Online supplementary file 1. Supplementary introduction

Several previous literature reviews have examined aspects of the arts in relation to anxiety and/or depressive symptoms in a variety of populations. Potential benefits of visual arts, music therapy and dance/movement therapy on anxiety and depression\(^1\) and pooled arts therapies on anxiety, depression and fatigue\(^2\) have been found in a population of breast cancer patients. Creative arts interventions (art, dance/movement, drama and music), especially when led by specialist creative arts therapists, have been shown to improve depression symptoms in older adults,\(^3\) who posited physical, inter-personal, cultural, cognitive and social mechanisms of action. More broadly, in the context of leisure activities, Fancourt et al\(^4\) identified and mapped over 600 mechanisms of action linking such activities to health outcomes, covering a range of psychological, biological, social and behavioural processes operating at individual (micro), group (meso) and societal (macro) levels to produce the Multi-Level Leisure Mechanisms Network. Dance/movement therapy specifically has been shown to be beneficial for depression in adult populations\(^5\) and improve social communication, cognition and quality of life in older people with mental health conditions,\(^6\) although evidence was uncertain for a benefit on primary mental health symptoms. Focusing on tai chi, Wang et al\(^7\) found a benefit on anxiety, depression and psychological wellbeing. Elsden & Roe,\(^8\) in a systematic review of observational studies, found an association between increased arts engagement and cultural participation and improved depression outcomes in adults. Dingle et al\(^9\) considered social group interventions more broadly for depression. It excluded interventions it classified as music therapy and art therapy, but included arts-based therapies, such as dancing and group singing, and found evidence that these interventions were beneficial for mild-to-moderate depression. Aalbers et al\(^10\) and Tang et al\(^11\) found that music therapy improved depression symptoms, while Lu et al\(^12\) found a benefit for anxiety symptoms, although these reviews were limited by the focus exclusively on randomised and controlled clinical trials which may be less appropriate in a community health setting.\(^13\) Group singing has been demonstrated to improve both physical and psychological wellbeing in people with and without chronic health condition.\(^14\) Art therapy has been shown to benefit both anxiety and depression,\(^15,16\) and in a realist review, Blomdahl et al\(^17\) posited a set of multidimensional therapeutic factors linking art therapy to improved depression outcomes.

References


Online supplementary file 2. Supplementary methods

The decision to search the combination of MEDLINE, EMBASE, Web of Science and Google Scholar was supported by Bramer et al, who found that this combination performed best, offering overall recall of 98.3% of eligible studies and offering 100% recall of eligible studies in 72% of systematic reviews analysed. The published search strategy from a prior systematic review of performing arts interventions was adapted to address anxiety and depression and expanded to include artistic modalities beyond the performing arts. No filters or limits were imposed on the search. Eligibility decisions were therefore all taken at the screening stage. No automation tools were used for any aspect of the review process. Endnote was used for reference management.

It was decided to include all arts in order to provide a full picture of the potential benefits of the arts in this population. The authors acknowledge martial arts are quite different in nature than the other arts, being at the interface of the arts as traditionally conceptualised and other domains such as exercise and relaxation. On balance, the authors, while acknowledging there are valid arguments in favour of and against including the martial arts in the scope of this review, decided to favour inclusion. This decision had little impact either way on the review, as only one martial arts study, on tai chi, was identified as eligible for this review.

Only quantitative studies were considered to be relevant to the specific research questions asked. Qualitative studies would address different questions about experience and process that were outside the scope of the current work. In the context of a community health intervention, following discussion with peers at academic events, it was decided to include both randomised and non-randomised studies. Randomised controlled trials may offer optimal internal validity, but often have inferior external validity, meaning that especially in a community health context, the consideration of both randomised and non-randomised studies often enriches the evidence base, while a focus on solely randomised controlled trials can lead to a mismatch of the evidence with the needs of community health practitioners, especially as the heterogeneity of patients encountered in clinical practice is seldom replicated in trials.

Information about study funding was instead captured during the risk of bias assessment process. In the event of missing or unclear information, this was stated to be the case and no assumptions were imposed. If missing data were sufficient to preclude inclusion in the narrative synthesis, the lead author would contact the corresponding author of the included study to seek further clarification. This was not the case for any included studies.
As meta-analysis was not feasible, all available results were collected for narrative synthesis and this was not constrained to specific measures or time points. Analysis was based on publicly verifiable data from study publications. In the event of any conference abstracts being eligible for inclusion, the lead author would contact the corresponding author of the conference abstract, provided contact details could be obtained, in order to request further data or information as appropriate. No automation tools were used for data extraction or tabulation. Any disagreements regarding data extraction were resolved by discussion.

Thematic narrative synthesis was conducted. Firstly, a study profile was conducted. Regarding grouping of studies for synthesis, results were analysed firstly by outcome domain and secondly within each outcome domain by artistic modality. The feasibility of conducting a meta-analysis was considered, since where appropriate a meta-analysis can enrich the findings by providing a quantitative pooled effective estimate to supplement the findings of the thematic narrative synthesis. However, meta-analysis should only be conducted where studies are sufficiently similar to offer meaningful statistical estimates. Following consideration, it was determined that substantial methodological and clinical heterogeneity, including recruitment strategies, country-level factors, study setting (such as community vs inpatient), population characteristics, the nature of the artistic intervention, the nature of the control arm and the assessment tools used for study outcome measures would preclude meaningful and robust meta-analysis, as the number of eligible studies is insufficient to generate the extensive range of subgroup analyses that would be required.

Therefore, thematic narrative synthesis was the sole analytical technique. This approach is common in the context of community health interventions where clinical and methodological heterogeneity between studies is commonplace and has been used in certain prior reviews on the arts in mental health. It was decided not to stratify the narrative synthesis by study design in order to achieve more effective synthesis given the relatively small number of studies eligible to answer the present research questions. As a thematic narrative synthesis approach was taken, the analysis was not constrained to a specific measure of effect size, but rather focused on a broader thematic synthesis of the narrative findings of included studies to draw conclusions. All studies that met the inclusion criteria for the systematic review were eligible for inclusion in narrative synthesis. When structuring the synthesis, it was noted which studies addressed which combinations of artistic modalities and outcome domains. The study characteristics, intervention characteristics, control characteristics and study results tables produced during data extraction were used to prepare for the narrative synthesis. Due to the narrative nature of the analysis, no data conversions or techniques to address missing summary statistics were required. The study results table is structured as appropriate for
narrative synthesis rather than meta-analysis. Synthesised results are presented in the narrative synthesis.

Due to the narrative nature of the analysis, exploration of potential causes of heterogeneity in the evidence base could not be conducted by statistical means, such as subgroup analysis or meta-regression, but rather through detailed narrative exploration of the studies and how they differed, for example in settings, population characteristics, interventions, controls and outcome measures. Due to the importance of these differences in interpreting the studies and advancing the field, these methodological considerations have been discussed in detail in the discussion section. Given the narrative method, as required because the evidence base was not suitable for meta-analysis, it was not feasible to assess reporting bias, sensitivity analysis or conduct certainty assessment. Given the exploratory rather than confirmatory nature of the review in the context of an evidence base that remains limited, certainty assessment would be inappropriate.

Specialist Unit for Review Evidence (SURE) checklists were used to systematically appraise risk of bias at the study level, using the appropriate checklist tailored to the methodology of each study. The use of SURE for risk of bias assessment in the context of arts interventions follows the precedent of two prior reviews by different review teams in different patient populations. It was considered advantageous to use a related set of tools across studies, rather than use different unrelated tools for studies with different methods. An alternative approach would have been to use the Cochrane risk of bias tools for randomised and non-randomised studies, but SURE was chosen in line with precedent from previous reviews in the field of the arts and health and ensure direct comparability of the issues with risk of bias in the evidence base between the present review and those by Barnish et al and Clare and Camic. It was not possible to conduct risk of bias assessment on any studies for which only a conference abstract was available, due to insufficient level of detail to conduct robust assessment. All risk of bias assessment was performed using standardised forms. No automation tools were used. The results of the critique are shown in Appendices 5 and 6. The results of the quality assessment were used to inform the interpretation and discussion of the findings.

References

2. Barnish MS, Barran SM. A systematic review of active group-based dance, singing, music therapy and theatrical interventions for quality of life, functional communication, speech,


6. Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions. https://handbook-5-1.cochrane.org/chapter_9/9_1_4_when_not_to_use_meta_analysis_in_a_review.htm#:~:text=Meta%2Danalyses%20of%20studies%20that,interpreted%20as%20having%20more%20credibility.


Online supplementary file 3. Detailed search strategies

All databases were searched in February 2022, within 6 months of initial submission. According to best practice recommendations, no filters and limits were applied in the search. Instead, eligibility was handled in the screening process.

**Medline**

"(exp Depression/ OR depression.mp OR exp Anxiety/ OR anxiety.mp OR exp Anxiety Disorders/ OR anxiety disorder*.mp OR exp Mental Health/ OR mental health.mp) AND (exp Art Therapy/ OR art*. therap*.mp OR creative art*.mp OR exp Art/ OR performing art*.mp OR exp Music Therapy/ or music therap*.mp OR exp Dancing/ OR dance.mp OR dancing.mp OR dance therap*.mp OR exp Singing/ OR singing.mp OR exp Drama/ OR drama.mp OR theat*.mp OR exp Tai Ji/ OR tai ji.mp OR tai chi.mp OR tai ji.mp OR exp Martial Arts/ OR martial art*.mp literary art*.mp OR exp Writing/ OR write.mp OR writing.mp OR exp Poetry as Topic/ OR poet*.mp OR graphic art*.mp OR painting.mp OR plastic art*.mp OR exp Sculpture/ OR sculpture*.mp OR decorative art*.mp OR fine art*.mp)"

**EMBASE**

"((exp depression/ OR depression.mp OR exp anxiety/ OR anxiety.mp OR exp anxiety disorder/ OR anxiety disorder*.mp OR exp mental health/ OR mental health.mp) AND (exp art therapy/ OR art*. therap*.mp OR creative art*.mp OR exp art/ OR exp performing arts/ OR performing art*.mp OR exp music therapy/ OR music therap*.mp OR exp dancing/ OR dance.mp OR dancing.mp OR dance therap*.mp OR exp singing/ OR exp choir (singing)/ OR singing.mp OR drama.mp OR theat*.mp OR exp Tai Chi/ or tai ji.mp OR tai chi.mp OR tai ji.mp OR exp Martial arts/ OR martial art*.mp OR exp performing arts/ OR performing arts*.mp OR exp drawing/ OR drawing.mp OR drawing.mp OR drawing therap*.mp OR exp writing/ OR writing.mp OR writing.mp OR poet*.mp OR graphic art*.mp OR exp painting/ OR painting.mp OR plastic art*.mp OR sculpting*.mp OR decorative art*.mp OR fine art*.mp)"

**Web of Science**

"((depression OR anxiety OR anxiety disorder* OR mental health) AND (art* therap* OR creative art* OR art* OR performing art* OR music therap* OR dancing OR dance OR dance therap* OR singing OR drama OR theat* OR tai ji OR tai chi OR tai jii OR martial art* OR write OR writing OR poet* OR graphic art* OR painting OR plastic art* OR sculpture* OR decorative art* OR fine art*))"

**Supplementary searches**

Searches on Google Scholar were conducted to supplement database searches. Google Scholar is seen as a useful supplementary tool for searching but does not support advanced search strategies. Separate searches were conducted for ‘depression’ and ‘anxiety’ respectively combined with ‘art’, ‘arts’, ‘art therapy’ and each principal artistic modality. The first 10 pages were scanned for each Google Scholar search, as per common practice when using Google Scholar as a supplementary tool in systematic reviewing.
Forward and backward citation chasing was conducted manually for all full-text screened articles and relevant review articles. Backward citation chasing was conducted using the bibliography in the published full text. Forward citation chasing was conducted using the ‘cited by’ function on Google Scholar.
Online supplementary file 4. Full list of included publications


Online supplementary file 5. Full list of excluded publications from full-text screening, with reasons


49. Cui B. A study on the promotion of Bel canto singing in the treatment of depression." Basic and Clinical Pharmacology and Toxicology 2020; 127(SUPPL 3): 138. OUTCOMES.


58. Doom L. The effectiveness of schema focused drama therapy for cluster C personality disorders: An exploratory study. Arts in Psychotherapy 2018; 61: 66-76. POPULATION.


63. Fancourt D, Mak HW. What barriers do people experience to engaging in the arts? Structural equation modelling of the relationship between individual characteristics and capabilities, opportunities, and motivations to engage? PLoS ONE 2020; 15(3): e0230487. OUTCOMES.


70. Gabai M. Contribution of music therapy in depressive syndromes. [French]." La Promotion Dentaire 1974; (25): 21-23. LANGUAGE.


75. Gopi D, Preetha AK. Effectiveness of music therapy on depressive symptoms among elderly in selected geriatric homes. Intern. J. Nursing Educ 2016; 8, 163–166. INTERVENTION.


82. Gussak D. Comparing the effectiveness of art therapy on depression and locus of control of male and female inmates. Arts in Psychotherapy 2009; 36(4): 202-207. POPULATION.

83. Gussak, D. The effects of art therapy on male and female inmates: Advancing the research base. Arts in Psychotherapy 2009; 36(1): 5-12. POPULATION.

84. Hackney ME, Earhart GM. Social Partnered Dance for People With Serious and Persistent Mental Illness A Pilot Study. Journal of Nervous and Mental Disease 2010; 198(1): 76-78. POPULATION.


120. Pylvanainen P, Lappalainen R. Change in body image among depressed adult outpatients after a dance movement therapy group treatment. Arts in Psychotherapy 2018; 59: 34-45. OUTCOMES.


137. Stacey G, Stickley T. The meaning of art to people who use mental health services. Perspectives in Public Health 2010; 130(2): 70-77. OUTCOMES.


139. Stickley T, Hui A. Social prescribing through arts on prescription in a UK city: Referrers' perspectives (part 2). Public Health 2012; 126(7): 580-586. STUDY DESIGN.


144. Tavormina R, Tavormina MGM. Overcoming depression with dance movement therapy: a case report. Psychiatria Danubina 2018; 30(Suppl 7): 515-520. STUDY DESIGN.


153. Windle E, et al. The experiences of patients in the synchrony group music therapy trial for long-term depression. The Arts in Psychotherapy 2020; 67: 101580. STUDY DESIGN.

154. Xie H, et al. The efficacy of combined music therapy and Tai Chi for major depressive disorder: Study protocol for a randomized controlled trial. Medicine 2021; 100(12): e25241. STUDY DESIGN.


156. Yang, J. Study on the effect of music therapy on middle-aged and elderly individuals with depression. Revista Brasileira de Medicina do Esporte 2021; 27: 47-49. INTERVENTION


## Online supplementary file 6. Study characteristics

<table>
<thead>
<tr>
<th>First author, year</th>
<th>Country</th>
<th>Design</th>
<th>Participants</th>
<th>Inclusion criteria</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aliberti &amp; Raiola, 2021</td>
<td>Italy</td>
<td>Pre-post single arm study</td>
<td>14 female dancers with an average age of 65 and at least mild depression – members of a dance studio that hadn’t been dancing for around 18 months</td>
<td>Aged over 50, no conditions that impair physical activity, Geriatric Depression Scale &gt; 5</td>
<td>Depression (Geriatric Depression Scale)</td>
</tr>
<tr>
<td>Atiwannapat et al, 2016</td>
<td>Thailand</td>
<td>Pilot single-blinded randomised controlled trial</td>
<td>14 adults with major depressive disorder (5 active music therapy, 5 receptive music therapy, 4 counselling) – psychiatry outpatients</td>
<td>Outpatients, ICD-10 diagnosis of major depressive disorder, score of &gt;= 7 on Montgomery-Asberg Depression Rating Scale (Thai version)</td>
<td>Montgomery-Asberg Depression Rating Scale (Thai version), Thai Depression Inventory, Short Form Health-Related Quality of Life survey (SF-36, Thai version)</td>
</tr>
<tr>
<td>Chen et al, 2016</td>
<td>China</td>
<td>Randomised-controlled trial</td>
<td>200 male prisoners with at least mild depression or anxiety (100 music therapy, 100 standard care) – prison population</td>
<td>Adult male prisoners with at least 6 months left on their term, anxiety score &gt;49 on State and Trait Anxiety Inventory (Chinese version) or depression score &gt;=14 on the Beck Depression Inventory (Chinese version). No severe physical disease or psychosis, no intellectual disability and able to understand questionnaires.</td>
<td>State and Trait Anxiety Inventory (Chinese version), Beck Depression Inventory (Chinese version), Texas Social Behavior Inventory (Chinese version)</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Design</td>
<td>Participants</td>
<td>Inclusion Criteria</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------</td>
<td>-------------------</td>
<td>--------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Ching-Teng et al, 2019</td>
<td>Taiwan</td>
<td>Quasi-experimental study</td>
<td>55 older adults with mild or moderate depression (29 art therapy, 26 standard of care) – living in long-term care institutes</td>
<td>Adults aged at least 65 years, fluent in Mandarin, able to hold spoons with either hand and sit continuously for 100 minutes, score &gt;=6 on Short Portable Mental Status Questionnaire, score &gt;= 5 on Geriatric Depression Scale Short Form, diagnosis of mild depression, able to communicate, no diagnosis of severe dementia or depression</td>
<td></td>
</tr>
<tr>
<td>Chou et al, 2004</td>
<td>China</td>
<td>Randomised controlled trial</td>
<td>14 older adults with major depression or dysthymia (7 tai chi, 7 control) – psycho-geriatric outpatients</td>
<td>Major depression or dysthymia (DMS-IV) but not ‘organic’ in nature, score &gt;= 16 on Center for Epidemiological Studies Depression Scale (Chinese version), aged at least 60, no regular exercise in past 6 months, no medical contraindication to exercise, preserved physical ability, score &gt;25 on Mini Mental State Examination</td>
<td></td>
</tr>
<tr>
<td>Authors</td>
<td>Location</td>
<td>Study Design</td>
<td>Participants</td>
<td>Inclusion Criteria</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------------------</td>
<td>-----------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Ciasca et al. 2018</td>
<td>Brazil</td>
<td>Randomised controlled single-blind trial</td>
<td>56 older women with major depressive disorder (31 art therapy, 25 standard care) – psychiatry outpatients</td>
<td>Lifetime diagnosis of major depressive disorder (DSM-IV), female, aged at least 60, ability to read and write, stable on pharmacotherapy, no dementia (Mini Mental State Examination &gt;=25 for those with 1-4 years of formal schooling and &gt;=27 for those with 5 or more years of formal schooling, not drug users, no degenerative conditions or systemic disorders associated with high morbidity or mortality</td>
<td></td>
</tr>
<tr>
<td>Krueger et al. 2019</td>
<td>United States</td>
<td>Pre-post single arm pilot study</td>
<td>32 adults with anxiety or depression – psychiatry outpatients</td>
<td>Depression or anxiety (Patient Health Questionnaire-9 or Generalised Anxiety Disorder-7), no psychosis in past 6 months, no current substance abuse</td>
<td></td>
</tr>
<tr>
<td>Nan &amp; Ho, 2017</td>
<td>Hong Kong (Special Administrative Region of China)</td>
<td>Randomised controlled trial</td>
<td>106 adults with major depression (53 clay art therapy, 53 nondirective visual art) – psychiatry outpatients recruited from integrated community centres for mental wellness</td>
<td>Adults with a diagnosis of major depressive disorder by a psychiatrist, aged 18-60, pharmacologically stabilised, receiving treatment for an integrated community centre for mental wellness</td>
<td></td>
</tr>
<tr>
<td>Pinninger et al., 2012</td>
<td>Australia</td>
<td>Randomised controlled trial</td>
<td>97 adults with depression – recruited via community advertisements</td>
<td>Adults (age 18 or over) with self-reported depression/psychological distress, not pregnant, no walking or balance problems</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Design/Type of Study</td>
<td>Participants</td>
<td>Inclusion Criteria</td>
<td>Outcomes Measures</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------------------</td>
<td>------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Polanco-Zuleta et al, 2021</td>
<td>Mexico</td>
<td>Mixed methods study (quantitative results extracted)</td>
<td>27 adults with major depression (13 pharmacotherapy plus dance, 14 pharmacotherapy) – psychiatric inpatients</td>
<td>Clinical record and diagnosis of major depressive disorder without psychosis, aged 18-65, not recurrent inpatients, no motor limitations, able to maintain heart rate between 65% and 85% during sessions</td>
<td>Beck Depression Inventory</td>
</tr>
<tr>
<td>Punkanen et al, 2014</td>
<td>Finland</td>
<td>Pre-post single-arm pilot study</td>
<td>21 adults with at least mild depression or anxiety – psychiatry outpatients</td>
<td>Primary diagnosis of depression (F32 or F33 on the ICD-10 and DMS-III-R), Beck Depression Inventory score &gt;=10, no history of repeated suicidal behaviour or psychosis, no acute and severe substance abuse, able to participate in assessments and conversation, able to speak Finnish. People with anxiety were also included.</td>
<td>Beck Depression Inventory, Hospital Anxiety and Depression Scale – Anxiety subscale, Satisfaction With Life Scale, Relationship Questionnaire</td>
</tr>
<tr>
<td>Pylvanainen et al, 2015</td>
<td>Finland</td>
<td>Non-randomised controlled trial – patient choice of arm</td>
<td>33 adults with depression (21 dance/movement therapy, 12 standard of care) – psychiatry outpatients</td>
<td>Diagnosis of depression, depression as primary symptom, no psychosis, no suicide attempts or plans, no severe personality disorder, no current alcohol or substance abuse, no debilitating somatic symptoms</td>
<td>Beck Depression Inventory-II, Hospital Anxiety and Depression Scale, Symptoms Check List-90, Clinical Outcomes in Routine Evaluation – Outcome Measure</td>
</tr>
</tbody>
</table>
## Online supplementary file 7. Intervention characteristics

<table>
<thead>
<tr>
<th>First author, year</th>
<th>Content</th>
<th>Leader</th>
<th>Location</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aliberti &amp; Raiola, 2021</td>
<td>Line dancing</td>
<td>Not stated</td>
<td>Dance studio</td>
<td>60-to-65-minute sessions 3 days a week for 3 months</td>
</tr>
<tr>
<td>Atiwannapat et al, 2016</td>
<td>Group music therapy (active arm – instrument choir playing, song writing, group performance and improvisation)</td>
<td>Board-certified music therapist and music therapy assistant</td>
<td>Not stated</td>
<td>60-minute sessions weekly for 12 weeks</td>
</tr>
<tr>
<td>Chen et al, 2016</td>
<td>Group music therapy (music and imagery, improvisation and song writing)</td>
<td>Music therapist (with a prison guard in attendance outside the circle)</td>
<td>Activity room in the prison</td>
<td>90-minute sessions twice weekly for 10 weeks</td>
</tr>
<tr>
<td>Ching-Teng et al, 2019</td>
<td>Art therapy (based on expressive therapy continuum and media dimension variables)</td>
<td>The researcher, who is a social worker and psychologist, trained in leading art therapy groups</td>
<td>Not stated</td>
<td>90-to-100-minute sessions once a week for 12 weeks</td>
</tr>
<tr>
<td>Chou et al, 2004</td>
<td>Tai Chi</td>
<td>Experienced Tai Chi practitioner</td>
<td>Not stated</td>
<td>45-minute sessions three times a week for 12 weeks</td>
</tr>
<tr>
<td>Ciasca et al, 2018</td>
<td>Art therapy (guided imagery, artistic output and verbal reflections) – group sessions but with an individual focus</td>
<td>Art therapist</td>
<td>Workshop facility in a university psychiatric department</td>
<td>90-minute sessions once a week for 20 weeks</td>
</tr>
<tr>
<td>Krueger et al, 2019</td>
<td>Improvisational theatre exercises (Get-on-Track)</td>
<td>The lead author, who is a licensed clinical psychologist</td>
<td>Not stated</td>
<td>120-minute sessions once a week for 4 weeks</td>
</tr>
<tr>
<td>Nan &amp; Ho, 2017</td>
<td>Clay art therapy</td>
<td>Art therapist and a community centre for mental wellness activity worker</td>
<td>Community centre for mental wellness</td>
<td>180-minute sessions once a week for 6 weeks</td>
</tr>
<tr>
<td>Pinninger et al, 2012</td>
<td>Argentine tango dance</td>
<td>Experienced registered tango instructor and trained tango-helpers</td>
<td>&quot;The Kirk&quot;, Surry Hills, Sydney – former church turned cultural venue</td>
<td>90-minute sessions once a week for 6 weeks</td>
</tr>
<tr>
<td>Study</td>
<td>Type of Intervention</td>
<td>Provider</td>
<td>Setting</td>
<td>Duration</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>Polanco-Zuleta et al, 2021</td>
<td>Dance programme based on Latin rhythms</td>
<td>Certified dance instructor</td>
<td>Facilities in a psychiatric hospital</td>
<td>50-minute sessions daily for 6-to-8 days</td>
</tr>
<tr>
<td>Punkanen et al, 2014</td>
<td>Dance/movement therapy</td>
<td>Dance/ movement therapist</td>
<td>Not stated</td>
<td>60-minute sessions twice a week for 10 weeks</td>
</tr>
<tr>
<td>Pylvanainen et al, 2015</td>
<td>Dance/movement therapy (plus individual counselling)</td>
<td>Psychologist and dance/movement therapist</td>
<td>Not stated</td>
<td>90-minute sessions once a week for 12 weeks</td>
</tr>
</tbody>
</table>
Online supplementary file 8. Supplementary discussion

Methodological considerations

The SURE checklists were used to appraise the quality of each included study. These indicated that methodological limitations were common and some applied to the majority of included studies.

Sample

There was a large variety in the locations of the studies included, both with regard to the country of study as well as the settings from which participants were recruited. Most studies (75%) were conducted in outpatient populations, but one study was conducted in psychiatric inpatients, one study was conducted in a prison population and another was conducted in a population living in long-term care institutes. Participants from nine countries across Europe, the Americas, Australasia and Asia were included, though no studies were multinational. This has implications in terms of route clinical provision for depression and/or anxiety in primary and community care, psychology and psychiatry services, but also in terms of societal familiarity with specific art forms and cultural views both around the arts and mental health. Indeed, cultural factors may play an important role in the onset and expression of mental health conditions, as well as influencing perceptions of mental health, which in turn may impact upon the course of the condition and treatment outcomes.

As noted by a performing arts review in a different clinical area, gender roles can play an important role in the arts, especially in the area of dance. While five studies assessed dance as an artistic modality, only one used a traditional partnered dance — in the form of Argentine tango — which may reduce the influence of gender. The Pinninger et al study was not very clear regarding the dancing roles taken by participants. It stated that dance was performed in the Argentinian close-embrace tradition, and that to facilitate this trained tango-helpers were available. In the absence of information otherwise, it may be inferred that participants danced their gendered roles, although this is uncertain. While following may be more beneficial than leading in partnered dance for participants with a neurodegenerative condition such as Parkinson's disease, it is less plausible that dancing role would play as great a role in a population with depression, although there may still be a smaller benefit for following over leading due to reduced cognitive load.

The present review – unlike prior reviews such as those by Tang et al, Tang et al and Dunphy et al which included studies in which depression was co-morbid with other conditions – focused specifically on primary anxiety and/or depression, that is to say when anxiety and/or depressive symptoms are not considered to be secondary to any other medical condition or health-related circumstance. However, there can be situations in which the symptoms are not secondary to
another medical condition or health-related circumstance, but in which other life circumstances may play a contributory role. In the study by Chou et al\textsuperscript{10} in a population of Chinese psycho-geriatric outpatients, ‘organic’, i.e. unexplained, depression was excluded. Chen et al\textsuperscript{2} considered a prison population and is known that mental health problems are especially common in prison settings,\textsuperscript{11} more so than in a comparable community-living sample. Meanwhile, Aliberti & Raiola\textsuperscript{12} examined a group of people who had been denied their regular line dancing group sessions as a result of Covid policies, which have been shown to have detrimental side effects on mental health.\textsuperscript{13-14}

Across the majority of studies, there was insufficient information presented to confirm that the sample size analysed provided sufficient statistical power to detect the expected effect. Out of 12 studies, only three studies (25\%) provided a rationale for how the target sample size was reached and in one of these studies\textsuperscript{15} there was under-recruitment of seven participants per arm versus the target sample size of 60 per arm. Sampling approaches differed across studies and included convenience sampling such as in Ching-Teng et al,\textsuperscript{3} while in Aliberti & Raiola,\textsuperscript{12} participants were sampled from a very narrow sampling frame in the form of one dance studio, which may impact upon the generalisability of findings. Moreover, there was some evidence of an imbalance in baseline characteristics between arms even among randomised controlled trials. In only two of the included randomised controlled trials\textsuperscript{2,10} was it clear that the arms were sufficiently similar in terms of their demographic and clinical characteristics at baseline.

\textbf{Outcome measures}

All outcome domains except functional communication were addressed within the identified evidence base with a variety of questionnaires and assessments used for the same construct, considering the relatively low overall number of studies. For depression, the most commonly used instruments were the Beck Depression Inventory\textsuperscript{16} and the Geriatric Depression Scale\textsuperscript{17} were used by two studies, but there was quite a wide spread of instruments across studies, limiting comparability of findings. Assessments of anxiety also used a variety of instruments – such as the Generalised Anxiety Disorder 7-Item scale,\textsuperscript{18} State and Trait Anxiety Inventory\textsuperscript{19} and the Hospital Anxiety and Depression Scale - Anxiety subscale.\textsuperscript{20} Furthermore, some studies used combined assessments of anxiety and depression.

There were only two studies assessing quality of life as a unitary concept and they took different approaches to the assessment of quality of life. Atiwannapat et al\textsuperscript{21} used the Short Form Health-Related Quality of Life survey (SF-36\textsuperscript{22}), which is a widely used disease-general quality of life index. In contrast, Krueger et al\textsuperscript{23} used the NeuroQol.\textsuperscript{24} This is a disease-specific quality of life index that was designed for use in neurological populations. No validation in a primary mental health population
not secondary to a neurological condition could be identified. It is therefore unknown how well this instrument captures changes in the quality of life of people with anxiety and/or depression.

No studies assessed functional communication, which is a considerable limitation in the evidence base, given the key role participation plays in the World Health Organization’s International Classification of Functioning, Disability, and Health (ICF). However, social participation was addressed by certain studies. However, no studies assessed social participation as a unitary concept, such as through the Social Participation Questionnaire, which has been validated in people with depression.25 Instead, studies focused on specific aspects of social participation, such as relationships and performance of activities of daily living, precluding a more holistic picture of the potential benefits of the arts on social participation. Furthermore, studies on wellbeing were few in number (n=3) and took different perspectives. This reflects the fact that wellbeing is a broad concept that can be approached from different directions, including the spiritual perspective, as seen in the Body–Mind–Spirit Well-Being Inventory26 as used by Nan & Ho.15 In contrast, the Satisfaction With Life scale,27 as used by Pinninger et al6 and Punkanen et al,28 takes a far more practically-focused perspective.

While reporting of outcome results was typically good, the analysis in Atiwanapat et al21 was not structured appropriately to give a pairwise comparison between active music therapy and counselling, while in Ching-Teng et al,3 only baseline scores and not follow-up scores were reported for the Short Portable Mental Status Questionnaire, Barthel index and Karnofsky Scale scores, precluding interpretation of the potential benefit of the art therapy intervention on these measures.

Allocation

Half of the studies were not randomised and studies differed considerably in how they sampled participants, including in one case12 from an established dance group in the community, albeit which had not practised together for 18 months at the time of study onset. The absence of randomisation may be justified on ethical grounds and also for practical reasons, such as instructor availability. Pragmatic and observational approaches to studying interventions may not be inferior,29 especially in community health contexts,30 where groups of participants studied in randomised controlled trials may not be representative of those encountered in routine practice, which may limit the external validity or generalisability of trial findings to practice settings. On the other hand, randomised controlled trials generally have superior internal validity compared to other designs, and therefore including both randomised and non-randomised studies is likely in this context to provide the most informative evidence base. It should be noted that there is often a trade off in study designs.
between internal and external validity. Blinding of the intervention group is not typically feasible in arts studies, which may introduce performance biases.

**Control groups**

Most studies (75%) included a control group. Among the nine studies with a control group, seven included a standard care, waiting-list control or established pharmacological management control group. These can all be broadly conceptualised as ‘usual care’, which while providing a prima facie unifying concept, can provide substantial limitations, due to a lack of detailed reporting of what constitutes usual care and differences both within and between countries due to health system differences and clinician preferences. Furthermore, in one study, control participants were receiving daily standard care from a long-term care institute, which represents a considerably more intense intervention than would typically be seen in a control group. In addition to a waiting-list control group, Pinninger et al considered a mindfulness meditation control group. Two studies included artistic control groups, although not group-based and active – Atiwanapat et al assessed receptive music therapy, while Nan & Ho considered nondirective visual art. Atiwanapat et al also considered a group counselling control group.

**Intervention**

All but one study provided the disciplinary background of the session leader, although limited information was given regarding level of experience. All interventions were delivered by professionals in the relevant art form (although studies varied in terms of whether the session leader was an experienced practitioner of the art form or a specialist instructor) with two exceptions. In Ching-Teng et al, the session leader was a social worker and psychologist, who had been trained in leading art therapy groups, but was not a professional art therapist, while in Krueger et al, improvisational theatre exercises were led by a clinical psychologist and it was not stated that the psychologist was specifically trained in theatre. Also, in these two studies, the session leader was the lead researcher, which raises questions about the independence of the intervention delivery and the analysis and write-up of the study. While there were studies in which not all participants received the intervention from the same session leader, there was no evidence that session leaders within the same arm of any study differed substantially in their disciplinary background in a way that would influence the methodological or theoretical focus or delivery of the session. Although session leader experience was not thoroughly reported, no study reporting using student and qualified therapists to deliver the intervention to different groups of participants within a study arm.
References


Online supplementary file 9. SURE critique checklist for experimental studies.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Does the study address a clearly focused question/hypothesis?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Was the population randomised?</td>
<td>Yes, drawing lots</td>
<td>Yes, computer-generated</td>
<td>Yes, not stated</td>
<td>Yes, odd vs even entry order</td>
<td>Yes, computer-generated random numbers</td>
<td>Yes, hat</td>
<td>No</td>
</tr>
<tr>
<td>Was allocation to intervention or comparator groups concealed?</td>
<td>Blinded outcome assessment</td>
<td>Blinded outcome assessment</td>
<td>Blinded outcome assessment</td>
<td>Single-blind (understood to relate to outcome assessment, but not clearly stated)</td>
<td>Not stated</td>
<td>Baseline assessment conducted prior to randomisation</td>
<td>No blinding of participants. It is not stated whether outcome assessment was blinded</td>
</tr>
<tr>
<td>Were interventions (and comparisons) well described and appropriate?</td>
<td>Generally yes (but location of intervention not stated)</td>
<td>Yes for the intervention, but No for the control</td>
<td>Generally yes (but location of intervention not stated)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No (location of intervention not stated and limited information about the control)</td>
</tr>
<tr>
<td>Was ethical approval sought and received?</td>
<td>Yes</td>
<td>Yes – but not in the host nation as not available</td>
<td>Not stated</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Question</td>
<td>Yes</td>
<td>Unclear</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>-----</td>
<td>---------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Was a trial protocol published?</td>
<td>Unclear</td>
<td>Yes</td>
<td>Unclear</td>
<td>Registered on clinical trial sites – unclear if a full protocol paper was published</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
<tr>
<td>Were the groups similar at the start of the trial?</td>
<td>Unclear – some trends to differences in baseline characteristics</td>
<td>Yes</td>
<td>Yes</td>
<td>Unclear – statistically significant difference on one parameter (age)</td>
<td>Unclear – some trends to differences in baseline characteristics</td>
<td>Unclear – statistical significance not presented</td>
<td>Unclear – statistically significant difference on one parameter</td>
</tr>
<tr>
<td>Was the sample size sufficient?</td>
<td>Unclear</td>
<td>Unclear – it is also stated that there is no control group, when there was a standard care control group</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear – under-recruited by 7 vs target of 60 in each arm</td>
<td>Yes</td>
<td>Unclear</td>
</tr>
<tr>
<td>Were participants properly accounted for?</td>
<td>Unclear</td>
<td>Yes</td>
<td>Unclear</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Are the statistical methods well described?</td>
<td>No – covariates for ANOVAs not listed</td>
<td>No – covariates for ANOVAs not listed</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Results appropriate and clear?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Is there any sponsorship/conflict of interest stated?</td>
<td>No conflicts declared. Funding declared from university source.</td>
<td>No conflicts declared. Funding declared from university source.</td>
<td>No statement regarding conflicts. Funding declared from university sources</td>
<td>No conflicts declared. No funding information provided.</td>
<td>Not stated</td>
<td>No conflicts declared. No funding information provided.</td>
<td>No conflicts declared. No funding information provided.</td>
</tr>
<tr>
<td>Did the authors identify any limitations?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Are the conclusions the same in the abstract and full text?</td>
<td>Yes</td>
<td>Unclear</td>
<td>Not applicable – no abstract</td>
<td>No, stronger in abstract</td>
<td>Unclear</td>
<td>No, stronger in full text</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Online supplementary file 10. SURE critique checklist for cohort studies.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the study design clearly stated?</td>
<td>No</td>
<td>Yes</td>
<td>Partly</td>
<td>Yes</td>
<td>Partly</td>
</tr>
<tr>
<td>Does the study address a clearly focused question?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Are the setting, locations and relevant dates provided?</td>
<td>Partly</td>
<td>Partly</td>
<td>Partly</td>
<td>Partly</td>
<td>Partly</td>
</tr>
<tr>
<td>Were participants fairly selected?</td>
<td>Unclear – random selection but from a narrow frame (one dance studio)</td>
<td>Unclear – convenience sampling</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
<tr>
<td>Are the measures of exposures and outcomes appropriate?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Was bias considered?</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Yes</td>
<td>Unclear</td>
</tr>
<tr>
<td>Is there a description of how the study size was arrived at?</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Are the statistical methods well described?</td>
<td>Yes</td>
<td>Unclear – some details missing</td>
<td>Unclear – some details missing</td>
<td>Unclear – some details missing</td>
<td>No</td>
</tr>
<tr>
<td>Is information provided on participant flow?</td>
<td>Partly</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Are the results well described?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Is there any sponsorship/ conflict of interest reported?</td>
<td>No conflict of interest declared. Unfunded.</td>
<td>No conflict of interest declared. Government funding.</td>
<td>No conflict of interest declared. No conflict of interest declared.</td>
<td>No conflict of interest declared.</td>
<td>No conflict of interest declared. No</td>
</tr>
<tr>
<td>Did the authors identify any limitations?</td>
<td>funding information provided.</td>
<td>Government and hospital funding.</td>
<td>funding information provided.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>-------------------------------</td>
<td>---------------------------------</td>
<td>-------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Online supplementary file 11. Control profile

<table>
<thead>
<tr>
<th>First author, year</th>
<th>Synopsis of control arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aliberti &amp; Raiola, 2021</td>
<td>No control arm</td>
</tr>
<tr>
<td>Atiwannapat et al, 2016</td>
<td>Receptive music therapy; Group counselling</td>
</tr>
<tr>
<td>Chen et al, 2016</td>
<td>Standard care (in prison context)</td>
</tr>
<tr>
<td>Ching-Teng et al, 2019</td>
<td>Standard care (daily care of the long-term care institutes)</td>
</tr>
<tr>
<td>Chou et al, 2004</td>
<td>Waiting-list control</td>
</tr>
<tr>
<td>Ciasca et al, 2018</td>
<td>Standard care (all participants across arms were stable on pharmacotherapy)</td>
</tr>
<tr>
<td>Krueger et al, 2019</td>
<td>No control arm</td>
</tr>
<tr>
<td>Nan &amp; Ho, 2017</td>
<td>Nondirective visual art</td>
</tr>
<tr>
<td>Pinninger et al, 2012</td>
<td>Mindfulness meditation; waiting-list control</td>
</tr>
<tr>
<td>Study</td>
<td>Intervention Description</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Polanco-Zuleta et al, 2021</td>
<td>Pharmacological treatment (all participants across arms received pharmacological treatment)</td>
</tr>
<tr>
<td>Punkanen et al, 2014</td>
<td>No control arm</td>
</tr>
<tr>
<td>Pylvanainen et al, 2015</td>
<td>Standard care (all participants across arms received standard care)</td>
</tr>
</tbody>
</table>
### Online supplementary file 12. Results of included studies

<table>
<thead>
<tr>
<th>First author, year</th>
<th>Study results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aliberti &amp; Raiola, 2021</td>
<td>Geriatric Depression Scale scores fell by a mean (95% CI) of 2.93 (2.27, 3.59) points, <em>p</em>&lt;0.05 after the line dancing intervention, indicating an improvement.</td>
</tr>
<tr>
<td>Atiwannapat et al, 2016</td>
<td>Montgomery-Asberg Depression Rating Scale (Thai version) scores were lower at 1 month (mean change from baseline -11.75 active, -10.5 receptive, -6 counselling), 3 months (-10.8 active, -5.5 receptive, -1 counselling) and 6 months (-14.4 active, -8.25 receptive, -1.5 counselling) for both active and receptive music therapy than for counselling, with the magnitude of effect being larger for active music therapy. Thai Depression Scale scores fell in both the active (mean change from baseline 1 month -6.5, 3 months -8.4, 6 months -10.6) and receptive music therapy (mean change from baseline 1 month -6.75, 3 months -8.25, 6 months -8) groups, but increased in the counselling group (mean change from baseline 1 month 5.5, 3 months 0.5, 6 months 1). Short Form Health-Related Quality of Life survey (SF=36, Thai version) scores increased in both the active (mean change from baseline 1 month 0.74, 3 months 8.36, 6 months 14.23) and receptive (mean change from baseline 1 month 16.39, 3 months 14.48, 6 months 13.53) music therapy groups and fell in the counselling group (mean change from baseline 1 month -12.94, 3 months -19.36, 6 months -13.44). Statistical significance was not reached for any outcome measure at any of the three time points, since the ANOVA comparisons were conducted across all three groups, and the magnitude of benefit of active over receptive music therapy was generally small and found more at the later time points, while both music therapy interventions were superior to counselling. On all three measures, participants’ scores improved on both music therapy interventions and deteriorated on the counselling intervention.</td>
</tr>
<tr>
<td>Chen et al, 2016</td>
<td>The music therapy group had lower State and Trait Anxiety Inventory (Chinese version) State (mean difference -8.05, <em>t</em>= 5.870, <em>p</em>&lt;0.001) and Trait (mean difference -8.51, <em>t</em>=6.933, <em>p</em>&lt;0.001), lower Beck Depression Inventory (Chinese version, mean difference -8.81, <em>t</em>=5.765, <em>p</em>&lt;0.001) and higher Texas Social Behavior Inventory (Chinese version, mean difference 7.54, <em>t</em>=3.437, <em>p</em>&lt;0.001)</td>
</tr>
<tr>
<td>Study</td>
<td>Findings</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Ching-Teng et al, 2019</td>
<td>The art therapy group had lower Geriatric Depression Scale scores at post-test than pre-test (mean difference -5.65, t=11.927, p&lt;0.001), while no significant effect was found in the standard care control group (mean difference -0.26, t=0.466, p=0.524). There was a trend (p=0.066) to a difference in religious profile between the groups, but this was unlikely to confound the results, nor were there significant baseline differences in the outcome measure scores between groups. Short Portable Mental Status Questionnaire, Barthel index and Karnofsky Scale scores were only presented at baseline.</td>
</tr>
<tr>
<td>Chou et al, 2004</td>
<td>The tai chi group had lower Center for Epidemiological Studies Depression Scale scores at the post-test than the pre-test (Mean, SD 15.3, 9.8 vs 32.0, 9.9), while the waiting-list control group had higher scores at the post-test than the pre-test (Mean, SD 39.1, 9.7 vs 32.7, 8.7). This indicates that the tai chi group improved, while the control group deteriorated. There was a significant group by time interaction (F=24.6, p&lt;0.01).</td>
</tr>
<tr>
<td>Ciasca et al, 2018</td>
<td>The art therapy group had superior results in Geriatric Depression Scale (p=0.007), Beck Anxiety Inventory (p=0.032) and Beck Depression Inventory (p=0.025) scores compared to standard care, improving on all measures, while control participants only improved on the Beck Depression Inventory. Geriatric Depression Scale post-test art therapy mean, SD 5.5, 3.4 vs pre-test mean, SD 8.6, 3.3; post-test standard care mean, SD 6.9, 3.53 vs pre-test mean, SD 7.5, 3.65. Beck Anxiety Inventory post-test art therapy mean, SD 9.6, 8.3 vs pre-test mean, SD 18.5, 13.6; post-test standard care mean 10.8, 10.26 vs pre-test mean, SD 13.7, 15.79. Beck Depression Inventory post-test art therapy mean, SD 12.3, 10.2 vs pre-test mean, SD 20.9, 10.8; post-test standard care mean, SD 14.4, 8.73 vs pre-test mean 15.9, 8.37.</td>
</tr>
<tr>
<td>Krueger et al, 2019</td>
<td>Scores on Patient Health Questionnaire-9 (Post mean, SD 10.00, 5.42 vs pre mean, SD 14.38, 6.83) and Generalised Anxiety Disorder-7 (Post mean, SD 9.63, 5.27 vs pre mean 12.81, 4.60) were statistically significantly lower after improvisational theatre (p&lt;0.001), indicating an improvement. However, the reduction in Neuro-Qol (Post mean, SD 140.53, 16.71 vs pre mean 143.03, 15.14) scores was not statistically significant (p=0.495).</td>
</tr>
<tr>
<td>Study</td>
<td>Results</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>Nan &amp; Ho, 2017</td>
<td>There were significant group*time interactions in favour of clay art therapy compared to nondirective visual art for the Beck Depression Inventory (<em>F</em>=5.26, <em>p</em>=0.008), General Health Questionnaire (<em>F</em>=5.61, <em>p</em>=0.005) and Body-Mind-Spirit Well-Being Inventory (<em>F</em>=4.38, <em>p</em>=0.014).</td>
</tr>
<tr>
<td>Pinninger et al, 2012</td>
<td>Depression levels using the Depression, Anxiety and Stress Scale-21 were significantly reduced after tango compared to waiting-list controls (<em>d</em>=0.50, <em>p</em>=0.010), although a similar effect was found for meditation (<em>d</em>=0.54, <em>p</em>=0.025). There was no significant effect across the three groups for the other outcome variables. For anxiety using the Depression, Anxiety and Stress Scale-21 the effect size was greater for tango versus waiting-list controls (<em>d</em>=0.52) than for tango vs meditation (<em>d</em>=0.30), while for the Satisfaction with Life Scale the effect size was also greater for tango versus waiting-list controls (<em>d</em>=0.36) than for tango vs meditation (<em>d</em>=0.09).</td>
</tr>
<tr>
<td>Polanco-Zuleta et al, 2021</td>
<td>Dance programme in addition to pharmacological therapy compared to pharmacological treatment alone was associated with a greater effect size for reduction for Beck Depression Inventory scores (<em>d</em>=0.91 vs <em>d</em>=0.73), although both groups had a statistically significant improvement (<em>p</em>&lt;0.01).</td>
</tr>
<tr>
<td>Punkanen et al, 2014</td>
<td>Following dance/movement therapy, Beck Depression Inventory scores reduced significantly (post mean, SD 10.50, 5.50 vs pre mean, SD 21.67, 5.26, <em>t</em>=10.40, <em>p</em>&lt;0.001). Secondary outcomes were analysed using MANOVA, followed by paired samples <em>t</em>-tests which revealed significant improvements on the Hospital Anxiety and Depression Scale – Anxiety subscale (<em>t</em>=2.27, <em>p</em>=0.04), Satisfaction With Life Scale (<em>t</em>= -3.58, <em>p</em>&lt;0.001) and parameters from the Relationship Questionnaire (at least <em>p</em>&lt;0.05).</td>
</tr>
<tr>
<td>Plyvanainen et al, 2015</td>
<td>Symptoms Check List-90 scores decreased statistically significantly for the dance/movement therapy group relative to the standard care control group during the intervention period (Estimate = -0.425, <em>p</em>=0.011) but not the follow-up period (Estimate = 0.031, <em>p</em>=0.086). Hospital Anxiety and Depression Scale scores decreased statistically significantly during the intervention period (Estimate = -6.295, <em>p</em>=0.024), but not the follow-up period (Estimate = 0.741, <em>p</em>=0.714). Beck Depression Inventory-II and Clinical Outcomes in Routine Evaluation – Outcome Measure scores did improve more in the dance/movement therapy group than in the standard care control group, however the difference did not reach statistical significance. It should be noted that participants in the control group had statistically significantly worse Beck Depression Inventory-II and Clinical...</td>
</tr>
<tr>
<td>Outcomes in Routine Evaluation – Outcome Measure than participants in the dance/movement therapy group.</td>
<td></td>
</tr>
</tbody>
</table>
Online supplementary file 13: Supplementary results

Dance was the artistic modality that was most frequently studied (five studies\(^1\text{-}^5\)). Types of dance intervention ranged from dance/movement therapy\(^4\text{-}^5\) to a dance programme based on Latin rhythms,\(^3\) Argentine tango\(^1\) and line dancing.\(^2\) Where the professional background of the session leaders was stated (four studies), in all cases a dance instructor or dance/movement therapist was involved in leading the sessions. Only two studies stated the venue in which sessions took place, but these were not comparable: a dance studio in the community\(^2\) and facilities in a psychiatric hospital.\(^3\)

Session durations were all within the range of 50-90 minutes, but frequency of sessions and course duration differed considerably between studies. Two studies\(^2,^4\) had no control group, two studies used standard care/pharmacotherapy as the control group,\(^3,^5\) while one study had two control groups comprising a waiting-list control and mindfulness meditation.\(^1\)

Three studies addressed art therapy.\(^6\text{-}^8\) The type of art interventions differed between studies. One study specifically assessed clay art therapy,\(^7\) one study assessed art therapy based on expressive therapy continuum and media dimension variables\(^8\) and the other assessed art therapy comprising guided imagery, artistic output and verbal reflections in the format of group sessions but with an individual focus.\(^6\) In two studies, the intervention was led by an art therapist,\(^6\text{-}^7\) while in one the session leader was a psychologist and social worker with training in art therapy.\(^8\) Two studies stated the venue in which sessions took place – these were a workshop facility in a university psychiatric department,\(^6\) and a community centre for mental wellness\(^7\) respectively. Session durations were between 90 and 100 minutes in two studies,\(^6,^8\) while sessions in one study\(^7\) were notably longer at 180 minutes. All interventions were weekly but the course duration ranged from six to 20 weeks. The control group was standard care in two studies\(^6,^8\) but a nondirective visual art intervention in one study.\(^7\)

Two studies addressed music therapy.\(^9,^10\) Both interventions were relatively comparable in their content and included song writing and instrumental improvisation and were led by music therapists. While the session venue was not stated in Atiwannapat et al,\(^9\) in Chen et al\(^10\) it was an activity room in a prison reflecting the prison population of the study, posing unique challenges to group dynamics as a prison guard was required to observe sessions, positioned outside the circle. The intervention in Chen et al\(^10\) was more intensive in its frequency than in Atiwannapat et al\(^9\) – 90 minutes twice weekly as opposed to 60 minutes weekly – but the course duration was slightly shorter, i.e. 10 weeks as opposed to 12 weeks. The control group was standard care in Chen et al,\(^10\) while in Atiwannapat et al\(^9\) there were two comparison groups – receptive music therapy and counselling – the former being treated as an intervention rather than control group in the authors’ statistical
analysis but not meeting the inclusion criteria for interventions in this systematic review, and hence being considered a control group for the present purposes.

One study assessed martial arts, in the form of tai chi\textsuperscript{11} led by an experienced tai chi practitioner in the form of 45-minute sessions three times a week for 12 weeks. The venue was not stated. Participants in the intervention group were compared against a waiting list control group who received tai chi three months later, after the study follow-up assessment.

One study assessed theatrical interventions, in the form of improvisational theatre exercises\textsuperscript{12} led by a clinical psychologist in the form of 120-minute sessions once a week for 4 weeks. The venue was not stated. There was no control group.

**Different artistic modalities**

No studies directly compared different group-based active artistic modalities. The volume of studies assessing each artistic modality was small, but the number of studies differed substantially between modalities. Dance was the most studied artistic modality (5 studies), while martial arts and theatrical interventions were each assessed by only one study. Functional communication was not assessed as an outcome in any of the included studies. Moreover, not all other outcome domains were assessed with regard to each of the artistic modalities. These factors preclude any conclusions being drawn on which group-based active artistic modalities are most effective for which specific outcome domains in a population of primary anxiety and/or depression.

**Effect of baseline anxiety and/or depression severity**

No studies presented an analysis stratified by level of baseline anxiety and/or severity. However, studies that only included major depressive disorder did not present markedly different patterns of results than studies that considered a broader depression and/or anxiety population. Since these studies were not conducted in comparable contexts, the effect sizes from these studies cannot be used to compare the magnitude of effect in these different populations. Therefore, no conclusions can be drawn within the present review as to whether baseline symptom severity influences the effectiveness of arts therapies.

**References**


