate response" as an inclusion criterion. Additionally, the duration of treatment with buprenorphine varied between the studies, but in general, the different treatment periods ranged from 1-8 weeks.

Outcome

Change in depression was assessed by either the Montgomery-Åsberg Depression Rating Scale (MADRS) (39) or the Hamilton Depression Rating Scale (HAM-D) (38). Six studies (46, 48-51) evaluated change from baseline while Lee et al. (47) compared trajectories between intervention groups. For the latter, change from baseline values were not included in the publication, but was registered on ClinicalTrials.gov for two out of three study sites (58, 59). See Table 2 and the subchapter Primary outcome for further details.

Risk of bias assessment

For our included studies, we evaluated five domains that might introduce bias to the results by using the Cochrane Risk-of-Bias tool 2.0 (RoB 2) (40), and an overview of our assessments is presented in Figure 2 and Figure 3. In Figure 2 the trials are sorted by risk of bias from low to high, and then by study size from large to small. Figure 3 outlines the overall risk of bias for all included studies for each of the five domains, correlating to the columns in Figure 2.

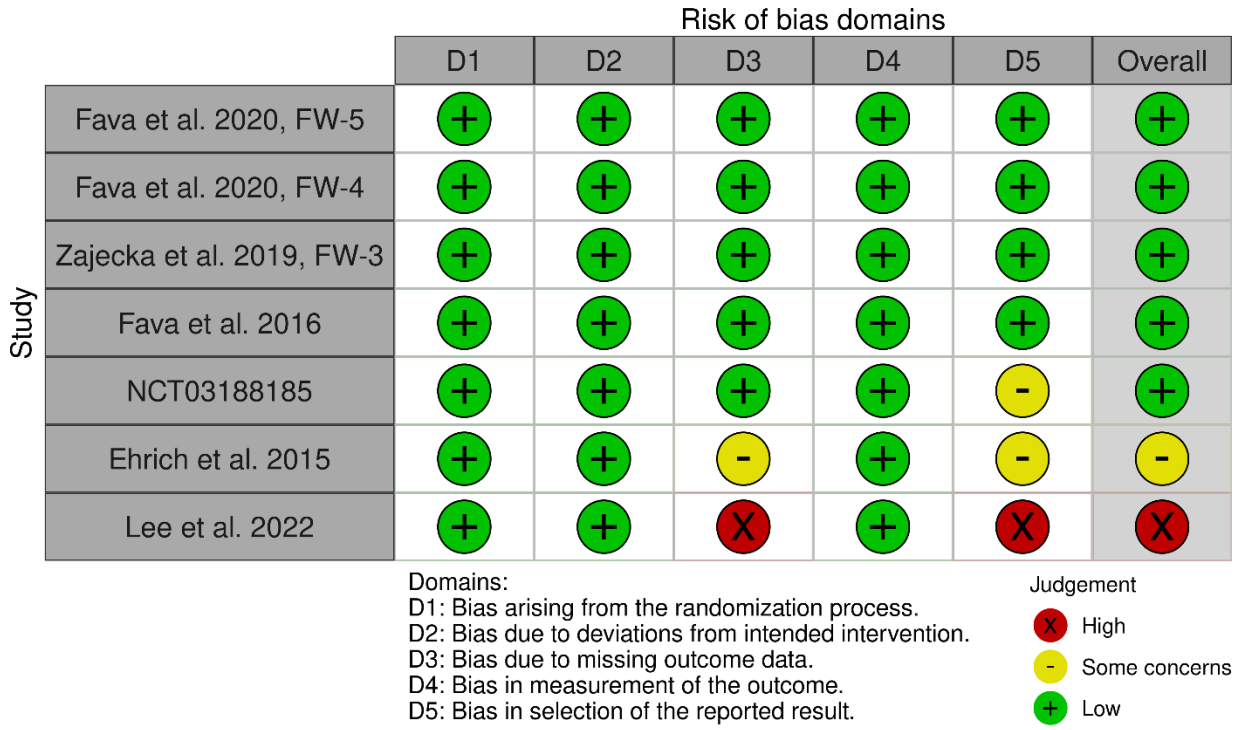


Figure 2 Risk of Bias assessment

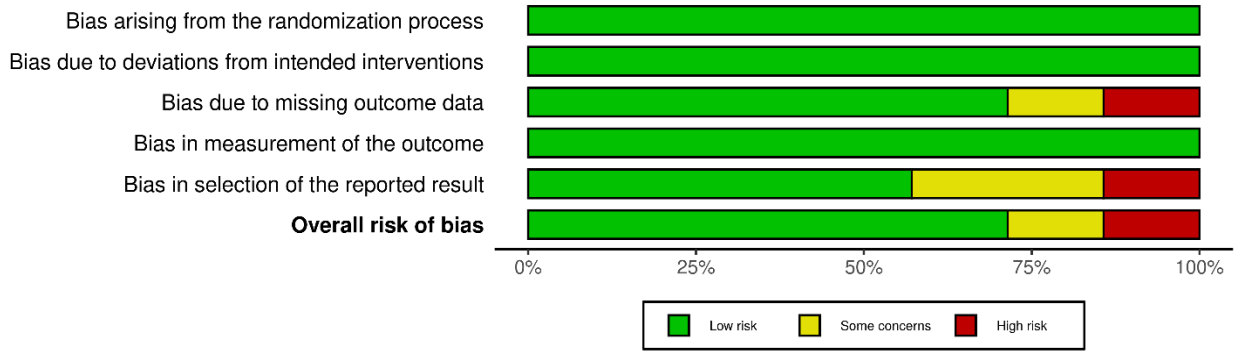


Figure 3 Risk of Bias summary plot

Randomization and allocation

Ehrich et al. (46), Fava et al. 2016 (50) and Zajecka et al. (48) did not report their method for generating a randomization sequence, nor did they explicitly state whether the allocation sequence was concealed until participants were assigned to their intervention group. Nevertheless, their studies were labelled as having triple or quadruple masking on ClinicalTrials.gov, suggesting that allocation was concealed for all studies.

Lee et al. (47) reported using an external consultant to generate a randomization sequence, though no details regarding allocation was further specified. Triple masking does however suggest that the sequence was concealed, like for the aforementioned studies. The remaining three studies (49, 51) described methods for all parts of the randomization process, stating that either an independent party prepared the sequence, or that randomization and treatment assignment was done via an independent interactive voice or web response system (IxRS). None of the studies reported any baseline differences that would suggest bias, and all were judged as having a low risk of bias in the randomization process.

Deviations from intended intervention

As triple or quadruple masking was reportedly utilized in all studies, neither participants nor investigators were assumed to be aware of the assigned intervention during the trials. Fava et al. 2016 (50) reported participants being analyzed according to the intervention they were formally assigned to rather than what they actually received, in other words following an intention-to-treat (ITT) analysis which is considered appropriate and lowers risk of bias score. Information from flow diagrams showed that Ehrich et al. (46) and Lee et al. (47) also analyzed according to ITT. Four studies (48, 49, 51) stated that efficacy was assessed for all participants who had received at least one dose of study drug and had at least one post-baseline measurement. In the trial by Zajecka et al. (48), data from one of 58 sites were excluded from analysis due to evidence of data-integrity issues. This was deemed appropriate and did not introduce bias. In conclusion, we judged all studies to have a low risk of bias when evaluating possible deviations from the intended intervention.

Missing outcome data

Two out of 32 participants in the study by Ehrich et al. (46) discontinued after the first dose of buprenorphine due to vomiting. It is uncertain how they handled the missing data as they stated that both safety and efficacy analyses were performed using results from all randomized participants, without specifying which analysis method they used to adjust for the missing outcome data. Hence, their study was judged as having some concerns in this domain.

Strategies for handling missing data were also undescribed for Lee et al. (47). Moreover, Lee et al. (47) listed “other reasons” as a subcategory for discontinuation, and this subcategory was only used for participants in the buprenorphine group. By comparison, “adverse effects” was a subcategory for both groups. As “other reasons” was not further specified, we do not know why eight people receiving buprenorphine left this study, and additionally, the discontinuation rate for participants receiving buprenorphine was higher than for the control group. Due to these reasons, we judged the risk of bias in handling of missing data as high in this trial.

Four studies reported using mixed models for repeated measures (MMRM) (60) to estimate change from baseline, namely Fava et al. 2016 (50), Zajecka et al. (48) and Fava et al. 2020 (51). This analysis method is based on likelihood and attempts to adjust for missing data without increasing bias (61); data from all previous measurements (from all participants) are used to calculate and predict a mean trajectory for the outcome (60, 62). In contrast, data imputation methods such as last observation carried forward (LOCF) are more prone to bias and may underestimate effect (60, 61). With MMRM as the chosen analysis method to handle missing data, we deemed the risk of bias as low for the four trials.

According to the statistical analysis plan, the unpublished trial (49) will use multiple methods for analysis; primary efficacy endpoint using MMRM, and sensitivity analysis using MMRM, an imputation model, LOCF and pattern-mixed models (PMM) (60). However, change in depression was supposedly analyzed using MMRM, and as such, risk of bias was judged as being low, though this assessment is purely based on the few results currently available and the statistical analysis plan provided.

Outcome measurements

Depression was assessed using validated standardized scales in all trials, and for all trials it was reported that assessors were blinded to the intervention statuses of the participants. Moreover, outcomes were measured at prespecified timepoints for all intervention groups; weekly in all trials (47-51) except for Ehrich et al. (46), where measurements were conducted daily. Thus, all studies were judged as having a low risk of bias when evaluating measurement of the outcome.

Selection of the reported results

As the study conducted by Ehrich et al. (46) was a pilot study, it mainly focused on evaluating the safety and tolerability of buprenorphine/samidorphan in participants diagnosed with MDD. The publication presented preliminary efficacy as a primary outcome, but we do not know if the results would have been included if they found no efficacy since only treatment emergent adverse events were prespecified on ClinicalTrials.gov. Consequently, the study was assigned an unclear risk of selective reporting.

For the unpublished trial (49), not all of their prespecified outcomes were reported on ClinicalTrials.gov, for instance “change from baseline to week three to the end of the treatment period” in MADRS-10 and MADRS-6 and scores in other outcome domains like Clinical Global Impression-Severity (CGI-S) and Snaith-Hamilton Pleasure Scale (SHAPS). Additionally, we do not know if there were any deviations from their prespecified analysis plan since the details regarding the conducted analysis on the results are still unpublished. Thus, there are currently some concerns about the risk of bias for selective reporting.

Lee et al. (47) did not present change from baseline in MADRS score for the overall study or the individual three sites in their publication. They included a graph of the MADRS scores over time, but we

could not derive exact values from this graph. They did, however, note the p-value for the differences between MADRS trajectories over time. There were three registered study records (58, 59, 63), though only two of them were updated with outcome results (58, 59). We did not have the total results from this trial since the outcome data for one of the sites (63), and thereby 36 out of 85 participants, were not made available. As such, the article by Lee et al. (47) was deemed as having a high risk of selective reporting.

The study by Fava et al. 2016 (50) lacked details regarding frequency of outcome measurements on ClinicalTrials.gov, but in the published article it was stated that depression was assessed at each study visit. We therefore deemed the risk of bias as low. For the trials by Zajecka et al. (48) and Fava et al. 2020 (51), the details registered on ClinicalTrials.gov correlated sufficiently with registered outcomes and results, and we did not find any indication of selective reporting.

Overall risk of bias

Although we found the unpublished trial (49) to have some concerns regarding selective reporting of results, the authors have uploaded change from baseline in MADRS-10 and response and remission rates which were their most important prespecified outcomes. Additionally, comprehensive versions of their study protocol and statistical analysis plan were made available, which supports the transparency of the trial. Only a fraction of the results was published on their clinicaltrials-site, possibly because ClinicalTrials.gov is not an ideal site for publishing all the results. Seeing as most of the prior Alkermes Inc. sponsored trials have thoroughly reported their findings in previous articles and supplementary, it is likely that most of the results of their prespecified outcomes in this currently unpublished trial will be revealed later if or when a proper article is issued. As such, the overall risk of bias was assessed as low for this trial. Furthermore, the overall risk of bias was deemed low for four more studies, namely the ones by Fava et al. 2016 (50), Zajecka et al. (48) and the FORWARD-4 and FORWARD-5 trials by Fava et al. 2020 (51). The trial by Ehrich et al. (46) had an overall uncertain risk of bias as the authors did not specify their analysis methods for handling missing data, and there were discrepancies between prespecified outcomes on ClinicalTrials.gov and their article. For the study by Lee et al. (47), we deemed the lack of information on the handling of missing data and the possibility of selective reporting to be serious, resulting in an overall judgement of high risk of bias.

GRADE assessment

To assess the quality of evidence across our primary studies, we applied a modified model (41) of the GRADE approach (42). We appraised the quality of evidence for change in depression, and the

evaluations for each of the five domains of GRADE are presented in Table 1. The secondary outcomes were not assessed because of insufficient reporting in the trials.

Table 1 GRADE assessment

GRADE domain	Judgement	Concerns about certainty domains
Methodological limitations of the studies (risk of bias)	Methodological limitations of the studies were assessed using the Cochrane's Risk-of-Bias tool 2.0 (RoB 2) (40). Randomization, allocation and blinding were mostly documented in all studies. The overall risk of bias was deemed low for five studies (49-51), despite one (49) of them not having registered all of the results yet as previously mentioned. Another trial (46) was assessed as having some concerns due to handling of missing outcome data and discrepancies between pre-registered and reported outcomes. The remaining study (47) had a high risk of bias in both handling of missing data and selective reporting (see Risk of bias assessment). As this last study had a relatively small population (85 participants), we concluded with an overall judgement of no serious methodological limitations, bordering to serious.	Not serious, borderline
Indirectness	Most of the populations, interventions, comparisons and outcomes in the trials provided direct evidence to our review question. Still there were minor discrepancies between the studies, for instance varying doses of buprenorphine, types of depression scales used and differing minimum depression scores required for inclusion in the trials, but all had clinically diagnosed MDD as defined by DSM-IV/V. One study (47) however, had a higher degree of indirectness in the evidence by using mainly fMRI, PET-scanning and transcranial magnetic stimulation in the assessment of depression, though MADRS was also evaluated. This study also had an age cutoff by only including people of 50 years of age and older. Despite this, the overall judgement was that there was no serious indirectness in the evidence as the outlier study still reported clinically direct evidence with a depression rating instrument and included a population with confirmed MDD.	Not serious
Imprecision	The selected studies yielded nearly 1 600 subjects in total. Imprecision across studies was evaluated through the confidence interval approach. Based on the available estimates, the majority of the confidence intervals were either fairly close to the null or enclosed the value. Where the CI included the null, the point estimate often indicated a beneficial effect of the intervention, but the CI still illustrated the possibility of no effect or even a negative effect by including positive values i.e. an increase in depression scores. The imprecision is deemed serious judging by the relatively wide intervals (~4)	Serious

	and the way several confidence intervals contain the null even in the larger studies (48-51).	
Inconsistency	Study design and population were similar between the trials, though treatment duration and daily dose varied. Generally, the reduction in depression was greater for participants receiving buprenorphine than placebo, but the difference between groups was modest. A dose of 2 mg buprenorphine/day showed statistically significant reductions in depression scores in three studies (50, 51), and higher doses did not grant markedly larger decline in depressive symptoms (46, 50). Furthermore, 1mg buprenorphine/day also showed a reduction in depression scores, but the difference from placebo was not statistically significant. Lastly, low-dose buprenorphine (0.5mg) was similar to placebo (47, 51). The conflicting findings between sites in the study by Lee et al. (47, 58, 59) did not result in a general concern of inconsistency, as comparison of overall MADRS-trajectories showed a reduction in depression for all participants. Ultimately, inconsistency across the studies was deemed not serious since we found similar effect sizes and direction of effect in most of the studies, with five (46, 49-51) reporting statistically significant improvements, and two (47, 48) not differing from placebo.	Not serious
Publication bias	There was a combination of both smaller (46, 47) and larger (48-51) published studies with the majority reporting a reduction in intensity of depressive symptoms. Though one of the study sites in the trial by Lee et al. (47) found an increase in depression scores in both the intervention and placebo groups, and Zajecka et al. (48) reported minimal reduction compared to placebo, bordering no effect difference. Additionally, we only found two unpublished trials; one finished in 2016 (64) and another finished in 2021 (49). The former contained incomplete results, lacking baseline MADRS-score which hindered us in evaluating the actual effect of the intervention and if there was any publication bias at play. As for the latter, it was too early to tell if it would be published or not. Considering that a thorough search was conducted, and that both positive and negative findings were published, the general impression was that no major publication bias was apparent.	Undetected

GRADE certainty of evidence template from Murad et al. (41)

The only domain we found to seriously affect the certainty of evidence was the imprecision in the results; while point estimates suggested that buprenorphine relieved depressive symptoms, the accompanying confidence intervals were large, and with the inclusion of the null (48-51) they embodied the possibility of no effect. No serious concerns were raised in the domains assessing methodological limitations, indirectness or inconsistency as the majority of trials had a low risk of bias (48-51), evaluated change in depression as a primary outcome using standardized scales and none of our included studies gave results

that were strikingly different from the others. Lastly, we did not find any major publication bias as trials with both negative and positive findings had been published. A summarized quality assessment of our findings is presented in Table 4.

Effect of intervention

Primary outcome

For our primary outcome of change in depression scores, there were mainly two types of effect measures collected; the mean differences from baseline to end of treatment (EOT) for each treatment group and the placebo-adjusted mean changes. Other associated numerical values such as standard error (SE), standard deviation (SD), p-values and confidence intervals (CI) were also included whenever available. Additionally, results from the different dosages and variants of depression scales have been listed, among other things. In relation to SPCD (49-51), the "overall" least square mean difference (LSMD) refers to the LSMD from baseline to EOT when combining both stage 1 and stage 2 data. Additionally, to correct for variability between weeks in some studies, the study investigators (51) averaged the scores from the last few weeks of treatment (from week 3 through EOT) in each stage and then assessed the respective differences in mean score between baseline and this averaged period. They combined the data from both stages and this combined value is referred to as the "average LSMD from baseline to week 3 through EOT".

All the extracted data for our primary outcome has been summarized in Table 2 below, which gives a comprehensive overview of the study details and key findings from all the studies. The studies are sorted by risk of bias from low to high, and subsequently by study size from large to small.

Table 2 Results table

Study	Study details	Dose, allocation	Key findings	Conclusion
Buprenorphine/samidorphan				
Fava et al. 2020 (51) NCT02218008 (FORWARD-5)	SPCD, N=406 Stage 1: 5-week treatment. Stage 2: 6-week treatment. Allocation: Stage 1: 2:2:9 ratio. Stage 2: 1:1:1 ratio from placebo non-responders. MADRS-10, MADRS-6	Stage 1: 1mg/1mg (1:1 ratio): N=63 2mg/2mg: N=63 Placebo: N=280 Stage 2: 1mg/1mg: N=62 2mg/2mg: N=63 Placebo: N=62	LS mean change from baseline to EOT in MADRS-10: <u>In stage 1:</u> 1mg/1mg: -10.3 (SE = 1.19) 2mg/2mg: -10.8 (SE = 1.22) Placebo: -9.2 (SE = 0.55) <u>In stage 2:</u> 1mg/1mg: -3.4 (SE = 0.98) 2mg/2mg: -3.6 (SE = 0.98) Placebo: -1.9 (SE = 0.96) LSMD between BUP/SAM and placebo from baseline to EOT in stage 1: 1mg/1mg: -1.1 (p = 0.382) 2mg/2mg: -1.6 (p = 0.220)	Even though reduction in depression was numerically larger for both buprenorphine groups compared to placebo, the difference between either of the buprenorphine groups and placebo was not statistically significant from baseline to EOT in both isolated stages and when combining the data from both stages (overall LMSD). The difference was however statistically significant between the 2mg/2mg group

Study	Study details	Dose, allocation	Key findings	Conclusion
			LSMD between BUP/SAM and placebo from baseline to EOT in stage 2: 1mg/1mg: -1.5 (p = 0.281) 2mg/2mg: -1.7 (p = 0.203) Overall LSMD between BUP/SAM and placebo from baseline to EOT (stage 1 and 2 combined): <u>MADRS-10:</u> 1mg/1mg: -1.3 (p = 0.165) 2mg/2mg: -1.7 (p = 0.076, 95% CI [-3.6, 0.2]) <u>MADRS-6:</u> 1mg/1mg: -0.8 (p = 0.262) 2mg/2mg: -1.3 (p = 0.061) Average LSMD between BUP/SAM and placebo from baseline to week 3 through EOT for 2mg/2mg group: MADRS-6: -1.5 (p = 0.018, 95% CI [-2.7, -0.3]) MADRS-10: -1.9 (p = 0.026, 95% CI [-3.6, -0.2])	and placebo through both MADRS-6 and MADRS-10 when looking at average change from baseline to week 3 through EOT (average LMSD).
Fava et al. 2020 (51) NCT02158533 (FORWARD-4)	SPCD, N=384 Stage 1: 5-week treatment. Stage 2: 6-week treatment. Allocation: Stage 1: 2:2:9 ratio. Stage 2: 1:1:1 ratio from placebo non-responders. MADRS-10	Stage 1: 0.5mg/0.5mg: N=59 2mg/2mg: N=60 Placebo: N=265 Stage 2: 0.5mg/0.5mg: N=56 2mg/2mg: N=56 Placebo: N=56	LS mean change from baseline to EOT: <u>In stage 1:</u> 0.5mg/0.5mg: -8.4 (SE = 1.49) 2mg/2mg: -13.0 (SE = 1.50) Placebo: -11.1 (SE = 0.67) <u>In stage 2:</u> 0.5mg/0.5mg: -4.8 (SE = 1.27) 2mg/2mg: -3.9 (SE = 1.13) Placebo: -2.2 (SE = 1.08) LSMD between BUP/SAM and placebo from baseline to EOT in stage 1: 0.5mg/0.5mg: 2.7 (p = 0.097) 2mg/2mg: -1.8 (p = 0.109, 95% CI [-4.1, 0.4]) LSMD between BUP/SAM and placebo from baseline to EOT in stage 2: 0.5mg/0.5mg: -2.4 (p = 0.151) 2mg/2mg: -3.2 (p = 0.041) Overall LSMD between BUP/SAM and placebo from baseline to EOT (stage 1 and 2 combined): 0.5mg/0.5mg: 0.2 (p = 0.881) 2mg/2mg: -2.5 (p = 0.025, 95% CI [-4.7, -0.3]) Average LSMD between BUP/SAM and placebo from baseline to week 3 through EOT: 2mg/2mg: -2.2 (p = 0.023, 95% CI [-4.1, -0.3])	All groups showed a reduction in MADRS-scores from baseline to EOT. However, the difference in depression scores was biggest between the 2mg/2mg group and placebo group in both isolated stages, although not statistically significant in stage 1. Low-dose treatment did not differ from placebo in terms of statistical significance in either stage. In the overall study when combining the data from both stages, post-hoc analysis showed statistically significant difference in depression scores between the 2mg/2mg group and the placebo group. Average LMSD between the 2mg/2mg group and placebo was also statistically significant.
Zajecka et al. 2019 (48) NCT02158546 (FORWARD-3)	Parallel RCT, N=297 4-week placebo run-in period, 6-week treatment. Allocation: 1:1 ratio. MADRS-10	2mg/2mg: N=148 Placebo: N=147 (N=149 randomized)	LS mean change from baseline to EOT: 2mg/2mg: -4.8 (SE = 0.67) Placebo: -4.6 (SE = 0.66) LSMD between BUP/SAM and placebo from baseline to EOT: 2mg/2mg: -0.3 (SE = 0.95; p = 0.782, 95% CI [-2.1, 1.6])	Both groups showed a reduction in depressive symptoms, but the difference between BUP/SAM and placebo was minimal and not statistically significant.

Study	Study details	Dose, allocation	Key findings	Conclusion
Fava et al. 2016 (50) NCT01500200	SPCD RCT, N=141 Two 5-week stages: 4-week treatment; 1-week taper. Allocation: Stage 1: 2:2:9 ratio. Stage 2: 1:1:1 ratio from placebo non-responders. HAM-D17, MADRS	Stage 1: 2mg/2mg: N=24 8mg/8mg: N=19 Placebo: N=98 Stage 2: 2mg/2mg: N=23 8mg/8mg: N=22 Placebo: N=20	LS mean change from baseline to EOT in stage 1: <u>HAM-D17:</u> 2mg/2mg: -9.3 (SE = 1.5) 8mg/8mg: - 6.6 (SE = 1.6) Placebo: -7.1 (SE = 0.6) <u>MADRS-10:</u> 2mg/2mg: -13.3 (SE = 2.2) 8mg/8mg: -11.3 (SE = 2.3) Placebo: -9.6 (SE = 0.9) LS mean change from baseline to EOT in stage 2: <u>HAM-D17:</u> 2mg/2mg: -5.2 (SE = 1.2) 8mg/8mg: -3.3 (SE = 1.1) Placebo: -1.5 (SE = 1.1) <u>MADRS-10:</u> 2mg/2mg: -8.8 (SE = 1.7) 8mg/8mg: -4.7 (SE = 1.7) Placebo: -2.1 (SE = 1.6) LSMD between BUP/SAM and placebo from baseline to EOT in stage 1: <u>HAM-D17:</u> 2mg/2mg: -2.2 (p = 0.168, 95% CI [-5.4, 0.9] 8mg/8mg: 0.5 (p = 0.787, 95% CI [-2.9, 3.8]) <u>MADRS-10:</u> 2mg/2mg: -3.7 (p = 0.020, 95% CI [-8.3, 0.9]) 8mg/8mg: -1.8 (p = 0.483, 95% CI [-6.7, 3.2]) LSMD between BUP/SAM and placebo from baseline to EOT in stage 2: <u>HAM-D17:</u> 2mg/2mg: -3.7 (p = 0.020, 95% CI [-6.9, -0.6]) 8mg/8mg: -1.9 (p = 0.241, 95% CI [-5.0, 1.3]) <u>MADRS-10:</u> 2mg/2mg: -6.7 (p = 0.005, 95% CI [-11.3, -2.0]) 8mg/8mg: -2.6 (p = 0.260, 95% CI [-7.2, 2.0] Overall LSMD between BUP/SAM and placebo from baseline to EOT (stage 1 and 2 combined): <u>HAM-D17:</u> 2mg/2mg: -2.8 (p = 0.014, 95% CI [-5.1, -0.6]) 8mg/8mg: -0.5 (p = 0.699, 95% CI [-2.8, 1.9]) <u>MADRS-10:</u> 2mg/2mg: -4.9 (p = 0.004, 95% CI [-8.2, -1.6]) 8mg/8mg: -2.1 (p = 0.233, 95% CI [-5.6, 1.4])	All groups showed a reduction in depression scores from baseline to EOT. In the overall study, 2mg/2mg dosage gave statistically significant reduction in depressive symptoms noted on both HAM-D17 and MADRS-10, compared to placebo. In the individual stages, the LSMD between the 2mg/2mg group and placebo from baseline to EOT was also statistically significant on both scales in stage 2 and on MADRS-10 in stage 1. Reduction in 8mg/8mg group did not reach statistical significance in either overall study or isolated stages compared to placebo.
Alkermes, Inc. (49) NCT03188185	SPCD, N=278 Stage 1: 5-week treatment. Stage 2: 6-week treatment. MADRS-10	Stage 1: 2mg/2mg: N=80 Placebo: N=198 Stage 2: 2mg/2mg: N=63 Placebo: N= 64	LS mean change in MADRS from baseline to EOT: <u>In stage 1:</u> 2mg/2mg: -13.9 (SE = 1.12) Placebo: -11.4 (SE = 0.70) <u>In stage 2:</u> 2mg/2mg: -4.7 (SE = 1.11) Placebo: -4.2 (SE = 1.06) LSMD between BUP/SAM and placebo from baseline to EOT in stage 1: 2mg/2mg: -2.5	Point estimates suggest greater reduction in depression for participants receiving BUP/SAM than placebo.

Study	Study details	Dose, allocation	Key findings	Conclusion
			LSMD between BUP/SAM and placebo from baseline to EOT in stage 2: 2mg/2mg: -0.5 Overall LSMD between BUP/SAM and placebo from baseline to EOT (stage 1 and 2 combined): 2mg/2mg: -1.5 (p = 0.128, 95% CI [-3.5, 0.4])	
Ehrich et al. 2015 (46) NCT01381107	Parallel RCT, N=32 7-day treatment. HAM-D17, MADRS	2mg/0.25mg → 4mg/0.5mg (8:1 ratio): N=14 4mg/4mg → 8mg/8mg (1:1 ratio): N=14 Placebo: N=4	Mean change from baseline to EOT, HAM-D17: 4mg/0.5mg: -5.0 (SD = 6.1) 8mg/8mg: -6.7 (SD = 3.4) Placebo: -1.0 (SD = 4.2) Mean change from baseline to EOT, MADRS: 4mg/0.5mg: -8.5 (SD = 7.4) 8mg/8mg: -11.5 (SD = 6.5) Placebo: -3.5 (SD = 5.8) Difference between 8mg/8mg and placebo: HAM-D17: p = 0.032 MADRS: p =0.054	The 8mg/8mg group had statistically significant reduction in depressive symptoms compared to placebo as evaluated by HAM-D17, and near statistically significant reduction by MADRS. Although the 4mg/0.5mg group showed reduction in depression, it was not statistically significant compared to placebo.
Buprenorphine				
Lee et al. 2022 (47) NCT02176291 NCT02181231 NCT02263248 (IRLGREY-B)	Parallel RCT, N=85 8-week treatment. Allocation: 2:1 ratio. Trajectory comparison: MADRS scores over time.	0.2mg/day → 1.2mg/day: N=55 Average dose: 0.59mg (SD ±0.33mg) Placebo: N=30	Change from baseline to EOT: Pittsburgh: BUP: 3.47 (SD = 8.94) Placebo: 4.09 (SD = 8.06) St. Louis: BUP: -1 (SD = 6.8) Placebo: -5.3 (SD = 8.8) No statistically significant difference between treatment groups: p = 0.85	There was no statistically significant difference between buprenorphine and placebo when comparing trajectories for both treatment groups.

Abbreviations: *BUP* buprenorphine; *CI* confidence interval; *EOT* end of treatment; *HAM-D* Hamilton Depression Rating Scale; *LS* least square; *LSMD* least square mean difference; *MADRS* Montgomery-Åsberg depression rating scale; *SAM* samidorphan; *SE* standard error; *SD* standard deviation; *SPCD* sequential parallel comparison design

We present a narrative description of the changes in depression scores in all the included studies in the following result section. For all four included sequential parallel comparison design (SPCD) trials, mean change in depression followed the same general pattern, with stage 1 showing larger mean reductions in depression scores than stage 2.

In the FORWARD-5 study by Fava et al. 2020 (51) the 2mg/2mg, 1mg/1mg and placebo groups demonstrated changes in MADRS-10 scores of -10.8, -10.3 and -9.2 respectively from baseline to EOT in stage 1 and changes of -3.6, -3.4 and -1.9 in stage 2. There was generally a larger reduction in scores for participants receiving BUP/SAM compared to placebo. The overall LSMD from placebo were -1.7 for the 2mg/2mg group (p = 0.076) and -1.3 for the 1mg/1mg group (p = 0.165). Furthermore, the average LSMD between 2mg/2mg BUP/SAM and placebo from baseline to week 3 through EOT reached

statistical significance in both MADRS-6 (average LSMD= -1.5; p = 0.018) and MADRS-10 (average LSMD = -1.9; p = 0.026).

In the FORWARD-4 trial by the same authors (51), the 0.5mg/0.5mg, 2mg/2mg and placebo groups showed MADRS-10 changes of -8.4, -13.0 and -11.1 respectively from baseline to EOT in stage 1. In stage 2 the corresponding changes were -4.8, -3.9 and -2.2. In this trial, LSMD from baseline to end of stage 1 was analyzed as part of their primary outcome, showing a difference of -1.8 for the 2mg/2mg group compared to placebo (p = 0.109). While the 0.5mg/0.5mg group had a general reduction in scores from baseline to EOT in both stages, the group had less reduction compared to placebo in stage 1, but there was no statistically significant difference between them (LSMD = 2.7, p = 0.097). For the 2mg/2mg treatment group the overall difference from placebo was statistically significant from baseline to EOT (overall LSMD = -2.5; p = 0.025) and from baseline to week 3 through EOT (average LSMD = -2.2; p = 0.023). For the low-dose group the overall difference from placebo from baseline to EOT was minimal with an LSMD of 0.2 (p = 0.881).

Zajecka et al. (48) also assessed BUP/SAM in a 2mg/2mg dosage compared to placebo, though unlike the forementioned trials they found the mean changes from baseline to end of trial to be quite similar between the groups. The MADRS-10 changes were -4.8 and -4.6 respectively for BUP/SAM and placebo, which resulted in a LSMD of -0.3 (p = 0.782). In our effect direction plot (Table 3) this minimal difference was interpreted as “no change”.

The reduction in depression was larger for both BUP/SAM groups than for placebo in Fava et al. 2016 (50) with MADRS-10 changes of -13.3 for 2mg/2mg and -11.3 for 8mg/8mg against -9.6 for placebo from baseline to EOT in stage 1, and -8.8 and -4.7 against -2.1 respectively in stage 2. In HAM-D17, the corresponding numbers were -9.3, -6.6 and -7.1 in stage 1 and -5.2, -3.3 and -1.5 in stage 2. From baseline to end of stage 1, the LSMD between 2mg/2mg and placebo was -3.7 (p = 0.020) in MADRS-10 and -2.2 (p = 0.168) in HAM-D17. As for 8mg/8mg, the placebo-adjusted values were -1.8 (p = 0.483) and -0.5 (p = 0.787) in the respective scales. In stage 2, placebo-adjusted LSMD were -6.7 (p = 0.005) for 2mg/2mg and -2.6 (p = 0.260) for 8mg/mg in MADRS-10, and -3.7 (p = 0.020) and -1.9 (p = 0.241) in HAM-D17. The overall, placebo-adjusted LSMD for 2mg/2mg was -4.9 (p = 0.004) in MADRS-10 and -2.8 (p = 0.014) in HAM-D17, and for the 8mg/mg group the placebo-adjusted mean differences were -2.1 in MADRS-10 and -0.5 in HAM-D17.

Change from baseline scores in the yet unpublished NCT03188185 (49) pointed toward 2mg/2mg reducing depressive symptoms more than placebo, with MADRS-10 changes of -13.9 for BUP/SAM and -11.4 for placebo in stage 1 and -4.7 vs. -4.2 in stage 2, resulting in LSMD -2.5 in stage 1 and -0.5 in

stage 2. Moreover, the overall LSMD from baseline to EOT was -1.5 for buprenorphine when compared to placebo, which was not statistically significant ($p = 0.128$).

In the pilot study by Ehrich et al. (46), the point estimates suggested greater reduction in depression for both groups receiving BUP/SAM compared to placebo. Mean change in HAM-D was -5.0, -6.7 and -1.0 for BUP/SAM 4mg/0.5mg, 8mg/8mg and placebo respectively, with the difference from placebo reaching statistical significance for the 8mg/8mg group ($p = 0.032$). In MADRS the changes from baseline were -8.5 (4mg/0.5mg), -11.5 (8mg/8mg) and -3.5 (placebo), and difference from placebo was near statistically significant for 8mg/8mg ($p = 0.054$).

Only Lee et al. (47) reported both positive and negative changes in MADRS-scores. With an average dose 0.59mg of buprenorphine, the intervention groups in the Pittsburgh (59) and St. Louis sites (58) had changes of +3.47 and -1 in scores respectively. In comparison, the placebo groups experienced changes of +4.09 in Pittsburgh (59) and -5.3 in St. Louis (58). Hence, there were some contrasting results. In St. Louis, the placebo group had a greater reduction in depression scores than the buprenorphine group while in Pittsburgh both treatment groups experienced an increase in scores. Results from the third location were unobtainable. However, when combining the results from all the participants from the different sites and comparing overall MADRS-trajectories over time, the buprenorphine group had an overall decrease in depression, though the change was not statistically significantly from placebo ($p = 0.85$).

Our findings are simplified in the following plot on the next page (Table 3) where arrow orientations illustrate the direction of effect, i.e. improvement, worsening or no change/conflicting results, while arrow sizes indicate the number of participants receiving buprenorphine in the trial. Note that the plot does not take p-values or magnitude of effect into account. The trials are sorted by risk of bias from low to high, and subsequently by study size from large to small. Five of the seven included studies (1 241 participants) (46, 49-51) showed some improvement in depressive symptoms when using buprenorphine compared to placebo while two trials (233 participants) (47, 48) had conflicting results. As previously mentioned, sign test (43) can be done to assess whether there was an indication of true effect in these results or if it happened by chance. In our case, a two-tailed test showed a p-value of 0.0625 which was not enough to reject the null hypothesis of equal proportions. However, this was only applicable for five studies since they were the only ones showing a clear direction of effect.

Table 3 Effect direction plot

Study	Study design	Change in depression
Fava et al. 2020, FW-5 (51)	RCT (SPCD)	▲
Fava et al. 2020, FW-4 (51)	RCT (SPCD)	▲
Zajecka et al. 2019, FW-3 (48)	Parallel RCT (placebo run-in)	◄►
Fava et al. 2016 (50)	RCT (SPCD)	▲
NCT03188185 (49)	RCT (SPCD)	▲
Ehrich et al. 2015 (46)	Parallel RCT	▲
Lee et al. 2022 (47)	Parallel RCT	◄►

Effect direction: upward arrow ▲= improvement, downward arrow ▼= worsening, sideways arrow ◄►= no change/conflicting results.
Sample size: participants receiving buprenorphine; large arrow ▲ >300; medium arrow ▲ 50-300; small arrow ▲ <50.
Study quality: denoted by row color; green = low risk of bias; yellow = some concerns; red = high risk of bias.
Effect direction plot template from Boon et al. (44)

A summary of our findings is presented in Table 4 below. The majority of studies showed that buprenorphine reduced depressive symptoms slightly more than placebo while two trials reported results that were unclear/conflicting in effect direction. Initially, quality of evidence from randomized trials is graded as high (65). However, the studies had relatively large confidence intervals which often included the null, even in the larger studies, and as detailed in our GRADE assessment (Table 1) this was a serious concern. Thus, we were only moderately confident that the estimated effects were close to the true effect of buprenorphine as treatment for MDD.

Table 4 Summary of findings

Outcome	Effect	Number of participants (studies)	Certainty of evidence
Reduction in depression, assessed with MADRS or HAM-D	Most studies showed some reduction in depression	1623 (7 randomized trials)	MODERATE ⨁⨁⨁◯ (due to serious imprecision)[1]

High certainty ⨁⨁⨁⨁, moderate certainty ⨁⨁⨁◯, low certainty ⨁⨁◯◯ and very low certainty ⨁◯◯◯.
[1]Serious imprecision because of large confidence intervals including the null even in the larger studies.
GRADE summary of findings table template by Murad et al. (41)

Secondary outcomes

Suicidal ideation

All studies reported screening for suicidal ideation (SI) using either the Columbia-Suicide Severity Scale (C-SSRS) (46, 48-51, 66) or Scale for Suicide Ideation (SSI) (47, 67) with subsequent weekly evaluations (daily evaluations for Ehrich et al. (46)). We have chosen to present this outcome narratively because they report their findings differently.

The pooled analysis of the two trials by Fava et al. 2020 (FORWARD-4 and FORWARD-5) (51) showed a lower incidence of suicidal ideation for participants receiving buprenorphine and samidorphan in a 2mg/2mg dosage compared to placebo, however the other trials did not report the same. Zajecka et al. (48) presented the incidence of SI in % at baseline and for postbaseline visits but did not point out any statistically significant difference in SI between groups, and Ehrich et al. (46) and Fava et al. 2016 (50) reported no difference in SI between groups without detailing further. One site (58) from Lee et al. (47) found mean changes of -0.2 (SD = 4.7) and -1.0 (SD = 3.0) from baseline in SSI for low-dose buprenorphine and placebo respectively, and the remaining study (49) has not reported any results yet.

Adverse events and opioid withdrawal

Regarding adverse events, buprenorphine was generally well tolerated, though with dizziness, headache, nausea, vomiting and constipation as common side effects (46-48, 50, 51). These are known adverse events (AE) of opioids (68-70). Fava et al. 2016 and Fava et al. 2020 (50, 51) specified that the most treatment-emergent AEs appeared within the first few days of treatment, and the former (50) suggested this might be due to the titration tempo. Ehrich et al. (46) noted that nausea and vomiting was more prominent for the group receiving buprenorphine and samidorphan in a 8:1 ratio.

With a 1:1 combination of buprenorphine and samidorphan, Zajecka et al. (48), Fava et al. 2016 (50) and Fava et al. 2020 (51) found no evidence of opioid withdrawal using the Clinical Opiate Withdrawal Scale (COWS) (71). Fava et al. 2016 (50) and Ehrich et al. (46) used visual analog scales (VAS) (72) to assess specific aspects of buprenorphine; the former (50) observed no differences in drug liking VAS scores between treatment groups, whereas the latter (46) reported that participants receiving BUP/SAM in a 8:1 ratio (4mg/0.5mg) had higher VAS scores for sedation and feeling high than the 1:1 ratio (8mg/8mg) group. Still, there was no evidence of opioid withdrawal for either ratio group. Lee et al. (47) did not address potential opioid withdrawal, and the unpublished study (49) has not yet presented any results on ClinicalTrials.gov, though assessment is a part of their study protocol.

Discussion

Overall, the majority of the included studies (46, 49-51) exhibited a decline in depressive symptoms as illustrated by the effect direction plot, except for two with unclear/conflicting results (47, 48). The intervention groups presented a mean change from baseline to EOT as values somewhere between +3.47 and -13.9 in MADRS (47, 49, 59) and between -3.3 and -9.3 in HAM-D17 (50). Previous studies have suggested that a reduction of 4-6 points in HAM-D17 is clinically meaningful while a reduction of 7-12 points is a clinically substantial change (73). As for MADRS, the corresponding numbers are approximately 6 and 12 respectively (74). Based on these numbers, the results in our studies indicate that buprenorphine may decently improve depressive symptoms. However, when adjusting the depression scores for placebo in our studies, the results dwindles considerably in numerical values with the least reduction as +4.3 change (47, 58) and biggest as -8.0 (46) in MADRS-scores and between -0.5 to -5.7 in HAM-D17 (46, 50), though, most of the studies showed placebo-adjusted point reductions of approximately 0-4. Furthermore, when maximum scores on MADRS and HAM-D17 are 60 and 52 respectively (38, 39), one has to question whether the low point changes shown in the results will make a tangible difference in the mental health of the potential recipients. However, when considering the results in this review, we should bear in mind that most of the trials were conducted by the same research group and their acquaintances, with Alkermes, Inc. as a major sponsor. This may present a limited point of view and assessment, and analyses from other groups or using different methodology could yield different results. It should also be noted that our primary studies examined various sample sizes, ranging from 32 to 406 participants, and for most studies, this was addressed as a limitation (75) as smaller sample sizes can increase the imprecision of the results (76). Nevertheless, based on the results presented in the included studies, our general impression is that buprenorphine is close to having no clinically meaningful effect as an antidepressant due to the minimal difference from placebo (77).

However, there seems to be small differences even between the most commonly used antidepressants and placebo as well. Administration of SSRIs and SNRIs have been found to reduce depressive symptoms in a statistically significant manner compared to placebo (78, 79), though the difference in actual effect sizes is relatively small (11, 79, 80). For instance, a meta-analysis evaluating the efficacy of fluoxetine (SSRI) and venlafaxine (SNRI) in patients with MDD found a difference of -3 in HAM-D point estimates compared to placebo, thereby favoring the aforementioned antidepressants (11). In a systematic review with a meta-analysis evaluating SSRI for MDD in 2017, the drug-placebo difference was -1.94 in HAM-D, favoring SSRI (10). The efficacy of SSRIs compared to placebo for MDD is smaller than one would expect based on how commonly used the medication is as an antidepressant (10, 11, 79, 81, 82). However, a plausible reason for the marginal difference in effect might be the strong placebo-response

which reduces the measured difference between the assessed antidepressant groups and the placebo groups. This results in a perceived minimal effect compared to placebo. In our included studies, four trials (49-51) tried to correct for this placebo-response by applying a sequential parallel comparison design (SPCD), ideally granting a larger difference between buprenorphine and placebo in stage 2. Though it is uncertain if this was achieved based on our reviewed trials. When comparing differences in MADRS-10 scores between 2mg/2mg BUP/SAM and placebo in the FORWARD-4 and FORWARD-5 trials by Fava et al. 2020 (51), there were minimal drug-placebo differences between the stages; -1.6 in stage 1 vs. -1.7 in stage 2 for FORWARD-5 and -1.8 in stage 1 vs. -3.2 in stage 2 for FORWARD-4. In the unpublished trial (49) the difference between buprenorphine and placebo was bigger in stage 1 (-2.5 in stage 1 vs. -0.5 in stage 2). As for the fourth SPCD trial by Fava et al. 2016 (50), they found a considerably larger difference between the stages, with a drug-placebo difference of -3.7 in stage 1 compared to -6.7 in stage 2. However, the latter could be attributed to the size of the trial as this was the smallest SPCD of them all. With these varying results, we cannot conclude whether SPCDs enhance potential differences between buprenorphine and placebo or not. Furthermore, while the drug-placebo differences for buprenorphine and SSRI are not directly comparable with each other due to the different scales used, we can still see that both have relatively small effect sizes compared to their respective scales i.e. MADRS and HAM-D. Nonetheless, when reviewing these similarities, one should bear in mind that our results are based on significantly fewer studies than the aforementioned systematic review (10).

When further assessing buprenorphine's potential antidepressant properties, some trials outside our included studies have also shown a reduction in depression scores during treatment with buprenorphine. Alleviation of depression with the use of low-dose (2x0.2mg/day) buprenorphine has been reported as early as 1981 (34); thirteen patients were examined and a close to 40% reduction in HAM-D score was found on average. Other newer, open-label trials (83-85) have observed an improvement with the use of buprenorphine as well. For instance, Bodkin et al. (85) examined 10 patients with treatment-resistant depression (TRD) who were treated with a daily average of 1.26mg buprenorphine, resulting in a mean HAM-D score of 28.1 to 10.7 from baseline to EOT (60.7% decrease from baseline, $p = 0.006$). However, only 7 out of 10 participants completed 4-6 weeks of treatment. The others dropped out due to side effects like nausea and malaise. Another trial (83) administered 0.8-2.0mg daily to six subjects with TRD for a week, noticing a mean change from 22.8 to 6.0 on HAM-D and 34.3 to 12.8 on BDI. Additionally, Karp et al. (84) studied 13-15 TRD patients who received an average dose of 0.4mg/day for 8 weeks and found a reduction in mean MADRS score from 27.0 at baseline to 9.5 at EOT. While the low-dose trials (83-85) showed considerable results on their own, there were no control groups present which makes it more difficult to assess the real influence of the medication, and the study groups were also substantially smaller than the RCTs in our review. An RCT including 162 patients suffering from advanced dementia

and depression (86) assessed the efficacy of transdermal (TDS) buprenorphine (maximum 10 µg/hour) and paracetamol tablets in separate groups compared to placebo and found no alleviation of depressive symptoms using these analgesics. On the contrary, the placebo group experienced a larger decrease on the Cornell Scale for Depression in Dementia (CSDD) (87), with a mean change of -3.30, compared to -0.66 in the active treatment group (buprenorphine or paracetamol). The authors emphasized caution when prescribing buprenorphine to this patient group, as 52% receiving buprenorphine discontinued due to adverse events. They also suggested that the lack of improvement might be because of the degree of side effects. Still, the results from this study are probably less applicable to the MDD population as a whole. In short, open label studies (83-85) have found buprenorphine in doses ranging from 0.4mg to 2.0mg to reduce depressive symptoms in patients with TRD, and as shown in the forementioned RCT (87) the medication should be used with caution for patients with advanced dementia.

Nonetheless, there are clear discrepancies between the studies with buprenorphine that must be addressed when comparing its efficacy, for instance the differences in dosages. In this review, the most commonly used dose in the RCTs was 2.0mg of buprenorphine (46, 48-51), but it could range from 0.2mg (47) to 8.0mg (50). Based on our results, the 2.0mg dose generally gave the biggest and most statistically significant changes over time (50, 51), compared to the other utilized doses and placebo. Only Ehrich et al. (46) reported bigger effect with larger doses of 4.0mg-8.0mg buprenorphine than 2.0mg. However, this study was more focused on testing the different BUP/SAM ratios i.e. 1:1 or 8:1 in which the 1:1 ratio contained the higher doses of buprenorphine and coincidently demonstrated larger effects and statistically significant results. Later Alkermes-sponsored studies adopted the 1:1 ratio, but at different dosages where 2.0mg frequently exhibited the most optimal results according to three trials (50, 51). Increasing the dose to for example 8.0mg (50) or using lower doses of 0.5mg or 1.0mg (51) does not appear to be more beneficial than 2.0mg, nor did it give statistically significant differences in scores. The 2.0mg dose also repeatedly reduced the depression score slightly more than placebo (46, 49-51), however the differences were minimal. The study by Zajecka et al. (48) was the only outlier, where the mean difference was -0.3 from placebo despite using 2.0mg buprenorphine. Authors suggested this might be attributed to the high placebo response. Although Lee et al. (47) had an average dose of 0.59mg, the results did not demonstrate the same effect as the low-dose open-label studies that were previously mentioned (83-85). Lee et al. (47) reported the least reduction in depression scores, and one of their sites was the only one to even present an increase in the numbers (59). Moreover, the placebo group at one of their locations (58) had a greater decline in depression scores (MADRS change of -5.3) compared to the active drug group (MADRS change of -1). As remarked in our assessment of bias and GRADE evaluation, this study had some weaknesses regarding evaluation of depression as it evaluated MADRS in relation to imaging diagnostics. Additionally, there were some concerns about the lack of information in the handling of missing data and

the possibility of skewed results due to a higher discontinuation rate in the buprenorphine group than placebo group. All in all, there seems to be no clearly established optimal dosage yet for depression, but the most frequently used dose was 2.0mg in our reviewed trials.

Another differing factor between the studies is the duration of treatment. The primary studies had varying treatment periods, ranging from one week (46) to 4-8 weeks (47-51) at most of continuous, active treatment. It is unknown if a longer period of daily administration would change the outcome. However, despite the short time span of only one week of treatment, Ehrich et al. observed a decline in depressive symptoms (46). This has also been witnessed in an open-label trial of the same time length (83). The open-label trials from Bodkin et al. and Karp et al. also noted a clinical improvement in symptoms after only one week of daily medication with low-dose buprenorphine (84, 85), though the biggest change from baseline was seen after just one week of treatment for the former (85) and after three weeks for the latter (84). The alleviation of symptoms was maintained during both trials, however, a reassessment done 8 weeks post-treatment in the study by Karp et al. showed a notable increase in MADRS scores again which suggests that continuous exposure to BUP may be required to maintain the effect (84). But at present, very few studies have done such a post-treatment re-evaluation in this field, for instance our primary studies have not re-evaluated the effects after discontinuation of treatment. Taken together, there seems to be an indication of rapid antidepressive effect, however, continuous treatment might be necessary for sustaining this improvement, and more trials are needed to assess the long-term and post-treatment effects of opioids on depression.

As for the secondary outcomes in our review, we found no evidence of opioid abuse or dependence, though this observation is solely based on the studies using a buprenorphine/samidorphan combination (46, 48, 50, 51). With its partial MOR agonism, buprenorphine has an abuse potential (56, 57). It is classified as a schedule III drug by the FDA (88) meaning it has a moderate to low potential for physical dependence or a high potential of physiological dependence. However, samidorphan's antagonistic properties on MOR may prevent abuse and addiction (56, 57) which is why the Alkermes-sponsored trials (46, 48-51) administered this substance with buprenorphine. Ehrich et al. evaluated the combination of buprenorphine and samidorphan in varying ratios in two RCTs (46). The first trial was conducted with opioid experienced participants without depression while the second trial assessed opioid-naïve subjects with MDD. The latter study was included in our review as a primary study. Based on the results, the authors concluded that the necessary ratio to achieve full blocking of subjective and physiological opioid effects was a 1:1 ratio of buprenorphine and samidorphan ($p \leq 0.001$ when compared to buprenorphine alone). In contrast, an 8:1 ratio showed an intermediate effect. Since the participants reported to feel more high and sedated by the 8:1 ratio, all consecutive trials sponsored by Alkermes (48-51) later applied the

1:1 ratio to prevent abuse and addiction. The trial by Lee et al. (47) did not evaluate this outcome despite using buprenorphine without a MOR antagonist, plausibly because the opioid effects might not be tangible due to the low dose administered (average 0.59mg/day). This may be supported by the findings from a previous open-label trial (84) assessing buprenorphine for depression, where no clinically significant withdrawal symptoms were observed when administering average doses of 0.4mg/day. All in all, a buprenorphine/samidorphan combination in a 1:1 ratio seems effective at preventing abuse and addiction, though we have not assessed the certainty of evidence for this outcome.

As for suicidal ideation (SI), the difference in degrees of reporting among the included studies limited the assessment of this outcome. The majority of studies did not report any significant changes (46-48, 50, 58), and only Fava et al. 2020 found a lower incidence of SI with buprenorphine when conducting a pooled analysis of both the FORWARD-4 and FORWARD-5 trials (51). Buprenorphine has been shown to reduce SI before; an RCT from 2016 (89) examined severely suicidal patients with heterogeneous diagnoses including MDD and borderline personality disorder after treatment with ultralow doses of buprenorphine and found a greater reduction in SI when using active treatment. Two case reports (90, 91) and another RCT (92) found the same with varying dosages of buprenorphine, however the participants did not have MDD and were drug dependent. While most studies showed no effect of buprenorphine on suicidal ideation for patients with MDD, the outcome ought to be explored further in future research.

Strengths and limitations of the review

As for the review, there are some strengths and limitations of the methodology that should be considered as well. A strong feature in the review is for instance the thorough literature search that was done. Since many synonyms were included in each concept, a broad search was performed which provided a considerable number of potential trials during the study selection. This increased the chances of finding all the available studies that fit the eligibility criteria. Initially, the plan was to include all opioid analgesics in our review which is why “opioids” was a main concept instead of just “buprenorphine”, however, after searching through the databases with our search strategy, we found that buprenorphine was the only opioid analgesic that was currently being tested as a potential antidepressant. As such, we focused mainly on buprenorphine after the search was done, and the review was reshaped accordingly. Only the search method and number of identified studies have remnants of this original plan, and no changes were made to the search strategy since the search was already finished and all the keywords for buprenorphine were already included in the original “opioid” concept. As a result, the search method contained an excessive amount of other opioid analgesics and the general terms for them as well and not just buprenorphine. However, this also increased the chances of finding the trials which utilized

buprenorphine without explicitly mentioning the name in the title or abstract, but only referring to it by generic terms such as “opioid” in the overview of the study.

In contrast, there are several weaknesses to the data synthesis method used in this review. Due to several factors, a meta-analysis was not performed despite being the most advantageous and illustrative method for summarizing the outcomes. Firstly, it was not feasible due to limited time and lack of prior experience with conducting such a procedure. Our current supervisors did not have firsthand experience with it either, and it was challenging to find an available supervisor for guidance in the required statistics as well. As such, the vote counting method (43) was applied, however, this approach only takes direction of effect into account. Other elements such as effect size and statistical significance are not considered despite being relevant for the interpretation of the results. To compensate for the lack of such details in the vote counting approach, a comprehensive result table was compiled to display these values together with other important study features. We also focused on presenting a more comprehensive, narrative description of the results. Furthermore, sign test was applied to evaluate the overall direction of effect in the presented results, thereby, adding another layer of understanding (43). However, only clearly defined effect directions are meant to be included while unclear/mixed effects are dismissed. If the pool of included studies is small, the power of the test might be limited (44). This was the case in our review since only five studies were included in the test, and as such, the result from our sign test was bordering on negligible in terms of importance as it is difficult to draw a definite conclusion from such a small sample pool. In sum, the chosen synthesis method does not fully represent the nuances in the results, but at least provides transparency in the assessment of them in our review.

Moreover, limited data collection and scarce statistical evaluation of the results represent additional limitations with the review. Most of the assessed effect estimates stem from EOT even though multiple, weekly evaluations were provided in the included studies. It is unlikely that solely using EOT-values encapsulates the intervention’s true impact since assessment through a single datapoint is susceptible to random fluctuations from both internal and external factors. An example is depressive symptoms which may naturally vary in intensity over time due to stressors in the person’s life. The SPCD trials (49-51) attempted to compensate for random variability by also calculating the means from baseline to week three through EOT, taking the average of the last few weeks to lessen this impact. To increase accuracy, one could compare evaluations week by week between studies or adopt the same approach as the SPCD studies and average the scores over several weeks, though the latter probably requires longer trials for better assessment. Only evaluating point estimates is another weakness in our review as the numbers may imply an effect for some studies, but the corresponding confidence intervals are large and often include the null, representing the possibility of no effect (93, 94) of buprenorphine. Even though we have

presented p-values for the results, we have not combined or evaluated them to a sufficient degree to draw an overall conclusion from them.

Conclusion

At first glance, buprenorphine might appear as a viable alternative to common antidepressants. Based on current literature, there seems to be a sizable change in depression scores when simply looking at the differences from baseline to EOT, especially in open-label trials, and the changes are often in or close to the double digits. Yet, when compared to control groups in the form of placebo, the difference shrinks as evidenced by the RCTs in our review. Five out of seven of our evaluated RCTs reported a reduction in depression scores in the intervention groups, however the difference from placebo was minimal. As such, one has to consider the clinical significance of these minor point reductions and whether the patients will truly gain a noticeable and meaningful improvement of their symptoms. At the same time, the included studies had several weaknesses that should be considered, for instance relatively small sample sizes, short durations of treatment, unknown optimal dosage and a lack of diversity in study investigators. Furthermore, there was a lack of post-treatment observation and evaluation of long-term usage of the medication in relation to depression, all of which is essential to provide a complete overview of buprenorphine's potential as an antidepressant. Additionally, there were many limitations with our evaluation of the results such as only comparing the effect estimates without numerically combining them and without taking statistical significance into proper consideration. Lastly, there were only a few currently available randomized controlled trials on buprenorphine as treatment for MDD in adult patients, and much is still unknown in this field of study. Thus, it is difficult to decide if one should recommend such a medication based on current literature. In conclusion, more research with larger and longer trials is needed to fully reveal buprenorphine's potential and efficacy and to remedy the shortcomings presented in this review.

References

1. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet Psychiatry. 2022;9(2):137-50.
2. World Health Organization. Depression [website]. Geneva: World Health Organization; 2021 [updated 13/09/2021; cited 2023 07.01.]. Available from: https://www.who.int/news-room/fact-sheets/detail/depression.

3. World Health Organization. ICD-11 for Mortality and Morbidity Statistics (Version : 02/2022) Geneva: World Health Organization; 2022 [cited 2023 09.01.]. Available from: https://icd.who.int/browse11/l-m/en.
4. American Psychiatric Association. Diagnostic And Statistical Manual of Mental Disorders, Fifth Edition. Arlington, VA: American Psychiatric Association; 2013.
5. Cleare A, Pariante C, Young A, Anderson I, Christmas D, Cowen P, et al. Evidence-based guidelines for treating depressive disorders with antidepressants: A revision of the 2008 British Association for Psychopharmacology guidelines. Journal of Psychopharmacology. 2015;29(5):459-525.
6. National Institute for Health and Care Excellence: Guidelines. Depression in adults: recognition and management. London: National Institute for Health and Care Excellence (NICE) Copyright © NICE 2019.; 2018.
7. Dale E, Bang-Andersen B, Sánchez C. Emerging mechanisms and treatments for depression beyond SSRIs and SNRIs. Biochemical Pharmacology. 2015;95(2):81-97.
8. Carlsson A, Corrodi H, Fuxe K, Hökfelt T. Effect of antidepressant drugs on the depletion of intraneuronal brain 5-hydroxytryptamine stores caused by 4-methyl-α-ethyl-meta-tyramine. European Journal of Pharmacology. 1969;5(4):357-66.
9. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry. 2006;163(1):28-40.
10. Jakobsen JC, Katakam KK, Schou A, Hellmuth SG, Stallknecht SE, Leth-Møller K, et al. Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and Trial Sequential Analysis. BMC Psychiatry. 2017;17(1):58.
11. Gibbons RD, Hur K, Brown CH, Davis JM, Mann JJ. Benefits From Antidepressants: Synthesis of 6-Week Patient-Level Outcomes From Double-blind Placebo-Controlled Randomized Trials of Fluoxetine and Venlafaxine. Archives of General Psychiatry. 2012;69(6):572-9.
12. Gartlehner G, Hansen RA, Morgan LC, Thaler K, Lux L, Van Noord M, et al. Comparative benefits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. Ann Intern Med. 2011;155(11):772-85.
13. Fava GA, Gatti A, Belaise C, Guidi J, Offidani E. Withdrawal Symptoms after Selective Serotonin Reuptake Inhibitor Discontinuation: A Systematic Review. Psychotherapy and Psychosomatics. 2015;84(2):72-81.
14. Kovich H, Kim W, Quaste AM. Pharmacologic Treatment of Depression. Am Fam Physician. 2023;107(2):173-81.
15. Riblet NB, Young-Xu Y, Shiner B, Schnurr PP, Watts BV. The efficacy and safety of buprenorphine for the treatment of depression: A systematic review and meta-analysis. Journal of Psychiatric Research. 2023;161:393-401.
16. Serafini G, Adavastro G, Canepa G, De Berardis D, Valchera A, Pompili M, et al. The Efficacy of Buprenorphine in Major Depression, Treatment-Resistant Depression and Suicidal Behavior: A Systematic Review. Int J Mol Sci. 2018;19(8).
17. Benarroch EE. Endogenous opioid systems. Current concepts and clinical correlations. 2012;79(8):807-14.
18. McDonald J, Lambert DG. Opioid receptors. BJA Education. 2015;15(5):219-24.
19. Le Merrer J, Becker JA, Befort K, Kieffer BL. Reward processing by the opioid system in the brain. Physiol Rev. 2009;89(4):1379-412.
20. Lutz P-E, Kieffer BL. Opioid receptors: distinct roles in mood disorders. Trends in Neurosciences. 2013;36(3):195-206.
21. Berrocoso E, Sánchez-Blázquez P, Garzón J, Mico JA. Opiates as antidepressants. Curr Pharm Des. 2009;15(14):1612-22.
22. Peciña M, Karp JF, Mathew S, Todtenkopf MS, Ehrich EW, Zubieta JK. Endogenous opioid system dysregulation in depression: implications for new therapeutic approaches. Mol Psychiatry. 2019;24(4):576-87.

23. Becker A, Grecksch G, Brödemann R, Kraus J, Peters B, Schroeder H, et al. Morphine self-administration in μ-opioid receptor-deficient mice. Naunyn-Schmiedeberg's Archives of Pharmacology. 2000;361(6):584-9.
24. Ghozland S, Matthes HWD, Simonin F, Filliol D, Kieffer BL, Maldonado R. Motivational Effects of Cannabinoids Are Mediated by μ-Opioid and κ-Opioid Receptors. The Journal of Neuroscience. 2002;22(3):1146-54.
25. Moles A, Kieffer BL, D'Amato FR. Deficit in attachment behavior in mice lacking the mu-opioid receptor gene. Science. 2004;304(5679):1983-6.
26. Pellissier LP, Gandía J, Laboute T, Becker JAJ, Le Merrer J. μ opioid receptor, social behaviour and autism spectrum disorder: reward matters. British Journal of Pharmacology. 2018;175(14):2750-69.
27. Filliol D, Ghozland S, Chluba J, Martin M, Matthes HWD, Simonin F, et al. Mice deficient for δ- and μ-opioid receptors exhibit opposing alterations of emotional responses. Nature Genetics. 2000;25(2):195-200.
28. Jelen LA, Stone JM, Young AH, Mehta MA. The opioid system in depression. Neurosci Biobehav Rev. 2022;140:104800.
29. Gavioli EC, Calo G. Nociceptin/orphanin FQ receptor antagonists as innovative antidepressant drugs. Pharmacol Ther. 2013;140(1):10-25.
30. Nobile B, Lutz P-E, Olie E, Courtet P. The Role of Opiates in Social Pain and Suicidal Behavior. In: Baca-Garcia E, editor. Behavioral Neurobiology of Suicide and Self Harm. Cham: Springer International Publishing; 2020. p. 197-210.
31. Cowan A, Lewis JW, Macfarlane IR. Agonist and antagonist properties of buprenorphine, a new antinociceptive agent. Br J Pharmacol. 1977;60(4):537-45.
32. Kress HG. Clinical update on the pharmacology, efficacy and safety of transdermal buprenorphine. European Journal of Pain. 2009;13(3):219-30.
33. Shulman M, Wai JM, Nunes EV. Buprenorphine Treatment for Opioid Use Disorder: An Overview. CNS Drugs. 2019;33(6):567-80.
34. Emrich HM, Günther R, Dose M. Current perspectives in the pharmacopsychiatry of depression and mania. Neuropharmacology. 1983;22(3, Part 2):385-8.
35. Namchuk AB, Lucki I, Browne CA. Buprenorphine as a Treatment for Major Depression and Opioid Use Disorder. Adv Drug Alcohol Res. 2022;2.
36. Lefebvre C, Glanville J, Briscoe S, Featherstone R, Littlewood A, Marshall C, Metzendorf M-I, Noel-Storr A, Paynter R, Rader T, Thomas J, Wieland LS. Technical Supplement to Chapter 4: Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston MS, Li T, Page MJ, Welch VA (eds). Cochrane Handbook for Systematic Reviews of Interventions Version 6.3 (updated February 2022). Cochrane, 2022. [cited 28/09/2023] Available from: www.training.cochrane.org/handbook.
37. Glanville J, Foxlee R, Wisniewski S, Noel-Storr A, Edwards M, Dooley G. Translating the Cochrane EMBASE RCT filter from the Ovid interface to Embase.com: a case study. Health Information & Libraries Journal. 2019;36(3):264-77.
38. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960;23(1):56-62.
39. Montgomery SA, Åsberg M. A New Depression Scale Designed to be Sensitive to Change. British Journal of Psychiatry. 1979;134(4):382-9.
40. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898.
41. Murad MH, Mustafa RA, Schünemann HJ, Sultan S, Santesso N. Rating the certainty in evidence in the absence of a single estimate of effect. Evidence Based Medicine. 2017;22(3):85-7.
42. Schünemann H, Brożek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. The GRADE Working Group, 2013. [cited 23/10/2023] Available from: https://guidelinedevelopment.org/handbook.
43. McKenzie JE, Brennan SE. Chapter 12: Synthesizing and presenting findings using other methods. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors).

Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane, 2023. [accessed: 7/10/2023] Available from www.training.cochrane.org/handbook.
44. Boon MH, Thomson H. The effect direction plot revisited: Application of the 2019 Cochrane Handbook guidance on alternative synthesis methods. Research Synthesis Methods. 2021;12(1):29-33.
45. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ. 2021;372:n160.
46. Ehrich E, Turncliff R, Du Y, Leigh-Pemberton R, Fernandez E, Jones R, et al. Evaluation of Opioid Modulation in Major Depressive Disorder. Neuropsychopharmacology. 2015;40(6):1448-55.
47. Lee HH, Blumberger DM, Lenze EJ, Anderson SJ, Barch DM, Black KJ, et al. Low-Dose Augmentation With Buprenorphine for Treatment-Resistant Depression: A Multisite Randomized Controlled Trial With Multimodal Assessment of Target Engagement. Biological Psychiatry Global Open Science. 2022;2(2):127-35.
48. Zajecka JM, Stanford AD, Memisoglu A, Martin WF, Pathak S. Buprenorphine/samidorphan combination for the adjunctive treatment of major depressive disorder: results of a phase III clinical trial (FORWARD-3). Neuropsychiatric Disease and Treatment. 2019;15:795-808.
49. Clinicaltrials.gov [Internet]. USA: Alkermes, Inc. [14/06/2017] Identifier: NCT03188185 - A Study of ALKS 5461 for Treatment Refractory Major Depressive Disorder (MDD). [cited 5/10/2023]. Available from: https://clinicaltrials.gov/study/NCT03188185?term=NCT03188185&rank=1.
50. Fava M, Memisoglu A, Thase ME, Bodkin JA, Trivedi MH, de Somer M, et al. Opioid Modulation With Buprenorphine/Samidorphan as Adjunctive Treatment for Inadequate Response to Antidepressants: A Randomized Double-Blind Placebo-Controlled Trial. American Journal of Psychiatry. 2016;173(5):499-508.
51. Fava M, Thase ME, Trivedi MH, Ehrich E, Martin WF, Memisoglu A, et al. Opioid system modulation with buprenorphine/samidorphan combination for major depressive disorder: two randomized controlled studies. Molecular Psychiatry. 2020;25(7):1580-91.
52. Fava M, Evins AE, Dorer DJ, Schoenfeld DA. 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(3):115-27.
53. Laursen DRT, Paludan-Müller AS, Hróbjartsson A. Randomized clinical trials with run-in periods: frequency, characteristics and reporting. Clin Epidemiol. 2019;11:169-84.
54. Association AP. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington, DC: American Psychiatric Association; 1994.
55. Association AP. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision: American Psychiatric Association; 2000.
56. Chaudhary AMD, Khan MF, Dhillon SS, Naveed S. A Review of Samidorphan: A Novel Opioid Antagonist. Cureus. 2019;11(7):e5139.
57. Bidlack J, Knapp BI, Deaver DR, Plotnikava M, Arnelle D, Quinn AM, et al. S165. ALKS 5461: Affinity, Potency, and Functional Activity of Buprenorphine and Samidorphan Alone and in Combination. Biological Psychiatry. 2018;83(9, Supplement):S412.
58. Clinicaltrials.gov [Internet]. Washington University School of Medicine, St. Louis: Eric Lenze [02/07/2014] Identifier: NCT02181231 - Buprenorphine Used With Treatment Resistant Depression in Older Adults (IRLGreyB). [cited 5/10/2023]. Available from: https://clinicaltrials.gov/study/NCT02181231?term=NCT02181231&rank=1.
59. Clinicaltrials.gov [Internet]. University of Pittsburgh: Jordan F. Karp [25/06/2014] Identifier: NCT02176291 - Incomplete Response in Late-Life Depression: Getting to Remission With Buprenorphine (IRLGREY-B). [cited 5/10/2023]. Available from: https://clinicaltrials.gov/study/NCT02176291?term=NCT02176291&rank=1.
60. Prakash A, Risser RC, Mallinckrodt CH. The impact of analytic method on interpretation of outcomes in longitudinal clinical trials. Int J Clin Pract. 2008;62(8):1147-58.

61. Barnes SA, Mallinckrodt CH, Lindborg SR, Carter MK. The impact of missing data and how it is handled on the rate of false-positive results in drug development. Pharm Stat. 2008;7(3):215-25.
62. Detry MA, Ma Y. Analyzing Repeated Measurements Using Mixed Models. JAMA. 2016;315(4):407-8.
63. Clinicaltrials.gov [Internet]. Toronto: Daniel Blumberger [10/10/2014] Identifier: NCT02263248 - Incomplete Response in Late-Life Depression: Getting to Remission With Buprenorphine ((IRLGREY-B)). [cited 5/10/2023]. Available from: https://clinicaltrials.gov/study/NCT02263248?term=NCT02263248&rank=1.
64. Clinicaltrials.gov [Internet]. University of Pittsburgh: Jordan F. Karp [09/03/2018] Identifier: NCT01407575 - Buprenorphine for Treatment Resistant Depression (BUP-TRD). [cited 5/10/2023]. Available from: https://clinicaltrials.gov/study/NCT01407575.
65. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. Journal of Clinical Epidemiology. 2011;64(4):401-6.
66. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry. 2011;168(12):1266-77.
67. Beck AT, Kovacs M, Weissman A. Assessment of suicidal intention: The Scale for Suicide Ideation. Journal of Consulting and Clinical Psychology. 1979;47(2):343-52.
68. Kurz A, Sessler DI. Opioid-Induced Bowel Dysfunction. Drugs. 2003;63(7):649-71.
69. Benyamin R, Trescot AM, Datta S, Buenaventura R, Adlaka R, Sehgal N, et al. Opioid complications and side effects. Pain Physician. 2008;11(2 Suppl):S105-20.
70. Pathan H, Williams J. Basic opioid pharmacology: an update. Br J Pain. 2012;6(1):11-6.
71. Wesson DR, Ling W. The Clinical Opiate Withdrawal Scale (COWS). J Psychoactive Drugs. 2003;35(2):253-9.
72. Aitken RC. Measurement of feelings using visual analogue scales. Proc R Soc Med. 1969;62(10):989-93.
73. Rush AJ, South C, Jain S, Agha R, Zhang M, Shrestha S, et al. Clinically Significant Changes in the 17- and 6-Item Hamilton Rating Scales for Depression: A STAR*D Report. Neuropsychiatr Dis Treat. 2021;17:2333-45.
74. Turkoz I, Alphs L, Singh J, Jamieson C, Daly E, Shawi M, et al. Clinically meaningful changes on depressive symptom measures and patient-reported outcomes in patients with treatment-resistant depression. Acta Psychiatrica Scandinavica. 2021;143(3):253-63.
75. Hengartner MP, Jakobsen JC, Sørensen A, Plöderl M. Efficacy of new-generation antidepressants assessed with the Montgomery-Asberg Depression Rating Scale, the gold standard clinician rating scale: A meta-analysis of randomised placebo-controlled trials. PLoS One. 2020;15(2):e0229381.
76. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. Journal of Biomedical Informatics. 2014;48:193-204.
77. Barrett B, Brown D, Mundt M, Brown R. Sufficiently Important Difference: Expanding the Framework of Clinical Significance. Medical Decision Making. 2005;25(3):250-61.
78. Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, et al. Antidepressant Drug Effects and Depression Severity: A Patient-Level Meta-analysis. JAMA. 2010;303(1):47-53.
79. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med. 2008;5(2):e45.
80. Linde K, Kriston L, Rücker G, Jamil S, Schumann I, Meissner K, et al. Efficacy and acceptability of pharmacological treatments for depressive disorders in primary care: systematic review and network meta-analysis. Ann Fam Med. 2015;13(1):69-79.
81. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy. New England Journal of Medicine. 2008;358(3):252-60.

82. Barbui C, Barth M, Cipriani A, Klostermann S, Kriston L, Linde K. Efficacy of selective serotonin reuptake inhibitors and adverse events: Meta-regression and mediation analysis of placebo-controlled trials. British Journal of Psychiatry. 2016;208(2):114-9.
83. Nyhuis PW, Gastpar M, Scherbaum N. Opiate Treatment in Depression Refractory to Antidepressants and Electroconvulsive Therapy. Journal of Clinical Psychopharmacology. 2008;28(5):593-5.
84. Karp JF, Butters MA, Begley AE, Miller MD, Lenze EJ, Blumberger DM, et al. Safety, tolerability, and clinical effect of low-dose buprenorphine for treatment-resistant depression in midlife and older adults. J Clin Psychiatry. 2014;75(8):e785-93.
85. Bodkin JA, Zornberg GL, Lukas SE, Cole JO. Buprenorphine Treatment of Refractory Depression. Journal of Clinical Psychopharmacology. 1995;15(1):49-57.
86. Erdal A, Flo E, Aarsland D, Ballard C, Slettebo DD, Husebo BS. Efficacy and Safety of Analgesic Treatment for Depression in People with Advanced Dementia: Randomised, Multicentre, Double-Blind, Placebo-Controlled Trial (DEP.PAIN.DEM). Drugs Aging. 2018;35(6):545-58.
87. Alexopoulos GS, Abrams RC, Young RC, Shamoian CA. Cornell Scale for Depression in Dementia. Biol Psychiatry. 1988;23(3):271-84.
88. Justice USDo. Controlled Substances - Alphabetical Order USA: Diversion Control Division, Drug Enforcement Administration, U.S. Department of Justice; [updated 07.11.23. Available from: https://www.deadiversion.usdoj.gov/schedules/orangebook/c_cs_alpha.pdf.
89. Yoram Yovell, M.D., Ph.D. ,, Gali Bar, Ph.D. ,, Moti Mashiah, M.D. ,, Yehuda Baruch, M.D. ,, Irina Briskman, M.D. ,, Jack Asherov, M.D. ,, et al. Ultra-Low-Dose Buprenorphine as a Time-Limited Treatment for Severe Suicidal Ideation: A Randomized Controlled Trial. American Journal of Psychiatry. 2016;173(5):491-8.
90. Striebel JM, Kalapatapu RK. The anti-suicidal potential of buprenorphine: a case report. Int J Psychiatry Med. 2014;47(2):169-74.
91. Ahmadi J, Abtahi S. Eight-Milligram Single Dose of Buprenorphine as an Expeditious Treatment for Severe Suicidal Ideation: A Case Report. Iranian Red Crescent Medical Journal. 2017;19(9):13-.
92. Ahmadi J, Sarani EM, Jahromi MS. Rapid effect of a single-dose buprenorphine on reduction of opioid craving and suicidal ideation: A randomized, double blind, placebo-controlled study. Ci Ji Yi Xue Za Zhi. 2020;32(1):58-64.
93. Frank O, Tam CM, Rhee J. Is it time to stop using statistical significance? Aust Prescr. 2021;44(1):16-8.
94. Sterne JA, Davey Smith G. Sifting the evidence-what's wrong with significance tests? Bmj. 2001;322(7280):226-31.

Appendix 1 – search strategy **(last updated 25th of August 2023):**

MEDLINE (via PubMed)

#	Searches
1.	"Depression"[Mesh]
2.	"Depressive Disorder"[Mesh]
3.	"Self-Injurious Behavior"[Mesh]
4.	depression[Title/Abstract]
5.	depressive[Title/Abstract]
6.	depressed[Title/Abstract]
7.	MDD[Title/Abstract]
8.	suicide[Title/Abstract]
9.	suicidal*[Title/Abstract]
10.	"self-harm*"[Title/Abstract]
11.	antisuicidal[Title/Abstract]
12.	"self-injur*"[Title/Abstract]
13.	SIB[Title/Abstract]
14.	"self-inflict*"[Title/Abstract]
15.	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14
16.	"Analgesics, Opioid"[Mesh]
17.	Opioid*[Title/Abstract]
18.	opiate*[Title/Abstract]
19.	narcotic*[Title/Abstract]
20.	buprenorphine[Title/Abstract]
21.	Buprenex[Title/Abstract]
22.	Subutex[Title/Abstract]
23.	Zubsolv[Title/Abstract]
24.	Probuphine[Title/Abstract]
25.	Sublocade[Title/Abstract]
26.	Belbuca[Title/Abstract]
27.	Butrans[Title/Abstract]
28.	Sixmo[Title/Abstract]
29.	Bunavail[Title/Abstract]

30. Buvidal[Title/Abstract]
31. Temgesic[Title/Abstract]
32. Norspan[Title/Abstract]
33. BUP[Title/Abstract]
34. ALKS-5461[Title/Abstract]
35. methadone[Title/Abstract]
36. Dolophine[Title/Abstract]
37. methadose[Title/Abstract]
38. Metadol[Title/Abstract]
39. Metadon[Title/Abstract]
40. Physeptone[Title/Abstract]
41. Diskets[Title/Abstract]
42. Suboxone[Title/Abstract]
43. #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40 OR #41 OR #42
44. "Drug Therapy"[Mesh]
45. treat*[Title/Abstract]
46. therap*[Title/Abstract]
47. pharmacotherap*[Title/Abstract]
48. "pharmacologic* intervention*"[Title/Abstract]
49. #44 OR #45 OR #46 OR #47 OR #48
50. "randomized controlled trial"[Publication Type]
51. "controlled clinical trial"[Publication Type]
52. "randomized"[Title/Abstract]
53. "placebo"[Title/Abstract]
54. "clinical trials as topic"[MeSH Terms:noexp]
55. "randomly"[Title/Abstract]
56. "trial"[Title]
57. #50 OR #51 OR #52 OR #53 OR #54 OR #55 OR #56
58. "animals"[MeSH Terms]
59. "humans"[MeSH Terms]
60. #58 NOT #59
61. #57 NOT #60

62. #15 AND #43 AND #49 AND #61

Embase (via Elsevier)

#	Searches
1.	'depression'/exp
2.	'suicidal behavior'/exp
3.	depression:ti,ab
4.	depressive:ti,ab
5.	depressed:ti,ab
6.	mdd:ti,ab
7.	suicide:ti,ab
8.	suicidal*:ti,ab
9.	'self-harm*':ti,ab
10.	antisuicidal:ti,ab
11.	'self-injur*':ti,ab
12.	sib:ti,ab
13.	'self-inflict*':ti,ab
14.	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13
15.	'narcotic analgesic agent'/exp
16.	opioid*:ti,ab
17.	opiate*:ti,ab
18.	narcotic*:ti,ab
19.	buprenorphine:ti,ab
20.	buprenex:ti,ab
21.	subutex:ti,ab
22.	zubsolv:ti,ab
23.	probuphine:ti,ab
24.	sublocade:ti,ab
25.	belbuca:ti,ab
26.	butrans:ti,ab
27.	sixmo:ti,ab
28.	bunavail:ti,ab

29. buvidal:ti,ab
30. temgesic:ti,ab
31. norspan:ti,ab
32. bup:ti,ab
33. 'alks 5461':ti,ab
34. methadone:ti,ab
35. dolophine:ti,ab
36. methadose:ti,ab
37. metadol:ti,ab
38. metadon:ti,ab
39. physeptone:ti,ab
40. diskets:ti,ab
41. suboxone:ti,ab
42. #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40 OR #41
43. 'drug therapy'/exp
44. treat*:ti,ab
45. therap*:ti,ab
46. pharmacotherap*:ti,ab
47. 'pharmacologic* intervention*':ti,ab
48. #43 OR #44 OR #45 OR #46 OR #47
49. 'randomized controlled trial'/exp
50. 'controlled clinical trial'/de
51. random*:ti,ab,tt
52. 'randomization'/de
53. 'intermethod comparison'/de
54. placebo:ti,ab,tt
55. compare:ti,tt OR compared:ti,tt OR comparison:ti,tt
56. (evaluated:ab OR evaluate:ab OR evaluating:ab OR assessed:ab OR assess:ab) AND (compare:ab OR compared:ab OR comparing:ab OR comparison:ab)
57. (open NEXT/1 label):ti,ab,tt
58. ((double OR single OR doubly OR singly) NEXT/1 (blind OR blinded OR blindly)):ti,ab,tt
59. 'double blind procedure'/de

60. (parallel NEXT/1 group*):ti,ab,tt
61. crossover:ti,ab,tt OR 'cross over':ti,ab,tt
62. ((assign* OR match OR matched OR allocation) NEAR/6 (alternate OR group OR groups OR intervention OR interventions OR patient OR patients OR subject OR subjects OR participant OR participants)):ti,ab,tt
63. assigned:ti,ab,tt OR allocated:ti,ab,tt
64. (controlled NEAR/8 (study OR design OR trial)):ti,ab,tt
65. volunteer:ti,ab,tt OR volunteers:ti,ab,tt
66. 'human experiment'/de
67. trial:ti,tt
68. #49 OR #50 OR #51 OR #52 OR #53 OR #54 OR #55 OR #56 OR #57 OR #58 OR #59 OR #60 OR #61 OR #62 OR #63 OR #64 OR #65 OR #66 OR #67
69. ((random* NEXT/1 sampl* NEAR/8 ('cross section*' OR questionnaire* OR survey OR surveys OR database OR databases)):ti,ab,tt) NOT ('comparative study'/de OR 'controlled study'/de OR 'randomised controlled':ti,ab,tt OR 'randomized controlled':ti,ab,tt OR 'randomly assigned':ti,ab,tt)
70. 'cross-sectional study' NOT ('randomized controlled trial'/exp OR 'controlled clinical trial'/de OR 'controlled study'/de OR 'randomised controlled':ti,ab,tt OR 'randomized controlled':ti,ab,tt OR 'control group':ti,ab,tt OR 'control groups':ti,ab,tt)
71. 'case control*':ti,ab,tt AND random*:ti,ab,tt NOT ('randomised controlled':ti,ab,tt OR 'randomized controlled':ti,ab,tt)
72. 'systematic review':ti,tt NOT (trial:ti,tt OR study:ti,tt)
73. nonrandom*:ti,ab,tt NOT random*:ti,ab,tt
74. 'random field*':ti,ab,tt
75. ('random cluster' NEAR/4 sampl*):ti,ab,tt
76. review:ab AND review:it NOT trial:ti,tt
77. 'we searched':ab AND (review:ti,tt OR review:it)
78. 'update review':ab
79. (databases NEAR/5 searched):ab
80. (rat:ti,tt OR rats:ti,tt OR mouse:ti,tt OR mice:ti,tt OR swine:ti,tt OR porcine:ti,tt OR murine:ti,tt OR sheep:ti,tt OR lambs:ti,tt OR pigs:ti,tt OR piglets:ti,tt OR rabbit:ti,tt OR rabbits:ti,tt OR cat:ti,tt OR cats:ti,tt OR dog:ti,tt OR dogs:ti,tt OR cattle:ti,tt OR bovine:ti,tt OR monkey:ti,tt OR monkeys:ti,tt OR trout:ti,tt OR marmoset*:ti,tt) AND 'animal experiment'/de
81. 'animal experiment'/de NOT ('human experiment'/de OR 'human'/de)

82. #69 OR #70 OR #71 OR #72 OR #73 OR #74 OR #75 OR #76 OR #77 OR #78 OR #79 OR #80 OR #81
83. #68 NOT #82
84. #14 AND #42 AND #48 AND #83

Web of Science

#	Searches
1.	TS=(depression)
2.	TS=(depressive)
3.	TS=(depressed)
4.	TS=(MDD)
5.	TS=(suicide)
6.	TS=(suicidal*)
7.	TS=("self-harm*")
8.	TS=(antisuicidal)
9.	TS=("self-injur*")
10.	TS=(SIB)
11.	TS=("self-inflict*")
12.	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11
13.	TS=(Opioid*)
14.	TS=(opiate*)
15.	TS=(narcotic*)
16.	TS=(buprenorphine)
17.	TS=(Buprenex)
18.	TS=(Subutex)
19.	TS=(Zubsolv)
20.	TS=(Probuphine)
21.	TS=Sublocade)
22.	TS=(Belbuca)
23.	TS=(Butrans)
24.	TS=(Sixmo)
25.	TS=(Bunavail)

26. TS=(Buvidal)
27. TS=(Temgesic)
28. TS=(Norspan)
29. TS=(BUP)
30. TS=(ALKS-5461)
31. TS=(methadone)
32. TS=(Dolophine)
33. TS=(methadose)
34. TS=(Metadol)
35. TS=(Metadon)
36. TS=(Physeptone)
37. TS=(Diskets)
38. TS=(Suboxone)
39. #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38
40. TS=(treat*)
41. TS=(therap*)
42. TS=(pharmacotherap*)
43. TS=("pharmacologic* intervention*")
44. #40 OR #41 OR #42 OR #43
45. TS=(random*)
46. TS=(placebo)
47. TI=(compare OR compared OR comparison)
48. AB=((evaluated OR evaluate OR evaluating OR assessed OR assess) AND (compare OR compared OR comparing OR comparison))
49. TS=(open NEAR/1 label)
50. TS=((double OR single OR doubly OR singly) NEAR/1 (blind OR blinded OR blindly))
51. TS=(parallel NEAR/1 group*)
52. TS=(crossover OR "cross over")
53. TS=((assign* OR match OR matched OR allocation) NEAR/6 (alternate OR group OR groups OR intervention OR interventions OR patient OR patients OR subject OR subjects OR participant OR participants))
54. TS=(assigned OR allocated)

55. TS=(controlled NEAR/8 (study OR design OR trial))
56. TS=(volunteer OR volunteers)
57. TI=(trial)
58. #45 OR #46 OR #47 OR #48 OR #49 OR #50 OR #51 OR #52 OR #53 OR #54 OR #55 OR #56 OR #57
59. TS=((random* NEAR/1 sampl* NEAR/8 ("cross section*" OR questionnaire* OR survey OR surveys OR database or databases)) NOT ("randomised controlled" OR "randomized controlled" OR "randomly assigned"))
60. TS=("case control*" AND random* NOT ("randomised controlled" OR "randomized controlled"))
61. TI=("systematic review" NOT (trial OR study))
62. TS=(nonrandom* NOT random*)
63. TS=("random field*")
64. TS=("random cluster" NEAR/4 sampl*)
65. AB=(review) NOT TI=(trial)
66. AB=("we searched") AND TI=(review)
67. AB=("update review")
68. AB=(databases NEAR/5 searched)
69. #59 OR #60 OR #61 OR #62 OR #63 OR #64 OR #65 OR #66 OR #67 OR #68
70. #58 NOT #69
71. #12 AND #39 AND #44 AND #70

CENTRAL – The Cochrane Library

#	Searches
1.	MeSH descriptor: [Self-Injurious Behavior] explode all trees
2.	MeSH descriptor: [Depressive Disorder] explode all trees
3.	MeSH descriptor: [Depression] explode all trees
4.	(depression):ti,ab,kw
5.	(depressive):ti,ab,kw
6.	(depressed):ti,ab,kw
7.	(MDD):ti,ab,kw
8.	(suicide):ti,ab,kw

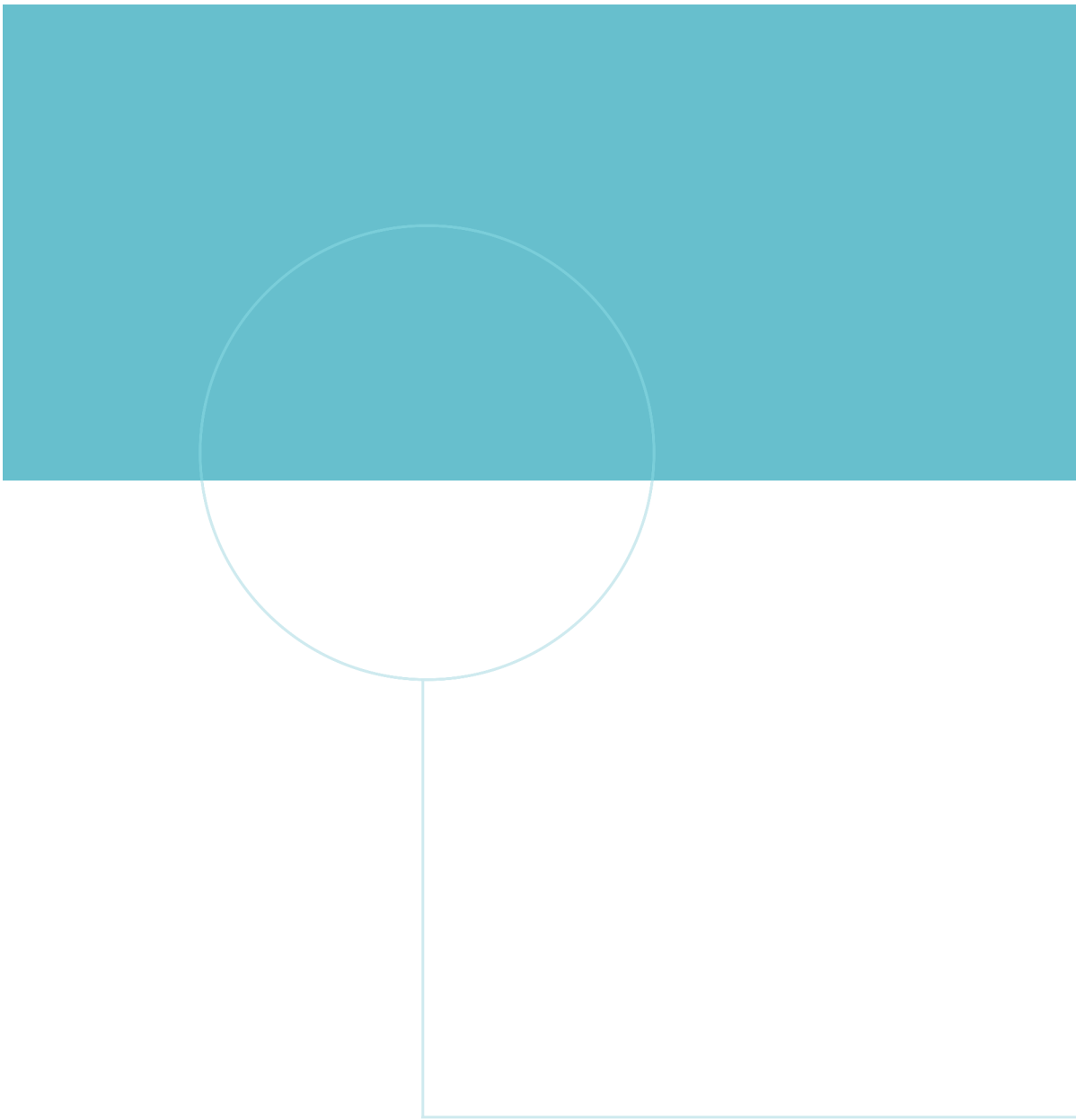
9. (suicidal*):ti,ab,kw
10. (“self-harm*”):ti,ab,kw
11. (antisuicidal):ti,ab,kw
12. (“self-injur*”):ti,ab,kw
13. (SIB):ti,ab,kw
14. (“self-inflict*”):ti,ab,kw
15. #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14
16. MeSH descriptor: [Drug Therapy] explode all trees
17. (treat*):ti,ab,kw
18. (therap*):ti,ab,kw
19. (pharmacotherap*):ti,ab,kw
20. (“pharmacologic* intervention*”):ti,ab,kw
21. #16 OR #17 OR #18 OR #19 OR #20
22. MeSH descriptor: [Analgesics, Opioid] explode all trees
23. (Opioid*):ti,ab,kw
24. (opiate*):ti,ab,kw
25. (narcotic*):ti,ab,kw
26. (buprenorphine):ti,ab,kw
27. (Buprenex):ti,ab,kw
28. (Subutex):ti,ab,kw
29. (Zubsolv):ti,ab,kw
30. (Probuphine):ti,ab,kw
31. (Sublocade):ti,ab,kw
32. (Belbuca):ti,ab,kw
33. (Butrans):ti,ab,kw
34. (Sixmo):ti,ab,kw
35. (Bunavail):ti,ab,kw
36. (Buvidal):ti,ab,kw
37. (Temgesic):ti,ab,kw
38. (Norspan):ti,ab,kw
39. (BUP):ti,ab,kw
40. (ALKS-5461):ti,ab,kw
41. (methadone):ti,ab,kw

42. (Dolophine):ti,ab,kw
43. (methadose):ti,ab,kw
44. (Metadol):ti,ab,kw
45. (Metadon):ti,ab,kw
46. (Physeptone):ti,ab,kw
47. (Diskets):ti,ab,kw
48. (Suboxone):ti,ab,kw
49. #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40 OR #41 OR #42 OR #43 OR #44 OR #45 OR #46 OR #47 OR #48
50. #15 AND #21 AND #49

Clinicaltrials.gov

Searches (Interventional study type)

(depression OR depressive OR depressed OR MDD OR suicide OR suicidal* OR "self-harm*" OR antisuicidal OR "self-injur*" OR SIB OR "self-inflict*") AND ((Opioid* OR opiate* OR narcotic* OR buprenorphine OR Buprenex OR Subutex OR Zubsolv OR Probuphine OR Sublocade OR Belbuca OR Butrans OR Sixmo OR Bunavail OR Buvidal OR Temgesic OR Norspan OR BUP OR ALKS-5461 OR methadone OR Dolophine OR methadose OR Metadol OR Metadon OR Physeptone OR Diskets OR Suboxone) AND (treat* OR therap* OR pharmacotherap* OR "pharmacologic* intervention*"))



NTNU

Norwegian University of
Science and Technology