BMJ Mental Health

Reappraising lithium, triiodothyronine and second-generation antipsychotic augmentation treatment for major depression: a systematic review and meta-analysis with bias-adjustment analyses

Tien-Wei Hsu [1,2,3] Chih-Wei Hsu [4] Trevor Thompson,[5] Andre F Carvalho,[6] Brendon Stubbs,[7,8,9] Ping-Tao Tseng,[10,11,12,13] Fu-Chi Yang,[14] Chia-Ling Yu,[15] Yu-Kang Tu,[16,17] Chih-Sung Liang [18,19]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjment-2026-302656).

For numbered affiliations see end of article.

Correspondence to
Dr Chih-Sung Liang; lcsyfw@gmail.com
Professor Yu-Kang Tu; yukangtu@ntu.edu.tw

T-WH and C-WH are joint first authors.
Y-KT and C-SL are joint senior authors.

Received 2 March 2026
Accepted 10 July 2026

To cite: Hsu T-W, Hsu C-W, Thompson T, *et al*. *BMJ Ment Health* 2026;**29**:1–9.

ABSTRACT

Background Pharmacological augmentation is commonly used for patients with major depressive disorder (MDD) who respond inadequately to antidepressant treatment. However, the robustness of evidence supporting lithium, second-generation antipsychotics (SGAs), and triiodothyronine (T3) augmentation for MDD remains uncertain.

Objective To reappraise the efficacy and robustness of evidence for pharmacological augmentation strategies for MDD, focusing on lithium, SGAs, and T3. Study selection and analysis: We systematically searched five electronic databases and included placebo-controlled RCTs in adults with MDD who received augmentation with lithium, SGAs, or T3. The primary outcome was response rate (defined as ≥50% depressive symptom reduction). Random-effects meta-analysis and trial sequential analysis (TSA) were conducted. We also performed publication-bias–adjustment analyses, including PET-PEESE and selection models. Subgroup analyses were conducted for individual SGAs.
: To reappraise the efficacy and robustness of evidence for pharmacological augmentation strategies for MDD, focusing on lithium, SGAs, and T3.

Findings Fifty-six RCTs were included (n=13616). SGA augmentation was associated with a higher response than placebo (k=45; reported as OR with 95% CI: 1.53; 1.42–1.65; I^2=0%), and the evidence was supported by TSA. Conversely, while lithium showed benefit in conventional meta-analysis (k=17; OR 2.06; 1.30–3.27; I^2=11.5%), this effect was not supported by TSA (accrued information size reached 8% of the required information size; 1.99, TSA-adjusted 95% CI 0.02–189.11) and became statistically non-significant in the PET-PEESE-adjusted and selection models. Additionally, T3 augmentation was not associated with significantly higher response than controls (k = 6; 1.19; 0.53–2.70). Among individual SGAs, only aripiprazole, quetiapine, brexpiprazole, and cariprazine demonstrated TSA-supported efficacy, whereas evidence for other SGAs was limited or inconclusive.

Conclusion The certainty and robustness of evidence for lithium and T3 augmentation remained limited. Evidence for T3 was sparse and did not show a significant advantage over control, while interpretation of the lithium findings is further constrained by the fact that most trials were conducted before treatment-resistant depression was more operationally defined in contemporary research. SGAs with TSA-supported benefits (aripiprazole, quetiapine, brexpiprazole and cariprazine) may be prioritised, pending individual patient considerations.

Study registration https://osf.io/27gp9

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Lithium, second-generation antipsychotics (SGAs) and triiodothyronine (T3) are widely used augmentation options for major depressive disorder when response to antidepressants is inadequate, and multiple guidelines have historically recommended them. However, the placebo-controlled evidence base for lithium and T3 largely comes from older, small trials with short durations, raising concerns about the certainty and robustness of their estimated effects.

WHAT THIS STUDY ADDS

⇒ Our findings suggested that SGA augmentation showed a robust response benefit supported by trial sequential analysis (TSA), whereas lithium's apparent benefit in conventional meta-analysis was not supported by TSA and became non-significant after small-study/publication-bias adjustments. T3 augmentation was not associated with a significantly higher response than placebo, and only aripiprazole, quetiapine, brexpiprazole and cariprazine had TSA-supported efficacy among individual SGAs.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Clinically, these findings support prioritising SGAs with TSA-confirmed benefits while interpreting lithium augmentation effects cautiously when evidence is drawn mainly from small, older placebo-controlled trials. For research and guidelines, the results highlight the need for larger, contemporary placebo-controlled trials to clarify the efficacy of lithium and less-studied SGAs and to strengthen the evidence base for augmentation decisions.

INTRODUCTION

Major depressive disorder (MDD) is a widespread chronic and recurrent condition with an estimated annual prevalence of 6% and a lifetime prevalence of 15%.[1 2] It is recognised as the leading cause of global disability, imposing substantial social and economic costs.[3 4] Despite the availability of antidepressant treatment, up to one-third of patients failed to achieve remission after at least two adequate trials of antidepressant therapy.[5] This subgroup is often operationalised as treatment-resistant depression (TRD); however, definitions vary across studies.[5] Individuals with inadequate antidepressant response tend to have greater functional impairment, higher healthcare utilisation and increased psychiatric/medical comorbidity than those who respond to first-line treatment.[6–8]

For patients with MDD who respond inadequately to an antidepressant, augmentation is a common next step: clinicians maintain the current antidepressant and add an agent with a complementary mechanism to increase the likelihood of response or remission.[9] Multiple guidelines have historically recommended lithium and second-generation antipsychotics (SGAs) as first-line augmentation choices.[10–12] A previous meta-analysis reported that lithium might be more effective than SGAs as an augmentation treatment for TRD, with a lower number needed to treat (NNT) (lithium NNT=5 vs SGA NNT=11).[13] Another review also described lithium as a cost-effective strategy with a well-established efficacy profile.[14] However, more recent comparative evidence has been less favourable to lithium. In the pragmatic lithium versus quetiapine in depression (LQD) randomised controlled trial, quetiapine augmentation was found to be more clinically effective than lithium augmentation for reducing depressive symptoms over 12 months and was also likely to be more cost-effective.[15] In parallel, triiodothyronine (T3) represents another common augmentation strategy, theoretically acting by modulating the hypothalamic-pituitary-thyroid axis to enhance monoaminergic transmission. The STAR*D trial suggested that T3 was comparable to lithium in efficacy and possessed a favourable tolerability profile. Notably, the CANMAT 2023 guideline recommended both T3 and lithium as second-line augmentation strategies, placing them after SGA brexpiprazole and aripiprazole. However, the evidentiary foundation supporting lithium and T3 augmentation largely derives from small, often single-centre randomised trials conducted two to three decades ago, frequently enrolling only 10–30 participants per arm and using relatively short treatment durations.[16–18] Consequently, the robustness of evidence for lithium and T3 remains uncertain.

Therefore, we aimed to reappraise the efficacy and certainty of evidence for lithium, T3 and SGA augmentation for MDD. In addition to conventional random-effects meta-analysis, we integrated trial sequential analysis (TSA) to evaluate the strength and reliability of the accumulating evidence and to assess whether available information size is sufficient for firm inferences. TSA applies principles similar to interim monitoring in a single randomised trial to cumulative meta-analysis: it estimates the amount of evidence required to support a reliable conclusion and evaluates whether the cumulative evidence has crossed boundaries for benefit, futility or inconclusiveness. This approach is particularly relevant when evidence is based on repeated testing of small trials because conventional meta-analysis may produce apparently significant findings before the accumulated sample size is sufficient. To further examine the sensitivity of findings to small-study effects and potential publication bias, we conducted bias-adjustment analyses using precision-effect test and precision-effect estimate with SE (PET-PEESE) and weight-function selection models.

METHODS

Prior to data analysis, we registered our protocol on Open Science Framework (https://osf.io/27gp9). The study adhered to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA)[19] statement to ensure transparent and comprehensive reporting (online supplemental appendix 1).

Eligibility criteria

We included randomised controlled trials (RCTs) enrolling adults with MDD who had an inadequate response to antidepressant therapy in the current episode, typically after one to three adequate courses and who received augmentation with lithium, SGAs or T3. The included studies were required to compare pharmacological augmentation strategies against placebo controls. To ensure the isolated efficacy of pharmacological augmentation, we strictly included only placebo-controlled RCTs. Eligible studies were also required to use validated clinician-rated scales to assess depressive symptoms, such as the 17-Item Hamilton Depression Rating Scale or the Montgomery-Åsberg Depression Rating Scale. For studies involving olanzapine, we accounted for the distinct design of olanzapine-fluoxetine combination (OFC) trials. In these protocols, olanzapine augmentation is evaluated by comparing OFC against fluoxetine monotherapy or fluoxetine plus placebo, rather than a traditional antidepressant plus olanzapine versus antidepressant plus placebo design. The exclusion criteria were: (1) studies without specific measurements for clinical response; (2) studies targeting non-TRD populations; (3) studies comparing interventions to treatment-as-usual or waitlist conditions and (4) studies not published in peer-reviewed journals.

Data sources and search

Two reviewers independently searched major databases, including MEDLINE, EMBASE, PsycInfo, the Cochrane Central Register of Controlled Trials (CENTRAL) and PubMed, from their inception to 25 December 2025, without language restrictions. Search terms included (depress* OR MDD OR major depress*) AND (resistan* OR refractor* OR non-respon* OR nonrespon* OR TRD OR fail* OR inadequate OR difficult) AND (combin* OR co-administ* OR augment* OR adjunct* OR add-on) AND (thyroid hormone OR T3 OR Triiodothyronine OR Lithium OR aripiprazole OR brexpiprazole OR cariprazine OR olanzapine OR risperidone OR quetiapine OR ziprasidone OR Lurasidone OR antipsychotic OR dopamine D2 receptors). We also reviewed the reference lists of the included studies and related systematic reviews to identify additional relevant citations.

Study selection

Two reviewers independently screened titles, abstracts and full-text articles. Disagreements were resolved through discussion and, if necessary, by consulting the corresponding authors. Online supplemental appendix 2 demonstrates the complete search strategies and online supplemental appendix 3 shows the reasons for exclusion.

Data extraction and outcome definition

Two reviewers independently extracted the data, and discrepancies were resolved by consensus or by consulting the corresponding authors. The primary outcome was the response rate, defined as the proportion of patients achieving a 50% reduction

in standardised depression scale scores from baseline to the study endpoint.

Quality assessment

The risk of bias (RoB) was assessed using the Cochrane risk-of-bias tool 2 for randomised trials. This tool evaluates five domains of potential bias: bias arising from the randomisation process, bias due to deviations from intended interventions, bias due to missing outcome data, bias in outcome measurement and bias in selection of reported results. Overall RoB was judged as low if all domains were rated low; as some concerns if at least one domain raised some concerns but none were high and as high if any domain was rated high or if multiple domains raised some concerns. Two reviewers independently conducted this assessment, and discrepancies were resolved by discussion or consulting the corresponding authors.

Data synthesis

We performed a random-effects meta-analysis to calculate the pooled OR and 95% CIs for the response rates of the three augmentation strategies. The DerSimonian-Laird method was used to estimate the between-study variance. To assess the potential range of effects in a future similar clinical setting, we calculated 95% prediction intervals (PI) for each treatment. Statistical heterogeneity was quantified using the I^2 statistic and the Cochran's Q test. For multiarm trials, each eligible active arm was treated as a separate comparison against the shared control group, with the control sample size and control events divided proportionally across these comparisons to avoid double-counting placebo participants; this split-control unit was used consistently across all analyses.

To evaluate the robustness of the cumulative evidence and control for the risk of false-positive (type I error) or false-negative (type II error) results, we conducted a TSA. TSA first estimates the required information size (RIS), which represents the meta-analytic sample size needed to detect or reject a prespecified intervention effect with adequate statistical power. It then compares the cumulative Z-curve from the accumulating trials with sequential monitoring boundaries. If the Z-curve crosses the benefit boundary, the evidence is considered sufficient to support a conclusive benefit; if it crosses the futility boundary, the available evidence suggests that the prespecified benefit is unlikely; if neither boundary is crossed and the RIS has not been reached, the evidence remains inconclusive. The RIS was estimated based on an anticipated OR=1.50, assuming a control group response rate of 30%. The anticipated effect was derived from the externally pooled SGA effect estimate[13] and the contemporaneous head-to-head comparator evidence[15 20] (quetiapine vs lithium augmentation). The statistical thresholds were established with a two-sided alpha of 5% and a beta of 10%. The RIS was further adjusted based on the estimated heterogeneity. We applied O'Brien-Fleming alpha-spending and beta-spending functions to determine the sequential monitoring boundaries. For the futility boundaries, a non-binding approach was implemented to maintain flexibility in interpreting negative results. TSA was also performed for individual SGAs to evaluate the evidentiary strength of specific pharmacological agents.

For lithium and SGA, we employed the PET and PEESE models to evaluate whether the observed clinical benefits remained significant after adjusting for small-study effects. Furthermore, weight-function selection models, including both two-step and four-step selection approaches with bootstrap CIs, were implemented to rigorously assess the impact of potential selective reporting based on p value thresholds. These analyses aimed to ensure the robustness of our effect size estimates against the risk of overestimation inherent in traditional meta-analytic methods.

We conducted five sensitivity analyses. (1) Because lithium trials almost exclusively enrolled patients after a single failed antidepressant trial, whereas SGA trials enrolled patients with one or multiple prior failures, we performed a subgroup analysis of SGA by number of prior treatment failures, online supplemental appendix 5. (2) For lithium augmentation, we repeated the meta-analysis after excluding the seven trials judged to be at high overall RoB, online supplemental appendix 6. (3) We repeated the analysis excluding olanzapine (OFC) augmentation studies, whose comparator design differs from that of other SGA trials, online supplemental appendix 7. (4) We re-ran the lithium TSA across a range of anticipated intervention effects (OR 1.5–2.0) to test the sensitivity of the required information size to the chosen anchor, online supplemental appendix 8. (5) We fitted an exploratory univariable meta-regression of publication year within the lithium subgroup, online supplemental appendix 9. All analyses were conducted using RStudio (V.4.3.0) with *meta* and *RTSA* packages. The full analysis code used to reproduce the main analyses is provided in the supplementary data (online supplemental appendix 10).

Patient and public involvement

Patients and the public were not involved in the design, conduct, reporting or dissemination plans of this study because it was a systematic review and meta-analysis of aggregated, de-identified clinical trial data. We plan to disseminate the findings to clinicians and lay audiences through academic presentations, lectures and social media.

Dissemination to participants and related patient and public communities

Dissemination of the work to the public and clinical community through social media and lectures is planned.

RESULTS

Study characteristics

Our searches resulted in 2439 potentially relevant citations (online supplemental figure 1). The complete search strategy and reasons for the exclusion of certain studies can be found in online supplemental appendices 2, 3. After removing duplicates and added head-search results, we included 56 RCTs (online supplemental appendix 4). Online supplemental tables 1–3 show the characteristics of the included lithium, SGA and T3 studies.

Across 17 lithium augmentation RCTs, a total of 631 participants were included (lithium, n=297; placebo, n=334), with a mean age of 43.3 years (SD 4.5) and 62.4% (14.2%) female participants. Across the 35 SGA augmentation RCTs, a total of 7450 participants were allocated to the SGA augmentation arms and 5294 participants to placebo. The mean participant age was 44.4 (3.7) years, and 66.1% (9.5%) of participants were female. Across five T3 augmentation RCTs, a total of 98 participants were allocated to the T3 augmentation arm and 143 participants to placebo. The mean participant age was 44.0 (3.3) years, and 63.0% (9.8%) of participants were female.

Quality of evidence

Overall RoB patterns differed across augmentation strategies (see online supplemental figures 2–4). For lithium augmentation trials, no study was rated low overall risk; 10/17 had some concerns and 7/17 were judged high overall risk (online

Table 1 Summary of the random-effects meta-analysis of the three augmentation strategies

Random-effects meta-analysis						
Treatment	***k***	**OR (95% CI)**	**95%PI**	**τ^2 (τ)**	**I^2**	**P value**
SGA	45	1.53 (1.42 to 1.65)	1.42–1.65	0.0000 (0.0000)	0.0%	<0.001
Lithium	17	2.06 (1.30 to 3.27)	0.95–4.47	0.0868 (0.2946)	11.5%	0.0043
T3	6	1.19 (0.53 to 2.70)	0.42–3.36	0.0614 (0.2478)	9.5%	0.6000

PI, prediction interval; SGA, second-generation antipsychotic; T3, triiodothyronine.;

supplemental figure 2). For SGA augmentation trials, most studies were rated low risk or some concerns, with only 1/35 judged high overall risk (online supplemental figure 3). All five T3 augmentation trials were rated as having some concerns overall (online supplemental figure 4).

Meta-analyses

Meta-analyses (table 1) showed that SGA augmentation was associated with significantly higher odds of response compared with control (k=45; OR 1.53, 95% CI 1.42 to 1.65; p<0.001) with no observed heterogeneity (τ^2 = 0 and I^2 = 0%). Lithium augmentation also demonstrated a significant benefit (k=17; OR 2.06, 95% CI 1.30 to 3.27; p=0.0043) with low heterogeneity (τ^2 = 0.09 and I^2 = 11.5%) compared with the control condition. T3 augmentation did not show a significantly different odds of response compared (k=6; OR1.19, 95% CI 0.53 to 2.70; p=0.600) with controls with low heterogeneity ($I\tau^2$ = 0.06 and ^{2} = 9.5%). The detailed forest plots for each meta-analysis were presented on online supplemental figures 5–7.

Effect size versus sample size

Figure 1 plots study-level effect size (log OR) against total sample size (log scale) across the three augmentation strategies. SGA trials were predominantly larger and clustered around modest positive effects, with most estimates lying above the null line and showing relatively limited dispersion across sample sizes. In contrast, lithium trials and T3 trials were generally smaller and displayed substantially greater variability, including several small studies with very large positive effects as well as studies near or below the null, suggesting a pronounced small-study pattern.

Trial sequential analysis

Figure 2 presents retrospective TSA for lithium (panel A) and SGA augmentation (panel B). For lithium, although the cumulative Z-curve crossed the conventional significance threshold, it did not cross the TSA monitoring boundary for benefit, and the accrued information was low (≈8% of the required information size; 631 of 8190). Accordingly, the TSA-adjusted CI was wide and included the null (OR 1.99, TSA-adjusted 95% CI 0.02 to 189.11), indicating that the current evidence remains inconclusive after adjusting for random errors and repeated testing. In contrast, for SGA augmentation, the cumulative Z-curve crossed the TSA boundary for benefit early and remained above it. The accrued information exceeded the required information size (accrued information siz (AIS) ≈12,856; >100% of RIS), with no heterogeneity (τ^2=0; I^2=0%). These TSA results support firm evidence for the efficacy of SGA augmentation (pooled OR 1.53, 95% CI 1.42 to 1.66). For T3 augmentation,

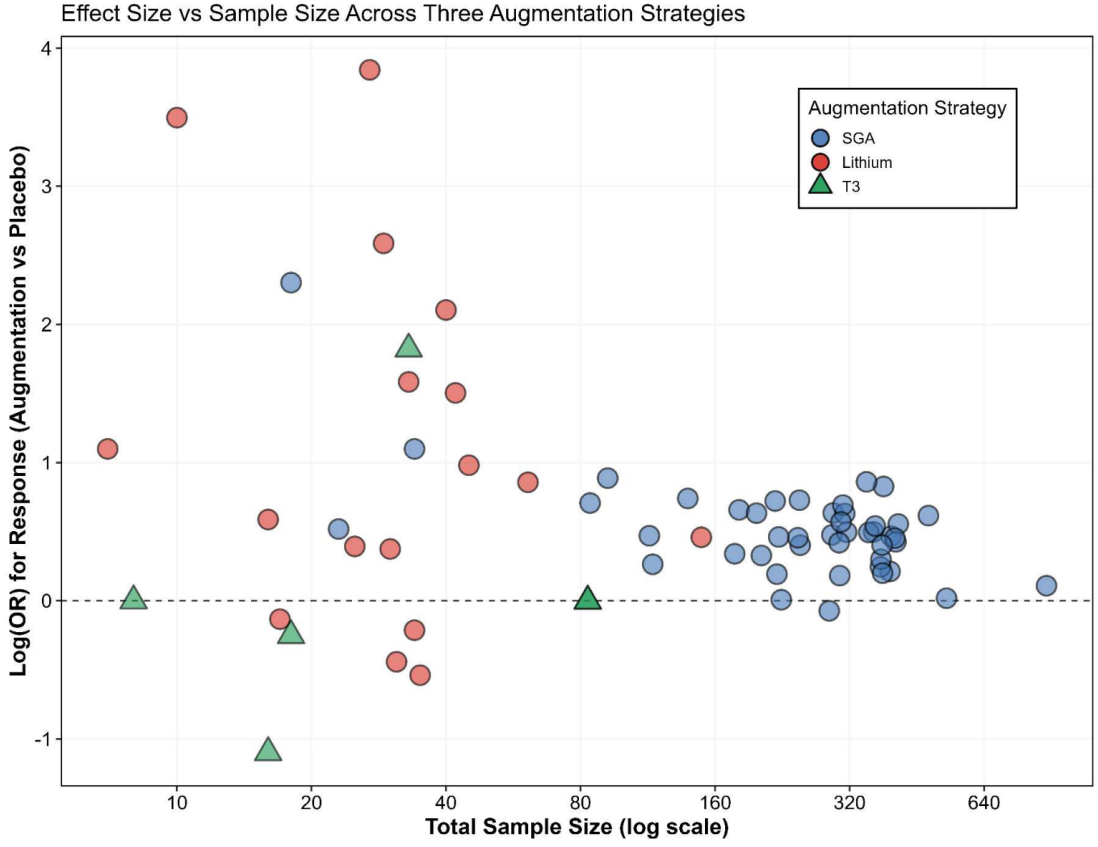

Figure 1 Effect size versus sample size in the three augmentation strategies. SGA, second-generation antipsychotic; T3, triiodothyronine.

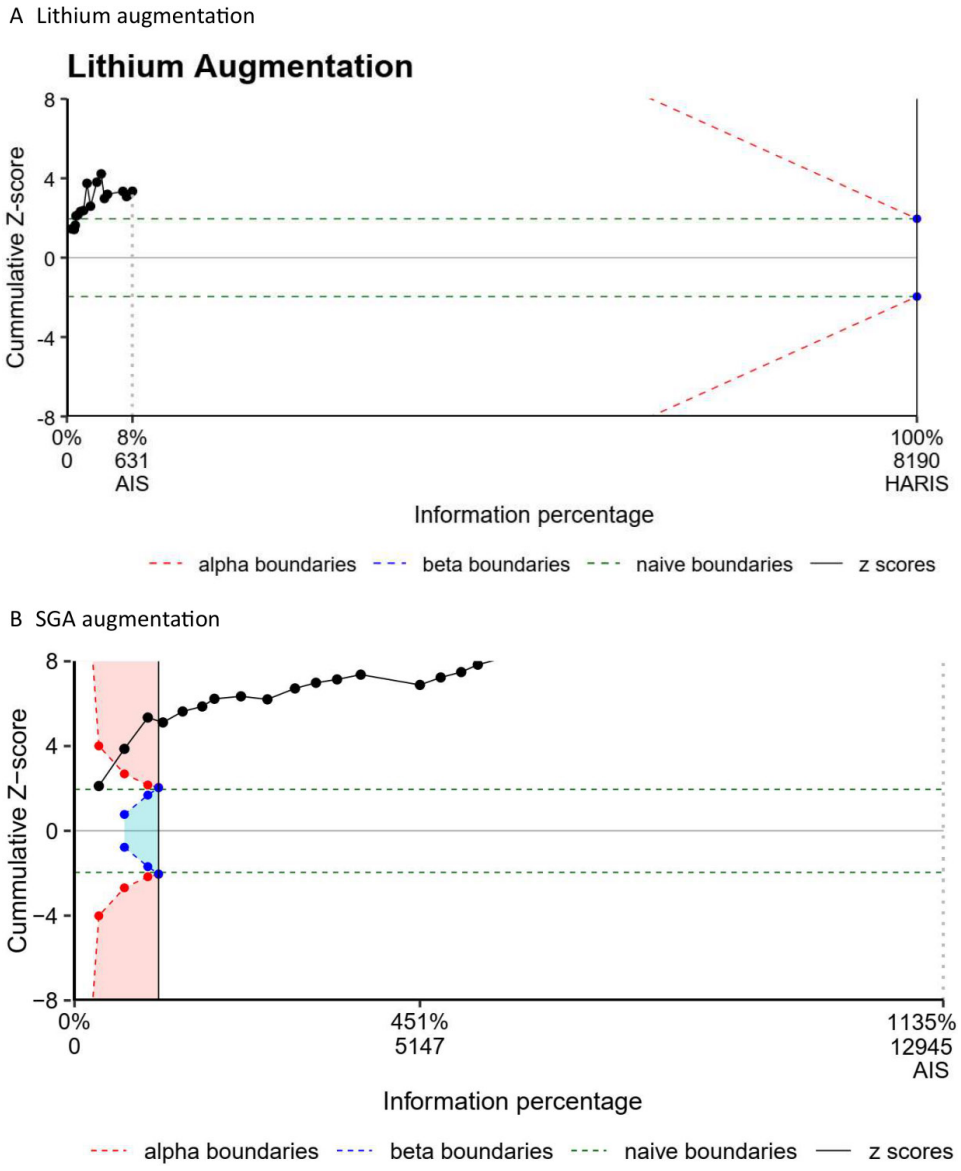

Figure 2 Trial sequential analysis for lithium and SGA augmentation for MDD. (A) Lithium augmentation, (B) SGA augmentation. MDD, major depressive disorder SGA, second-generation antipsychotic. AIS, accrued information size, HARIS, heterogeneity-adjusted required information size

the cumulative Z-curve crossed neither the TSA monitoring boundary for benefit nor the futility boundary, with accrued information reaching approximately 4% of the heterogeneity-adjusted required information size and a TSA-adjusted CI of 0.00–2844.95 (online supplemental figure 8).

Publication bias adjustment

Figure 3 summarises publication bias-adjusted estimates for lithium and SGA augmentation. In panel A, the standard random-effects model indicated significant benefits for both strategies (lithium: OR 2.06, 95% CI 1.30 to 3.27; SGA: OR 1.53, 95% CI 1.42 to 1.65). After small-study adjustment, the lithium effect was attenuated and no longer statistically significant under PET (OR 1.08, 95% CI 0.49 to 2.41) and PEESE (OR 1.42, 95% CI 0.86 to 2.34). For SGA, PET similarly reduced the estimate toward the null (OR 1.12, 95% CI 0.87 to 1.44), whereas PEESE remained statistically significant (OR 1.40, 95% CI 1.26 to 1.57).

Panel B shows selection-model analyses with bootstrap CIs. For lithium, selection models yielded smaller and non-significant effects (two-step: OR 1.51, 95% CI 0.98 to 2.52; four-step: OR 1.10, 95% CI 0.72 to 1.81), consistent with potential publication bias or small-study effects driving the larger random-effects estimate. In contrast, SGA effects were comparatively robust: the two-step selection model closely matched the random-effects estimate (OR 1.50, 95% CI 1.29 to 1.78), and the four-step model remained significant although attenuated (OR 1.27, 95% CI 1.06 to 1.62).

Subgroup analyses for individual SGA

Across individual SGAs (online supplemental table 4), aripiprazole (k=8, OR 1.85, 95% CI 1.54 to 2.21; 95% PI 1.49–2.29; I^2=0%; p<0.001), risperidone (k=4, OR 2.09, 95% CI 1.78 to 2.46; 95% PI 1.08–4.05; I^2=0%; p<0.001), quetiapine (k=6, OR 1.47, 95% CI 1.21 to 1.79; 95% PI 1.05–2.07; I^2=0%; p=0.004), brexpiprazole (k=10, OR 1.46, 95% CI 1.23 to 1.74; 95% PI 1.22–1.75; I^2=0%; p<0.001) and cariprazine (k=9, OR 1.33, 95% CI 1.18 to 1.51; 95% PI 1.11–1.60; I^2=0%; p<0.001) were associated with significantly higher odds of response compared with control. In contrast, olanzapine was

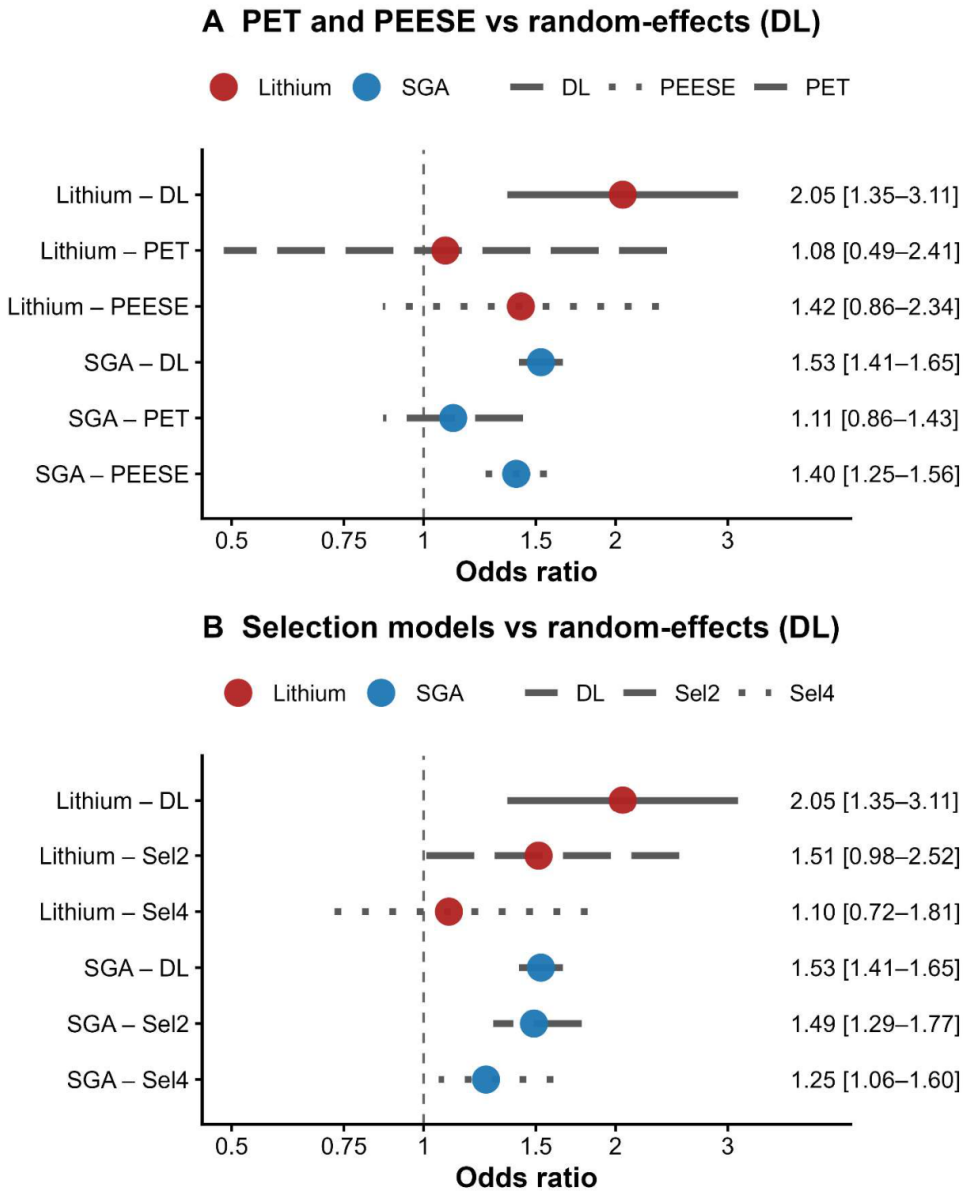

Figure 3 Bias-adjusted effect estimates for lithium and SGA augmentation. DL, DerSimonian-Laird; PET, precision-effect test; PEES, precision-effect estimatPrecision-Effect Test and Precision-Effect Estimate with SE; SGA, second-generation antipsychotic.

not significant (k=5, OR 1.42, 95% CI 0.83 to 2.43; p=0.142). Single-trial estimates were significant for lumateperone (OR 1.85, 95% CI 1.25 to 2.74; p=0.002) but not for ziprasidone (OR 2.10, 95% CI 0.98 to 4.50; p=0.057) or perospirone (OR 1.40, 95% CI 0.76 to 2.60; p=0.280). The forest plots for each SGA were presented on online supplemental figures 9–17.

TSA by individual SGA are summarised in online supplemental table 5. Evidence was TSA-confirmed (crossing the alpha monitoring boundary early) for aripiprazole (OR 1.85, 95% CI 1.55 to 2.21; AIS 219%), quetiapine (OR 1.47, 95% CI 1.11 to 1.94; AIS 111%), brexpiprazole (OR 1.46, 95% CI 1.23 to 1.74; AIS 189%) and cariprazine (OR 1.34, 95% CI 1.14 to 1.56; AIS 270%). For the remaining agents, no TSA monitoring boundary was crossed, and estimates were either non-significant or became imprecise after TSA adjustment, consistent with insufficient accrued information: olanzapine (OR 1.42, TSA-adjusted 95% CI 0.03 to 67.14; AIS 1%), ziprasidone (OR 2.10, TSA-adjusted 95% CI 0.00 to 5062.14; AIS 12%), lumateperone (OR 1.85, TSA-adjusted 95% CI 0.97 to 3.55; AIS 42%) and perospirone (OR 1.40, TSA-adjusted 95% CI 0.24 to 8.11; AIS 16%). Risperidone remained statistically significant with a Shao-Wei-adjusted CI (OR 2.10, 95% CI 1.39 to 3.14; AIS 37%), but did not cross a TSA monitoring boundary, indicating that the cumulative evidence has not yet reached TSA-defined conclusiveness. The TSA plots for each SGA were presented on online supplemental figures 18–26.

All five sensitivity analyses supported the primary findings: among SGA trials, the subgroup analysis comparing patients who had failed two versus one prior antidepressant trial showed no significant subgroup difference (p=0.44; online supplemental appendix 5); excluding lithium trials at high RoB yielded a similar lithium estimate (OR=2.04; 95% CI 0.98 to 4.25; online supplemental appendix 6); excluding olanzapine augmentation studies left the SGA estimate essentially unchanged (OR=1.54; 95% CI 1.43 to 1.66; online supplemental appendix 7); the lithium TSA remained inconclusive across anticipated effects of OR 1.5–1.8 (information fraction ≤78%), reaching the required information size only near the unadjusted estimate (≈OR 2.0;

online supplemental appendix 8) and publication year was not associated with the lithium effect (slope = −0.020 per year; p=0.41; R^2 = 0%; online supplemental appendix 9).

DISCUSSION

In this meta-analysis of 56 randomised controlled trials, we found that SGA augmentation was associated with significantly higher odds of response versus placebo with minimal heterogeneity, and TSA corroborated the benefit with the accrued information size exceeding the required information size. In contrast, although lithium augmentation was significant in conventional meta-analysis, TSA suggested the available information remained insufficient: the cumulative evidence did not cross the monitoring boundary for benefit, and the TSA-adjusted CI was wide and included the null. Bias-adjustment analyses for small-study effects and potential publication bias generally attenuated the lithium estimate toward the null, whereas SGA estimates were comparatively stable across adjustment approaches. T3 augmentation did not demonstrate significant efficacy. Among individual SGAs, aripiprazole, quetiapine, brexpiprazole and cariprazine had TSA-supported evidence, while evidence for other agents remained limited, underscoring meaningful differences in the strength of evidence across specific SGAs. Importantly, the present study primarily evaluates response efficacy and the robustness of response-efficacy evidence, rather than the full comparative benefit-risk profile of lithium, T3 and SGAs.

McIntyre *et al* noted[21] that the evidence base for lithium augmentation is better established in patients with partial response to tricyclic antidepressants (TCAs) and monoamine oxidase inhibitors (MAOIs), rather than being principally studied in rigorously defined TRD. Consistent with this, under our inclusion criteria, most placebo-controlled lithium RCTs were conducted before 2005, when consensus definitions of TRD were lacking and eligibility criteria were generally broader. This historical context may partly explain the relatively higher estimated efficacy observed for lithium augmentation in our analyses. Alternatively, after adjusting for small-study effects, the efficacy estimate for lithium augmentation in TRD was attenuated and no longer statistically significant, consistent with possible publication bias. Over the past two decades, RCTs of lithium augmentation have predominantly adopted head-to-head designs which did not meet our inclusion criteria. Among the nine head-to-head studies excluded during the screening process, two directly compared lithium augmentation with an SGA and both involved quetiapine, Bauer *et al*[20] conducted a non-inferiority trial comparing quetiapine augmentation with lithium augmentation for TRD and found no statistically significant difference in depressive symptom outcomes between the two treatments. Nonetheless, at week 6, quetiapine augmentation showed slightly higher response and remission rates than lithium augmentation (response: 52.4% vs 46.2%; remission: 31.9% vs 27.1%). Another large RCT conducted in the UK reported that, compared with lithium augmentation for TRD, quetiapine augmentation led to significantly lower depressive symptom severity and a lower overall (cumulative) burden of depression in the quetiapine group. The study also suggested that quetiapine augmentation is probably more cost-effective than lithium.[15] Taken together, these findings suggest that the apparent benefit of lithium augmentation in TRD may be largely driven by older placebo-controlled trials with broader case definitions, whereas contemporary comparative evidence tends to favour quetiapine (or at least shows no clear advantage for lithium) in more pragmatically defined TRD populations.

From a statistical perspective, the SGA evidence base is far more stable: trial numbers are substantially larger (k=45), show no evidence of heterogeneity (I^2=0%) and the cumulative evidence crosses TSA benefit boundaries with >100% of the required information size. In contrast, lithium trials are smaller (k=17), display a pronounced small-study pattern, fail to cross TSA monitoring boundaries (AIS~8%) and the apparent benefit is attenuated to non-significance after publication-bias/small-study adjustment—suggesting the pooled signal is more vulnerable to random error and selective reporting. For SGAs, PET yielded a non-significant estimate, whereas PEESE remained significant. Following the conventional PET-PEESE decision rule, we retained the PEESE estimate. Given the large number of trials, absence of heterogeneity and TSA-supported conclusiveness, the PET non-significance in this context is more likely to reflect over-correction in the presence of a true underlying effect than a genuine absence of efficacy. From a psychopharmacological perspective, the efficacy signal of SGA augmentation is biologically plausible because SGAs converge on serotonin-dopamine receptor mechanisms that can complement monoamine reuptake inhibition. For dopamine, agents with D2/D3 partial agonism (eg, aripiprazole, brexpiprazole, cariprazine) can stabilise dopaminergic tone—functioning as a functional antagonist when dopaminergic signalling is excessive and as an agonist when it is deficient—which has been proposed to be relevant to motivational and reward-related depressive symptoms.[22 23] For serotonin, many SGAs combine 5-HT2A antagonism with 5-HT1A agonism/partial agonism, a receptor pairing repeatedly implicated in antidepressant mechanisms (including downstream facilitation of cortical monoamine transmission and network-level effects on affect regulation).[22 23]

To date, the US Food and Drug Administration has approved six SGA for the augmentation treatment of MDD—aripiprazole, quetiapine, brexpiprazole, cariprazine, lumateperone and olanzapine/fluoxetine. In our analyses, aripiprazole, brexpiprazole, cariprazine and quetiapine augmentation were associated with significantly higher odds of response compared with placebo augmentation and confirmed by TSA. However, the efficacy of olanzapine and lumateperone was not supported by conclusive TSA evidence. To date, only one RCT of lumateperone[24] has been published; similarly, the evidence for ziprasidone and perospirone is also limited, and therefore TSA could not provide supportive or conclusive evidence for these agents. In RCTs of olanzapine augmentation, the TRD inclusion criteria were generally more stringent than those used in trials of other SGAs, typically requiring <25%–30%[25 26] improvement during an open-label SSRI/SNRI lead-in phase, compared with the more common threshold of <50% improvement used in many other SGA augmentation RCTs.[27 28] Enrolling a more treatment-resistant TRD population may help explain why olanzapine did not show a significant benefit in the odds of response and consequently failed to meet the required information size and significance boundaries in the TSA. Risperidone represents a second-tier SGA (after aripiprazole, brexpiprazole, quetiapine and cariprazine) in terms of evidence quality; although it remained significant after Shao-Wei adjustment, the AIS was only 37% and TSA boundaries were not crossed, so the signal remains provisional and warrants larger placebo-controlled RCTs.

Strengths and limitations of this study

We synthesised a large RCT evidence base across multiple augmentation strategies and complemented conventional meta-analysis with TSA and multiple small-study/publication-bias

adjustments, providing a more stringent assessment of evidential conclusiveness and robustness than standard pooling alone. Second, we evaluated agent-specific effects within SGAs and quantified differences in evidence certainty (eg, TSA-confirmed vs non-confirmed agents), which improves clinical interpretability beyond class-level conclusions. However, our study has several limitations. First, most lithium and T3 augmentation RCTs were published between the 1980s and the early 2000s, whereas SGA augmentation trials have largely been published from 2005 to the present. Accordingly, comparisons between these two evidence bases should account for era effects, including changes in MDD diagnostic criteria, evolving definitions of TRD and shifts in first-line antidepressant classes used. The placebo-controlled lithium evidence base was derived predominantly from trials using TCAs or MAOIs as background antidepressants, with very limited representation of SSRI-based treatment. Thus, the available evidence pertains mainly to lithium augmentation in older antidepressant contexts and does not adequately address its efficacy in contemporary selective serotonin reuptake inhibitor (SSRI)/serotonin–norepinephrine reuptake inhibitor (SNRI) -based practice. Second, across the included trials, there was also notable variability in augmentation dose or target serum levels and treatment duration. Such differences may have contributed to clinical heterogeneity and may limit the direct comparability of pooled effect estimates across studies and agents. Third, for the purpose of TSA, we did not include head-to-head trials directly comparing lithium with SGAs, and we did not perform a network meta-analysis. In addition, because tolerability and safety outcomes were sparsely and inconsistently reported across trials, particularly for lithium, we did not meta-analyse all-cause discontinuation, discontinuation due to adverse events or adverse events. Therefore, our findings should not be interpreted as establishing a comprehensive clinical preference among augmentation strategies across all patients. To date, the published head-to-head evidence is largely limited to comparisons with quetiapine, which we have discussed above. We did consider whether transitivity concerns could be mitigated by restricting the network to more contemporary trials or by incorporating publication era as a covariate. However, these approaches were ultimately deemed unsuitable because they would have left very sparse lithium and T3 data, further weakened network connectivity and introduced additional instability into indirect comparisons. Fourth, the control design differed across augmentation trials. Most SGA studies compared SGA plus antidepressant with placebo plus antidepressant, whereas the olanzapine studies compared olanzapine plus fluoxetine with fluoxetine alone. Although this comparison is still relevant from an augmentation-efficacy perspective, the lack of a matched placebo condition in the olanzapine trials means that these studies were not fully equivalent to the other placebo-controlled SGA trials and may have differed in placebo-related effects. Finally, placebo-controlled augmentation trials typically exclude patients at high suicide risk. Therefore, the present findings primarily inform antidepressant response in trial-eligible populations and do not address the potential antisuicidal benefits of lithium which may remain clinically relevant in real-world practice.

CONCLUSION AND IMPLICATIONS

SGA augmentation demonstrated a robust response benefit, with TSA providing conclusive support. For lithium, the available information was insufficient for firm inference in TSA and effect estimates were sensitive to bias-adjustment approaches for small-study effects and potential publication bias. These findings primarily address response efficacy and evidence robustness, rather than the full comparative benefit-risk profile of lithium, T3 and SGAs. Clinically, SGAs with TSA-supported efficacy (aripiprazole, quetiapine, brexpiprazole and cariprazine) may be prioritised only in terms of the certainty of response-efficacy evidence. Treatment selection should still incorporate remission, tolerability, adverse-effect burden, monitoring requirements, comorbidities and patient preference.

Author affiliations
[1]Department of Psychiatry, E-Da Dachang Hospital, Kaohsiung, Taiwan
[2]Departmnt of Psychiatry, E-Da Hospital, Kaohsiung, Taiwan
[3]School of Medicine, College of Medicine, I-Shou University, Kaohsiung, Taiwan
[4]Department of Psychiatry, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung, Taiwan
[5]Centre for Chronic Illness and Ageing, University of Greenwich, London, UK
[6]IMPACT (Innovation in Mental and Physical Health and Clinical Treatment) Strategic Research Centre, Deakin University, Melbourne, Victoria, Australia
[7]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[8]Comprehensive Centre for Clinical Neurosciences and Mental Health (C3NMH), Medical University of Vienna, Vienna, Austria
[9]Clinical Division of Social Psychiatry, Department of Psychiatry & Psychotherapy, Medical University of Vienna, Vienna, Austria
[10]Institute of Biomedical Sciences, National Sun Yat-sen University, Kaohsiung, Taiwan
[11]Institute of Precision Medicine, National Sun Yat-sen University, Kaohsiung, Taiwan
[12]Department of Psychology, College of Medical and Health Science, Asia University, Taichung, Taiwan
[13]Prospect Clinic for Otorhinolaryngology & Neurology, Kaohsiung, Taiwan
[14]Department of Neurology, Tri-Service General Hospital, Taipei City, Taiwan
[15]Department of Pharmacy, Linkou Chang Gung Memorial Hospital, Taoyuan, Taiwan
[16]Institute of Epidemiology & Preventive Medicine, National Taiwan University, Taipei City, Taiwan
[17]Department of Dentistry, National Taiwan University Hospital, Taipei City, Taiwan
[18]Department of Psychiatry, Tri-Service General Hospital Beitou Branch, Taipei, Taiwan
[19]Department of Psychiatry, School of Medicine, College of Medicine, National Defense Medical University, Taipei, Taiwan

Contributors T-WH and C-WH contributed equally to this work and are joint first authors. Y-KT and C-SL contributed equally to this work and are joint last/corresponding authors. C-SL and Y-KT conceived and designed the study. T-WH, C-LY, C-WH and P-TT selected the articles, extracted the data and assessed the risk of bias. C-LY did the systemic search. T-WH and C-WH wrote the first draft of the manuscript. Y-HL, TT, AFC, BS, F-CY and Y-KT interpreted the data and contributed to the writing of the final version of the manuscript. C-SL and T-WH have accessed and verified the data. C-SL and Y-KT were responsible for the decision to submit the manuscript. All authors confirmed that they had full access to all the data in the study and accept responsibility to submit for publication. C-SL, as the study guarantor, accepts full responsibility for the finished work and/or the conduct of the study, had access to the data and controlled the decision to publish.

Funding It study was supported by Taiwan National Science and Technology Council (109-2314-B-182A-009-MY2, 111-2314- B-182A-027-, and 112-2314-B-182-070-MY3), the Chang Gung Medical Foundation (CMRPG8P0631, CORPG8P0561, and BMRPJ30), and E-Da Dachang Hospital (EDCHP114006) .

Competing interests No, there are no competing interests.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The data that support the findings of this study are available from the corresponding author (CSL) upon reasonable request.

ORCID iDs
Tien-Wei Hsu https://orcid.org/0000-0003-4136-1251
Chih-Wei Hsu https://orcid.org/0000-0002-8650-4060
Chih-Sung Liang https://orcid.org/0000-0003-1138-5586

REFERENCES

1 Bromet E, Andrade LH, Hwang I, *et al*. Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med* 2011;9:90.
2 Vandeleur CL, Fassassi S, Castelao E, *et al*. Prevalence and correlates of DSM-5 major depressive and related disorders in the community. *Psychiatry Res* 2017;250:50–8.
3 Vos T, Lim SS, Abbafati C, *et al*. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* 2020;396:1204–22.
4 Pérez-Sola V, Roca M, Alonso J, *et al*. Economic impact of treatment-resistant depression: A retrospective observational study. *J Affect Disord* 2021;295:578–86.
5 Rybak YE, Lai KSP, Ramasubbu R, *et al*. Treatment-resistant major depressive disorder: Canadian expert consensus on definition and assessment. *Depress Anxiety* 2021;38:456–67.
6 Heerlein K, De Giorgi S, Degraeve G, *et al*. Real-world evidence from a European cohort study of patients with treatment resistant depression: Healthcare resource utilization. *J Affect Disord* 2022;298:442–50.
7 McIntyre RS, Prieto R, Schepman P, *et al*. Healthcare resource use and cost associated with timing of pharmacological treatment for major depressive disorder in the United States: a real-world study. *Curr Med Res Opin* 2019;35:2169–77.
8 McIntyre RS, Rosenbluth M, Ramasubbu R, *et al*. Managing medical and psychiatric comorbidity in individuals with major depressive disorder and bipolar disorder. *Ann Clin Psychiatry* 2012;24:163–9.
9 Fava M, Targum SD. Augmentation and combination strategies to treat the residual symptoms of major depressive disorder. *Psychiatry (Edgmont)* 2007;4:16–8.
10 Bauer M, Severus E, Möller H-J, *et al*. Pharmacological treatment of unipolar depressive disorders: summary of WFSBP guidelines. *Int J Psychiatry Clin Pract* 2017;21:166–76.
11 Lam RW, Kennedy SH, Adams C, *et al*. Canadian Network for Mood and Anxiety Treatments (CANMAT) 2023 Update on Clinical Guidelines for Management of Major Depressive Disorder in Adults: Reseau canadien pour les traitements de l'humeur et de l'anxiete (CANMAT). *Can J Psychiatry* 2023;69:641–87.
12 Malhi GS, Bell E, Bassett D, *et al*. The 2020 Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. *Aust N Z J Psychiatry* 2021;55:7–117.
13 Vázquez GH, Bahji A, Undurraga J, *et al*. Efficacy and Tolerability of Combination Treatments for Major Depression: Antidepressants plus Second-Generation Antipsychotics vs. Esketamine vs. Lithium. *J Psychopharmacol* 2021;35:890–900.
14 Acero-González A, Guzman Y, Proaños NJ, *et al*. Lithium Augmentation in Treatment-Resistant Depression: A Qualitative Review of the Literature. *Pharmacotherapy* 2025;45:688–701.
15 Cleare AJ, Kerr-Gaffney J, Goldsmith K, *et al*. Clinical and cost-effectiveness of lithium versus quetiapine augmentation for treatment-resistant depression: a pragmatic, open-label, parallel-group, randomised controlled superiority trial in the UK. *Lancet Psychiatry* 2025;12:276–88.
16 Girlanda F, Cipriani A, Agrimi E, *et al*. Effectiveness of lithium in subjects with treatment-resistant depression and suicide risk: results and lessons of an underpowered randomised clinical trial. *BMC Res Notes* 2014;7:731.
17 Katona CL, Abou-Saleh MT, Harrison DA, *et al*. Placebo-controlled trial of lithium augmentation of fluoxetine and lofepramine. *Br J Psychiatry* 1995;166:80–6.
18 Stein G, Bernadt M. Lithium augmentation therapy in tricyclic-resistant depression. A controlled trial using lithium in low and normal doses. *Br J Psychiatry* 1993;162:634–40.
19 Page MJ, McKenzie J, Bossuyt P. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *MetaArXiv* [Preprint].
20 Bauer M, Dell'osso L, Kasper S, *et al*. Extended-release quetiapine fumarate (quetiapine XR) monotherapy and quetiapine XR or lithium as add-on to antidepressants in patients with treatment-resistant major depressive disorder. *J Affect Disord* 2013;151:209–19.
21 McIntyre RS, Alsuwaidan M, Baune BT, *et al*. Treatment-resistant depression: definition, prevalence, detection, management, and investigational interventions. *World Psychiatry* 2023;22:394–412.
22 McEvoy J, Citrome L. Brexpiprazole for the Treatment of Schizophrenia: A Review of this Novel Serotonin-Dopamine Activity Modulator. *Clin Schizophr Relat Psychoses* 2016;9:177–86.
23 Pae CU, Forbes A, Patkar AA. Aripiprazole as adjunctive therapy for patients with major depressive disorder: overview and implications of clinical trial data. *CNS Drugs* 2011;25:109–27.
24 Durgam S, Earley WR, Kozauer SG, *et al*. Adjunctive Lumateperone in Patients With Major Depressive Disorder: Results From a Randomized, Double-Blind, Phase 3 Trial. *Am J Psychiatry* 2025;182:1072–82.
25 Corya SA, Williamson D, Sanger TM, *et al*. A randomized, double-blind comparison of olanzapine/fluoxetine combination, olanzapine, fluoxetine, and venlafaxine in treatment-resistant depression. *Depress Anxiety* 2006;23:364–72.
26 Thase ME, Corya SA, Osuntokun O, *et al*. A randomized, double-blind comparison of olanzapine/fluoxetine combination, olanzapine, and fluoxetine in treatment-resistant major depressive disorder. *J Clin Psychiatry* 2007;68:224–36.
27 Hobart M, Skuban A, Zhang P, *et al*. Efficacy and safety of flexibly dosed brexpiprazole for the adjunctive treatment of major depressive disorder: a randomized, active-referenced, placebo-controlled study. *Curr Med Res Opin* 2018;34:633–42.
28 Sachs GS, Yeung PP, Rekeda L, *et al*. Adjunctive Cariprazine for the Treatment of Patients With Major Depressive Disorder: A Randomized, Double-Blind, Placebo-Controlled Phase 3 Study. *Am J Psychiatry* 2023;180:241–51.