



The Role of Social and Physiological Variables on Older Adults' Cognitive Improvement after a Group Singing Intervention: The Sing4Health Randomized Controlled Trial

Iolanda Galinha^{a, c}, Diego Pinal^b, M. Luísa Lima^c, and Antonio Labisa-Palmeira^d

^aUniversidade Autónoma de Lisboa, CIP-UAL/UALG, Lisbon, Portugal; ^bUniversidade do Minho, Braga, Portugal; ^cISCTE CIS IUL, Lisbon, Portugal; ^dUniversidade de Lisboa & CIDEFES, Universidade Lusófona, Lisboa, Portugal

ARTICLE INFO

Received 4 December 2020
Accepted 19 April 2021
Available online 21 May 2021

Keywords:

Executive functions
Verbal memory
Respiratory function
Inflammaging
Loneliness

ABSTRACT

Population aging prompts efforts to help older adults maintain a high quality of life and independence. Group singing (GS) has shown benefits on social, physical, and cognitive domains, being a cost-effective strategy to reach these goals. Nevertheless, randomized controlled trials (RCT) evaluating systematic GS interventions and the interplay between their effects on those domains are scarce. Thus, the present RCT assessed whether a 34-session GS intervention boosts older adults' cognitive functions (executive functions and verbal memory) and whether it has an impact on systemic inflammation markers. Additionally, it intended to determine whether changes in social or physiological domains mediated the intervention effects on cognition. One hundred and forty-nine participants were allocated to a waiting-list group, which kept their usual activity levels, or an intervention group. Blinded outcome assessments were conducted at baseline, post-intervention, and follow-up. Results confirmed positive effects of GS on cognition, especially marked for verbal memory. Evidence of GS's protective effect on systemic inflammation is also provided. Mediation analyses revealed a complex interaction of GS intervention engagement and respiratory function and social wellbeing changes with the magnitude of cognitive improvement. These results highlight socialization and respiratory functions as critical pieces of multidimensional GS interventions, maximizing their benefits on older adults' cognition.

El papel de las variables sociales y fisiológicas en la mejora cognitiva de los adultos mayores tras una intervención de canto en grupo: el ensayo controlado aleatorio *Sing4Health*

RESUMEN

El envejecimiento poblacional requiere de esfuerzos para ayudar a los adultos mayores a mantener una mayor calidad de vida e independencia. El canto en grupo (CG) ha mostrado beneficios en los dominios social, físico y cognitivo, siendo una estrategia eficiente para alcanzar estos objetivos. Sin embargo, son escasos los ensayos controlados aleatorios (ECA) que evalúan intervenciones sistemáticas de CG y la interacción de sus efectos en dichos dominios. Así, el presente ECA evaluó si una intervención de CG de 34 sesiones mejora las funciones cognitivas en adultos mayores (funciones ejecutivas y memoria verbal) y si tiene efectos en los marcadores de inflamación sistémica. Además, se quiso determinar si los cambios en los dominios social y fisiológico mediaban los efectos de la intervención en la cognición. Se asignaron 149 participantes a un grupo en lista de espera, que mantuvo sus niveles de actividad habituales, o a un grupo de intervención. Se realizaron evaluaciones ciegas de los resultados al inicio, después de la intervención y durante el seguimiento. Los resultados confirmaron los efectos positivos del CG en la cognición, especialmente sobre la memoria verbal. También aportaron evidencias acerca del efecto protector del CG en la inflamación sistémica. Los análisis de mediación revelaron una compleja interacción de, por un lado, la adherencia a la intervención del CG y los cambios en la función respiratoria y el bienestar social y, por otro, la magnitud de la mejora cognitiva. Estos resultados sitúan a la socialización y la función respiratoria como piezas clave en intervenciones multidimensionales de CG, maximizando los beneficios de estas sobre la cognición en adultos mayores.

Palabras clave:

Funciones ejecutivas
Memoria verbal
Función respiratoria
Inflamaging
Soledad

Cite this article as: Galinha, I., Pinal, D., Lima, M. L., & Labisa-Palmeira, A. (2021). The role of social and physiological variables on older adults' cognitive improvement after a group singing intervention: The Sing4Health randomized controlled trial. *Psychosocial Intervention*, 30(3), 123-138. <https://doi.org/10.5093/pi2021a3>

Funding: This study was supported by a grant from Portugal's Participatory Budget (Project n.º 626, OPP 2017; Dispatch n.º 11409 C/2017). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Correspondence: iolanda.galinha@gmail.com (I. Galinha).

ISSN: 1132-0559/© 2021 Colegio Oficial de la Psicología de Madrid. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

In 2018, for the first time, the number of adults aged over 65 years exceeded that of children below 5 years old, with United Nations' projections suggesting the former will be the double of the latter by 2050 (United Nations, 2019). These demographic changes impose a significant burden on current health and social support systems (Harper, 2014) as well as on households' economic welfare (Cavallo et al., 2018). Consequently, there is a growing interest in actions that boost older adults' quality of life, cognitive integrity, and physical, mental, and social wellbeing, enabling them to keep engaged in social, economic, cultural, spiritual, and civic activities (Fernández-Ballesteros et al., 2013; Foster & Walker, 2015). Group singing (GS) programs stand as a cost-effective strategy to fulfill such criteria (e.g., Coulton et al., 2015), since they have been associated with longevity and positive lifespan trajectories of mortality (Bygren et al., 1996) as well as with higher levels of quality of life in different psychological, social, and physical domains (Clift et al., 2018; Williams et al., 2018).

GS effects on the social domain were reported, for instance, in a study by Davidson et al. (2014). Results of that study showed that completing a structured 8-week singing program was associated with an increase in older adults' social satisfaction and wellbeing. Similarly, several studies by Pearce and colleagues compared participation in adult (participants' age range: 18 to 83) singing programs with participation in crafts and other activities, and showed that the singing programs created faster cohesion and bonding in groups of previously unfamiliar individuals (Pearce et al., 2015; Pearce et al., 2017). To condense these different aspects of social relationships, participation, and engagement, Keyes (1998) coined the construct "social wellbeing", which was defined as "the appraisal of one's circumstance and functioning in society" (p. 122). Such appraisal is based on five challenges that impact quality of life: i) social integration, ii) social acceptance, iii) social contribution, iv) social actualization, and v) social coherence (Keyes, 1998). Given the aforementioned findings, it could be suggested that participation in GS activities may benefit older adults' social wellbeing (SWB). Furthermore, SWB seems to be inversely related to loneliness, since frequent contact with significant others reduce loneliness (Lee & Ishii-Kuntz, 1987). Hence, not surprisingly, GS effects include reduced loneliness, as concluded by Teater and Baldwin (2014) from semi-structured interviews with 5 participants in the Golden Oldies program. Those participants reported a reduction in their social isolation and increased social contact. Additional evidence is presented in a recent quantitative study of a 12-week intervention in residential care centers, which includes one-hour weekly sessions of singing and discussion. That study found a reduction of loneliness across 24 participants in the intervention (Millett & Fiocco, 2020; for opposite results, see Davidson et al., 2014).

Positive effects of GS have also been observed on physiological variables like respiratory function, as concluded in recent reviews in healthy (Kang et al., 2018) and clinical populations (Goldenberg, 2018). For example, GS and toning (i.e., voicing an open vowel sound at a given pitch during an exhalation) were shown to positively affect respiratory function in young adults (Bernardi et al., 2017). Regarding older adults, Mazalli et al. (2019) observed a statistically significant increase in inspiratory capacity and peak inspiration flow in 11 seniors after taking part in 16 bi-weekly singing sessions with respiratory exercises in their care center. Similarly, 42 older adults from senior living communities showed statistically significant increases in respiratory muscle strength resulting from a 12-week program, including deep-breathing training and song learning and singing (Fu et al., 2018).

The respiratory function, as reflected by forced expiratory volume in 1 s (FEV1) and forced vital capacity (FVC), was shown to be inversely related to C-Reactive Protein (CRP) levels in a sample of 7,753 participants between 20 and 80 years old (Nerpin et al., 2018). This inverse (bidirectional) relationship between respiratory function markers and CRP levels holds true also in prospective and

longitudinal studies for as long as 13 years (Ahmadi-Abhari et al., 2014; Hancox et al., 2016). CRP is commonly used as a marker of systemic inflammation (SI), which tends to be higher in older than in younger adults (Frank & Caceres, 2015), which led some authors to coin the term *inflammaging*. It is defined as an imbalance between inflammatory and anti-inflammatory processes due to a combination of immunosenescence, genetic, and lifestyle factors, which results in a low-grade, chronic, systemic, asymptomatic, and sterile status promoting frailty, morbidity, and mortality (Franceschi et al., 2000; Franceschi et al., 2007). Despite this association between respiratory function markers and SI markers, and the observed effects of GS on respiratory function, to the best of our knowledge, there are no previous studies assessing the effects of systematic GS interventions on SI levels in older adults. Nonetheless, there is preliminary evidence that singing may impact the inflammatory response. In that regard, Fancourt et al. (2016) assessed inflammatory response markers before and after a singing session through saliva samples in groups of cancer patients, bereaved carers, and carers. In all groups, one hour of GS resulted in an increase of the inflammatory markers. Providing the preliminary character of their results, Fancourt et al. refer the need for further systematic studies on the impact of GS interventions on inflammatory response.

Finally, GS interventions' effect on cognition has also been studied, albeit with mixed results. On the one hand, the 12-week intervention presented in Millett and Fiocco's (2020) work failed to find differences related to participation in any of the cognitive functions assessed: general cognition, executive function, and short-term and episodic verbal memory. On the other hand, a 12-week program including deep-breathing training and song learning and singing yielded respiratory function improvements accompanied by better performance in immediate verbal recall and verbal fluency tests, but no change in executive functions (Fu et al., 2018). Similarly, a cross-sectional analysis carried out on 1,101 surveys from the Longitudinal Aging Study Amsterdam found that habitually participating in singing activities and/or playing an instrument was associated with better executive functioning (Mansens et al., 2018). Given that executive functions (EFs) and memory are essential for the maintenance of health-enhancing behaviors and an adequate functional status (Allan et al., 2016; Bell-McGinty et al., 2002; Royall et al., 2004), and that they are among the cognitive capacities capitalizing most of the decline observed in normal aging (Park et al., 2002; Rosselli & Torres, 2019), it is necessary to clarify the effects that singing interventions may have on cognition and whether such effects are capacity-specific or rather general.

It is noteworthy that the different social and physiological aspects on which GS effects have been studied seem to modulate cognition in older adults. For instance, the social causation hypothesis posits that better social support and larger networks may delay the onset and rate of cognitive declines (Liao et al., 2016; Miceli et al., 2019). Indeed, higher social activity frequency has been associated with better executive functioning, working memory, visuospatial abilities, and processing speed (Kelly et al., 2017). Furthermore, recent systematic reviews highlight the role of loneliness in the relationship between social aspects and cognition (Boss et al., 2015; Cacioppo & Hawkley, 2009; Hawkley & Cacioppo, 2010). Also, loneliness was shown to be a predictor of general cognitive status and the incidence of Alzheimer's disease (Tilvis et al., 2004; Wilson et al., 2007). There is also evidence of a relationship between SI and cognitive function in general, and EFs in particular. Heringa et al. (2014) revealed a negative relationship between SI indices and measures of processing speed, attention, and EFs while controlling for cardiovascular risk factors effects. Similar results were observed in another study controlling for sociodemographic factors, depression, and other clinical covariates (Tegele et al., 2016). Furthermore, composite scores of SI were predictive of decline in EFs performance in prospective studies spanning 20 years (Beydoun et al., 2019; Walker et al., 2019).

However, work relating respiratory function to cognition in older adults is scarce, with weak but positive relations being reported (for a systematic review and a meta-analysis; see [Duggan et al., 2020](#); [Duggan et al., 2019](#)).

In sum, the available evidence highlights the benefits that GS interventions may have on SWB ([Davidson et al., 2014](#); [Pearce et al., 2015](#); [Pearce et al., 2017](#)) as well as in the reduction of older adults' loneliness ([Millett & Fiocco, 2020](#); [Teater & Baldwin, 2014](#)). Similar positive effects of GS interventions on respiratory function were also observed ([Fu et al., 2018](#); [Mazalli et al., 2019](#)). Nevertheless, it is still unclear whether GS interventions may modify older adults' SI levels as well as whether those interventions promote their general cognition or specific cognitive functions. Furthermore, recent research revealed a complex interplay between cognition and SWB (for a review see [Kelly et al., 2017](#)) and loneliness (for a review see [Boss et al., 2015](#)), as well as between cognition and physiological aspects ([Duggan et al., 2020](#); [Heringa et al., 2014](#); [Tegeler et al., 2016](#)). However, no previous study has addressed the interaction of GS interventions in these different domains, to the best of our knowledge.

Consequently, the aim of the present study is two-fold. First, it aims to assess whether a singing intervention of 34 two-hour sessions is associated with changes in the level of SI markers and whether it has positive effects on verbal memory, EFs, or both cognitive functions' performance. Secondly, it intends to explore the relationship between potential beneficial effects of GS on the cognitive, social, and physiological level. Specifically, it assessed whether engagement in the intervention was a predictor of cognitive change, and whether such relationship between GS and cognition was mediated by the intervention effects in social processes related to loneliness and wellbeing or in physiological variables related to respiratory function and SI. In this way, based on previous research, we expected that engagement with GS sessions would provide an increase in SWB ([Davidson et al., 2014](#)) and a reduction in older adults' loneliness ([Millett & Fiocco, 2020](#)), which following the social causation hypothesis ([Liao et al., 2016](#)) may be related with an increase in cognitive performance at the end of the intervention or, at least, with a smoother decline of cognitive functioning than that observed for a control group not participating in singing activities. Similarly, previous GS interventions observed an improvement of respiratory function ([Fu et al., 2018](#)), which led us to expect increased levels of respiratory function at the end of the intervention as well as in comparison with a control group. Further, given the inverse relationship of respiratory function and SI ([Nerpin et al., 2018](#)) and preliminary data pointing to an increase of the immune response as a result of participating in a GS session ([Fancourt et al., 2016](#)), we also expected reduced markers of SI at the end of the intervention and compared with a control group. Finally, as the expected SI and respiratory function changes have been associated with improvements in cognitive function ([Beydoun et al., 2019](#); [Duggan et al., 2020](#), respectively), we also expected physiological changes due to engagement in the GS intervention to contribute to the buildup of cognitive improvements associated with the intervention.

Method

Design

This RCT (Sing4Health; trial registration number: NCT03985917) included a randomized experimental phase (T0 to T1 with a within- and a between-groups comparison) and a non-randomized quasi-experimental follow-up phase (T1 to T2 within-groups comparison). The RCT included two groups, an immediate GS intervention group (IG), and a waiting-list group (WLG). For the first phase, WLG participants maintained their participation in other daycare center

intervention activities like sports (aerobics and water aerobics), artistic activities (theatre, painting, and music), or crafts (crocheting) except GS (constituting an active control group). The IG completed 34 two-hour GS sessions (see [Galinha, Farinha, et al., 2020](#) for the detailed RCT protocol). After a two-month interval, in the second phase, the WLG participants switched to complete the intervention program while the former IG participated in other activities, excluding GS.

Details about the structure of the sessions comprised in the intervention can be found in the Supplementary Information as well as in the intervention protocol ([Galinha, Farinha, et al., 2020](#)). In short, the intervention was carried out in a rehearsal room with a chair for each participant and an extra cleared space for standing-up exercises and activities. Each intervention session comprised: i) relaxation exercises and vocal warm-up (e.g., exercises promoting effective breathing and group dynamics), ii) a vocal technique component, iii) repertoire rehearsals, iv) 20-minute breaks for socialization, v) development of a show to be presented to the public and families, and vi) assessment of each participant's performance.

Participants completed a battery of psycho-emotional, psychosocial, and cognitive tests as well as the collection of biological samples to determine several health indicators after enrollment in the RCT (T0, February 2019), at the end of the first round of intervention (T1, July 2019), and at the end of the second round of intervention (T2, February 2020) (see [Figure 1](#)). Data collection at each time point extended for 10 weekdays, so 10 to 15 participants were assessed per day in the facilities of the social care institution (SCI). Assessments began early in the morning to acquire blood samples and respiratory function and balance data while participants were fasting. After a breakfast provided by the SCI, cognitive assessments were carried out followed by a 15-20 min break. Finally, psychosocial questionnaires were completed and semi-structured interviews conducted. Each phase of data collection was performed by different independent trained researchers who were blinded to participants allocation to the IG or WLG. All RCT protocols followed the Declaration of Helsinki principles and were approved by the Ethics Committee of the Centro de Investigação em Psicologia at Universidade Autónoma de Lisboa (Approval Nr. 12-09-2018). After a briefing session with the intervention and research teams, signed informed consent from each participant was obtained before their participation in the RCT. All RCT data was collected and stored following European and national directives and legislation for vulnerable human participants and for sensitive data (see [Galinha, Farinha, et al., 2020](#)).

Sample

The Sing4Health RCT protocol ([Galinha, Farinha, et al., 2020](#)) planned the IG and WLG to have 70 participants each. A post hoc power analysis showed that this sample size was enough to detect large effect sizes ($\eta^2 = .14$) with an 88% power for repeated measures ANCOVAs. Further, a dropout of 15 participants per group would have reduced the statistical power to 79%. For the RCT implementation, 149 participants aged 60 or over were recruited from daycare centers in January 2019. Participants were allocated by the scientific coordinators to the IG ($n = 75$) or the WLG ($n = 74$) using the IBM® SPSS® Statistics randomization tool to split a numerically ordered list of participants into (initially) two equal groups. Twelve IG participants dropped out during the first two sessions of intervention due to scheduling incompatibility with the intervention sessions, allocation of friends or partners to a different experimental group and health-related issues like illness, scheduled treatments, operations, recovery from operations, or taking care of dependent relatives. Thus, following the same randomization procedure, 14 participants were relocated from the WLG to the IG (see [Figure 1a](#) and [1b](#) with the CONSORT flow diagram) to ensure enough statistical power to

test, at least, the intervention effects within the IG. Therefore, the final sample consisted of an IG of 89 individuals and a WLG of 60 participants. Within the IG, 77 (86.5%) participants completed T0 and T1 assessments, albeit only 54 (60.7%) completed at least 75% of the intervention sessions. For the follow-up assessment at T2, 57 (64.0%) participants of the IG were available. Within the WLG, 47 (78.3%) participants completed T0 and T1 assessments, and 30 (50%) the T2 assessment after their participation in the GS intervention. Twenty-three (38.3%) participants of this group completed at least 75% of the intervention sessions.

All participants met the following inclusion criteria: i) being older than 60 years of age, ii) being retired, and, iii) lack of participation in other structured interventions in the previous four months. The only exclusion criterion was the presence of severe diagnosed impairments precluding involvement in GS activities (e.g., severe sensory or mobility impairments).

Materials and Procedure

At T0, all participants filled a brief questionnaire on sociodemographic data, including information about sex, education level, birth date, and diagnosed medical conditions. Furthermore, at T0, T1, and T2, participants completed the Portuguese adaptations of the following tests (for further details on the RCT protocols, see Galinha, Farinha, et al., 2020):

- Montreal Cognitive Assessments (MoCA; Freitas et al., 2011, 2012) for screening dysfunction in six cognitive domains (i.e., EFs, visuospatial skills, short-term memory, language, attention, concentration, and working memory, and temporal and spatial orientation). Participants with fewer than 12 years of formal education received one extra point, as recommended in the original work by Nasreddine et al. (2005). Higher scores indicate better performance.

- Digit-Symbol subtest of the Weschler Adult Intelligence Scale – III (DS-WAIS; Wechsler, 1997) to assess cognitive and motor speed, planning ability, visual memory, visuomotor coordination, and attention. Higher scores indicate better performance.

- Immediate recall on the Rey Auditory Verbal Learning Test (RAVLT; Cavaco et al., 2015) to measure verbal learning, immediate memory, and retention of information. The number of words correctly retrieved in the fifth attempt was considered for analysis.

- Trail Making Test part A (TMT; Cavaco et al., 2013) to measure processing speed. Since some participants did not conclude the test within the allowed 5 minutes, a composite score calculated as the number of points correctly connected divided by the time expended to join them was calculated for each participant.

- An adaptation of the UCLA Loneliness Scale (UCLALS; Pocinho et al., 2010) consisting of four items to be responded on a 4-point Likert scale. This test provides a score operationalizing the feeling of being cut off from others. Higher values indicate more loneliness.

- A shortened version of the Social Well-being Scale (Lages-Silva, 2016) consisting of six items rated on a 6-point Likert scale. The larger the total score, the higher the level of social wellbeing. This measure showed low internal consistency at T0, T1, and T2 (Cronbach's α : T0 = .34, T1 = .32, and T2 = .48). Thus, a composite score was created by adding scores on the subscales of satisfaction with social relationships from the World Health Organization Quality of Life – Bref (Vaz Serra et al., 2006), and social participation from the World Health Organization Quality of Life – Old (Pereira Vilar, 2015) tests. Both tests' items should be responded on a 5-point Likert scale, with higher values indicating greater satisfaction or quality of life in the measured dimension. The composite score for SWB ranged between 13 and 71 points and showed adequate internal consistency at T0, T1, and T2 (Cronbach's α : T0 = .64, T1 = .71, and T2 = .75).

A fasting blood sample was collected in the morning for each participant at T0, T1, and T2. Samples were analyzed by a certified

laboratory to assess the physiological effects of the intervention. These analyses determined:

- C-reactive protein (CRP) levels. Before entering further analyses, obtained values were multiplied by 10 and log-transformed to increase their distribution normality.

- Erythrocyte Sedimentation Rate (ESR). Obtained values were transformed to a percentage of the expected value according to the formula $\text{age} / 2$ for men; and $(\text{age} + 10) / 2$ in women (Miller et al., 1983).

- Each participant also completed a spirometry at T0, T1, and T2. A professional portable spirometer (Medikro Pro) validated according to the American Thoracic Society (ATS) and European Respiratory Society (ETS) criteria (Miller et al., 2005) was used. This spirometer monitors ambient conditions to ensure automated and constant calibration, allowing to precisely assess the following parameters of respiratory function:

- Forced vital capacity (FVC), defined as the total volume of air exhaled during a forced breath. Additionally, the Tiffeneau-Pinelli index was calculated as the ratio between air exhaled volume in the first second of a forced breath (FEV1) and FVC (i.e., FEV1/FVC).

- The flow of air exhaled from 25 % to 75 % of the FVC in a forced breath (FEF25-75).

All spirometry data was determined via the Medikro Spirometry Software. For the statistical analyses, the percentages over recommended reference values proposed by *Direção Geral da Saúde* (2016) and the National Health and Nutrition Survey (NHANES III) were used.

Data Preparation

To deal with extreme values, Mahalanobis distance (MD) was calculated in a multivariate space with scores at T0, T1, and T2 for MoCA, RAVLT, DS-WAIS, UCLALS, as well as the total number of words produced in the fluency subtest of MoCA, the calculated points per second value for TMT, the composite score for social wellbeing, the percentage to predicted values for FVC, FEV1/FVC, FEF25-75, and ESR, and the log-transformed CRP levels. Multivariate outliers were defined as individuals showing an MD exceeding the value with an associated probability of .001 in a χ^2 distribution with as many degrees of freedom as variables in the multivariate space (i.e., 36). Four multivariate outliers were identified and discarded for further analysis. Additionally, any individual scores deviating from the variable mean at least 2.2 times the interquartile range were considered univariate outliers (Hoaglin et al., 1986; Hoaglin & Iglewicz, 1987). Univariate outliers were then “winsorized” to reduce the effect of extreme values in the analyses (Tukey, 1962).

Regarding missing data, first, participants' mean in a given test was used to replace missing values for a single item in that same test (three participants with one missing response for MoCA, four for SWB, two for WHOQoL-OLD, and six for WHOQoL-BREF). Second, an analysis of the data matrix indicates a total of 21.99 % missing data across observations. Little's missing completely at random (MCAR) test results confirmed that there was no pattern in the missing data, $\chi^2(1759) = 1811.097$, $p = .189$. Consequently, missing values were estimated separately for IG and WLG (Sullivan et al., 2018) through multiple imputation. A predictive mean matching (PPM) procedure, as implemented in SPSS (version 21; IBM), was used to run 100 imputations with the following predictors: scores for MoCA, RAVLT, DS-WAIS, UCLALS, the total number of correct words produced in the fluency subtest of MoCA, and the composite score of SWB, all at T0. Weight at T0, T1, and T2; and age, sex, years of education, height, and the number of completed sessions were also used as predictors. None of the predictors presented missing values. The mean value of the 100 imputations was used for subsequent analysis after checking that the variables' distributions were similar before and after imputation.

Finally, the standardized residuals of a regression between pre-GS intervention and post-GS intervention scores (T0 and T1 for the IG, T1 and T2 for the WLG) for each variable were calculated. In this way, analyses could be performed with a measure that controls a given variable's level before intervention and is informative of the change in that variable from pre-intervention to post-intervention (Cohen et al., 2003).

Statistical Analyses

Statistical analyses were performed following an intention to treat (ITT) approach. Thus, the initial randomization was considered regardless of the number of intervention sessions completed by the participants. Analyses were also run following a per-protocol (i.e., intervention was considered complete if an individual participated in at least 75% of the sessions) approach for comparison purposes, and their results are reported in the supplementary information.

Statistical analyses were run with data from 145 participants: 57 in the WLG and 88 in the IG. An alpha level of .05 was maintained for all tests. All statistical analyses were performed using IBM SPSS Statistics 21 and the PROCESS macro for mediation and moderation

analyses (Hayes, 2017; Hayes & Preacher, 2014; Hayes & Rockwood, 2017).

Group homogeneity after randomization. To test the homogeneity between IG and WLG at T0, independent samples *t*-tests were run comparing the mean MoCA scores of IG and WLG groups. Additionally, since age, Kolmogorov-Smirnov's $Z(145) = 0.083$, $p = 0.016$, the number of years of formal education, $Z(139) = 0.314$, $p < 0.001$, and DSWAIS score, $Z(145) = 0.093$, $p = 0.004$, were non-normally distributed, a series of Mann-Whitney tests comparing IG and WLG for these variables at T0 was run. Sex distribution equality between groups was tested through a χ^2 test.

GS intervention effect on cognitive, physiological, and social variables. To evaluate the GS intervention effects on cognition as compared with a control group, mixed repeated-measures ANOVAs with the between-subjects factor Group (IG vs. WLG) and the within-subjects factor Time (T0 vs. T1) were run on TMT points per second, and RAVLT correctly remembered words for list 5. ANOVAs were used as they have been shown to be still valid under conditions of non-normality of data (Blanca et al., 2017; Khan & Rayner, 2003). No sphericity violations were encountered, and consequently no adjustments were necessary. Additionally, to test for physiological effects and check the presence

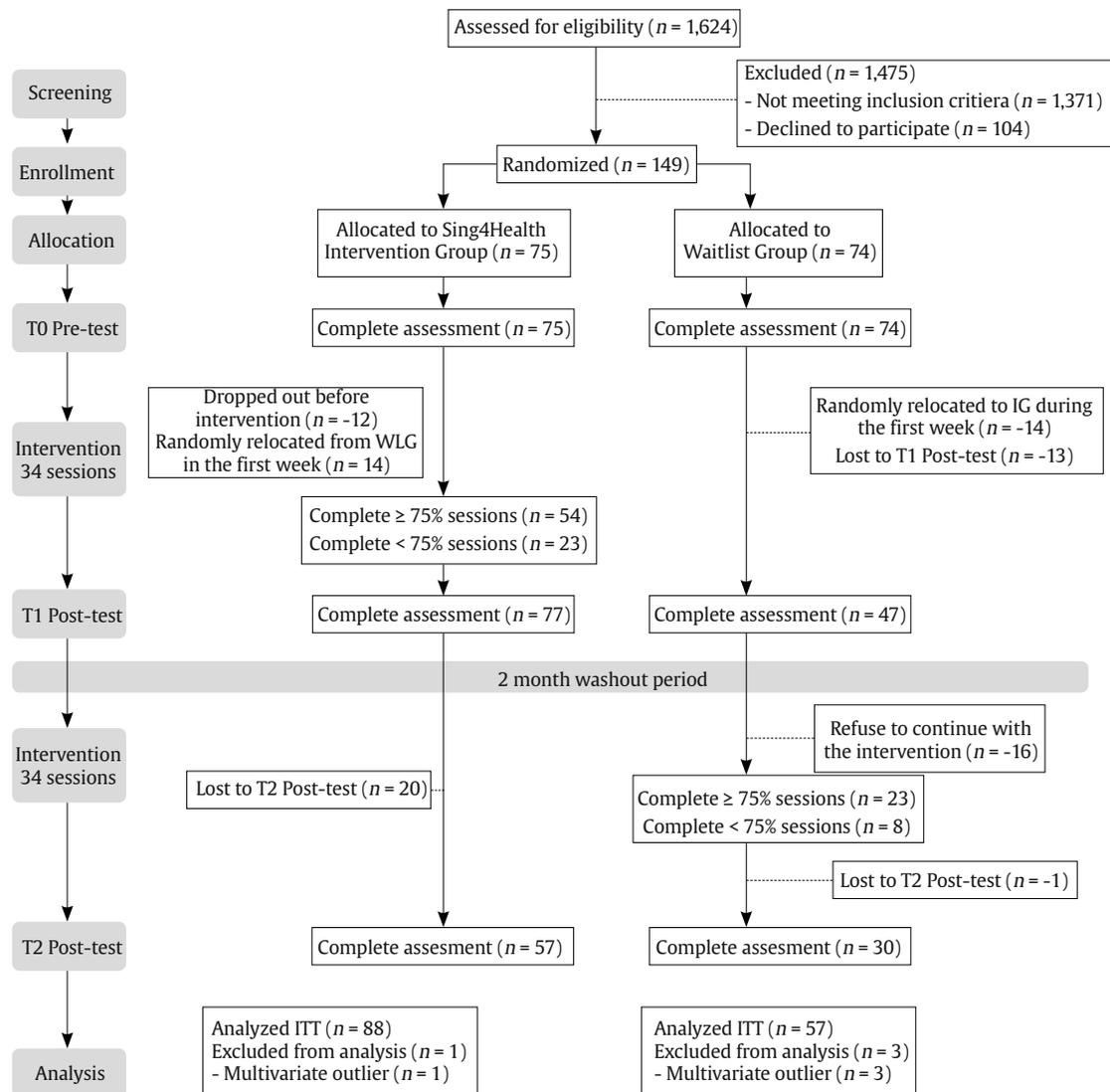


Figure 1a. Flowchart Showing Recruitment and Progress through the Sing4Health RCT.

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract	1a	Identification as a randomised trial in the title	1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	3
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	4-8
	2b	Specific objectives or hypotheses	8-9
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	9-10
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	11
Participants	4a	Eligibility criteria for participants	11
	4b	Settings and locations where the data were collected	10
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	9-10, SI
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	12-14
	6b	Any changes to trial outcomes after the trial commenced, with reasons	13
Sample size	7a	How sample size was determined	10-11
	7b	When applicable, explanation of any interim analyses and stopping guidelines	---
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	11
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	11
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	11
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	11
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	10
	11b	If relevant, description of the similarity of interventions	---
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	16-17
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	16-17
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	11
	13b	For each group, losses and exclusions after randomisation, together with reasons	11, Figure 1
Recruitment	14a	Dates defining the periods of recruitment and follow-up	10-11
	14b	Why the trial ended or was stopped	---
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Table 1, 18
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	15, 18-22, Tables
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	Tables, 18-22
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	---
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	21-23
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	---
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	29
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	29-30
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	23-28
Other information			
Registration	23	Registration number and name of trial registry	2
Protocol	24	Where the full trial protocol can be accessed, if available	9
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	1

Figure 1b. CONSORT 2010 Checklist of Information to Include when Reporting a Randomised Trial¹.

¹We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

Table 1. Homogeneity of the Experimental Groups on Sociodemographic and Cognitive Variables

	IG (<i>n</i> = 88)		WLG (<i>n</i> = 57)		Statistic	<i>p</i> -value	Effect size
	Mean	<i>SD</i>	Mean	<i>SD</i>			
Age	76.56	8.73	76.35	8.73	<i>U</i> = 2500.0	.974	< .001
Years of formal education	4.83	3.33	4.77	3.48	<i>U</i> = 2205.0	.577	.002
MoCA at T0	22.83	18.25	22.11	15.16	<i>t</i> (143) = 1.177	.241	.199
DSWAIS at T0	17.77	5.58	16.65	5.68	<i>U</i> = 2502.0	.981	< .001

Note. Sociodemographic and cognitive test means and standard deviations (*SD*) for the IG and WLG. Test statistics, associated *p* and estimated effect size (i.e., η^2 for Mann-Whitney tests and Cohen's *d* for *t*-test) values for the tests assessing between group differences at the start of the RCT protocol can also be found in the table. MoCA = Montreal Cognitive Assessment; DSWAIS = digit-symbol subtest of the Weschler Adult Intelligence Scale III.

of potential social effects of the GS interventions, mixed repeated-measures ANOVA with the same factors were run on: CRP levels, and the percentage to predicted values of ESR, FVC, FEV1/FVC, and FEF25-75, as well as UCLALS scores and the SWB composite score.

Further, to study the effects of the GS intervention, a series of Wilcoxon signed-ranks tests comparing pre- with post-GS intervention scores in the experimental variables were performed to further explore changes in cognitive, social, and physiological domains. To that end, T0 and T1 for IG and T1 and T2 for WLG were used as pre- and post-intervention assessments, respectively. Note that for this analysis data from the total sample of 145 participants was used as a single group.

Maintenance of GS intervention effects at 6-month follow-up.

To study the maintenance over time of the potential GS intervention effects, a repeated-measures ANOVA with the within-subjects factor time (T0 vs. T1 vs. T2) was run on IG's scores for the variables mentioned above. Greenhouse-Geisser corrections were applied where appropriate.

Relationship between cognitive, physiological, and social GS intervention effects. Finally, standardized residuals from the regression between pre- and post-GS intervention scores were used in an exploratory mediation analysis evaluating the relationship between physiological, social, and cognitive effects of the GS intervention. According to our hypothesis, multiple mediation was tested using model 4 on the PROCESS macro for SPSS. This analysis allows examining the relationships between one independent variable (X: number of completed GS intervention sessions), one dependent variable (Y: TMT points per second or RAVLT standardized residuals) and two or more simultaneous mediator variables (MVsocial: UCLALS scores and SWB composite score standardized residuals or MVphysiological: CRP levels and the percentage to predicted values for ERS, FVC, FEV1/FVC or FEF25-75 standardized residuals). Further, given the exploratory nature of these analyses, interaction effects between the independent and mediator variables

were also tested. When statistically significant, such interactions are interpreted as moderation effects of the predictor on the relationship between the mediator and the model outcome.

For exploratory mediation analyses, mean centering of the variables was used for interactions between X and M variables. Further, 10,000 bootstrap resamplings were performed and a 95% bias-corrected confidence interval (CI) was used to evaluate indirect effects (Preacher & Hayes, 2008). These indirect effects are considered statistically significant whenever zero is not included in the CI. Data is reported for unadjusted analysis since no control variables were included in the models. Additionally, all reported results are standardized coefficients for indirect and direct effects.

Results

Group Homogeneity after Randomization

Sex distribution was similar in IG (*n* = 88) and WLG (*n* = 57), $\chi^2(1) = 2.039$, *p* = 0.153, Cohen's *W* = 0.153. Also, *t*-test on MoCA scores at T0 and Mann-Whitney tests on age, years of formal education, and DSWAIS failed to reach statistical significance (see Table 1 for details). These results demonstrated the homogeneity of groups after participants' random allocation and before the start of the intervention.

GS Intervention Effect on Cognitive, Physiological, and Social Variables

Descriptive statistics for the IG (*n* = 88) and WLG (*n* = 57) as well as the mixed repeated-measures ANOVAs *F* statistic, *p* values of main and interaction effects, and estimated effect sizes can be found on Table 2. Analyses on TMT points per second failed to reach significance for group or time main effects as well as for the interaction between these factors. Regarding RAVLT correctly remembered words for the

Table 2. Group Singing Intervention Effects Comparison with an Active Control Group

	IG (<i>n</i> = 88)		WLG (<i>n</i> = 57)		Group			Time			Interaction		
	T0	T1	T0	T1	<i>F</i> (1, 143)	<i>p</i>	η^2	<i>F</i> (1, 143)	<i>p</i>	η^2	<i>F</i> (1, 143)	<i>p</i>	η^2
TMT	0.31 ± 0.20	0.32 ± 0.19	0.29 ± 0.16	0.30 ± 0.18	0.62	.432	.004	1.00	.319	.007	0.01	.974	< .001
RAVLT	8.08 ± 3.03	8.18 ± 2.67	7.75 ± 3.42	6.57 ± 3.70	3.90	.050	.033	6.39	.013	.043	8.99	.003	.059
UCLALS	9.23 ± 3.17	8.55 ± 2.99	9.07 ± 3.26	8.58 ± 2.33	0.02	.888	< .001	5.22	.024	.035	0.14	.709	.001
SWB	48.65 ± 6.44	47.83 ± 7.34	48.88 ± 7.85	45.63 ± 6.35	0.88	.349	.006	13.04	<.001	.084	4.68	.032	.032
ESR	59.35 ± 43.04	65.63 ± 40.66	64.73 ± 39.81	68.30 ± 39.85	0.43	.511	.003	2.10	.150	.014	0.16	.691	.001
CRP	0.37 ± 0.63	0.34 ± 0.56	0.32 ± 0.55	0.52 ± 0.69	0.57	.450	.004	2.17	.143	.015	4.41	.037	.030
FVC	60.37 ± 12.27	60.13 ± 11.67	54.84 ± 12.27	56.72 ± 11.97	6.29	.013	.042	0.67	.416	.005	1.11	.293	.008
FEV1/FVC	105.41 ± 13.27	103.68 ± 11.38	104.65 ± 11.99	103.36 ± 12.11	0.09	.753	.001	1.69	.196	.012	0.04	.848	< .001
FEF25-75	83.42 ± 33.51	82.68 ± 32.79	69.81 ± 31.14	70.72 ± 27.50	8.35	.004	.055	0.01	.976	< .001	0.07	.790	< .001

Note. IG and WLG mean and standard deviation for the cognitive, social, and physiological variables at T0 and T1, as well as the *F* statistic, associated *p* and estimated effect size (η^2) values for a repeated-measures ANOVA with the between subject factor group (IG vs. WLG) and the within-subjects factor time (T0 vs. T1). TMT = Trail Making Test correctly linked points per second; RAVLT = correct words remembered for the fifth list; UCLALS = adapted UCLA Loneliness Scale scores; SWB = composite score of social wellbeing; ESR = erythrocyte sedimentation rate percentage to predicted values; CRP = log 10 transformed C-reactive protein levels; FVC = forced vital capacity percentage to predicted values; FEV1/FVC = forced expiratory volume in 1 second on FVC percentage to predicted values; FEF25-75 = forced expiratory flow between 25% and 75% of the FVC percentage to predicted values.

fifth presentation of the list, significant main effects of group and time were observed, as well as a significant interaction between group and time. Post-hoc comparisons showed that only the WLG had higher scores at T0 compared with T1 (mean difference = 1.19, 95% CI [0.53, 1.85], $p = 0.001$, $\eta^2 = .081$), thus indicating a decline of verbal memory over time. Further, at T1, scores from the IG were higher than those of the WLG (mean difference = 1.61, 95% CI [0.57, 2.66], $p = .003$, $\eta^2 = .061$), hence, indicating a positive effect of GS in maintaining verbal memory capacities.

Regarding physiological variables, no significant main effects or interactions were observed for ERS percentage to predicted values. CRP levels, however, showed a significant interaction between group and time, while the main effects were non-significant. Post-hoc comparisons indicated that only the WLG had lower CRP levels on T0 than T1 (mean difference = -0.20, 95% CI [0.38, -0.03], $p = .023$, $\eta^2 = .035$). Hence, the GS intervention may be protective, avoiding increases in SI.

Table 3. Baseline Assessment Comparison against post Group Singing Intervention Assessment

<i>n</i> = 145	Pre-intervention	Post-intervention	Wilcoxon Signed-Rank test		
	Median	Median	Z	<i>p</i> -value	<i>r</i>
TMT	0.301	0.327	-1.91	.056	-.112
RAVLT	7	9	-4.47	< .001	-.262
UCLA	8.98	8.51	-1.74	.082	-.102
SWB	47	48.35	-1.26	.207	-.074
ESR	62.92	57.03	-0.56	.572	-.033
CRP	0.576	0.576	-0.37	.711	-.022
FVC	58	56	-2.14	.032	-.126
FEV1/FVC	103.17	99.61	-4.07	< .001	-.239
FEF25-75	72.59	63.66	-2.68	.007	-.157

Note. Median for the experimental variables before and after the group singing intervention across participants, together with the Z statistic, associated *p* and standardized effect size (i.e., *r*) values from the Wilcoxon signed rank test contrasting both assessments. TMT = Trail Making Test correctly linked points per second; RAVLT = correct words remembered for the fifth list; UCLA = adapted UCLA Loneliness Scale scores; SWB = composite score of social wellbeing; ESR = erythrocyte sedimentation rate percentage to predicted values; CRP = log 10 transformed C-reactive protein levels; FVC = forced vital capacity percentage to predicted values; FEV1/FVC = forced expiratory volume in 1 second on FVC percentage to predicted values; FEF25-75 = forced expiratory flow between 25% and 75% of the FVC percentage to predicted values.

Regarding respiratory function, no significant main or interaction effects involving time were observed for the percentage to predicted values of FVC. Nevertheless, a main effect of group indicated that the IG presented larger air volumes than the WLG (mean difference = 4.47, 95% CI [0.95, 7.99]). Similarly, for FEF25-75 there were no

significant main or interaction effects of time but a main effect of group, indicating that the IG presented more expiratory flow than the WLG (mean difference = 12.78, 95% CI [4.04, 21.53]). No significant main effects of group or time nor interactions were observed for FEV1/FVC values.

As regards social dimensions, the repeated-measures ANOVA on UCLALS scores showed a significant effect of time, indicating that both groups presented higher scores on T0 than T1 (mean difference = 0.581, 95% CI [0.08, 1.08]). No effect of group or interactions were observed involving UCLALS. For the composite score on SWB, a significant main effect of time and a significant interaction of group and time were found. Post-hoc comparisons showed that only participants in the WLG presented higher composite scores of SWB at T0 than at T1 (mean difference = 3.25, 95% CI [1.52, 4.98], $p < .001$, $\eta^2 = .088$); thus, pointing towards a GS effect in maintaining adequate levels of SWB.

Finally, median for pre-intervention and post-intervention assessments for all participants (*n* = 145) as well as the Z statistic, associated *p*, and estimated effect size values from the Wilcoxon signed-rank tests can be found in Table 3. These tests revealed a marginally significant effect for TMT's points per second and a significant effect for RAVLT correctly remembered words at the 5th repetition. In both cases, scores were better for post- than pre-intervention assessments, supporting a positive effect of GS on cognitive function.

Regarding physiological variables, no statistically significant effects were found for ERS or CRP levels. The effects on respiratory function variables were significant for FVC, FEV1/FVC, and FEF25-75. In all three cases, values were higher at pre- than post-intervention assessments, contrary to GS's hypothesized effects on respiratory function.

As regards social variables, no statistically significant differences were observed for the social dimension variables, namely, UCLALS and the composite score of SWB.

Maintenance of GS Intervention Effects at 6-Month Follow-up

Descriptive statistics of the IG (*n* = 88) at the three time points of the RCT, as well as the repeated-measures ANOVA *F* statistic, associated *p*, and estimated effect sizes values can be found in Table 4. Analyses on TMT's points per second across the three assessment points failed to reach significance. Likewise, no significant main effect of time was observed for RAVLT scores. These results indicate that cognitive scores were stable across all time points on the RCT.

Table 4. Comparison between T0, T1 and T2 Scores in the Experimental Variables to Assess Maintenance of Intervention Effects over Time

<i>n</i> = 88	T0		T1		T2		<i>F</i> (2, 174)	Time	
	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>		<i>p</i> -value	η^2
TMT	0.312	0.198	0.324	0.187	0.332	0.193	0.98	.377	.001
RAVLT	8.08	3.03	8.18	2.67	8.10	2.68	0.08	.920	.011
UCLA	9.23	3.17	8.55	2.99	8.58	2.64	3.62	.029	.040
SWB	48.65	6.44	47.83	7.34	47.35	7.25	1.98	.141	.022
ESR	59.35	43.04	65.63	40.66	62.12	31.92	0.91	.404	.010
CRP	0.371	0.632	0.335	0.556	0.589	0.534	8.59	< .001	.164
FVC	60.37	12.27	60.13	11.67	62.24	14.22	1.41 ¹	.247	.016
FEV1/FVC	105.41	13.27	103.68	11.38	87.69	18.33	54.20¹	< .001	.384
FEF25-75	83.42	33.51	82.68	32.79	65.53	29.16	14.46 ¹	.247	.143

Note. ¹Greenhouse-Geisser correction was applied for FVC (degrees of freedom: 1.544, 134.341, $\epsilon = 0.772$), for FEV1/FVC (degrees of freedom: 1.753, 152.518, $\epsilon = .877$), and for FEF25-75 (degrees of freedom: 1.799, 156.488, $\epsilon = .899$).

IG mean and standard deviation (*SD*) for the cognitive, social and physiological variables at T0, T1, and T2 and the *F* statistic, associated *p* values and estimated effect sizes for a repeated-measures ANOVA with the within-subjects factor Time (T0 vs. T1 vs. T2). TMT = Trail Making Test correctly linked points per second; RAVLT = correct words remembered for the fifth list; UCLA = adapted UCLA Loneliness Scale scores; SWB = composite score of social wellbeing; ESR = erythrocyte sedimentation rate percentage to predicted values; CRP = log 10 transformed C-reactive protein levels; FVC = forced vital capacity percentage to predicted values; FEV1/FVC = forced expiratory volume in 1 second on FVC percentage to predicted values; FEF25-75 = forced expiratory flow between 25% and 75% of the FVC percentage to predicted values.

Regarding physiological variables, there were no significant effects of time for ERS values. In contrast, a statistically significant effect of time for CRP was found, showing that participants at T0 ($p = .010$) and T1 ($p < .001$) presented lower levels of CRP than at T2 (means difference and 95% CI, for T0-T2 = -0.219, from -0.396 to -0.041; for T1-T2 = -0.254, from -0.408 to -0.101). These results may indicate that the effects of GS on SI may be transient. For the respiratory function, no significant main effect of Time was found for FVC, but statistically significant effects of time on FEV1/FVC and FEF25-75 were observed. In both cases, participants showed better indices of respiratory function at T0 ($p < .001$ in both cases) and T1 ($p < .001$ in both cases) than at T2 (FEV1/FVC: means difference T0-T2 = 17.727, 95% CI, 12.764 to 22.690; means difference T1-T2 = 15.996, 95% CI [10.971, 21.020]; FEF25-75: means difference T0-T2 = 17.885, 95% CI [8.745, 27.026]; means difference T1-T2 = 17.154, 95% CI [9.382, 24.927]).

Finally, as regards the social domain, no significant effect of time was observed for the composite score of SWB. A statistically significant main effect of time for UCLALS scores was observed; nonetheless, pairwise comparisons failed to reach significance (all $p > .76$).

Relationship between the Cognitive, Physiological and Social GS Intervention Effects

The multiple mediation analysis ($n = 145$) using the number of GS completed sessions as the predictor, the number of words correctly remembered at the 5th repetition of the RAVLT as the outcome, and UCLALS and the composite score of SWB as mediator variables (Figure 2a) revealed that the total effect model was significant, $R^2 = .041$, $F(1, 143) = 6.052$, $p = .015$, as was the full model with mediators, $R^2 = .060$, $F(3, 141) = 2.995$, $p = 0.033$. The direct effect (path c') of GS completed sessions on RAVLT standardized residuals was significant (point estimate = -0.015, $p = .015$); however, none of the indirect effects were significant. Hence, no mediation effects seem to exist for social dimension variables on the relationship between completed sessions in the intervention and cognitive change. However, the interaction between GS completed sessions and SWB standardized residuals was significant, $F(1, 140) = 6.007$, $p = .015$, pointing to a moderation effect of GS on the effects of SWB change over RAVLT change.

Using the same predictor and outcome variables, but CRP levels, and the percentage to predicted values for ERS, FVC, FEV1/

FVC and FEF25-75 standardized residuals as mediator variables (Figure 2b), the total effect model was also significant, $R^2 = .041$, $F(1, 143) = 6.052$, $p = .015$, as was the full model with mediators, $R^2 = .152$, $F(6, 138) = 4.134$, $p = .001$. The direct effect (path c') of GS completed sessions on RAVLT standardized residuals failed to reach significance (point estimate = -0.008, $p = .186$), as well as the indirect effects. Thus, none of the physiological variables, either respiratory or SI, seem to mediate between completed sessions in the intervention and cognitive change. However, the interaction between GS completed sessions and FEV1/FVC standardized residuals was significant, $F(1, 137) = 5.329$, $p = .022$, indicating a moderation effect of GS sessions on effects of changes in respiratory function over RAVLT change.

Multiple mediation was also tested ($n = 145$) using the same predictor and social mediator variables but with TMT points per second standardized residuals as outcome variable (Figure 3a). The total effect model in this analysis failed to reach significance, $R^2 = .003$, $F(1, 143) = 0.433$, $p = .511$, although the full model with mediators was significant, $R^2 = .060$, $F(3, 141) = 2.865$, $p = .039$. The direct effect (path c') of GS completed sessions on TMT standardized residuals was not significant (point estimate = 0.002, $p = .741$), as were the indirect effects. Hence, no mediation effects may exist for social dimension variables on the relationship between completed sessions in the intervention and TMT performance change. No significant interactions between GS completed sessions and social mediator variables were found either.

Regarding the model with standardized residuals of physiological indices as mediator variables of the relationship between completed GS sessions and TMT performance change (Figure 3b), the total effect model also failed to reach significance, $R^2 = .003$, $F(1, 143) = 0.433$, $p = .511$, as well as the full model with mediators, $R^2 = .039$, $F(6, 138) = 0.934$, $p = .472$. Thus, neither the direct effect (point estimate = 0.007, $p = .325$) nor the indirect effects were significant. Thus, none of the physiological variables seems to mediate the relationship between completed sessions in the intervention and TMT performance. Interestingly, the interaction between GS completed sessions and FEV1/FVC standardized residuals was again significant, $F(1, 137) = 5.005$, $p = .027$, indicating a moderation effect of GS sessions on the effects of changes in the respiratory function over TMT performance change.

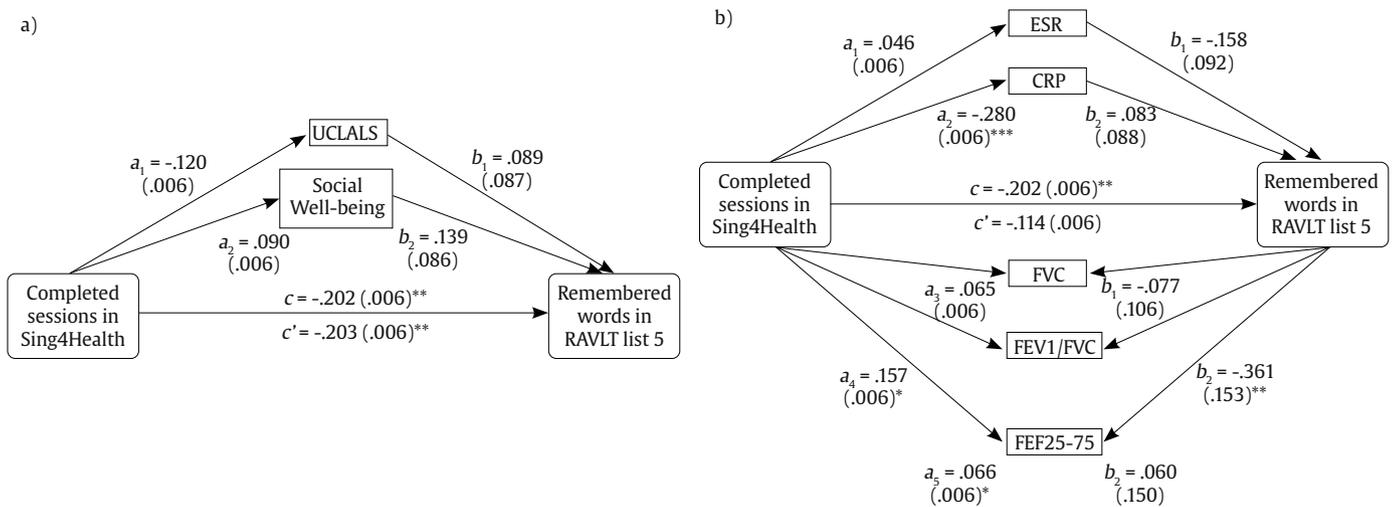


Figure 2. Conceptual diagram showing the multiple mediation models for total, direct and indirect effects of the number of completed sessions in the GS intervention on the standardized residuals of the regression of the number of words correctly remembered at the 5th repetition of the RAVLT between pre- and post-intervention assessments (a) with the social variables as mediators and (b) the physiological variables as mediators. * $p < .050$, ** $p < .025$, *** $p < .01$.

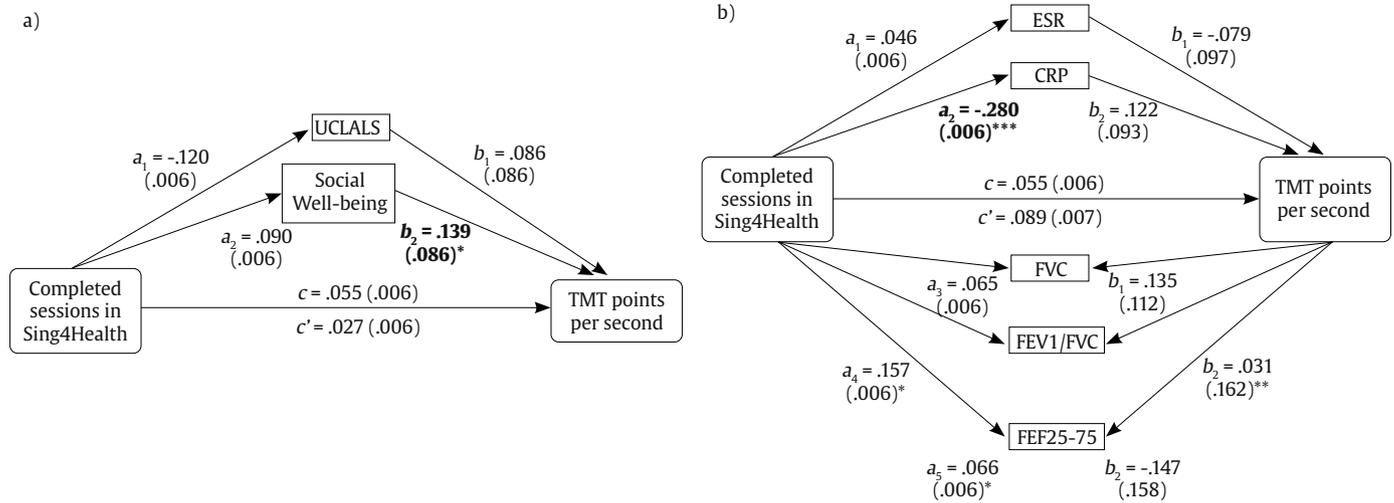


Figure 3. Conceptual diagram showing the multiple mediation models for total, direct and indirect effects of the number of completed sessions in the GS intervention on the standardized residuals of the regression of the number of points per second in the TMT between pre-and post-intervention assessments (a) with the social variables as mediators and (b) the physiological variables as mediators.

* $p < .050$, ** $p < .025$, *** $p < .01$.

Discussion

The present study confirmed that GS interventions showed a positive effect on cognitive (verbal memory), social (social well-being), and physiological (CRP) variables.

Group Singing Intervention Effect on Cognitive, Physiological, and Social Variables

Cognitive function. Verbal memory and EFs were evaluated prior to and after the GS intervention since previous studies showed mixed results on the effects of this kind of interventions on these cognitive domains. In the current study, verbal memory, as measured by RAVLT scores, increased from baseline to post-intervention assessments when considering the whole sample. Further, participants in the IG showed better verbal memory in comparison with an active waiting-list control group at the end of the program's first phase. Interestingly, in mediation analyses the number of GS sessions completed by participants predicted the scores on verbal memory. However, results indicated that taking part in more sessions was related to lower benefits on verbal memory. As regards EFs, measured as performance in the TMT, only a marginally significant effect showing better EFs performance at post-intervention than at baseline assessment.

Present results, therefore, indicate that GS effects on cognition may depend on the cognitive domain evaluated since positive effects were observed for verbal memory while EFs showed little improvement. In contrast with these results, Millett and Fiocco (2020) had 24 residential care users complete 12 weekly 60-minute sessions, including singing and discussion about a given topic, without any change on EFs or verbal memory from pre- to post-intervention assessments. However, present results are consistent with previous evidence provided by Fu et al. (2018), who found increased scores on immediate but not delayed verbal memory in the 10-word recall test in 45 participants of a GS intervention carried out in different senior living communities. In that study, participants completed 12 75-minute weekly sessions, including 10 min of stretching, deep-breathing, and vocal exercise, 50 min of song-singing, and 15 min of informal socialization. Millett and Fiocco's intervention was shorter than the current intervention, but only differed from Fu et al.' intervention in the absence of specific components for breathing training and informal socialization. Therefore, it may be the case

that including respiratory and social activities in GS sessions made them more stimulating for the participants, and in that way cognitive effects of GS interventions are modulated or depend on the inclusion of social and respiratory function aspects in their session.

Interestingly, no difference was observed for EFs performance in Fu et al.'s (2018) study, while in the present study a statistically marginal trend towards better EFs performance in the post-intervention than the baseline assessment was observed. Furthermore, it is worth noting that albeit small and non-significant, the relation between number of sessions and cognitive improvement in EFs was positive. This may indicate that a more intense and longer intervention may be needed to yield positive GS intervention effects on EFs. Taken together with results on verbal memory, it could be the case that verbal memory improvement already occurred earlier in the program development, when initially learning the repertoire, and that later sessions are just accompanied by a more modest improvement in verbal memory and incipient improvements on EFs. Hence, given that health and cognitive integrity may be related to the adherence to the intervention (Dingle et al., 2019) and that cognitive demand may vary according to the specific design of GS intervention sessions, future studies assessing the role of such variables in the link between the number of GS sessions and change in older adults' cognition are warranted.

Physiological health indicators. Regarding physiological effects, GS intervention's impact on low-grade SI seems to be transient and mild. Indeed, no effects were observed for ESR, albeit C-reactive protein levels were shown to be stable in the IG while they increased with time in the active control WLG. Interestingly, the number of sessions attended during the GS intervention predicted the magnitude of change in CRP levels. This way, the more sessions completed, the slower the increase in CRP levels. The IG showed, nonetheless, higher CRP levels at a 6-month follow-up assessment than at baseline and after participation in the intervention.

Previous research found that even a single singing session affected adults' cytokine levels, augmenting its activity (Fancourt et al., 2016). The sample in that study consisted of cancer patients and bereaved and non-bereaved carers. Arguably, such participants may be in situations that increase their stress levels. Authors suggested that singing produces a reduction in cortisol that prompts a subsequent decrease in glucocorticoid suppression, thus enabling an increase in the pro-inflammatory immune activity. Current results showed that

a more sustained (i.e., 34 sessions over 17 weeks) GS intervention might serve to maintain low levels of inflammatory response during participation in the GS intervention, hence, with transient effects that disappear soon after intervention ends. Consequently, the present study showed only a marginal but positive effect of GS interventions on systemic inflammation. Participation in GS may therefore be considered a protective factor against *inflammaging* related processes.

Regarding the respiratory function, IG and active control WLG in the present study differ in FVC and FEF25-75 already at baseline, with levels of these parameters being generally higher in the intervention than the control group in the first study phase. Furthermore, in contrast with our initial hypotheses, all three respiratory function indices used in the study decreased from pre- to post-intervention assessments when considering GS intervention participants in both study phases. Nonetheless, for the IG of the first study phase, expiratory flow and the Tiffeneau-Pinelli index (i.e., FEV1/FVC) showed higher levels at T0 and T1 than at T2, indicating that to a certain degree, the GS intervention prevented a steeper decline of respiratory function with time in the IG. This interpretation should be taken with caution since the comparison between the IG and the WLG did not show a significant interaction between group and time.

Present results are, therefore, at odds with the evidence summarized in recent systematic reviews showing positive effects of singing interventions on the respiratory function across adulthood (Kang et al., 2018) and clinical populations (Goldenberg, 2018). Likewise, they are in contrast with the positive results of a GS intervention including 16 bi-weekly singing sessions with respiratory exercises for older adults in care homes (Mazalli et al., 2019) and a 12-week program including deep-breathing training and song learning and singing for 42 older adults in senior living communities (Fu et al., 2018). It is well known that frequency and intensity of singing activities determine the effect of these activities on respiratory function (Sundberg & Rossing, 1990). Therefore, given that the current intervention is longer than those of previous studies, it could be argued that the main difference should be in the intensity and quality of breathing exercises between interventions or the starting health conditions of participants. Nonetheless, details on respiratory exercises and training provided in previous studies are scarce, and a comparison with the current intervention design difficult. Thus, future detailed and systematic analyses of quantity, quality, and type of breathing, vocal and tonal exercises, and training provided in GS interventions for older adults are needed to shed light on causes behind these contradictory results.

Social indicators. Regarding social aspects, present analyses showed that loneliness decreased during the four months between T0 and T1 assessments regardless of the intervention group, while values on the composite score of SWB were only reduced in the active control WLG.

The effects on loneliness are in partial agreement with Millett and Fiocco's (2020) study, in which a reduction in loneliness was observed when comparing pre- and post-intervention assessments of the GS intervention participants. Previous survey and interview-based studies also reported reduced loneliness resulting from participation in GS interventions (Teater & Baldwin, 2014). In contrast, Davidson et al. (2014) failed to find differences in loneliness scores between pre- and post-intervention assessments for 29 participants who completed 8 weekly 90-minutes sessions, including informal socialization, physical warm-ups, and GS. Nevertheless, the relatively short intervention and small sample size in Davidson et al.'s study, may have hindered their power to detect changes in loneliness. In the present study, loneliness reductions were found regardless of group. It could be argued that social breaks and group dynamics in the current GS intervention were not enough to reduce participants loneliness beyond the fact of being considered and invited to participate in a RCT,

which may, in turn, have had an impact on reducing loneliness also in the WLG. Further, as suggested for the respiratory exercise component of GS intervention, future studies should carefully consider the social components of their interventions in order to provide more data to elucidate the impact of GS interventions on loneliness.

Regarding SWB scores, reductions from T0 to T1 were only observed for the WLG group. Therefore, the lack of significant differences between evaluations observed for older adults completing the GS intervention (i.e., IG) indicates the potential protective effect of the intervention on SWB. Although no previous study of GS interventions assessed SWB using a composite score, Davidson et al.'s (2014) work and research by Pearce et al. (2015) and Pearce et al. (2017) reported older adults in GS activities to have greater satisfaction with their social relationships as well as more interactions and closeness to others. In the present study, such feelings may have been associated with the maintenance of SWB levels in the IG since they tap on some of the five dimensions listed in Keyes' definition of SWB (Keyes, 1998). These results, therefore, warrant future studies that systematically assess the different dimensions of SWB to shed light on the specific benefits that GS interventions convey. Nonetheless, another study from this project will further analyze the effects of GS interventions of social indicators, including loneliness and SWB, testing the mediation effects of social identification with the singing group and subjective well-being (Galinha, García-Martín, et al., 2020).

Relationship between Cognitive, Physiological, and Social Group Singing Intervention Effects

Finally, mediation analyses allowed us to unmask some relations between cognitive, social, and physiological changes associated with the number of sessions completed in the GS intervention. In contrast to our hypothesis, no mediation effects were observed. Still, several relations were found: the magnitude of change on FEV1/FVC between pre- and post-intervention assessments predicted the degree of change in this same period for RAVLT scores. Similarly, magnitude of pre- to post-intervention change on SWB predicted the amount of change in TMT performance between pre- and post-intervention assessments. The former relationship indicates that participants with larger than expected reduction in their respiratory function exhibited a larger than expected increase in their verbal memory scores. The latter relationship demonstrates that participants with lower than expected reductions in SWB showed the largest TMT performance increases. Furthermore, the number of sessions completed in the GS intervention moderates SWB change effects on verbal memory. Likewise, the number of sessions completed in the GS intervention also moderates the aforementioned respiratory function effects on verbal memory, as well as the effects of FE1/FVC on EFs. Hence, these analyses highlighted a map of complex relations between studied variables, indicating that further exploration of the connection between the effects of GS interventions on different cognitive domains and those on respiratory and social domains is needed.

Different factors may limit the impact of the present research. The loss of participants between T1 and T2 or the lack of adherence of several participants to the intervention may have hampered the evaluation or the consolidation of intervention effects. Nonetheless, these facts could be related to the characteristics of the sample. For the sake of ecological validity, the current study's sample includes a diverse community-based elderly group who was not proactively searching for singing activities nor has bonds or formed a choir before the intervention. Indeed, the study includes participants of a low socio-economic background with low levels of formal education and low MoCA scores. These variables and others, such as subjective well-being or perceived physical and mental health, can affect intervention outcomes (for an extended discussion, see Galinha, Fernandes, et al., 2020). Additionally, variables such as

gender distribution in intervention groups may limit generalizability of present results, since more females than males were present in the sample. Consequently, future studies should address the aforementioned and related variables' effects on GS interventions' benefits and outcomes. Future studies will also benefit from more control on the activities the control group and the intervention group are participating in, since unbalanced participation in those activities may bias group comparison results.

In summary, results of the present study confirmed and extend previous findings on the positive effects of GS on cognitive functions and showed preliminary evidence of a protective effect of GS interventions on systemic inflammation. Furthermore, they highlight that different dimensions of these interventions may promote different aspects of cognitive function. Additionally, mediation analyses failed to support the initial hypothesis that social and physiological factors may mediate the effects of GS on cognition, but provided initial evidence supporting the moderating role that engagement in GS interventions may play in the relationship between respiratory function and SWB and cognitive improvement. Consequently, our results highlight the importance of considering informal socialization and breathing exercises in the design of multidimensional GS interventions as a way to maximize their benefits on older adults' cognition, social engagement, and physiological health, thus being a potential cost-effective way to promote cognitive function, social well-being, and health in vulnerable elderly populations from day care centers.

Conflict of Interest

The authors of this article declare no conflict of interest.

Acknowledgments

To Anabela Pires (Singer) for the artistic direction of the intervention and Sérgio Fontão (maestro) and Pedro Baião (pianist) for the singing group intervention implementation. To Maria D'Assis Ribamar and Alexandra Antunes from the SCM of Almada; and Etelvina Ferreira, Neusa Freixinho, Maria Teresa Barata from the SCM of Lisboa for the promotion and logistics support of the intervention, and the recruitment, contacts, transportation, and support of participants' mobility. To the City of Almada and OPART TNSC (National Opera House) for providing theatres and rehearsal rooms.

References

Ahmadi-Abhari, S., Kaptoge, S., Luben, R. N., Wareham, N. J., & Khaw, K. T. (2014). Longitudinal association of C-reactive protein and lung function over 13 years. *American Journal of Epidemiology*, *179*(1), 48-56. <https://doi.org/10.1093/aje/kwt208>

Allan, J. L., McMinn, D., & Daly, M. (2016). A bidirectional relationship between executive function and health behavior: Evidence, implications, and future directions. *Frontiers in Neuroscience*, *10*, 386. <https://doi.org/10.3389/fnins.2016.00386>

Bell-McGinty, S., Podell, K., Franzen, M., Baird, A. D., & Williams, M. J. (2002). Standard measures of executive function in predicting instrumental activities of daily living in older adults. *International Journal of Geriatric Psychiatry*, *17*(9), 828-834. <https://doi.org/10.1002/gps.646>

Bernardi, N. F., Snow, S., Peretz, I., Orozco Perez, H. D., Sabet-Kassouf, N., & Lehmann, A. (2017). Cardiorespiratory optimization during improvised singing and toning. *Scientific Reports*, *7*(1), 8113. <https://doi.org/10.1038/s41598-017-07171-2>

Beydoun, M. A., Weiss, J., Obhi, H. K., Beydoun, H. A., Dore, G. A., Liang, H., Evans, M. K., & Zonderman, A. B. (2019). Cytokines are associated with longitudinal changes in cognitive performance among urban adults. *Brain, Behavior, and Immunity*, *80*, 474-487. <https://doi.org/10.1016/j.bbi.2019.04.027>

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Datos no normales: ¿es el ANOVA una opción válida? *Psicothema*, *29*(4), 552-557. <https://doi.org/10.7334/psicothema.2016.383>

Boss, L., Kang, D.-H., & Branson, S. (2015). Loneliness and cognitive function in the older adult: a systematic review. *International Psychogeriatrics*, *27*(04), 541-553. <https://doi.org/10.1017/S1041610214002749>

Bygren, L. O., Konlaan, B. B., & Johansson, S. E. (1996). Attendance at cultural events, reading books or periodicals, and making music or singing in a choir as determinants for survival: Swedish interview survey of living conditions. *British Medical Journal*, *313*(7072), 1577-1580. <https://doi.org/10.1136/bmj.313.7072.1577>

Cacioppo, J. T., & Hawkley, L. C. (2009). Perceived social isolation and cognition. *Trends in Cognitive Sciences*, *13*(10), 447-454. <https://doi.org/10.1016/j.tics.2009.06.005>

Cavaco, S., Gonçalves, A., Pinto, C., Almeida, E., Gomes, F., Moreira, I., Fernandes, J., & Teixeira-Pinto, A. (2013). Trail Making Test: Regression-based norms for the Portuguese population. *Archives of Clinical Neuropsychology*, *28*(2), 189-198. <https://doi.org/10.1093/arclin/acs115>

Cavaco, S., Gonçalves, A., Pinto, C., Almeida, E., Gomes, F., Moreira, I., Fernandes, J., & Teixeira-Pinto, A. (2015). Auditory verbal learning test in a large nonclinical Portuguese population. *Applied Neuropsychology: Adult*, *22*(5), 321-331. <https://doi.org/10.1080/23279095.2014.927767>

Cavallo, E., Sánchez, G., & Valenzuela, P. (2018). Gone with the wind: Demographic transitions and domestic saving. *Review of Development Economics*, *22*(4), 1744-1764. <https://doi.org/10.1111/rode.12518>

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences. In *Journal of Visual Languages & Computing* (3rd ed.). Lawrence Erlbaum Associates Ltd. <https://doi.org/10.4324/9780203774441>

Coulton, S., Clift, S., Skingley, A., & Rodriguez, J. (2015). Effectiveness and cost-effectiveness of community singing on mental health-related quality of life of older people: Randomised controlled trial. *British Journal of Psychiatry*, *207*(3), 250-255. <https://doi.org/10.1192/bjp.bp.113.129908>

Davidson, J. W., McNamara, B., Rosenwax, L., Lange, A., Jenkins, S., & Lewin, G. (2014). Evaluating the potential of group singing to enhance the well-being of older people. *Australasian Journal on Ageing*, *33*(2), 99-104. <https://doi.org/10.1111/j.1741-6612.2012.00645.x>

Dingle, G. A., Clift, S., Finn, S., Gilbert, R., Groarke, J. M., Irons, J. Y., Bartoli, A. J., Lamont, A., Launay, J., Martin, E. S., Moss, H., Sanfilippo, K. R., Shipton, M., Stewart, L., Talbot, S., Tarrant, M., Tip, L., & Williams, E. J. (2019). An agenda for best practice research on group singing, health, and well-being. *Music & Science*, *2*. <https://doi.org/10.1177/2059204319861719>

Direção Geral da Saúde (2016). Especificações técnicas para a realização de uma espirometria. Orientação da Direção Geral da Saúde. <http://nocs.pt/wp-content/uploads/2017/10/i022868.pdf>

Duggan, E. C., Graham, R. B., Piccinin, A. M., Jenkins, N. D., Clouston, S., Muniz-Terrera, G., & Hofer, S. M. (2020). A systematic review of pulmonary function and cognition in aging. *The Journals of Gerontology: Series B*, *75*(5), 937-952. <https://doi.org/10.1093/geronb/gby128>

Duggan, E. C., Piccinin, A. M., Clouston, S., Koval, A. V., Robitaille, A., Zammit, A. R., Wu, C., Brown, C. L., Lee, L. O., Finkel, D., Beasley, W. H., Kaye, J., Terrera, G. M., Katz, M., Lipton, R. B., Deeg, D., Bennett, D. A., Björk, M. P., Johansson, B., ... Hofer, S. M. (2019). A Multi-study coordinated meta-analysis of pulmonary function and cognition in aging. *The Journals of Gerontology: Series A*, *74*(11), 1793-1804. <https://doi.org/10.1093/gerona/glz057>

Fancourt, D., Williamson, A., Carvalho, L. A., Steptoe, A., Dow, R., & Lewis, I. (2016). Singing modulates mood, stress, cortisol, cytokine and neuropeptide activity in cancer patients and carers. *Ecancermedicalscience*, *10*, 631. <https://doi.org/10.3332/ecancer.2016.631>

Fernández-Ballesteros, R., Robine, J. M., Walker, A., & Kalache, A. (2013). Active aging: A global goal [Special Issue]. *Current Gerontology and Geriatrics Research*, *29*8012. <https://doi.org/10.1155/2013/298012>

Foster, L., & Walker, A. (2015). Active and successful aging: A European policy perspective. *Gerontologist*, *55*(1), 83-90. <https://doi.org/10.1093/geront/gnu028>

Franceschi, C., Bonafé, M., Valensin, S., Olivieri, F., de Luca, M., Ottaviani, E., & de Benedictis, G. (2000). Inflamm-aging – an evolutionary perspective on immunosenescence. *Annals of the New York Academy of Sciences*, *908*, 244-254. <https://doi.org/10.1111/j.1749-6632.2000.tb06651.x>

Franceschi, C., Capri, M., Monti, D., Giunta, S., Olivieri, F., Sevini, F., Panourgia, M. P., Invidia, L., Celani, L., Scurti, M., Cevenini, E., Castellani, G. C., & Salvio, S. (2007). Inflammaging and anti-inflammaging: A systemic perspective on aging and longevity emerged from studies in humans. *Mechanisms of Ageing and Development*, *128*(1), 92-105. <https://doi.org/10.1016/j.mad.2006.11.016>

Frank, M. O., & Caceres, B. A. (2015). Inflammaging: A concept analysis. *The Journal for Nurse Practitioners*, *11*(2), 258-261. <https://doi.org/10.1016/j.nurpra.2014.08.005>

Freitas, S., Simões, M. R., Alves, L., & Santana, I. (2011). Montreal Cognitive Assessment (MoCA): Normative study for the Portuguese population. *Journal of Clinical and Experimental Neuropsychology*, *33*(9), 989-996. <https://doi.org/10.1080/13803395.2011.589374>

Freitas, S., Simões, M. R., Alves, L., & Santana, I. (2012). Montreal cognitive assessment: Influence of sociodemographic and health variables. *Archives of Clinical Neuropsychology*, *27*(2), 165-175. <https://doi.org/10.1093/arclin/acr116>

Fu, M. C., Belza, B., Nguyen, H., Logsdon, R., & Demorest, S. (2018). Impact of group-singing on older adult health in senior living communities:

- A pilot study. *Archives of Gerontology and Geriatrics*, 76, 138-146. <https://doi.org/10.1016/j.archger.2018.02.012>
- Galinha, I. C., Farinha, M., Lima, M. L., & Palmeira, A. L. (2020). Sing4Health: Protocol of a randomized controlled trial of the effects of a singing group intervention on the well-being, cognitive function and health of older adults. *BMC Geriatrics*, 20(1), 354. <https://doi.org/10.1186/s12877-020-01686-6>
- Galinha, I. C., Fernandes, H. M., Lima, M. L., & Palmeira, A. L. (2020). Intervention and mediation effects of a community-based singing group on older adults' perceived physical and mental health: The Sing4Health randomized controlled trial. *Universidade Autónoma de Lisboa*. Manuscript submitted for publication.
- Galinha, I. C., García-Martín, M. A., Lima, M. L., & Palmeira, A. L. (2020). Sing4Health: Randomized controlled trial of a singing group program and the effects in the social and subjective well-being of older adults. *Universidade Autónoma de Lisboa*, Manuscript submitted for publication.
- Goldenberg, R. B. (2018). Singing lessons for respiratory health: A literature review. *Journal of Voice*, 32(1), 85-94. <https://doi.org/10.1016/j.jvoice.2017.03.021>
- Hancox, R. J., Gray, A. R., Sears, M. R., & Poulton, R. (2016). Systemic inflammation and lung function: A longitudinal analysis. *Respiratory Medicine*, 111, 54-59. <https://doi.org/10.1016/j.rmed.2015.12.007>
- Harper, S. (2014). Economic and social implications of aging societies. *Science*, 346(6209), 587-591. <https://doi.org/10.1126/science.1254405>
- Hawkey, L. C., & Cacioppo, J. T. (2010). Loneliness matters: A theoretical and empirical review of consequences and mechanisms. *Annals of Behavioral Medicine*, 40(2), 218-227. <https://doi.org/10.1007/s12160-010-9210-8>
- Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Publications.
- Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, 67(3), 451-470. <https://doi.org/10.1111/bmsp.12028>
- Hayes, A. F., & Rockwood, N. J. (2017). Regression-based statistical mediation and moderation analysis in clinical research: Observations, recommendations, and implementation. *Behaviour Research and Therapy*, 98, 39-57. <https://doi.org/10.1016/j.brat.2016.11.001>
- Heringa, S. M., van den Berg, E., Reijmer, Y. D., Nijpels, G., Stehouwer, C. D. A., Schalkwijk, C. G., Teerlink, T., Scheffer, P. G., van den Hurk, K., Kappelle, L. J., Dekker, J. M., & Biessels, G. J. (2014). Markers of low-grade inflammation and endothelial dysfunction are related to reduced information processing speed and executive functioning in an older population – the Hoorn Study. *Psychoneuroendocrinology*, 40, 108-118. <https://doi.org/10.1016/j.psyneuen.2013.11.011>
- Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 82(400), 1147-1149. <https://doi.org/10.1080/01621459.1987.10478551>
- Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81(396), 991-999. <https://doi.org/10.1080/01621459.1986.10478363>
- Johnson, J. K., Louhivuori, J., Stewart, A. L., Tolvanen, A., Ross, L., & Era, P. (2013). Quality of life (QOL) of older adult community choral singers in Finland. *International Psychogeriatrics*, 25(7), 1055-1064. <https://doi.org/10.1017/S1041610213000422>
- Kang, J., Scholp, A., & Jiang, J. J. (2018). A Review of the physiological effects and mechanisms of singing. *Journal of Voice*, 32(4), 390-395. <https://doi.org/10.1016/j.jvoice.2017.07.008>
- Kelly, M. E., Duff, H., Kelly, S., McHugh Power, J. E., Brennan, S., Lawlor, B. A., & Loughrey, D. G. (2017). The impact of social activities, social networks, social support and social relationships on the cognitive functioning of healthy older adults: A systematic review. *Systematic Reviews*, 6(1), 259. <https://doi.org/10.1186/s13643-017-0632-2>
- Keyes, C. L. M. (1998). Social well-being. *American Sociological Association*, 61(2), 121-140. <https://www.jstor.org/stable/2787065>
- Khan, A., & Rayner, G. (2003). Robustness to non-normality of common tests for the many-sample location problem. *Journal of Applied Mathematics and Decision Sciences*, 7(4), 187-206. https://doi.org/10.1207/s15327612jamd0704_1
- Lages Silva, A. (2016). *Bem-estar social na idade adulta: adaptação e validação das Social Well-being Scales no contexto português*. Universidade Lusófona do Porto.
- Lee, G. R., & Ishii-Kuntz, M. (1987). Social interaction, loneliness, and emotional well-being among the elderly. *Research on Aging*, 9(4), 459-482. <https://doi.org/10.1177/0164027587094001>
- Liao, J., Muniz-Terrera, G., Head, J., & Brunner, E. J. (2016). Dynamic longitudinal associations between social support and cognitive function: A prospective investigation of the directionality of associations. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 73(7), 1233-1243. <https://doi.org/10.1093/geronb/gbw135>
- Mansens, D., Deeg, D. J. H., & Comijs, H. C. (2018). The association between singing and/or playing a musical instrument and cognitive functions in older adults. *Aging and Mental Health*, 22(8), 964-971. <https://doi.org/10.1080/13607863.2017.1328481>
- Mazalli, J., Petreça, D. R., & Campos, R. (2019). O uso do canto como terapêutica na reabilitação respiratória em idosos institucionalizados. *Journal Health NPEPS*, 4(1), 31-46. <https://doi.org/10.30681/252610103318>
- Miceli, S., Maniscalco, L., & Matranga, D. (2019). Social networks and social activities promote cognitive functioning in both concurrent and prospective time: Evidence from the SHARE survey. *European Journal of Ageing*, 16(2), 145-154. <https://doi.org/10.1007/s10433-018-0486-z>
- Miller, A., Green, M., & Robinson, D. (1983). Simple rule for calculating normal erythrocyte sedimentation rate. *British Medical Journal*, 286(6361), 266-266. <https://doi.org/10.1136/bmj.286.6361.266>
- Miller, M. R., Crapo, R., Hankinson, J., Brusasco, V., Burgos, F., Casaburi, R., Coates, A., Enright, P., van der Grinten, C. P. M., Gustafsson, P., Jensen, R., Johnson, D. C., MacIntyre, N., McKay, R., Navajas, D., Pedersen, O. F., Pellegrino, R., Viegi, G., & Wanger, J. (2005). General considerations for lung function testing. *European Respiratory Journal*, 26(1), 153-161. <https://doi.org/10.1183/09031936.05.00034505>
- Millett, G., & Fiocco, A. J. (2020). A pilot study implementing the JAVA Music Club in residential care: Impact on cognition and psychosocial health. *Aging & Mental Health*. Advance online publication. <https://doi.org/10.1080/13607863.2020.1758919>
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695-699. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>
- Nerpin, E., Jacinto, T., Fonseca, J. A., Alving, K., Janson, C., & Malinovsky, A. (2018). Systemic inflammatory markers in relation to lung function in NHANES, 2007-2010. *Respiratory Medicine*, 142, 94-100. <https://doi.org/10.1016/j.rmed.2018.07.011>
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, 17(2), 299-320. <https://doi.org/10.1037/0882-7974.17.2.299>
- Pearce, E., Launay, J., & Dunbar, R. I. M. (2015). The ice-breaker effect: Singing mediates fast social bonding. *Royal Society Open Science*, 2(10), 150221. <https://doi.org/10.1098/rsos.150221>
- Pearce, E., Launay, J., Maccarron, P., & Dunbar, R. I. M. (2017). Tuning in to others: Exploring relational and collective bonding in singing and non-singing groups over time. *Psychology of Music*, 45(4), 496-512. <https://doi.org/10.1177/0305735616667543>
- Pereira Vilar, M. M. (2015). *Avaliação da qualidade de vida em adultos idosos: Estudos de adaptação, validação e normalização do WHOQOL-OLD para a população portuguesa*. Universidade de Coimbra.
- Pocinho, M., Farate, C., & Dias, C. A. (2010). Validação psicométrica da escala UCLA-loneliness para idosos portugueses. *Interações: Sociedade e Novas Modernidades*, 18(18), 65-77. <http://www.interacoes-ismt.com/index.php/revista/article/view/A08J/304>
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879-891. <https://doi.org/10.3758/BRM.40.3.879>
- Rosselli, M., & Torres, V. L. (2019). Executive dysfunction during normal and abnormal aging. In A. Ardila, S. Fatima, & M. Rosselli (Eds.), *Dysexecutive syndromes* (pp. 155-175). Springer International Publishing. https://doi.org/10.1007/978-3-030-25077-5_8
- Royall, D. R., Palmer, R., Chiodo, L. K., & Polk, M. J. (2004). Declining executive control in normal aging predicts change in functional status: The Freedom House Study. *Journal of the American Geriatrics Society*, 52(3), 346-352. <https://doi.org/10.1111/j.1532-5415.2004.52104.x>
- Sullivan, T. R., White, I. R., Salter, A. B., Ryan, P., & Lee, K. J. (2018). Should multiple imputation be the method of choice for handling missing data in randomized trials? *Statistical Methods in Medical Research*, 27(9), 2610-2626. <https://doi.org/10.1177/0962280216683570>
- Sundberg, J., & Rossing, T. D. (1990). The science of singing voice. *The Journal of the Acoustical Society of America*, 87(1), 462. <https://doi.org/10.1121/1.399243>
- Teater, B., & Baldwin, M. (2014). Singing for successful ageing: The perceived benefits of participating in the golden oldies community-arts programme. *British Journal of Social Work*, 44(1), 81-99. <https://doi.org/10.1093/bjsw/bcs095>
- Tegeler, C., O'Sullivan, J. L., Bucholtz, N., Goldeck, D., Pawelec, G., Steinhagen-Thiessen, E., & Demuth, I. (2016). The inflammatory markers CRP, IL-6, and IL-10 are associated with cognitive function-data from the Berlin Aging Study II. *Neurobiology of Aging*, 38, 112-117. <https://doi.org/10.1016/j.neurobiolaging.2015.10.039>
- Tilvis, R. S., Kahonen-Vare, M. H., Jolkkonen, J., Valvanne, J., Pitkala, K. H., & Strandberg, T. E. (2004). Predictors of cognitive decline and mortality of aged people over a 10-year period. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 59(3), M268-M274. <https://doi.org/10.1093/gerona/59.3.M268>
- Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 33(1), 1-67. <https://doi.org/10.1214/aoms/1177704711>
- United Nations. (2019). World population prospects 2019. In *Department of Economic and Social Affairs. World Population Prospects 2019* (Issue 141). <http://www.ncbi.nlm.nih.gov/pubmed/12283219>

- Vaz Serra, A., Canavarro, M. C., Simões, M. R., Pereira, M., Gameiro, S., Quartilho, M. J., Rijo, D., Carona, C., & Paredes, T. (2006). Estudos psicométricos do Instrumento de Avaliação da Qualidade de Vida da Organização Mundial de Saúde (WHOQOL-Bref) para Português de Portugal. *Psiquiatria Clínica*, 27, 41-49.
- Walker, K. A., Gottesman, R. F., Wu, A., Knopman, D. S., Gross, A. L., Mosley, T. H., Selvin, E., & Windham, B. G. (2019). Systemic inflammation during midlife and cognitive change over 20 years: The ARIC Study. *Neurology*, 92(11), E1256-E1267. <https://doi.org/10.1212/WNL.0000000000007094>
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third edition* (3rd ed.). The Psychological Corporation.
- Williams, E., Dingle, G., Clift, S. (2018). A systematic review of mental health and well-being outcomes of group singing for adults with a mental health condition. *European Journal of Public Health*, 28(6), 1035 -1042. <https://doi.org/10.1093/eurpub/cky115>
- Wilson, R. S., Krueger, K. R., Arnold, S. E., Schneider, J. A., Kelly, J. F., Barnes, L. L., Tang, Y., & Bennett, D. A. (2007). Loneliness and risk of Alzheimer disease. *Archives of General Psychiatry*, 64(2), 234-240. <https://doi.org/10.1001/archpsyc.64.2.234>

Appendix

Supplementary Information

Sing4Health session structure

The intervention comprised 34 sessions delivered twice a week for 2 hours including group singing and breaks for socializing. The sessions were carried out in a rehearsal room at local theaters in Almada (Portugal) and Lisbon (Portugal) as well as on the OPART TNSC (National Opera House) in Lisbon (Portugal). The rooms used for the intervention have capacity for 35 seated individuals and an additional space for exercises and activities that require participants to stand up or move. The equipment in the room included a piano, files with lyrics, and water bottles for each participant. Further, for social components of the intervention a separate coffee-break room was used.

Intervention sessions were guided by the intervention team, which includes a maestro (Master in music specialized in choral direction by the Escola Superior de Música de Lisboa) and a pianist (Master in jazz performance by the Royal Welsh College of Music and Drama). Further, the intervention involved a substitute for the maestro (Degree in music in the community by the Escola Superior de Educação e Música de Lisboa) as well as an artistic director (professional singer and Master in psychology by the Universidade Autónoma de Lisboa), an invited artist (tenor resident opera singer of Teatro Nacional de São Carlos), older adults' assistants (Degree in social education, specialized in rehabilitation), and older adults' carers (professional carers from the daycare centers).

Each of the sessions in the intervention included the following components:

- Physical exercises for relaxation and vocal warm up; participants start with 15 mins of relaxation exercises and vocal warm-up exercises in which physical aspects of group singing are practiced and discussed to promote body awareness, healthy body posture, and understanding of effective breathing during singing. The sessions include sitting and standing up periods, as well as movements to follow the rhythm and mobility around the room to sing. This component was led by the intervention team (i.e., pianist, maestro) aided by the older adults' assistants.
- Vocal technique training; since no musical experience was required, participants were taught fundamental aspects of singing, sound, voice mechanics, and group singing. Such aspects include correct and healthy vocal production skills, the use of breathing techniques and listening skills. Methodologies to teach participants how to sing in different parts or voices were based on hearing and repetition. Participants were also encouraged to focus and learn to synchronize with each other. This component was led by the intervention team aided by older adults' assistants.
- Rehearsal of a repertoire, which involve memorization and interpretation of songs and lyrics using different strategies (e.g.,

singing lyrics or musical notation, call and response methods, etc.). This component also involved reviewing previously learned songs and learning to harmonize the individual voices to sound as one. The repertoire included songs proposed by the intervention team and the participants, which were selected by the whole group after discussion. The component is led by the intervention team aided by older adults' assistants.

- Social dynamics and socializing breaks; a break of about 20 minutes for a snack and socialization was included. Social interaction was promoted encouraging group members to share emotions, ideas, and expectations about participation in the intervention, along with group dynamics, work oriented towards a common goal, and group discussions about the meaning and cultural and historical context of the songs. This component was also led by the intervention team aided by older adults' assistants.
- Creation and presentation of a public performance; the final goal for the participants was to prepare a final performance for the general public, for which family members and older adults from other daycare centers were invited. Further, a singer was invited to sing with the group in this public performance. This component was also led by the intervention team aided by the older adults' assistants.
- Evaluation of participants performance; participants' ability for tuning was assessed at the beginning and at the end of the program. Additionally, the intervention team monitored the evolution of each participant's skills, attendance, and achievement of objectives. This component was led by the intervention team with the aid of assistant researchers and the older adults' assistants.

Results following a per Protocol Approach

Statistical analyses following a per protocol approach were run with data from 110 participants: 57 in the WLG and 53 in the IG. As for the intention to treat approach, an alpha level of .05 was maintained for all tests, which were performed using IBM SPSS Statistics 21 and the PROCESS macro for mediation and moderation analyses (Hayes, 2017; Hayes & Preacher, 2014; Hayes & Rockwood, 2017). Complete results could be found in Tables A1 to A3. Below, we report the main differences in the findings.

Comparison of the IG with the WLG. For the composite score of SWB, the factor Group additionally showed a significant main effect, $F(1,108) = 4.032, p = .047, \eta^2 = .036$, indicating that the IG has higher SWB than the WLG regardless of the time (T0 or T1).

For the CRP levels, the IG showed lower scores than the WLG at T1 ($p = .050, \eta^2 = .035$) in a significant interaction between group and time, $F(1, 108) = 5.650, p = .019$.

Table A1. Group Singing Intervention Effects Comparison with an Active Control Group as per Protocol

	IG (n = 53)		WLG (n = 57)		Group			Time			Interaction		
	T0	T1	T0	T1	F(1, 108)	p	η^2	F(1, 108)	p	η^2	F(1, 108)	p	η^2
TMT	0.30 ± 0.18	0.34 ± 0.19	0.29 ± 0.16	0.30 ± 0.18	0.49	.484	.005	3.71	.057	.033	1.17	.282	.011
RAVLT	8.30 ± 3.03	8.23 ± 2.84	7.75 ± 3.42	6.57 ± 3.70	3.72	.056	.033	6.39	.013	.056	5.05	.027	.045
UCLALS	8.85 ± 3.24	7.84 ± 2.91	9.07 ± 3.26	8.58 ± 2.33	0.97	.328	.009	7.40	.008	.064	0.91	.342	.008
SWB	49.38 ± 6.33	49.83 ± 6.89	48.88 ± 7.85	45.63 ± 6.35	4.03	.047	.036	5.40	.022	.048	9.44	.003	.080
ESR	62.17 ± 42.60	66.87 ± 40.20	64.73 ± 39.81	68.30 ± 39.85	0.08	.772	.001	1.37	.245	.012	0.03	.874	0.001
CRP	0.39 ± 0.60	0.28 ± 0.57	0.32 ± 0.55	0.52 ± 0.69	0.79	.376	.007	0.51	.478	.005	5.65	.019	.050
FVC	61.91 ± 11.74	61.42 ± 12.42	54.84 ± 12.27	56.72 ± 11.97	8.47	.004	.073	0.39	.532	.004	1.13	.290	.010
FEV1/FVC	107.89 ± 13.01	105.38 ± 11.00	104.65 ± 11.99	103.36 ± 12.11	2.00	.160	.018	1.97	.164	.018	0.21	.652	.002
FEF25-75	90.19 ± 36.70	87.51 ± 35.12	69.81 ± 31.14	70.72 ± 27.50	12.91	< .001	.107	0.06	.800	.001	0.27	.607	.002

Table A2. Baseline Assessment Comparison against post Group Singing Intervention Assessment as per Protocol

<i>n</i> = 78	Pre-intervention		Post-intervention		Wilcoxon Signed-Rank test	
	Median	Median	<i>Z</i>	<i>p</i> -value	<i>r</i>	
TMT	0.30	0.33	-2.52	.012	-.202	
RAVLT	8	8	-0.99	.320	-.079	
UCLA	8	8	-2.12	.034	-.170	
SWB	49.00	49.50	-1.63	.104	-.131	
ESR	62.57	54.05	-0.124	.901	-.010	
CRP	0.57	0.42	-0.056	.955	-.004	
FVC	59.78	58.53	-1.479	.139	-.118	
FEV1/FVC	106.00	105.00	-2.182	.029	-.175	
FEF25-75	78.68	72.00	-2.044	.041	-.164	

Table A3. Comparison between T0, T1 and T2 Scores in the Experimental Variables to Assess Maintenance of Intervention Effects over Time as per Protocol

<i>n</i> = 53	T0		T1		T2		Time		
	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>	<i>F</i> (2, 104)	<i>p</i> -value	η^2
TMT	0.30	0.18	0.34	0.19	0.31	0.19	2.86	.062	.052
RAVLT	8.30	3.03	8.23	2.84	8.01	2.99	0.36	.698	.007
UCLA	8.85	3.24	7.84	2.91	8.34	3.07	4.18	.018	.074
SWB	49.38	6.33	49.83	6.89	49.90	7.47	0.25	.781	.005
ESR	62.17	42.60	66.87	40.20	56.63	32.66	1.73	.182	.032
CRP	0.39	0.60	0.28	0.57	0.44	0.56	1.88	.158	.035
FVC	61.91	11.74	61.42	12.42	60.28	15.49	0.54 ¹	.546	.010
FEV1/FVC	107.89	13.01	105.38	11.00	89.75	21.47	27.53¹	< .001	.346
FEF25-75	90.19	36.70	87.51	35.11	64.46	33.67	15.91¹	< .001	.234

¹Greenhouse-Geisser correction was applied for FVC (degrees of freedom: 1.597, 83.064, $\epsilon = 0.799$), for FEV1/FVC (degrees of freedom: 1.717, 89.295, $\epsilon = 0.859$), and for FEF25-75 (degrees of freedom: 1.742, 90.560, $\epsilon = 0.871$).

Baseline assessment comparison with post-intervention assessment. Regarding RAVLT scores, there was no significant difference between pre- and post- intervention scores ($Z = -0.995$, $p = .320$). For UCLALS scores there was a significant main effect of time, indicating loneliness were higher at pre- than at post-intervention assessments ($Z = -2.117$, $p = .034$, $r = -.169$). For FVC there were no significant effects of Time ($Z = -1.479$, $p = 0.139$).

Maintenance of effects in the IG at follow up. UCLALS scores were lower at T1 than at T0, $F(2, 104) = 4.178$, $p = .018$; post hoc

comparison, $p = .011$, $\eta^2 = .151$. For CRP levels, there was no main effect of time, $F(2, 104) = 1.879$, $p = .158$.

Relationship between the cognitive, physiological and social GS intervention effects. None of the regressions, moderations, or mediations studied in the exploratory mediation analyses were significant when run following a per-protocol approach. This fact may indicate a lack of statistical power or that those effects are associated with participants that do not fully complete the intervention.