Systematic Review and Meta-Analysis

Medicine

OPEN

Art therapy for core symptoms in schizophrenia
An umbrella review

Zhongyun Zhang, MN[a], Yunchuan Liu, BSN[a], Yao Yang, BSN[b], Weifen Qiu, MN[a,*]

Abstract

Background: Evidence on art therapy for core symptoms of schizophrenia remains uncertain. We synthesized systematic reviews (SRs) and assessed their methodological quality and outcome-level certainty.

Methods: Eleven databases were searched through December 2025 for SRs of art therapy targeting positive, negative, or cognitive symptoms. Effect estimates were restricted to randomized controlled trials (RCTs) or RCT-only subsets. SR methods were appraised with A MeaSurement Tool to Assess systematic Reviews 2, reporting completeness with Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020, and evidence certainty with Grading of Recommendations Assessment, Development and Evaluation. Primary-study risk-of-bias judgments were drawn from source reviews rather than independently reassessed in original trial reports. Primary-study overlap was assessed using the corrected covered area (CCA).

Results: Thirteen SRs covered music, dance, drama, horticultural, visual arts, and multimodal therapies. Eleven were rated low or critically low on A MeaSurement Tool to Assess systematic Reviews 2. Of 28 effect estimates, 3 had moderate, 16 low, and 9 very low Grading of Recommendations Assessment, Development and Evaluation certainty; none had high certainty. Negative-symptom estimates generally favored art therapy, but shared primary studies limited their independence, and selected estimates were sensitive to blinded assessment and publication bias. Positive-symptom findings were inconclusive. Only 1 eligible SR synthesized cognitive outcomes, although additional primary studies reported them. Overlap was slight within the positive- and negative-symptom domains (CCA, 3.43% and 3.65%, respectively); cognitive CCA was inapplicable because only 1 SR contributed.

Conclusion: Art therapy, particularly music therapy, may benefit negative symptoms, but low certainty and limitations in source reviews preclude firm clinical conclusions. Its role as an adjunctive intervention therefore remains tentative pending stronger, more independent evidence. Better-conducted, prospectively registered SRs and adequately powered RCTs are needed.

Abbreviations: AMSTAR-2 = A MeaSurement Tool to Assess systematic Reviews 2, CI = confidence interval, GRADE = Grading of Recommendations Assessment, Development and Evaluation, MD = mean difference, RCTs = randomized controlled trials, SMD = standardized mean difference, SRs = systematic reviews, WCST-Cc = Wisconsin Card Sorting Test–Correctly Completed Categories.

Keywords: art therapy, cognitive symptoms, negative symptoms, positive symptoms, schizophrenia, umbrella review

1. Introduction

Schizophrenia is a severe and chronic mental disorder affecting approximately 1% of the global population, characterized by positive symptoms (hallucinations, delusions), negative symptoms (apathy, avolition, anhedonia, and social withdrawal), and cognitive dysfunction (deficits in attention, memory, and executive function).[1,2] These symptoms severely compromise patients' social and occupational functioning and quality of life, imposing a heavy burden on those affected, their families, and healthcare systems worldwide.[3] Antipsychotic medications remain the cornerstone of treatment for schizophrenia

This study was supported by Kunming University of Science and Technology through the project entitled 'Effects of Trajectory-Based Mandala Drawing on Cognitive Function and Alexithymia in Patients with Schizophrenia' (Project No. KUST-AN2023022Y).

The authors have no conflicts of interest to declare.

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

This manuscript has not been published or submitted elsewhere, in whole or in part.

Generative AI tools were used to assist with language editing, consistency checks, and preparation of visual materials. All outputs incorporated into the manuscript were critically reviewed and verified by the authors, who take full responsibility for the content.

PROSPERO registration number: CRD420251181698.

Supplemental Digital Content is available in the online version of this article (http://dx.doi.org/10.1097/MD.0000000000051551).

[a] Anning First People's Hospital Affiliated to Kunming University of Science and Technology, Anning, Yunnan, China, [b] School of Nursing, Dali University, Dali, Yunnan, China.

** Correspondence: Weifen Qiu, Department of Psychiatry, Anning First People's Hospital Affiliated to Kunming University of Science and Technology, No. 2 Ganghe South Road Jinfang Subdistrict, Anning, Yunnan 650300, P. R. China (e-mail: qiuweifen66@163.com).*

Copyright © 2026 the Author(s). Published by Wolters Kluwer Health, LLC. This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Zhang Z, Liu Y, Yang Y, Qiu W. Art therapy for core symptoms in schizophrenia: An umbrella review. Medicine 2026;105:47(e51551).

Received: 23 June 2026 / Received in final form: 28 October 2026 / Accepted: 29 October 2026

http://dx.doi.org/10.1097/MD.0000000000051551

but demonstrate limited efficacy for negative symptoms and cognitive impairment.[4,5] In addition, adverse effects – including metabolic syndrome, extrapyramidal symptoms, and sedation – frequently compromise adherence.[6–8] These challenges have prompted growing interest in adjunctive non-pharmacological interventions.[9]

Art therapy is defined as the systematic use of creative arts modalities (visual arts, music, dance/movement, drama, and horticultural activities) within a therapeutic framework. Art therapy has emerged as a promising adjunctive approach.[10,11] Art therapy, delivered by trained therapists and grounded in psychological theories, facilitates nonverbal emotional expression, interpersonal interaction, and cognitive engagement.[12,13] These features are particularly relevant for patients with schizophrenia, who often experience profound deficits in verbal communication and social motivation.[14,15]

Numerous randomized controlled trials (RCTs) and systematic reviews (SRs) have investigated the effects of art therapy on schizophrenia symptoms. Several SRs report that music therapy can improve negative symptoms, social functioning, and mental health status.[16,17] A primary study has also examined cognitive functioning following mandala painting therapy.[18]Visual art therapy shows potential benefits for negative symptoms and emotional regulation.[19] Dance/movement therapy can improve psychotic symptoms and social function,[20] while horticultural therapy has been investigated for psychiatric symptoms and cognitive functioning.[21,22] However, the research findings are inconsistent, with some studies reporting significant improvements in positive symptoms,[19,23] whereas others have found no statistically significant effect.[16,24] Thus, review-level evidence regarding cognitive outcomes remains sparse and inconclusive.[25]

The methodological quality of existing SRs varies considerably. Separately, the primary studies synthesized within these reviews are frequently limited by small sample sizes, high or unclear risk of bias, heterogeneous intervention protocols, and insufficient follow-up.[16,25,26] Four gaps remain in the review-level evidence. First, findings across SRs differ in the direction, magnitude, and statistical precision of symptom estimates. These differences require interpretation in relation to the contributing study designs, outcome measures, and assessment time points. Critically, Laws and Conway [24] showed that benefits in terms of negative symptoms did not remain with restriction to studies with blinded assessment, raising questions about whether improvements reflect genuine effects or methodological bias. Second, evidence is fragmented across art therapy modalities, limiting the comparative interpretation of reported effects and their certainty. Reviews of music therapy,[16,23,25] visual art therapy,[19,26] dance therapy,[20,27] drama therapy,[12] and horticultural therapy[22,28] exist in isolation. Clinicians require an integrated summary of the reported effects and their certainty across art therapy modalities. Third, the methodological quality of SRs has not been systematically appraised using validated tools such as Assessing the Methodological Quality of Systematic Reviews 2 (AMSTAR-2).[29] A consistent outcome-level assessment of evidence certainty is needed to support interpretation across reviews and symptom domains. Fourth, cognitive dysfunction – which is recognized as a core symptom domain affecting up to 85% of patients with schizophrenia[30] and a key determinant of functional outcomes – remains critically under-investigated. A preliminary review of existing SRs suggests that cognitive outcomes are systematically under-reported and under-analyzed. The extent and certainty of review-level evidence for cognitive outcomes therefore warrant systematic examination.

An umbrella review – which is a SR of SRs[31] – can address the above challenges by enabling integration, quality appraisal, and evidence grading within a unified framework. This methodology is appropriate when sufficient secondary evidence exists; the objective is to evaluate trustworthiness and consistency of existing syntheses; and substantial heterogeneity or contradictions exist across reviews.[31,32] Therefore, considering the contradictory findings, heterogeneous methodological quality, and lack of cross-modality comparisons in existing SRs – conditions for which the umbrella review methodology is particularly appropriate – we conducted the present study with the aim to systematically integrate and critically appraise the clinical efficacy of different art therapy modalities on positive, negative, and cognitive symptoms of schizophrenia as well as to resolve discrepancies across reviews.

2. Methods

The protocol for this umbrella review was registered with the International Prospective Register of Systematic Reviews (registration number CRD420251181698) and can be accessed at https://www.crd.york.ac.uk/prospero/. Following registration, the review methods were extended in the following respects. Outcome-level Grading of Recommendations Assessment, Development and Evaluation (GRADE) assessment, including consideration of publication bias, was undertaken, although the registration record stated that certainty of evidence and bias due to missing results would not be assessed. Limited trial-level recalculations were undertaken for selected randomized-trial subsets; no meta-analysis of pooled estimates across SRs was performed. Additional post hoc work comprised an exploratory Spearman rank-correlation analysis and a targeted, nonsystematic contextual search for primary studies reporting cognitive outcomes. These additional components were incorporated to strengthen transparency and interpretability. The latter search informed the discussion but did not contribute additional studies to the umbrella-review synthesis or its GRADE assessments.

This umbrella review was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement[33] and conducted following Cochrane methodological guidance.[34] The PRIOR checklist for overviews of reviews was also followed and is provided as Supplementary Material. This umbrella review used data from published SRs and did not involve new participant-level data; therefore, ethics committee approval and individual informed consent were not required.

2.1. Inclusion and exclusion criteria

2.1.1. Inclusion criteria.

(1) Population: Patients diagnosed with schizophrenia using the Diagnostic and Statistical Manual of Mental Disorders, the International Classification of Diseases, or the Chinese Classification and Diagnostic Criteria of Mental Disorders, without restrictions on sex, age, ethnicity, or illness duration.
(2) Intervention: Any form of art therapy (visual arts, music, dance/movement, drama therapy, horticultural therapy, painting therapy, drawing therapy, poetry, creative work, handicrafts, calligraphy, and theatrical art).
(3) Comparator: Any control condition (usual treatment, active interventions, waitlist, or no treatment).
(4) Outcomes: Core symptoms of schizophrenia (positive, negative, and cognitive symptoms).
(5) Study design: We included SRs and meta-analyses that synthesized data from clinical trials. For the purpose of this umbrella review, a SR was defined as a literature review that employed a comprehensive and reproducible search strategy, explicit eligibility criteria, and critical appraisal of the risk of bias of included primary studies. Meta-analyses that pooled data from clinical trials using statistical synthesis were also eligible if they met these methodological criteria.

2.1.2. Exclusion criteria.

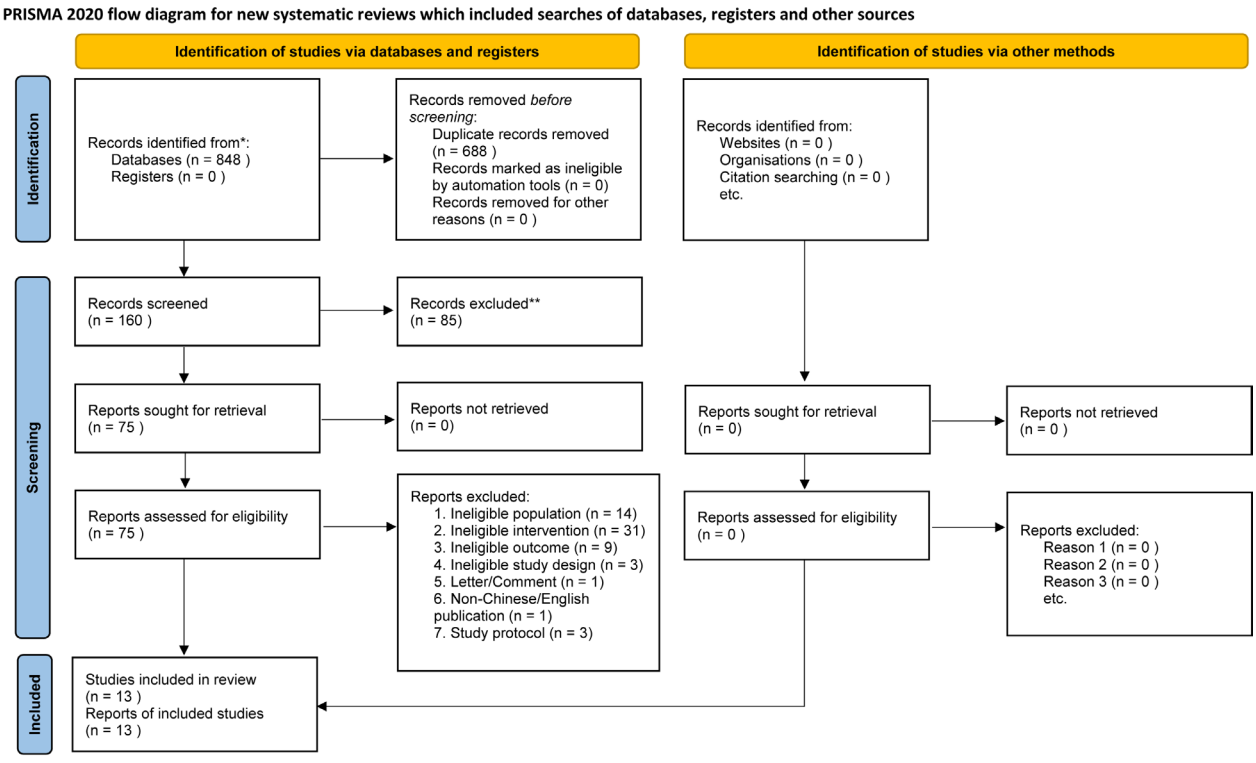
(1) Study protocols, proposals, or conference abstracts.
(2) Studies with full text unavailable.
(3) Duplicate publications; the most comprehensive or earliest version was retained.
(4) SRs including mixed psychiatric populations, even with schizophrenia subgroup data.
(5) SRs that did not include any randomized controlled trials among their primary studies.
(6) Non-English or non-Chinese articles.

2.2. Search strategy

We searched the following 11 databases from inception to December 28, 2025: the Joanna Briggs Institute Evidence Synthesis database, PubMed, Embase, the Cochrane Library, Web of Science, Cumulative Index to Nursing and Allied Health Literature, and PsycINFO; plus the Chinese databases: China National Knowledge Infrastructure, Wanfang Data, Chinese Science and Technology Periodical Database, and Chinese Biomedical Literature Database. We used a combination of MeSH/Emtree terms and free-text words, including schizophrenia, art therapy, music therapy, painting, yoga, dance therapy, poetry, drama therapy, psychodrama, handicrafts, calligraphy, positive symptoms, hallucinations, delusions, negative symptoms, anhedonia, alogia, apathy, cognitive symptoms, cognitive impairment, and cognitive dysfunction. Corresponding Chinese keywords were applied. We manually searched the reference lists of relevant reviews. The PubMed search strategy is provided in Supplementary File 1, Supplemental Digital Content 1. To determine whether the review-level cognitive evidence gap reflected an absence of primary research, we additionally conducted a targeted nonsystematic contextual search (PubMed, relevant journal websites, Chinese-language bibliographic sources, and reference lists of eligible reviews and recent scoping reviews; searched on August 19, 2026). This search was not part of the preregistered umbrella-review selection process and did not alter the eligible SR set, overlap assessment, GRADE ratings, or quantitative synthesis.

2.3. Study selection and data extraction

Records were independently screened by 2 reviewers, with disagreements resolved by a 3rd reviewer. After duplicates were removed, titles and abstracts were screened, followed by a full-text review. We used a standardized form to capture the 1st author, publication year, study design, number of primary studies and sample size, intervention and comparator details, quality assessment tool, outcome instruments, and measures. Data extraction was performed independently and was cross-checked. During study selection, we identified potential primary study overlap by comparing bibliographic details (1st author, publication year, and trial registry numbers, where available) of the primary studies listed in the reference lists of eligible SRs. When multiple reviews included the same primary studies, we retained all relevant reviews to enable cross-review concordance assessment, and we quantified the degree of overlap using the corrected covered area during data synthesis. At the outcome level, we identified primary study overlap by extracting citation details (1st author, publication year, and sample size) of all primary studies included in each review. For each outcome domain, K denoted the number of contributing SRs, U the number of unique primary studies, and N the total number of primary-study occurrences in the citation matrix. CCA was calculated as $(N - U)/[U \times (K - 1)] \times 100\%$. Citation-matrix



Source: Page MJ, et al. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.
This work is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

Figure 1. PRISMA 2020 flow diagram of systematic review identification, screening, and inclusion. PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

coverage, calculated as $N/(K \times U) \times 100\%$, was reported separately. CCA values were interpreted as follows: <5% slight, 5%–9.9% moderate, 10%–14.9% high, and ≥15% very high overlap.[32] CCA was not applicable when only 1 SR contributed to an outcome domain.

When discrepant data were identified across SRs (e.g., divergent effect estimates or conflicting directions of effect for the same outcome), 2 reviewers independently cross-checked the original review reports to verify the extracted values, outcome definitions, and statistical models. Persistent discrepancies were resolved through discussion with a 3rd senior reviewer (WQ) to reach consensus.

For consistent presentation, effect estimates were aligned to a common sign convention in which positive values favor art therapy and negative values favor the comparator. The sign assigned to each estimate was determined from the source-reported treatment contrast, the direction of the outcome scale, and, where applicable, the definition of change from baseline. Estimates for which negative values favored art therapy were multiplied by −1, whereas estimates for which positive values favored art therapy were retained. For an original estimate θ with a 95% confidence interval (CI) [*L*, *U*], sign reversal yielded −θ with a CI [−*U*, −*L*]. This transformation preserved the absolute effect magnitude, confidence-interval width, and statistical significance. Direction-coding decisions were recorded for each estimate in the supplementary source-data table.

Effect estimates were extracted from the source reviews and their published corrections. For Tseng et al, the RCT-only estimates were recalculated from the corrected trial-level Hedges' *g* values and CIs reported in the erratum.[23] For You and Wang, the randomized-trial subset was recalculated from the posttreatment means, standard deviations, and sample sizes reported in the source review. Directly extracted and recalculated estimates were distinguished in the outcome-level evidence summary. Mean differences (MDs) remained in the units of the corresponding outcome scales and were not directly compared numerically with standardized effects. Cross-review comparisons were descriptive and considered differences in interventions, outcome measures, and assessment time points.

2.4. Quality assessment

Two reviewers independently assessed each SR, with discrepancies resolved through discussion.

2.4.1. Methodological quality. We used AMSTAR-2[29] to assess methodological quality, which comprises 16 items (7 critical: items 2, 4, 7, 9, 11, 13, and 15). Items were rated as "Yes," "Partial Yes," or "No" according to the tool's rating guidelines; "Partial Yes" was applied only to items for which the AMSTAR-2 tool defines this response option. The overall SR quality was classified as high, moderate, low, or critically low. For critical items, a "No" was counted as a critical flaw, whereas "Partial Yes" was treated as a partial deficiency without elevating the rating to low or critically low. Overall ratings were assigned as: high (no or 1 noncritical weakness), moderate (>1 noncritical weakness but no critical flaws), low (1 critical flaw), and critically low (>1 critical flaw).

2.4.2. Reporting completeness. Reporting completeness was assessed using the PRISMA 2020 checklist[33] with its 27 numbered items assessed as 42 separate items and subitems. Following the 3-category approach described by Xue et al,[35] reporting was classified as fully reported (Y), partially reported (PY), or not reported (N). These categories were assigned values of 1, 0.5, and 0, respectively, for descriptive summarization. The intermediate value distinguished partial reporting from complete or absent reporting, and each assessed item or subitem received equal weight. For each SR, the overall reporting completeness score was calculated as the sum of its scores divided by 42 and multiplied by 100. For each item or subitem, weighted reporting completeness was calculated as the sum of scores across the 13 SRs divided by 13 and multiplied by 100. Scores were summarized descriptively without categorical cutoffs. Reporting completeness was assessed separately from review-level methodological quality, primary-study risk of bias, and outcome-level certainty of evidence.

2.4.3. Certainty of evidence. We independently reassessed the certainty of every outcome-level effect estimate using a uniform outcome-level framework.[36] Ratings reported by the source SRs were used only as contextual information and were not adopted automatically. Two reviewers independently evaluated risk of bias, inconsistency, indirectness, imprecision, and publication bias. Certainty of the evidence was classified as high, moderate, low, or very low.

2.4.4. Risk of bias of primary studies included in the SRs. We did not independently reassess the risk of bias of each primary trial from its original report. For every outcome-level effect estimate, 2 reviewers extracted the most directly applicable primary-study risk-of-bias information reported in the source SR and determined whether that information applied to the trials contributing to the specific outcome and time point. When an SR reported trial-level or domain-level judgments, these judgments were mapped to the identifiable contributing trials. When only a review-wide primary-study quality profile was available, we used the information most applicable to the relevant trial set and documented the resulting uncertainty. A conservative risk-of-bias judgment was applied when the available primary-study information indicated serious or very serious methodological limitations.

2.5. Rationale for the umbrella review methodology

Umbrella review methodology was used to integrate and critically appraise existing SRs across art therapy modalities and core symptom domains.[31] Differences in interventions, outcome measures, and assessment time points were considered in the narrative synthesis. No meta-analysis of pooled estimates across SRs was performed. Limited trial-subset recalculations were undertaken to restrict selected estimates to randomized trials. For Tseng et al,[23] corrected trial-level Hedges' *g* values and 95% CIs were extracted from the published erratum. Trials classified as nonrandomized in the source review were excluded. Standard errors were reconstructed from the reported CIs, and the retained effects were combined using an inverse-variance random-effects model with DerSimonian–Laird estimation of between-study variance and normal-approximation CIs. For You and Wang,[37] posttreatment SANS data from 5 trials reported as randomized were pooled using an inverse-variance fixed-effect model with Hedges' *g*; 1 study classified as semi-randomized was excluded. Study selection, inputs, formulas, software, and calculation results are documented in Supplementary Materials. Cross-review concordance was summarized descriptively, taking account of primary-study overlap and outcome-level certainty of evidence. Because some primary studies appeared in >1 review, the number of reviews reporting a benefit was treated as a descriptive count rather than as independent corroboration. Participant numbers were presented for individual effect estimates and were not summed across reviews.

2.6. Certainty of evidence assessment

The certainty of evidence for each effect estimate was independently appraised by 2 reviewers using the GRADE approach. Because the outcome-level estimates were derived from RCTs or RCT-only subgroups, each body of evidence initially started at

high certainty. Risk-of-bias judgments were based on the source SRs' assessments of the primary trials contributing to the specific outcome and time point. When outcome-specific risk-of-bias summaries were unavailable, trial-level or domain-level information reported by the source SR was mapped to the contributing studies; when such mapping was not possible, the most applicable primary-study quality information was used conservatively and the information gap was explicitly documented.

Inconsistency was assessed by considering the direction and magnitude of effect estimates across the primary studies contributing to each outcome estimate, as reported in the source SR, together with the overlap of CIs, I^2, and relevant clinical and methodological heterogeneity. I^2 values of approximately 50% or greater prompted closer examination but did not automatically result in downgrading. Evidence was downgraded by 1 level for serious inconsistency when substantial unexplained heterogeneity was accompanied by clinically meaningful variation in effect magnitude or limited overlap of CIs. A 2-level downgrade was reserved for extreme unexplained heterogeneity accompanied by major differences in effect magnitude or direction. For estimates derived from a single primary study, no downgrade for inconsistency was applied because between-study heterogeneity could not be assessed. Indirectness was assessed against the prespecified population, intervention, comparator, and outcome. Imprecision was assessed using a minimally contextualized approach, with the 95% CI as the primary consideration.[38] We examined whether the CI crossed the null effect or encompassed effect magnitudes that would support materially different interpretations, together with the amount of information available and the optimal information size where it could be estimated. Outcome-specific minimally important differences were used when available. When no validated minimally important difference was available, we did not infer adequate precision solely because the CI excluded the null; instead, we considered the range of plausible effect magnitudes and the degree of information-size shortfall. Evidence was downgraded by 1 level for serious imprecision and by 2 levels only when the available information was extremely sparse and the CI remained compatible with markedly different effect magnitudes, or when the available information was far below the relevant optimal information size. Publication bias was downgraded only when supported by outcome-specific statistical evidence or a strong, documented suspicion of missing evidence; inability to assess publication bias because of few studies did not itself result in downgrading. No upgrading factors were applied. Disagreements were resolved through discussion with a 3rd senior reviewer (WQ). The complete outcome-level decision record, including the interpretation of CIs and information-size considerations, is provided in Supplementary Materials.

2.7. Exploratory quality–effect size analysis

We conducted a post hoc exploratory Spearman rank-correlation analysis to examine the association between review-level methodological quality and reported effect magnitude. Negative symptoms were the primary domain, and positive symptoms were examined in a secondary analysis. AMSTAR-2 categories were coded as critically low = 1, low = 2, moderate = 3, and high = 4, with average ranks assigned to ties.

The analysis used the retained outcome-level estimates, including 2 RCT-only estimates recalculated from the corrected trial-level effects reported for Tseng et al[23] and 1 randomized-trial estimate recalculated from the summary outcome data reported by You and Wang.[37] All other estimates were taken directly from the source reviews. Each SR contributed 1 estimate per domain. The longest time category available for each outcome was selected, with Lee et al's short-term estimates substituted in sensitivity analyses. Analyses were restricted to standardized mean differences (SMDs) or Hedges' g; scale-dependent MDs were excluded. Absolute effect magnitudes were ranked, so the analysis did not distinguish beneficial from adverse directions. Spearman's rho was calculated as the Pearson correlation of the 2 rank variables. Conventional 2-sided P values used the t approximation. Exact reference-model P values were obtained by exhaustive permutation, counting absolute correlations at least as large as observed. Leave-one-out analyses assessed sensitivity to individual SRs. These analyses did not account for covariance from overlapping primary studies; coefficients and P values were interpreted descriptively. Full inputs and calculations are provided in Supplementary Materials.

Given the small number of included SRs, the presence of tied AMSTAR-2 categories, and the lack of an identifiable covariance structure among overlapping primary studies, formal adjustment for cross-review dependence was not feasible. Accordingly, Spearman's rho and associated P values should be regarded as descriptive and hypothesis-generating rather than confirmatory estimates.

3. Results

3.1. Results of the literature search

A total of 848 potentially relevant records were identified. After 688 duplicate records were removed, 160 records remained for title and abstract screening. Following a stepwise screening process, 13 SRs met the inclusion criteria,[12,16,17,19,20,22–28,37] of which 3 were in Chinese[20,26,37] and 10 in English.[12,16,17,19,22–25,27,28] Figure 1 presents the literature screening process and results. A list of records that appeared to meet the inclusion criteria after title and abstract screening but were subsequently excluded after full-text review, together with the primary reason for exclusion, is provided in Supplementary File 2, Supplemental Digital Content 2.

3.2. Characteristics of included SRs)

The 13 included SRs were published between 2002 and 2025. The number of primary studies included in each review ranged from 1 to 35, and the total sample sizes ranged from 43 to 3037 participants. Nine SRs included only RCTs,[12,17,19,20,24–28] whereas 4 also included nonrandomized studies.[16,22,23,37] The outcome-level estimates were restricted to RCTs or RCT-only subgroups. The characteristics of the included SRs are summarized in Table 1.

3.3. Methodological quality assessment

We assessed the methodological quality of the 13 included SRs using the AMSTAR-2 tool. Overall, the quality was found to be suboptimal. Specifically, only 1 SR (7.69%)[25] was rated high quality; 1 SR (7.69%)[28] was rated moderate; 4 SRs (30.77%)[19,20,24,27] were rated low quality, and the remaining 7 SRs (53.85%)[12,16,17,22,23,26,37] were rated critically low quality. Among critical domains, compliance was notably poor for item 2 (protocol registration), item 4 (comprehensiveness of literature search), item 7 (provision of a list of excluded studies with justifications), item 13 (consideration of risk of bias in the interpretation of results), and item 15 (investigation of publication bias). Table 2 presents the detailed AMSTAR-2 ratings for each included SR. Primary-study risk-of-bias assessments reported by the source SRs are summarized in Supplementary File 3, Supplemental Digital Content 3. Commonly reported limitations included inadequate or unclear allocation concealment and insufficient blinding; the applicability of these assessments to individual outcome estimates was considered separately in the GRADE assessment.

Table 1

Characteristics of the included systematic reviews.

Author (year)	Country	Trial (sample size)	Diagnostic criteria for schizophrenia	Intervention		Quality assessment tool	Outcomes (measurement tools)	Main conclusions
				Experiment	Control			
He XC, 2025	China	19 (1134)	ICD-10, DSM-5, CCMD-3	Dance therapy/dance training	TAU, routine care, routine rehabilitation activities	Cochrane RoB Tool 5.1.0	Positive symptoms (PANSS-PSS) Negative symptoms (PANSS-NSS/BNSS/SANS) Quality of life (Q-LES-Q/SQLS) General psychotic symptoms (PANSS-GPS, BPRS, BDI) Social functioning (GAF/SDSS/SSPI) Self-efficacy (GSES)	Dance therapy can alleviate psychotic symptoms, improve social function and self-efficacy but has no significant effect on quality of life. Effects on cognition, motor function, and self-esteem remain unclear.
Liu DJ, (2017)	China	9 (784)	CCMD-2/3, ICD-10, DSM-5	Medication + TAU + art therapy (self-expression, music, dance, painting)	Medication + TAU + basic rehabilitation	Cochrane RoB Tool 5.1.0	Positive symptoms (PANSS) Negative symptoms (PANSS) Functional rehabilitation (GAF) Self-esteem (SES)	Adjunctive art therapy improves negative symptoms and self-esteem but has no significant effect on positive symptoms or functional rehabilitation.
You ZY, (2002)	China	11 (603)	CCMD-II-R, DSM-III-R	Music therapy + antipsychotics	Antipsychotics alone	Jadad scores	Negative symptoms (SANS) General psychotic symptoms (BPRS)	The short-term effect of adjunctive music therapy for chronic schizophrenia is favorable, but its long-term efficacy requires further investigation.
Lam L, (2023)	China	13 (1114)	DSM-III, DSM-IV, DSM-V, ICD-10 CCMD-2;CCMD-3	Adjunct multiple-guided group music therapy	TAU	Cochrane RoB tool	positive and negative symptoms (PANSS/SAPS/SANS/BPRS) Brain activity (EEG/fMRI)	This review demonstrated that both objective and subjective assessments captured reductions in symptom severity and improvements in psychosocial functioning. Group music therapy, irrespective of delivery format, yielded encouraging adjunctive benefits relative to treatment as usual in individuals with chronic schizophrenia. Given that music therapy is low-cost, noninvasive, and associated with minimal adverse effects, its wider implementation in schizophrenia care is warranted. Future studies should adopt rigorous longitudinal designs with larger sample sizes to determine whether regular, long-term passive music listening, active singing and music making, and group-based music therapy confer comparable therapeutic benefits after hospital discharge.
Laws KR, (2019)	UK	9(948)	DSM-V	Art therapies (dance/movement, music, art) as an adjunct to pharmacotherapy	TAU (no additional art therapy)	Cochrane RoB Tool 2.0	Total symptoms (PANSS, BPRS) Positive symptoms (PANSS, SAPS, BPRS) Negative symptoms (PANSS, SANS, BPRS)	Meta-regression revealed larger negative symptom reductions in trials with a higher proportion of women and younger patients. Critically, however, this effect was absent in trials employing blinded outcome assessment.
Jia RY, (2020)	China	18 (1212)	DSM-III-R, DSM-IV, ICD-10, CCMD-II, CCMD-III	Music therapy + antipsychotics/psychiatric routine treatment	Antipsychotics alone/TAU/waitlist/standard care	Jadad scores	Total symptoms (PANSS/BPRS) Positive symptoms (PANSS/SAPS/BPRS) Negative symptoms (PANSS/SANS/BPRS)	Adjunct music therapy improved total, negative, and depressive symptoms and quality of life in patients with schizophrenia relative to controls; however, positive symptoms were unaffected.
Tseng PT, (2016)	China	12 (804)	SADS, DSM-III-R, ICD-10, DSM-IV-TR, DSM-IV, CCMD-2, CCMD-3,	Standard treatment + adjunctive music therapy	Standard treatment alone	Jadad score	Positive symptoms (PANSS/BPRS) Negative symptoms (PANSS/SANS) General psychopathology (PANSS/BPRS) Mood symptoms (SDS)	Adjunctive music therapy significantly ameliorated positive, negative, and mood symptoms, with particularly pronounced effects observed in patients with a chronic disease course.

(*Continued*)

Table 1

(*Continued*)

Author (year)	Country	Trial (sample size)	Diagnostic criteria for schizophrenia	Intervention		Quality assessment tool	Outcomes (measurement tools)	Main conclusions
				Experiment	Control			
Ren J, (2013)	UK	1 (43)	DSM-IV	Dance therapy + routine care	Routine care + supportive counseling	Cochrane RoB Tool (Higgins 2011)	Mental state Total scores (PANSS) Positive symptoms (PANSS) Negative symptoms (PANSS) Satisfaction with care Quality of life (MANSA)	Dance therapy combined with routine care can significantly improve negative symptoms but had no significant effects on total/positive symptoms, satisfaction, or quality of life. Evidence is insufficient owing to small sample size and high dropout.
Geretsegger M, (2017)	Norway, China,	18 (1215)	DSM-III-R, DSM-IV, ICD-10, CCMD-2, CCMD-3	Music therapy + standard care	Standard care alone	Cochrane RoB Tool (Higgins 2011)	General mental state (PANSS/BPRS) Negative symptoms (SANS) positive Symptoms (SAPS) Depression and anxiety (SDS/HAMD/CDSS/SAS) General functioning (GAF), functioning (IADL), social functioning (SDSS), cognitive functioning (PASAT/CCPT/WMS/CMT/BCST/WCST-Cc) Behavior (NOSIE) Recipient of care satisfaction (CSQ) Quality of life (GWB/SPG/SSQ)	Adjunctive music therapy improves global state, total/negative symptoms, social functioning, depression, anxiety, and some cognitive functions. No significant effect on positive symptoms. Dosage (≥20 sessions) and quality of therapy are key factors.
Du SC, (2024)	China	31 (3037)	DSM-III-R, ICD-10, DSM-IV, CCMD-2, CCMD-3, DSM-IV-TR	Visual art therapy + standard treatment	Standard treatment only (mainly antipsychotics)	Cochrane RoB tool	Positive symptoms (PANSS/BPRS/SAPS) Negative symptoms (PANSS/SANS/BPRS) Anxiety (SAS/HAMA/SCL-90) Depression (SDS/HAMD/CDSS)	Visual art therapy significantly ameliorated positive and negative symptoms, depression, and anxiety.
Ruddy R, (2007)	UK	5 (210)	Any criteria for schizophrenia or schizophrenia-like illnesses	Psychodrama + medication + inpatient stay	Medication + inpatient stay	Cochrane Handbook (Higgins 2005)	Negative symptoms (SANS) Positive symptoms (SPNS) Self-Esteem (SES) Feeling Inferior Scale (FIS)	Randomized studies are possible. The benefits or harms of drama therapy remain unclear owing to poor reporting and limited data.
Lee YW, (2024)	China	35 (2899)	Criteria by Leucht et al. (PANSS/BPRS cutoffs)	Horticultural therapy + standard treatment	Standard treatment only (mainly antipsychotics)	Cochrane Risk of Bias tool (Higgins 2011)	Positive and Negative Symptoms Scale (PANSS positive subscale) Negative symptoms (PANSS negative subscale/SANS) Depression (SDS/HAMD/CDSS) Anxiety (SAS/HAMA) Total symptoms (PANSS total/BPRS)	Horticultural therapy is efficacious in improving symptoms and emotional well-being in patients with schizophrenia.
Lu S, Zhao YJ, (2021)	China	23 (2024)	CCMD-2-R, ICD, DSM	Horticultural therapy (planting, making bonsai, etc)	Usual care (medication + conventional activities)	Cochrane RoB Tool (RCTs)/ JBI Tool (quasi-experimental)	Symptoms (PANSS/BPRS/SANS) Rehabilitation Outcomes (IPROS) Quality of Life (SQLS/GQLI-74) Social Functioning (SSPI/PSP/SDSS)	Horticultural therapy has positive effects on symptoms, rehabilitation outcomes, quality of life, and social functioning. Nonhospital environments show better therapeutic effects.

Main conclusions summarize the statements of the source reviews and do not represent recommendations of the present umbrella review.

BPRS = Brief Psychiatric Rating Scale, CCMD = Chinese Classification of Mental Disorders, CCPT = Conners' Continuous Performance Test, CDSS = Calgary Depression Scale for Schizophrenia, CMT = Clinical Memory Test, DSM = Diagnostic and Statistical Manual of Mental Disorders, GAF = Global Assessment of Functioning, GQLI-74 = Generic Quality of Life Inventory-74, HAMA = Hamilton Anxiety Rating Scale, HAMD = Hamilton Depression Scale, ICD = International Classification of Diseases, IPROS = Inpatient Psychiatric Rehabilitation Outcome Scale, JBI = Joanna Briggs Institute, PANSS = Positive and Negative Syndrome Scale, PSP = Personal and Social Performance Scale, SANS = Scale for Assessment of Negative Symptoms, SAPS = Scale for the Assessment of Positive Symptoms, SAS = Self-rating Anxiety Scale, SCL-90 = Symptom Checklist-90, SDS = Self-Rating Depression Scale, SDSS = Social Disability Screening Schedule, SES = Self-Esteem Scale, SQLS = Schizophrenia Quality of Life Scale, SSPI = Scale of Social Function in Psychosis Inpatients, TAU = treatment as usual, WCST-Cc = Wisconsin Card Sorting Test–Categories completed.

Table 2

AMSTAR-2 assessment results.

	AMSTAR-2																
Study	***Q*1**	***Q*2**	***Q*3**	***Q*4**	***Q*5**	***Q*6**	***Q*7**	***Q*8**	***Q*9**	***Q*10**	***Q*11**	***Q*12**	***Q*13**	***Q*14**	***Q*15**	***Q*16**	**Quality**
Lu S, Zhao YJ, (2021)	Y	N	Y	Y	Y	Y	PY	Y	Y	N	Y	N	N	Y	N	Y	CL
Lee YW (2024)	Y	Y	Y	Y	Y	Y	PY	Y	Y	N	Y	N	N	Y	Y	Y	L
Lam L (2023)	Y	N	N	PY	N	N	PY	Y	Y	N	NMA	NMA	N	N	NMA	Y	CL
Keith R Laws (2019)	Y	N	Y	PY	N	N	PY	Y	Y	N	Y	Y	Y	Y	Y	N	L
Jia RY (2020)	Y	N	Y	PY	Y	Y	PY	Y	Y	N	Y	N	N	Y	Y	Y	CL
Tseng PT (2016)	Y	N	Y	PY	Y	N	PY	Y	Y	N	Y	N	N	Y	Y	N	CL
Ren J, 2013	Y	PY	Y	Y	Y	Y	Y	Y	Y	Y	NMA	NMA	Y	N	NMA	Y	M
Geretsegger M (2017)	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	H
Du SC (2024)	Y	N	Y	PY	Y	Y	PY	Y	Y	N	Y	N	Y	Y	Y	Y	L
Ruddy R (2007)	Y	N	Y	PY	Y	Y	PY	Y	Y	N	NMA	NMA	N	Y	NMA	Y	CL
He XC (2025)	Y	N	Y	PY	Y	Y	PY	Y	Y	N	Y	N	Y	Y	Y	N	L
Liu DJ (2017)	Y	N	Y	PY	Y	Y	PY	Y	Y	N	Y	N	N	Y	N	N	CL
You ZY (2002)	Y	N	Y	PY	N	N	PY	Y	Y	N	Y	N	N	Y	N	N	CL

*Q*1: Did the research questions and inclusion criteria incorporate all PICO components?
*Q*2: Did the report explicitly state that the review methods were established a priori, and were any substantial deviations from the protocol justified?
*Q*3: Did the review authors provide a rationale for their selection of study designs?
*Q*4: Did the review authors employ a comprehensive literature search strategy?
*Q*5: Was study selection performed in duplicate?
*Q*6: Was data extraction performed in duplicate?
*Q*7: Did the review authors provide a list of excluded studies alongside justifications for each exclusion?
*Q*8: Did the review authors describe the included studies in sufficient detail?
*Q*9: Did the review authors use an appropriate method to assess the risk of bias (RoB) in individual included studies?
*Q*10: Did the review authors report the funding sources of the included studies?
*Q*11: If a meta-analysis was conducted, did the review authors use appropriate statistical methods to combine the results?
*Q*12: If a meta-analysis was conducted, did the review authors evaluate the potential impact of risk of bias in individual studies on the meta-analytic results or other evidence synthesis?
*Q*13: Did the review authors consider risk of bias in individual studies when interpreting the review findings?
*Q*14: Did the review authors adequately explain and discuss any heterogeneity observed in the results?
*Q*15: If quantitative synthesis was conducted, did the review authors sufficiently examine publication bias (small-study bias) and discuss its likely impact on the results?
*Q*16: Did the review authors disclose all potential conflicts of interest, including any funding received for the review?
Critical domains: *Q*2, *Q*4, *Q*7, *Q*9, *Q*11, *Q*13, and *Q*15.
High: No or 1 noncritical weakness.
Moderate: >1 noncritical weakness.
Low: 1 critical flaw, with or without noncritical weaknesses.
Critically low: >1 critical flaw, with or without noncritical weaknesses.
AMSTAR-2 = A MeaSurement Tool to Assess systematic Reviews 2, CL = critically low, H = high, L = low, M = moderate, N = no, NMA = no meta-analysis conducted, PY = partial yes, Y = yes.

3.4. Reporting completeness

Across the 13 included SRs, the mean reporting completeness score was 69.3% (range, 38.1%–88.1%; median, 67.9%). No SR fully reported all 42 assessed items and subitems. Eight items or subitems received full scores across all reviews: the abstract (item 2), rationale (item 3), objectives (item 4), information sources (item 6), effect measures (item 12), results of individual studies (item 19), interpretation of findings (item 23a), and implications for practice, policy, and future research (item 23d). The lowest weighted reporting completeness was observed for the processes used to determine eligibility for each synthesis (item 13a, 11.5%), methods for tabulating or visually displaying results (item 13c, 15.4%), registration information (item 24a, 15.4%), and amendments to registration or protocol information (item 24c, 15.4%). Certainty assessment methods (item 15), certainty of evidence results (item 22), and availability of data, code, and other materials (item 27) each had weighted reporting completeness of 30.8%. Detailed ratings are presented in Table 3 and visualized in Figure 2.

3.5. Certainty of evidence assessment

Evidence certainty was assessed for 28 effect estimates from 10 SRs, covering positive symptoms (n = 10), negative symptoms (n = 11), and cognitive functioning (n = 7). Overall, 3 estimates (10.7%) were rated moderate certainty, 16 (57.1%) were rated low certainty, and 9 (32.1%) were rated very low certainty; no estimate was rated high certainty.

For positive symptoms, 1 of the 10 estimates was rated moderate certainty, 3 were rated low certainty, and 6 were rated very low certainty. For negative symptoms, 2 of the 11 estimates were rated moderate certainty, 7 were rated low certainty, and 2 were rated very low certainty. Cognitive functioning was represented by 7 estimates from 1 SR: 6 were rated low certainty and 1 was rated very low certainty. Risk-of-bias downgrading was applied to 26 of the 28 estimates on the basis of primary-study assessments reported in the source SRs; AMSTAR-2 ratings were not used to determine these outcome-level judgments. Fourteen of the 28 estimates were downgraded by 1 level for serious inconsistency. These judgments considered the direction and magnitude of effects, overlap of CIs, and clinical or methodological heterogeneity alongside the I^2 statistic. Imprecision was downgraded for 16 estimates: 15 by 1 level for serious concerns and 1 by 2 levels for very serious imprecision, the latter reflecting a single-study cognitive estimate with only 30 participants, a wide CI, and an information size far below illustrative optimal information size estimates. These judgments were based on the width and clinical interpretability of the 95% CI together with the available information size. Publication bias was downgraded by 1 level for 3 estimates supported by outcome-specific statistical evidence or a documented suspicion of missing studies. Three of the 13 included SRs did not contribute outcome-level estimates to Table 4. Ruddy et al[12] reported no pooled effect estimate; Lam et al[17] was a qualitative SR without meta-analytic pooling; and Lu et al[22] pooled overall symptom severity, rehabilitation outcomes, quality of life, and social functioning but

Table 3

Reporting completeness of the included systematic reviews assessed using PRISMA 2020.

Section/ topic	Items	He XC, 2025	Liu DJ, 2017	You ZY,2002	Lam L,2023	Laws KR, 2019	Jia RY,2020	Tseng PT, 2016	Ren J , 2013	Geretsegger M, 2017	Du SC, 2024	Ruddy R, 2007	Lee YW,2024	Lu S, Zhao YJ,2021	Weighted item reporting completeness (%)
Title	*Q*1. Title	Y	Y	Y	Y	Y	Y	Y	N	N	Y	PY	Y	Y	80.8
Abstract	*Q*2. Abstract	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100
Introduc- tion	*Q*3. Rationale	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100
	*Q*4. Objectives	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100
Methods	*Q*5. Eligibility criteria	Y	Y	PY	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	96.2
	*Q*6. Information sources	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100
	*Q*7. Search strategy	Y	N	N	N	PY	Y	N	Y	Y	Y	Y	Y	Y	65.4
	*Q*8. Selection process	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	84.6
	*Q*9. Data collection process	Y	Y	N	Y	N	Y	N	Y	Y	Y	Y	Y	PY	73.1
	*Q*10a. Data items	PY	PY	PY	PY	PY	PY	Y	Y	Y	PY	Y	PY	PY	65.4
	*Q*10b. Data items	PY	PY	N	PY	Y	Y	PY	Y	Y	PY	Y	PY	PY	65.4
	*Q*11. Study risk of bias assessment	Y	Y	PY	Y	PY	PY	PY	Y	Y	Y	Y	Y	Y	84.6
	*Q*12. Effect measures	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100.0
	*Q*13a. Synthesis methods	PY	PY	N	N	N	N	N	N	N	N	PY	N	N	11.5
	*Q*13b. Synthesis methods	Y	PY	N	N	Y	Y	Y	Y	Y	Y	Y	Y	PY	76.9
	*Q*13c. Synthesis methods	N	N	N	N	N	N	N	N	Y	N	Y	N	N	15.4
	*Q*13d. Synthesis methods	Y	Y	PY	PY	Y	Y	Y	Y	Y	Y	PY	Y	Y	88.5
	*Q*13e. Synthesis methods	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	69.2
	*Q*13f. Synthesis methods	PY	N	N	N	N	Y	N	Y	Y	Y	Y	Y	PY	53.8
	*Q*14. Reporting bias assessment	PY	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	N	65.4
	*Q*15. Certainty assessment	N	N	N	N	N	N	N	Y	Y	Y	N	Y	N	30.8
Results	*Q*16a. Study selection	Y	PY	PY	Y	PY	Y	PY	Y	Y	Y	PY	Y	Y	80.8
	*Q*16b. Study selection	PY	N	PY	PY	N	N	N	Y	Y	PY	Y	PY	PY	46.2
	*Q*17. Study characteristics	PY	PY	PY	PY	Y	Y	Y	Y	Y	Y	Y	Y	Y	84.6
	*Q*18. Risk of bias in studies	PY	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	88.5
	*Q*19. Results of individual studies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100.0
	*Q*20a. Results of syntheses	Y	Y	PY	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	96.2
	*Q*20b. Results of syntheses	Y	Y	Y	PY	Y	Y	Y	PY	Y	Y	Y	Y	Y	92.3
	*Q*20c. Results of syntheses	N	N	N	N	Y	Y	Y	N	Y	Y	N	Y	Y	53.8
	*Q*20d. Results of syntheses	N	N	Y	N	N	Y	N	N	Y	Y	N	Y	Y	46.2
	*Q*21. Reporting biases	Y	N	N	N	Y	Y	Y	N	N	Y	N	Y	N	46.2
	*Q*22. Certainty of evidence	N	N	N	N	N	N	N	Y	Y	Y	N	Y	N	30.8
Discus- sion	*Q*23a. Discussion	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100.0
	*Q*23b.Discussion	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	PY	96.2
	*Q*23c.Discussion	Y	PY	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	88.5
	*Q*23d.Discussion	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100.0
Other infor- mation	*Q*24a.Registration and protocol	N	N	N	N	N	N	N	N	N	Y	N	Y	N	15.4
	*Q*24b.Registration and protocol	N	N	N	N	N	N	N	Y	Y	PY	Y	PY	N	30.8
	*Q*24c.Registration and protocol	N	N	N	N	N	N	N	Y	Y	N	N	N	N	15.4
	*Q*25. Support	N	PY	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	65.4
	*Q*26.Competing interests	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	76.9
	*Q*27. Availability of data, code and other materials	N	N	N	Y	N	PY	PY	N	N	Y	N	Y	N	30.8
Overall reporting completeness score (%)		61.9%	52.4%	38.1%	54.8%	61.9%	77.4%	66.7%	79.8%	88.1%	88.1%	76.2%	88.1%	67.9%	

The 3 reporting categories were adapted from Xue et al.[35] and coded as 1, 0.5, and 0, respectively, for descriptive summarization. The 27 numbered PRISMA 2020 items were assessed as 42 separate items and subitems, each receiving equal weight. For each SR, the overall reporting completeness score was calculated as (nY + 0.5 × nPY)/42 × 100%. For each item or subitem, weighted reporting completeness was calculated as (nY + 0.5 × nPY)/13 × 100%. Here, nY and nPY denote the numbers of fully and partially reported ratings within the relevant column or row, respectively. Percentages were rounded to 1 decimal place. These scores summarize reporting completeness. No categorical cutoffs were applied.

N = not reported, PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses, PY = partially reported, SR = systematic review, Y = fully reported.

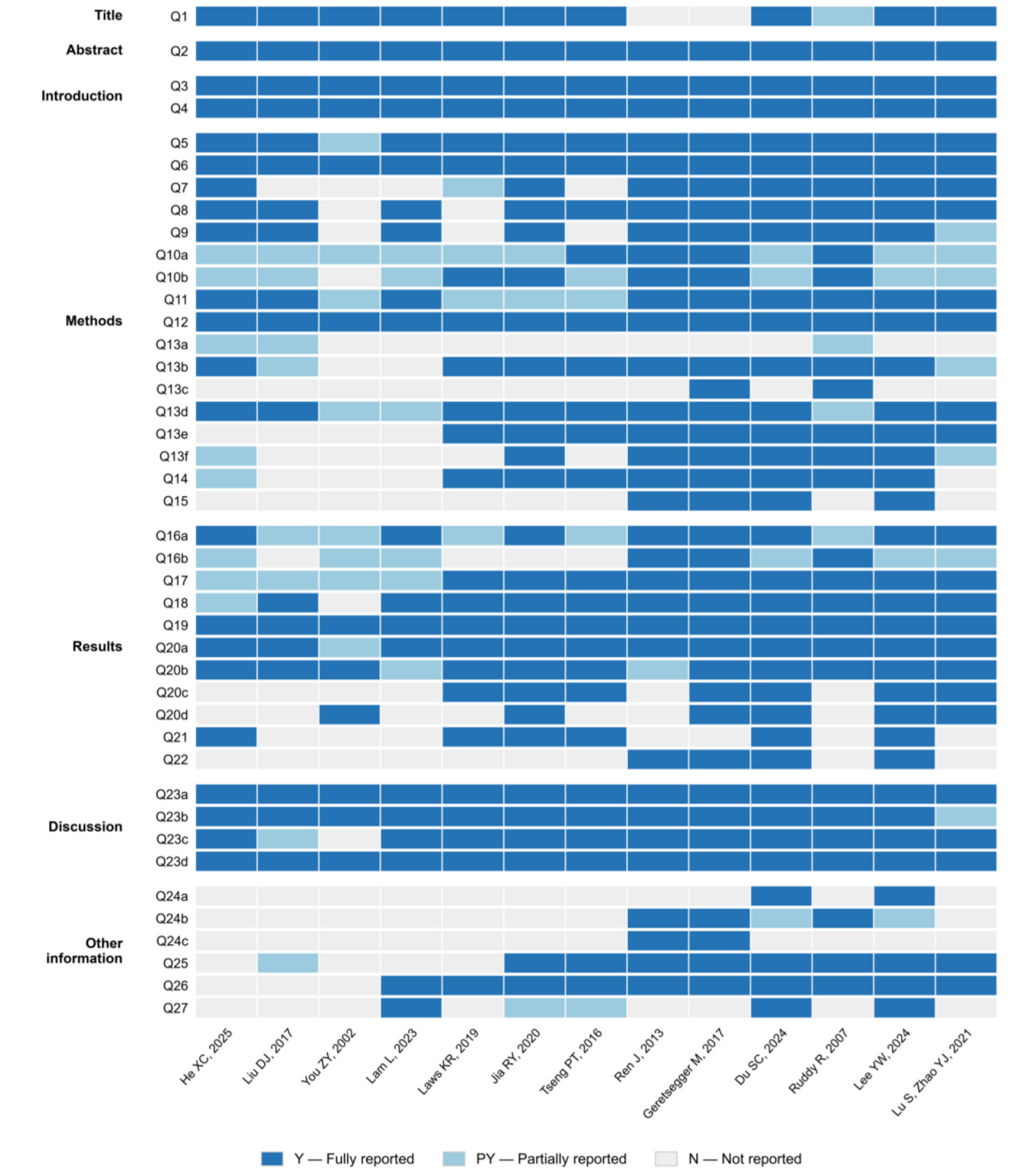


Figure 2. Reporting completeness of the included systematic reviews assessed using PRISMA 2020. Each column represents 1 systematic review, and each row represents one of the 42 PRISMA 2020 items and subitems assessed. Rows are grouped into Title, Abstract, Introduction, Methods, Results, Discussion, and Other Information. Cell colors indicate fully reported (Y; 1 point; dark blue), partially reported (PY; 0.5 points; light blue), or not reported (N; 0 points; light gray) information. The numerical codes summarize reporting completeness and do not indicate methodological quality, risk of bias, or certainty of evidence. PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

did not report pooled estimates for the positive, negative, or cognitive symptom domains assessed in this umbrella review. These 3 reviews were therefore not represented in Table 4. Detailed judgments, including the available primary-study risk-of-bias information and documented information gaps for each estimate, are provided in Table 4 and Supplementary File 4, Supplemental Digital Content 4.

For positive symptoms, 9 SRs contributed 65 primary-study occurrences representing 51 unique primary studies, yielding a CCA of 3.43%. For negative symptoms, 10 SRs contributed

93 occurrences representing 70 unique primary studies, yielding a CCA of 3.65%. Both values indicated slight overlap. Cognitive outcomes were reported by 1 SR containing 3 unique primary studies; CCA was therefore not applicable. The citation matrices and study-normalization decisions are provided in Supplementary File 5, Supplemental Digital Content 5. Repeated inclusion of primary studies limited the independence of the review-level estimates and may have contributed to cross-review concordance.[32]

3.6. Summary of findings for primary outcomes

3.6.1. Positive symptoms. Nine SRs contributed 10 positive-symptom estimates to Table 4 and Figure 3. Five estimates from 4 SRs had CIs excluding the null in favor of art therapy: He et al,[20] Tseng et al,[23] Du et al,[19] and the short- and long-term estimates from Lee et al.[28] The estimate from He et al was rated moderate certainty; the other 4 were rated very low certainty. The 5 estimates from Liu et al,[26] Laws et al,[24] Jia et al,[16] Ren et al,[27] and Geretsegger et al[25] had CIs that included the null; 3 were rated low certainty and 2 very low certainty.

These findings were interpreted in relation to effect magnitude, confidence-interval width, and certainty of evidence. Differences in statistical significance across reviews did not necessarily indicate conflicting underlying effects. Variation in the estimates, predominantly low or very low certainty, and overlapping primary studies limited conclusions regarding the effects of art therapy on positive symptoms.

3.6.2. Negative symptoms. Ten SRs[16,19,20,23–28,37] contributed 11 negative-symptom estimates to Table 4 and Figure 3. All retained estimates favored art therapy, with CIs excluding the null. Two estimates were rated moderate certainty, 7 low certainty, and 2 very low certainty. He et al[20] reported an SMD of –0.67 (95% CI –1.14 to –0.21; low-certainty evidence). The RCT-only estimate from Jia et al[16] was an SMD of –0.61 (95% CI –0.80 to –0.42; moderate-certainty evidence), and the RCT-only estimate recalculated for Tseng et al[23] was Hedges' g = 0.79 (95% CI 0.46–1.11; low-certainty evidence). Du et al[19] reported an SMD of 0.697 (95% CI 0.514–0.880; low-certainty evidence). Lee et al[28] reported short-term and long-term SMDs of 0.681 (95% CI 0.395–0.967) and 1.389 (95% CI 0.935–1.842), respectively; both were rated low certainty. For You and Wang,[37] the estimate was recalculated from the 5-trial subset reported as randomized (300 participants), excluding 1 study classified as semi-randomized in the source review, yielding Hedges' g = –0.63 (95% CI –0.86 to –0.40; very low certainty). Effect directions were harmonized for Figure 3.

The favorable direction of these estimates represented descriptive agreement across reviews, with primary-study overlap limiting their independence. Effect magnitudes varied, and 9 of the 11 estimates were rated low or very low certainty. Laws and Conway[24] reported that the apparent benefit was not retained in trials with blinded outcome assessment; their trim-and-fill analysis also yielded a nonsignificant estimate (g = –0.01, 95% CI – 0.13 to 0.10). These findings warrant caution in interpreting the apparent benefits for negative symptoms.

3.6.3. Cognitive functioning. Only 1 SR[25] examined cognitive outcomes. Geretsegger et al[25] assessed multiple domains: attention (Paced Auditory Serial Addition Test), vigilance and attention (Conners' Continuous Performance Test), memory (Wechsler Memory Scale; Clinical Memory Test), and abstract thinking (Berg's Card Sorting Test; Wisconsin Card Sorting Test–Correctly Completed Categories, WCST-Cc). Significant improvements were observed in attention assessed using the Paced Auditory Serial Addition Test (SMD = 0.72, 95% CI 0.22–1.21; low-certainty evidence), memory assessed using the Clinical Memory Test (SMD = 0.58, 95% CI 0.06–1.09; low-certainty evidence), and medium-term abstract thinking assessed using the WCST-Cc (MD = 1.18, 95% CI 0.33–2.03; very-low-certainty evidence). No significant differences were observed for vigilance and attention assessed using the Conners' Continuous Performance Test, memory assessed using the Wechsler Memory Scale, abstract thinking assessed using the Berg's Card Sorting Test, or short-term abstract thinking assessed using the WCST-Cc (MD = –0.02, 95% CI –0.07 to 0.03); these estimates were rated as low-certainty evidence.

Cross-review concordance for cognitive functioning could not be assessed because only 1 eligible SR addressed this domain. A nonsystematic contextual search identified 4 randomized trials[39–42] reporting cognitive outcomes that had not been synthesized in any included SR. Three evaluated group painting interventions (assessed with the MATRICS Consensus Cognitive Battery or the Mattis Dementia Rating Scale-2, including painting combined with aripiprazole,[39–41] and 1 evaluated dance therapy (assessed with the WCST and Montreal Cognitive Assessment).[42] The 3 painting trials were included in the visual-art SR by Du et al[19] for symptom or emotional outcomes, but their cognitive findings were not extracted or pooled. The dance trial[42] was published after the eligible cognitive SR and was therefore not covered by any included review. These findings suggest that limited review-level coverage partly reflects unsynthesized cognitive outcome data. Additional records comprised a 2025 conference abstract reporting randomized allocation, for which no full study report was available,[43] and a protocol for combined transcranial direct current stimulation and music therapy without outcome data.[44] These additional records were considered solely as contextual information and were not included in the evidence synthesis or GRADE assessment.

3.6.4. Overall concordance across outcome domains. Of the 28 retained outcome-level estimates, 25 were rated low or very low certainty, and 3 were rated moderate certainty (Fig. 3). Positive-symptom findings remained inconclusive. Negative-symptom estimates consistently favored art therapy, although effect magnitudes varied, and most estimates were supported by low- or very-low-certainty evidence. Cognitive outcomes were represented by 1 SR. Apparent agreement across reviews may partly reflect shared primary studies.

3.6.5. Post hoc exploratory quality–effect size analysis. Across the 9 SRs with standardized negative-symptom estimates, AMSTAR-2 category was not correlated with effect magnitude (Spearman's rho = 0.055; 2-sided approximate P = .889; exact permutation P = .889). Substituting Lee et al's short-term estimate did not materially alter this result (rho = –0.018; approximate P = .963; exact P = .971), and leave-one-out analyses yielded correlations ranging from –0.209 to 0.327, none of which was statistically significant under either reference model. Similarly, in the secondary analysis of 7 SRs with standardized positive-symptom estimates, no association was observed (rho = –0.270; approximate P = .558; exact P = .557), and results were unchanged after substituting Lee et al's short-term estimate. Full inputs, rank calculations, and sensitivity analyses are provided in Supplementary File 6, Supplemental Digital Content 6.

4. Discussion

4.1. Interpretation of principal findings

Negative-symptom estimates consistently favored art therapy across the reporting SRs. However, 9 of the 11 estimates were rated low or very low certainty, and overlapping primary studies limited the independence of this pattern. These findings suggest potential benefit but leave substantial uncertainty regarding its magnitude and robustness. Mechanistically, art therapy provides structured opportunities for nonverbal expression and social

Table 4

Summary of the effectiveness of art therapy on core symptoms in schizophrenia: findings from included systematic reviews.

Outcomes	Study	Studies (participants)	Effect estimation (95% CI)	Downgrading factors					Certainty of the evidence (GRADE)
				Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	
Positive symptoms	He XC, 2025	6 (369)	MD = −0.69 [−0.86 to −0.52]	−1	0	0	0	0	⊕⊕⊕○ Moderate[a]
Positive symptoms	Liu DJ, 2017	6 (573)	SMD = −0.32 (−0.72 to 0.07)	0	−1	0	−1	0	⊕⊕○○ Low[b,d]
Positive symptoms	Laws KR, 2019	6 (851)	Hedges' *g* = −0.10 [−0.34 to 0.15]	−1	−1	0	−1	0	⊕○○○ Very low[a,b,d]
Positive symptoms	Jia R, 2020	8 (602)	SMD = −0.24 (−0.61 to 0.13)	−1	−1	0	−1	0	⊕○○○ Very low[a,b,d]
Positive symptoms	Tseng PT, 2016	3 (220)	Hedges' *g* = 0.63 (0.16–1.10)	−1	−1	0	−1	0	⊕○○○ Very low[a,b,d]
Positive symptoms	Ren J, 2013	1 (43)	MD = 2.50 [−0.67 to 5.67]	−1	0	0	−1	0	⊕⊕○○ Low[a,d]
Positive symptoms	Geretsegger M, 2017	1 (96)	SMD = −0.18 [−0.60 to 0.24]	−1	0	0	−1	0	⊕⊕○○ Low[a,d]
Positive symptoms	Du SC, 2024	22 (2033)	SMD = 0.407 [0.233–0.581]	−1	−1	0	0	−1	⊕○○○ Very low[a,b,e]
Positive symptoms	Lee YW, 2024	5 (412)	(Short-term) SMD = 0.695 [0.038–1.351]	−1	−1	0	−1	0	⊕○○○ Very low[a,b,d]
		7 (682)	(Long-term) SMD = 0.667 [0.077–1.258]	−1	−1	0	−1	0	⊕○○○ Very low[a,b,d]
Negative symptoms	He XC, 2025	7 (427)	SMD = −0.67 [−1.14 to −0.21]	−1	−1	0	0	0	⊕⊕○○ Low[a,b]
Negative symptoms	Liu DJ, 2017	6 (573)	SMD = −0.39 (−0.68 to −0.09)	0	−1	0	0	0	⊕⊕⊕○ Moderate[b]
Negative symptoms—SANS RCT subset	You ZY, 2002	5 (300)	Hedges' *g* = −0.63 (−0.86 to −0.40)	−2	0	0	0	−1	⊕○○○ Very low[a,e]
Negative symptoms	Laws KR, 2019	9 (948)	Hedges' *g* = −0.42 [−0.70 to −0.14]	−2	0	0	0	−1	⊕○○○ Very low[a,e]
Negative symptoms	Jia R, 2020—RCT subgroup	11 (775)	SMD = −0.61 [−0.80 to −0.42]	−1	0	0	0	0	⊕⊕⊕○ Moderate[a]
Negative symptoms	Tseng PT, 2016	8 (607)	Hedges' *g* = 0.79 (0.46–1.11)	−1	−1	0	0	0	⊕⊕○○ Low[a,b]
Negative symptoms	Ren J, 2013	1 (43)	MD = −4.40 [−8.15 to −0.65]	−1	0	0	−1	0	⊕⊕○○ Low[a,d]
Negative symptoms	Geretsegger M, 2017	3 (177)	SMD = −0.55 [−0.87 to −0.24]	−1	−1	0	0	0	⊕⊕○○ Low[a,b]
Negative symptoms	Du SC, 2024	26 (2539)	SMD = 0.697 [0.514–0.880]	−1	−1	0	0	0	⊕⊕○○ Low[a,b]
Negative symptoms	Lee YW, 2024	8 (752)	(Short-term) SMD = 0.681 [0.395–0.967]	−1	−1	0	0	0	⊕⊕○○ Low[a,b]
		11 (1090)	(Long-term) SMD = 1.389 [0.935–1.842]	−1	−1	0	0	0	⊕⊕○○ Low[a,b]

(*Continued*)

Table 4

(*Continued*)

Outcomes	Study	Studies (participants)	Effect estimation (95% CI)	Downgrading factors					Certainty of the evidence (GRADE)
				Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	
Cognitive functioning	Geretsegger M, 2017	1 (67)	Attention (PASAT) SMD = 0.72 [0.22–1.21]	–1	0	0	–1	0	⊕⊕◯◯ Low[a,d]
		1 (67)	(Vigilance and attention (CCPT)) SMD = 0.25 [–0.23 to 0.74]	–1	0	0	–1	0	⊕⊕◯◯ Low[a,d]
		1 (67)	Memory (WMS) SMD = 0.43 [–0.06 to 0.92]	–1	0	0	–1	0	⊕⊕◯◯ Low[a,d]
		1 (60)	Memory (CMT) SMD = 0.58 [0.06–1.09]	–1	0	0	–1	0	⊕⊕◯◯ Low[a,d]
		1 (67)	Abstract thinking (BCST) SMD = 0.09 [–0.39 to 0.58]	–1	0	0	–1	0	⊕⊕◯◯ Low[a,d]
		2 (90)	Abstract thinking (WCST-Cc), short term, MD = –0.02 (–0.07 to 0.03)	–1	0	0	–1	0	⊕⊕◯◯ Low[a,d]
		1 (30)	Abstract thinking (WCST-Cc) medium term, MD = 1.18 [0.33–2.03]	–1	0	0	–2	0	⊕◯◯◯ Very low[a,d]

For You et al. (2002), the estimate was recalculated from posttreatment SANS data for 5 trials (300 participants) reported as randomized in the source review. Sun et al. (1999), classified as semi-randomized in the source review, was excluded. The recalculation is documented in Supplementary File 6, Supplemental Digital Content 6. For Tseng et al. (2016), RCT-only estimates were recalculated from the corrected trial-level effects reported in the erratum. Participant totals were summed from the original review's study-characteristics table and were not independently verified as outcome-specific analysis denominators.

GRADE Working Group grades of evidence.

High certainty: We are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: We are moderately confident in the effect estimate; the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low certainty: Our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect.

Very low certainty: We have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect.

All effect estimates were derived from randomized controlled trials (RCTs) or RCT-only subgroups and therefore started at high certainty. Downgrading factors were coded as follows: 0, not downgraded; –1, downgraded by 1 level; and – 2, downgraded by 2 levels. A value of 0 indicates that no downgrade was applied. When a domain could not be assessed reliably because of the small number of contributing studies, a value of 0 should not be interpreted as evidence that the concern was absent. Superscript letters identify the domain(s) contributing to downgrading: [a] risk of bias; [b] inconsistency; [c] indirectness; [d] imprecision; and [e] publication bias. Risk-of-bias judgments were based on assessments of the primary studies reported in the source systematic reviews; AMSTAR-2 ratings were not used as a direct proxy for primary-study risk of bias. The direction of benefit follows the coding and scale direction used in each source review and should not be inferred from the sign of the effect estimate alone.

BCST = Berg's Card Sorting Test, CCPT = Conners' Continuous Performance Test, CI = confidence interval, CMT = Clinical Memory Test, GRADE = Grading of Recommendations, Assessment, Development and Evaluation, MD = mean difference, PASAT = Paced Auditory Serial Addition Test, SMD = standardized mean difference, WCST-Cc = Wisconsin Card Sorting Test–Correctly Completed Categories, WMS = Wechsler Memory Scale.

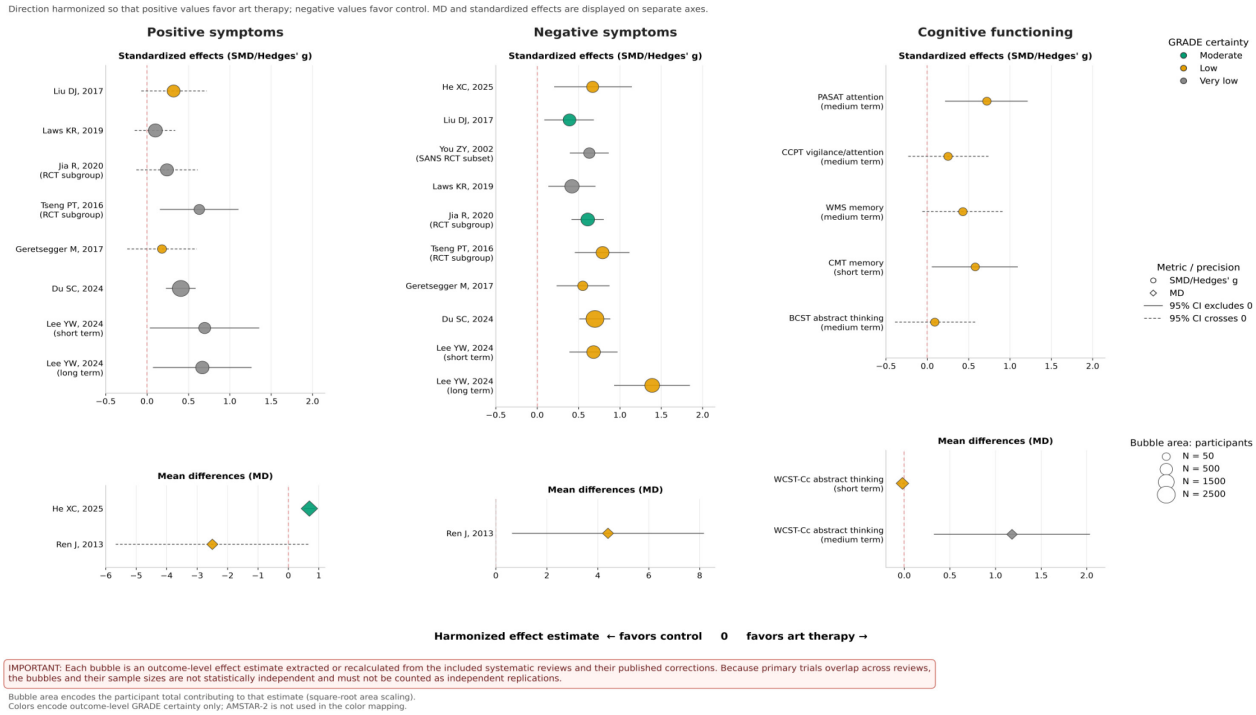


Figure 3. Evidence map of 28 outcome-level estimates. Source data and direction-coding decisions are provided in Supplementary File 7, Supplemental Digital Content 7. Each bubble represents an an outcome-level effect estimate derived from an included systematic review. Any recalculated trial subsets are documented in Supplementary Files 4 and 7, Supplemental Digital Content 4 and 7. Bubble size reflects the participant count assigned to each estimate using a compressed square-root scale. For the recalculated Tseng subsets, counts were derived from the source review's study-characteristics table and may differ from outcome-specific analysis denominators. Horizontal lines indicate 95% confidence intervals (CIs); solid lines indicate CIs that exclude zero, and dashed lines indicate CIs that include zero. Colors indicate outcome-level GRADE certainty: moderate, low, or very low. AMSTAR-2 ratings were not used to determine bubble colors or outcome-level GRADE ratings. Effect directions were harmonized so that positive values favor art therapy and negative values favor the comparator. Direction was determined from the source-reported treatment contrast, scale direction, and definition of change from baseline. Estimates for which negative values favored art therapy were multiplied by −1, with the CI transformed from [L, U] to [−U, −L]; estimates already expressed in the direction of benefit were retained. Sign reversal preserved the absolute effect magnitude, CI width, and statistical significance. Standardized effects and unstandardized mean differences are displayed separately. Because some primary studies contribute to >1 review or estimate, the bubbles do not represent independent bodies of evidence, and their participant counts do not define a unique cumulative sample size. Effect estimates are interpreted jointly with their 95% confidence intervals and outcome-level GRADE ratings. Of the 28 estimates, 3 were rated moderate certainty, 16 low certainty, and 9 very low certainty. A favorable point estimate, a large bubble, or agreement across bubbles does not by itself establish efficacy. AMSTAR-2 = A MeaSurement Tool to Assess systematic Reviews 2, CI = confidence interval, GRADE = Grading of Recommendations Assessment, Development and Evaluation, MD = mean difference, SMD = standardized mean difference, SR = systematic review.

interaction, which may support motivation and engagement.[11,13] Neurobiologically, music therapy may influence neural systems involved in emotion and social processing, although the mechanisms underlying any clinical benefit remain uncertain.[45,46] Similarly, horticultural therapy exposes patients to natural and multisensory environments, which may promote engagement and counteract apathy, although this mechanism has not been directly examined in schizophrenia.[47,48] However, this evidence is undermined by methodological limitations. Laws and Conway[24] reported a significant pooled reduction in negative symptoms across all trials (Hedges' $g = -0.42$, 95% CI −0.70 to −0.14); however, in trials with blinded outcome assessment, the pooled estimate was not statistically significant (Hedges' $g = -0.15$, 95% CI −0.37 to 0.06). Furthermore, in a trim-and-fill sensitivity analysis, the effect size became nonsignificant ($g = -0.01$, 95% CI −0.13 to 0.10). The attenuation of the estimates in these analyses reduces confidence in the robustness of the observed benefit; it does not establish whether any benefit is specific to art therapy or attributable to nonspecific factors.

Evidence regarding positive symptoms remains markedly heterogeneous and inconclusive. Of the 10 retained estimates, 1 was rated moderate certainty, 3 low certainty, and 6 very low certainty. Five estimates had CIs excluding the null in favor of art therapy, whereas the remaining 5 included the null. This discordance, together with the predominantly low or very low certainty of the evidence, limits confidence in a consistent benefit across art therapy modalities or in the magnitude of any effect. In the broader clinical context, dopaminergic dysfunction has been implicated in positive symptoms, for which antipsychotic medication remains a central treatment.[49,50] Music therapy, one of the modalities examined in this review, has also been investigated as an adjunctive intervention.[51]

Cognitive dysfunction, affecting up to 85% of patients,[30] represents a critical clinical concern in this population. Cognitive functioning remains a major evidence-synthesis gap, with only 1 eligible SR providing quantitative data.[25] Additional randomized trials of group painting and dance interventions have reported cognitive outcomes that were not synthesized in the included SRs.[39–42] An updated, prospectively registered SR focused on cognition is needed to integrate these findings and assess their certainty. Further adequately powered RCTs with active comparators, blinded outcome assessment, standardized cognitive batteries such as the MATRICS Consensus Cognitive Battery, and longer follow-up would help clarify the magnitude and durability of potential benefits.

4.2. Methodological quality and post hoc exploratory quality–effect analysis

Systematic appraisal using AMSTAR-2 identified variation in review-level methodological quality and common

methodological limitations. Applying AMSTAR-2 categories to the retained standardized estimates provided no statistically significant evidence of a monotonic association between AMSTAR-2 category and negative-symptom effect magnitude (rho = 0.055; approximate $P = .889$; exact permutation $P = .889$). This correlation remained near zero across sensitivity analyses using alternative time points and leave-one-out exclusion. No association was observed for positive symptoms either (rho = −0.270; approximate $P = .558$; exact $P = .557$). Thus, although these data did not show a monotonic quality–effect relationship, the small number of reviews and overlapping primary studies mean that they cannot distinguish a true absence of association from insufficient power.

Several factors limit interpretation: the small number of SRs, tied AMSTAR-2 categories, heterogeneity in interventions and outcome definitions, and shared primary studies across reviews. The analyses ranked absolute effect magnitudes and therefore did not distinguish beneficial from adverse directions, and covariance arising from overlapping primary studies could not be estimated from the available aggregate data. Accordingly, all coefficients and P values were interpreted descriptively. The MATISSE trial[52] (N = 417) offers a complementary perspective: art therapy did not differ significantly from standard care at 24 months; however, a single trial cannot determine whether the larger estimates reported in lower-quality reviews are attributable to methodological limitations. Collectively, these findings support a hypothesis-generating rather than confirmatory interpretation of the quality–effect relationship.

4.3. Clinical implications, limitations, and future directions

Art therapy may be considered as an adjunct to usual care, with its use informed by patient preferences and the limitations of the available evidence. Consistent with the emphasis of the American Psychiatric Association practice guideline on psychosocial rehabilitation as part of comprehensive treatment,[53] art therapy may serve as an adjunctive, recovery-oriented intervention for patients with prominent negative symptoms. From a nursing perspective, the nonverbal nature of art therapy makes it particularly suited for patients with limited verbal communication capacity.[14] Clinicians should clearly communicate the limitations of the current evidence when recommending art therapy. With respect to policymaking, investment in high-quality research is needed before incorporating art therapy into standard treatment protocols for schizophrenia.

This umbrella review has several limitations. Eligibility was restricted to reviews of schizophrenia populations, art therapy interventions, and core symptom outcomes. Reviews of mixed psychiatric populations, including that by Chu et al,[54] and reviews of interventions or outcomes outside this scope were excluded. These criteria may limit the applicability of the findings to broader rehabilitation settings, where art therapy is often delivered as part of a comprehensive program. Restriction to Chinese- and English-language publications may also have limited the geographical and cultural coverage of the evidence.

The included evidence had important methodological limitations. Eleven of the 13 source SRs were rated low or critically low using AMSTAR-2, and many contributing trials were small, lacked blinded outcome assessment, or had short follow-up periods. Primary-trial risk of bias was not independently reassessed; outcome-level GRADE judgments relied on information reported by the source SRs and may therefore reflect errors, differing assessment methods, or incomplete outcome-specific reporting. Although measured overlap across reviews was slight, some primary studies contributed to >1 estimate. Cross-review agreement and the exploratory rank-correlation analysis therefore did not provide fully independent observations.

Several continuous symptom and cognitive measures lacked validated minimally important differences. Imprecision judgments consequently required consideration of confidence-interval width and available information size and retained an element of judgment. The descriptive PRISMA 2020 reporting-completeness scores assigned equal weight to assessed items and subitems; this may obscure differences in their importance, while assessments of partial reporting may vary between reviewers. Scores for older SRs may also reflect changes in reporting expectations over time.

Future research could examine whether effects vary by intervention dose, modality, and patient characteristics, alongside adequately controlled studies of potential therapeutic mechanisms.

5. Conclusions

Across the included SRs, negative-symptom estimates consistently favored art therapy, but most were supported by low- or very-low-certainty evidence, and some drew on overlapping primary studies. Positive-symptom findings remained inconclusive, while cognitive outcomes were quantitatively synthesized in only 1 eligible SR. The magnitude and robustness of potential benefits across art therapy modalities therefore remain uncertain.

Review-level methodological quality was rated low or critically low for 11 of 13 SRs using AMSTAR-2, and outcome-level certainty was rated low or very low for 25 of 28 estimates using GRADE. The post hoc rank-correlation analysis did not detect a monotonic association between review quality and negative-symptom effect magnitude; this result remains exploratory given the small number of SRs and overlapping primary studies. Art therapy may be considered as an adjunct to usual care, taking patient preferences and the uncertainty of current evidence into account. An updated SR of cognitive outcomes and adequately powered RCTs with active comparators, blinded outcome assessment, standardized cognitive measures, and longer follow-up are needed.

Acknowledgments

We thank LetPub (www.letpub.com.cn) for its linguistic assistance during the preparation of this manuscript. We also thank Dr Qin Liu for her guidance on umbrella review methodology. This study was supported by Kunming University of Science and Technology through the project entitled 'Effects of Trajectory-Based Mandala Drawing on Cognitive Function and Alexithymia in Patients with Schizophrenia' (Project No. KUST-AN2023022Y).

Author contributions

Conceptualization: Zhongyun Zhang, Weifen Qiu.
Data curation: Zhongyun Zhang, Yunchuan Liu, Weifen Qiu.
Formal analysis: Zhongyun Zhang, Yunchuan Liu, Weifen Qiu.
Funding acquisition: Weifen Qiu.
Investigation: Zhongyun Zhang, Yao Yang.
Methodology: Zhongyun Zhang, Weifen Qiu.
Project administration: Zhongyun Zhang, Yao Yang.
Resources: Yunchuan Liu, Weifen Qiu.
Software: Zhongyun Zhang, Yao Yang.
Supervision: Yunchuan Liu, Weifen Qiu.
Validation: Yunchuan Liu, Yao Yang, Weifen Qiu.
Visualization: Zhongyun Zhang.
Writing – original draft: Zhongyun Zhang.
Writing – review & editing: Zhongyun Zhang, Weifen Qiu.

References

[1] Sawa A, Snyder SH. Schizophrenia: diverse approaches to a complex disease. Science. 2002;296:692–5.
[2] Zanelli J, Mollon J, Sandin S, et al. Cognitive change in schizophrenia and other psychoses in the decade following the first episode. Am J Psychiatry. 2019;176:811–9.
[3] Charlson FJ, Baxter AJ, Cheng HG, Shidhaye R, Whiteford HA. The burden of mental, neurological, and substance use disorders in China and India: a systematic analysis of community representative epidemiological studies. Lancet. 2016;388:376–89.
[4] Fusar-Poli P, Papanastasiou E, Stahl D, et al. Treatments of negative symptoms in schizophrenia: meta-analysis of 168 randomized placebo-controlled trials. Schizophr Bull. 2015;41:892–9.
[5] Stępnicki P, Kondej M, Kaczor AA. Current concepts and treatments of schizophrenia. Molecules. 2018;23:2087.
[6] Velligan DI, Sajatovic M, Hatch A, Kramata P, Docherty JP. Why do psychiatric patients stop antipsychotic medication? A systematic review of reasons for nonadherence to medication in patients with serious mental illness. Patient Prefer Adherence. 2017;11:449–68.
[7] Hasan A, Falkai P, Wobrock T, et. World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of schizophrenia, part 1: update 2012 on the acute treatment of schizophrenia and the management of treatment resistance. World J Biol Psychiatry. 2012;13:318–78.
[8] Samaras K, Correll CU, Mitchell AJ, De Hert M. Diabetes risk potentially underestimated in youth and children receiving antipsychotics. JAMA Psychiatry. 2014;71:209–10.
[9] Cella M, Roberts S, Pillny M, et al. Psychosocial and behavioural interventions for the negative symptoms of schizophrenia: a systematic review of efficacy meta-analyses. Br J Psychiatry. 2023;223:321–31.
[10] Regev D, Cohen-Yatziv L. Effectiveness of art therapy with adult clients in 2018-what progress has been made? Front Psychol. 2018;9:1531.
[11] Chiang M, Reid-Varley WB, Fan X. Creative art therapy for mental illness. Psychiatry Res. 2019;275:129–36.
[12] Ruddy RA, Dent-Brown K. Drama therapy for schizophrenia or schizophrenia-like illnesses. Cochrane Database Syst Rev. 2007;2007:CD005378.
[13] Case C, Dalley T. The Handbook of Art Therapy. Routledge; 2014.
[14] Samochowiec J, Szulc A, Bieńkowski P, et al. Polish Psychiatric Association consensus statement on non-pharmacological methods in the treatment of negative symptoms of schizophrenia. Psychiatr Pol. 2021;55:719–42.
[15] Fernandez-Egea E, Mucci A, Lee J, Kirkpatrick B. A new era for the negative symptoms of schizophrenia. Br J Psychiatry. 2023;223:269–70.
[16] Jia R, Liang D, Yu J, et al. The effectiveness of adjunct music therapy for patients with schizophrenia: a meta-analysis. Psychiatry Res. 2020;293:113464.
[17] Lam L, Chang WC, Grimmer K. Treatment effects of adjunct group music therapy in inpatients with chronic schizophrenia: a systematic review. Front Psychiatry. 2023;14:1215578.
[18] Wang Z, Zhong YH, Cai LL, Zhang HX, Liao JL. Effect of mandala painting therapy on improving cognitive function and self-worth in patients with residual schizophrenia. J Qiqihar Med Univ. 2020;41:1221–3. (Chinese).
[19] Du SC, Li CY, Lo YY, et al. Effects of visual art therapy on positive symptoms, negative symptoms, and emotions in individuals with schizophrenia: a systematic review and meta-analysis. Healthcare (Basel). 2024;12:1156.
[20] He XC, Liu XL, Xie ZB, Xia ZC, Meng YY, Yang Y. Meta-analysis of the effect of dance therapy on rehabilitation outcomes in patients with schizophrenia. Chin Evid Based Nurs. 2025;11:800–8. (Chinese).
[21] Kong SL, Zhao JW, Zhang YH. Effect of horticultural therapy on cognitive function in community-dwelling patients with chronic schizophrenia. Chin J Prev Med. 2019;20:342–6.
[22] Lu S, Zhao Y, Liu J, Xu F, Wang Z. Effectiveness of horticultural therapy in people with schizophrenia: a systematic review and meta-analysis. Int J Environ Res Public Health. 2021;18:964.
[23] Tseng PT, Chen YW, Lin PY, et al. Significant treatment effect of adjunct music therapy to standard treatment on the positive, negative, and mood symptoms of schizophrenic patients: a meta-analysis. BMC Psychiatry. 2016;16:16.
[24] Laws KR, Conway W. Do adjunctive art therapies reduce symptomatology in schizophrenia? A meta-analysis. World J Psychiatry. 2019;9:107–20.
[25] Geretsegger M, Mössler KA, Bieleninik L, Chen XJ, Heldal TO, Gold C. Music therapy for people with schizophrenia and schizophrenia-like disorders. Cochrane Database Syst Rev. 2017;5:CD004025.
[26] Liu DJ, Cai JJ, Wang N, et al. Systematic review of art therapy for symptom improvement and functional recovery in patients with schizophrenia. J Hebei Med Univ. 2017;38:529–34. (Chinese).
[27] Ren J, Xia J. Dance therapy for schizophrenia. Cochrane Database Syst Rev. 2013:CD006868.
[28] Lee YW, Chen TT, Hsu CW, et al. Efficacy of horticultural therapy on positive, negative, and affective symptoms in individuals with schizophrenia: a systematic review and meta-analysis of randomized controlled trials. Healthcare (Basel). 2024;12:2104.
[29] Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:j4008.
[30] Goonathilake P, Ediriweera D, Ruban R, Isuru A. Prevalence and correlates of cognitive impairment in schizophrenia: a cross-sectional study from a teaching hospital southern Sri Lanka. BMC Psychiatry. 2022;22:716.
[31] Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc. 2015;13:132–40.
[32] Pieper D, Antoine SL, Mathes T, Neugebauer EA, Eikermann M. Systematic review finds overlapping reviews were not mentioned in every other overview. J Clin Epidemiol. 2014;67:368–75.
[33] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
[34] Cumpston M, Li T, Page MJ, et al. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Database Syst Rev. 2019;10:ED000142.
[35] Xue H, Li YX, Xiao YS, Fan WH, He HX. Repetitive transcranial magnetic stimulation for Alzheimer's disease: an overview of systematic reviews and meta-analysis. Front Aging Neurosci. 2024;16:1383278.
[36] Pollock A, Farmer SE, Brady MC, et al. An algorithm was developed to assign GRADE levels of evidence to comparisons within systematic reviews. J Clin Epidemiol. 2016;70:106–10.
[37] You ZY, Wang JZ. Meta-analysis of music therapy as an adjuvant treatment for chronic schizophrenia. Acta Acad Med Sin. 2002;24:564–7. (Chinese).
[38] Zeng L, Brignardello-Petersen R, Hultcrantz M, et al. GRADE guidance 34: update on rating imprecision using a minimally contextualized approach. J Clin Epidemiol. 2022;150:216–24.
[39] Xu JQ, Wang G. Effects of group painting art intervention on negative symptoms, cognitive status, and social functioning in adult schizophrenia patients during rehabilitation phase. Guizhou Med J. 2022;46:1289–90.
[40] Yuan XP. Effects of group painting art therapy on mental symptoms, cognitive function, and sleep quality in patients with schizophrenia. J Clin Pathol Res. 2023;43:1183–9.
[41] You H, Niu ZQ, Wu XW. Effect of group painting intervention combined with aripiprazole on mental symptoms and cognitive function in patients with schizophrenia. Chin J Convalesc Med. 2023;32:633–6.
[42] Kong Y, Min H, Zhu X, Zhang L, Hu J. Clinical study of dance art therapy on hospitalized patients with chronic schizophrenia. Medicine (Baltimore). 2024;103:e37393.
[43] Gao Y, Yang K, Du D. The therapeutic effect of expressive art therapy on cognitive function in patients with schizophrenia. Schizophr Bull. 2025;51(Suppl 1):S26–7.
[44] Wei Y, He S, Luo P, et al. Combining transcranial direct current stimulation with music therapy improves cognitive function in schizophrenia: study protocol for a randomized, double-blind, sham-controlled clinical trial. Front Psychiatry. 2025;16:1543789.
[45] He H, Yang M, Duan M, et al. Music intervention leads to increased insular connectivity and improved clinical symptoms in schizophrenia. Front Neurosci. 2017;11:744.
[46] Yang M, He H, Duan M, et al. The effects of music intervention on functional connectivity strength of the brain in schizophrenia. Neural Plast. 2018;2018:2821832.
[47] Ulrich RS, Simons RF, Losito BD, Fiorito E, Miles MA, Zelson M. Stress recovery during exposure to natural and urban environments. J Environ Psychol. 1991;11:201–30.

[48] Kaplan S. The restorative benefits of nature: toward an integrative framework. J Environ Psychol. 1995;15:169–82.
[49] McCutcheon RA, Reis Marques T, Howes OD. Schizophrenia – an overview. JAMA Psychiatry. 2020;77:201–10.
[50] Jauhar S, Johnstone M, McKenna PJ. Schizophrenia. Lancet. 2022;399:473–86.
[51] Wang MQ, Lu SS, Hao L, Shi ZC, Li XF, Su L. Research progress of music therapy for core symptoms of schizophrenia. Chin J Convalesc Med. 2024;33:91–4. (Chinese).
[52] Crawford MJ, Killaspy H, Barnes TR, et al. Group art therapy as an adjunctive treatment for people with schizophrenia: multicentre pragmatic randomised trial. BMJ. 2012;344:e846.
[53] Keepers GA, Fochtmann LJ, Anzia JM, et al. The American Psychiatric Association practice guideline for the treatment of patients with schizophrenia. Am J Psychiatry. 2020;177:868–72.
[54] Chu KY, Huang CY, Ouyang WC. Does Chinese calligraphy therapy reduce neuropsychiatric symptoms: a systematic review and meta-analysis. BMC Psychiatry. 2018;18:62.