AUREUM ACADEMIC JOURNAL
VOL. 3, NO. 1, March 2026
https://doi.org/10.48161/aaj.v3n1a95

The Use of Visual Arts in Healthcare Settings: Improving Patient Experience and Emotional Recovery

Kumaresu Murugasu[1*], Denesh Sooriamoorthy[2], Abishov Nurzhan[3], Ahmad Pitra Jumli[4]

[1] School of Global Hospitality & Tourism, Asia Pacific University of Technology & Innovation (APU), Kuala Lumpur, Malaysia;
[2] School of Engineering, Asia Pacific University of Technology and Innovation, Kuala Lumpur, 57000, Malaysia;
[3] Tourism, International University of Tourism and Hospitality, Turkestan, Kazakhstan.
[4] School of Hospitality & Creative Arts, Management and Science University, Shah Alam, Malaysia.

* **Corresponding author:** kumaresu@apu.edu.my.

ABSTRACT: Hospitals increasingly use visual arts to support person-centered care, yet controlled evidence remains limited for programs that combine environmental artwork with facilitated art-making. This study estimated the program-level effectiveness of a combined visual-arts-enhanced care service and examined its implementation and explanatory processes. A multicentre pragmatic cluster-randomized trial with an embedded longitudinal mixed-methods process evaluation allocated 24 adult inpatient wards in six hospitals to culturally responsive environmental artwork plus up to four facilitated art-making sessions or to an attention- and time-matched non-art wellbeing program, both alongside standard care. Among 681 patients screened, 492 enrolled. We measured patient experience and reductions in psychological distress as co-primary outcomes at baseline, mid-intervention, discharge (the primary endpoint), and six-week follow-up. We integrated small-sample-corrected multilevel intention-to-treat models, multiple imputation, and multiplicity control with clinical records, implementation data, interviews, focus groups, observations, and reflective responses. At discharge, the intervention group had a 4.8-point higher person-centered care score (95% CI 1.8–7.8; g=.26) and a 1.7-point lower Hospital Anxiety and Depression Scale total score (95% CI −2.9 to −0.5; g=−.22). Follow-up estimates were smaller and included no effect. Positive affect and care satisfaction showed the clearest secondary differences, whereas pain, quality of life, opioid exposure, length of stay, and readmission did not differ reliably. Exploratory process estimates were associated with perceived control, emotion regulation, meaning-making, and social connection but did not establish causal mechanisms. Qualitative findings identified personhood, temporary distance from distress, choice, and connection alongside fatigue, cultural mismatch, and performance pressure. The combined program was associated with modest short-term experiential and psychological advantages during admission, contingent on voluntary participation, cultural relevance, facilitation quality, and coordination with ward routines.

Keywords: visual arts; patient experience; psychological distress; arts in health; person-centered care.

I. INTRODUCTION

Hospital care is experienced through its physical, interpersonal, and clinical environments. As a result, healthcare settings have incorporated visual art through commissioned works, photographic and digital installations, and facilitated drawing, painting, collage, and printmaking. Reviews indicate that patients and staff often value visual art when it is accessible and relevant, but the evidence remains heterogeneous in intervention content, comparator quality, outcome selection, and follow-up [1].

Environmental artwork and participatory art-making represent related but distinct exposures. Environmental art can provide low-demand attentional focus, orientation, and conversational material, whereas participatory activity adds authorship, choice, and embodied engagement. Patients actively interpret, approach, avoid, and discuss hospital art rather than receiving it passively [2]. Evidence from pediatric and adult settings also shows that responses vary by placement, content, illness, age, sensory tolerance, and opportunities for interaction [3], [4]. These findings caution against assuming that any artwork is therapeutic merely because it is displayed in a clinical setting.

Participatory art-making may offer non-verbal expression, structured attention, and a temporary sense of competence or control. Reviews of adult art-therapy and arts-based interventions report possible benefits for anxiety, depression, coping, and quality of life, but effect estimates vary substantially across populations, facilitators, treatment intensity, and comparators [5]. Economic review has likewise found insufficient evidence that short-term experiential benefits translate into service savings [8]. Controlled designs using an attention-matched comparator are therefore important because supportive contact, structured time, and expectancy can themselves influence patient-reported outcomes.

Mechanistic evidence suggests several plausible, overlapping processes. Brief art-making has been associated with reduced cortisol without requiring prior artistic experience [9]. The Bodymind Model emphasizes sensory-motor engagement, symbolic transformation, emotional activation, and reflective integration [10], while healthcare-design research links environmental interventions with stress regulation, perceived control, and social interaction [11]. These perspectives support examination of aesthetic engagement, positive distraction, perceived control, emotion regulation, meaning-making, and social connectedness, while recognizing that measured mediators remain non-randomized. Researchers usually evaluate environmental and participatory visual arts separately; they emphasize immediate mood and satisfaction more often than sustained effects; usual-care comparisons do not isolate art-specific processes from attention and structured activity; and effectiveness studies rarely integrate implementation quality, adverse experiences, cultural responsiveness, clinical records, and stakeholder explanations.

The study addressed these limitations through a multicentre, pragmatic, cluster-randomized effectiveness trial with a longitudinal, mixed-methods process evaluation. We randomized wards because environmental artwork altered shared spaces and could not be confined to individually allocated patients. We deliberately delivered environmental access and optional facilitated making as a single service model that could accommodate different levels of patient energy and interest. The active control matched attention and contact time without visual-art content. The central contribution was a program-level estimate of the combined service beyond supportive contact, accompanied by a careful account of implementation, process variation, heterogeneity, and adverse experience. The design did not estimate the independent contribution of either arts component.

II. RESEARCH PROBLEM, AIM, OBJECTIVES, AND QUESTIONS

1. RESEARCH PROBLEM

Hospitals increasingly use visual-arts programs, but decision makers lack multicentre, active-controlled evidence that estimates program effects while accounting for ward-level delivery, implementation fidelity, cultural fit, adverse experience, and retained outcomes. Existing studies frequently use self-selected samples, weak comparators, or single-site designs, making it difficult to distinguish art-specific effects from supportive contact and clinical recovery.

The research problem was therefore the absence of a cluster-randomized, implementation-sensitive evaluation of a combined environmental and participatory visual-arts service. The study estimated the combined program's effect and did not attempt to attribute outcomes to either component independently.

2. RESEARCH AIM

The study aimed to estimate the program-level effects of the combined visual-arts-enhanced care service on patient experience and psychological-distress reduction relative to an attention- and time-matched non-

art wellbeing program, and to examine implementation conditions and exploratory process associations linked with variation in response.

- To estimate program-level effects on patient experience, psychological-distress reduction as an indicator of emotional recovery, and prespecified psychological, clinical, and healthcare-service outcomes from baseline to discharge and six-week follow-up.
- To explore a restricted set of temporally ordered process associations—positive distraction, perceived control, emotion regulation, meaning-making, and social connectedness—and assess whether baseline distress and cultural congruence were associated with variation in the co-primary outcomes; these analyses were not powered for confirmatory mechanism or interaction testing.
- To evaluate reach, dose, fidelity, acceptability, feasibility, culturally responsive adaptation, contamination, and adverse experiences across hospitals and wards.
- To integrate patient-reported, clinical, observational, reflective, and stakeholder evidence to explain convergence, divergence, negative cases, and the practical conditions under which visual arts supported or constrained care.

3. RESEARCH QUESTIONS

- RQ1. What program-level effects occurred in patient experience and psychological-distress reduction, as indicators of emotional recovery, at discharge and at the six-week follow-up relative to the active control condition?
- RQ2. What between-group differences occurred in positive affect, stress, pain, sleep, recovery self-efficacy, quality of life, satisfaction, medication use, length of stay, readmission, and adverse experiences?
- RQ3. What exploratory associations linked prespecified processes and participant characteristics with stronger, weaker, or more sustained outcomes?
- RQ4. How did patients, families, healthcare professionals, and artist-facilitators describe implementation, benefit, burden, cultural fit, and boundary conditions, and how did these accounts converge with or challenge the quantitative findings?

III. LITERATURE REVIEW, RESEARCH GAP, AND ORIGINALITY

1. VISUAL ENVIRONMENTS AND THE EXPERIENCE OF HOSPITAL CARE

Supportive-design research proposes that healthcare environments are more restorative when they reduce avoidable stressors, support social connection, and increase perceived control. In hospital rooms, artwork seems most relevant when it is visible, comprehensible, and available for voluntary engagement, rather than placed primarily for circulation or staff benefit [12]. Stress-recovery theory offers a foundational account of rapid affective responses to visual environments [13], but applying it to inpatient care requires attention to illness, culture, sensory tolerance, and clinical context.

Environmental art is better treated as an opportunity for attention, recognition, or conversation than as a standardized treatment dose. Placement, glare, line of sight, content, and the right to disengage determine whether it functions as a resource. Cultural responsiveness requires consultation about representation, symbolism, language, religious sensitivity, and the risk of reducing identity to decorative motifs. Adverse responses—including overstimulation, unwanted associations, and perceived stereotyping—are relevant implementation outcomes rather than exceptional cases.

2. ART-MAKING, EMOTIONAL REGULATION, AND RECOVERY

Facilitated art-making changes the patient's role from observer to agent through selection, mark-making, pacing, and decisions about privacy or sharing. These activities may interrupt threat-focused attention and preserve a sense of capability, but they can also evoke fatigue, perfectionism, painful memory, or unwanted exposure. Supportive art-making therefore requires clear non-diagnostic boundaries, voluntary participation, adaptable materials, and safeguarding pathways.

Emotion-regulation theory distinguishes attentional deployment, appraisal, and response modulation [14]. In hospital care, visual art may support attentional redirection, symbolic reappraisal, and paced

sensory-motor activity. Meaning-making and social connectedness broaden the model beyond symptom reduction by linking illness with biography, values, and relationships. These constructs informed prespecified exploratory process analyses, while additional boundary conditions were retained when they emerged qualitatively.

3. PERSON-CENTRED CARE AND CULTURAL RESPONSIVENESS

Person-centered care requires an enabling culture, communication, access, integration, and meaningful engagement [15]. Healthcare staff describe it as an organizational practice rather than an isolated bedside technique [16]. Visual arts may contribute when patients can choose whether and how to engage, facilitators adapt activities to physical and cognitive capacity, and patients and the community select artwork.

The available evidence indicates that benefit depends less on visual intensity than on fit, choice, privacy, and the protection of refusal. Programs can reproduce institutional control when participation is framed as expected, content is selected without patient input, or completed work is displayed without explicit permission. These risks connect cultural responsiveness directly with intervention governance.

4. CRITICAL SYNTHESIS AND RESEARCH GAP

Overall, visual environments and participatory art-making may influence care experience, affect, and coping, but limited evidence, small samples, single-site designs, inconsistent comparators, short follow-up, incomplete fidelity reporting, and underreporting of adverse or indifferent responses limit conclusions. Researchers often use patient-experience measures interchangeably with satisfaction, and they rarely report program costs. Mechanisms are frequently proposed post hoc rather than measured in temporal order.

The study's central originality was an active-controlled, cluster-randomized evaluation of a combined visual-arts service beyond matched supportive attention. We used process evaluation to interpret delivery, cultural fit, burden, and heterogeneity; we did not present it as a separate source of causal proof. We reported experiential, psychological, clinical-utilization, implementation, and resource outcomes separately to prevent overgeneralization across outcome domains.

IV. VISUAL ARTS–EMOTION–RECOVERY FRAMEWORK

The Visual Arts–Emotion–Recovery (VAER) Framework integrated Supportive Design and Stress-Recovery Theory, Emotion-Regulation Theory, and Person-Centred Care. The framework distinguished the combined program from proximal engagement, agency and relational processes, psychological regulation, and two separate co-primary outcomes. The trial treated environmental artwork and facilitated art-making as coordinated components whose independent effects could not be identified.

Prespecified process constructs included aesthetic engagement, positive distraction, perceived control, emotion regulation, meaning-making, and social connectedness. We measured early processes during exposure and later processes before follow-up outcomes. Qualitative themes on recognition, timing, fatigue, and performance pressure helped refine boundary conditions rather than retroactively redefine the statistical pathways.

We treated baseline distress, service line, illness severity, cultural congruence, prior arts engagement, admission duration, ward workflow, pre-existing artwork, and facilitation quality as contextual conditions. Process and moderator analyses were exploratory because 24 wards were insufficient for confirmatory multilevel mediation or extensive interaction testing.

FIGURE 1 presents a simplified analytical structure. Solid arrows indicate prespecified program-to-process relations, while dashed arrows indicate exploratory process-to-outcome associations. The contextual band represents conditions that could influence exposure, interpretation, and response. The diagram is explanatory rather than causal and allows direct effects, partial explanation, non-response, and adverse experience.

https://doi.org/10.48161/aaj.v3n1a95

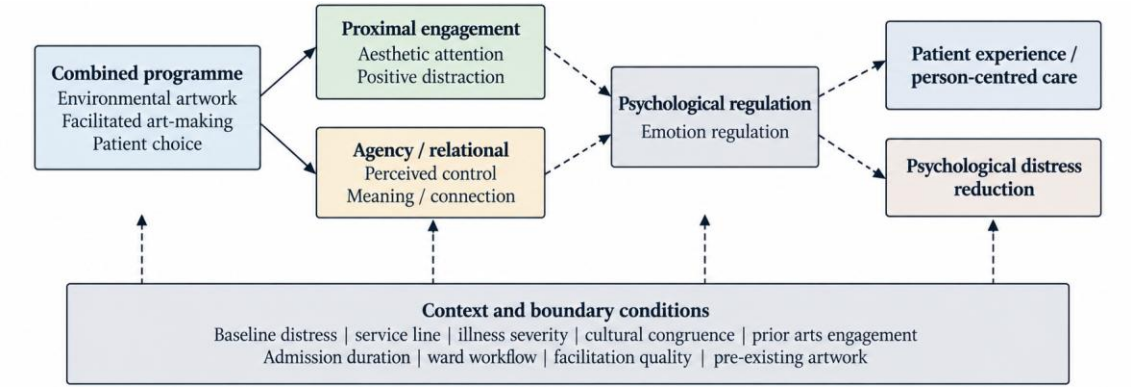


FIGURE 1. Simplified Visual Arts–Emotion–Recovery Framework. Solid arrows identify prespecified program-to-process relations; dashed arrows identify exploratory process-to-outcome associations. No arrow establishes a unique causal mechanism.

V. METHODOLOGY

1. DESIGN AND REPORTING LOGIC

The study used a two-arm, multicentre, pragmatic cluster-randomized controlled effectiveness design with an embedded longitudinal explanatory mixed-methods process evaluation. We used the inpatient ward as the unit of allocation because environmental artwork altered shared spaces, and individual randomization would have caused substantial contamination. Quantitative effectiveness analysis was primary; qualitative and implementation strands explained delivery, experience, adaptation, and boundary conditions. We organized contextual determinants using the updated Consolidated Framework for Implementation Research [17], and we evaluated the program as a complex intervention under Medical Research Council guidance [18]. The protocol (version 2.1, 15 September 2021), intervention manual (version 1.3, 20 September 2021), and statistical analysis plan (version 1.0, 30 September 2021) were version-controlled before ward allocation and are summarized in the separate Supplementary Material.

We conducted fieldwork in six hospitals within one national health system. Each hospital contributed four adult inpatient wards, producing a balanced network of oncology, cardiopulmonary or general medical, and orthopedic or neurological rehabilitation services. Hospitals differed in size, urbanity, staffing stability, pre-existing artwork, and prior arts programming. We randomized 24 wards, 12 per condition. Eight artist-facilitators and six wellbeing facilitators delivered the allocated programs; each completed condition-specific training, safeguarding induction, and weekly supervision. Each ward implemented its condition for 14 weeks, and participants were followed up for six weeks after discharge.

2. PARTICIPANTS, ELIGIBILITY, AND RECRUITMENT

Eligible patients were aged 18 years or older, admitted to a participating ward, medically stable at screening, expected to remain in hospital for at least five days, able to provide informed consent, and able to complete study procedures in one of three supported languages with interpreter assistance where required. Exclusion criteria were delirium or severe cognitive impairment, acute behavioral risk, infection-control restrictions that prevented safe participation, rapidly deteriorating medical status, or concurrent enrolment in an intensive creative-arts therapy program. Prior artistic experience was not required.

Research nurses screened admission lists and approached potentially eligible patients within 48 hours of admission. Recruitment occurred after ward allocation; nurses therefore knew the condition, although they used a standardized eligibility and invitation script, logged every approach, and recorded reasons for non-participation. Across wards, the median enrolment rate among eligible patients was 84% (IQR 79%–89%; range 72%–94%). Participation was voluntary and independent of clinical care. Patients who declined art-making remained eligible for outcome follow-up because the estimand concerned ward-level assignment

rather than receipt of a minimum artistic dose. Process-evaluation sampling sought variation in hospital, service line, exposure, cultural congruence, and quantitative response profile.

3. RANDOMIZATION, ALLOCATION, BLINDING, AND CONTAMINATION CONTROL

Randomization followed ward recruitment and baseline environmental audit. An independent statistician prepared the allocation schedule using permuted blocks within hospital and service-line strata. Allocation was released to implementation teams only after baseline ward preparation. Because the intervention altered visible ward environments and involved facilitated sessions, patients, ward staff, and facilitators necessarily knew the assigned condition. Questionnaire assessors who were not involved in delivery, clinical-record abstractors, and the primary statistician were masked to allocation labels until the primary models were finalized. In a masking audit, assessors correctly inferred allocation for 58% of sampled assessments, indicating incomplete masking that was only slightly above chance. Intervention materials were stored on allocated wards, artist-facilitators did not work in control wards, and control staff were asked not to introduce new organized arts activities. We documented existing artwork and patient-initiated hobbies rather than prohibiting them.

Reporting followed the CONSORT extension for cluster-randomized trials, including the timing of cluster and individual recruitment, ward-level allocation, reasons for non-participation, cluster retention, individual follow-up, intraclass correlation, and the distinction between observed and imputed samples [19]. All 24 wards remained in the study, although staffing disruptions and temporary renovation affected delivery on four wards.

4. INTERVENTION AND ACTIVE CONTROL

The visual-arts-enhanced care program combined environmental and participatory components. After review for cultural relevance, accessibility, religious sensitivity, glare, infection control, and representational balance, staff installed patient-visible artworks in corridors, communal areas, treatment waiting spaces, and selected rooms. Patients could ask staff to cover, reposition, or replace them. The combined design reflected routine service delivery; accordingly, every effectiveness estimate refers to assignment to the complete service and cannot be attributed separately to environmental artwork, facilitated art-making, or their interaction.

The participatory component offered up to four 35–45 minute individual or small-group sessions over seven to ten days. Eight professional artist-facilitators received a 12-hour healthcare induction, six hours of safeguarding and infection-control training, and weekly 60-minute supervision. A manualized sequence covered invitation and choice, low-risk sensory entry, self-directed making, optional reflection, and decisions about privacy, display, storage, or disposal. The study prohibited diagnostic interpretation, compulsory disclosure, public display without consent, and therapeutic claims beyond the facilitator role. We adapted materials, posture, duration, and prompts for pain, fatigue, mobility, dexterity, and preference.

The active control matched interpersonal attention and approximate duration. Six wellbeing facilitators offered structured conversation, neutral reading, puzzles, brief non-art reminiscence, and recovery goal-setting. They received eight hours of training in matched contact, condition separation, and escalation procedures. The intervention prohibited visual art-making, guided image interpretation, study-provided art materials, and discussion of the intervention collection. Standard medical, nursing, rehabilitation, psychosocial, and pharmacological care continued in both conditions.

Table 1. TIDieR-aligned description of the intervention and active control.

Domain	Visual-arts-enhanced care	Active non-art wellbeing control	Fidelity indicator
Ward environment	Curated, culturally responsive artworks in patient-visible locations; optional multilingual captions; replacement or covering pathway	No new study-provided artwork; pre-existing environment retained and audited	Installation, visibility, glare, access, and response-time audit
Contact dose	Up to four 35–45 min sessions over 7–10 days	Up to four 35–45 min contacts over 7–10 days	Session number, duration, timing, and interruptions
Core activity	Drawing, watercolor, collage, printmaking, and low-mess mixed media	Conversation, neutral reading, puzzles, reminiscence, and goal-setting	Activity checklist and facilitator log
Provider / training	Eight professional artist-facilitators; 12-h healthcare induction; 6-h safeguarding/infection-control training; weekly 60-min supervision	Six wellbeing facilitators; 8-h training in matched contact, condition separation, and escalation; weekly supervision	Training completion; supervision attendance; provider-level deviations; ≥80% adherence threshold
Agency safeguards	Choice of medium, prompt, pace, privacy, display, storage, disposal, and refusal	Choice of non-art activity and discussion topic; refusal respected	Observed choice opportunity and refusal documentation
Tailoring	Adaptation for fatigue, pain, mobility, dexterity, infection control, language, and cultural preference	Adaptation of contact length and non-art activity for clinical capacity	Adaptation type and preservation of core functions
Facilitation boundary	Supportive art-making; no diagnostic image interpretation	Supportive contact; no visual-art discussion or making	Supervision review and deviation log
Clinical integration	Scheduling around care; escalation pathway for distress and risk	Scheduling around care; standard escalation pathway	Coordination record, missed-session reason, adverse-event review
Concomitant care	Standard medical, nursing, rehabilitation, psychosocial, and pharmacological care	Same standard care	EHR and co-intervention record
Prohibited elements	Diagnostic interpretation, compulsory disclosure, public display without permission, or claims of clinical treatment	Visual-art discussion, art-making, guided imagery, or study-provided art materials	Deviation log; immediate retraining for prohibited-content exposure

All participants continued to receive standard clinical care. Component adaptations were permitted when they preserved patient choice, non-diagnostic facilitation, and condition separation.

5. MEASUREMENT SCHEDULE AND VARIABLES

Assessments occurred at baseline within 48 hours of admission (T0), mid-intervention on day 4 or 5 (T1), at discharge or on day 10 if admission continued (T2), and six weeks after discharge (T3). Discharge was the primary endpoint. We extracted electronic clinical outcomes through 30 days after discharge. We

administered patient-reported measures by tablet, paper, or trained interviewer, depending on physical ability and preference; we recorded the mode and interpreter use for sensitivity analysis.

Psychological distress reduction was used as a limited indicator of emotional recovery and was measured with the Hospital Anxiety and Depression Scale total score. HADS was treated as a non-diagnostic measure of anxiety and depressive distress, not as a comprehensive measure of emotional, social, or functional recovery [20]. We measured patient experience/person-centered care with the Person-centered Climate Questionnaire–Patient version, transformed to a 0–100 score; it assesses safety, everydayness, and hospitality as experienced by patients [21]. The two outcomes were co-primary because a patient could experience a more person-centered care environment without a large reduction in distress, or report reduced distress without rating the broader care climate more positively.

Secondary patient-reported outcomes were PROMIS Positive Affect, the four-item Perceived Stress Scale, 0–10 pain intensity and interference ratings, PROMIS Sleep Disturbance, PROMIS Self-Efficacy for Managing Symptoms, EQ-5D-5L index and visual analog scale, and a global 0–100 care-satisfaction rating. We selected PROMIS measures for standardized item development [22], including established sleep short forms [23], and used the EQ-5D-5L as a concise, generic health-status measure [24]. The satisfaction rating supplemented, but did not replace, the validated PCQ-P and was interpreted as a limited global indicator. Clinical outcomes were daily oral morphine-equivalent dose among analgesic users, length of stay, and unplanned 30-day readmission. Illness severity was a prespecified, service-standardized composite of the National Early Warning Score 2, the Charlson Comorbidity Index, and the recorded care-dependency level; higher scores indicated greater severity. Adverse experiences included physical discomfort, emotional escalation, conflict about artwork, infection-control incidents, and events judged related to program participation.

Mechanism measures were brief study-specific scales derived from the VAER Framework. Expert review and cognitive interviewing preceded field use. The six-factor ordinal model showed acceptable fit (CFI=.943, TLI=.934, RMSEA=.049, SRMR=.047). Omega coefficients ranged from .77 to .89 at T1 and from .79 to .90 at T2; coefficients within each questionnaire language remained ≥.74. Partial scalar invariance across language versions and longitudinal metric invariance across T1–T2 were retained after freeing two item thresholds. These results support internal use within this dataset but do not constitute independent validation, criterion validity, or broad cross-cultural generalizability. The scales were therefore used only in exploratory process analyses and require replication in independent samples. Baseline distress and cultural congruence were the two primary moderators; remaining interactions were supplementary and exploratory. Implementation outcomes included reach, dose, fidelity, adaptation, acceptability, feasibility, contamination, resource use, and adverse experience; we interpreted them using RE-AIM concepts [25].

We recorded resource use prospectively from the provider perspective. Micro-costing included artwork licensing or commissioning, mounting, materials, facilitator and supervision time, staff coordination, training, replacement requests, and infection-control adaptations.

Table 2. Outcomes, instruments, scoring, and measurement schedule.

Construct	Measure / source	Scale and direction	T0	T1	T2	T3
Patient experience (co-primary)	PCQ-P	0–100; higher is better	●	●	●	●
Psychological distress reduction (co-primary; emotional-recovery indicator)	HADS total	0–42; lower better	●	●	●	●
Positive affect	PROMIS short form	T-score; higher better	●	●	●	●
Perceived stress	PSS-4	0–16; lower better	●	●	●	●
Pain	NRS intensity/interference	0–10; lower better	●	●	●	●

Construct	Measure / source	Scale and direction	T0	T1	T2	T3
Sleep disturbance	PROMIS short form	T-score; lower is better	●		●	●
Recovery self-efficacy	PROMIS short form	T-score; higher is better	●		●	●
Quality of life	EQ-5D-5L / EQ-VAS	Index and 0–100; higher better	●		●	●
Satisfaction	Global care-satisfaction rating (supplementary)	0–100; higher better			●	
Clinical utilization	EHR: OME, LOS, readmission	Dose, days, binary	●		●	30 d
Early mechanisms	Engagement, distraction, control	Standardized scores		●	●	
Later mechanisms	Regulation, meaning, connection	Standardized scores			●	
Implementation	Logs, audits, observations, interviews	Reach, dose, fidelity, acceptability	Cont.	Cont.	Cont.	●

T0: baseline; T1: day 4–5; T2: discharge or day 10; T3: six-week follow-up; OME: oral morphine equivalent; LOS: length of stay.

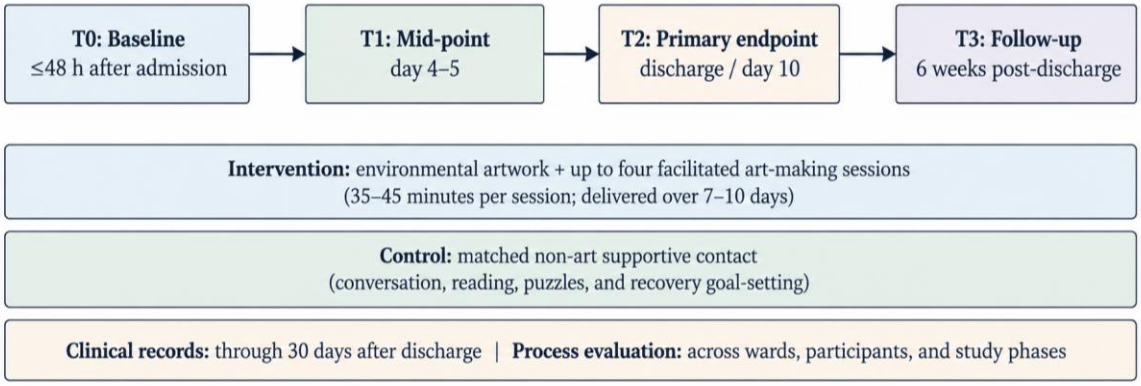


FIGURE 2. Intervention and measurement timeline. We extracted electronic clinical outcomes through 30 days after discharge.

6. *SAMPLE-SIZE DETERMINATION*

The target sample was 480 patients across 24 wards. Computational power calculations assumed a mean cluster size of 20, ward intraclass correlation of .03, correlation of .55 among repeated observations, 18% loss to six-week follow-up, two-sided familywise alpha of .05 divided across the two co-primary discharge outcomes, and moderate variation in cluster size. The design provided approximately 80% power to detect a standardized intervention-by-time difference of d=.30 for either co-primary outcome. Calculations accounted for cluster-size variation, covariate adjustment, and attrition in accordance with cluster-trial guidance [26]. The enrolled sample of 492 exceeded the individual target but did not increase the number of independent clusters; interaction and mediation analyses therefore remained less precise than the primary analysis.

7. QUANTITATIVE ANALYSIS

Analyses followed the intention-to-treat principle. For continuous outcome Y_{ijt} for participant i in ward j at occasion t, the principal model was $Y_{ijt} = \beta0 + \beta1(\text{condition}_j) + \beta2(\text{time}_t) + \beta3(\text{condition}_j \times \text{time}_t) + \beta4(\text{baseline outcome}_{ij}) + \beta5X_{ij} + u_j + v_{ij} + \varepsilon_{ijt}$, where X represented prespecified covariates, u_j was the ward random intercept, and v_{ij} the participant random intercept. A heterogeneous first-order autoregressive residual structure represented within-participant dependence. Kenward–Roger degrees of freedom were used because only 24 wards were randomized. We used generalized mixed models for binary outcomes, a log-linked model with robust standard errors for length of stay, and a two-part model for opioid exposure. Estimates are reported with standard errors, degrees of freedom, 95% confidence intervals, Hedges g, ward ICCs, and observed denominators.

We addressed missing data with 50 multilevel chained-equation imputations, including ward, allocation, outcomes, baseline covariates, clinical variables, contact dose, transfer status, assessment mode, and retention indicators [27]. We reviewed trace plots, distributional checks, Monte Carlo error, and the fraction of missing information. Each co-primary discharge outcome used a two-sided alpha of .025. False-discovery-rate adjustment controlled multiplicity across secondary patient-reported outcomes [28]; clinical-utilization outcomes formed a separate descriptive family. A 5-point PCQ-P difference was prespecified as a pragmatic group-level threshold (approximately 0.3 baseline SD), while a 3-point HADS reduction was used only for responder description, not as a validated universal minimal important difference.

Because the number of wards constrained multilevel process modeling, four temporally ordered pathways were examined as exploratory conditional associations. Monte Carlo confidence intervals were pooled across imputed datasets, and no mediator was interpreted as uniquely causal [29]. Full pathway estimates, alternative specifications, and sensitivity analyses are reported in the Supplementary Material.

The two primary moderator analyses examined baseline distress and cultural congruence. We also examined service line, age, gender, previous arts engagement, illness severity, and planned admission duration. Sensitivity analyses included complete-case, no-imputation models; ward fixed effects; adjustment for contact time; per-protocol contrasts; exclusion of contaminated controls; inverse-probability retention weighting; survey-mode adjustment; and leave-one-ward-out analysis. Convergence, residual distributions, influence, collinearity, variance components, and covariance-structure fit were checked before interpretation; full diagnostics are supplied separately.

8. QUALITATIVE AND IMPLEMENTATION PROCESS EVALUATION

The qualitative corpus comprised interviews with 42 patients (34 intervention and eight control), 12 family members, and six artist-facilitators; four focus groups involved 18 healthcare professionals. Patient interviews were distributed across all six hospitals (six to eight per site), all three service lines (14 per service line), and contrasting exposure and response profiles. Interviews lasted a median of 39 minutes (IQR 31–51), family interviews 34 minutes (IQR 27–45), facilitator interviews 52 minutes (IQR 44–61), and focus groups 63 minutes (range 55–74). All interviews, focus groups, and quotations used in the analysis were conducted in English; no translated quotations were included. Recruitment continued until the corpus provided sufficient information power across sites, roles, exposure, cultural-congruence profiles, and negative cases; the final eight interviews refined theme boundaries without adding a new explanatory theme.

Analysis used the Framework Method [30]. A health-psychology researcher led coding, an arts-in-health researcher examined artistic and facilitation processes, and a qualitative methodologist reviewed cultural interpretation and disconfirming evidence. The analysts independently coded 12 contrasting transcripts, developed a provisional matrix combining VAER categories with inductive codes, and refined definitions through reflexive discussion. Memos documented prior commitments to arts-in-health, possible confirmation bias, clinical assumptions, and changes prompted by negative cases. A patient-advisory reviewer challenged overly positive interpretations. Each final theme was supported across at least three hospitals and more than one stakeholder group; the published quotations were selected from five hospitals to avoid site dominance. The Supplementary Material includes the participant matrix, codebook, quotation audit, and negative-case log.

Mixed-methods integration used common identifiers, connected sampling, interim case summaries, research-question joint displays, and explicit examination of convergence, complementarity, expansion, and disagreement. Integration followed established principles for connecting design, methods, interpretation, and reporting [31]. The Pillar Integration Process supported transparent movement from strand-specific findings to integrated interpretations [32], and reporting was aligned with qualitative and mixed-methods standards in psychology [33]. Quantitative estimates determined magnitude and uncertainty; qualitative evidence explained experience, context, implementation, and contradiction rather than replacing outcome estimates.

9. ETHICS, SAFETY, CONFIDENTIALITY, AND DATA GOVERNANCE

The protocol received approval from the coordinating health-research ethics committee and from research-governance offices at all participating hospitals before recruitment. Written informed consent covered questionnaires, electronic-record linkage, interviews, observations, and optional description of artwork. Participation was voluntary, refusal did not affect clinical care, and participants could stop an art-making session while remaining in outcome follow-up. Separate written permission governed photography, display, or later use of artwork; no image or identifiable description was reproduced without specific consent.

Study identifiers replaced names, and consent and contact records were stored separately from research data. Electronic files were encrypted on access-controlled institutional systems, paper records were kept in locked storage, and qualitative records were de-identified to reduce deductive disclosure through rare diagnoses, cultural references, professional roles, or artwork descriptions. Access was role-based, transfers were logged, and the linkage key was restricted to the data manager. Facilitators used a safeguarding protocol for acute distress, disclosure of risk, infection control, and physical fatigue. A clinician independent of programme delivery reviewed adverse-event relatedness without access to outcome trends.

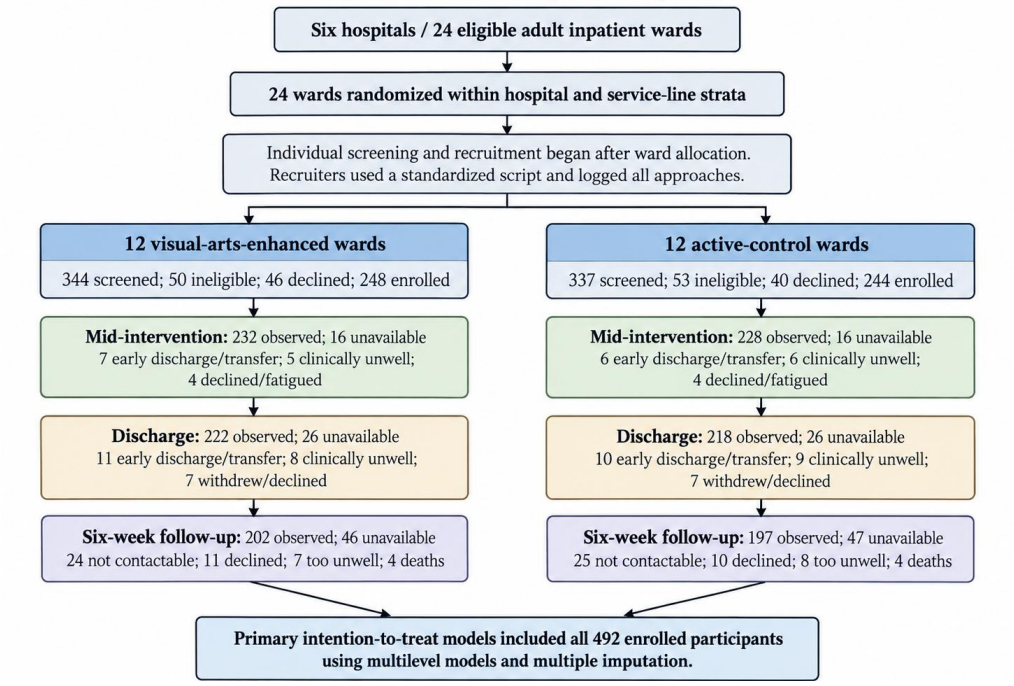


FIGURE 3. Cluster-randomized participant flow. Individual screening and recruitment began after ward allocation; principal reasons for unavailable assessments are shown. All randomized wards and enrolled participants remained in the intention-to-treat estimand.

VI. RESULTS

1. RECRUITMENT, RETENTION, AND BASELINE PROFILE

All 24 wards completed the study. Across intervention wards, 344 patients were screened; 50 did not meet eligibility criteria and 46 declined, leaving 248 enrolled. Across control wards, 337 were screened; 53 were ineligible and 40 declined, leaving 244 enrolled. In total, 681 patients were screened, 578 were eligible, and 492 enrolled. Mid-intervention data were observed for 460 participants (93.5%), discharge data for 440 (89.4%), and six-week follow-up data for 399 (81.1%). Loss to follow-up was associated with greater illness severity, unplanned transfer, and general-medical admission, but not with allocation or baseline PCQ-P after adjustment. Baseline HADS was modestly higher among participants lost at six weeks, supporting inclusion of distress and transfer indicators in the imputation model.

Participants had a mean age of 57.8 years (SD=14.6); 54.5% identified as women, 44.7% as men, and 0.8% used another description or preferred not to respond. Oncology accounted for 34.1% of the cohort, cardiopulmonary or general medicine for 32.9%, and rehabilitation for 33.0%. Mean baseline HADS total was 15.4 (SD=7.1), and mean PCQ-P was 64.8 (SD=17.2). Forty-one percent reported visual-arts participation at least several times per year, whereas 27% reported no voluntary art-making since school. Participant-level standardized mean differences were below .10 for all prespecified variables. Ward size, monthly admissions, and pre-existing patient-visible artwork were also broadly balanced, although the small number of clusters limited the usefulness of significance testing at baseline. Ward-level intraclass correlations were .031 for PCQ-P and .036 for HADS total.

Table 3. Baseline ward and participant characteristics and standardized mean differences.

Characteristic	Visual-arts care (n=248)	Active control (n=244)	SMD
Ward beds, mean (SD)	24.6 (4.3)	24.3 (4.5)	0.07
Monthly ward admissions, mean (SD)	71.4 (18.6)	70.2 (19.1)	0.06
Pre-existing patient-visible artworks, mean (SD)	5.8 (3.4)	6.1 (3.6)	−0.09
Age, years, mean (SD)	57.6 (14.7)	58.0 (14.5)	−0.03
Women, n (%)	137 (55.2)	131 (53.7)	0.03
University education, n (%)	91 (36.7)	86 (35.2)	0.03
Oncology service, n (%)	85 (34.3)	83 (34.0)	0.01
Cardiopulmonary/general medicine, n (%)	81 (32.7)	81 (33.2)	−0.01
Rehabilitation service, n (%)	82 (33.1)	80 (32.8)	0.01
PCQ-P 0–100, mean (SD)	64.7 (17.4)	64.9 (17.0)	−0.01
HADS total 0–42, mean (SD)	15.5 (7.2)	15.3 (7.0)	0.03
PROMIS positive affect T-score	43.8 (8.9)	44.1 (8.7)	−0.03
Pain intensity 0–10	4.4 (2.3)	4.3 (2.2)	0.04
PROMIS sleep disturbance T-score	56.1 (8.4)	55.8 (8.5)	0.04
Previous arts engagement ≥ several times/year	103 (41.5)	99 (40.6)	0.02
High cultural preference congruence	151 (60.9)	146 (59.8)	0.02
Illness-severity index, standardized	0.02 (0.98)	−0.02 (1.02)	0.04

SMD: standardized mean difference. Absolute SMD values below .10 indicate close baseline balance.

2. *INTERVENTION EXPOSURE, FIDELITY, AND ACCEPTABILITY*

Environmental installation fidelity averaged 88% across ward audits; ten of 12 intervention wards met the prespecified 85% threshold throughout implementation, while two fell below threshold during renovation and a prolonged infection-control review. Among intervention participants, 195 of 248 (78.6%) received at least three art-making sessions, 139 (56.0%) received all four, 21 received one or two, and 32 declined or were medically unable to participate. Median art-making contact time was 126 minutes (IQR 84–156). In control wards, 187 of 244 participants (76.6%) received at least three matched contacts, with median contact time of 121 minutes (IQR 82–151).

Facilitator adherence to core art-making procedures averaged 86%, and independent observation confirmed that an explicit choice opportunity was offered in 90% of observed sessions. Forty-seven adaptations were documented: 22 concerned timing around clinical care, 13 altered materials for infection-control or dexterity needs, eight shortened or repositioned activity because of fatigue or mobility, and four adapted language or cultural prompts. Sixteen control participants (6.6%) reported a new visual-art activity during admission, most commonly family-provided colouring materials or personal sketching. Contamination remained limited but was higher on rehabilitation wards with existing recreational activity.

Mean acceptability was 4.1/5 among intervention patients who completed the rating (n=205), 3.9/5 among family respondents (n=62), and 3.8/5 among healthcare professionals (n=53). Staff ratings were higher where facilitators joined morning coordination and lower where sessions competed with therapy, medication, or protected rest. Fourteen patients requested that an artwork be covered, repositioned, or replaced because of glare, visual intensity, cultural mismatch, or distressing association. The median response time was two working days (range same day to five days); two requests were delayed by facilities approval. These

deviations were retained as implementation findings rather than treated as protocol failures. Total incremental provider cost was £99,800 across 12 intervention wards (£8,317 per ward; £203 per assigned participant). Excluding one-time curation and installation, recurrent delivery cost was £124 per assigned participant. No short-term utilization saving offset these costs.

Table 4. Implementation, fidelity, acceptability, and resource-use outcomes.

Indicator	Result	Interpretation
Ward adoption	12/12 intervention wards initiated both components	Complete cluster adoption; four wards had temporary delivery disruption
Environmental fidelity	88% overall; 10/12 wards maintained ≥85%	Generally adequate; renovation and infection-control review reduced fidelity
Participants receiving ≥3 sessions	195/248 (78.6%)	Most received a meaningful dose, but one fifth did not
Participants receiving all 4 sessions	139/248 (56.0%)	Full dose was not achieved by nearly half
Median art-making contact time	126 min (IQR 84–156)	Lower than the maximum planned dose
Facilitator core-procedure adherence	86%	Adequate with documented departures and adaptations
Choice opportunity observed	90% of observed sessions	Person-centred delivery was common but not universal
Control participants receiving ≥3 contacts	187/244 (76.6%)	Attention matching was credible
Control contamination	16/244 (6.6%)	Low-level, mainly patient- or family-initiated art activity
Patient acceptability	4.1/5 (n=205 respondents)	Positive overall with refusal, fatigue, and non-response retained
Artwork change requests	14; median response 2 working days	Cultural, sensory, and operational fit required active management
Documented adaptations	47	Timing, material, mobility, infection-control, language, and cultural adaptation
Incremental programme cost	£99,800 total; £8,317 per intervention ward; £203 per assigned intervention participant	Provider-perspective descriptive cost; recurrent delivery excluding initial installation was £124 per assigned participant

IQR: interquartile range. Fidelity estimates used structured audits, facilitator logs, and independent observations.

3. CONFIRMATORY CO-PRIMARY OUTCOMES

At discharge, the model-based PCQ-P mean was 73.2 in visual-arts-enhanced care and 68.4 in active control. The adjusted difference was +4.8 points (95% CI 1.8–7.8; g=.26; p=.002), close to the prespecified 5-point pragmatic group threshold. The mid-intervention difference was +2.7 (95% CI 0.4–5.0), while the six-week estimate attenuated to +2.1 (95% CI −0.4–4.6). The largest discharge contrast concerned everydayness and recognition of personhood; the safety domain changed less.

HADS total decreased in both conditions during admission. At discharge, model-based means were 12.8 and 14.5, yielding an adjusted difference of −1.7 points (95% CI −2.9 to −0.5; g =- .22; p = .006). This was a small group-level difference and below the 3-point responder threshold used descriptively. The mid-intervention contrast was −0.6 (95% CI −1.5 to 0.3), and the six-week contrast was −0.8 (95% CI −1.9 to 0.3). Anxiety and depression subscales were directionally similar but did not justify separate confirmatory claims.

Complete-case, retention-weighted, ward-fixed-effect, and leave-one-ward-out estimates were similarly directed. Per-protocol contrasts were larger but remained vulnerable to self-selection. Excluding 16 control participants exposed to new visual art changed discharge estimates by less than 0.3 points. A HADS reduction of at least three points occurred in 49% of intervention and 41% of control participants; the adjusted responder confidence interval included no effect. All primary models converged, standardized residuals showed no material departure from assumptions, ward-level variance remained non-zero, and imputation Monte Carlo errors were below 5% of the corresponding standard errors.

Table 5. Confirmatory multilevel estimates for the co-primary outcomes.

Outcome / time	Model-based means intervention / control	Adjusted difference (95% CI)	SE	KR df	Ward ICC	Hedges g	Clinical interpretation
PCQ-P: mid-intervention	69.1 / 66.4	+2.7 (0.4, 5.0)	1.17	20.8	.031	.15	Below 5-point pragmatic threshold
PCQ-P: discharge	73.2 / 68.4	+4.8 (1.8, 7.8)	1.53	20.6	.031	.26	Near 5-point pragmatic threshold; small effect
PCQ-P: six-week follow-up	70.6 / 68.5	+2.1 (−0.4, 4.6)	1.28	20.2	.031	.12	Uncertain; CI includes no effect
HADS: mid-intervention	14.1 / 14.7	−0.6 (−1.5, 0.3)	0.46	21.4	.036	−.08	Small and uncertain
HADS: discharge	12.8 / 14.5	−1.7 (−2.9, −0.5)	0.61	21.0	.036	−.22	Below 3-point responder threshold
HADS: six-week follow-up	13.4 / 14.2	−0.8 (−1.9, 0.3)	0.56	20.7	.036	−.11	Uncertain; CI includes no effect

Positive PCQ-P differences and negative HADS differences favor visual-arts-enhanced care. Discharge was the primary endpoint; we used alpha=.025 for each co-primary discharge outcome. KR: Kenward–Roger. Observed n was 460 at mid-intervention, 440 at discharge, and 399 at follow-up; all 492 participants contributed to the multiply imputed intention-to-treat models.

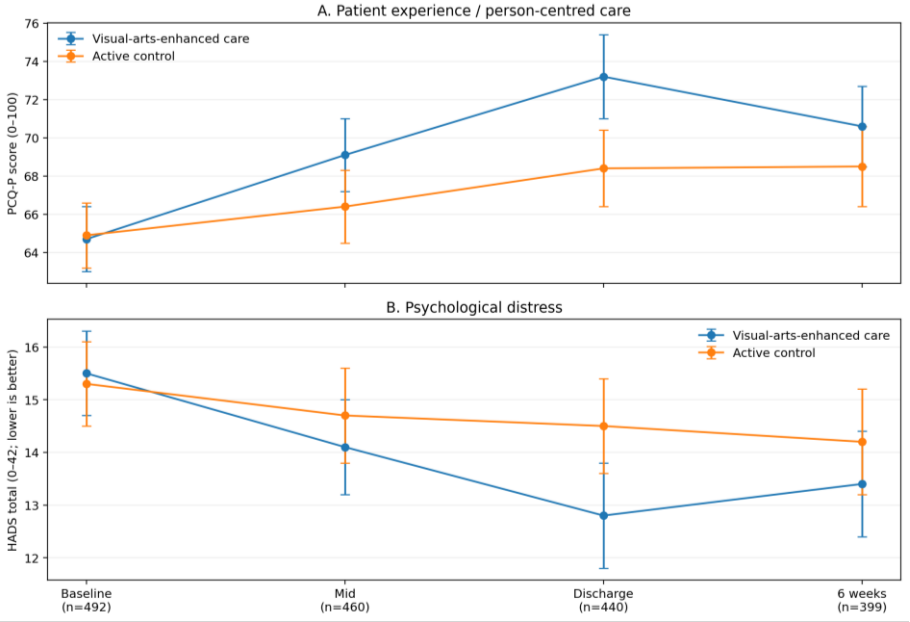


FIGURE 4. Model-based trajectories for the co-primary outcomes. Panel A shows PCQ-P patient-experience scores; Panel B shows HADS psychological-distress scores. Error bars are 95% confidence intervals, and observed sample sizes are shown beneath each time point.

4. SECONDARY PATIENT-REPORTED, CLINICAL, AND SERVICE OUTCOMES

At discharge, the clearest secondary differences after false-discovery-rate adjustment were higher positive affect and greater care satisfaction. Perceived stress, sleep disturbance, recovery self-efficacy, and EQ-VAS estimates favoured the intervention but were small and did not all remain reliable after multiplicity adjustment. Pain intensity and EQ-5D-5L showed no clear contrast. At six weeks, all secondary patient-reported estimates were smaller, and confidence intervals included no effect. These analyses were secondary and do not establish broad or durable recovery.

Clinical-utilization outcomes did not differ reliably. Opioid exposure, length of stay, and 30-day readmission estimates were close to no effect. Ten intervention participants (4.0%) and seven control participants (2.9%) experienced a session-related adverse event or burden. Events included fatigue, frustration, nausea aggravated by visual concentration, and two emotionally activating memories that led to session discontinuation and clinical review. All were mild or moderate; neither condition was attributed to a serious event.

Table 6. Secondary patient-reported outcomes at discharge and six-week follow-up.

Outcome	Discharge effect (95% CI)	FDR q	Six-week effect (95% CI)	FDR q
PROMIS positive affect, T-score	+2.4 (0.7, 4.1)	.028	+1.1 (−0.5, 2.7)	.268
PSS-4 perceived stress	−0.7 (−1.3, −0.1)	.047	−0.4 (−1.0, 0.2)	.326
Pain intensity, 0–10	−0.2 (−0.5, 0.1)	.286	0.0 (−0.3, 0.3)	.918
PROMIS sleep disturbance, T-score	−1.8 (−3.4, −0.2)	.071	−0.5 (−2.1, 1.1)	.611
PROMIS recovery self-efficacy, T-score	+1.9 (0.2, 3.6)	.073	+1.1 (−0.6, 2.8)	.351
EQ-5D-5L index	+.014 (−.004, .032)	.216	+.007 (−.011, .025)	.611

Outcome	Discharge effect (95% CI)	FDR q	Six-week effect (95% CI)	FDR q
EQ-VAS, 0–100	+2.8 (0.1, 5.5)	.088	+1.6 (−1.1, 4.3)	.388
Care satisfaction, 0–100	+5.9 (2.8, 9.0)	.006	—	—

Effects are adjusted mean differences. Positive values favor intervention for positive affect, self-efficacy, EQ-5D-5L, EQ-VAS, and satisfaction; negative values favor intervention for stress, pain, and sleep disturbance. FDR: false-discovery rate.

Table 7. Clinical-utilization, adverse-event, and cost outcomes.

Outcome	Intervention	Control	Adjusted effect (95% CI)	Interpretation
Opioid exposure among analgesic users	—	—	Rate ratio .96 (.84, 1.10)	No reliable difference
Length of stay	—	—	−0.1 days (−0.6, 0.4)	No reliable difference
30-day readmission	10.9%	11.5%	OR .94 (.56, 1.58)	No reliable difference
Session-related adverse event or burden	10/248 (4.0%)	7/244 (2.9%)	Descriptive	Uncommon; all mild or moderate
Incremental provider cost	£203 per assigned participant	Not applicable	£99,800 total intervention cost	No short-term utilization offset

Clinical-utilization outcomes formed a separate descriptive family. OR: odds ratio. Resource costs were descriptive provider-perspective estimates, not a formal cost-effectiveness analysis.

5. EXPLORATORY PROCESS ANALYSIS

Program assignment was associated with greater aesthetic engagement, positive distraction, and perceived control at mid-intervention, but indirect estimates were small. The positive-distraction pathway to discharge HADS was uncertain, while pathways through perceived control and emotion regulation, meaning-making, and social connectedness produced small associations in the expected direction. Direct program estimates remained, mediators were correlated, and results were sensitive to facilitation-quality adjustment. Full estimates are reported in the Supplementary Material.

These process findings were interpreted as conditional associations rather than confirmation of causal mechanisms. Reverse-order models fitted less well but could not exclude mediator–outcome confounding.

6. EXPLORATORY MODERATOR AND SUBGROUP FINDINGS

Baseline distress and cultural congruence showed plausible but statistically uncertain interaction patterns. More distressed participants had a larger discharge HADS contrast, and participants reporting high cultural or personal congruence had a larger PCQ-P contrast, but both interaction confidence intervals included no effect after multiplicity adjustment.

Service line, age, gender, prior arts engagement, illness severity, and planned admission duration produced small and imprecise interactions. All moderator findings were exploratory because the study was powered for overall co-primary effects rather than subgroup differences; complete estimates are provided in the Supplementary Material.

7. QUALITATIVE FINDINGS AND NEGATIVE CASES

Five themes explained how the program was experienced. The first, restoring personhood within clinical routines, described art as a reminder that the patient retained taste, memory, skill, and relationships beyond the clinical record. One oncology patient stated, “For half an hour I was not the woman in bed six; I was choosing colors again” (P17, intervention, oncology). A nurse similarly observed, “The picture gave us

something to ask about that was not pain or discharge" (H06, oncology). The theme was strongest when the activity arose from patient choice rather than staff enthusiasm.

The second theme, creating a workable distance from distress, captured focused sensory attention without claiming that anxiety or sadness disappeared. A rehabilitation patient explained, "It did not remove the worry. It gave the worry somewhere to sit while I worked" (P31, intervention, rehabilitation). Immediate distraction was useful during waiting, pain, and anticipatory anxiety, but several participants described benefit as temporary. Patients with continuing diagnostic uncertainty often returned to distress soon after the session, consistent with the attenuation of quantitative effects after discharge.

The third theme, choice, cultural recognition, and control, showed why curation and facilitation mattered. "I liked being able to say no and change the picture. That mattered more than liking every image" (P08, intervention, general medicine). Cultural recognition strengthened trust when it was plural and contextualized, whereas generic heritage motifs could feel tokenistic. One patient requested removal of a highly patterned work: "It was too busy when I was nauseated; I wanted the wall plain" (P39, intervention, oncology). These accounts challenged the assumption that greater visual stimulation was inherently supportive.

The fourth theme, making connection visible, described artwork as a relational object. A family member said, "We talked about home because of the drawing, not about the next test" (F05, family, rehabilitation). Healthcare professionals reported that art sometimes revealed preferences and strengths that improved conversation, but they rejected clinical interpretation without permission. "The image opened a door, but it was not our job to decide what it meant" (H14, general medicine). Small-group activity supported some rehabilitation patients and felt exposing to others who preferred privacy.

The fifth theme, timing, fatigue, and performance pressure, defined the principal boundary conditions. Fourteen interviewed patients had declined or stopped at least one session because of pain, fatigue, treatment schedules, low mood, or fear of being judged. "I was tired and thought I had to make something good, so it became another task" (P22, intervention, cardiopulmonary). An artist-facilitator reflected, "Sometimes the most person-centered decision was to shorten the session or not begin it" (A03). Negative cases clustered where sessions were offered during clinical disruption, materials were insufficiently adapted, or patients believed participation was expected. These cases supported opt-in delivery, flexible timing, and explicit permission to stop.

Table 8. Qualitative themes, supporting evidence, and boundary conditions.

Theme	Illustrative quotation(s)	Convergent evidence	Boundary / divergent evidence
Restoring personhood	"I was not the woman in bed six; I was choosing colors again" (P17).	PCQ-P everydayness; staff observations of non-clinical conversation	Benefits weakened when staff framed participation as expected
Workable distance from distress	"It gave the worry somewhere to sit while I worked" (P31).	Short-term HADS and positive-affect pattern; reflective cards	Relief was temporary for patients facing continuing diagnostic uncertainty
Choice, cultural recognition, and control	"Being able to say no … mattered more than liking every image" (P08).	Choice fidelity; cultural-congruence pattern	Busy, tokenistic, or personally distressing images prompted avoidance or removal
Making connection visible	"We talked about home because of the drawing" (F05).	Social-connectedness process estimate; family and staff accounts	Group participation felt exposing for some patients

Theme	Illustrative quotation(s)	Convergent evidence	Boundary / divergent evidence
Timing, fatigue, and performance pressure	"It became another task" (P22).	Missed sessions, adaptations, workflow observations	Fatigue, treatment schedules, and fear of judgment constrained dose and benefit

P denotes patient, F family member, H healthcare professional, and A artist-facilitator. All quotations were from English-language interviews and were de-identified. We compared themes across hospital, service line, stakeholder role, cultural congruence, intervention dose, and quantitative response profile; negative cases refined theme definitions by distinguishing voluntary engagement from perceived performance pressure.

8. MIXED-METHODS INTEGRATION BY RESEARCH QUESTION

For RQ1, the discharge difference in patient experience converged with accounts of personhood, choice, and relational recognition. The distress result converged with descriptions of manageable attentional distance and temporary relief. The weaker six-week estimates aligned with reduced exposure, ongoing illness, and the absence of a sustained post-discharge component. Participants with persistent clinical uncertainty often described meaningful sessions without durable symptom change.

For RQ2, qualitative evidence supported short-term positive affect and satisfaction while clarifying why pain, quality of life, medication, length of stay, and readmission remained unchanged. Patients and staff described art as altering the quality and meaning of time in hospital rather than the biological course of illness or the operational criteria for discharge. The null service outcomes therefore bounded rather than contradicted the experiential findings.

For RQ3, accounts of control, meaning, and connection aligned with the exploratory process estimates, but negative cases showed that the same activity could generate fatigue, exposure, or pressure. Baseline distress and cultural congruence produced plausible subgroup patterns, yet the interaction intervals were wide. Integration therefore refined the VAER Framework toward conditional and heterogeneous processes rather than confirming a uniform mechanism or subgroup effect.

For RQ4, implementation evidence showed complementary but inseparable functions. Environmental work provided low-demand access for patients unable or unwilling to join sessions, while facilitated making offered authorship and expression unavailable through viewing alone. Neither component benefited every patient, and delivery quality varied with staffing, timing, facilities responsiveness, and adaptation. The integrated interpretation was that the combined program produced modest short-term experiential and psychological advantages when delivered as optional, culturally responsive, and operationally coordinated care; the design did not identify which component was responsible.

Table 9. Research-question-level mixed-methods joint display.

RQ	Quantitative pattern	Qualitative explanation / challenge	Integrated inference
RQ1	Modest discharge differences; follow-up estimates attenuated	Personhood, choice, and temporary regulatory relief; persistent illness constrained retention	Short-term experiential and distress advantages, not durable recovery
RQ2	Positive affect and satisfaction clearest; clinical utilization null	Art changed the quality of hospital time rather than disease course or discharge criteria	Experiential significance did not imply medical or service effects
RQ3	Small exploratory process associations; moderator interactions imprecise	Control, meaning, connection, cultural fit, and negative cases refined pathways	Associations in expected directions, not confirmed mechanisms or subgroups

RQ	Quantitative pattern	Qualitative explanation / challenge	Integrated inference
RQ4	Fidelity and dose were adequate but incomplete; adverse burdens were uncommon	Timing, facilities response, mobility, privacy, and voluntary choice shaped experience	Effectiveness depended on optional, culturally responsive, coordinated implementation

The joint display distinguishes adjusted effects, stakeholder explanation, contradictory evidence, and integrated interpretation.

VII. DISCUSSION

This multicenter cluster-randomized evaluation found small program-level differences at the prespecified discharge endpoint: patient experience was 4.8 points higher on the PCQ-P and psychological distress 1.7 HADS points lower than under matched supportive contact. The PCQ-P estimate approached the pragmatic 5-point threshold, whereas the HADS estimate remained below the 3-point responder threshold. Both follow-up estimates were smaller and uncertain. The pattern therefore indicates a short-term, experiential, group-level psychological response during admission rather than durable individual recovery.

These effect sizes are smaller than those often reported in uncontrolled or single-site arts studies and are compatible with reviews describing promising but heterogeneous effects [5]–[8]. The active comparator reduced the likelihood that contact time, structured attention, or expectancy alone explained the result. Pragmatic inclusion of patients with limited interest, incomplete dose, and varied diagnoses also produced a more conservative estimate than an efficacy study restricted to highly engaged participants.

Secondary outcomes showed a selective profile. Positive affect and care satisfaction were most consistent with the co-primary experience result; pain, generic quality of life, medication exposure, length of stay, and readmission were not reliably altered. Therefore, statistical significance, individual clinical importance, experiential meaning, implementation quality, and economic value require separate interpretation. The intervention changed how some patients experienced time, choice, and interaction in hospital without affecting disease course or service use.

Implementation findings further qualified effectiveness. Only 56% received all four sessions, environmental fidelity averaged 88%, and requests to modify artwork were not always resolved immediately. The provider cost was £203 per assigned intervention participant, with no short-term utilization offset. These findings do not negate experiential value, but they prevent claims of cost saving or comprehensive clinical benefit.

The exploratory process findings were consistent with the VAER Framework, although they explained only a small proportion of the contrast between programs. Positive distraction was linked primarily with immediate experience; perceived control, emotion regulation, meaning-making, and social connection were more closely associated with retained outcomes. Because the mediators were correlated, measured with study-specific scales, and not randomized, and because those scales have not been independently validated, the findings identify plausible associations rather than confirmed causal mechanisms.

Environmental access offered low-demand engagement, while facilitated making added authorship and expression. Their combined delivery may have created several routes to recognition and agency, but the bundled design precluded component attribution. The VAER Framework is consequently best treated as a parsimonious map of processes and boundary conditions rather than a validated causal chain.

Familiar content was neither necessary nor sufficient: unfamiliar work could be meaningful when contextualized, whereas familiar motifs could feel stereotyped or intrusive. Cultural responsiveness therefore involved plural representation, interpretive openness, and the ability to refuse, cover, or replace an image. The interaction estimate remained uncertain, so this conclusion rests on integrated quantitative and qualitative evidence rather than a confirmed subgroup effect.

Within person-centered care, aesthetic and narrative agency appeared to create an additional medium for recognition. Patients selected materials, controlled the pace, decided whether to keep work private, and used images to redirect the conversation. These practices complemented, but did not replace, clinical communication, safety, or treatment.

Morning coordination reduced interruptions, whereas medication rounds, therapy, nausea, fatigue, and protected rest reduced dose. Material adaptation, infection-control approval, privacy, language support, and facilities response affected whether the programme was experienced as respectful or burdensome.

Healthcare organizations may consider environmental curation, facilitated art-making, and governance as distinct implementation functions. Curation requires patient-visible placement, glare and infection-control review, representational plurality, and a responsive replacement pathway. Participatory activity requires trained practitioners who understand consent, mobility, fatigue, distress escalation, and non-diagnostic boundaries. Governance requires explicit protection of privacy, ownership, refusal, and non-art alternatives.

The active control was clinically informative because conversation, reading, puzzles, and goal-setting also provided attention and relief from routine. The estimated effect therefore represents the additional contribution of the combined visual-arts service beyond matched supportive contact. It does not justify positioning visual art as a universal symbol of compassionate care.

Resource implications remained unresolved. The programme required curation, mounting, materials, facilitation, coordination, supervision, and response to replacement requests. A formal economic evaluation would require longer follow-up and explicit valuation of patient experience, equity, and maintenance, rather than assuming reduced length of stay would finance the program.

1. STRENGTHS, LIMITATIONS, AND TRANSFERABILITY

Strengths included multicentre ward-level randomization, an attention- and time-matched comparator, repeated measurement, cluster-aware intention-to-treat analysis with small-sample correction, multiple imputation, multiplicity control, active fidelity assessment, clinical-record outcomes, and mixed-methods integration that retained negative cases. The design separated experiential and psychological outcomes from healthcare utilization and did not use qualitative testimony as proof of statistical effect.

Several limitations constrain interpretation. Individual recruitment occurred after ward allocation, and recruitment nurses could not be blinded; standardized scripts and approach logs reduced but did not eliminate differential recruitment. Patients, ward staff, and facilitators knew assignment because the environmental and participatory components were visible, creating possible expectancy, performance, and social-desirability effects. Questionnaire assessors were formally masked, but allocation was correctly inferred for 58% of audited assessments, so masking was incomplete. The active comparator reduced, but could not remove, unequal expectations about the attractiveness of the two conditions or bias in self-reported co-primary outcomes.

The programme bundled environmental artwork and facilitated art-making. The trial estimated joint effectiveness and could not identify whether one component, their interaction, or choice between them produced the observed difference. Exposure was heterogeneous: 32 assigned patients completed no art-making session, and pre-existing artwork and patient-initiated hobbies produced low-level contamination. A factorial or dismantling design would be required for component-specific attribution.

The follow-up lasted six weeks, limiting inferences about durability, sustained psychological recovery, rehabilitation, repeated healthcare use, or maintenance after the inpatient program ended. HADS measured psychological distress rather than comprehensive emotional recovery, and PCQ-P captured perceived care climate rather than the full range of patient experience. Most outcomes were self-reported. The study-specific process and cultural-congruence scales showed acceptable internal performance and partial invariance but require independent validation before confirmatory or cross-cultural use. Mediation remained vulnerable to unmeasured confounding, and 24 wards provided limited power for interactions and subgroup analyses.

Cross-diagnostic heterogeneity strengthened pragmatic relevance but reduced condition-specific precision. Findings do not generalize to intensive care, paediatrics, emergency care, psychiatric inpatient settings, severe cognitive impairment, or very short admissions. The English-only qualitative corpus excluded perspectives that required interview translation, despite multilingual questionnaire support. Qualitative volunteers may also have been more willing to discuss the program, although sampling included refusal, deterioration, cultural mismatch, and adverse experience. Transferability depends on preserving choice, trained non-diagnostic facilitation, cultural responsiveness, privacy, flexible timing, and safeguarding rather than reproducing a fixed artwork collection.

VIII. CONCLUSION

Among medically stable adult inpatients, assignment to a combined visual-arts-enhanced care program was associated with modestly higher patient-experience scores and lower psychological distress at discharge than an attention- and time-matched non-art wellbeing program. The patient-experience estimate approached a pragmatic group-level threshold, whereas the distress difference was small. Because both six-week estimates were uncertain, the findings do not establish durable emotional recovery. Positive affect and care satisfaction showed the clearest secondary differences, while pain, generic quality of life, opioid exposure, length of stay, and readmission did not differ reliably.

Exploratory process estimates and qualitative accounts indicated possible roles for choice, perceived control, emotion regulation, meaning-making, and social connection, but they did not establish unique mechanisms. Fatigue, clinical interruption, cultural mismatch, visual overstimulation, and performance pressure constrained acceptability for some patients. The program is therefore best understood as an optional, context-sensitive component of person-centered inpatient care. Its observed effect applies to the combined service, not to either component independently, and its value depends on culturally responsive curation, privacy, trained non-diagnostic facilitation, flexible timing, and access to non-art alternatives.

Author Contributions

The author conducted the conceptualization, methodology, data analysis, investigation, writing, review, editing, and final approval of the manuscript.

Funding

This research received no external funding.

Data Availability

The dataset will be available from the author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.

REFERENCES

[1]. L. Lankston, P. Cusack, C. Fremantle, and C. Isles, "Visual art in hospitals: Case studies and review of the evidence," Journal of the Royal Society of Medicine, vol. 103, no. 12, pp. 490–499, 2010, doi: 10.1258/jrsm.2010.100256.

[2]. S. L. Nielsen, L. B. Fich, K. K. Roessler, and M. F. Mullins, "How do patients actually experience and use art in hospitals? The significance of interaction: A user-oriented experimental case study," International Journal of Qualitative Studies on Health and Well-being, vol. 12, no. 1, art. 1267343, 2017, doi: 10.1080/17482631.2016.1267343.

[3]. A. M. Ullán and M. H. Belver, "Visual arts in children's hospitals: Scoping review," HERD: Health Environments Research & Design Journal, vol. 14, no. 4, pp. 339–367, 2021, doi: 10.1177/19375867211003494.

[4]. A. Fudickar, D. Konetzka, S. M. L. Nielsen, and K. Hathorn, "Evidence-based art in the hospital," Wiener Medizinische Wochenschrift, vol. 172, nos. 9–10, pp. 234–241, 2022, doi: 10.1007/s10354-021-00861-7.

[5]. D. Regev and L. Cohen-Yatziv, "Effectiveness of art therapy with adult clients in 2018—What progress has been made?" Frontiers in Psychology, vol. 9, art. 1531, 2018, doi: 10.3389/fpsyg.2018.01531.

[6]. A. Abbing, A. Ponstein, S. van Hooren, L. de Sonneville, H. Swaab, and E. Baars, "The effectiveness of art therapy for anxiety in adults: A systematic review of randomized and non-randomized controlled trials," PLOS ONE, vol. 13, no. 12, art. e0208716, 2018, doi: 10.1371/journal.pone.0208716.

[7]. J. T. Bosman, Z. M. Bood, M. Scherer-Rath, H. Dörr, N. Christophe, M. A. G. Sprangers, and H. W. M. van Laarhoven, "The effects of art therapy on anxiety, depression, and quality of life in adults with cancer: A systematic literature review," Supportive Care in Cancer, vol. 29, no. 5, pp. 2289–2298, 2021, doi: 10.1007/s00520-020-05869-0.

[8]. L. Uttley, A. Scope, M. Stevenson, A. Rawdin, E. Taylor Buck, A. Sutton, J. Stevens, E. Kaltenthaler, K. Dent-Brown, and C. Wood, "Systematic review and economic modelling of the clinical effectiveness and cost-effectiveness of art therapy among people with non-psychotic mental health disorders," Health Technology Assessment, vol. 19, no. 18, pp. 1–120, 2015, doi: 10.3310/hta19180.

[9]. G. Kaimal, K. Ray, and J. Muniz, "Reduction of cortisol levels and participants' responses following art making," Art Therapy, vol. 33, no. 2, pp. 74–80, 2016, doi: 10.1080/07421656.2016.1166832.
[10]. J. Czamanski-Cohen and K. L. Weihs, "The Bodymind Model: A platform for studying the mechanisms of change induced by art therapy," The Arts in Psychotherapy, vol. 51, pp. 63–71, 2016, doi: 10.1016/j.aip.2016.08.006.
[11]. T. O. Iyendo, P. C. Uwajeh, and E. S. Ikenna, "The therapeutic impacts of environmental design interventions on wellness in clinical settings: A narrative review," Complementary Therapies in Clinical Practice, vol. 24, pp. 174–188, 2016, doi: 10.1016/j.ctcp.2016.06.008.
[12]. C. C. Andrade and A. S. Devlin, "Stress reduction in the hospital room: Applying Ulrich's theory of supportive design," Journal of Environmental Psychology, vol. 41, pp. 125–134, 2015, doi: 10.1016/j.jenvp.2014.12.001.
[13]. R. S. Ulrich, R. F. Simons, B. D. Losito, E. Fiorito, M. A. Miles, and M. Zelson, "Stress recovery during exposure to natural and urban environments," Journal of Environmental Psychology, vol. 11, no. 3, pp. 201–230, 1991, doi: 10.1016/S0272-4944(05)80184-7.
[14]. J. J. Gross, "Emotion regulation: Current status and future prospects," Psychological Inquiry, vol. 26, no. 1, pp. 1–26, 2015, doi: 10.1080/1047840X.2014.940781.
[15]. M. J. Santana, K. Manalili, R. J. Jolley, S. Zelinsky, H. Quan, and M. Lu, "How to practice person-centred care: A conceptual framework," Health Expectations, vol. 21, no. 2, pp. 429–440, 2018, doi: 10.1111/hex.12640.
[16]. G. M. Fix, C. S. VanDeusen Lukas, M. E. Bolton, J. N. Hill, C. Mueller, L. LaVela, and B. G. Bokhour, "Patient-centered care is a way of doing things: How healthcare employees conceptualize patient-centred care," Health Expectations, vol. 21, no. 1, pp. 300–307, 2018, doi: 10.1111/hex.12615.
[17]. L. J. Damschroder, C. M. Reardon, M. A. O. Widerquist, and J. Lowery, "The updated Consolidated Framework for Implementation Research based on user feedback," Implementation Science, vol. 17, art. 75, 2022, doi: 10.1186/s13012-022-01245-0.
[18]. K. Skivington et al., "A new framework for developing and evaluating complex interventions: Update of Medical Research Council guidance," BMJ, vol. 374, art. n2061, 2021, doi: 10.1136/bmj.n2061.
[19]. M. K. Campbell, G. Piaggio, D. R. Elbourne, and D. G. Altman, "CONSORT 2010 statement: Extension to cluster randomized trials," BMJ, vol. 345, art. e5661, 2012, doi: 10.1136/bmj.e5661.
[20]. I. Djukanovic, J. Carlsson, and K. Årestedt, "Is the Hospital Anxiety and Depression Scale (HADS) a valid measure in a general population 65–80 years old? A psychometric evaluation study," Health and Quality of Life Outcomes, vol. 15, art. 193, 2017, doi: 10.1186/s12955-017-0759-9.
[21]. D. Edvardsson, P. O. Sandman, and B. Rasmussen, "Swedish language Person-centered Climate Questionnaire—Patient version: Construction and psychometric evaluation," Journal of Advanced Nursing, vol. 63, no. 3, pp. 302–309, 2008, doi: 10.1111/j.1365-2648.2008.04709.x.
[22]. D. Cella et al., "The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008," Journal of Clinical Epidemiology, vol. 63, no. 11, pp. 1179–1194, 2010, doi: 10.1016/j.jclinepi.2010.04.011.
[23]. L. Yu, D. J. Buysse, A. Germain, D. E. Moul, A. Stover, N. E. Dodds, K. L. Johnston, and P. A. Pilkonis, "Development of short forms from the PROMIS sleep disturbance and Sleep-Related Impairment item banks," Behavioral Sleep Medicine, vol. 10, no. 1, pp. 6–24, 2012, doi: 10.1080/15402002.2012.636266.
[24]. M. Herdman et al., "Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L)," Quality of Life Research, vol. 20, no. 10, pp. 1727–1736, 2011, doi: 10.1007/s11136-011-9903-x.
[25]. R. E. Glasgow, T. M. Harden, B. Gaglio, B. Rabin, M. L. Smith, G. C. Porter, M. G. Ory, and P. A. Estabrooks, "RE-AIM planning and evaluation framework: Adapting to new science and practice with a 20-year review," Frontiers in Public Health, vol. 7, art. 64, 2019, doi: 10.3389/fpubh.2019.00064.
[26]. C. Rutterford, A. Copas, and S. Eldridge, "Methods for sample size determination in cluster randomized trials," International Journal of Epidemiology, vol. 44, no. 3, pp. 1051–1067, 2015, doi: 10.1093/ije/dyv113.
[27]. I. R. White, P. Royston, and A. M. Wood, "Multiple imputation using chained equations: Issues and guidance for practice," Statistics in Medicine, vol. 30, no. 4, pp. 377–399, 2011, doi: 10.1002/sim.4067.
[28]. Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," Journal of the Royal Statistical Society: Series B, vol. 57, no. 1, pp. 289–300, 1995, doi: 10.1111/j.2517-6161.1995.tb02031.x.
[29]. T. J. VanderWeele, "Mediation analysis: A practitioner's guide," Annual Review of Public Health, vol. 37, pp. 17–32, 2016, doi: 10.1146/annurev-publhealth-032315-021402.

[30]. N. K. Gale, G. Heath, E. Cameron, S. Rashid, and S. Redwood, "Using the framework method for the analysis of qualitative data in multi-disciplinary health research," BMC Medical Research Methodology, vol. 13, art. 117, 2013, doi: 10.1186/1471-2288-13-117.

[31]. M. D. Fetters, L. A. Curry, and J. W. Creswell, "Achieving integration in mixed methods designs—Principles and practices," Health Services Research, vol. 48, no. 6, pt. 2, pp. 2134–2156, 2013, doi: 10.1111/1475-6773.12117.

[32]. R. E. Johnson, A. L. Grove, and A. Clarke, "Pillar Integration Process: A joint display technique to integrate data in mixed methods research," Journal of Mixed Methods Research, vol. 13, no. 3, pp. 301–320, 2019, doi: 10.1177/1558689817743108.

[33]. H. M. Levitt, M. Bamberg, J. W. Creswell, D. M. Frost, R. Josselson, and C. Suárez-Orozco, "Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology," American Psychologist, vol. 73, no. 1, pp. 26–46, 2018, doi: 10.1037/amp0000151.