Clinical effectiveness of music interventions for dementia and depression in elderly care (MIDDEL): Australian cohort of an international pragmatic cluster-randomised controlled trial

Felicity A Baker, Young-Eun C Lee, Tanara Vieira Sousa, Phoebe A Stretton-Smith, Jeanette Tamplin, Vigdis Sveinsdottir, Monika Geretsegger, Jo Dugstad Wake, Jörg Assmus, Christian Gold

Summary

Background Dementia and depression are highly prevalent and comorbid conditions among older adults living in care homes and are associated with individual distress and rising societal costs. Effective, scalable, and feasible interventions are needed. Music interventions have shown promising effects, but the current evidence base is inconclusive. The present study aimed to determine the effectiveness of two different music interventions on the depressive symptoms of people with dementia living in residential aged care.

Methods We implemented a 2×2 factorial cluster-randomised controlled trial to determine whether group music therapy (GMT) is more effective than no GMT with standard care, or recreational choir singing (RCS) is more effective than no RCS with standard care, for reducing depressive symptoms and other secondary outcomes in people with dementia with mild to severe depressive symptoms living in residential aged care. Care home units with at least ten residents were allocated to GMT, RCS, GMT plus RCS, or standard care, using a computer-generated list with block randomisation (block size four). The protocolised interventions were delivered by music therapists (GMT) and community musicians (RCS). The primary outcome was Montgomery-Åsberg Depression Rating Scale score at 6 months, assessed by a masked assessor and analysed on an intention-to-treat basis using linear mixed-effects models, which examined the effects of GMT versus no-GMT and RCS versus no-RCS, as well as interaction effects of GMT and RCS. We report on the Australian cohort of an international trial. This trial is registered with ClinicalTrials.gov, NCT03496675, and anzctr.org.au, ACTRN12618000156280.

Findings Between June 15, 2018, and Feb 18, 2020, we approached 12 RAC facilities with 26 eligible care home units and, excluding six units who could not be enrolled due to COVID-19 lockdowns, we screened 818 residents. Between July 18, 2018, and Nov 26, 2019, 20 care home units were randomised (318 residents). Recruitment ceased on March 17, 2020, due to COVID-19. The primary endpoint, available from 20 care home units (214 residents), suggested beneficial effects of RCS (mean difference −4·25, 95% CI −7·89 to −0·62; p=0·0221) but not GMT (mean difference −0·44, −4·32 to 3·43; p=0·8224). No related serious adverse events occurred.

Interpretation Our study supports implementing recreational choir singing as a clinically relevant therapeutic intervention in reducing depressive symptoms for people with dementia in the Australian care home context.

Funding National Health and Medical Research Council, Australia.

Lancet Healthy Longev 2022; 3: e153–65

See **Comment** page e124

Faculty of Fine Arts and Music, The University of Melbourne, Melbourne, VIC, Australia (F A Baker PhD, Y-E C Lee PhD, T Vieira Sousa PhD, P A Stretton-Smith MMusThrp, J Tamplin PhD); **Norwegian Academy of Music, Oslo, Norway** (F A Baker); **NORCE Norwegian Research Centre, Bergen, Norway** (V Sveinsdottir PhD, M Geretsegger PhD, J Dugstad Wake PhD, J Assmus PhD, C Gold PhD); **Department of Clinical and Health Psychology, University of Vienna, Vienna, Austria** (C Gold)

Correspondence to:
Prof Felicity Baker, Faculty of Fine Arts and Music, The University of Melbourne, Melbourne 3006, VIC, Australia
felicity.baker@unimelb.edu.au

Introduction

Dementia and depression are highly prevalent and comorbid conditions in older adults living in residential aged care (RAC).[1,2] In 2019–20, more than half of all people living in Australian RAC had dementia, and depression was the most common health condition affecting care home residents with dementia.[2] Depression in dementia is associated with individual distress, increased cognitive and functional impairment, and increased mortality.[1] Neuropsychiatric symptoms, particularly depression and agitation, are also strong predictors of poor quality of life in residents with dementia,[3] and contribute to increased carer stress and high health-care and societal costs.[1,4] There is growing focus on non-pharmacological interventions as first-line treatments for depression and other neuropsychiatric symptoms in dementia due to the limited efficacy and adverse effects of psychotropic medication.[1]

Increasing evidence suggests that music interventions are promising and potentially cost-effective non-pharmacological approaches for people with dementia.[4–9] They show potential to reduce symptoms of depression and agitation, while improving quality of life and wellbeing. Music interventions can be beneficial even for people in more advanced stages of dementia,[6] possibly

Research in context

Evidence before this study

A 2018 Cochrane systematic review of randomised controlled trials of music-based therapeutic interventions (active, or passive or receptive) involving people with dementia found moderate-quality evidence that the interventions reduced depressive symptoms at the end of treatment, but not in the long term. However, it is difficult to draw firm conclusions about the clinical significance of the effect of specific music-based interventions on depression for people with dementia due to the heterogeneity of the type, frequency and duration of interventions, and the range of outcome measures used to assess depressive symptoms.

In our updated review of the literature in June, 2021, we searched PsycINFO and PubMed for randomised controlled trials published between July 1, 2017, to June 29, 2021, using the search string "dementia OR Alzheimer (s) AND depress* AND music therapy OR music interventions", with no language restrictions.

Our search identified three small randomised controlled trials (sample sizes between 50 and 62), which measured the effects of active group music interventions on depressive symptoms in people with dementia. The populations included in all three trials differed from our participants, because they only included people with mild to moderate dementia. Two trials examined active 12-week group music interventions; the first focused on 60-min weekly percussion instrument playing with familiar music delivered by a trained music facilitator, and the second involved a 2 h weekly singing intervention delivered by a professional choir conductor. Neither study found significant effects of the music intervention on depressive symptoms. In the third trial, credentialed music therapists delivered a 30–60-min, small group music therapy intervention three times per week for 2 weeks, and trained nursing assistants to incorporate music activities into their daily caregiving routine in a 3-day training course. Results showed clinically important reductions in depressive symptoms after 2 weeks of music therapy. Depression scores increased during a 2-week wash-out period but appeared to stabilise following the 2 weeks of nursing-assistant delivered music activities. Further, existing literature highlights the need for a systematic investigation of active music interventions within the residential aged care context.

Added value of this study

MIDDEL examined the two most widely used active music interventions for people with dementia in residential aged care: group music therapy (GMT) and recreational choir singing (RCS). These interventions were more intensive and standardised than those reported in previous literature. RCS reduced depressive symptoms at the end of the 6-month intervention period; positive effects were also observed on secondary outcome measures of neuropsychiatric symptoms and quality of life. Importantly, the singing intervention effects on depression outcomes were sustained in the long term (12 months).

Implications of all the available evidence

Taken together with earlier evidence, intensive, active music intervention with primary focus on singing designed and supervised by credentialed music therapists can significantly improve symptoms of depression in care home residents with dementia and depressive symptoms. In particular, the component of structured group singing can improve depression and neuropsychiatric symptoms and is easily scalable.

Findings of the Australian cohort of the MIDDEL trial are timely because the Australian Government is currently revising its aged care policies and funding models as a result of the recent Royal Commission into Aged Care. The Commission's recommendations were that music therapy programmes be included as a compulsory service of every aged care provider. Our findings therefore have important policy and practice implications.

due to relative preservation of brain areas relevant to music memory.[10] Group music therapy (GMT) and recreational choir singing (RCS) are among the most widely used and studied active music interventions for people with dementia.[9] Both involve a combination of biological, psychological (cognitive and emotional), and social mechanisms thought to be associated with improved mood and depressive symptoms.[11] Although GMT has a therapeutic focus and includes a broader range of music therapy methods, RCS focuses primarily on singing in larger groups and is more scalable.[11] However, there have been few rigorous and high-powered studies on how these existing interventions can be used in the care of people with dementia and depression living in residential aged care (RAC).[12] There is therefore a need for a large-scale randomised controlled trial to systematically evaluate the clinical effectiveness of these interventions across a range of severity levels of dementia and depression in RAC settings.

The music interventions for dementia and depression in elderly care (MIDDEL) trial aimed to identify the main effects of GMT or RCS, compared with the absence of GMT or RCS, with standard care; any interaction effects of GMT and RCS; and the predictive effects on intervention efficacy of clinical characteristics, specifically severity of dementia (selected as the most important from a larger number of characteristics specified for the international trial).[11] We hypothesised GMT to be superior to no GMT and RCS superior to no RCS in reducing depression symptoms over 6 months.[11] Given the significant potential for contamination between intervention arms of trials within RAC contexts, a cluster-randomised design was used.[13]

Methods

Study design

MIDDEL is a cluster-randomised controlled trial which was conducted in RAC settings in Melbourne, VIC,

Australia, and is also being conducted in Germany, the Netherlands, Norway, Turkey, and the UK. The international trial commenced on April 12, 2021, and the results of this trial will be reported subsequently. MIDDEL uses a 2×2 factorial design to examine effects of GMT, RCS, both, and neither over 12 months, for residents living with dementia and depressive symptoms. The MIDDEL protocol, resource use, and intervention fidelity protocols are published elsewhere and describe the neuropsychological underpinnings of the project.[11,14,15] The protocol is available online. There were two changes to the protocol; first, expansion of the inclusion criteria to a Clinical Dementia Rating (CDR) score of 0·5–3 to include a broader spectrum of people with dementia (mild to severe) and, second, exclusion of the CDR as a secondary outcome at follow-up due to the additional burden on participants and the limited expectations for change over time (appendix p 1). These changes were implemented before the enrolment of the first study participant (appendix p 1).

Australia, the first country to receive funding, commence the trial, and complete recruitment and follow-up assessments, enrolled ten large private RAC facilities, which were already subdivided into two-to-three self-contained units. In metropolitan Melbourne, the majority of RAC facilities are private. The trial was stopped early due to extended COVID-19 lockdowns in Australia. The decision to analyse and publish the data from the Australian cohort was based on an assessment of the benefits and risks of early publication, in consideration of the project period and delays in funding and trial commencement in Europe, and was endorsed by the data and safety monitoring committee.

Ethical approval was obtained from the Medicine and Dentistry Human Ethics Sub-Committee at The University of Melbourne, Australia (Jan 12, 2018, Ethics ID 1750400).

Participants

The 20 largest care home units were recruited from the pool of care home units available for inclusion. We implemented a staggered rollout of the study; four care home units commenced every 3 to 4 months, with a total of five waves of four care home units randomised per wave. Individual RAC facilities were selected by the RAC providers based on size and staff capacity to support the project. All potentially eligible residents were screened for eligibility by a neuropsychologist. Formal written consent was obtained for all residents before their unit was randomised. Additionally, oral assent was sought from both residents and their next of kin. Residents' capacity to provide informed consent was assessed informally by assessors and confirmed by nursing staff. For residents not able to provide informed consent, their legal guardian or next of kin provided consent on their behalf. After consent was received, eligible residents were enrolled, and the full baseline assessment was completed.

Eligibility criteria were defined at two levels: care home units (cluster) and individual participants (residents). Eligible care home units had at least ten eligible and consenting residents and were separate and self-contained to avoid potential intervention contamination. Individual residents met the following inclusion criteria: aged 65 years or older; had dementia as indicated by a CDR score of 0·5–3 and a Mini-Mental State Examination (MMSE) score of 26 or less; mild to severe depressive symptoms (8 or more on the Montgomery-Åsberg Depression Rating Scale, MADRS). Residents in short-term care, with known diagnoses of schizophrenia or Parkinson's disease, severe hearing impairment, or an inability to tolerate sitting in a chair for at least part of the sessions were excluded. In line with the pragmatic design of the trial, other neurological or psychiatric illnesses or poor or no knowledge of English language skills were not reasons for exclusion; family and staff assisted communication for participants without English proficiency. Permanent RAC staff, employed for at least 0·4 full-time equivalent at the time the site was randomised, were also recruited as participants to assess the secondary outcome of staff burden. Agency staff were excluded.

For the **protocol** see http://dx.doi.org/10.1136/bmjopen-2018-023436

See **Online** for appendix

Randomisation and masking

Block randomisation (block size of four) was performed using a computer-generated randomisation list generated by CG. Four care home units were randomised at the same time (forming a wave), to ensure allocation concealment. All seven assessors were masked to treatment allocation until all follow-up assessments were completed and masking success was verified by asking assessors whether they inadvertently discovered the unit's allocation.[11] Masked assessors had to rely partly on information from proxy respondents (care staff). After assignment, interventionists, residents, and care staff could no longer be masked.

Interventions

Care home units were cluster-randomised to one of four conditions: only GMT, only RCS, both GMT and RCS, or standard care (neither GMT nor RCS). Note that this design leads to two groups with GMT (GMT and GMT plus RCS), which were joined as GMT in the analyses, and two groups with RCS (RCS and GMT plus RCS). For the single-intervention groups, 45-min sessions of GMT or RCS were planned for implementation twice per week for 3 months and thereafter once per week for 3 months (39 sessions over 26 weeks). For the combined GMT and RCS group, twice per week GMT and twice per week RCS were planned for 3 months and thereafter once per week for each intervention for 3 months (78 sessions over 26 weeks). Interventions were planned to start immediately after randomisation; in cases where this was not possible for organisational reasons, the 6-month intervention period could be extended beyond the

6-month assessment.[15] Standardised treatment protocols, training manuals, and fidelity checklists were developed for each intervention.[14] Interventionists received training and ongoing supervision from music therapists with more than 20 years of experience.

GMT, facilitated by a credentialed music therapist, was designed as small, closed groups with consistent membership of eight to ten participants. Usually this meant that the residents were split into two GMT groups per unit; these were often based on existing communities within each unit (eg, corresponding with the location of residents' rooms or characteristics such as dementia stage, or both). Where possible, group formation was also informed by other considerations (eg, cultural background, music preferences, session time). Informed by the principles of personhood and person-centred care, GMT aimed to meet the psychosocial needs of each individual resident in the moment and included familiar song singing, music-stimulated reminiscence, improvising on percussion instruments, and spontaneous or directed movement to music.[11,14] RCS was designed for larger open groups of 15–20 participants (ie, one choir per unit) and was open to other residents, not participating in the trial, who wished to join. Facilitated by community musicians with previous experience in leading ensembles, RCS sessions were structured around song singing, using familiar and preferred repertoire with lyrics displayed on a screen.[14] Brief physical and vocal warm-ups and learning new musical material were also incorporated into sessions based on the needs and abilities of the group. RCS aimed to foster connectedness, emotional wellbeing, and enjoyment of group music-making.[11,14] Participants in clusters randomised to standard care did not receive music interventions but continued to engage in the leisure programmes provided by the RAC facilities, such as group games (word games, bingo, or trivia), entertainment (concerts or movies), classes (arts and crafts, or exercise), gardening, and outings; these programmes were also open to intervention participants. Participation in such activities was recorded. For more detailed descriptions of the interventions, including proposed mechanisms and process-outcome relations, refer to the study protocol and intervention fidelity protocol.[11,14]

Outcomes

All outcomes were assessed at baseline, and at months 3, 6, and 12. Outcome measures were selected based on core outcomes for residents and have good reliability and validity.[11]

For more on **core outcome** see https://www.comet-initiative.org/

The primary endpoint was depressive symptoms score at 6 months, measured using the MADRS.[16] This was based on MADRS being used successfully in previous studies of music interventions and in people living with dementia, and its sensitivity to change compared with alternative scales.[11] The choice of 6 months as the primary endpoint was based on the life expectancy of residents. We were interested in long-term effects and therefore aimed to measure the longest meaningfully measurable time frame. We expected attrition due to death to be higher at 12 months compared with 6 months.[11] The MADRS is a reliable and valid 10-item scale with items rated from 0 (no depression) to 6 (severe depression) based on observed signs and symptoms in the week earlier.[16] Total score ranges from 0 to 60, and higher scores indicate higher severity of depressive symptoms. MADRS was scored by an external masked assessor but this scoring relied partly on information from care staff.

Residents' secondary outcomes included neuropsychiatric symptoms and quality of life from unmasked proxy (care staff) or self-reports. Neuropsychiatric symptoms were assessed by proxy using the Neuropsychiatric Inventory Questionnaire (NPI-Q),[17] a 12-item measure assessing severity of symptoms (from mild [1] to severe [3]) and caregiver distress (not distressing [0] to extreme or very severe distress [5]). Quality of life measures included the disease-specific Quality of Life in Alzheimer's Disease (QoL-AD; self-rated),[18] and the generic EuroQol (EQ-5D-5L; proxy).[19] The QoL-AD is a 13-item scale assessing aspects of quality of life from physical health to memory and ability to do things for fun. The QoL-AD has self-rated and proxy versions, and scores range from 13 to 52, with higher scores indicating better quality of life.[18] The EQ-5D-5L was included as an additional quality of life measure to derive the quality-adjusted life-years. The EQ-5D-5L consists of two parts: the EQ-5D-5L descriptive system and the EQ Visual Analogue Scale (EQ-VAS). The EQ-5D-5L descriptive system assesses health in five dimensions (mobility, self-care, usual activities, pain or discomfort, and anxiety or depression) and five levels from no problems (1) to extreme problems (5). The EQ-5D-5L was adjusted by societal weights using the Australian crosswalk value set estimated by the authors. The EQ-VAS indicates overall perceived health today on a vertical visual analogue scale ranging from the worst (0) to the best imaginable health (100).[19] Additional measures for residents assessed at baseline included the CDR[20] and MMSE,[21] administered and scored by masked assessors (neuropsychologists).

The effect of the interventions on staff burden was assessed at the four timepoints using the 10-item, self-assessment Professional Care Team Burden Scale (PCTB).[22] The PCTB scores range from 0 to 40 with higher scores indicative of higher burden. The PCTB has shown good validity and high internal consistency and reliability.[22]

Additional outcomes described in the protocol for the full trial[11] were not analysed here as the study was not sufficiently powered to assess these outcomes due to the limited sample size; results of cost-effectiveness analyses will be published separately. Exploratory post-hoc outcomes included the Agitation/aggression, Mood, and Frontal subscales of the NPI-Q[23] and the Objective Burden subscale of the PCTB.[22]

Responses to the music interventions and any adverse events during sessions were reported in

session notes by facilitators. Deaths and acute hospitalisations were reported by the RAC facility and recorded by assessors.

Statistical analysis

While the international trial aims to recruit 1000 participants with a view to detecting small effect sizes and

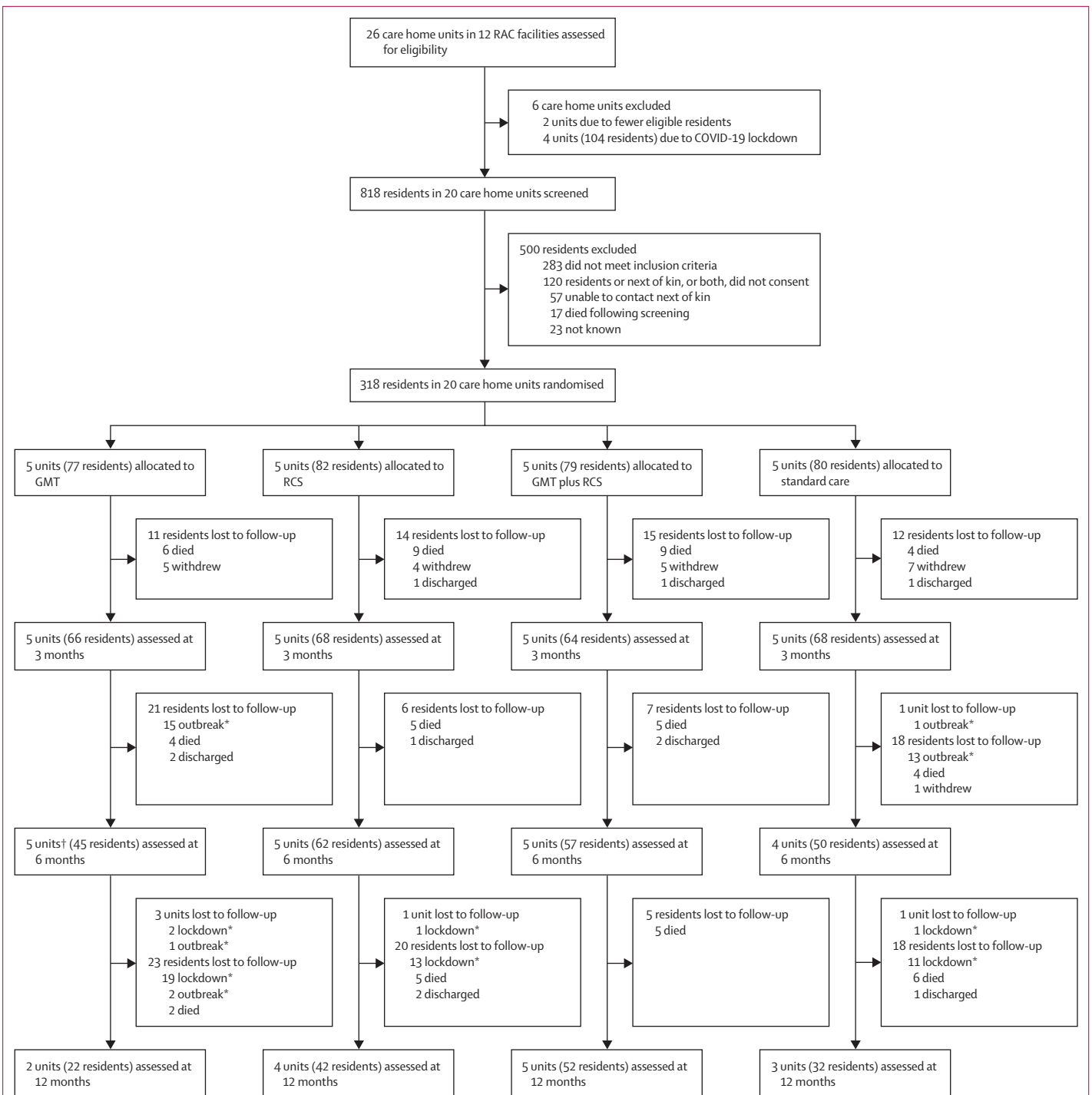

Figure 1: **Recruitment and follow-up of care home units and residents**

Care home unit (cluster). Recruitment and follow-up of staff is reported in a separate flow diagram in the appendix (p 9). RAC=residential aged care. GMT=group music therapy. RCS=recreational choir singing. *Due to COVID-19; ie, lockdown, RAC facility in lockdown due to COVID-19 safety measures; outbreak, COVID-19 outbreak at the RAC facility. †Only two residents assessed at one care home unit due to COVID-19 outbreak during 6-month follow-up.

subgroup effects,[11] we determined that the Australian trial had sufficient power to justify a separate analysis. With 214 residents across 20 care home units at the primary endpoint and an intraclass correlation coefficient of 0·09 as observed, the trial had 80% power to detect a medium-to-large effect size at a two-sided significance level of 0·025 (ie, Bonferroni-corrected from 0·05 for the two main comparisons; Cohen's d=0·65), corresponding to 5 MADRS points with SD 7·6. The power calculations reported here (for the Australian cohort) were post hoc.

Participants' sociodemographic and clinical characteristics at baseline were summarised by study group and across the whole sample, using descriptive methods.

	Total	GMT	RCS	GMT plus RCS	Standard care
Care home unit	20	5	5	5	5
Number of beds per unit*	54·7 (11·1)	56·0 (13·6)	58·4 (12·3)	49·8 (9·6)	54·4 (10·5)
Number of full-time equivalent care staff per unit*†	28·1 (5·4)	30·1 (7·8)	28·3 (6·7)	24·5 (4·8)	29·5 (2·6)
Residents enrolled per unit*	15·9 (3·5)	15·2 (5·3)	16·4 (3·4)	16·0 (3·2)	16·0 (2·9)
Staff enrolled per unit*	6·6 (3·8)	7·0 (3·8)	5·2 (3·0)	7·4 (5·7)	6·6 (3·0)
Residents enrolled	318	77	82	79	80
Sex					
Female‡	219 (69%)	53 (69%)	59 (72%)	52 (66%)	55 (69%)
Male	99 (31%)	24 (31%)	23 (28%)	27 (34%)	25 (31%)
Age, years*	86·5 (7·2)	86·0 (7·5)	87·1 (7·0)	85·8 (7·9)	87·2 (6·5)
Marital status‡					
Single or unmarried	20/313 (6%)	4/77 (5%)	3/80 (4%)	9/77 (12%)	4/79 (5%)
Married	99/313 (32%)	20/77 (26%)	26/80 (33%)	30/77 (39%)	23/79 (29%)
Separated or divorced	25/313 (9%)	5/77 (7%)	6/80 (8%)	9/77 (12%)	5/79 (6%)
Widow or widower	167/313 (53%)	47/77 (61%)	45/80 (56%)	28/77 (36%)	47/79 (60%)
Not known	2/313 (1%)	1/77 (1%)	0/80	1/77 (1%)	0/79
Country of birth‡					
Australia	189/313 (60%)	44/76 (58%)	48/81 (59%)	48/78 (62%)	49/78 (63%)
UK	32/313 (10%)	11/76 (15%)	7/81 (9%)	5/78 (6%)	9/78 (12%)
European countries	62/313 (20%)	15/76 (20%)	21/81 (26%)	15/78 (19%)	11/78 (14%)
Australasian countries	14/313 (5%)	4/76 (5%)	4/81 (5%)	2/78 (3%)	4/78 (5%)
Other countries	15/313 (5%)	2/76 (3%)	1/81 (1%)	7/78 (9%)	5/78 (6%)
Not known	1/313 (<1%)	0/76	0/81	1/78 (1%)	0/78
First language‡					
English	233/307 (76%)	57/76 (75%)	61/79 (77%)	55/75 (73%)	60/77 (78%)
Other language (but having good knowledge of English)	48/307 (16%)	11/76 (15%)	11/79 (14%)	15/75 (20%)	11/77 (14%)
Other language (and having poor or no knowledge of English)	26/307 (9%)	8/76 (11%)	7/79 (9%)	5/75 (7%)	6/77 (8%)
Highest level of education completed‡					
Primary education or less	66/312 (21%)	9/77 (12%)	22/81 (27%)	17/75 (23%)	18/79 (23%)
Secondary education	167/312 (54%)	47/77 (61%)	41/81 (51%)	43/75 (57%)	36/79 (46%)
Tertiary or further education	38/312 (12%)	7/77 (9%)	10/81 (12%)	10/75 (13%)	11/79 (14%)
Not known	41/312 (13%)	14/77 (18%)	8/81 (10%)	5/75 (7%)	14/79 (18%)
Clinical diagnosis of dementia‡					
Alzheimer's disease	104/318 (33%)	32/77 (42%)	22/82 (27%)	26/79 (33%)	24/80 (30%)
Other dementia types	57/318 (18%)	10/77 (13%)	16/82 (20%)	15/79 (19%)	16/80 (20%)
Unspecified dementia	157/318 (49%)	35/77 (45%)	44/82 (54%)	38/79 (48%)	40/80 (50%)
Severity of dementia‡					
Very mild or mild (CDR 0·5 or 1)	64/317 (20%)	16/77 (21%)	14/81 (17%)	12/79 (15%)	22/80 (28%)
Moderate (CDR 2)	113/317 (36%)	26/77 (34%)	27/81 (33%)	29/79 (37%)	31/80 (39%)
Severe (CDR 3)	140/317 (44%)	35/77 (45%)	40/81 (49%)	38/79 (48%)	27/80 (34%)
MMSE score*§	8·0 (7·6)	7·3 (7·7)	8·2 (7·6)	7·3 (7·1)	9·2 (8·1)
MADRS score*	18·3 (7·6)	17·6 (8·1)	19·0 (7·5)	19·9 (7·2)	16·7 (7·0)

(Table 1 continues on next page)

	Total	GMT	RCS	GMT plus RCS	Standard care
(Continued from previous page)					
Staff members enrolled	131	36	26	37	32
Sex					
Female‡	108/131 (82%)	25/36 (69%)	25/26 (96%)	33/37 (89%)	25/32 (78%)
Male	23/131 (18%)	11/36 (31%)	1/26 (4%)	4/37 (11%)	7/32 (22%)
Age, years*¶	43·3 (11·6)	43·3 (13·5)	47·3 (13·1)	42·1 (7·3)	42·0 (11·7)
Staff category‡					
Registered nurse	6/131 (5%)	1/36 (3%)	1/26 (4%)	1/37 (3%)	3/32 (9%)
Enrolled nurse	3/131 (2%)	0/36	0/26	2/37 (5%)	1/32 (3%)
Personal care attendant	114/131 (87%)	31/36 (86%)	25/26 (96%)	32/37 (87%)	26/32 (81%)
Leisure staff	8/131 (6%)	4 (11%)	0/26	2/37 (5%)	2/32 (6%)
Work experience, years*¶	8·0 (4·8%)	7·5 (4·3%)	9·4 (6·8%)	7·2 (3·2%)	8·3 (5·4%)

Data are n, n/N (%), or mean (SD). Clinical diagnosis of dementia categories include Alzheimer's disease; other dementia types include vascular dementia, frontotemporal dementia, dementia with Lewy bodies, dementia in other diseases classified elsewhere, mixed dementia; and unspecified dementia includes unspecified dementia and other (eg, noted cognitive impairment). GMT=group music therapy. RCS=recreational choir singing. CDR=Clinical Dementia Rating. MMSE=Mini-Mental State Examination. MADRS=Montgomery-Åsberg Depression Rating Scale. *Kruskal-Wallis Test. †Only 12 units provided information about number of staff. ‡Pearson Chi-Square test. §MMSE was available in 316 residents. ¶Only 46 staff members disclosed their age and 103 their work experience in years.

Table 1: **Baseline characteristics**

The main comparisons, GMT versus no-GMT and RCS versus no-RCS,[11] were chosen to maximise the statistical power by fully exploiting the factorial design: the comparison of GMT versus no-GMT includes data from the two groups receiving GMT (GMT and GMT plus RCS) versus the two groups receiving only RCS or standard care; whereas the comparison of RCS versus no RCS includes data from the two groups receiving RCS (RCS and GMT plus RCS) versus the two groups receiving only GMT or standard care. We used an intention-to-treat approach with a longitudinal linear mixed-effects model for the outcome at all time points depending on time. We used GMT versus no GMT; RCS versus no RCS; as well as all two-way and three-way interactions, with a nested random intercept for individual and unit (model formula, madrs~gmt*rcs*time, random=~1|id|unit, where madrs contains the data of all timepoints, gmt and rcs are factors with two levels, and time is a factor with four levels; analogous for other outcomes).[13] In the time domain we used simple contrasts leading to an effect estimate for both treatments, as well as their interactions for each follow-up timepoint. The normality assumption for continuous outcomes was checked by Q-Q plots of residuals. The general significance level was set to 0·05; for the primary analyses, where we had two comparisons (GMT *vs* no GMT, RCS *vs* no RCS), we therefore used a marginal Bonferroni level of 0·025. Computations were performed using R, version 4.1.0, with package nlme 3.1. Graphics were produced with Matlab 2021a.

Prespecified subgroup analyses for the primary outcome were performed for people with moderate to severe dementia (CDR ≥2) and for those with adequate attendance (≥50%);[11] additional exploratory subgroup analyses were conducted for the intention-to-treat sample, excluding the last wave, which was affected by COVID-19 lockdown.

Sessions delivered per care home unit and sessions attended per resident are presented by wave and by group for the whole intervention period, and at months 1–3; months 4–6; and after 6 months post-randomisation (appendix pp 3 and 4). The number and percentages of sessions adequately conducted according to the manual are also presented (appendix p 4). Randomly selected recordings of music intervention sessions were assessed for fidelity by experienced music therapy clinicians. Interventions were considered adequately conducted if at least 11 of 14 (GMT) or at least seven of nine (RCS) mandatory items on the fidelity checklists were completed.

In Australia, the COVID-19 pandemic impacted the study. Lockdowns of RAC from March, 2020, until the end of 2020 restricted the assessors' and interventionists' access to RAC sites, thereby disrupting recruitment, implementation of the interventions, and data collection. This disruption resulted in reduced sample size and data missing not at random. We interpreted these as extenuating circumstances and implemented mitigating strategies, including modified assessment procedures and changes to the analysis plan to reduce the consequences of these changes on sample size. Due to COVID-19 lockdowns, only the 3-month endpoint was valid for all 20 clusters for analyses. 19 (95%) clusters completed the 6-month (primary) endpoint; however, one of the clusters had only two residents assessed. At the 12-month endpoint, only 14 (70%) of the 20 clusters had assessment completed. 104 participants screened were eligible for the last wave (care home units 21–24), but baseline assessments were discontinued with the extended COVID-19 lockdown.[15]

This trial is registered with ClinicalTrials.gov, NCT03496675, and anzctr.org.au, ACTRN12618000156280.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Between June 15, 2018, and Feb 18, 2020, we invited six RAC providers to participate in the study, of which four agreed. From these, two RAC providers with different

	Individual groups				Factorial comparisons		
	GMT	RCS	GMT plus RCS	Standard care	GMT	RCS	RCS × GMT
Intention-to-treat sample							
MADRS, baseline	77, 17·6 (8·1)	82, 19·0 (7·5)	79, 19·9 (7·2)	80, 16·7 (7·0)			
3 months	66, 13·0 (8·3)	68, 12·5 (8·2)	64, 13·2 (7·0)	68, 14·4 (7·1)	−1·98 (−5·42 to 1·45), 0·2578	−4·14 (−7·54 to −0·75), 0·0170	2·07 (−2·79 to 6·92), 0·4040
6 months (primary endpoint)	45, 12·0 (9·7)	62, 11·4 (8·1)	57, 12·4 (6·7)	50, 13·7 (7·8)	−0·44 (−4·32 to 3·43), 0·8224	−4·25 (−7·89 to −0·62), 0·0221	0·86 (−4·38 to 6·11), 0·7479
12 months	22, 9·6 (8·3)	42, 11·1 (8·7)	52, 11·3 (9·1)	32, 13·3 (9·0)	−2·51 (−7·39 to 2·37), 0·3140	−5·48 (−9·7 to −1·27), 0·0109	2·98 (−3·23 to 9·19), 0·3472
NPI-Q severity, baseline	77, 11·3 (7·8)	82, 11·4 (7·1)	79, 11·6 (6·4)	80, 8·9 (5·5)			
3 months	66, 8·2 (6·8)	68, 6·7 (5·2)	64, 8·0 (5·4)	68, 7·9 (5·3)	−1·85 (−4·55 to 0·85), 0·1799	−3·67 (−6·34 to −1), 0·0072	3·22 (−0·6 to 7·03), 0·0988
6 months	45, 6·6 (6·4)	62, 7·0 (6·5)	57, 7·4 (5·3)	50, 7·2 (5·2)	−1·32 (−4·37 to 1·73), 0·3956	−2·61 (−5·47 to 0·25), 0·0738	1·81 (−2·31 to 5·94), 0·3888
12 months	22, 6·6 (5·3)	42, 7·0 (6·5)	50, 7·5 (5·9)	32, 7·9 (6·5)	−1·27 (−5·11 to 2·57), 0·5181	−4·0 (−7·32 to −0·69), 0·0181	2·42 (−2·47 to 7·32), 0·3321
NPI-Q caregiver distress, baseline	77, 13·5 (12·7)	82, 12·4 (11·8)	79, 13·0 (9·9)	80, 9·3 (8·8)			
3 months	66, 7·4 (8·3)	68, 6·6 (8·0)	64, 9·0 (8·4)	68, 8·4 (7·8)	−4·86 (−8·84 to −0·88), 0·0169	−4·87 (−8·8 to −0·94), 0·0154	6·91 (1·29 to 12·54), 0·0162
6 months	45, 5·5 (6·8)	62, 7·5 (8·8)	57, 6·4 (5·8)	50, 7·2 (9·1)	−3·44 (−7·93 to 1·06), 0·1342	−2·39 (−6·6 to 1·83), 0·2670	1·98 (−4·1 to 8·06), 0·5235
12 months	22, 7·0 (7·0)	42, 6·8 (8·4)	50, 7·6 (7·9)	32, 9·4 (8·4)	−3·26 (−8·92 to 2·4), 0·2589	−5·24 (−10·12 to −0·35), 0·0359	4·01 (−3·21 to 11·22), 0·2767
EQ-5D-5L index (AUS Crosswalk), baseline	76, 0·36 (0·32)	81, 0·37 (0·30)	79, 0·32 (0·28)	79, 0·38 (0·29)			
3 months	66, 0·42 (0·30)	68, 0·49 (0·27)	64, 0·40 (0·31)	68, 0·43 (0·31)	0·01 (−0·12 to 0·14), 0·8976	0·07 (−0·06 to 0·2), 0·2656	−0·07 (−0·25 to 0·12), 0·4911
6 months	45, 0·43 (0·28)	62, 0·46 (0·29)	57, 0·40 (0·28)	50, 0·43 (0·27)	−0·03 (−0·17 to 0·12), 0·7309	0·02 (−0·11 to 0·16), 0·7315	0·01 (−0·19 to 0·21), 0·9373
12 months	22, 0·46 (0·29)	42, 0·42 (0·26)	52, 0·35 (0·25)	32, 0·41 (0·32)	0·01 (−0·18 to 0·19), 0·9383	0·01 (−0·15 to 0·17), 0·8890	−0·04 (−0·28 to 0·2), 0·7363
EQ-5D-VAS, baseline	76, 58·4 (18·4)	82, 53·3 (19·7)	78, 51·0 (15·3)	80, 57·3 (18·9)			
3 months	66, 59·8 (17·8)	68, 64·9 (18·0)	64, 60·7 (21·9)	68, 56·1 (16·8)	1·77 (−6·15 to 9·69), 0·6608	12·18 (4·36 to 20·01), 0·0023	−3·49 (−14·69 to 7·71), 0·5413
6 months	45, 63·2 (15·4)	62, 65·4 (20·0)	57, 57·6 (17·1)	50, 60·7 (18·6)	−0·89 (−9·82 to 8·04), 0·8452	9·27 (0·89 to 17·65), 0·0305	−4·86 (−16·95 to 7·23), 0·4310
12 months	22, 59·1 (16·5)	41, 65·0 (14·9)	52, 57·7 (15·3)	32, 66·1 (16·7)	−10·24 (−21·48 to 1), 0·0746	6·68 (−3·06 to 16·42), 0·1793	2·89 (−11·45 to 17·22), 0·6932
QoL-AD score, baseline	48, 26·9 (7·1)	53, 28·1 (6·9)	47, 26·6 (6·6)	58, 26·7 (7·9)			
3 months	39, 27·1 (6·4)	45, 27·9(6·4)	38, 26·2 (5·7)	48, 26·9 (7·5)	0·83 (−5·18 to 6·84), 0·7868	0·3 (−5·65 to 6·26), 0·9207	0·3 (−5·65 to 6·26), 0·8269
6 months	26, 26·1 (6·5)	37, 27·4(7·2)	22, 27·0(5·2)	31, 27·0 (7·0)	−2·07 (−8·9 to 4·75), 0·5514	0·65 (−6 to 7·29), 0·8490	0·65 (−6 to 7·29), 0·6426
12 months	10, 24·1 (7·2)	17, 26·8 (5·8)	10, 24·4 (8·2)	13, 28·2 (5·4)	−2·55 (−11·8 to 6·71), 0·5897	−5·89 (−14·29 to 2·51), 0·1695	−5·89 (−14·29 to 2·51), 0·5015
PCTB scale, baseline	36, 8·8 (3·8)	26, 8·1 (4·2)	36, 9·3 (3·7)	32, 10·3 (3·6)			
3 months	25, 9·9 (3·7)	16, 8·9 (5·2)	19, 8·3 (4·3)	23, 8 (4·7)	3·36 (0·27 to 6·46), 0·0341	3·01 (−0·43 to 6·45), 0·0873	−5·11 (−9·78 to −0·44), 0·0327
6 months	20, 9·4 (3·4)	18, 10·6 (5·0)	14, 9·1 (5·6)	12, 7·9 (4·5)	2·93 (−0·67 to 6·53), 0·1118	4·83 (1·08 to 8·58), 0·0121	−5·53 (−10·62 to −0·44), 0·0342
12 months	11, 9·5 (2·5)	11, 8·5 (3·9)	14, 10·6 (4·8)	8, 9·5 (4·8)	1·49 (−2·81 to 5·79), 0·4978	1·09 (−3·3 to 5·47), 0·6267	−0·51 (−6·33 to 5·31), 0·8644

(Table 2 continues on next page)

	Individual groups				Factorial comparisons		
	GMT	RCS	GMT plus RCS	Standard care	GMT	RCS	RCS × GMT
(Continued from previous page)							
Subsample with CDR ≥2							
MADRS, baseline	61, 18·8 (8·4)	67, 20 (7·7)	67, 21·2 (7)	58, 17·8 (7·3)			
3 months	50, 12·9 (8·6)	53, 12·7 (7·4)	54, 14 (7)	48, 16·6 (6·8)	−4·38 (−8·33 to −0·43), 0·0301	−5·99 (−9·87 to −2·11), 0·0025	4·61 (−0·83 to 10·06), 0·0973
6 months	32, 12·2 (8·1)	49, 11·4 (7·6)	47, 13·4 (6·5)	38, 15·6 (7·5)	−2·03 (−6·49 to 2·43), 0·3733	−6·2 (−10·29 to −2·11), 0·0031	3·1 (−2·8 to 9·01), 0·3038
12 months	18, 8·3 (7·1)	30, 12·9 (9·2)	43, 12 (9·1)	23, 15·5 (8·8)	−5·29 (−10·79 to 0·22), 0·0602	−5·79 (−10·67 to −0·91), 0·0203	4·51 (−2·47 to 11·5), 0·2058
Subsample with ≥50% attendance*							
MADRS, baseline	27, 15·8 (7·2)	39, 18·4 (7·2)	25, 19·2 (6·7)	80, 16·6 (7)			
3 months	27, 10·3 (6·8)	39, 10·9 (8·6)	25, 11·8 (6)	68, 14·4 (7·1)	−3·06 (−7·52 to 1·39), 0·1783	−5·12 (−9·04 to −1·2), 0·0108	3·24 (−3·51 to 9·99), 0·3470
6 months	24, 11 (6·7)	37, 10·8 (8)	25, 12·4 (5·8)	50, 13·7 (7·8)	−1·32 (−6·02 to 3·37), 0·5807	−4·29 (−8·4 to −0·18), 0·0415	2·41 (−4·52 to 9·34), 0·4954
12 months	17, 7·5 (6·5)	21, 11·7 (9·8)	22, 9·1 (7·5)	32, 13·3 (9)	−4·78 (−10·12 to 0·55), 0·0794	−4·7 (−9·6 to 0·21), 0·0609	2·87 (−4·89 to 10·63), 0·4692
Intention-to-treat sample excluding wave 5 (COVID-19 lockdown)†							
MADRS, baseline	55, 15·1 (6·1)	65, 19·2 (7·7)	68, 20·6 (7·1)	65, 17·3 (7·2)			
3 months	49, 11·4 (7·1)	55, 11·9 (8·2)	53, 13·8 (7)	55, 14·7 (7·3)	−0·93 (−4·84 to 2·99), 0·6423	−4·53 (−8·31 to −0·75), 0·0189	1·54 (−3·9 to 6·98), 0·5795
6 months	43, 11·3 (8·3)	50, 11·1 (8·1)	50, 12·8 (6·9)	50, 13·7 (7·8)	0·06 (−3·98 to 4·1), 0·9768	−4·53 (−8·4 to −0·65), 0·0225	0·6 (−4·99 to 6·19), 0·8336
12 months	22, 9·6 (8·3)	31, 11·1 (9·4)	45, 11·6 (9·1)	32, 13·3 (9)	−1·67 (−6·63 to 3·3), 0·5110	−5·51 (−10·06 to −0·96), 0·0178	2·03 (−4·53 to 8·59), 0·5444

Data are n, mean (SD); or, effect (95% CI), p value. Effects estimates based on linear mixed-effects models. GMT=group music therapy. RCS=recreational choir singing. MADRS=Montgomery-Åsberg Depression Rating Scale. NPI-Q=Neuropsychiatric Inventory Questionnaire. EQ-5D-5L=EuroQoL 5 levels adjusted by societal weights using the Australian crosswalk value set estimated by the authors. EQ-5D-VAS=EuroQoL Visual Analogue Scale. QoL-AD=Quality of Life in Alzheimer's Disease. PCTB=Professional Care Team Burden Scale. *The per-protocol sample with at least 50% attendance included residents with at least 19 sessions if randomly assigned to group 1 or 2; with 39 sessions or more if randomly assigned to group 3; and all residents in group 4. †Wave 5 was affected by COVID-19 lockdown at the time of the primary endpoint.

Table 2: **Observed outcomes and effects estimates from linear mixed-effects models**

RAC facilities participated, one did not, and one commenced recruitment and baseline assessment but COVID-19 prevented enrolment and randomisation. Therefore, 12 RAC facilities with 26 eligible care home units were invited and, excluding six units who could not be enrolled due to COVID-19 lockdowns, 818 residents and 155 staff members were screened. Between July 18 and Nov 26, 2019, 20 care home units with 318 residents and 131 staff members were randomly assigned in five waves (figure 1; appendix p 9). A mean of 15·9 residents (SD 3·5) and 6·6 staff (SD 3·8) members were enrolled in each cluster (appendix p 1).

Most residents were female (219 [69%] of 318), and had a mean age of 86·5 years (SD 7·2), similar to the demographic profile of residents in Australian RAC reported in the literature.[24] Dementia types were mostly unspecified (157 [49%]). 140 (44%) residents were at a severe stage of dementia (CDR 3) at baseline, and cognitive impairment was severe for all residents (MMSE mean score 8·0, SD 7·6; table 1). Of the 131 staff members enrolled, most staff were female (108 [82%]), working as personal care assistants (114 [87%]), with a mean of 8 years (SD 4·8) experience working in RAC (table 1). Baseline characteristics were similar across groups (table 1). No care home units dropped out after randomisation; dropout of residents occurred due to death, withdrawal, discharge from the RAC facility, and COVID-19 lockdowns or outbreaks (figure 1). Residents who dropped out before the primary endpoint were similar on most variables to those who remained, but had more severe cognitive impairment (MMSE) and depressive symptoms (MADRS; appendix p 2). Assessor masking was successful, except for one case of accidental unmasking at 3 months; the assessor was subsequently replaced to ensure masking for the remaining assessments.

On average, residents attended 22·2 GMT sessions (SD 13·1) and 20·0 RCS sessions (SD 13·3; appendix p 4); people with mild dementia (CDR ≤1) attended fewer sessions than residents with moderate or severe dementia (appendix p 5). We noted that additional residents often attended RCS (due to a keen interest in music); they were likely to participate actively, which contributed to encouraging others to sing along and to the experience of the group as a choir. 14% of all sessions delivered were assessed for fidelity (77/574 GMT; 52/320 RCS). From these, 77 (100%) GMT and 49 (94%) RCS sessions were adequately conducted according to the manual (appendix p 4). Participation in other

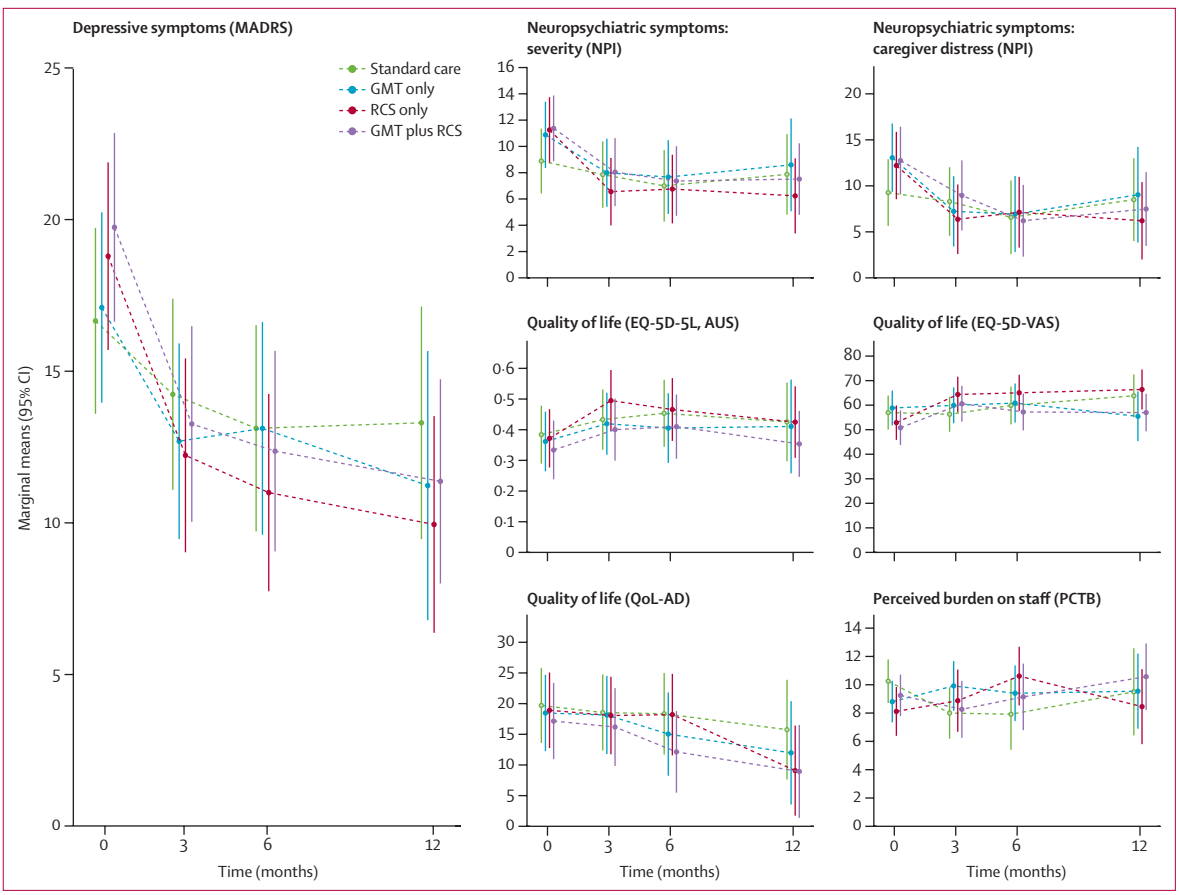

Figure 2: **Effects of music interventions**
Marginal means from linear mixed-effects models are shown. MADRS=Montgomery-Åsberg Depression Rating Scale.GMT=group music therapy. RCS=recreational choir singing. NPI=Neuropsychiatric Inventory. EQ-5D, AUS=EuroQol, Australian weights. EQ-VAS=EuroQol Visual Analogue Scale. QoL-AD=Quality of Life in Alzheimer's Disease. PCTB=Professional Care Team Burden Scale.

additional recreational activities at the unit was common and remained balanced across groups (appendix p 6).

The primary statistical analysis of effects showed that RCS was superior to no RCS at all timepoints (mean difference at 3 months –4·14, 95% CI –7·54 to –0·75, p=0·0170; 6 months –4·25, –7·89 to –0·62, p=0·0221; 12 months –5·48, –9·7 to –1·27, p=0·0109; table 2). GMT was not superior to no GMT (table 2). There was no interaction effect between GMT and RCS (table 2). A descriptive comparison of the four groups illustrated that RCS alone had the best outcomes and standard care had the worst outcomes throughout the follow-up period (figure 2). An exploratory subgroup analyses excluding wave 5 (affected by COVID-19 lockdown at the primary endpoint) reduced the rate of missing data and showed similar results to the main analysis (table 2), suggesting that the findings were robust.

Secondary outcomes are presented in table 2 and figure 2. For severity of neuropsychiatric symptoms, RCS tended to be superior to no RCS at all timepoints, reaching statistical significance at months 3 and 12; there were no effects of GMT versus no GMT. With respect to distress of neuropsychiatric symptoms for caregivers, GMT was superior to no GMT at 3 months, and RCS was superior to no RCS at months 3 and 12. No effects were found for generic or disease-specific quality of life, except for the EQ-5D-VAS for which RCS was superior to no RCS at 3 months (mean difference 12·18) and 6 months (mean difference 9·27). For perceived burden on staff there was an effect of GMT at 3 months (mean difference 3·36) and of RCS at 6 months (mean difference 4·83), showing that both interventions increased the perceived burden on staff. Results of exploratory post-hoc outcomes are shown in the appendix (p 7).

In the subset of residents with moderate to severe dementia (CDR ≥2), RCS was superior to no RCS at all timepoints; GMT was superior to no GMT at 3 months (table 2). For residents with adequate attendance (≥50%), effects for both interventions were larger than in the intention-to-treat sample but remained non-significant for GMT. There might be a complex three-way interaction between severity, attendance, and intervention (appendix p 10).

Unrelated serious adverse events during the first 6 months included deaths (ten cases in GMT, 14 in RCS, 14 in GMT plus RCS, and eight in standard care) and acute hospital admissions (seven cases in GMT, 18 in RCS, 14 in GMT plus RCS, and ten in standard care). Two related non-serious adverse events were reported by GMT therapists (appendix p 8).

Discussion

In this large 2×2 factorial cluster-randomised controlled trial of two active music interventions for care home residents with dementia and mild to severe depressive symptoms, RCS led to a clinically meaningful reduction in depressive symptoms. This effect was observed at the primary endpoint (6 months), as well as at an intermediate assessment (3 months) and long-term follow-up (12 months), and was robust in several subgroup analyses. In contrast to an earlier Cochrane review[6] of music-based interventions, the current study showed that RCS had a larger effect size at the end of intervention (Cohen's d –0·56 *vs* –0·27) and a lasting or even increasing effect beyond completion of the intervention (Cohen's d –0·82 *vs* –0·03; figure 3). RCS was also effective in reducing the severity of neuropsychiatric symptoms and associated caregiver distress at months 3 and 12. RCS increased quality of life at months 3 and 6; however, this was only found in one of three measures of quality of life. These findings corroborate previous smaller-scale research in showing that music can be a useful medium to improve mood, behavioural problems, and quality of life in older adults with dementia.[6]

The second music intervention, GMT, was offered in smaller, closed groups with a stronger focus on one-to-one interaction and a mixed use of singing, playing, and talking. For GMT, a small point difference with a wide CI means that we do not know if GMT was substantially better or worse than standard care for depression. Most secondary outcomes did not show clear benefits of GMT, with the exception of distress associated with neuropsychiatric symptoms for caregivers at 3 months. The finding that music interventions are effective for reducing distress of neuropsychiatric symptoms is in line with a previous study.[25] Both interventions increased perceived burden on staff, possibly due to the additional workload caused by implementing them. In a setting where burden is already high, this might outweigh any burden reduction caused by decreased neuropsychiatric symptoms.

Subgroup analyses suggested GMT to be more beneficial in later-stage dementia than in earlier stages, in line with a previous study that informed MIDDEL.[12] In residents with moderate to severe dementia, both GMT and RCS were beneficial in reducing depression at 3 months and RCS was also effective at 6 months; the difference between subgroups might be mediated by the number of sessions attended. In relation to previous research examining effects of music interventions on depressive symptoms in people with dementia more generally,[5–8] the present findings add more specific knowledge on types of interventions and severity of dementia (figure 3).

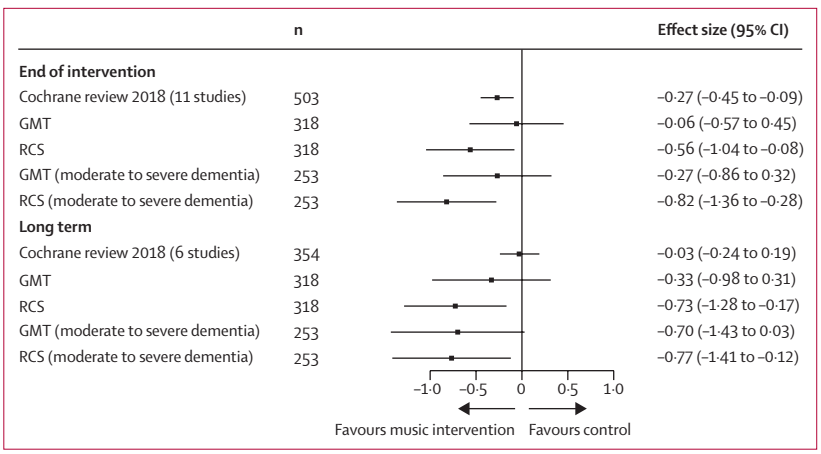

Figure 3: **Effects of music interventions on depressive symptoms (effect sizes from this study compared with earlier findings)**

Effect sizes (Cohen's d) from this study were calculated by dividing linear-mixed effects. Linear mixed-effects results by baseline SD. Effect sizes from previous findings are from Gold and colleagues.[11] End of intervention refers to 6 months in MIDDEL; varying time points in Cochrane. Long-term refers to 12 months in MIDDEL; varying time points in Cochrane. The Cochrane Review included various types of music interventions.[13] GMT=group music therapy. RCS=recreational choir singing.

Although the combined GMT and RCS group had twice the dose of interventions, compared with GMT or RCS alone, we did not see a further increase of effects. This absence of an interaction between GMT and RCS interaction is difficult to interpret, especially for secondary outcomes, and will require replication. The study was not designed to investigate dose–effect relationships or therapeutic mechanisms. The relation between interventions might be complex, given that GMT and RCS have partly different therapeutic intentions.[11,14]

Major strengths of this trial include the assessor-masked, randomised controlled design and high statistical power due to its large sample size and the factorial design. Furthermore, the care home units included in the trial were located across Melbourne, reflecting diverse socioeconomic and multicultural communities; from Melbourne's population of 5 million, 33·3% were born overseas, and 16·8% of immigrants were older than 65 years compared with 12·0% from the general population of Australia.[26] Interventionists were well trained, supervised by experienced clinicians, and adhered to intervention protocols. Our success in implementing the trial, despite the significant impact of COVID-19, was underpinned by strong collaborative relationships with management who encouraged staff involvement in the study.

The trial also has some important limitations. Although the Australian sample was sufficiently powered to detect medium-to-large effects, we cannot exclude possible smaller effects of GMT, or any effects in relevant subgroups that will only be meaningful to analyse with

the full international dataset. These will be addressed in the larger international trial. The decision to report the results of the Australian trial alone was based on the time lag between trial commencement in Australia and Europe (several years, due to timing of funding and COVID-19). The trial was also heavily impacted by the ongoing COVID-19 pandemic during the periods of enrolment, interventions, and follow-up, which had a negative impact on sample size, number of sessions delivered, attrition (figure 1), staff stress, and might have impacted resident depression levels. Assessors reported that care home staff had few opportunities to engage in more specific training and had difficulty understanding terminology when asked to report on resident behaviour, which might have impacted the reliability of the data. Furthermore, many residents were too advanced in their disease to be able to complete self-report questions. Therefore, one-third of quality of life (QoL-AD) data was sourced from care staff; care staff are known to often underestimate quality of life compared with residents' self-report.[27] The experience from and analysis of the Australian cohort afforded important learnings about trial implementation, including intervention delivery and organisation of assessments, which will benefit the European cohort.[15]

Based on the components of RCS, the findings suggest that group singing is particularly effective. An important component of RCS might be behavioural activation, which is effective both as a component of therapy and as a standalone intervention.[28,29] Behavioural activation aims to increase pleasant activities and positive interactions with the social environment. Therefore, the reduction in depressive symptoms might be tied to an experience of fun and lower cognitive load on the participant within RCS, compared with GMT. Furthermore, RCS can be conducted in large groups and is easily scalable. This scalability has important implications for practice, as the reach of RCS is not limited by the availability of intervention providers with extensive and specialised training, and it can be implemented for large numbers of residents at relatively low cost. In contrast to a previous study,[12] the RCS intervention in this trial was designed by music therapists and delivered by community musicians, who were supervised by credentialed music therapists to assist them in considering repertoire decisions, understanding positive and negative responses to music, and developing skills in empathy and person-centred care with people with dementia. The clinical implication is that the scalable RCS approach is valuable, even when considering the costs associated with regular supervision by a credentialed music therapist. The suggestion that effects of both music interventions were larger in residents with more severe dementia warrants further research.

Music interventions are being developed continuously both outside and within trials. The intervention protocols in this trial included opportunities for participant influence (ie, surveying participants' musical preferences, adapting to participant responses), supporting the notion of designing interventions based on participants' needs.[11] Although this might be true for many complex interventions and person-centred care in general, it does open pathways for continuous development, based on feedback from stakeholders ranging from residents to nurses and managers, in line with the person-based approach to intervention development.[30]

In conclusion, MIDDEL showed that RCS was beneficial for older care home residents with dementia and depressive symptoms. Group singing led to clinically important improvements in depression, as well as neuropsychiatric symptoms and generic quality of life. No clear evidence was found for the more specialised intervention, GMT. For care staff, both interventions might reduce distress associated with neuropsychiatric symptoms but were shown to increase perceived burden. It remains to be established in further work how GMT affects depression and other outcomes in RAC residents with dementia.

Contributors

CG initiated the study. FAB, Y-ECL, MG, and JA contributed to development of concept design. FAB obtained funding and led the implementation of the trial. FAB, JT, and Y-ECL developed fidelity measures, intervention manuals, and trained interventionists. Y-ECL supervised assessors. JT and FAB supervised interventionists. Y-ECL and PAS-S liaised with industry partners. FAB, Y-ECL, and PAS-S were involved in setting up the study conduct in each site. PAS-S conducted the GMT sessions. FAB, JT, Y-ECL and PAS-S completed fidelity checks. CG, MG, JDW, JA, and VS were responsible for data management activities. CG and JA conducted the main statistical analysis. CG, JA, and TVS verified all underlying data. All authors had access to the data, and contributed to interpreting the data, drafting the manuscript, and approving the final version of the manuscript. All authors had final responsibility for the decision to submit for publication.

Declaration of interests

All authors declare no competing interests.

Data sharing

De-identified datasets (participant codes and outcome scores) generated during or analysed, or both, during the MIDDEL trial will be stored in a publicly available repository (NSD, Norwegian Centre for Research Data, https://www.nsd.no/en/).

Acknowledgments

We acknowledge our industry partners Bupa, BlueCross, and Royal Freemasons' Homes whose commitment to the project has enabled us to deliver this clinical trial. In particular, we would like to thank all general managers, research site coordinators, lifestyle and personal care staff, and residents and families at the residential aged care facilities where this research took place. We also thank Imogen Clark for her involvement in the trial during its planning and early stages of implementation, Charly Harrison for his involvement as a recreational choir singing facilitator, as well as David Berlowitz, David Story, Nicola Lautenschlager, Cassandra Szoeke, Justine Schneider, and Martin Orrell for providing us with feedback on earlier drafts of the manuscript. This work was supported by the Australian National Health and Medical Research Council (grant number APP1137853) and by intramural support from NORCE Norwegian Research Centre. The study has been undertaken independently of any input from funding bodies. Funders have had no role in the study to date.

References

1 Livingston G, Sommerlad A, Orgeta V, et al. Dementia prevention, intervention, and care. *Lancet* 2017; **390:** 2673–734.

2 AIHW. Dementia in Australia. Australian Institute of Health and Welfare. 2021. https://www.aihw.gov.au/reports/dementia/dementia-in-aus (accessed Dec 20, 2021).

3 Wetzels RB, Zuidema SU, de Jonghe JF, Verhey FR, Koopmans RT. Determinants of quality of life in nursing home residents with dementia. *Dement Geriatr Cogn Disord* 2010; **29:** 189–97.

4 Livingston G, Kelly L, Lewis-Holmes E, et al. A systematic review of the clinical effectiveness and cost-effectiveness of sensory, psychological and behavioural interventions for managing agitation in older adults with dementia. *Health Technol Assess* 2014; **18:** 1–226.

5 Ray KD, Götell E. The use of music and music therapy in ameliorating depression symptoms and improving well-being in nursing home residents with dementia. *Front Med* 2018; **5:** 287.

6 van der Steen JT, Smaling HJA, van der Wouden JC, Bruinsma MS, Scholten RJPM, Vink AC. Music-based therapeutic interventions for people with dementia. *Cochrane Database Syst Rev* 2018; **7:** CD003477.

7 Liu M-N, Liou Y-J, Wang W-C, et al. Group music interventions using percussion instruments to reduce anxiety among elderly male veterans with Alzheimer's disease. *Med Sci Monit* 2021; **27:** e928714.

8 Pongan E, Tillmann B, Leveque Y, 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.

9 Soufineyestani M, Khan A, Sufineyestani M. Impacts of music intervention on dementia: a review using meta-narrative method and agenda for future research. *Neurol Int* 2021; **13:** 1–17.

10 Jacobsen JH, Stelzer J, Fritz TH, Chételat G, La Joie R, Turner R. Why musical memory can be preserved in advanced Alzheimer's disease. *Brain* 2015; **138:** 2438–50.

11 Gold C, Eickholt J, Assmus J, et al. Music Interventions for Dementia and Depression in ELderly care (MIDDEL): protocol and statistical analysis plan for a multinational cluster-randomised trial. *BMJ Open* 2019; **9:** e023436.

12 Werner J, Wosch T, Gold C. Effectiveness of group music therapy versus recreational group singing for depressive symptoms of elderly nursing home residents: pragmatic trial. *Aging Ment Health* 2017; **21:** 147–55.

13 Cook AJ, Delong E, Murray DM, Vollmer WM, Heagerty PJ. Statistical lessons learned for designing cluster randomized pragmatic clinical trials from the NIH Health Care Systems Collaboratory Biostatistics and Design Core. *Clin Trials* 2016; **13:** 504–12.

14 Baker FA, Tamplin J, Clark IN, Lee YC, Geretsegger M, Gold C. Treatment fidelity in a music therapy multi-site cluster randomized controlled trial for people living with dementia: the MIDDEL project intervention fidelity protocol. *J Music Ther* 2019; **56:** 125–48.

15 Baker FA, Stretton-Smith PA, Sousa TV, et al. Resource assessment in trials undertaken in residential care homes: experiences from the Australian MIDDEL cluster randomised controlled trial research team. *Contemp Clin Trials Commun* 2020; **20:** 100675.

16 Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979; **134:** 382–89.

17 Kaufer DI, Cummings JL, Ketchel P, et al. Validation of the NPI-Q, a brief clinical form of the Neuropsychiatric Inventory. *J Neuropsychiatry Clin Neurosci* 2000; **12:** 233–39.

18 Logsdon RG, Gibbons LE, McCurry SM, Teri L. Quality of life in Alzheimer's disease: patient and caregiver reports. *J Ment Health Aging* 1999; **5:** 21–32.

19 EuroQol Research Foundation. EQ-5D-5L User Guide 2019. https://euroqol.org/publications/user-guides (accessed Dec 20, 2021).

20 Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 1993; **43:** 2412–14.

21 Folstein MF, Robins LN, Helzer JE. The Mini-Mental State Examination. *Arch Gen Psychiatry* 1983; **40:** 812.

22 Auer S, Graessel E, Viereckl C, Kienberger U, Span E, Luttenberger K. Professional Care Team Burden (PCTB) scale—reliability, validity and factor analysis. *Health Qual Life Outcomes* 2015; **13:** 17–25.

23 Trzepacz PT, Saykin A, Yu P, et al. Subscale validation of the neuropsychiatric inventory questionnaire: comparison of Alzheimer's disease neuroimaging initiative and national Alzheimer's coordinating center cohorts. *Am J Geriatr Psychiatry* 2013; **21:** 607–22.

24 Amare AT, Caughey GE, Whitehead C, et al. The prevalence, trends and determinants of mental health disorders in older Australians living in permanent residential aged care: implications for policy and quality of aged care services. *Aust N Z J Psychiatry* 2020; **54:** 1200–11.

25 Ridder HM, Stige B, Qvale LG, Gold C. Individual music therapy for agitation in dementia: an exploratory randomized controlled trial. *Aging Ment Health* 2013; **17:** 667–78.

26 Australian Bureau of Statistics. Data by region. https://dbr.abs.gov.au/region.html?lyr=gccsa&rgn=2GMEL (accessed Dec 20, 2021).

27 Griffiths AW, Smith SJ, Martin A, Meads D, Kelley R, Surr CA. Exploring self-report and proxy-report quality-of-life measures for people living with dementia in care homes. *Qual Life Res* 2020; **29:** 463–72.

28 Orgeta V, Brede J, Livingston G. Behavioural activation for depression in older people: systematic review and meta-analysis. *Br J Psychiatry* 2017; **211:** 274–79.

29 Jacobson NS, Dobson KS, Truax PA, et al. A component analysis of cognitive-behavioral treatment for depression. *J Consult Clin Psychol* 1996; **64:** 295–304.

30 Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. *J Med Internet Res* 2015; **17:** e30.