



OPEN ACCESS

EDITED BY

Ricardo Valentim,
Federal University of Rio Grande do
Norte, Brazil

REVIEWED BY

Aliete Cunha-Oliveira,
Coimbra Nursing School, Portugal
Janaina Valentim,
Science and Technology of Rio Grande do
Norte (IFRN), Brazil
Lucía Sanjuán Núñez,
Universitat Autònoma de Barcelona, Spain
Juciano de Sousa Lacerda,
Federal University of Rio Grande do
Norte, Brazil

*CORRESPONDENCE

Katarzyna Grebosz-Haring
✉ katarzyna.grebosz-haring@plus.ac.at
Mette Kaasgaard
✉ mkaasgaard@health.sdu.dk

RECEIVED 08 April 2024

ACCEPTED 18 June 2024

PUBLISHED 11 July 2024

CITATION

Kaasgaard M, Grebosz-Haring K, Davies C,
Musgrave G, Shiraam J, McCrary JM and
Clift S (2024) Is it premature to formulate
recommendations for policy and practice,
based on culture and health research? A
robust critique of the CultureForHealth (2022)
report. *Front. Public Health* 12:1414070.
doi: 10.3389/fpubh.2024.1414070

COPYRIGHT

© 2024 Kaasgaard, Grebosz-Haring, Davies,
Musgrave, Shiraam, McCrary and Clift. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Is it premature to formulate recommendations for policy and practice, based on culture and health research? A robust critique of the CultureForHealth (2022) report

Mette Kaasgaard^{1,2*}, Katarzyna Grebosz-Haring^{3,4*},
Christina Davies⁵, George Musgrave⁶, Jahnusha Shiraam⁷,
J. Matt McCrary^{8,9} and Stephen Clift^{10,11}

¹Pulmonary Research Unit (PLUZ), Department of Medicine, Zealand University Hospital, Naestved, Denmark, ²Department of Regional Health Research, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark, ³Interuniversity Organisation Science and Art, Paris Lodron University Salzburg, Mozarteum University Salzburg, Salzburg, Austria, ⁴Department of Art History, Musicology and Dance Studies, Paris Lodron University Salzburg, Salzburg, Austria, ⁵Centre for Arts, Mental Health and Wellbeing, School of Allied Health and School of Humanities, The University of Western Australia, Perth, WA, Australia, ⁶Institute for Creative and Cultural Entrepreneurship (ICCE), Goldsmiths, University of London, London, United Kingdom, ⁷Music and Health Research Institute (MHRI), University of Ottawa, Ottawa, ON, Canada, ⁸Department of Human Genetics, Hannover Medical School, Hanover, Germany, ⁹Prince of Wales Clinical School, University of New South Wales, Kensington, NSW, Australia, ¹⁰Sidney De Haan Research Centre for Arts and Health, Canterbury Christ Church University, Canterbury, United Kingdom, ¹¹International Centre for Community Music, York St John University, York, United Kingdom

Introduction: Arts and health practice and research has expanded rapidly since the turn of the millennium. A World Health Organization scoping review of a large body of evidence claims positive health benefits from arts participation and makes recommendations for policy and implementation of arts for health initiatives. A more recent scoping review (CultureForHealth) also claims that current evidence is sufficient to form recommendations for policy and practice. However, scoping reviews of arts and health research—without critical appraisal of included studies—do not provide a sound basis for recommendations on the wider implantation of healthcare interventions.

Methods: We performed a detailed assessment of 18 Randomised Controlled Trials (RCTs) on arts-based interventions included in Section 1 of the CultureForHealth report using the Joanna Briggs Institute Critical Appraisal Tool for RCTs (2023).

Results: The 18 RCTs included demonstrated considerable risks of bias regarding internal and statistical conclusion validity. Moreover, the trials are substantially heterogeneous with respect to settings, health-issues, interventions, and outcomes, which limits their external validity, reliability, and generalisability.

Conclusions: The absence of a critical appraisal of studies included in the CultureForHealth report leads to an overinterpretation and overstatement of the health outcomes of arts-based interventions. As such, the CultureForHealth review is not a suitable foundation for policy recommendations, nor for formulating guidance on implementation of arts-based interventions for health.

KEYWORDS

culture, arts and health, scoping reviews, evidence, health policy

Summary

The CultureForHealth report (1) maps the literature in the field of culture, well-being and health in order to inform policy recommendations for Europe. The scoping review methodology employed did not include critical screening of the research studies included. In this paper, we report a critical assessment of 18 randomised controlled trials (RCTs) cited in Section 1 on “Culture and Health” of the CultureForHealth report using the Joanna Briggs Institute (JBI) Critical Appraisal Tool for RCTs (2022). The appraisals reveal considerable risks of bias across all trials, which limit their internal and statistical conclusion validity. The absence of a critical appraisal of studies included in the CultureForHealth report leads to an overinterpretation and overstatement of the health outcomes of arts-based interventions. As such, the CultureForHealth review is not a suitable foundation for policy recommendations, nor for formulating guidance on implementation of arts-based interventions for health.

Introduction

Since the turn of the millennium, there has been increasing attention internationally towards the potential wellbeing and health benefits of engagement with culture and creative arts activities. This interest has been motivated by the perceived need to draw on community assets outside the traditional medical field to address growing challenges in population health and demands made on healthcare systems due to funding constraints. Moreover, there is increasing recognition that medical science may face limitations in dealing with progressive long-term conditions and health inequalities and that greater efforts are needed to address social determinants of health across the life course (2). Thus, engagement with culture and creative arts are suggested as potential resources to support health through prevention, promotion, care, and treatment (3).

Efforts have been made to review the growing international body of research on culture, arts, and health in scoping and narrative evidence reviews, notably, by the All-Party Parliamentary Group for Arts, Wellbeing and Health, UK (2017, 2023) (4, 5) and in reports reviewing evidence, e.g., from Europe (5–10), the US (11), and Australia (12). One considerable boost to further developments in the field has been given by a scoping review published by the World Health Organization (WHO) in 2019, summarising findings from over 3,000 studies (13). In addition, the WHO has supported the establishment of collaborating centres for arts and health research at University College London (UCL), and additional centres in the Steinhardt School at New York University and Edgehill University (UK). In 2023, the WHO and the Jameel Arts & Health Lab (New York, USA) announced a special Lancet Global Series on the health benefits of the arts (14–16) which “will show the scientific basis of the arts’ role in health with rigour, and help position artists and scientists as necessary partners towards health and wellbeing for all” (16).

In 2022, Culture Action Europe (17) published a further scoping review - the CultureForHealth report (1) — to support: “Bottom-Up Policy Development for Culture & Wellbeing in the EU” (17). The report aimed “to synthesise existing evidence on the

positive effect of arts and cultural activities on health and wellbeing” (17) and “to inform policy recommendations for Europe” (p. 24) (1). Studies published between 2005 and November 2021 were identified for the review following a search strategy using PubMed, Scopus, and other sources (see p. 26). The authors of the CultureForHealth report acknowledge limitations with their search strategy and that “our search terms may not have covered all possible valuable aspects of our focus theme very accurately” (p. 25). Section 1 of four sections in the report focuses on health benefits from cultural and arts participation and includes details of 137 empirical studies (including controlled trials and quasi-experimental, observational, qualitative, and mixed methods studies) and reviews (including systematic, scoping, and narrative reviews). Sections 2 to 4 are concerned with culture and subjective wellbeing, community wellbeing, and COVID-19, and are not considered in this paper.

The CultureForHealth report also includes a large section describing key challenges to public health across Europe and the authors state (p. 5) (1) that culture could help to effectively tackle these challenges:

1. The need for an increased focus on health promotion and disease prevention.
2. A growing mental health crisis.
3. The need to support the broader health and wellbeing of young people.
4. Ongoing changes to the labour markets, patterns of work and the economy.
5. An ageing population.
6. The association between ill health and patterns of inequality.
7. The need to promote active citizenship.
8. The mental health challenges faced by forcibly displaced people.

The findings and emerging initiatives associated with the CultureForHealth project have been showcased at various events (see weblinks in the Appendix p. 2), a guide tailored for practitioners has been published (18), and several activities are already in the process of implementation (see weblinks in the Appendix p. 1).

The WHO and the CultureForHealth scoping reviews, and other reports, do summarise a large body of empirical evidence on the benefits of arts initiatives, especially music and dance. However, scoping reviews alone are generally not a satisfactory basis for recommending healthcare interventions (19–21). Accordingly, serious concerns have been raised regarding the limitations of the WHO report (22, 23) for its lack of critical appraisal of the studies included and for its willingness to take conclusions drawn in primary research studies at face value. Further critical papers have stressed the need for treating findings and conclusions from research and evidence reviews within arts and health with considerable caution until strong evidence has been established (22, 24–28).

In this paper, we present the results of a critical appraisal of RCTs included in Section 1 in the CultureForHealth report. We focus on RCTs as these are widely regarded as providing the most robust source of evidence on effects of interventions on health outcomes and are, moreover, central to systematic critical reviews and acknowledged frameworks for clinical guidelines for safe, effective, and evidence-based healthcare interventions (29–33). In

addition, we analyse the impact of this appraisal on the validity of the report conclusions and policy recommendations.

Methods

Inclusion and analysis procedure

Between February and April 2023, we identified all RCTs on arts/cultural interventions included in Section 1 of the CultureForHealth report (1). Table 2 in the report lists 20 RCTs, but one of the sources identified as such is not a randomised trial, and a second is not a trial report. This leaves 18 trials and data extraction was undertaken to describe characteristics of these RCTs. The RCTs were concerned primarily with singing, dance, and music listening interventions. We then undertook an appraisal of the RCTs using the revised JBI tool to assess risk of bias in RCTs (34) with respect to internal and statistical conclusion validity. The tool consists of thirteen questions with associated guidance (see Table 1), based on the Joanna Briggs Critical Appraisal Tool for the Assessment of Risk of Bias for Randomised Controlled Trials (34).

To ensure accurate appraisals, a two-stage strategy was employed:

Stage 1: Between July 2023 and November 2023, each trial report was appraised independently by two members of the research team for each domain (singing: MK, SC; dance: KG-H, SC; music listening and games: JS, SE). The assessors then met online and discussed their ratings and, where differences of opinion had arisen, an agreed judgment was reached through discussion and re-reading of the papers. Where two team members did not resolve differences, another member of the team was involved as moderator (JMM).

Stage 2: In addition, in December 2023, all papers were read again and appraised against each JBI question (34) by a single team member (SC), based on appropriate data extraction from all papers (Supplementary Table 1). This helped to ensure that a reasonable relativity in judgement could be achieved, as assessments were made against the same standards. These judgements were then independently scrutinised by two other team members (MK, KHG) and any difference of opinion discussed and resolved. Findings from the second strategy were then used to moderate agreed ratings arising from the first strategy.

Results

Characteristics of the RCTs

Section 1 of the CultureForHealth report includes reference to 18 RCTs (see Table 2). The art-form investigated varied: nine of the 18 trials were on group singing (35–43); five trials examined group dancing (44–48); three trials involved a musical intervention (49–51); and the final trial examined effects of games and painting (52). Moreover, the control arm(s) varied substantially from other arts-/culture-based activities to no intervention, or usual care (standard, health-care-based treatment).

The trials were conducted in 12 different countries: UK: 3 (35, 37, 41), China: 3 (40, 50, 51), Greece: 2 (46, 47), the USA:

2 (45, 48), and one each in Brazil (36), France (42), Germany (43), Iran (52), Italy (49), Portugal (44), Singapore (38), and Switzerland (39), i.e., only nine trials were conducted within the European Union. There was substantial heterogeneity regarding study durations, settings, health issues, participants, and outcome measure, and the RCTs varied considerably in size [smallest study (41): eighteen participants; largest (35): 258 participants]. Only five trials reported a prospective power calculation with target sample sizes achieved. The rest were either under-powered or did not report a prospective power calculation. In addition, nine trials were explicitly described as pilot studies (35, 39, 41, 48), as a “pioneer” study (36), or as “exploratory” studies (37, 40, 43, 46). Only three trials (35, 39, 41) reported on achievement of minimal clinically important differences (MCID) related to study outcomes and study findings observed.

Pre-registration, ethics, and CONSORT

Eleven trials provide details of pre-registration, including the trials’ register number, so that the protocol is accessible. However, for seven trials, there is no indication that the study was pre-registered. These include two trials conducted in China (40, 51), both studies in Greece (46, 47), one in the USA (48), and those in Italy (49) and Portugal (44). All trials apart from one (51) report ethical approval, however, only seven provide an ethics committee reference number. All reports indicate that participants gave informed consent. Seven make explicit reference to CONSORT guidelines (53) and report a standard CONSORT flow diagram. However, only one includes a CONSORT checklist (39). Two trials refer to CONSORT guidelines, but the flow diagram is either incomplete (46) or is non-standard (43). Five trial reports (40, 42, 48, 50, 52) do not explicitly refer to CONSORT but do include a participant flow chart (42, 44, 47, 50, 52). Finally, two reports (46, 51) make no reference to CONSORT and do not include any participant flow diagram.

Notably, authors in all trial reports acknowledge substantial limitations to their study and recommend further research with the conduct of large-scale trials.

Assessment of the 18 RCTs using the JBI Critical Appraisal Tool (2023)

Table 3 reports the consensus JBI tool (34) assessments of each of the 18 RCTs. Supplementary Table 1 provides example quotations from the trial reports to clarify the variations in ratings of the RCTs for questions in the JBI tool.

JBI questions with high ratings

Table 3 shows that all trials were rated positively with respect to three questions: Q6: in all trials, treatment groups were treated identically apart from the intervention(s) of interest; Q8: in all trials, outcomes were measured in the same way; and Q13: for all trials, a standard and appropriate parallel groups design

TABLE 1 JBI critical appraisal tool for the assessment of risk of bias for RCTs (34): questions and guidance.

Question	Domain and potential bias	Level of appraisal*	Guidance on what reviewers should cheque
1: Was true randomisation used for assignment of participants to treatment groups?	Selection and allocation	Study	Was a true chance (random) procedure used? For example, was a list of random numbers used? Was a computer-generated list of random numbers used? Was a statistician, external to the research team, consulted for the randomisation sequence generation?
2: Was allocation to groups concealed?	Selection and allocation	Study	Was an appropriate allocation concealment procedure used? For example, was central randomisation used? Were sequentially numbered, opaque, and sealed envelopes used?
3: Were treatment groups similar at the baseline?	Selection and allocation	Study	Are the participants from the compared groups similar with regard to the characteristics that may explain the effect? NB: Do not only consider the <i>P</i> value for testing of differences.
4: Were participants blind to treatment assignment?	Administration of intervention	Study	Was an appropriate blinding procedure used? Were participants aware of the treatment arm they were allocated to?
5: Were those delivering the treatment blind to treatment assignment?	Administration of intervention	Study	Were those delivering the treatment unaware of the assignments of participants to the compared groups?
6: Were treatment groups treated identically other than the intervention of interest?	Administration of intervention	Study	Are there other exposures or treatments occurring at the same time as the cause? Is it plausible that the effect may be explained by other exposures or treatments occurring at the same time as the cause?
7: Were outcome assessors blind to treatment assignment?	Assessment of outcomes	Outcome	Were those assessing the treatment's effects on outcomes unaware of the assignments of participants to the compared groups?
8: Were outcomes measured in the same way for treatment groups?	Assessment of outcomes	Outcome	Was the same instrument or scale used? Was the measurement timing the same? Were the measurement procedures and instructions the same?
9: Were outcomes measured in a reliable way?	Assessment of outcomes	Outcome	NB: This question is about the reliability of the measurement performed in the study, and not about the validity of the measurement instruments/scales used in the study.
10: Was follow-up complete and, if not, were differences between groups in terms of their follow-up adequately described and analysed?	Participant retention	Result	Was follow-up complete? If there are differences between groups with regard to the loss to follow-up (numbers/proportions and reasons), was there an analysis of patterns of loss to follow-up? NB: Question 10 is not about intention-to-treat [ITT] analysis. This is covered by Q11.
11: Were participants analysed in the groups to which they were randomised?	Statistical conclusion validity	Result	Were participants analysed in the groups to which they were initially randomised, regardless of whether they participated in those groups and regardless of whether they received the planned interventions (ITT analysis)? NB: The ITT analysis is a type of statistical analysis recommended in the Consolidated Standards of Reporting Trials (CONSORT) statement on best practices in trials reporting, and it is considered a marker of good methodological quality of the analysis of results of a randomised trial.
12: Was appropriate statistical analysis used?	Statistical conclusion validity	Result	Reviewers should cheque the following aspects: if the assumptions of the statistical tests were respected; if appropriate statistical power analysis was performed; if appropriate effect sizes were used; if appropriate statistical methods were used given the nature of the data and the objectives of statistical analysis (e.g., association between variables, prediction, survival analysis).
13: Was the trial design appropriate?	Statistical conclusion validity	Study	Was standard RCT design (individual randomisation, parallel groups) adopted. If not, what form of trial was conducted and what rationale is offered?

*In the 2023 revision of the JBI Appraisal Tool for the Assessment of Risk of Bias for Randomised Controlled Trials (RCTs), Questions 1–6 and 13 operate at the study level. Questions 7–9 relate to outcomes assessed and so mixed assessments are possible. Questions 10–12 are addressed in relation to each outcome result. In the current evaluation, however, RCTs were appraised at a study level for questions 10–12.

was employed. In two cases, however, the trials involved three arms [Fancourt and Perkins (37)—intervention, active control and treatment as usual; Qin (51)—music listening, painting and usual treatment]. All other trials involved two-arm intervention-control designs.

JBI questions with low ratings

All trials were rated as being at high risk of bias for two criteria: Q4: for all trials, participants were not blinded to the conditions they were allocated to, and Q5: for all trials, deliverers

TABLE 2 Characteristics of the 18 RCTs cited in Section 1 of the CultureForHealth Report (1).

RCTs included	References	Country	Target group	Intervention and control	Outcome measure(s)	Participants randomised
RCTs on group singing						
Bonilha et al.	(36)	Brazil	People with Chronic Obstructive Lung Disease (COPD)	Group singing vs. Handcraft	Maximal respiratory pressures, spirometry measures, breathlessness and quality of life (SGRQ)	43
Coulton et al.	(35)	UK	Older people 60+ years	Group singing vs. no intervention	Mental health-related quality of life (SF-12), anxiety and depression (HADS), quality of life (EQ-5D)	258
Fancourt and Perkins	(37)	UK	Mothers with symptoms of post-natal depression	Group singing vs. group play vs. usual treatment	Depression (EPDS)	153
Feng et al.	(38)	Singapore	Older people with early cognitive impairment	Group singing vs. health education	Composite measure of cognitive function (CCTS)	93
Ganzoni et al.	(39)	Switzerland	Patients with acquired or congenital structural heart disease	Group singing and breathing exercises vs. usual treatment	Maximal respiratory pressures and quality of life (MHLFQ)	22
Liu et al.	(40)	China	Patients with Chronic Obstructive Lung Disease (COPD) and mild depression	Group singing vs. health education	Depression (HADS) and quality of life (CCQ)	60
Philip et al.	(41)	UK	Patients with Chronic Obstructive Lung Disease (COPD)	Group singing in person and online vs. usual treatment	SF-36 physical and mental components, balance confidence (ABC scale), anxiety (GAD-7), depression (PHQ-9), COPD quality of life (CAT), breathlessness (MRC)	18
Pongan et al.	(42)	France	Patients with mild cognitive disorder 60+ years	Group singing vs. painting	Three measures of chronic pain (NRS, SVS, BPI), anxiety (STAI), depression (GDS), quality of life (EQ-5D)	65
Wulff et al.	(43)	Germany	Mothers with post-partum depression	Group singing vs. usual treatment	Depression (EPDS), bonding (PBQ), anxiety (STAI)	120
RCTs on group dancing						
Cruz-Ferreira et al.	(44)	Portugal	Older women 65+ years of age	Group creative dance vs. no intervention	Composite measure of physical fitness (SFT)	68
Duncan et al.	(45)	USA	Patients with clinically defined Parkinson's	Community Argentine Tango vs. usual treatment	Parkinson's severity (MDS-UPDRS), balance (MiniBESTest), gait (FOG-Q), upper extremity function (9HPT), distance walked in 6 min 6MWT	62
Lazarou et al.	(46)	Portugal	Elders 55–75 years with mild cognitive impairment	International ballroom dancing vs. no intervention	Multiple tests of cognitive function and mood and depression (e.g., MMSE, RBMT, BDI, PSS, RAVLT)	154
Kalsatou et al.	(47)	Greece	Patients with schizophrenia	Greek traditional dancing vs. usual treatment	Multiple tests of functional capacity (e.g., 6MWT, SST, BBS), mental assessment and quality of life (GAFS, QLESQ)	31
Marquez et al.	(48)	USA	Late middle-aged and older Latinos 55+ years	Latin dancing vs. health education	Multiple tests of physical health (e.g., BMI, blood pressure), cognitive function (e.g., TMT, DST)	57

(Continued)

TABLE 2 (Continued)

RCTS included	References	Country	Target group	Intervention and control	Outcome measure(s)	Participants randomised
RCTs on music						
Caprilli et al.	(49)	Italy	Children 4–13 years undergoing venipuncture	Live music and parental support vs. parental support	Observation of child distress (OSBD-A), child self-reported pain (VAS with six faces)	108
Huang et al.	(50)	China	Healthy medical students receiving orthodontic treatment	Customised brainwave music vs. CBT vs. no intervention	Self-reported pain (VAS), multiple measures of brain electrical activity	36
Qin et al.	(51)	China	Hospital patients with ankylosing spondylitis	Personalised Chinese music vs. painting vs. usual treatment	Quality of life (GQO-LI-8&4)	120
RCTs on games and painting						
Forouzandeh et al.	(52)	Iran	Children 3–12 years undergoing surgery	Interactive games vs. painting vs. usual treatment	Observation of child anxiety (mYPAS)	172

6MWT, 6-Min Walk Test; 9HPT, Nine-Hole Peg Test; ABC, Activity-specific Balance Confidence scale; BBS, Berg Balance Scale; BDI, Beck Depression Inventory; BMI, Body Mass Index; BPI, Brief Pain Inventory; CAT, COPD assessment Test; CCQ, Clinical COPD Questionnaire; CCTS, Composite Cognitive Test Score; DST, Digit Span Test; EPDS, Edinburgh Postnatal Depression Scale; EQ-5D, EuroQoL Five Dimensions scale; FOG-Q, Freezing of Gait Questionnaire; GAD-7, Generalised Anxiety Disorder scale; GAFS, Global Assessment of Functioning Scale; GDS, Geriatric Depression Scale; GQO-LI-74, Generic Quality of Life Inventory; HADS, Hospital Anxiety and Depression Scale; MDS-UPDRS, composite measure of function in Parkinson's; MHLFQ, Minnesota Living with Heart Failure Questionnaire; MiniBESTest, a measure of balance in Parkinson's; MMSE, Mini Mental State Examination; MRC, Medical Research Council breathlessness scale; mYPAS, Modified Yale Preoperative Anxiety Scale; NRS, Numerical Rating Scale; OSBD-A, Amended Form of the Observational Scale of Behavioural Distress; PBQ, Postpartum Bonding Questionnaire; PHQ-9, Patient Health Questionnaire scale; PSS, Perceived Stress Scale; QLESQ, Quality of Life Enjoyment and Satisfaction Scale; RAVLT, Rey Auditory Verbal Learning Test; RBMT, Rivermead Behavioural Memory Test; SF-12, York Short Form 12 Health Survey; SF-36, Short Form 36 Health Survey; SFT, Senior Fitness Test; SGRQ, St George's Respiratory Questionnaire; SST, Sit to Stand Test; STAI, State Trait Anxiety Inventory; SVS, Subjective Visual Scale; TMT, Trail Making Test; VAS, Visual Analogue Scale.

of interventions were not blinded to the activity (e.g., singing or dancing).

JB1 questions with varied ratings

For the remaining questions in the JB1 tool, assessments varied:

- Q1: Seven trials (35, 38, 39, 41, 44, 46, 48, 50) provide sufficient information that a satisfactory randomisation procedure was followed. In the remainder, we judged this to be unclear. Given, however, that all studies are described as RCTs, it was not possible to conclude that true randomisation did not take place for any study.
- Q2: Concealment was judged to have taken place where a satisfactory randomisation process was described. Otherwise, concealment was unclear in most cases. For two trials, however, the authors explicitly state that concealment of allocation from researchers did not happen (37, 43).
- Q3: For 12 trials, trial groups appeared to be equivalent at baseline, with no significant differences on any outcome measure reported. For five trials, however, especially where sample sizes were small, marked differences in outcome means were apparent, even though these were reported as not statistically significant.
- Q7: On the blinding of outcome assessors, Yes/No ratings were given for 11 of the 18 trials because multiple outcomes were assessed. Where the outcomes were objective assessments (e.g., lung function, cognitive function), the assessors were blinded. However, where outcomes were participant reported

- (e.g., quality of life, ratings of risk of developing depression), the assessments were not blind. For six trials, single or multiple outcomes were patient reported and so "No" rating is given on blinding of assessment at study level. In only two trials (49, 52) were outcome assessments satisfactorily blinded.
- Q9: In two trials (49, 52), information was provided on the reliability of the data reported. These are both studies of children undergoing medical procedures where an observer assessed their levels of anxiety/fear. For a further study (44), reliability estimates for one of the outcomes assessed was presented but not for all of them. In all other trials, there is no direct evidence that the data reported was reliable. In most cases, however, it is clear that previously validated assessment procedures or questionnaires and scales were employed.
- Q10: For 14 of the 18 trials, follow-up of participants throughout the trial was complete with very little or no attrition. For four trials (35, 36, 46, 48), however, attrition was 15% or greater, and so judged as substantial.
- Q11: In most trials, participants were judged to have been analysed as initially randomised, either because an intention-to-treat analysis was explicitly undertaken, or because there was little or no attrition in the course of the study. For three studies (36, 46, 48), however, this was not the case.
- Q12: For five trials (35, 39, 42, 45, 49), the statistical analysis reported was judged to be appropriate. For a further four trials, the appropriateness of the analysis was judged to be unclear. For the remaining trials, we had substantial reservations regarding the analysis undertaken, which compromise the statistical conclusion validity of the study. These concerns are elaborated in the Discussion section below.

TABLE 3 Assessments of 18 RCTs cited in Section 1 of the CultureForHealth Report (1) using the JBI assessment tool (34).

Item	1	2	3	4	5	6	7	8	9	10	11	12	13
RCTs	Truly randomised	Allocation concealed	Similar at baseline	Participants blind to treatment	Deliverers blind to treatment	Groups treated identically	Outcome assessors blind*	Outcomes measured same way	Outcome measure(s) Reliable	Follow-up complete	Participants analysed as allocated	Statistical analysis appropriate	Design appropriate
Bonilha et al. (36)	Yellow	Yellow	Yellow	Red	Red	Green	Green	Red	Red	Red	Red	Red	Green
Coulton et al. (35) ^{a,b}	Green	Green	Green	Red	Red	Green	Red	Red	Red	Red	Green	Green	Green
Fancourt and Perkins (37)	Yellow	Red	Green	Red	Red	Green	Red	Red	Red	Green	Green	Red	Green
Feng et al. (38) ^c	Green	Green	Green	Red	Red	Green	Green	Red	Red	Red	Green	Yellow	Green
Ganzoni et al. (39) ^b	Green	Green	Yellow	Red	Red	Green	Green	Red	Red	Red	Green	Green	Green
Liu et al. (40)	Yellow	Yellow	Green	Red	Red	Green	Red	Red	Red	Red	Green	Red	Green
Philip et al. (41) ^a	Green	Green	Yellow	Red	Red	Green	Green	Red	Red	Red	Green	Red	Green
Pongan et al. (42) ^{a,b}	Yellow	Yellow	Green	Red	Red	Green	Green	Red	Red	Red	Green	Green	Green
Wulff et al. (43) ^c	Yellow	Red	Green	Red	Red	Green	Red	Red	Red	Red	Green	Yellow	Green
Cruz-Ferreira et al. (44)	Green	Green	Green	Red	Red	Green	Green	Red	Red	Red	Green	Yellow	Green
Duncan et al. (45) ^{a,b}	Yellow	Yellow	Green	Red	Red	Green	Green	Red	Red	Red	Green	Green	Green
Kalsatou et al. (47)	Yellow	Yellow	Green	Red	Red	Green	Green	Red	Red	Red	Green	Yellow	Green
Lazarou et al. (46)	Green	Green	Yellow	Red	Red	Green	Green	Red	Red	Red	Red	Red	Green
Marquez et al. (48)	Yellow	Yellow	Yellow	Red	Red	Green	Green	Red	Red	Red	Red	Red	Green
Caprilli et al. (49)	Yellow	Yellow	N/A	Red	Red	Green	Red	Red	Green	Green	Green	Green	Green
Huang et al. (50)	Green	Green	Green	Red	Red	Green	Green	Red	Red	Red	Green	Red	Green
Qin et al. (51)	Yellow	Yellow	Green	Red	Red	Green	Red	Red	Red	Red	Green	Red	Green
Forouzandeh et al. (52)	Yellow	Yellow	Green	Red	Red	Green	Green	Red	Red	Red	Green	Red	Green

Red = No, Yellow = Unclear, Green = Yes, Q3, N/A = no baseline assessment. *Patient-reported outcomes not blind. ^aQ11, Intention-to-treat analysis undertaken; ^bQ12, Prospective power calculation/sample achieved; ^cQ12, Prospective power calculation/sample not achieved.

Discussion

In the present study, we undertook a critical appraisal of the 18 RCTs included in Section 1 of the CultureForHealth report (1), which the authors of the report had not carried out. We analysed the basic characteristics and assessed risks of bias, using the JBI appraisal tool (34). The RCTs were characterised by substantial heterogeneity and high risks of bias, affecting both internal and external validity, and, hence, compromising the claims and recommendations stated within the CultureForHealth report.

Characteristics of the included RCT

There are key challenges with the basic conduct and reporting of many of the trials included in the CultureForHealth report. Firstly, approximately half of the trials had not been pre-registered, some did not conform to CONSORT guidance, and one trial failed to report ethical approval. These considerations alone should have resulted in extreme caution in even including these trials for consideration in a review. Secondly, the RCTs are characterised by high study heterogeneity regarding all aspects, as highlighted in Table 2. Thirdly, given that the outcomes assessed are mostly based on patient-reported outcomes, the clinical relevance of findings reported is not always clear, with no reference to established cut-off values or MCID. Additionally, distinguishing between primary and secondary outcomes is often not clear. Fourthly, many trials are extremely small and under-powered, and, furthermore, involve predominantly females, so the extent to which the findings could be generalised even to males is questionable. Finally, descriptions of the arts-based interventions used are frequently lacking, including any disease-specific adaptations of the intervention, besides a frequent absence of appropriate and validated checklists for reporting within research on healthcare interventions [e.g., CONSORT (53)].

These challenges severely limit both study replication to confirm or disconfirm findings and the reliable translation of a particular intervention to clinical and/or public health settings. Notably, nine trials are explicitly described as being a “pilot”, “pioneer”, or “exploratory”. As such, our analysis demonstrates that before even considering the assessments of the trials using the JBI tool, there are aspects of the trials which immediately raise questions over external validity and generalisability.

JBI appraisals of the RCTs

There are several issues affecting all or most of the trials which potentially introduce substantial risks of bias with respect to the outcomes. In all cases, both participants in interventions and those facilitating them were not blinded. This is unavoidable for arts-based activities of all kinds, but nevertheless, awareness introduces the potential for expectation and social desirability biases (54, 55). A further source of non-blinding arises with outcome assessment. Where objective measures are taken by a member of the research team, all trial reports properly describe the assessors as blinded, in all trials but one (52). Some of the outcome measures are

participant-reported, which is exclusively the case in six trials (35, 37, 40, 43, 49, 51). Participants were, thus, aware not only of the nature of the intervention, but were also asked to report on the impact of the activity.

A further potential source of bias arises in 15 out of 18 trials as there is no reporting of reliability of data gathered to evaluate the interventions. The JBI appraisal is very clear that the issue is not whether the measures employed had been previously validated, but whether reliability estimates for the data itself are reported. This could readily have been done in two ways: estimates of internal consistency for scales used, or examining correlations between baseline and follow-up assessments, but in only three trials is data reliability reported (35, 44, 50, 52).

In addition to “internal validity”, the JBI appraisal involves assessing “statistical conclusion validity”. For most trials, follow-up is complete (Q10), or levels of attrition are very low, but for four cases, attrition is quite substantial (35, 36, 46, 48). This represents a potential source of bias for these trials, as the sample followed up differs from the initially randomised sample. For four trials, however, an intention-to-treat analysis was explicitly employed, and so participants were analysed as randomised (Q11) (35, 41, 42, 45). For the remainder of the trials where follow up was complete, we judged that participants were analysed as allocated.

The picture is much more varied for Q12, however, on whether the statistical analysis reported was appropriate. Six out of 18 trials were considered to meet exacting demands for statistical analysis (35, 38, 39, 42, 43, 45). Ganzoni et al. (39), for example, are meticulous regarding their account of the intention-to-treat analysis undertaken with reference to a prospective power calculation and use of two-tailed tests. In addition, they include a clear statement of testing for normality and matching the statistical tests employed to the measurement characteristics of the outcome variables.

For the remaining twelve trials, however, there is no prospective power calculation, and in most cases no reference to MCID scores or effect sizes. Accounts of the statistical analysis adopted may appear satisfactory, but with details lacking (e.g., no information on whether *t*-tests were one-tailed or two-tailed), and problems with the reporting of results (e.g., a failure to report *t*-values but only *p*-values).

In two UK trial reports (37, 41), there are also concerns over the details of the statistical strategy adopted, and we discuss the approach adopted in detail to illustrate the threats to validity of the conclusions drawn. Fancourt and Perkins (37), for example, report no differences in depression across three arms of their trial (singing, play, and usual care) after 10 weeks for their total sample of mothers with scores on the Edinburgh Postnatal Depression Scale of ten or greater. They then focus on a smaller sample of mothers with scores of 13 or greater, and, again, find no differences between the trial arms at 10 weeks. Finally, they focus on changes over the first 6 weeks of the trial and find an apparent faster reduction in depression scores over the first 6 weeks of the trial. However, across the first 6 weeks, there was no difference between change for the singing and play groups. Nevertheless, the conclusion reached focuses on the rate of change in the singing group, and it is claimed that “evidence that singing interventions could speed the rate of recovery in women

affected by symptoms of PND... could have clinical relevance". (p. 120) (37).

In the case of Philip et al. (41), the appropriate use of non-parametric techniques where data was not normally distributed, and the intention-to-treat approach are both excellent features of the analysis undertaken, but the use of one-tailed criterion may be criticised as too liberal, as for an exploratory study a two-tailed approach would be recommended (56). This is especially the case given that the reported p -value for changes in a measure of depression is 0.049—at the very limit for rejecting the null hypothesis, and, thus, far from convincing.

Context of the CultureForHealth report

Societies and healthcare systems all over the world face unprecedented challenges in public health, and pressures on health care services and escalating costs (57). Evidence-based medicine, despite extraordinary advances, is also limited in what it can offer people with enduring and progressive health conditions. It is undeniable that cross-sector collaboration and inter-disciplinary working is needed to meet these demands. Improvements in public health will only come about by addressing the root causes of ill-health and health inequities, which has been clear since the seminal work of 19th century social reformers (58), and which currently are amply reinforced by the work of Marmot on the social determinants of health (59).

The arts may be one possible sector to integrate into a holistic strategy for improving health and wellbeing in all sectors of society and across the lifespan (3). The field of arts and health practice and research has indeed made considerable efforts over the last quarter century, and the trials we are considering demonstrate the commitment, passion, and collaborative energy of funding bodies, healthcare professionals, creative artists, and researchers to explore new frontiers in arts and health interventions.

Most of the trials cited in the CultureForHealth report, however, only focus on group singing and dancing, and address specific health issues related to ageing, chronic health conditions, and mental health. As such, the studies may have relevance to addressing just three of the public health challenges identified in the report: the need for greater focus on health promotion, a growing mental health crisis, and an ageing population. In contrast, however, they have little or no relevance to remaining challenges which the report suggests that the arts can help address, such as the need to address the health and wellbeing of young people, addressing health inequalities, and the needs of forcibly displaced people. Moreover, the trials are by no means representative of the international work in arts and health, and many other trials are missing in the report due to its limited search strategy (for example, the report includes three trials concerned with COPD, but misses other key trials published within their time envelope) (60–62). Nevertheless, it is clear that this, albeit selective, corpus demonstrates how global the interest in “creative health” is, with studies as far

afield from east to west as China and Brazil, and also in Europe from north to south in the UK and Greece, although no considerations are given in the report on any potential challenges related to drawing conclusions across heterogeneous cultural contexts.

Although scoping reviews do not necessarily involve critical appraisal of the studies included (19, 20), the CultureForHealth report's limited research strategy results in even higher risk of bias even at a basic methodological level. Specifically, the report does not offer critical considerations regarding methodological study type and quality, risks of bias, or descriptions of interventions employed. Moreover, the report does not offer critical considerations on the clinical relevance of outcomes and findings to assess the validity of the evidence claims stated, for example regarding the claim that singing improves respiratory function which, however, has not been demonstrated (62–64).

Moreover, no consideration is given to the dangers of drawing conclusions based on underpowered trials (65). Instead, the authors take findings at face value, and, given the lack of any basic scrutiny, the reporting only meets the very first step in Bloom's taxonomy (66). Moreover, as demonstrated in the present paper, the included primary RCTs do not meet current standards for good practice regarding developing, investigating, reviewing, and evaluating the benefits of healthcare interventions, nor for rigorous synthesis [e.g., the GRADE framework (67, 68)].

Notably, in all 18 trial reports, the authors themselves give due attention to the limitations of their studies and the need for further research. This is, however, not clearly transmitted in the report. Moreover, despite previous criticism regarding the methodology and conclusions of the WHO report (22–24), the CultureForHealth report does not address these concerns and presents no clear aim to enhance the methodological quality of research in the field. Immediate confidence in the criticality of reported findings is further compromised e.g., by the identification of a source described as a “systematic review” on singing and health (1, 69), which, however, is no more than a spreadsheet of selected studies, even depicting incorrect information, e.g., regarding singing and respiratory function (62, 69).

Taken together, the CultureForHealth report (1), similarly to the WHO report (13), does not express any hesitations or cautions regarding conclusions and recommendations stated, nor does it sufficiently consider the following aspects:

- Quality and certainty of the evidence (29, 70);
- Standard frameworks for synthesising the body of evidence regarding complex and health-care interventions, e.g., c.f., the GRADE framework (30, 31, 71);
- Patient safety (32, 33);
- Standards for development/definition of a core outcome set (COS) in health care interventions (72, 73);
- Standards for evaluating the implementation of healthcare interventions, including acceptability, fidelity, feasibility, scalability, and sustainability (74).

Thus, based on findings presented in the present paper and the previous concerns raised, we stress the limitations of scoping reviews and grey literature reports which further perpetuate a lack of scientific rigour and trustworthiness within the field. What is needed are systematic reviews which have undergone thorough external peer-review and which properly assess factors relevant to practice and policy development (22, 75). We welcome the new Lancet Global Series initiative (14–16) and encourage further high-quality, rigorous research, alongside a fruitful and ambitious academic discussion to form a qualified basis for informing policy and practice. However, we have also found a need to express concerns to the initial opinion piece coming from the Jameel Arts & Health Lab (15, 76).

Strengths and limitations of this study

The present paper builds upon previous critiques of reports and research within arts and health, but is the first study to present in-depth scrutiny of the primary studies on which evidence-claims, conclusions, and recommendations are made in the CultureForHealth report. We have approached the exercise of critique with a proper sense of humility and propriety based on an acknowledged framework, and our paper exemplifies the steps and thoroughness needed for assessment and evaluation of primary studies as a basis for drawing conclusions ahead of formulating any evidence-based recommendations for policy or practice. We did not assess or evaluate other study types included in the report (e.g., quasi-experimental and qualitative studies) and further critical scrutiny may be warranted (23). Given the focus in this paper on the treatment of RCTs, our critique provides a constructive outline and perspectives for future research and provides useful information for readers of the CultureForHealth report.

Conclusion

The CultureForHealth report substantially fails to meet current standards for good practice regarding evaluation of healthcare interventions. The report is not a suitable foundation for policy or practice recommendations nor for current scaled-up implementation of arts for health initiatives. Future trials should adhere to established high-quality standards for the development and evaluation of healthcare interventions, and robust, critical systematic reviews and meta-analyses are needed as a basis for evaluation of the field before considering policy formulation and practice guidelines.

Author contributions

MK: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualisation, Writing – original draft. KG-H: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Supervision, Validation, Writing – original draft. CD: Conceptualisation, Methodology, Resources,

Validation, Writing – review & editing. GM: Conceptualisation, Methodology, Resources, Validation, Writing – review & editing. JS: Data curation, Resources, Validation, Writing – review & editing. JM: Conceptualisation, Data curation, Formal analysis, Methodology, Resources, Validation, Writing – review & editing. SC: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualisation, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. KG-H was supported by Land Salzburg. The open access publication costs for article were covered by Paris Lodron University Salzburg, Salzburg, Austria. The funders had no role in the conceptualization, design, literature searches, analysis, interpretations, preparation of the manuscript or decision to publish.

Acknowledgments

The authors would like to thank: Simon Eldaoud, student (Music and Health Research Institute (MHRI), University of Ottawa, Canada), for assistance in the appraisal of selected RCTs; Helena Daffern, Professor, PhD (School of Physics, Engineering and Technology at the University of York, United Kingdom), for feedback on an earlier draft of this paper; Arne C. Bathke, Professor, PhD, Dean (Faculty of Digital and Analytical Sciences, Paris Lodron University Salzburg, Austria), for advice on statistical issues and feedback on an earlier draft of this paper; Timothy Barker, PhD (School of Public Health, Faculty of Health and Medical Sciences, The University of Adelaide, Evidence-based Healthcare Research, JBI Adelaide GRADE Centre, Australia), for guidance on the use of the JBI assessment tool; students (Interuniversity Organisation Science and Art, Department of Musicology and Dance Studies, Paris Lodron University Salzburg, Mozarteum University Salzburg, Salzburg, Austria), for co-assessment of the RCTs on dancing: Antonella Dilber, Hanna Hrubá, Lu Lorenz, Khatia Maisarudze, Marie Pichler, Tamara Ranxl, Enya Marie Rieder, Emilia Schatzl, Lisa Stelzer, Johanna Wunsch, Berthold Beyerlein, Barbara Bischof, Mark Hebell, Onjeon Lee, Miranda Lipovica, Lily Marie Ludwig, Maryam Rostamivand, Dustin Waskow, Eila Büche, Elena Grolmusz, Kaori Shimada, Rojas Trcka, Rebecca Naß, Ronja Rühmkorff, Elena Schwalbe.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Zbranca R, D'Amato M, Blaga O, Kiss K, Dascal MD, Yakobson D, et al. *CultureForHealth Report. Culture's contribution to health and well-being*. A report on evidence and policy recommendations for Europe. CultureForHealth. Culture Action Europe. (2022). ISBN: 978-2-9603183-0-2. Available online at: https://www.cultureforhealth.eu/app/uploads/2023/02/Final_C4H_FullReport_small.pdf (accessed March 15, 2024).
- Goldblatt P. *Health inequalities, lives cut short*. UCL Institute of Health Equity. (2024). Available online at: <https://www.instituteofhealthequity.org/resources-reports/health-inequalities-lives-cut-short/read-the-report.pdf> (accessed February 23, 2024).
- Clift S. Creative arts as a public health resource: moving from practice-based research to evidence-based practice. *Perspect Public Health*. (2012) 132:120–7. doi: 10.1177/1757913912442269
- The All-Party Parliamentary Group on Arts, Health and Wellbeing (APPG AHW). *Creative Health: The Arts for Health and Wellbeing*. All-Party Parliamentary Group on Arts, Health and Wellbeing Inquiry Report (2017). Available online at: https://ncch.org.uk/uploads/Creative_Health_Inquiry_Report_2017_-_Second_Edition.pdf (accessed March 15, 2024).
- The National Centre for Creative Health (NCCH) and the All-Party Parliamentary Group on Arts, Health and Wellbeing (APPG AHW). *Creative Health Review*. (2023). Available online at: <https://ncch.org.uk/creative-health-review> (accessed March 15, 2024).
- Fancourt D, Bone JK, Bu F, Mak HW, Bradbury A. *The Impact of Arts and Cultural Engagement on Population Health: Findings from Major Cohort Studies in the UK and USA 2017 – 2022*. London: UCL: Research Department of Behavioural Science and Health Institute of Epidemiology & Health Care, University College London (2023). Available online at: <https://sbbresearch.org/wp-content/uploads/2023/03/Arts-and-population-health-FINAL-March-2023.pdf> (accessed March 15, 2024).
- Bone JK, Fancourt D. *Arts Culture & the Brain: a Literature Review and New Epidemiological Analyses*. London: Arts Council England (2022). Available online at: [https://www.artscouncil.org.uk/arts-culture-brain?relax=\\$09](https://www.artscouncil.org.uk/arts-culture-brain?relax=$09) (accessed March 15, 2024).
- Fancourt D, Warran K, Aughterson H. *Research and analysis - Evidence summary for policy: The role of arts in improving health & wellbeing*. GOV.UK: Department for Culture, Media and Sport & Department for Digital, Culture, Media & Sport (2020). Available online at: https://assets.publishing.service.gov.uk/media/5f9812268fa8f543f786b37f/DCMS_report_April_2020_final__1_.pdf (accessed March 15, 2024).
- Jespersen KV, Gebauer L, Vuust P. *White Paper: Music interventions in health care. Center for Music in the Brain, Aarhus University, Denmark - in collaboration with SoundFocus, Oticon Medical, WS Audiology and Danish Sound Cluster, Denmark*. Available online at: https://musicinthebrain.au.dk/fileadmin/MusicInTheBrain/MusicInTheBrain_Whitepaper_2022_web.pdf (accessed March 15, 2024).
- Bonde LO, Stensæth K, Ruud E. *Music and Health. A Comprehensive Model*. Department of Communication and Psychology, Aalborg University, Denmark (2023). Available online at: <https://prod-audxp-cms-001-app.azurewebsites.net/media/fwmnuadw/music-and-health-2023-final.pdf> (accessed March 15, 2024).
- Chapline J, Johnson JK. *The National Endowment for the Arts Guide to Community-Engaged Research in the Arts and Health*. National Endowment for the Arts. Office of Research & Analysis (2016). Available online at: <https://www.arts.gov/sites/default/files/Guide-to-Community-Engaged-Research-in-the-Arts-and-Health-March2017.pdf> (accessed March 15, 2024).
- Davies C, Pescud M. *The Arts and Creative Industries in Health Promotion. An Evidence Check rapid review*. Australia: Sax Institute for The Victorian Health Promotion Foundation (2020). Available online at: https://www.saxinstitute.org.au/wp-content/uploads/20.12_Evidence-Check_arts-and-health-promotion.pdf
- Fancourt D, Finn S. *What is the evidence on the role of the arts in improving health and well-being? A scoping review*. WHO Regional Office for Europe. Copenhagen (2019) (Health Evidence Network (HEN) synthesis report No. 67). Available online at: <https://www.who.int/europe/publications/i/item/9789289054553>
- Dow R, Warran K, Letrondo P, Fancourt D. The arts in public health policy: progress and opportunities. *Lancet Public Health*. (2023) 8:e155–60. doi: 10.1016/S2468-2667(22)00313-9
- Sajani N, Fietje N. The Jameel arts & health lab in collaboration with the WHO–lancet global series on the health benefits of the arts. *Lancet*. (2023) 402:1732–4. doi: 10.1016/S0140-6736(23)01959-1
- World Health Organization, Departmental news New York, United States of America. *Ground-breaking research series on health benefits of the arts*. (2023). Available online at: <https://www.who.int/news/item/25-09-2023-ground-breaking-research-series-on-health-benefits-of-the-arts> (accessed February 15, 2024).
- Culture Action Europe. *CultureForHealth*. (2021). Available online at: <https://cultureactioneurope.org/projects/cultureforhealth/> (accessed March 3, 2024).
- D'Amato M, Dowden S, Smith C. *Culture for Health and Well-being Compendium - A Guide for Practitioners*. CultureForHealth. Culture Action Europe (2023). Available online at: https://www.cultureforhealth.eu/app/uploads/2023/06/C4H_Compendium_V4_LP.pdf (accessed March 15, 2024).
- Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. (2018) 169:467–73. doi: 10.7326/M18-0850
- Peters MD, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: Scoping reviews. In: *JB I Manual for Evidence Synthesis*. JBI. (2020). Available online at: <https://jbi-global-wiki.refined.site/space/MANUAL/4687342/Chapter+11+%3A+Scoping+reviews> (accessed January 30, 2024).
- Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol*. (2018) 18:143. doi: 10.1186/s12874-018-0611-x
- Clift S, Phillips K, Pritchard S. The need for robust critique of research on social and health impacts of the arts. *Cult Trends*. (2021) 30:442–59. doi: 10.1080/09548963.2021.1910492
- Shiraam J, Comeau G, Swirp M. Identification and illustration of means to a critical assessment of music and health research literature. *Healthcare*. (2023) 11:807. doi: 10.3390/healthcare11060807
- Clift S, Grebosz-Haring K, Thun-Hohenstein L, Schuchter-Wiegand AK, Bathke A, Kaastra M. The need for robust critique of arts and health research: the treatment of the Gene Cohen et al. (2006) paper on singing wellbeing and health in subsequent evidence reviews. *Arts Health*. (2024) 22:1–19. doi: 10.1080/17533015.2023.2290075
- Clift S, Grebosz-Haring K, Thun-Hohenstein L, Schuchter-Wiegand AK, Bathke A. The need for robust critique of arts and health research: An examination of the Goldbeck and Ellerkamp (2012) randomised controlled trial of music therapy for anxiety in children, and its treatment in four systematic reviews. *Front Psychol*. (2022) 13:64. doi: 10.56883/ajmt.2024.64
- Grebosz-Haring K, Thun-Hohenstein L, Schuchter-Wiegand AK, Bathke AC, Clift S. The need for robust critique of arts and health research: dance-movement therapy, girls, and depression. *Ann N Y Acad Sci*. (2023) 1525:128–39. doi: 10.1111/nyas.15006
- McCrary JM, Redding E, Altenmüller E. Performing arts as a health resource? An umbrella review of the health impacts of music and dance participation. *PLOS ONE*. (2021) 16:e0252956. doi: 10.1371/journal.pone.0252956
- McCrary JM, Großbach M, Altenmüller E. Are culturally vibrant communities healthier? Relationships between performing arts activity and health outcomes in the 500 largest US cities. *Perspect Public Health*. (2022) 142:94–101. doi: 10.1177/17579139221081397
- Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol*. (2017) 87:4–13. doi: 10.1016/j.jclinepi.2017.05.006
- Young M, Smith MA. *Standards and Evaluation of Healthcare Quality, Safety, and Person-Centered Care*. StatPearls. Treasure Island (FL): StatPearls Publishing (2022). Available online at: <https://www.ncbi.nlm.nih.gov/books/NBK576432/> (accessed March 15, 2024).
- Erickson SM, Wolcott J, Corrigan JM, Aspden P. *Patient Safety: Achieving a New Standard for Care. 4: Health Care Data Standards In Institute of Medicine (US) Committee on Data Standards for Patient Safety*. Washington (DC): National Academies Press (US). (2004).
- Tabish SA. Editorial: Standards for better health. *Int J Health Sci*. (2009) 3:5–8.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2024.1414070/full#supplementary-material>

33. World Health Organization (WHO). *Quality of care*. Available online at: https://www.who.int/health-topics/quality-of-care#tab=protect&relax=Stab_1 (accessed March 15, 2024).
34. Barker TH, Stone JC, Sears K, Klugar M, Tufanaru C, Leonardi-Bee J, et al. The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. *JBI Evid Synth*. (2023) 21:494–506. doi: 10.11124/JBIES-22-00430
35. Coulton S, Clift S, Skingley A, Rodriguez J. Effectiveness and cost-effectiveness of community singing on mental health-related quality of life of older people: randomised controlled trial. *Br J Psychiatry*. (2015) 207:250–5. doi: 10.1192/bjp.bp.113.129908
36. Bonilha AG, Onofre F, Vieira ML, Prado MY, Martinez JA. Effects of singing classes on pulmonary function and quality of life of COPD patients. *Int J Chron Obstruct Pulmon Dis*. (2008) 4:1. doi: 10.2147/COPD.S4077
37. Fancourt D, Perkins R. Effect of singing interventions on symptoms of postnatal depression: three-arm randomised controlled trial. *Br J Psychiatry*. (2018) 212:119–21. doi: 10.1192/bjp.2017.29
38. Feng L, Romero-Garcia R, Suckling J, Tan J, Larbi A, Cheah I, et al. Effects of choral singing versus health education on cognitive decline and aging: a randomized controlled trial. *Aging*. (2020) 12:24798–816. doi: 10.18632/aging.202374
39. Ganzoni C, Arslani K, Pfister O, Freese M, Strobel W, Mueller C, et al. Choir singing improves respiratory muscle strength and quality of life in patients with structural heart disease – HeartChoir: a randomised clinical trial. *Swiss Med Wkly*. (2020) 150:w20346. doi: 10.4414/sm.w.2020.20346
40. Liu H, Song M, Zhai ZH, Shi RJ, Zhou XL. Group singing improves depression and life quality in patients with stable COPD: a randomized community-based trial in China. *Qual Life Res*. (2019) 28:725–35. doi: 10.1007/s11136-018-2063-5
41. Philip KE, Lewis A, Jeffery E, Buttery S, Cave P, Cristiano D, et al. Moving singing for lung health online in response to COVID-19: experience from a randomised controlled trial. *BMJ Open Respir Res*. (2020) 7:e000737. doi: 10.1136/bmjresp-2020-000737
42. Pongan E, Tillmann B, Leveque Y, Trombert B, Getenet JC, Auguste N, et al. Can musical or painting interventions improve chronic pain, mood, quality of life, and cognition in patients with mild Alzheimer's Disease? Evidence from a randomized controlled trial. *J Alzheimers Dis*. (2017) 60:663–77. doi: 10.3233/JAD-170410
43. Wulff V, Hepp P, Wolf OT, Fehm T, Schaal NK. The influence of maternal singing on well-being, postpartum depression and bonding – a randomised, controlled trial. *BMC Preg Childb*. (2021) 21:501. doi: 10.1186/s12884-021-03933-z
44. Cruz-Ferreira A, Marmeleira J, Formigo A, Gomes D, Fernandes J. Creative dance improves physical fitness and life satisfaction in older women. *Res Aging*. (2015) 37:837–55. doi: 10.1177/0164027514568103
45. Duncan RP, Earhart GM. Randomized controlled trial of community-based dancing to modify disease progression in Parkinson disease. *Neurorehabil Neural Repair*. (2012) 26:132–43. doi: 10.1177/1545968311421614
46. Lazarou I, Parastatidis T, Tsolaki A, Gkioka M, Karakostas A, Douka S, et al. International ballroom dancing against neurodegeneration: a randomized controlled trial in greek community-dwelling elders with mild cognitive impairment. *Am J Alzheimers Dis Dementiasr*. (2017) 32:489–99. doi: 10.1177/1533317517725813
47. Kaltsou A, Kouidi E, Fountoulakis K, Sipka C, Theochari V, Kandyli D, et al. Effects of exercise training with traditional dancing on functional capacity and quality of life in patients with schizophrenia: a randomized controlled study. *Clin Rehabil*. (2015) 29:882–91. doi: 10.1177/0269215514564085
48. Marquez DX, Wilson R, Aguiñaga S, Vásquez P, Fogg L, Yang Z, et al. Regular Latin dancing and health education may improve cognition of late middle-aged and older Latinos. *J Aging Phys Act*. (2017) 25:482–9. doi: 10.1123/japa.2016-0049
49. Caprilli S, Anastasi F, Grotto RPL, Abeti MS, Messeri A. Interactive music as a treatment for pain and stress in children during venipuncture: a randomized prospective study. *J Dev Behav Pediatr*. (2007) 28:399–403. doi: 10.1097/DBP.0b013e31811ff8a7
50. Huang R, Wang J, Wu D, Long H, Yang X, Liu H, et al. The effects of customised brainwave music on orofacial pain induced by orthodontic tooth movement. *Oral Dis*. (2016) 22:766–74. doi: 10.1111/odi.12542
51. Qin Y. Effect of music therapy intervention on physical functions and mental health of patients with ankylosing spondylitis. *Psychiatr Danub*. (2020) 32:403–10. doi: 10.24869/psyd.2020.403
52. Forouzandeh N, Drees F, Forouzandeh M, Darakhshandeh S. The effect of interactive games compared to painting on preoperative anxiety in Iranian children: a randomized clinical trial. *Compl Ther Clin Pract*. (2020) 40:101211. doi: 10.1016/j.ctcp.2020.101211
53. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. (2010) 340:c332–c332. doi: 10.1136/bmj.c332
54. Tack M, Tuller DM, Struthers C. Bias caused by reliance on patient-reported outcome measures in non-blinded randomized trials: an in-depth look at exercise therapy for chronic fatigue syndrome. *Fatigue Biomed Health Behav*. (2020) 8:181–92. doi: 10.1080/21641846.2020.1848262
55. Savović J, Jones H, Altman D, Harris R, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technol Assess*. (2012) 16:1–82. doi: 10.3310/hta16350
56. Ludbrook J. Should we use one-sided or two-sided *P* values in tests of significance? *Clin Exp Pharmacol Physiol*. (2013) 40:357–61. doi: 10.1111/1440-1681.12086
57. World Health Organization (WHO). *Urgent health challenges for the next decade*. (2020). Available online at: <https://www.who.int/news-room/photo-story/photo-story-detail/urgent-health-challenges-for-the-next-decade> (accessed March 15, 2024).
58. Govender P, Medvedyuk S, Raphael D. 1845 or 2023? Friedrich Engels's insights into the health effects of Victorian-era and contemporary Canadian capitalism. *Social Health Illn*. (2023) 45:1609–33. doi: 10.1111/1467-9566.13676
59. Institute of Health Equity. Available online at: <https://www.instituteoftheequity.org/home> (accessed March 15, 2024).
60. Lord VM, Cave P, Hume VJ, Flude EJ, Evans A, Kelly JL, et al. Singing teaching as a therapy for chronic respiratory disease – a randomised controlled trial and qualitative evaluation. *BMC Pulm Med*. (2010) 10:41. doi: 10.1186/1471-2466-10-41
61. Lord VM, Hume VJ, Kelly JL, Cave P, Silver J, Waldman M, et al. Singing classes for chronic obstructive pulmonary disease: a randomized controlled trial. *BMC Pulm Med*. (2012) 12:69. doi: 10.1186/1471-2466-12-69
62. Kaasgaard M, Rasmussen DB, Andreasson KH, Hilberg O, Løkke A, Vuust P, et al. Use of Singing for Lung Health as an alternative training modality within pulmonary rehabilitation for COPD: a randomised controlled trial. *Eur Respir J*. (2022) 59:2101142. doi: 10.1183/13993003.01142-2021
63. Kaasgaard M, Rasmussen DB, Løkke A, Vuust P, Hilberg O, Bodtger U. Physiological changes related to 10 weeks of singing for lung health in patients with COPD. *BMJ Open Respir Res*. (2022) 9:e001206. doi: 10.1136/bmjresp-2022-001206
64. Lewis A, Philip KEJ, Lound A, Cave P, Russell J, Hopkinson NS. The physiology of singing and implications for 'Singing for Lung Health' as a therapy for individuals with chronic obstructive pulmonary disease. *BMJ Open Respir Res*. (2021) 8:e000996. doi: 10.1136/bmjresp-2021-000996
65. Moore A, Fisher E, Eccleston C. Flawed, futile, and fabricated—features that limit confidence in clinical research in pain and anaesthesia: a narrative review. *Br J Anaesth*. (2023) 130:287–95. doi: 10.1016/j.bja.2022.09.030
66. Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR. *Taxonomy of educational objectives: The classification of educational goals. Book 1, Cognitive domain*. New York, NY: Longmans, Green and CO LTD, London (1956). Available online at: https://eclass.uoa.gr/modules/document/file.php/PPP242/Benjamin%20S.%20Bloom%20-%20Taxonomy%20of%20Educational%20Objectives%2C%20Handbook%201_%20Cognitive%20Domain-Addison%20Wesley%20Publishing%20Company%20%281956%29.pdf (accessed March 15, 2024).
67. Meneses-Echavez JF, Bidonde J, Yepes-Nuñez JJ, Poklepović Peričić T, Puljak L, Bala MM, et al. Evidence to decision frameworks enabled structured and explicit development of healthcare recommendations. *J Clin Epidemiol*. (2022) 150:51–62. doi: 10.1016/j.jclinepi.2022.06.004
68. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. *Cochrane Handbook for Systematic Reviews of Interventions version 6.4*. Cochrane (2023). Available online at: www.training.cochrane.org/handbook (accessed March 15, 2024).
69. Hagemann V. *Benefits of Collective Singing: An International Systematic Review on "Choral Research" - Curated Research* | #BenefitsOfSinging Toolkit. (2021). Available online at: [https://docs.google.com/spreadsheets/d/17GGqRDercBznwugi_hHilpCrcnk8ctCZZ0jv9rOosE/edit?pli=protect&relax=\\$1#gid=protect&relax=\\$0](https://docs.google.com/spreadsheets/d/17GGqRDercBznwugi_hHilpCrcnk8ctCZZ0jv9rOosE/edit?pli=protect&relax=$1#gid=protect&relax=$0) (accessed December 2, 2024).
70. Schünemann HJ. Interpreting GRADE's levels of certainty or quality of the evidence: GRADE for statisticians, considering review information size or less emphasis on imprecision? *J Clin Epidemiol*. (2016) 75:6–15. doi: 10.1016/j.jclinepi.2016.03.018
71. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ*. (2021) 374:n2061. doi: 10.1136/bmj.n2061
72. Crocker H, Peters M, Foster C, Black N, Fitzpatrick R. A core outcome set for randomised controlled trials of physical activity interventions: development and challenges. *BMC Public Health*. (2022) 22:389. doi: 10.1186/s12889-022-12600-7
73. Kirkham JJ, Williamson P. Core outcome sets in medical research. *BMJ Med*. (2022) 1:e000284. doi: 10.1136/bmjmed-2022-000284
74. Klaim M, Kapp S, Hudson P, Chapman W, Denehy L, Story D, et al. Implementability of healthcare interventions: an overview of reviews and development of a conceptual framework. *Implement Sci*. (2022) 17:10. doi: 10.1186/s13012-021-01171-7
75. Hillier S, Grimmer-Somers K, Merlin T, Middleton P, Salisbury J, Tooher R, et al. FORM: an Australian method for formulating and grading recommendations in evidence-based clinical guidelines. *BMC Med Res Methodol*. (2011) 11:23. doi: 10.1186/1471-2288-11-23
76. Clift S, Bathke A, Daffern H, Davies C, Grebosz-Haring K, Kaasgaard M, et al. WHO–Lancet Global Series on health and the arts. *Lancet*. (2024) 403:1335. doi: 10.1016/S0140-6736(24)00246-0