Music Therapy in Infancy and Neurodevelopmental Outcomes in Preterm Children
A Secondary Analysis of the LongSTEP Randomized Clinical Trial

Łucja Bieleninik, Dr habil; Ingrid Kvestad, PhD; Christian Gold, PhD; Andreas Størksen Stordal, PhD; Jörg Assmus, PhD; Shmuel Arnon, MD; Cochavit Elefant, PhD; Mark Ettenberger, PhD; Tora Söderström Gaden, PhD; Dafna Haar-Shamir, MA; Tonje Håvardstun, PsyD; Marcela Lichtensztejn, MA; Julie Mangersnes, MA; Anne-Marthe Nygård Wiborg, PsyD; Bente Johanne Vederhus, PhD; Claire M. Ghetti, PhD

Abstract

IMPORTANCE Preterm children are at risk for neurodevelopment impairments.

OBJECTIVE To evaluate the effect of a music therapy (MT) intervention (parent-led, infant-directed singing) for premature children during the neonatal intensive care unit (NICU) stay and/or after hospital discharge on language development at 24 months' corrected age (CA).

DESIGN, SETTING, AND PARTICIPANTS This predefined secondary analysis followed participants in the LongSTEP (Longitudinal Study of Music Therapy's Effectiveness for Premature Infants and Their Caregivers) randomized clinical trial, which was conducted from August 2018 to April 2022 in 8 NICUs across 5 countries (Argentina, Colombia, Israel, Norway, and Poland) and included clinic follow-up visits and extended interventions after hospital discharge.

INTERVENTION Participants were children born preterm (<35 weeks' gestation) and their parents. Participants were randomized at enrollment to MT with standard care (SC) or SC alone; they were randomized to MT or SC again at discharge. The MT was parent-led, infant-directed singing tailored to infant responses and supported by a music therapist and was provided 3 times weekly in the NICU and/or in 7 sessions across 6 months after discharge. The SC consisted of early intervention methods of medical, nursing, and social services, without MT.

MAIN OUTCOME AND MEASURES Primary outcome was language development, as measured by the Bayley Scales of Infant and Toddler Development, Third Edition (BSID-III) language composite score, with the remaining BSID-III composite and subscale scores as the secondary outcomes. Group differences in treatment effects were assessed using linear mixed-effects models using all available data.

RESULTS Of 206 participants (103 female infants [50%]; mean [SD] GA, 30.5 [2.7] weeks), 51 were randomized to MT and 53 to SC at enrollment; at discharge, 52 were randomized to MT and 50 to SC. A total of 112 (54%) were retained at the 24 months' CA follow-up. Most participants (79 [70%] to 93 [83%]) had BSID-III scores in the normal range (≥85). Mean differences for the language composite score were −2.36 (95% CI, −12.60 to 7.88; *P* = .65) for the MT at NICU with postdischarge SC group, 2.65 (95% CI, −7.94 to 13.23; *P* = .62) for the SC at NICU and postdischarge MT group, and −3.77 (95% CI, −13.97 to 6.43; *P* = .47) for the MT group at both NICU and postdischarge. There were no significant effects for cognitive or motor development.

CONCLUSIONS AND RELEVANCE This secondary analysis did not confirm an effect of parent-led, infant-directed singing on neurodevelopment in preterm children at 24 months' CA; wide CIs suggest, however, that potential effects cannot be excluded. Future research should determine the

Key Points

Question What is the effect of parent-led, music therapist–supported, infant-directed singing, initiated during the neonatal intensive care unit stay and/or through 6 months after hospital discharge, on language development at 24 months' corrected age?

Findings In this secondary analysis of a randomized clinical trial involving 112 children, receiving music therapy along with standard care up to 6 months of age led to language development at 24 months' corrected age that was similar to development in children who received standard care only.

Meaning Findings of this study suggest that, although safe and well accepted by parents, parent-led, infant-directed singing compared with standard care did not result in greater language development benefits at 24 months' corrected age in preterm children.

+ **Supplemental content**

Author affiliations and article information are listed at the end of this article.

(continued)

Abstract (continued)

MT approaches, implementation time, and duration that are effective in targeting children at risk for neurodevelopmental impairments and introducing broader measurements for changes in brain development.

TRIAL REGISTRATION ClinicalTrials.gov Identifier: NCT03564184

Introduction

Preterm birth is a major public health challenge worldwide affecting approximately 15 million neonates and their parents each year.[1] Prematurity has a heavy burden on mortality, short-term severe morbidity, as well as long-term mental health and neurodevelopment of survivors.[2] Although approximately 90% of preterm newborns survive without neurodevelopmental impairments,[2,3] the brain is susceptible to dysfunction after preterm birth. Known neurodevelopmental impairments include poor cognitive and behavioral development[4,5] and worse academic outcomes during childhood,[6,7] posing challenges in adulthood.[7,8]

After preterm birth, the overwhelming auditory environment of the neonatal intensive care unit (NICU) may adversely influence plasticity of the auditory brain system, increasing the risk for language, attention, and cognitive deficits in preterm infants.[9,10] Conversely, early exposure to adult speech in the NICU has been associated with better language and cognitive development at 7 and 18 months in preterm children.[11] Parental voice benefits preterm infant vocalization,[12] and infant-directed singing specifically captures infant attention and promotes expressive vocabulary and regulation.[13-17] Meta-analyses have shown beneficial short-term effects of music therapy (MT) on infant heart rate, respiratory rate, oral feeding volume, and stress level.[18,19] An increasing number of studies describe the role of MT in promoting beneficial neuroplasticity in preterm children,[20-22] and recent evidence shows promising short-term neurophysiological outcomes.[23-25] Three studies examined the effect of music interventions on neurodevelopment in preterm infants.[26-28] Although all 3 studies suggested no effect on neurodevelopmental outcomes, 1 study demonstrated the benefit of early music listening for emotional outcomes at ages 12 and 24 months.[28] All 3 studies used MT intervention delivered during NICU hospitalization. We could not identify any studies that used the Bayley Scales of Infant and Toddler Development, Third Edition (BSID-III) to examine the effect of MT delivered both during NICU hospitalization and after discharge from the hospital.

The Longitudinal Study of Music Therapy's Effectiveness for Premature Infants and Their Caregivers (LongSTEP) was a pragmatic randomized clinical trial (RCT) that evaluated the effects of MT through parent-led, infant-directed singing in the NICU and/or through 6 months after discharge.[29] Previously reported results of this trial indicated no clinically important effects on primary (parent-infant bonding) or secondary (parental stress and mental health and child development) outcomes at 6 months' and 12 months' corrected age (CA).[30] The current predefined secondary analysis evaluated longer-term effects on preplanned secondary outcomes: neurodevelopment at 24 months' CA as measured by the BSID-III. The aim of this study was to evaluate the effect of the MT intervention during NICU stay and/or after hospital discharge on language development at 24 months' CA. We predefined language development, as measured by the BSID-III language composite score, as the primary outcome; the remaining BSID-III composite and subscale scores were the secondary outcomes. We hypothesized that receiving music therapist–supported, parent-led, infant-directed singing would enable attuned musical interactions that promote mutual regulation and thereby facilitate neurodevelopment over time, leading to improved language, cognitive, and motor outcomes at 2 years of age.

Methods

Study Design and Participants

Children and their parents participated in the LongSTEP study, a 2 × 2 factorial, international, multicenter, assessor-blind pragmatic RCT conducted in 7 level-III and 1 level-IV NICUs across Argentina, Colombia, Israel, Norway, and Poland from August 2018 to April 2022. The original trial protocol is provided in Supplement 1. The intervention protocol, baseline characteristics of participants, and results from 2 independent feasibility studies have been published elsewhere.[29,31-34] The Regional Committees for Medical and Health Research Ethics in Norway and the relevant ethics committees in each participating country approved the study. Written informed consent was obtained from parents or caregivers. We followed the Consolidated Standards of Reporting Trials (CONSORT) reporting guideline.

Between August 2018 and April 2020, we approached and recruited preterm infants and their parents during their NICU stay. Eligible infants were born before 35 weeks' gestation, determined by medical staff to be sufficiently medically stable to start the intervention, and likely to stay at the NICU longer than 2 weeks from the time of recruitment. Eligible parents were those who could provide written, site-specific informed consent; were willing to actively participate in MT sessions; lived within reasonable commuting distance from the NICU to facilitate participation after discharge; and had sufficient understanding of the respective national language and mental capacity to complete the intervention and questionnaires.[29]

Randomization

Eligible participants were enrolled and underwent baseline assessments. The LongSTEP trial's factorial design included 2 phases: NICU hospitalization and follow-up after hospital discharge, with 2 randomization time points. At the NICU, participants were first randomized to MT plus standard care (SC) or SC alone. A computer-generated randomization list with a 1:1 ratio, in block sizes of 2 or 4 varying randomly, was used and stratified by site. The randomization sequence was generated by a researcher with no clinical involvement (C. Gold) and distributed using an online system (Sealed Envelope; Sealed Envelope Ltd). The allocation status was entered into a web-based system for electronic data capture and clinical data management for multicenter clinical trials (OpenClinica; OpenClinica, LLC). In cases of multiple pregnancies, we randomized and included only the first-born infant while all multiple-birth siblings received the same intervention. Before discharge, participants were randomized again to postdischarge MT or SC in a 1:1 ratio, using the same procedures.

Procedures

MT Intervention

The approach to MT, parent-led, infant-directed singing, in the LongSTEP trial[29] was individualized to each family and focused on mutual beneficial effects for both children and parents. The approach builds on MT theory[35-38] and research[39,40] integrated with concepts from developmental psychology and neuroscience.[33] It is (1) resource oriented; (2) a musical dialogic process, with therapeutic intent aimed at the parent-infant dyad or triad; (3) culturally relevant; (4) focused on a high level of parent involvement to foster parental identity; (5) individualized; (6) progressively developmentally appropriate based on the infant's emerging awareness and responsiveness over time; and (7) infant directed.[33] Music was adapted to the parent's culture, musical preferences, and parental abilities but was still guided by specific principles to ensure similar intervention for the included participants.[33,41]

The MT intervention consisted of 3 weekly individual sessions of approximately 30 minutes each (with a maximum of 27 sessions) during NICU hospitalization and 7 individual sessions of approximately 45 minutes each across the first 6 months after discharge to home.[29] Interventions were provided by 11 music therapists with or in the final stage of completing a master's degree in MT. These music therapists had clinical experience and training in the use of MT in neonatal contexts and

received training on the protocol and supervision from the study core team to enhance protocol adherence. Treatment fidelity evaluations demonstrated that music therapists across the sites delivered the intervention with satisfactory adherence to the guiding principles.[41]

Standard Care

Standard care included early intervention methods of medical, nursing, and social services, with the exception of MT approaches. Families were asked not to engage in MT outside of the study context through the 12-month study period.[29]

Outcomes and Blinding

This analysis used baseline and discharge data[32] along with neurodevelopmental data gathered at 24 months' CA. As mentioned, the BSID-III language composite score was set as the primary outcome, and the other composite and scaled scores were the secondary outcomes.

Neurodevelopment was assessed using the BSID-III when the children were approximately 2 years of age.[42] The BSID-III, an individually administered instrument used worldwide for children up to 42 months of age, is considered the gold standard for evaluating early child development in clinical work and research[43] and is used extensively with preterm infants.[44] The BSID-III includes 3 assessor-evaluated developmental domains: cognitive, language, and motor, which are expressed in composite scores standardized with a population mean (SD) of 100 (15). The BSID-III also includes subscales (receptive and expressive language; fine and gross motor), which are expressed in scaled scores ranging from 1 to 19, with a population mean (SD) of 10 (3). Higher scores indicate better development. Two parent-reported BSID-III subscales on adaptive and socioemotional development were not included in the current study.

Assessments were completed by 11 qualified independent assessors (2 from Poland, 3 from Israel, 1 from Argentina, 1 from Colombia, and 4 from Norway) following the manual without local adjustments to the items. All, except for the Polish assessors, had prior experience with conducting BSID-III assessments. Prior to the study, assessors received individualized virtual training, including standardization procedures (eAppendix in Supplement 2), from senior psychologists (Ł.B., I.K.). Assessments were scheduled at 24 months' CA (within a 2-week window). Parents were approached by the site investigator or assessors by post, email, and/or telephone or text messaging approximately 4 weeks before the infant reached 24 months' CA to schedule an appointment within the target window period. Due to COVID-19–related challenges, the window period was extended to 32 months' CA. Age was adjusted for prematurity for all, also for those older than 24 months at assessment,[45] when converting raw scores into scaled and composite scores using the US norms. Assessments were completed at the hospital, clinic, research institution, or in exceptional cases the family's home. Regardless of location, all aspects of the environment were controlled as much as possible in order to obtain a reliable assessment of the child's development. Parents or caregivers were present during the assessment. When assessments were completed, the assessor provided feedback and recommendations to parents according to local standards, ensuring appropriateness to the participant's culture.

Assessments were performed by assessors who were blind to participants' group randomization. After assessments were completed, assessors were asked whether they inadvertently had become aware of the randomization. Success of blinding was verified in all assessments.

Statistical Analysis

The target sample size was calculated for the main outcome of the LongSTEP trial (parent-infant bonding),[29] with 80% power for a medium effect size (Cohen d = 0.5) and 20% attrition in a 2-sided test with 5% significance level. Therefore, the present study would also be able to detect medium-sized effects (corresponding with a mean [SD] score of 7.5 [15] points) on the BSID-III.

Descriptive statistics for participants with data at follow-up (mean [SD]; No. [%]) were generated for all variables included in this study. Treatment effects were investigated using linear mixed-effects models (LMEs) with site as a random effect. All tests were 2-sided with a $P < .05$ significance level. LMEs make optimal use of all available data, providing the basis for the main intention-to-treat analysis. Sensitivity intention-to-treat analysis was based on multiply imputed complete datasets. Post hoc, we explored clinical and demographic factors of language composite scores and associated intervention effects in LMEs. We conducted analyses using R, version 4.3.0 (R Project for Statistical Computing).

Results

A total of 206 families consented to participate and were randomized to 4 intervention groups at enrollment (51 to MT with SC; 53 to SC at NICU) and discharge (52 to MT; 50 to SC). Of the original sample (103 females [50%], 103 males [50%]; mean [SD] gestational age, 30.5 [2.7] weeks; mean [SD] birth weight, 1400.5 [432.8] grams), 112 (54%) were included in the follow-up at 24 months' CA (mean [SD] CA at assessment, 25.54 [1.72] months) (**Figure 1**). Baseline characteristics of children and parents were well balanced between the intervention groups (**Table 1**)[46]; we found no difference between those who were followed up and those who were lost to follow-up (eTable 1 in Supplement 2). Retention rates varied across countries but not by intervention group; 1 site had high retention (12 of 54 [22%]), which shows that it is possible to maintain contact with families over a long period (eTable 2 in Supplement 2). Intervention statistics and adverse event data are published elsewhere.[30]

Observed values of the BSID-III composite and subscale scores in the 4 intervention groups were similar (**Table 2**; **Figure 2**). Across groups, the mean (SD) language composite score was 94.7 (19.3), with a range from 53 to 135. The mean (SD) for the cognitive composite score was 100.8

Figure 1. Study Flow Diagram of Families in the LongSTEP Trial
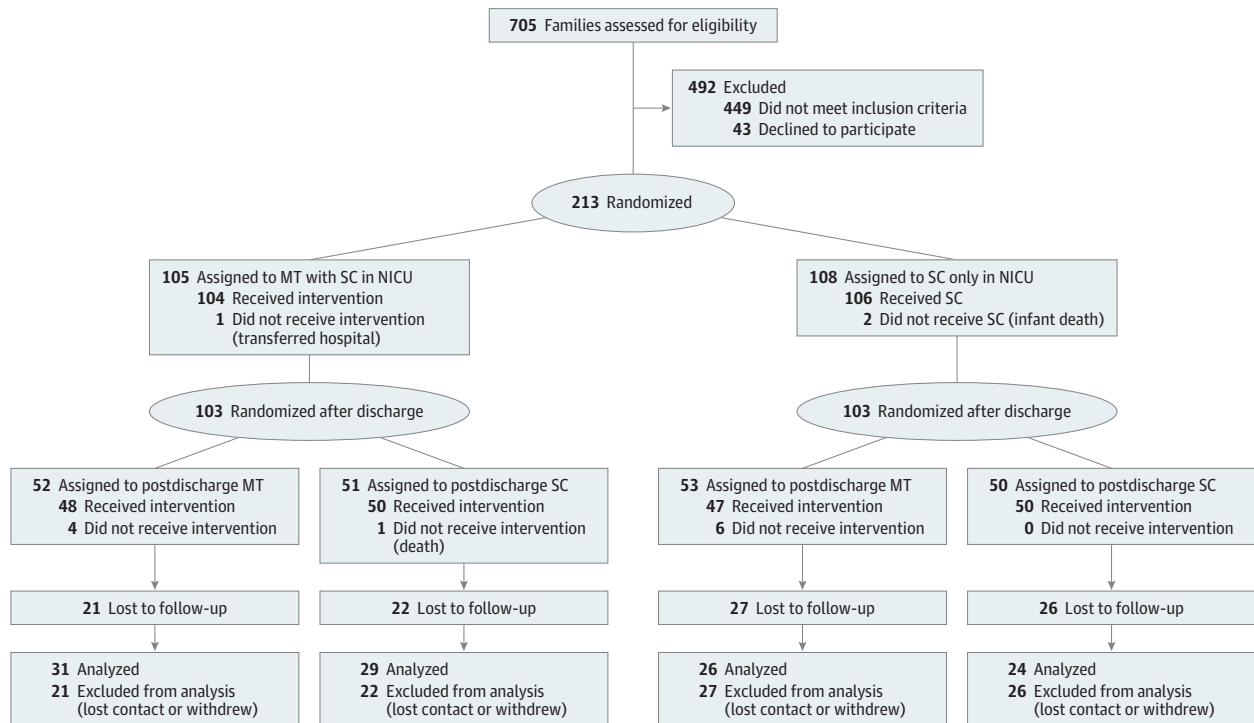
MT indicates music therapy; NICU, neonatal intensive care unit; SC, standard care.

(20.6), with a range from 55 to 145, and the mean (SD) for the motor composite score was 95.0 (17.9), with a range from 46 to 154. Most participants (79 [70%] to 93 [83%]) had BSID-III scores in the normal range (≥85). In the overall sample, 33 infants (30%) demonstrated language composite scores below the normal range (<85), while 19 (17%) and 29 (26%) had cognitive and motor scores below the normal range. Mean differences for the language composite score were −2.36 (95% CI, −12.60 to 7.88; P = .65) for the MT at NICU with postdischarge SC group, 2.65 (95% CI, −7.94 to 13.23; P = .62) for the SC at NICU and postdischarge MT group, and −3.77 (95% CI, −13.97 to 6.43; P = .47) for the MT group at both NICU and postdischarge, all compared with the SC group (Table 2). The mean differences for secondary outcomes were all not significant (Table 2; Figure 2).

Male sex and oxygen supplementation were risk factors for lower language composite scores (eTable 3 and eFigure 1 in Supplement 2) but were not associated with intervention effects (eTable 4 and eFigure 2 in Supplement 2). Sensitivity analyses using multiple imputation (eTable 5 in Supplement 2) did not show significant effects.

Table 1. Baseline Characteristics Across 4 Intervention Groups[a]

Characteristic	SC group Total No.	SC group No. (%)	MT at NICU with postdischarge SC group Total No.	MT at NICU with postdischarge SC group No. (%)	SC at NICU and postdischarge MT group Total No.	SC at NICU and postdischarge MT group No. (%)	MT at both NICU and postdischarge group Total No.	MT at both NICU and postdischarge group No. (%)
Child sex								
Female	50	25 (50)	51	23 (45)	53	30 (57)	52	25 (48)
Male	50	25 (50)	51	28 (55)	53	23 (43)	52	27 (52)
Singleton pregnancy	50	33 (66)	51	38 (75)	53	43 (81)	52	31 (60)
Cesarean delivery route	50	43 (86)	51	39 (76)	53	44 (83)	52	42 (81)
Birth weight, mean (SD), g	50	1422.86 (428.20)	51	1364.04 (426.30)	52	1424.02 (464.08)	52	1391.27 (420.94)
GA at birth, mean (SD), wk	50	30.89 (2.61)	51	30.37 (2.59)	53	30.48 (2.79)	52	30.18 (2.61)
<28		6 (12)		9 (18)		9 (17)		12 (23)
28-32		23 (46)		23 (45)		19 (36)		22 (42)
32 to <35		21 (42)		19 (37)		25 (47)		18 (35)
PMA enrollment, mean (SD), wk	50	33.11 (1.69)	51	33.32 (2.34)	53	33.1 (1.69)	52	32.77 (2.06)
Apgar score at 5 min, mean (SD)	49	8.53 (1.43)	50	8.42 (1.67)	53	8.68 (1.09)	49	8.8 (1.14)
Weight enrollment, mean (SD), g	48	1582.25 (390.92)	51	1675.49 (494.89)	52	1640.48 (372.01)	52	1577.38 (406.50)
Estimated severity of IVH[b]								
Cranial ultrasonography not indicated		22 (44)		17 (33)		21 (40)		27 (52)
None	50	25 (50)	51	27 (53)	52	24 (46)	52	21 (40)
Grade 1-2		3 (6)		5 (10)		6 (12)		4 (8)
Grade 3-4		0		2 (4)		1 (2)		0
Maternal age, mean (SD), y	50	33.4 (5.58)	51	32.14 (5.15)	51	32.22 (5.71)	51	34.08 (5.34)
Maternal educational level, mean (SD), y	48	15.56 (2.62)	50	15.08 (3.85)	50	15.9 (2.77)	51	16.49 (2.63)
Maternal civil status[c]								
Married or living with partner	50	45 (90)	50	47 (94)	53	47 (89)	52	49 (94)
Maternal work status[d]								
Full-time or self-employed	50	39 (78)	51	34 (67)	53	40 (75)	52	37 (71)
Paternal age, mean (SD), y	47	36.57 (7.14)	48	35.17 (6.73)	50	34.48 (5.79)	49	36.53 (5.08)
Paternal educational level, mean (SD), y	44	14.98 (3.11)	47	14.3 (3.42)	49	15.29 (3.42)	47	15.64 (3.23)
Paternal work status[d]								
Full-time or self-employed	47	43 (91)	48	44 (92)	50	49 (98)	49	44 (90)

Abbreviations: GA, gestational age; IVH, intraventricular hemorrhage; MT, music therapy; NICU, neonatal intensive care unit; PMA, postmenstrual age; SC, standard care.

[a] For 206 participants who were randomized twice.

[b] IVH was diagnosed by cranial ultrasound and graded according to Papile et al.[46]

[c] Same-sex parents were invited to participate in the study, but none were enrolled.

[d] Other status included part-time, homemaker or stay-at-home parent, and student.

Table 2. Mean Bayley Scales of Infant and Toddler Development, Third Edition (BSID-III) Composite and Subscale Scores in the 4 Intervention Groups and Mean Differences

BSID-III domains	SC group (n = 24) [Reference]	MT at NICU with postdischarge SC group (n = 29)			SC at NICU and postdischarge MT group (n = 26)			MT at both NICU and postdischarge group (n = 31)			No. of participants
	Mean scores (SD)	Mean scores (SD)	Mean difference (95% CI)[a]	P value	Mean scores (SD)	Mean difference (95% CI)	P value	Mean scores (SD)	Mean difference (95% CI)	P value	
Language composite	97.25 (12.61)	93.52 (22.50)	−2.36 (−12.60 to 7.88)	.65	98.73 (18.10)	2.65 (−7.94 to 13.23)	.62	91.06 (21.41)	−3.77 (−13.97 to 6.43)	.47	110
Expressive subscale	9.00 (2.50)	8.38 (3.53)	−0.41 (−2.13 to 1.31)	.64	9.38 (3.40)	0.54 (−1.24 to 2.31)	.55	8.26 (3.51)	−0.34 (−2.05 to 1.37)	.69	110
Receptive subscale	10.04 (2.22)	9.34 (4.48)	−0.49 (−2.48 to 1.50)	.63	10.12 (3.13)	0.28 (−1.78 to 2.33)	.79	8.65 (4.19)	−1.06 (−3.03 to 0.92)	.29	110
Cognitive composite	102.71 (13.27)	98.79 (20.77)	−3.04 (−14.07 to 7.57)	.58	104.81 (20.17)	3.49 (−7.89 to 14.47)	.54	98.75 (24.76)	−2.84 (−13.85 to 7.56)	.60	111
Motor composite	97.17 (16.24)	94 (18.18)	−3.25 (−12.54 to 6.04)	.49	101.04 (17.72)	1.97 (−7.56 to 11.50)	.68	89.60 (17.91)	−7.38 (−16.66 to 1.88)	.12	108
Gross motor subscale	8.88 (2.68)	8.07 (2.57)	−0.95 (−2.58 to 0.69)	.25	9.46 (3.59)	0.21 (−1.46 to 1.89)	.80	7.77 (3.26)	−1.29 (−2.92 to 0.34)	.12	108
Fine motor subscale	10.08 (3.41)	9.86 (3.94)	−0.08 (−2.00 to 1.85)	.94	10.85 (3.67)	0.56 (−1.41 to 2.53)	.57	8.84 (3.44)	−1.01 (−2.92 to 0.90)	.29	109

Abbreviations: MT, music therapy; NICU, neonatal intensive care unit; SC, standard care.

[a] Results from linear mixed-effects models with site as random effect.

Discussion

Results suggest that children who received MT at NICU with postdischarge SC demonstrated similar developmental scores as children who received SC only. This finding means that we could not confirm a beneficial or detrimental MT effect on neurodevelopment in these premature children. Wide CIs suggest, however, that meaningful effects in either direction cannot be excluded.

Although we hypothesized that MT has an effect on language development at 24 months' CA, the lack of favorable findings is consistent with recent studies demonstrating no clinically meaningful effects of creative MT on the BSID-III in 82 preterm children at 24 months' CA[26] and no effect of parent-led, infant-directed singing supported by a music therapist on BSID-III measured at age 2 to 3 years.[27] Both the current study and recent studies used individually tailored MT during NICU hospitalization, emphasizing the communicative capacity of the infant. The results are also similar to those of a study of music-listening intervention during NICU hospitalization not involving a music therapist,[28] demonstrating no effect on the BSID-III cognitive, language, and motor domains in preterm children at 12 months. Hence, in terms of effects of music-based interventions provided during NICU hospitalization on BSID-III scores at 24 months' CA in preterm children, the findings of

Figure 2. Mean Bayley Scales of Infant and Toddler Development, Third Edition Composite and Subscale Scores and Treatment Effects in the 4 Intervention Groups
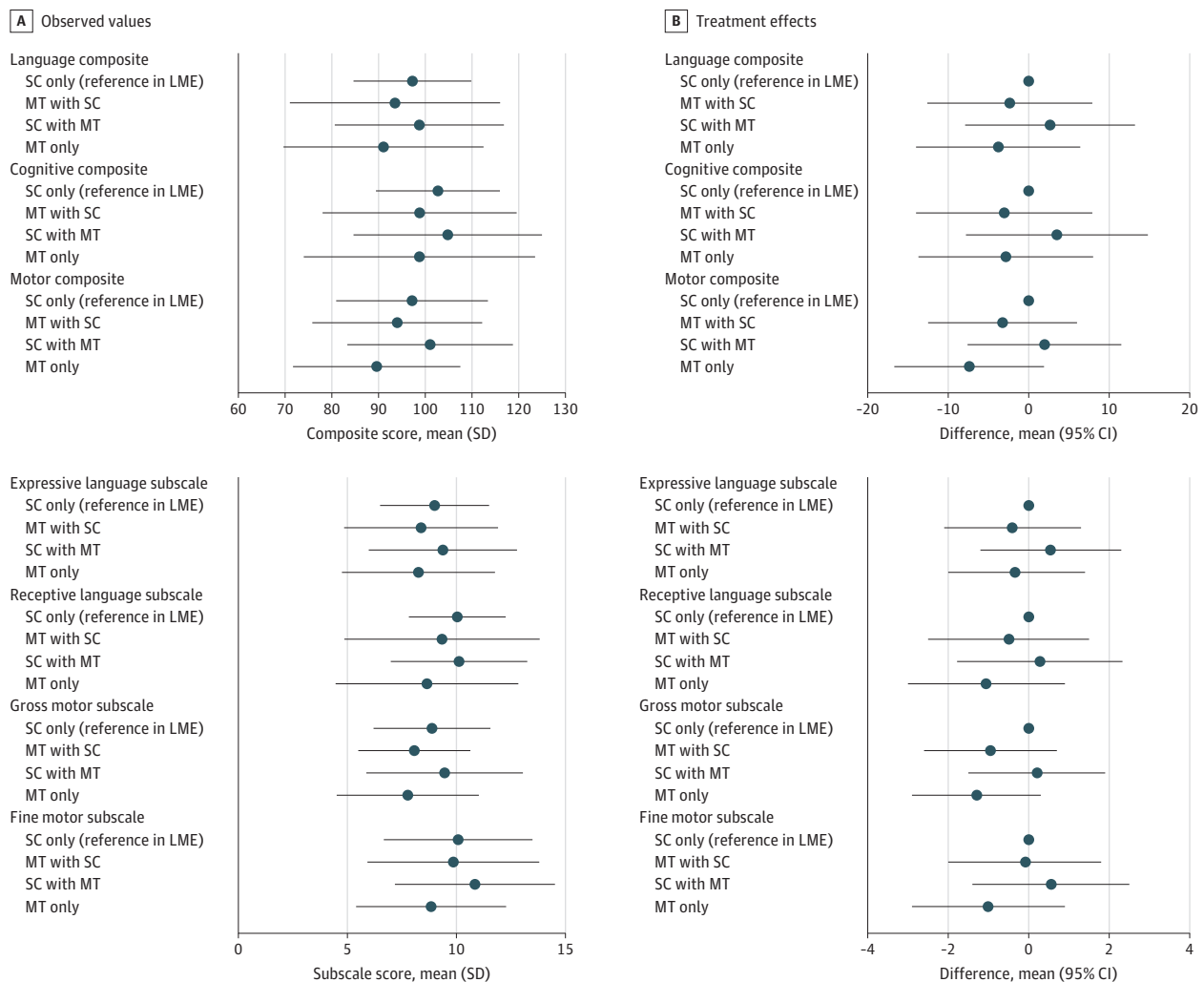
LME indicates linear mixed-effects models; MT, music therapy; and SC, standard care. Circles represent mean composite and subscale scores and mean differences, and whiskers represent SDs for the composite and subscale scores and 95% CIs for the mean differences.

this analysis confirmed previous null findings (eFigure 3 in Supplement 2).[26,28] Notably, in all studies, including the present work, mean BSID-III scores were within the normal range, questioning the need for additional intervention in the study populations. An asset of this study is the ability to test the potential effects of MT in the NICU and/or MT after discharge compared with SC. Estimates suggest no difference in outcomes across condition. To our knowledge, no study has examined the effect of postdischarge MT on BSID-III scores in preterm infants, and we were unable to compare the current findings with previous results.

Although we hypothesized that music therapist–supported, parent-led, infant-directed singing would support infant neurodevelopment over time, we were not able to prove such an effect in the current study. There could be several possible explanations for the null findings. First, previous reports suggested that infants in the sample were medically stable and mature at enrollment,[41] which was underscored by a large proportion of children having BSID-III scores within the normal range at the 24-month follow-up. Hence, this sample was not the group of premature infants most vulnerable to neurodevelopmental impairment, which may have limited the opportunity for intervention effects. Second, the instability of early child development measures[47] may lead to challenges in identifying group differences. This explanation is of particular importance in the case of some but not all preterm children with an increased risk of delayed development in the first years of life.[48] In the current study, the wide CIs of the effect estimates and comparably larger SDs in the intervention groups could indicate large variability of the outcomes, suggesting that some but not all infants benefitted from the MT intervention. Moreover, although the BSID-III is considered a gold standard instrument in early childhood, it might not be sufficiently sensitive to capture the hypothesized subtle changes in the brain from MT.[49,50] Thus, follow-up in later childhood using tools with higher sensitivity and specificity (eg, Wechsler Intelligence Scale for Children, Fifth Edition; Stanford-Binet Intelligence Scales, Fifth Edition) may yield different findings. Alternatively, other procedures, such as heart rate variability, electroencephalogram, or functional near-infrared spectroscopy, could be more suited to identify the effect of MT on brain functioning.[16,17,51] Future studies should focus on identifying those infants who could benefit from early MT and measuring the effects with a wider array of neurodevelopmental and neurophysiological instruments.

Strengths and Limitations

Key strengths of the LongSTEP trial were the pragmatic approach in setting (a diversity of clinical settings) and the intervention completion, increasing the generalizability of the findings. The pragmatic approach, however, could also represent a limitation, introducing noise in the data. The onset of the COVID-19 pandemic and lockdowns aggravated this risk in several ways. Retaining families to participate in follow-up is already challenging,[52] but it is even more so during the COVID-19 pandemic, leading to a low study retention rate (54%) and, as a consequence, underpowered analyses for detecting the effect of MT on neurodevelopmental outcomes.[53,54] Retention differed by participating country, probably due to differences in clinical practice. Notably, 1 site had high retention (12 of 54 = 22%), showing that it is possible in principle to keep contact with families over a longer time span. Moreover, pandemic-related lockdowns affected the quality of the BSID-III training, standardization, and quality control process, impairing the reliability of the measures. The onset of the pandemic also extended the assessment window from 24 to 32 months' CA (within a 2-week window). However, although 53 children were assessed outside the window period, the mean (SD) CA at assessment was 25.54 (1.72) months, suggesting minimal deviation from the original age window. Furthermore, although testing was planned to be administered in designated rooms at the clinic or hospital, some assessments were also performed at the family's home due to lockdowns. Noise introduced by the COVID-19 pandemic and high attrition rate increased the risk of type II error and the risk of our inability to detect effects that were actually present.

Conclusions

In this secondary analysis of a pragmatic RCT, a music therapist–supported, parent-led, infant-directed singing intervention did not show effects on neurodevelopment at 24 months' CA. Wide CIs suggest that meaningful effects in either direction cannot be excluded. Future research with a focused RCT design is warranted to determine the MT approaches, implementation time, and duration that are effective in targeting groups at risk for neurodevelopmental impairment and introducing broader measurements for changes in brain development.

Corresponding Author: Ingrid Kvestad, PhD, NORCE Norwegian Research Centre AS, PO Box 22 Nygårdstangen, 5838 Bergen, Norway (inkv@norceresearch.no).

Author Affiliations: GAMUT—The Grieg Academy Music Therapy Research Centre, NORCE Norwegian Research Centre AS, Bergen, Norway (Bieleninik, Gold, Assmus, Gaden); Department of Clinical and Health Psychology, Institute of Psychology, Faculty of Social Sciences, University of Gdańsk, Gdańsk, Poland (Bieleninik); Regional Centre for Child and Youth Mental Health and Child Welfare, NORCE Norwegian Research Centre AS, Bergen, Norway (Kvestad); Department of Clinical and Health Psychology, University of Vienna, Vienna, Austria (Gold); NORCE Energy, NORCE Norwegian Research Centre AS, Bergen, Norway (Stordal); Department of Mathematics, University of Bergen, Bergen, Norway (Stordal); Meir Medical Center, Kfar-Saba, Israel (Arnon, Haar-Shamir); Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel (Arnon); University of Haifa, Haifa, Israel (Elefant); Hospital Universitario Fundación Santa Fe de Bogotá, Bogotá, Colombia (Ettenberger); Clínica de la Mujer, Bogotá, Colombia (Ettenberger); Department of Children and Youth Clinic, Haukeland University Hospital, Bergen, Norway (Håvardstun, Wiborg, Vederhus); Facultad de Ciencias de la Salud, Universidad de Ciencias Empresariales y Sociales, Buenos Aires, Argentina (Lichtensztejn); Oslo University Hospital, Oslo, Norway (Mangersnes); GAMUT—The Grieg Academy, Department of Music, University of Bergen, Bergen, Norway (Ghetti).

REFERENCES

1. Chawanpaiboon S, Vogel JP, Moller AB, et al. Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *Lancet Glob Health*. 2019;7(1):e37-e46. doi:10.1016/S2214-109X(18)30451-0

2. World Health Organization. *Born Too Soon: The Global Action Report on Preterm Birth*. World Health Organization; 2012.

3. Blencowe H, Lee AC, Cousens S, et al. Preterm birth-associated neurodevelopmental impairment estimates at regional and global levels for 2010. *Pediatr Res*. 2013;74(suppl 1):17-34. doi:10.1038/pr.2013.204

4. Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJ. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA*. 2002;288(6):728-737. doi:10.1001/jama.288.6.728

5. Delobel-Ayoub M, Arnaud C, White-Koning M, et al; EPIPAGE Study Group. Behavioral problems and cognitive performance at 5 years of age after very preterm birth: the EPIPAGE Study. *Pediatrics*. 2009;123(6):1485-1492. doi:10.1542/peds.2008-1216

6. Twilhaar ES, de Kieviet JF, Aarnoudse-Moens CS, van Elburg RM, Oosterlaan J. Academic performance of children born preterm: a meta-analysis and meta-regression. *Arch Dis Child Fetal Neonatal Ed*. 2018;103(4): F322-F330. doi:10.1136/archdischild-2017-312916

7. Aarnoudse-Moens CS, Weisglas-Kuperus N, van Goudoever JB, Oosterlaan J. Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. *Pediatrics*. 2009;124(2):717-728. doi:10.1542/peds.2008-2816

8. Allotey J, Zamora J, Cheong-See F, et al. Cognitive, motor, behavioural and academic performances of children born preterm: a meta-analysis and systematic review involving 64 061 children. *BJOG*. 2018;125(1):16-25. doi:10.1111/1471-0528.14832

9. Lahav A, Skoe E. An acoustic gap between the NICU and womb: a potential risk for compromised neuroplasticity of the auditory system in preterm infants. *Front Neurosci*. 2014;8(381):381. doi:10.3389/fnins.2014.00381

10. Bertsch M, Reuter C, Czedik-Eysenberg I, et al. The "sound of silence" in a neonatal intensive care unit–listening to speech and music inside an incubator. *Front Psychol*. 2020;11:1055. doi:10.3389/fpsyg.2020.01055

11. Caskey M, Stephens B, Tucker R, Vohr B. Adult talk in the NICU with preterm infants and developmental outcomes. *Pediatrics*. 2014;133(3):e578-e584. doi:10.1542/peds.2013-0104

12. Caskey M, Stephens B, Tucker R, Vohr B. Importance of parent talk on the development of preterm infant vocalizations. *Pediatrics*. 2011;128(5):910-916. doi:10.1542/peds.2011-0609

13. Filippa M, Panza C, Ferrari F, et al. Systematic review of maternal voice interventions demonstrates increased stability in preterm infants. *Acta Paediatr*. 2017;106(8):1220-1229. doi:10.1111/apa.13832

14. Saliba S, Esseily R, Filippa M, Kuhn P, Gratier M. Exposure to human voices has beneficial effects on preterm infants in the neonatal intensive care unit. *Acta Paediatr*. 2018;107(7):1122-1130. doi:10.1111/apa.14170

15. Saliba S, Gratier M, Filippa M, Devouche E, Esseily R. Fathers' and mothers' infant directed speech influences preterm infant behavioral state in the NICU. *J Nonverbal Behav*. 2020;44:437-451. doi:10.1007/s10919-020-00335-1

16. Markova G, Nguyen T, Hoehl S. Neurobehavioral interpersonal synchrony in early development: the role of interactional rhythms. *Front Psychol*. 2019;10:2078. doi:10.3389/fpsyg.2019.02078

17. Nguyen T, Reisner S, Lueger A, Wass SV, Hoehl S, Markova G. Sing to me, baby: infants show neural tracking and rhythmic movements to live and dynamic maternal singing. *Dev Cogn Neurosci*. 2023;64:101313. doi:10.1016/j.dcn.2023.101313

18. Yue W, Han X, Luo J, Zeng Z, Yang M. Effect of music therapy on preterm infants in neonatal intensive care unit: systematic review and meta-analysis of randomized controlled trials. *J Adv Nurs*. 2021;77(2):635-652. doi:10.1111/jan.14630

19. Bieleninik Ł, Ghetti C, Gold C. Music therapy for preterm infants and their parents: a meta-analysis. *Pediatrics*. 2016;138(3):e20160971. doi:10.1542/peds.2016-0971

20. Chorna O, Filippa M, De Almeida JS, et al. Neuroprocessing mechanisms of music during fetal and neonatal development: a role in neuroplasticity and neurodevelopment. *Neural Plast*. 2019;2019:3972918. doi:10.1155/2019/3972918

21. Haslbeck FB. Music therapy with preterm infants—theoretical approach and first practical experience. *Music Therapy Today*. 2004;5(4):1-15.

22. Haslbeck FB, Bassler D. Music from the very beginning—a neuroscience-based framework for music as therapy for preterm infants and their parents. *Front Behav Neurosci*. 2018;12(112):112. doi:10.3389/fnbeh.2018.00112

23. Haslbeck FB, Jakab A, Held U, Bassler D, Bucher HU, Hagmann C. Creative music therapy to promote brain function and brain structure in preterm infants: a randomized controlled pilot study. *Neuroimage Clin*. 2020;25:102171. doi:10.1016/j.nicl.2020.102171

24. Lordier L, Loukas S, Grouiller F, et al. Music processing in preterm and full-term newborns: a psychophysiological interaction (PPI) approach in neonatal fMRI. *Neuroimage*. 2019;185:857-864. doi:10.1016/j.neuroimage.2018.03.078

25. Lordier L, Meskaldji DE, Grouiller F, et al. Music in premature infants enhances high-level cognitive brain networks. *Proc Natl Acad Sci U S A*. 2019;116(24):12103-12108. doi:10.1073/pnas.1817536116

26. Haslbeck FB, Bucher HU, Bassler D, Hagmann C, Natalucci G. Creative music therapy and neurodevelopmental outcomes in pre-term infants at 2 years: a randomized controlled pilot trial. *Front Pediatr*. 2021;9:660393. doi:10.3389/fped.2021.660393

27. Kostilainen K, Hugoson P, Haavisto A, et al. No impact of parental singing during the neonatal period on cognition in preterm-born children at 2-3 years. *Acta Paediatr*. 2023;112(7):1471-1477. doi:10.1111/apa.16788

28. Lejeune F, Lordier L, Pittet MP, et al. Effects of an early postnatal music intervention on cognitive and emotional development in preterm children at 12 and 24 months: preliminary findings. *Front Psychol*. 2019; 10:494. doi:10.3389/fpsyg.2019.00494

29. Ghetti C, Bieleninik Ł, Hysing M, et al. Longitudinal Study of Music Therapy's Effectiveness for Premature Infants and Their Caregivers (LongSTEP): protocol for an international randomised trial. *BMJ Open*. 2019;9(8): e025062. doi:10.1136/bmjopen-2018-025062

30. Ghetti CM, Gaden TS, Bieleninik Ł, et al. Effect of music therapy on parent-infant bonding among infants born preterm: a randomized clinical trial. *JAMA Netw Open*. 2023;6(5):e2315750. doi:10.1001/jamanetworkopen. 2023.15750

31. Bieleninik Ł, Ghetti C. Music therapy for preterm infants and their parents: a path forward for research in Poland. *Annee Psychol*. 2019;22(2):135-150.

32. Gaden TS, Ghetti C, Kvestad I, et al. Short-term music therapy for families with preterm infants: a randomized trial. *Pediatrics*. 2022;149(2):e2021052797. doi:10.1542/peds.2021-052797

33. Gaden TS, Ghetti C, Kvestad I, Gold C. The LongSTEP approach: theoretical framework and intervention protocol for using parent-driven infant-directed singing as resource-oriented music therapy. *Nord J Music Ther*. 2022;31(2):107-132. doi:10.1080/08098131.2021.1921014

34. Ghetti CM, Vederhus BJ, Gaden TS, et al. Longitudinal Study of Music Therapy's Effectiveness for Premature Infants and Their Caregivers (LongSTEP): feasibility study with a Norwegian cohort. *J Music Ther*. 2021;58(2): 201-240. doi:10.1093/jmt/thaa023

35. Loewy J. NICU music therapy: song of kin as critical lullaby in research and practice. *Ann N Y Acad Sci*. 2015; 1337(1):178-185. doi:10.1111/nyas.12648

36. Shoemark H. Time together: a feasible program to promote parent-infant interaction in the NICU. *Music Ther Perspect*. 2018;36(1):6-16. doi:10.1093/mtp/mix004

37. Shoemark H. Translating infant-directed singing into a strategy for the hospitalized family. In: Edwards J, ed. *Music Therapy and Parent-Infant Bonding*. Oxford University Press; 2011:161-178. doi:10.1093/acprof:oso/ 9780199580514.003.0012

38. Mondanaro JF, Ettenberger M, Park L. Mars rising: music therapy and the increasing presence of fathers in the NICU. *Music Med*. 2016;8(3):96-107. doi:10.47513/mmd.v8i3.440

39. Loewy J, Stewart K, Dassler AM, Telsey A, Homel P. The effects of music therapy on vital signs, feeding, and sleep in premature infants. *Pediatrics*. 2013;131(5):902-918. doi:10.1542/peds.2012-1367

40. Haslbeck FB, Bassler D. Clinical practice protocol of creative music therapy for preterm infants and their parents in the neonatal intensive care unit. *J Vis Exp*. 2020;(155). doi:10.3791/60412-v

41. Gaden TS, Gold C, Assmus J, et al. Treatment fidelity in a pragmatic clinical trial of music therapy for premature infants and their parents: the LongSTEP study. *Trials*. 2023;24(1):160. doi:10.1186/s13063-022-06971-w

42. Bayley N. *Bayley Scales of Infant and Toddler Development*. 3rd ed. Pearson; 2006.

43. Walder DJ, Sherman JC, Pulsifer MB. Neurodevelopmental assessment. In: Mowder BA, Rubinson F, Yasik AE, eds. *Evidence-Based Practice in Infant and Early Childhood Psychology*. John Wiley & Sons; 2009:167-205. doi:10.1002/9781118269602.ch6

44. Luttikhuizen dos Santos ES, de Kieviet JF, Königs M, van Elburg RM, Oosterlaan J. Predictive value of the Bayley scales of infant development on development of very preterm/very low birth weight children: a meta-analysis. *Early Hum Dev*. 2013;89(7):487-496. doi:10.1016/j.earlhumdev.2013.03.008

45. Romeo DM. Correcting for prematurity with the Bayley Scales of Infant Development. *Dev Med Child Neurol*. 2018;60(8):736-737. doi:10.1111/dmcn.13741

46. Papile LA, Burstein J, Burstein R, Koffler H. Incidence and evolution of subependymal and intraventricular hemorrhage: a study of infants with birth weights less than 1,500 gm. *J Pediatr*. 1978;92(4):529-534. doi:10.1016/ S0022-3476(78)80282-0

47. Liu L, Johnson HL, Cousens S, et al; Child Health Epidemiology Reference Group of WHO and UNICEF. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet*. 2012;379(9832):2151-2161. doi:10.1016/S0140-6736(12)60560-1

48. Jansson-Verkasalo E, Ruusuvirta T, Huotilainen M, et al. Atypical perceptual narrowing in prematurely born infants is associated with compromised language acquisition at 2 years of age. *BMC Neurosci*. 2010;11(88):88. doi:10.1186/1471-2202-11-88

49. Johnson S, Moore T, Marlow N. Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used? *Pediatr Res*. 2014;75(5):670-674. doi:10.1038/pr.2014.10

50. Anderson PJ, Burnett A. Assessing developmental delay in early childhood - concerns with the Bayley-III scales. *Clin Neuropsychol*. 2017;31(2):371-381. doi:10.1080/13854046.2016.1216518

51. Varisco G, Van Der Wal WR, Bakker-Bos J, Kommers D, Andriessen P, Van Pul C. *Effect of Music Therapy Interventions on Heart Rate Variability in Premature Infants*. IEEE; 2022:678-681. doi:10.1109/EMBC48229.2022.9871017

52. Mas C, Gérardin P, Chirpaz E, et al. Follow-up at two years of age and early predictors of non-compliance in a cohort of very preterm infants. *Early Hum Dev*. 2017;108:1-7. doi:10.1016/j.earlhumdev.2017.03.001

53. Natalucci G, Latal B, Koller B, et al; Swiss EPO Neuroprotection Trial Group. Effect of early prophylactic high-dose recombinant human erythropoietin in very preterm infants on neurodevelopmental outcome at 2 years: a randomized clinical trial. *JAMA*. 2016;315(19):2079-2085. doi:10.1001/jama.2016.5504

54. Youn YA, Shin SH, Kim EK, et al. Preventive intervention program on the outcomes of very preterm infants and caregivers: a multicenter randomized controlled trial. *Brain Sci*. 2021;11(5):575. doi:10.3390/brainsci11050575

SUPPLEMENT 1.
Trial Protocol

SUPPLEMENT 2.
eAppendix. Standardization Procedures of Neurodevelopmental Assessment With the Bayley Scales of Infant and Toddler Development, 3rd Edition (BISD-III)
eTable 1. Infant Baseline Characteristics of Those Followed-Up Versus Lost to Follow-Up
eTable 2. Completers by Country and Intervention Group
eTable 3. Clinical and Demographic Predictors of the Bayley Scales of Infant and Toddler Development, 3rd Edition (BISD-III) Language Composite Score
eFigure 1. Estimates of Clinical and Demographic Predictors of the Bayley Scales of Infant and Toddler Development, 3rd Edition (BISD-III) Language Composite Score
eTable 4. Intervention Effects by Selected Clinical/Demographic Predictors of the Bayley Scales of Infant and Toddler Development, 3rd Edition (BISD-III) Language Composite Score
eTable 5. Sensitivity Analysis: Estimates of Intervention Effects on the Bayley Scales of Infant and Toddler Development, 3rd Edition (BISD-III) Language Composite Score Using Multiple Imputation
eFigure 2. Estimates of Intervention Effects on the Bayley Scales of Infant and Toddler Development, 3rd Edition (BISD-III) Language Composite Score by Selected Clinical/Demographic Predictors
eFigure 3. Effects of Music Interventions on the Bayley Scales of Infant and Toddler Development, 3rd Edition (BISD-III) at 24 Months Across Studies
eReferences

SUPPLEMENT 3.
Data Sharing Statement