Cochrane Library

Cochrane Database of Systematic Reviews

Musical and vocal interventions to improve neurodevelopmental outcomes for preterm infants (Review)

Haslbeck FB, Mueller K, Karen T, Loewy J, Meerpohl JJ, Bassler D

Haslbeck FB, Mueller K, Karen T, Loewy J, Meerpohl JJ, Bassler D.
Musical and vocal interventions to improve neurodevelopmental outcomes for preterm infants.
Cochrane Database of Systematic Reviews 2023, Issue 9. Art. No.: CD013472.
DOI: 10.1002/14651858.CD013472.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Musical and vocal interventions to improve neurodevelopmental outcomes for preterm infants

Friederike B Haslbeck[1], Katharina Mueller[2], Tanja Karen[1], Joanne Loewy[3], Joerg J Meerpohl[4], Dirk Bassler[1]

[1]Department of Neonatology, University Hospital Zurich, Zurich, Switzerland. [2]Zentrum für Kinder und Jugendmedizin, University Freiburg, Freiburg, Germany. [3]Mount Sinai Health System, The Louis Armstrong Center for Music & Medicine, New York City, USA. [4]Institute for Evidence in Medicine, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

Contact: Friederike B Haslbeck, friederike.haslbeck@usz.ch.

Editorial group: Cochrane Neonatal Group.
Publication status and date: New, published in Issue 9, 2023.

Citation: Haslbeck FB, Mueller K, Karen T, Loewy J, Meerpohl JJ, Bassler D. Musical and vocal interventions to improve neurodevelopmental outcomes for preterm infants. *Cochrane Database of Systematic Reviews* 2023, Issue 9. Art. No.: CD013472. DOI: 10.1002/14651858.CD013472.pub2.

ABSTRACT

Background

Preterm birth interferes with brain maturation, and subsequent clinical events and interventions may have additional deleterious effects. Music as therapy is offered increasingly in neonatal intensive care units aiming to improve health outcomes and quality of life for both preterm infants and the well-being of their parents. Systematic reviews of mixed methodological quality have demonstrated ambiguous results for the efficacy of various types of auditory stimulation of preterm infants. A more comprehensive and rigorous systematic review is needed to address controversies arising from apparently conflicting studies and reviews.

Objectives

We assessed the overall efficacy of music and vocal interventions for physiological and neurodevelopmental outcomes in preterm infants (< 37 weeks' gestation) compared to standard care. In addition, we aimed to determine specific effects of various interventions for physiological, anthropometric, social-emotional, neurodevelopmental short- and long-term outcomes in the infants, parental well-being, and bonding.

Search methods

We searched Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE, Embase, CINAHL, PsycINFO, Web of Science, RILM Abstracts, and ERIC in November 2021; and Proquest Dissertations in February 2019. We searched the reference lists of related systematic reviews, and of studies selected for inclusion and clinical trial registries.

Selection criteria

We included parallel, and cluster-randomised controlled trials with preterm infants < 37 weeks` gestation during hospitalisation, and parents when they were involved in the intervention. Interventions were any music or vocal stimulation provided live or via a recording by a music therapist, a parent, or a healthcare professional compared to standard care. The intervention duration was greater than five minutes and needed to occur more than three times.

Data collection and analysis

Three review authors independently extracted data. We analysed the treatment effects of the individual trials using RevMan Web using a fixed-effects model to combine the data. Where possible, we presented results in meta-analyses using mean differences with 95% CI. We

performed heterogeneity tests. When the I^2 statistic was higher than 50%, we assessed the source of the heterogeneity by sensitivity and subgroup analyses. We used GRADE to assess the certainty of the evidence.

Main results

We included 25 trials recruiting 1532 infants and 691 parents (21 parallel-group RCTs, four cross-over RCTs). The infants gestational age at birth varied from 23 to 36 weeks, taking place in NICUs (level 1 to 3) around the world. Within the trials, the intervention varied widely in type, delivery, frequency, and duration. Music and voice were mainly characterised by calm, soft, musical parameters in lullaby style, often integrating the sung mother's voice live or recorded, defined as music therapy or music medicine. The general risk of bias in the included studies varied from low to high risk of bias.

Music and vocal interventions compared to standard care

Music/vocal interventions do not increase oxygen saturation in the infants during the intervention (mean difference (MD) 0.13, 95% CI -0.33 to 0.59; P = 0.59; 958 infants, 10 studies; high-certainty evidence). Music and voice probably do not increase oxygen saturation post-intervention either (MD 0.63, 95% CI -0.01 to 1.26; P = 0.05; 800 infants, 7 studies; moderate-certainty evidence). The intervention may not increase infant development (Bayley Scales of Infant and Toddler Development (BSID)) with the cognitive composition score (MD 0.35, 95% CI -4.85 to 5.55; P = 0.90; 69 infants, 2 studies; low-certainty evidence); the motor composition score (MD -0.17, 95% CI -5.45 to 5.11; P = 0.95; 69 infants, 2 studies; low-certainty evidence); and the language composition score (MD 0.38, 95% CI -5.45 to 6.21; P = 0.90; 69 infants, 2 studies; low-certainty evidence). Music therapy may not reduce parental state-trait anxiety (MD -1.12, 95% CI -3.20 to 0.96; P = 0.29; 97 parents, 4 studies; low-certainty evidence).

The intervention probably does not reduce respiratory rate during the intervention (MD 0.42, 95% CI -1.05 to 1.90; P = 0.57; 750 infants; 7 studies; moderate-certainty evidence) and post-intervention (MD 0.51, 95% CI -1.57 to 2.58; P = 0.63; 636 infants, 5 studies; moderate-certainty evidence). However, music/vocal interventions probably reduce heart rates in preterm infants during the intervention (MD -1.38, 95% CI -2.63 to -0.12; P = 0.03; 1014 infants; 11 studies; moderate-certainty evidence). This beneficial effect was even stronger after the intervention. Music/vocal interventions reduce heart rate post-intervention (MD -3.80, 95% CI -5.05 to -2.55; P < 0.00001; 903 infants, 9 studies; high-certainty evidence) with wide CIs ranging from medium to large beneficial effects. Music therapy may not reduce postnatal depression (MD 0.50, 95% CI -1.80 to 2.81; P = 0.67; 67 participants; 2 studies; low-certainty evidence). The evidence is very uncertain about the effect of music therapy on parental state anxiety (MD -0.15, 95% CI -2.72 to 2.41; P = 0.91; 87 parents, 3 studies; very low-certainty evidence). We are uncertain about any further effects regarding all other secondary short- and long-term outcomes on the infants, parental well-being, and bonding/attachment. Two studies evaluated adverse effects as an explicit outcome of interest and reported no adverse effects from music and voice.

Authors' conclusions

Music/vocal interventions do not increase oxygen saturation during and probably not after the intervention compared to standard care. The evidence suggests that music and voice do not increase infant development (BSID) or reduce parental state-trait anxiety. The intervention probably does not reduce respiratory rate in preterm infants. However, music/vocal interventions probably reduce heart rates in preterm infants during the intervention, and this beneficial effect is even stronger after the intervention, demonstrating that music/vocal interventions reduce heart rates in preterm infants post-intervention. We found no reports of adverse effects from music and voice. Due to low-certainty evidence for all other outcomes, we could not draw any further conclusions regarding overall efficacy nor the possible impact of different intervention types, frequencies, or durations. Further research with more power, fewer risks of bias, and more sensitive and clinically relevant outcomes are needed.

PLAIN LANGUAGE SUMMARY

Can music and vocal interventions benefit preterm infants and their parents?

Key messages

• Music and vocal interventions probably reduce heart rates in preterm infants compared to standard care during the intervention. This beneficial effect was even more substantial and confident after the intervention suggesting a long-lasting relaxing and stabilising effect.

• We found no harmful effects from music and voice. However, many studies did not explicitly explore the possibility of unwanted effects.

• We found no evidence of any other clear beneficial or harmful effects of the interventions on the infants, their parents, and parent-infant bonding. More good-quality evidence is needed to draw further clear conclusions.

What is a preterm infant?

Preterm infants are newborns born before the gestational age of 37 weeks and often have to be treated for weeks to months in the stressful environment of a neonatal intensive care unit to survive.

Why examine the potential benefits of music and vocal interventions for preterm infants and their parents?

Preterm infants are at risk for various health issues. Preterm birth is a traumatic event for the parents as well. Therefore, complementary approaches such as music and vocal interventions are increasingly used in neonatal care to improve physical and mental health in preterm infants and their parents. However, various studies and reviews show ambiguous results in the efficacy of a variety of music and vocal interventions. A more comprehensive and rigorous systematic review is needed to address conflicting data and reviews.

What did we want to find out?

We wanted to find out if music and vocal interventions benefit:

• the health and development of the preterm infant

• the mental health of the parents and their bonding with the infant

We wanted to know which types, delivery, duration, and frequency of music and vocal interventions would best support infants and parents. We aimed to find out if the intervention can cause any harmful effects.

What did we do?

We searched for studies that compared:

• music and vocal interventions for preterm infants (and parents) compared to usual standard care in the neonatal unit that did not include any music or vocal interventions.

We compared and summarised their results and rated our confidence in the evidence, based on factors such as study methods and sizes.

What did we find?

We found 25 studies that involved 1532 preterm infants and 691 parents. The biggest study was in 272, and the smallest was in 17 preterm infants. Within the studies from around the world, mainly the immediate effect of music and voice was examined in the moments of intervention and minutes post-intervention, whereas two studies wanted to know if there would be a beneficial effect on long-term development at two years. Most studies were funded by University/Health Department/Hospital research funds and local medical/health foundations. The reported music and vocal interventions varied widely in type, delivery, frequency, and duration. They were mainly characterised by calm, soft, musical parameters in lullaby style, often integrating the mother's voice live or recorded, defined as music therapy when provided by a music therapist within a therapeutic relationship or music medicine when delivered as "medicine" by medical and healthcare professionals.

Main results

In preterm infants (and their parents), compared to standard care without any music and vocal interventions:

• Music and voice make no difference to the oxygen saturation during the intervention (10 studies with 958 infants) and may make no difference after the intervention (7 studies with 800 infants).

• Music and voice may make no difference in the respiratory rate during the intervention (7 studies with 750 infants) and after the intervention (5 studies with 636 infants).

• Music and voice may lead to a beneficial reduction in infants' heart rates (11 studies with 1014 infants). This beneficial effect was even more substantial and confident after the intervention, leading to a medium-to-large beneficial reduction in the heart rate (5 studies with 636 infants).

• We are uncertain if the intervention may influence infant long-term development at two years of age (2 studies with 69 infants).

• We are uncertain about the possible effect of music therapy on parental state-trait anxiety (4 studies with 97 participants) and postnatal depression (2 studies with 67 infants).

• We are very uncertain about a possible effect on parental state anxiety (3 studies with 87 parents).

• We found no studies which reported harmful effects of music or voice.

What are the limitations of the evidence?

We are confident that music and voice do not reduce oxygen saturation during the intervention compared to standard care. We are confident in our results of the substantial beneficial effect on the heart rate in preterm infants after the intervention. There are not enough rigorous studies (many small studies with poor recording standards) to be certain about the results of all other outcomes that we assessed in the infants and their parents. There is further uncertainty about music delivery and for which duration and frequency music works best.

How up-to-date is this evidence?

The evidence is up-to-date to 12 November 2021.

SUMMARY OF FINDINGS

Summary of findings 1. Summary of findings table - Music and vocal interventions for preterm infants and their parents

Music and vocal interventions for preterm infants and their parents

Patient or population: preterm infants and their parents
Setting: NICUs level 1-3 in Europe, Middle East, USA, Asia, Australia, South America
Intervention: music/vocal interventions
Comparison: standard care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with standard care	Risk with music/vocal interventions				
OXYGEN SATURATION DURING intervention (higher = favourable)	The mean OXYGEN SATURATION DURING intervention (higher = favourable) ranged from **92.76 to 98.2**	MD **0.13 higher** (0.33 lower to 0.59 higher)	-	958 (10 RCTs)	⊕⊕⊕⊕ High[a]	Music/vocal interventions do not increase oxygen saturation during intervention.[a]
OXYGEN SATURATION POST-intervention (higher = favourable) assessed with: up to follow-up: 30 minutes	The mean OXYGEN SATURATION POST-intervention (higher = favourable) ranged from **92.33 to 96.54**	MD **0.63 higher** (0.01 lower to 1.26 higher)	-	800 (7 RCTs)	⊕⊕⊕⊖ Moderate[b]	Music/vocal interventions probably do not increase oxygen saturation post-intervention.[b]
INFANT DEVELOPMENT: Bayley Scales of Infant and Toddler Development-III (BSID-III) assessed with: COGNITIVE composition score Scale from: 0 to 200 (higher = favourable)	The mean INFANT DEVELOPMENT: Bayley Scales of Infant and Toddler Development-III ranged from **98.6 to 102.8**	MD **0.35 higher** (4.85 lower to 5.55 higher)	-	69 (2 RCTs)	⊕⊕⊖⊖ Low[b,c]	Music/ vocal interventions may not increase infant development.[b,c]
INFANT DEVELOPMENT: Bayley Scales of Infant and Toddler Development-III (BSID III) assessed with: MOTOR composition score Scale from: 0 to 200 (higher = favourable) follow-up: 2 patient years	The mean INFANT DEVELOPMENT: Bayley Scales of Infant and Toddler Development-III ranged from **97.6 to 102.1**	MD **0.17 lower** (5.45 lower to 5.11 higher)	-	69 (2 RCTs)	⊕⊕⊖⊖ Low[b,c]	Music/vocal interventions may not increase infant development.[b,c]

INFANT DEVELOPMENT: Bayley Scales of Infant and Toddler Development-III (BSID III) assessed with: LANGUAGE composition score Scale from: 9 to 200 (higher = favourable) follow-up: 2 patient years	The mean INFANT DEVELOPMENT: Bayley Scales of Infant and Toddler Development-III ranged from **91.62 to 92.4**	MD **0.38 higher** (5.45 lower to 6.21 higher)	-	69 (2 RCTs)	⊕⊕⊖⊖ Low[b,c]	Music/vocal interventions may not increase infant development.[b,c]
PARENTAL ANXIETY: STAI-T assessed with: State Trait Anxiety Inventory Scale from: 6 to 80 (lower = favourable)	The mean PARENTAL ANXIETY: STAI-T ranged from **12 to 38.85**	MD **1.12 lower** (3.2 lower to 0.96 higher)	-	97 (4 RCTs)	⊕⊕⊖⊖ Low[c,d]	Music/ vocal interventions may not reduce parental state-trait anxiety.[c,d]
RESPIRATORY RATE DURING intervention (lower = favourable)	The mean RESPIRATORY RATE DURING intervention (lower = favourable) ranged from **41.2 to 53**	MD **0.42 higher** (1.05 lower to 1.9 higher)	-	750 (7 RCTs)	⊕⊕⊕⊖ Moderate[b]	Music/vocal interventions probably do not reduce respiratory rate during intervention.[b]
RESPIRATOTY RATE POST-intervention (lower = favourable) assessed with: up to follow-up: 30 minutes	The mean RESPIRATOTY RATE POST-intervention (lower = favourable) ranged from **48.5 to 60.33**	MD **0.51 higher** (1.57 lower to 2.58 higher)	-	636 (5 RCTs)	⊕⊕⊕⊖ Moderate[b]	Music/vocal interventions probably do not reduce respiratory rate postintervention.[b]
HEART RATE DURING intervention (lower = favourable)	The mean HEART RATE DURING intervention (lower = favourable) ranged from **131.4 to 158.81**	MD **1.38 lower** (2.63 lower to 0.12 lower)	-	1014 (11 RCTs)	⊕⊕⊕⊖ Moderate[e]	Music/vocal interventions probably reduce heart rate during the intervention period.[e]
HEART RATE POST-intervention (lower = favourable) assessed with: up to follow-up: 30 minutes	The mean HEART RATE POST-intervention (lower = favourable) was **138.19 to 161.4**	MD **3.8 lower** (5.05 lower to 2.55 lower)	-	903 (9 RCTs)	⊕⊕⊕⊕ High[a]	Music/vocal interventions reduce heart rate postintervention.[a]
PARENTAL WELL-BEING: EPDS assessed with: Edinburgh Postnatal Depression Scale Scale from: 0 to 30 (higher = favourable)	The mean PARENTAL WELL-BEING: EPDS ranged from **7.83 to 8.08**	MD **0.5 higher** (1.8 lower to 2.81 higher)	-	67 (2 RCTs)	⊕⊕⊖⊖ Low[c,f]	Music/ vocal interventions may not reduce postnatal depression.[c,f]
PARENTAL ANXIETY: STAI-SKD assessed with: State Anxiety Inventory	The mean PARENTAL ANXIETY: STAI-SKD ranged from **8.5 to 43.79**	MD **0.15 lower** (2.72 lower to 2.41 higher)	-	87 (3 RCTs)	⊕⊖⊖⊖ Very low[b,c,g]	The evidence is very uncertain about the effect of music/ vocal interven-

Scale from: 6 to 80 (lower = favourable)				tions on parental state anxiety.[b,c,g]

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).

CI: confidence interval; **MD:** mean difference

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

See interactive version of this table: https://gdt.gradepro.org/presentations/#/isof/isof_question_revman_web_433263637383142806.

[a] The evidence is certain. We did not downgrade since sensitivity analysis demonstrates that reducing analysis to studies without high risk of bias did not change the overall results.
[b] The confidence intervals for the effect are consistent with both an appreciable benefit and appreciable harm. We downgraded by one level for imprecision.
[c] Under 100 participants in total. We downgraded by one level.
[d] Three of four studies have a high risk of bias. We downgraded by one level.
[e] Half of the studies have a high risk of bias. We downgraded by one level.
[f] A high risk of bias in detection bias in one of two studies. We downgraded by one level.
[g] Two of three studies have a high risk of bias. We downgraded by one level.

BACKGROUND

Approximately 15 million infants are born preterm yearly, constituting more than 10% of all infants born worldwide (Chawanpaiboon 2019). Advances in technology and treatments have increased survival rates and reduced morbidity in preterm infants. However, preterm birth interferes with brain maturation, and subsequent clinical events and interventions may have additional deleterious effects (Ment 2008; Stoll 2015). Therefore, various non-pharmacological, therapeutic, or individual developmental care interventions have emerged that aim to improve health outcomes and quality of life for preterm infants and their parents (Aita 2021; Symington 2006). Music as therapy is one such intervention and is used increasingly in neonatal intensive care units (NICU). This has been studied in observational and experimental designs (Van der Heijden 2016; Yue 2021).

The sense of foetal hearing has been shown to develop as early as 16 weeks' gestation (Hepper 1994). Auditory perception has already developed when preterm infants are born. Studies suggest that the foetus responds to sound at least as early as 25 weeks to 27 weeks of gestational age (Clark-Gambelunghe 2015; Hepper 1994; Monson 2018). Intrauterine sounds encompass characteristics of organised sounds that are highly musical in nature. The maternal heartbeat, for instance, is rhythmic, and the foetus primarily hears the musical parameters of speech: melody, rhythm, prosody (patterns of stress and intonation), phonemes (sounds that distinguish one word from another), and pitch contour of the maternal voice and external voices (Moon 2013; Partanen 2013; Philbin 2017). Music promotes neuronal activation, and many researchers suggest that musical learning starts prior to birth (Huotilainen 2010; Perani 2010). In preterm birth, the enclosed intrauterine environment optimal for foetal growth and maturation is abandoned too early. Aside from other stressful experiences, such as separation from the mother and father, preterm infants must also adjust to the unusual - and potentially noxious - sound environment of an intensive care unit (Kuhn 2013; Park 2014; Rossetti 2013).

Appropriate auditory stimulation and social contact for preterm infants are desirable (Anderson 2018). Music therapy may provide environmental and socio-emotional enrichment through meaningful auditory stimulation and social contact (Anderson 2018; Haslbeck 2018; Loewy 2015; Shoemark 2015). This may be particularly warranted following preterm birth, as preterm infants are at risk of neurodevelopmental impairment, parents are at risk of post-traumatic stress disorders, and parents and preterm babies risk attachment and bonding difficulties (Forcada-Guex 2006; Korja 2011; Ruiz 2018). However, the precise effects of various musical and vocal stimulation types on short- and long-term outcomes in preterm infants and their parents remain ambiguous.

Description of the condition

Preterm birth is a significant determinant of neurodevelopmental delay, and the resulting impairment can have adverse long-term health effects (Pierrat 2017; Twilhaar 2018). It is sometimes associated with negative quality of life consequences, and an increased financial burden for the family and healthcare system (Lakshmanan 2021). Preterm infants face a range of morbidities, such as bradycardia, apnoea, anaemia, and respiratory distress syndrome. These infants are at risk of brain injury, and may have reduced white- and grey-matter volumes (Inder 2005). Such brain-structure abnormalities are associated with long-term neurodevelopmental impairments, including motor dysfunction, cerebral palsy, cognitive and behavioural problems, and deficits in executive function (Twilhaar 2018; Woodward 2006). Factors such as environmental noise and sensory deprivation (e.g. the lack of the regular intrauterine rhythms of the maternal heartbeat and the maternal voice) may also impact neurodevelopment negatively (Heim 1999; Lahav 2014; McMahon 2012). For many parents, preterm birth is a traumatic and lasting experience. They struggle with numerous problems and concerns, such as the uncertainty of the infant's future, feelings of fear, guilt, loss, grief, and confusion (Flacking 2007; Jotzo 2005; Roque 2017). These reactions may increase parental stress, adversely affect the stress-coping behaviour of their infant, and impair the formation of a secure attachment (Forcada-Guex 2006; Korja 2011; Malouf 2022).

Description of the intervention

Various musical and vocal interventions have been evaluated for efficacy in preterm infants (Haslbeck 2012; Mohan 2021; Van der Heijden 2016). They can be directed towards the infant (with or without parental involvement), to an entire family, or even applied within the whole NICU. The interventions aim to relax, stabilise and stimulate the infant and their parents (Hanson-Abromeit 2008).

Auditory stimulation for preterm infants and their parents incorporates calm music sung softly or played on an instrument. Examples include lullabies; improvised music; popular, New Age, classical, or family indigenous music; or song-of-kin, songs or sounds entrained to infant vital signs (i.e. synchronised with breathing or heart rate pattern) or based on the acoustic intrauterine environment (womb sounds, heartbeats, and parents' voices) (Hanson-Abromeit 2008; Haslbeck 2012; Loewy 2015; Mondanaro 2016). Music therapists, parents, nurses, doctors, nurses, and other healthcare professionals deliver the specific stimulation to the infants (and sometimes to their parents). These interventions are provided in addition to standard care in the NICU and are either performed live or recorded. The intervention is defined as music therapy when a trained music therapist provides the music within a therapeutic relationship and process facilitating personally tailored music experiences (Bradt 2015). Family-integrating music therapy approaches (Haslbeck 2020; Loewy 2015; Shoemark 2015), may be most appropriate to address family-centred recommendations in neonatal care, where the parents are seen as the most valuable resource for the infant (Lancet Child Adolescent Health 2019). In contrast, in music medicine, the music is administered by medical or healthcare professionals for passive listening (Dileo 1999).

How the intervention might work

The quality of early auditory experiences may have a direct influence on the plasticity of the brain's auditory regions and may affect cortex development in infants (Yan 2003). Both auditory overstimulation and sensory deprivation in the NICU may adversely affect preterm infants' short- and long-term neurobehavioural development, as the infants are already susceptible to neurodevelopmental impairment (Pineda 2014; Wachman 2011). Studies at the interface of music science and neuroscience suggest that music might promote neurobiological processes in humans, including the modulation of synaptic plasticity (linked to learning and memory), and might facilitate the differentiation, activation, readjustment, and growth of neurones (Abbott 2002; Rickard 2005; Sacks 2007). For instance, music

can alter brain activity in core structures involved in processing emotions (Koelsch 2014). Auditory stimulation, therefore, is recommended to enhance psychological and physiological health in preterm infants (Jobe 2014; Shoemark 2015).

Several systematic reviews suggest that musical and vocal interventions may stabilise and soothe preterm infants demonstrating beneficial effects on their behavioural states, physiological parameters, sleep quality, oral feeding, pain, and maternal anxiety (Anderson 2018; Hartling 2009; Haslbeck 2012; Hodges 2010; Mohan 2021; Standley 2012; Tramo 2011; Van der Heijden 2016; Yue 2021). The Van der Heijden 2016 review suggested that music may improve heart rate, sleep, feeding, and sucking outcomes in preterm infants. A meta-analysis by Bieleninik 2016 could not confirm or refute beneficial effects on those outcomes but did find a favourable impact of music on the infants' respiratory rate, and additionally demonstrated a reduction of maternal anxiety when parents were integrated into the music intervention process.

Why it is important to do this review

A number of systematic reviews have demonstrated ambiguous results for the efficacy of various types of auditory stimulation on preterm infants. Most of the reviews focused on a specific topic (e.g. maternal voice (Krueger 2010); music (Hartling 2009); or music interventions carried out by or in consultation with a trained music therapist (Bieleninik 2016)). The authors of these reviews concluded that the heterogeneity and clinical diversity of the included studies prevented the drawing of definitive conclusions about the impact of auditory stimulation on preterm infants (Hartling 2009; Haslbeck 2012; Hodges 2010; Krueger 2010; Standley 2012; Van der Heijden 2016). A more recent meta-analysis (Yue 2021), reported the significant positive influence of any music intervention on preterm infants respiratory rate, heart rate, oral feeding volume, stress level, and maternal anxiety. However, detailed reported criteria for considering studies for inclusion and assessment of the certainty of evidence are missing. Therefore, a more comprehensive and rigorous systematic review is needed to address existing controversies arising from apparent conflicting studies and reviews. Firstly, we evaluate the overall efficacy of auditory stimulation. Then, by analysing the impact of various types of auditory stimulation systematically with subgroup analysis, and by focusing on the methodological quality of the included studies, we may be able to provide better guidance. We may be able to determine how to use these interventions most effectively to promote specific outcomes in preterm infants and their parents (e.g. live versus recorded versions; sung versus instrumental; choices made in rendering decisions regarding length and time of intervention, associated keys, etc.). The current review should assist health professionals in neonatal care to make practical, evidence-based decisions about the use of musical and vocal interventions for preterm infants and their parents. If such a low-cost, low-risk intervention is demonstrated to be effective in supporting preterm infants' neurodevelopment and parental well-being, the findings could have significant clinical implications for this vulnerable patient population.

OBJECTIVES

We assessed the overall efficacy of music and vocal interventions for physiological and neurodevelopmental outcomes in preterm infants (< 37 weeks' gestation), compared to standard care. In addition, we aimed to determine specific effects of various music and vocal interventions for physiological, anthropometric, social-emotional, neurodevelopmental short- and long-term outcomes in preterm infants, parental well-being, and bonding.

METHODS

Criteria for considering studies for this review

Types of studies

We included parallel, cluster, and factorial randomised controlled trials. To avoid bias by carry-over effects, we included only the first phase of cross-over trials for short-term outcomes.

Types of participants

We included preterm infants of less than 37 weeks' gestational age, during hospitalisation and parents, only when they were involved in the music or vocal intervention (listening to it with their infant, or providing it for their infant or themselves, in relation to their infant, e.g. to support singing to their infant).

Types of interventions

We included any music or vocal stimulation, provided live or by recording, addressing either the infant alone or also the parents. The music intervention could be combined with another intervention, such as skin-to-skin care, but only if both arms of the study received the additional intervention. We included studies that examined a combination of interventions versus only music or voice. We also included studies that compared one type of music or voice to another type of music or voice and analysed them separately. Music interventions during painful procedures were included and would have been analysed separately. A parent, music therapist, musician, doctor, nurse, or other health professional or caregiver could deliver the intervention. The intervention must include musical elements, such as rhythm and melody, or sounds based on the acoustic intrauterine environment, e.g. womb sounds, heartbeats, and the human voice.

The intervention duration must comprise at least five minutes and the intervention must be delivered/administered at least three times for inclusion in the review. The intervention period may include any time from birth to hospital discharge. We compared the interventions with standard care without musical or vocal stimulation. We excluded auditory stimulation with white noise (noise with constant amplitude throughout the audible frequency range) or digital signals. These stimulation types lack musical parameters such as melody and prosody.

Types of outcome measures

Primary outcomes

Short-term outcome in preterm infants:

- Change in mean oxygen saturation during and post-intervention[a]

Long-term outcome in preterm infants:

- Infant development (assessed using Bayley Scales of Infant and Toddler Development (BSID-II and III),[b] focusing on mean mental development index (MDI) scores and psychomotor

development index (PSI) scores at two years of corrected age (Johnson 2008))

Outcomes in parents:

- Change in anxiety[b], defined as mean State-Trait Anxiety Inventory Score with 20 items and a four-point Likert Scale (Spielberger 1983).

Secondary outcomes

Short-term outcomes in preterm infants:

- Heart rate: beats per minute during and post-intervention[a] (measured by pulse oximetry or electrocardiogram);
- Respiratory rate during and post-intervention[a]: inspiration per minute[a] (measured by, e.g. electric strain-gauges, thoracic impedance plethysmography, nasal air-flow sensor and spirometers);
- Heart rate variability during and post-intervention[a] (measured by low-frequency power (ms^2/Hz); high-frequency power (ms^2/Hz); low frequency/high-frequency ratio, reflecting the balance between sympathetic and parasympathetic tone);
- Behavioural outcomes[b] (measured with behavioural numerical scores or scales for neonates, e.g. Assessment of Preterm Infant Behaviour (Als 2005));
- Hospitalisation (days);
- Adverse effects, including severe apnoea during the intervention requiring stimulation by the neonatal care team; and
- Weight gain (kg/day).

Long-term outcomes in preterm infants:

- Neurodevelopment (assessed by standardised follow-up examinations, e.g. intelligence quotient (Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R) (Park & Demakis 2017); Kaufmann Assessment Battery for Children (K-ABC II) at five years of corrected age (Melchers 2009)

Outcomes in parents:

- Well-being (measured with e.g. the Edinburgh Postnatal Depression Scale);
- Attachment[b] (measured with standardised scales, e.g. Postpartum Bonding Questionnaire (Hoffenkamp 2015)).

[a]assessed up to 30 minutes before, during, and 30 minutes after each musical intervention or control condition; reported at study level as mean changes or assessed after the last measurement round of musical intervention or control condition

[b]assessed before and after the whole intervention or control period

Search methods for identification of studies

The Neonatal Group Information Specialist developed search strategies in consultation with the authors. The MEDLINE strategy was translated, using appropriate syntax, for other databases. Search strategies combine intervention terms with standard terms for the neonatal population. Methodological filters were used to limit retrieval to randomised controlled trials and systematic reviews. Searches were conducted without date, language, or publication status limits.

Database search results were reduplicated using a combination of methods as follows: deduplication of MEDLINE and Embase results in OVID; all results reduplicated in EndNote and Covidence.

Trial registries reference lists of included studies and related systematic reviews, and conference proceedings were searched.

Note: The timeline for this publication was disrupted by the COVID-19 pandemic and staffing issues at the Cochrane Neonatal editorial base. As a result, publication of this review has been delayed, and the literature search is more than one year old. We will endeavour to undertake an updated search within the next calendar year.

Electronic searches

The following databases were searched November 1 to 12, 2021. Search strategies and dates of each search are presented in: Appendix 1; Appendix 2; Appendix 3; Appendix 4; Appendix 5; Appendix 6; Appendix 7; Appendix 8; Appendix 9.

- Cochrane Central Register of Controlled Trials (CENTRAL 2021, Issue #11) via CRS (Cochrane Register of Studies);
- Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Daily and Versions(R) (1946 to October 29, 2021);
- Embase (OVID) (1974 to October 29, 2021);
- CINAHL (1981 to November 1, 2021);
- PsycINFO (1806 to November 1, 2021);
- Web of Science (1982 to November 1, 2021);
- RILM Abstracts of Music Literature (1967 to November 1, 2021);
- ERIC (Educational Resources Information Center; 1966 to 15 November 1, 2021).

The following databases were searched in 2019 and strategies are presented in Appendix 10.

- Cochrane Central Register of Controlled Trials (CENTRAL 2019, Issue #11) in the Cochrane Library;
- Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Daily and Versions(R) (1946 to 15 November 2019);
- CINAHL (1981 to 15 November 2019);
- PsycINFO (1806 to 15 November 2019);
- Web of Science (1982 to 15 November 2019);
- RILM Abstracts of Music Literature (1967 to 15 November 2019);
- ProQuest Dissertations & Theses A&I (1637 to 15 November 2019);
- ERIC (Educational Resources Information Center; 1966 to 15 November 2019).

The search strategies and sources used in 2021 differ slightly from those run in 2019 - mainly by the addition of field qualifiers in Cochrane Central, CINAHL, PsycINFO, Web of Science, and RILM Abstracts; the addition of Embase; and the omission of ProQuest Dissertations, which we were unable to access in 2021 due to technical difficulties. Results of the 2021 search were reduplicated against results from 2019 and results of both searches are represented in the PRISMA flow diagram (Figure 1).

Figure 1. Study flow diagram

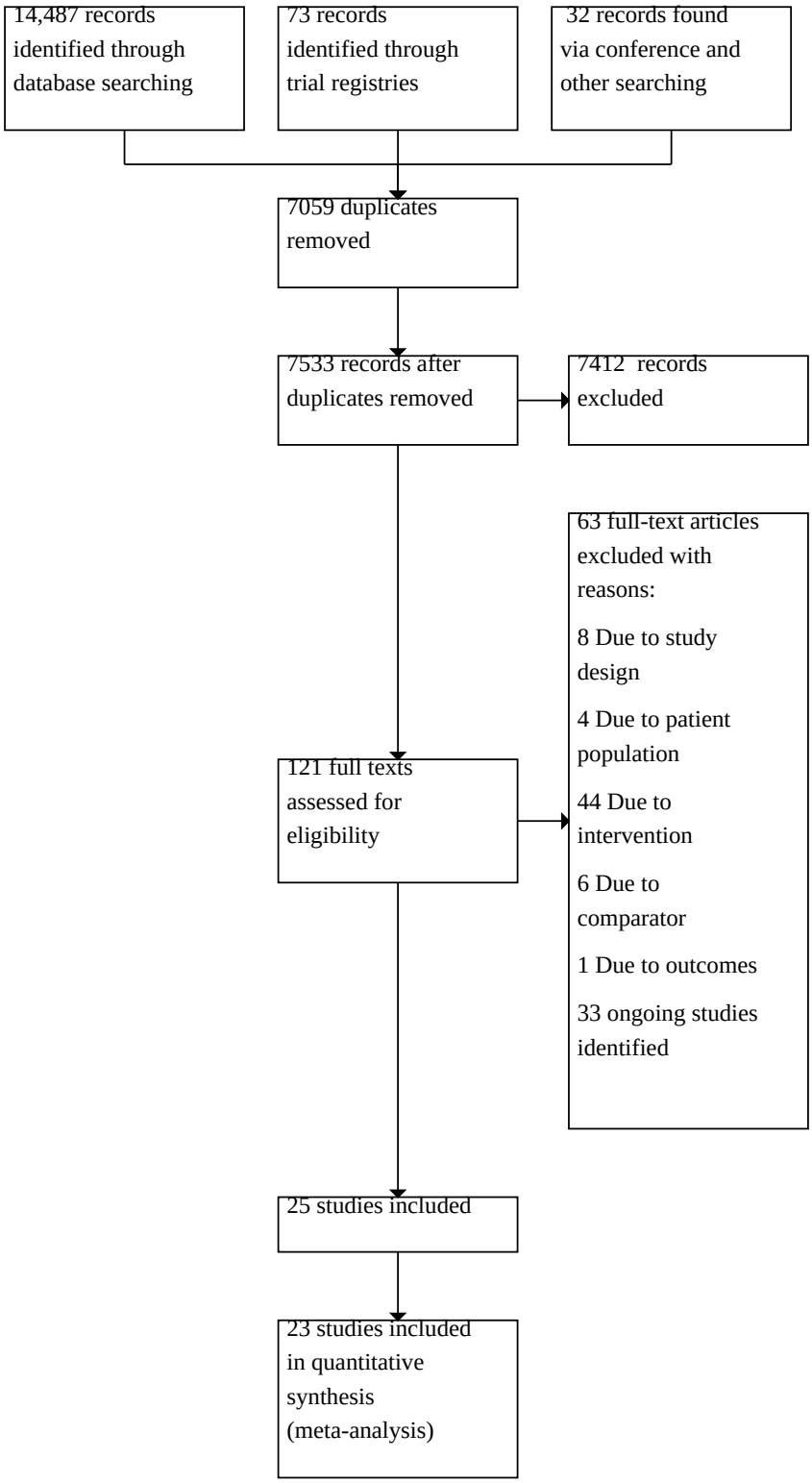


Figure 1. (Continued)

Searching other resources

The following trial registers were searched on 01 November 2021 and search strategies are presented in:

- International Clinical Trials Registry Platform (ICTRP), The World Health Organisation (https://trialsearch.who.int/Default.aspx);
- U.S. National Library of Medicine's https://clinicaltrials.gov;
- ISRCTN Registry (https://www.isrctn.com).

Conference abstracts were searched for the following conferences:

- Paediatric Academic Societies;
- European Society for Paediatric Research;
- European Association of Music Therapy;
- American Music Therapy Association;
- World Federation of Music Therapy.

We reviewed the reference lists of all included studies and related systematic reviews in an effort to identify studies not captured by database searches.

Data collection and analysis

We used the standard methods and criteria of the Cochrane Collaboration and its Neonatal Group to assess the methodological quality of the trials:

FH and DB were not involved in those activities for Haslbeck 2020a. TK and KM assessed this study instead. JL was not engaged in assessing methodological quality of the trials in general, so she was not involved in any assessment of her research (Loewy 2013). JM determined the final overall inclusion and exclusion criteria.

Selection of studies

We downloaded all titles and abstracts retrieved by electronic searching to the review management software Covidence to manage the results of our search. Two review authors (FH, KM) independently assessed study eligibility for inclusion in the review according to the prespecified selection criteria. They screened titles and abstracts to remove obviously irrelevant reports. The review authors linked together multiple reports of the same study so that each study rather than each report was the unit of interest in the review. They examined full-text reports to establish the compliance of studies with the eligibility criteria. If trial eligibility was unclear, they resolved discrepancies through discussion with the other review authors to reach a consensus. They listed all excluded studies with reasons for exclusion. They recorded the selection process in sufficient detail to complete a flow diagram (Figure 1) and Characteristics of excluded studies table.

Data extraction and management

Three review authors (FH, KM, TK) independently conducted data extraction using and adapting the most recent version of the Cochrane data collection form (Higgins 2019), in Covidence. They pretested the adapted version with a subset of five studies before general application. They used the adapted form to decide trial inclusion or exclusion and to extract data from eligible trials.

They (FH, KM, TK) extracted the following characteristics from each included study.

- Administrative details: study authors; published or unpublished, year of publication, CRS ID, sponsorship source;
- Study characteristics: study design type, study grouping, study setting, number of study centres, location, and contact;
- Participants: number randomised, the number lost to follow-up/withdrawn, number analysed, age, gender and further baseline characteristics in the infants and parents, inclusion and exclusion criteria, the reason for dropouts, reasons for exclusion, sample size calculation;
- Interventions: detailed description of music and voice type (music therapy or music medicine, musical parameters, instruments, music genre or music piece/voice genre or particular chosen voice, music selection, intervention alone or combined), dose, duration, frequency, mode of delivery (live, infant-directed entrained/recorded or standardised and dB level);
- Outcomes as mentioned under Types of outcome measures.

If any queries arose or when data appeared to be missing, they requested additional information from the authors of the original reports, e.g. when outcomes of interest were not reported. They described ongoing studies identified by their search, when available, detailing the authors, study reference, study name, methods, participants, interventions, outcomes, starting date, and contact information. If there were disagreements when comparing extracted data, they resolved them in consultation with the other review authors. Two review authors (FH, KM) entered and cross-checked data using Review Manager Web (RevMan Web 2022).

Assessment of risk of bias in included studies

Three review authors (FH, KM, TK) independently assessed the risk of bias (low, high, or unclear) of all included trials using the Cochrane Risk of bias tool (Higgins 2017), for the following domains.

• Sequence generation (selection bias);

• Allocation concealment (selection bias);

• Blinding of participants and personnel (performance bias);

• Blinding of outcome assessment (detection bias);

• Incomplete outcome data (attrition bias);

• Selective reporting (reporting bias);

• Any other bias.

We looked for evidence of bias or methodological differences between trials.

We resolved any disagreements through discussion with the other review authors to reach a consensus. See Appendix 11 for a more detailed description of risk of bias for each domain.

Measures of treatment effect

The treatment effects of the individual trials were analysed using Review Manager Web (RevMan Web 2022). For dichotomous data, we planned to use risk ratio (RR) and risk difference (RD) with 95% confidence intervals (CIs). If the difference between groups had been statistically significant, we would have calculated the number needed to treat for an additional beneficial outcome (NNTB) and the number needed to treat for an additional harmful outcome (NNTH), with their respective CIs. We evaluated continuous data by assessing the mean difference (MD) with its 95% CI when measured in the same way between trials. If studies had reported the same outcome but measured it in different ways, we would have used the standardised mean difference (SMD) with its 95% CI. Where summary statistics would have been missing, we would have derived them from the accompanying P values.

We analysed short-term cross-over trials to determine if there would be no significant risk of a carry-over effect. We calculated an effect estimate using the generic inverse variance method described in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2019). We incorporated cross-over trials into meta-analyses using the methods described in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2019). If we had identified cluster trials, we would have incorporated them using generic variance methods for analysis (Higgins 2019).

Unit of analysis issues

We performed the primary analysis per individual randomised. If the following issues had occurred, we planned to address them according to the methods described below.

Cluster-randomised trials

In cluster-randomised trials, groups of participants rather than individuals are randomised to different interventions. Because of this, participant data can no longer be assumed to be independent of one another. Unfortunately, some cluster-randomised trials are not analysed correctly, i.e. do not take into account that the unit of allocation (the group) is different from the unit of analysis (the individual). If this clustering is ignored, there is a unit of analysis error, which means that the resulting P values and 95% CIs will be artificially small and lead to an inappropriately increased weight in the meta-analysis. If cluster-randomised trials had failed to report results based on appropriate analyses such as the multi-level model or variance component analysis, we would have used the methods described in the *Cochrane Handbook for Systematic Reviews of Interventions* (chapter 16.3.3) (Higgins 2019), to reanalyse these trials with appropriate consideration of the intra-cluster (or intraclass) correlation coefficient (ICC) to estimate the effective sample size. Sensitivity analyses would have been performed to explore whether there were any differences in effects between cluster- and individually randomised trials.

Cross-over trials

Cross-over trials are suitable for evaluating interventions with a temporary effect in the treatment of stable conditions. The principal problem is that of carry-over (a type of period-by-intervention interaction). Since we believe that some carry-over from period one to period two cannot be precluded in these trials in our setting, we included only the data from the first period (as suggested in chapter 16.4.3 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2019)). We requested the data from the first phase of the cross-over trial from the authors to avoid bias in the carry-over effect. In cross-over trials, we assessed whether the short-term cross-over design is suitable, whether there is a carry-over effect, whether only first-period data are available, whether the analysis is correct, and whether the results are comparable with those from parallel-group trials.

Studies with more than two intervention groups (multi-arm studies)

If more than one comparison arm from the same trial was eligible for inclusion in the same meta-analysis, we combined the live or recorded music intervention groups to create a single pairwise comparison so that the same participants did not contribute data to the meta-analysis more than once according to the methods described in the *Cochrane Handbook for Systematic Reviews of Interventions (*chapter 16.5.4) (Higgins 2019). We included all music and vocal intervention groups meeting our inclusion criteria in our synthesis since all interventions were relevant to our review. We (FH & KM) used the formulae in Table 6.5.a (6.5.2.10 Combining groups#section-6-5-2-10) to combine numbers into a single sample size, mean, and SD for each intervention group. We calculated independently, compared results, and where differences occurred, we calculated again to reach a consensus. To reflect the fact that comparisons within multi-arm studies are correlated, we adjusted the standard error of each two-arm comparison from a multi-arm study. We used the method proposed by Rücker and Schwarzer which uses back-calculated standard errors in the weighted least-square estimator to reflect the within-study correlation (Rücker 2012; Rücker 2014; Rücker 2016). In one study (Namjoo 2021), which compared a recorded versus a live lullaby intervention, we appropriately reduced the sample size of the control group so that the same participants did not contribute data to the meta-analysis more than once according to the methods described in the *Cochrane Handbook for Systematic Reviews of Interventions* (chapter 16.5.4) (Higgins 2019).

Multiple measurements of outcomes

When primary outcomes were assessed at more than one time point in our time ranges, we used the data from the latest time point available in our analyses. We did not plan to adjust for multiplicity in our review based on multiple outcome measurements. Considering the heterogeneity of time points for short-time outcomes, we identified two different, but overall used time points: a) during intervention; b) post-intervention. Since both time points are clinically relevant and were available in most of the included studies, we decided to conduct these two analyses: a) during intervention; b) post-intervention in parallel. Since not all authors reported the data for the different time points in detail, the first author (FH) contacted and requested the missing data.

Dealing with missing data

We contacted the authors whenever we detected that data and statistics were missing or incomplete to request further information. However, when data were missing due to dropouts, we included the reported infants and examined the effect of losses in a sensitivity analysis according to the risk of bias. We contacted

the primary investigators. If authors were unable or unwilling to provide the data, we still included the study in the review and explicitly stated that data were missing.

Assessment of heterogeneity

We describe the clinical diversity and methodological variability of the evidence in the review text and with study tables, which describe study characteristics including design features, population characteristics, and intervention details. To assess statistical heterogeneity, we visually inspected forest plots and described the direction and magnitude of the effect and the degree of overlap between confidence intervals. We estimated treatment effects in individual trials and examined heterogeneity between trials by inspecting forest plots and quantifying the impact of heterogeneity by using the I^2 statistic, a measure that describes the proportion of variation in point estimates that is due to variability across studies rather than sampling error (Deeks 2017). We interpreted the results as follows:

- Less than 25%: no heterogeneity;
- 25% to 49%: low heterogeneity;
- 50% to 74%: moderate heterogeneity;
- 75% to 100%: high heterogeneity.

Assessment of reporting biases

We assessed reporting bias by comparing the stated primary outcomes and secondary outcomes and reported outcomes. Where study protocols were available, we compared these to the full publications to determine the likelihood of reporting bias. Studies using the interventions in a potentially eligible infant population but not reporting on any of the primary and secondary outcomes would have been documented in the Characteristics of included studies table.

For outcomes reported by more than 10 studies, we planned to prepare a funnel plot to assess possible reporting bias. If publication bias had been suggested by a significant asymmetry of the funnel plot on visual assessment, we would have incorporated this in our assessment of the certainty of evidence (Egger 1997). If our review included few studies eligible for meta-analysis, the ability to detect publication bias would be largely diminished, and we would simply note our inability to rule out possible publication bias or small study effects.

Data synthesis

We used the standard methods of Cochrane and Cochrane Neonatal to perform statistical analysis (neonatal.cochrane.org/resources-review-authors). The treatment effects of all infants in the eligible trials were analysed. If we identified multiple studies that we considered to be sufficiently similar, we performed meta-analysis using Review Manager Web (RevMan Web 2022). We used a fixed-effects model to combine the data. We planned to calculate average estimates of RR and RD with 95% CIs for any meta-analyses where required. We used the MD with 95% CIs for continuous outcomes that were measured in the same way between trials. We planned to calculate the standardised mean difference (SMD) with 95% CIs to combine trials that would have measured the same outcome but used different scales where required. Individual trials were interpreted separately when a meta-analysis appeared to be inappropriate based on clinical judgement and the I^2 heterogeneity test (i.e. when $I^2 > 80\%$). When the I^2 statistic was higher than 50%, we reported the finding and assessed the source of the heterogeneity (e.g. differences in study quality, participants, intervention regimens, or outcome assessments) by sensitivity and subgroup analysis (Subgroup analysis and investigation of heterogeneity).

Subgroup analysis and investigation of heterogeneity

Tests for subgroup differences in effects were interpreted with caution given the potential for confounding with other study characteristics. In particular, subgroup analyses with fewer than five studies per category are unlikely to be adequate to ascertain the valid differences in effects and were not highlighted in our results. When subgroup comparisons were possible, stratified meta-analysis and a formal statistical test for interaction were conducted to examine subgroup differences that could account for effect heterogeneity (e.g. Cochran's Q test, meta-regression) (Higgins 2019).

According to the heterogeneity of auditory stimulation types, we planned to compare the following modalities separately, if a sufficient number of studies were identified.

- **Auditory stimulation**
  - Spoken voice;
  - Sung voice;
  - Music without a voice;
  - Womb sounds;
  - Rhythmic sounds; or
  - Breathing sounds.
- **Auditory stimulation**
  - Live, infant-directed, or entrained music; or
  - Recorded or standardised music.
- **Musical decision or selection**
  - By parent; or
  - Random, unidentified, or unknown.
- **Duration of intervention**
  - Between five and 10 minutes; or
  - More than 10 minutes.
- **Frequency of intervention**
  - Between three and seven times; or
  - At least eight times.
- **Auditory stimulation**
  - Alone; or
  - Combined with other interventions (e.g. skin-to-skin care).
- **Painful procedure**
  - With auditory stimulation; or
  - Without auditory stimulation.
- **Gestational age**
  - Extremely preterm (less than 28 weeks gestation);
  - Very preterm (28 to 32 weeks gestation);
  - Moderate to late preterm (32 to 37 weeks gestation).

Given that studies in a variety of settings may not have reliable gestational age and may therefore use birth weight categories, we planned to include infants categorised as follows.

- Low birth weight (LBW) infants – defined as infants with birth weight < 2500 g;
- Very low birth weight (VLBW) infants – defined as infants with birth weight < 1500 g;
- Extremely low birth weight (ELBW) infants – defined as infants with birth weight < 1000 g.

We planned to group infants with birth weights 1500 to 2499 g with moderate preterm infants, infants 1000 to 1499 g with very preterm infants, and infants < 1000 g with extremely preterm infants.

However, since there were insufficient studies or details to be able to distinguish the listed modalities clearly, we only assessed possible differences between subgroups of recorded versus live music and frequency of intervention by using the formal test for subgroup differences in Review Manager Web (RevMan Web 2022).

Sensitivity analysis

We planned to perform sensitivity analyses where sufficient data were available to explore methodological heterogeneity. We considered characteristics of bias (Assessment of risk of bias in included studies). It is worthwhile to note that double-blinding has scarcely been possible in intervention designs with live music. Consequently, it is all the more crucial that the outcome assessors are blind to the data. In our sensitivity analyses, we excluded trials with a high risk of bias for any of the following: allocation concealment, adequate randomisation, and blinding of outcome assessment (Schulz 1994; Schulz 2000). Additionally, we considered the characteristics of participants (e.g. participants with high morbidities). Given that there is no formal statistical test that can be used for sensitivity analysis, we made informal comparisons between the different ways of estimating the effect under different assumptions. Changes in the P values should not be used to judge whether there is a difference between the main analysis and sensitivity analysis, since statistical significance may be lost with fewer studies included. We reported sensitivity analysis results narratively.

Summary of findings and assessment of the certainty of the evidence

We used the GRADE approach, as outlined in the GRADE Handbook (Schünemann 2013), to assess the certainty of evidence of the following (clinically relevant) outcomes.

- Oxygen saturation during and post-intervention;
- Infant development assessed using the Bayley Scales of Infant and Toddler Development at two years;
- Parental anxiety assessed using the State-Trait-Anxiety Inventory Scores;
- Respiration rate during and post-intervention;
- Heart rate during and post-intervention;
- Parental well-being assessed with the Edinburgh Postnatal Depression Scale;
- Parental anxiety assessed using the State-Anxiety Inventory Scores.

Two review authors (FH, KM) independently assessed the certainty of the evidence for each of the outcomes above. We considered evidence from randomised controlled trials as high certainty, downgrading the evidence to one level for serious (or two levels for very serious) limitations based upon the following: design (risk of bias), consistency across studies, directness of the evidence, precision of estimates, and presence of publication bias. We used the GRADEproGDT Guideline Development Tool to create Summary of findings 1 to report the certainty of the evidence.

The GRADE approach results in an assessment of the certainty of a body of evidence in one of the following four grades.

- High: we are very certain that the true effect lies close to that of the estimate of the effect.
- Moderate: we are moderately certain in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
- Low: our certainty in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
- Very low: we have very little certainty in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

RESULTS

Description of studies

For a full description of studies please see the Characteristics of included studies; Characteristics of excluded studies; Characteristics of studies awaiting classification; Characteristics of ongoing studies.

Results of the search

Database searches identified 14,487 records; trial registry searches identified 73 records; and conference abstract searching identified 32 records (14,592 total). After removing 7059 duplicates, 7533 records were screened. We excluded 7412 records during screening and reviewed 121 full texts for eligibility. During the full-text screening, 33 ongoing studies were identified and no awaiting classification studies found. We excluded 63 studies, finally including 25 trials; for details see Figure 1.

Included studies

Participants

Twenty-five studies recruiting 1532 infants and 691 parents, of which 122 mothers of two studies (Namjoo 2021; Wirth 2016), were included to use their (recorded) voice as an intervention without assessing additional parental outcomes. Study sample sizes ranged from 17 (Calabro 2003; Caparros-Gonzalez 2018), to 272 (Loewy 2013). Most of the studies included preterm infants, whereas 10 studies additionally included parents, of which only three music therapy studies (Kehl 2021; Loewy 2013; Menke 2021), included fathers in the analysis. The gestational age at birth for recruitment varied from 23 to 36 weeks, with only four music therapy studies including extremely preterm infants (Epstein 2021; Haslbeck 2021; Kraft 2021; Menke 2021) (Table 1).

Setting

We identified 21 parallel-group RCTs and 4 cross-over RCTs (Epstein 2021; Johnston 2007; Kraft 2021; Loewy 2013). All trials were single-centre studies except one multi-centre trial (Loewy 2013). Nine RCTs examined two intervention arms, two RCTs evaluated three intervention arms (Loewy 2013; Kucuk Alemdar 2020), and the remaining RCTs compared one intervention with the control

condition. The trials took place in NICUs from levels one to three. Most of the studies have been conducted since the 2010s except for four studies from earlier years (Calabro 2003; Cevasco 2008; Johnston 2007; White-Traut 1988). The studies were performed by researchers from around the world, mainly from Europe (n = 8), followed by the Middle East (n = 6), USA (n = 4), Asia (n = 3), Australia (n =1), South America (n = 1) and Canada (n = 1) (see Table 1). Seven studies reported being funded by University/Health Department/ Hospital research funds (Farhat 2010; Jabraeili 2016; Johnston 2007; Kraft 2021; Kucuk Alemdar 2020; Liao 2021; Namjoo 2021). In addition, local medical/health foundations funded five projects (Calabro 2003; Kehl 2021; Loewy 2013; White-Traut 1988; Wirth 2016), and National Science Foundations supported two studies (Caparros-Gonzalez 2018; Lafferty 2021); the remaining studies did not mention the sponsorship source.

Intervention

We identified a range of music and vocal interventions varying widely in intervention type, delivery, frequency, and duration across studies mainly characterised by calm, soft, musical parameters in lullaby style, often integrating the sung mother's voice live or recorded. Seventeen music medicine studies were conducted (we also included the vocal intervention delivered by medical personnel as music medicine), and seven studies provided music therapy of whom six studies were based on Creative Music Therapy with preterm infants and their parents (Haslbeck 2020), or the Rhythm, Breath, and Lullaby approach (Loewy 2013). In these family-centred live music approaches, entrainment to the infant's breathing rhythm is consciously deployed as a technique. The music therapist continually adapts and tailors the music to the infant's and parents' needs and musical heritage in an ongoing individualised therapeutic reciprocal, family-integrating, and interactive process. Additional to these six live music therapy studies, two music medicine studies provided live music for the infants (n = 8), whereas the remaining two-thirds of the studies (n = 16) played recorded music and one study (n = 1) recorded the spoken voice. One of the studies (Namjoo 2021), compared a recorded lullaby to a live version. In a little over half of the studies (n = 14), the musical selection was random, unidentified, or unknown. In contrast, the remaining studies (n = 9) indicated that the parents selected the song/music in at least one intervention arm. More than two-thirds of the studies (n =17) used the sung voice (mostly lullabies) in at least one intervention arm whereas the remaining one-third used (additionally) spoken voice as the choice of intervention type (n = 7). Pure recorded instrumental music was rare (n = 4) characterised as calm and sedative music played by wind and string instruments (harp, string, flute) (Calabro 2003; Caparros-Gonzalez 2018; Lejeune 2019). One study (Lafferty 2021), provided piano music (Mozart's double piano sonata), and a few studies played womb sounds and rhythmic or breathing sounds. Most of the interventions were delivered exclusively whereas four studies provided music therapy during skin-to-skin care as an integral part of standard care in the unit (Epstein 2021; Ettenberger 2014; Haslbeck 2021; Menke 2021). Two studies were conducted while the infants were lying in the arm of the mother (Namjoo 2021; White-Traut 1988). Two studies examined the effects of music on pain (Johnston 2007; Yu 2021). Moreover, the dose of the interventions ranged from five to 30 minutes with most studies (n = 20) providing the intervention for more than 10 minutes and more than two-thirds of studies (n = 16) at least eight times. For further details see Table 2; Table 3; Table 4.

Excluded studies

After the abstract/title and records screening, we excluded 63 full-text studies for the following reasons.

- Eight of the 63 studies were identified as not suitable due to their study design.
- Four studies displayed unmet criteria of the patient population.
- Forty-four studies used an intervention outside our study purpose; for example, the intervention was delivered only once.
- Six studies were identified as not suitable due to the comparator.
- One study, was excluded, due to unmet criteria of the data outcomes.

For further information see the Characteristics of excluded studies and Figure 1.

Ongoing studies

There are 33 ongoing studies. For further details of those studies, please see the Characteristics of ongoing studies.

Risk of bias in included studies

The general risk of bias in the included studies varied from low (e.g. for allocation concealment) to high risk of bias (e.g. for blinding outcome assessors), often remaining unclear since detailed descriptions were missing for many risk of bias areas. The risk of bias is described in detail in the Characteristics of included studies and summarised in Figure 2; Figure 3.

Figure 2. Risk of bias graph

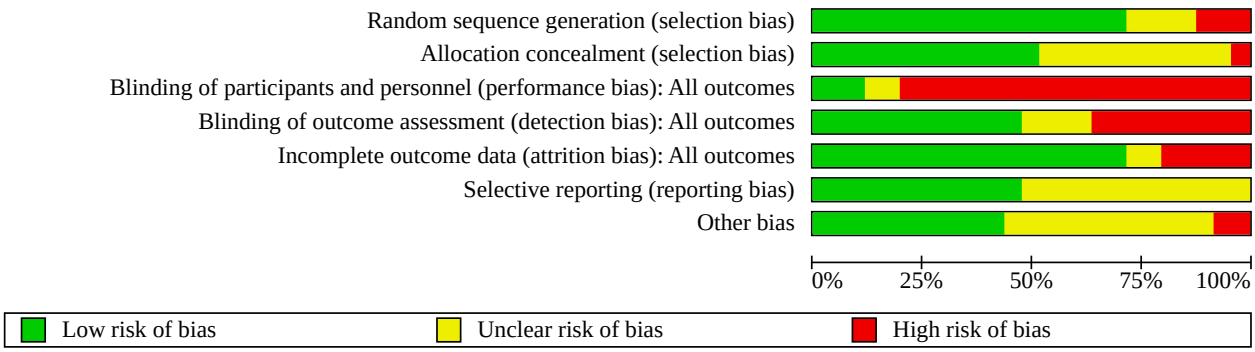


Figure 3. Risk of bias summary

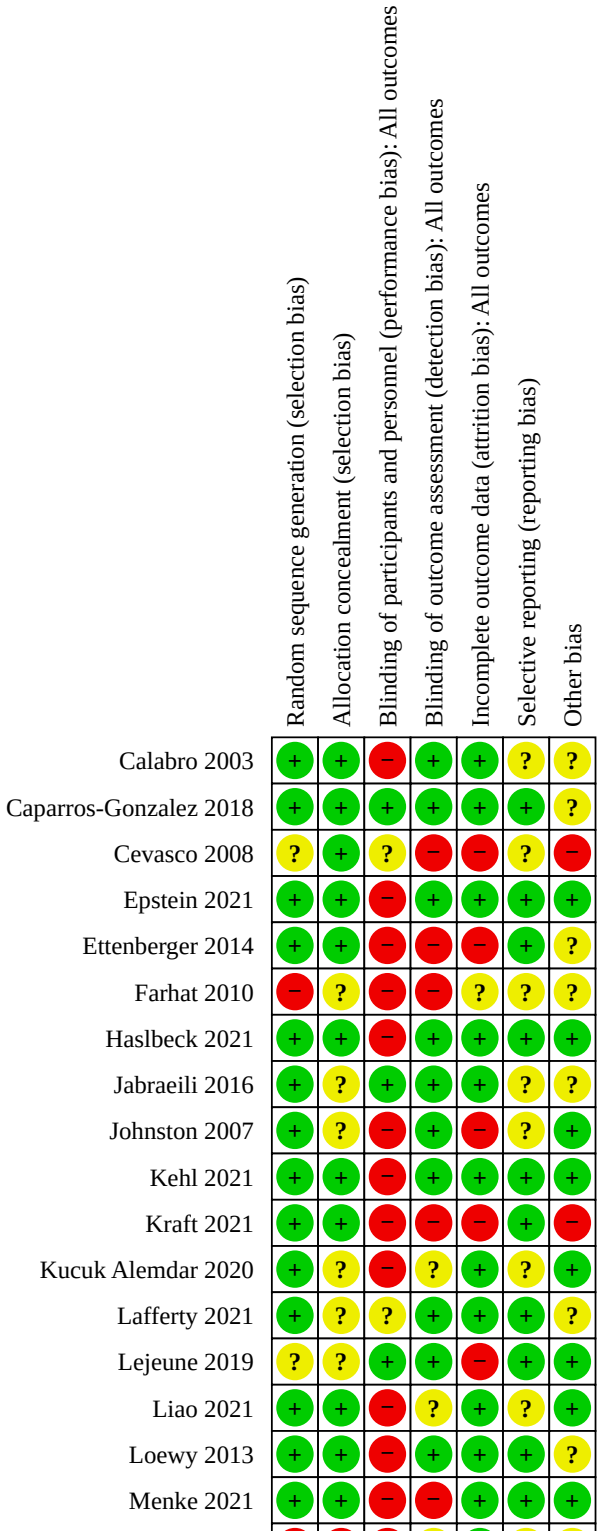
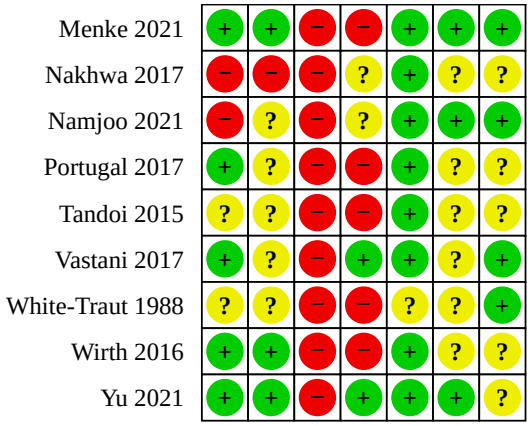


Figure 3. (Continued)



Allocation

Most of the studies clearly reported random sequence generation, except for five studies (Cevasco 2008; Lejeune 2019; Tandoi 2015; White-Traut 1988; Wirth 2016), in which the sequence generation was not clearly described, and three studies with a high risk of bias concerning random sequencing (Farhat 2010; Nakhwa 2017; Namjoo 2021).

For concealment of allocation, the risk of bias was low in 11 studies and unclear in 13 studies not reporting sufficient detail on if/how allocation concealment was achieved. Only one study displayed a high risk of allocation concealment (Nakhwa 2017) (see Figure 2; Figure 3).

Blinding

Blinding of both participants and personnel could not be achieved through the study design in 22 of the 25 studies, leading to a high risk of performance bias as anticipated in our study protocol (Haslbeck 2019). Three studies using prerecorded music involved delivering a 'sham' intervention (silent prerecorded track) to the control group to achieve blinding of personnel (Caparros-Gonzalez 2018; Jabraeili 2016; Lejeune 2019). Twelve studies reported blinding of outcome assessment. Nevertheless, nine studies demonstrated a high risk of detection bias by not blinding outcome assessors and, in four studies, there were insufficient details to determine if detection bias occurred (Kucuk Alemdar 2020; Liao 2021; Nakhwa 2017; Namjoo 2021).

Incomplete outcome data

Four studies were at high risk of attrition bias because of incomplete assessment of the trial cohort (Cevasco 2008; Ettenberger 2014; Johnston 2007; Lejeune 2019) and, in four studies, completeness of outcome data remained unclear (Farhat 2010; Kraft 2021; White-Traut 1988). The remaining studies displayed a low risk for attrition bias.

Selective reporting

In more than half of the included studies, we were not able to assess reporting bias since we could not identify study protocols or other sources of prespecified outcomes of most of the trials. We assessed eleven studies as having a low risk of selective reporting and considered one study to have a high risk of reporting bias due to missing outcome data (Kraft 2021), compared with the trial register entry of the trial.

Other potential sources of bias

Many studies showed no other sources of bias (n =11). We considered twelve studies to have an unclear risk of bias and two studies to have a high risk of bias due to limited baseline characteristics of the included mothers (Kraft 2021), and due to differences in reporting of the number of randomised preterm infants in the paper (Cevasco 2008).

Effects of interventions

See: **Summary of findings 1** Summary of findings table - Music and vocal interventions for preterm infants and their parents

See: Summary of findings 1

We included 25 studies, of which 23 studies were included in the meta-analyses. In two studies, data were incompletely reported and therefore not eligible to be included (Nakhwa 2017; White-Traut 1988). For short-term infant physiological outcomes, we used two different endpoints (during the intervention and post-intervention). We analysed studies during painful procedures separately. Where sufficient studies were available, we conducted subgroup analysis for recorded versus live musical/vocal interventions or for intervention frequency differences. There were insufficient studies for building further subgroups meaningfully.

Primary outcomes

Short-term outcome in preterm infants

Oxygen saturation

We identified 14 studies that reported measuring oxygen saturation. In one study (Ettenberger 2014), outcome data were not sufficiently reported. After receiving the requested data from the authors, the amount of missing data did not allow for inclusion of the study in an analysis, so we conducted meta-analyses using 13 studies (see Table 1).

During intervention

Ten studies provided data for this outcome (Calabro 2003; Epstein 2021; Farhat 2010; Jabraeili 2016; Kucuk Alemdar 2020; Liao 2021; Loewy 2013; Namjoo 2021; Portugal 2017; Tandoi 2015). Music and vocal interventions did not increase oxygen saturation in preterm infants during the intervention compared to standard care (mean difference (MD) 0.13, 95% CI -0.33 to 0.59; P = 0.59; 958 participants, 10 studies; high-certainty evidence; Analysis 1.1).

As heterogeneity was moderate (Chi^2 = 24.68, df = 10 (P = 0.006); I^2 = 59%), we conducted a sensitivity analysis excluding data from four studies with high selection or detection bias (Farhat 2010; Namjoo 2021; Portugal 2017; Tandoi 2015) so that heterogeneity was reduced to a low level (Chi^2 = 5.99, df = 5 (P = 0.31); I^2 = 17%). However, after sensitivity analysis, the overall effect on oxygen saturation showed no substantial changes (P = 0.45 without high-risk bias studies versus P = 0.59 with all included studies). Subgroup analysis comparing recorded versus live music showed no substantial changes either (recorded music: P = 0.40, live music: P = 0.24) nor clear subgroup differences (Chi^2 = 1.78, df = 1 (P = 0.18), I^2 = 43.9%; Analysis 1.1).

During intervention with heel lance

One study (Yu 2021) with 64 infants compared oxygen saturation during recorded maternal voice during heel lance with standard care without discovering a significant difference.

Post-intervention

Seven studies assessed oxygen saturation post-intervention (Calabro 2003; Caparros-Gonzalez 2018; Jabraeili 2016; Liao 2021; Loewy 2013; Namjoo 2021; Portugal 2017). Music and vocal interventions probably do not increase oxygen saturation post-intervention compared to standard care (MD 0.63, 95% CI -0.01 to 1.26; P = 0.05; 800 infants, 7 studies; moderate-certainty evidence; Analysis 1.2). Heterogeneity was low (Chi^2 = 7.65, df = 7 (P = 0.36); I^2 = 8%) and subgroup analysis comparing recorded versus live intervention revealed no clear subgroup differences (Chi^2 = 0.00, df = 1 (P = 0.98), I^2 = 0%) (see Analysis 1.2).

Post-intervention with heel lance

Two studies evaluated oxygen saturation after heel lance (Johnston 2007; Yu 2021). The evidence suggested that music and vocal interventions result in no substantial difference in oxygen saturation after the intervention compared to standard care after heel lance (MD 0.75, 95% CI -0.02 to 1.51; P = 0.06; 100 infants, 2 studies; Analysis 1.3). Heterogeneity was high (Chi^2 = 3.51, df = 1 (P = 0.06); I^2 = 71%). After removing the study with high attrition bias (Johnston 2007), the remaining study (Yu 2021), suggested a favourable effect of the music intervention on oxygen saturation compared to standard care (MD 0.93, 95% CI 0.14 to 1.72; P = 0.02; 64 infants; one study; Analysis 1.3).

Long-term outcome in preterm infants

Infant development

Two studies provided data on all three Bayley Scale composition scores (cognitive, motor, language) at two years of age (Haslbeck 2021; Lejeune 2019) (Table 1). Music/vocal interventions may not increase infant development or the cognitive composition score (MD 0.35, 95% CI -4.85 to 5.55; P = 0.90; 69 infants, 2 studies; low-certainty evidence; Analysis 1.4), the motor composition score (MD -0.17, 95% CI -5.45 to 5.11; P = 0.95; 69 infants, 2 studies; low-certainty evidence; Analysis 1.4), or the language composition score (MD 0.38, 95% CI -5.45 to 6.21; P = 0.90; 69 infants, 2 studies; low-certainty evidence; Analysis 1.4), although the CIs of all subscores included meaningful effects in both directions. Heterogeneity between the studies was minimal (I^2 = 0%).

Outcomes in parents

Change in state-trait-anxiety

Four music therapy studies (Ettenberger 2014; Kehl 2021; Kraft 2021; Menke 2021), reported data on parental anxiety after the whole treatment period, measured with the State-Trait-Anxiety Inventory (Table 1). Music and vocal interventions may not reduce parental state-trait anxiety (MD -1.12, 95% CI -3.20 to 0.96; P = 0.29; 97 parents, 4 studies; low-certainty evidence; Analysis 1.5), although the CIs included meaningful effects in favour of the music therapy group. Heterogeneity between the studies was minimal (I^2 = 0%) (see Analysis 1.5).

Secondary outcomes

Short-term outcomes in preterm infants

Respiratory rate

We identified 10 studies that reported measuring respiratory rate by e.g. electric strain-gauges, thoracic impedance plethysmography, nasal air-flow sensor, and spirometers. We conducted meta-analyses using nine studies. One study (Yu 2021), evaluated music during and after heel lance; these results are reported narratively (see Table 5).

During intervention

Seven studies provided data on respiratory rate during the intervention demonstrating that music and vocal interventions probably do not reduce respiratory rate compared to standard care (Calabro 2003; Epstein 2021; Kucuk Alemdar 2020; Loewy 2013; Portugal 2017; Tandoi 2015; Wirth 2016) (MD 0.42, 95% CI -1.05 to 1.90; P = 0.57; 750 infants, 7 studies; moderate-certainty evidence; Analysis 1.9). Heterogeneity was low (Chi^2 = 11.72, df = 6 (P = 0.07); I^2 = 49%; Analysis 1.9).

During intervention with heel lance

One study (Yu 2021) with 64 infants compared the respiratory rate during heel lance with the recorded maternal voice to standard care without discovering significant differences.

Post-intervention

Five studies (Calabro 2003; Caparros-Gonzalez 2018; Loewy 2013; Portugal 2017; Wirth 2016) provided data on respiratory rate post-intervention demonstrating that music and vocal interventions probably do not reduce respiratory rate compared to standard care (MD 0.51, 95% CI -1.57 to 2.58; P = 0.63; 636 infants, 5 studies; moderate-certainty evidence; Analysis 1.10). Heterogeneity was low (Chi^2 = 7.66, df = 4 (P = 0.10); I^2 = 48%; Analysis 1.10).

Post-intervention with heel lance

One study (Yu 2021) with 64 infants compared the respiratory rate after heel lance with the recorded maternal voice to standard care without discovering significant differences.

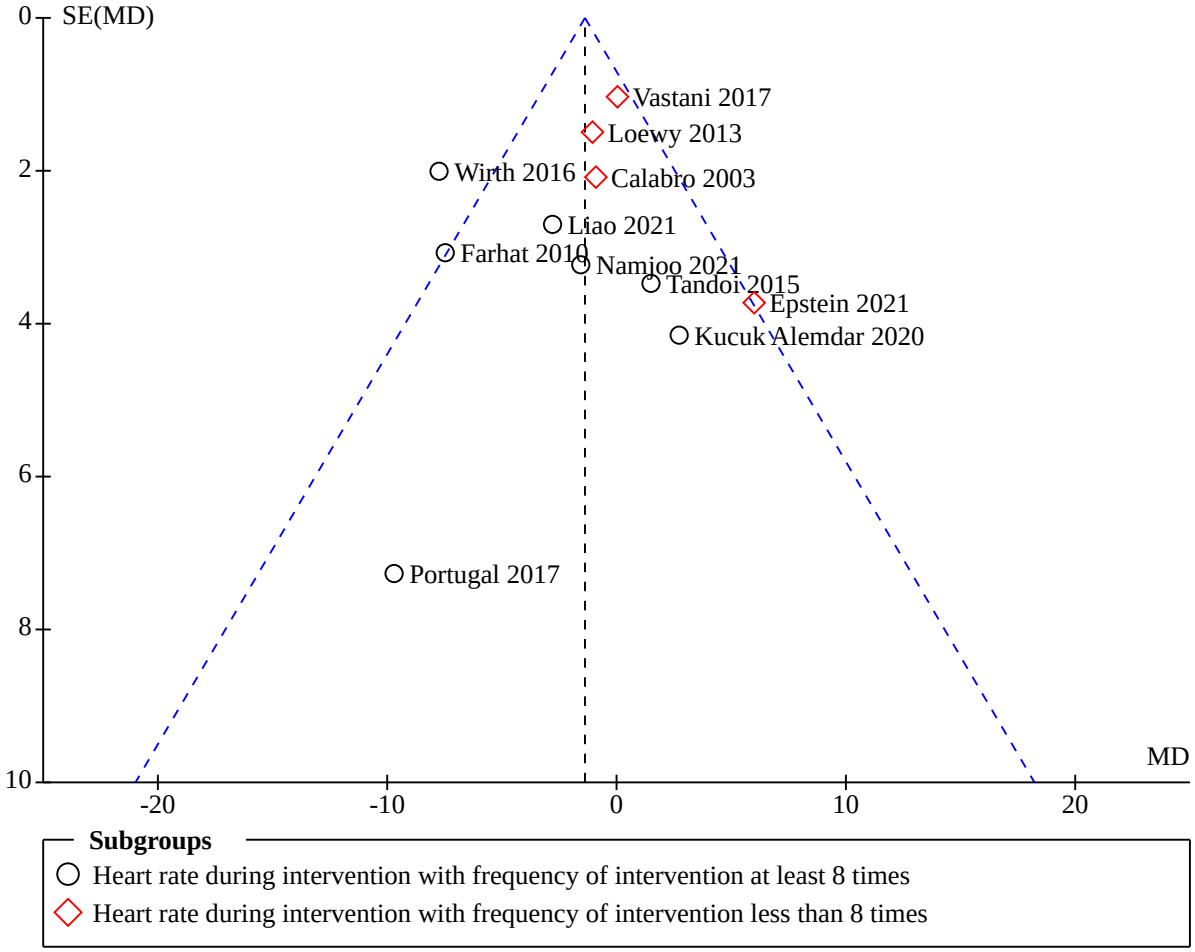
Heart rate

We identified 14 studies that reported measuring heart rate by pulse oximetry or electrocardiogram. In one of these studies (Ettenberger 2014), outcome data were not sufficiently reported. After receiving the requested data from the authors the amount of missing data did not allow us to include the study in an analysis, so we conducted meta-analyses using 13 studies (see Table 5).

During intervention

Eleven studies (Calabro 2003; Epstein 2021; Farhat 2010; Kucuk Alemdar 2020; Liao 2021; Loewy 2013; Namjoo 2021; Portugal 2017; Tandoi 2015; Vastani 2017; Wirth 2016) provided data on this outcome. Music and vocal interventions probably reduce heart rates in preterm infants during the intervention compared to standard care (MD -1.38, 95% CI -2.63 to -0.12; P = 0.03; 1014 infants, 11 studies; moderate-certainty evidence; Analysis 1.6) with the CI ranging from small to medium effects favouring the intervention. Heterogeneity was moderate (Chi^2 = 23.15, df = 10 (P = 0.01); I^2 = 57%). Significant subgroup differences (Chi^2 = 8.66, df = 1 (P = 0.003), I^2 = 88.5%) indicated a greater reduction in the heart rate for music/vocal intervention with a frequency of at least 8 times (P = 0.0003). Investigating the related funnel plot in Figure 4 did not yield clear asymmetry, thus there was no clear indication of a risk of non-reporting bias.

Figure 4.



During intervention with heel lance

One study (Yu 2021) with 64 infants compared the heart rate during heel lance with the recorded maternal voice to standard care without discovering a significant difference.

Post-intervention

Nine studies (Calabro 2003; Caparros-Gonzalez 2018; Farhat 2010; Liao 2021; Loewy 2013; Namjoo 2021; Portugal 2017; Vastani 2017; Wirth 2016), provided data on heart rate post-intervention. Music and vocal interventions reduced heart rate post-intervention compared to standard care (MD -3.80, 95% CI -5.05 to -2.55; P = P < 0.00001; 903 infants, 9 studies; high-certainty evidence; Analysis 1.7). Wide confidence intervals ranged from a medium to large effect favouring the intervention (Analysis 1.7). Heterogeneity was moderate (Chi^2 = 18.50, df = 9 (P = 0.03); I^2 = 51%). Sensitivity analysis excluding studies with high selection or detection bias reduced heterogeneity (Chi^2 = 4.68, df = 4 (P = 0.32); I^2 = 15%) and the overall effect size (P < 0.0001), but still demonstrated a substantial beneficial effect.

Subgroup differences (Chi^2 = 5.20; df = 1; P = 0.02; I^2 = 80.8%) indicated a greater reduction in the heart rate post-intervention for recorded music and vocal intervention than for live music without substantial changes in the favouring of music and vocal interventions (Analysis 1.7). However, after sensitivity analysis, excluding studies with high selection or detection bias, we found no subgroup differences (Chi^2 = 0.09, df = 1 (P = 0.76), I^2 = 0%).

Post-intervention with heel lance

Two studies (Johnston 2007; Yu 2021) with 100 infants, evaluated heart rate post-intervention with heel lance showing no substantial difference in heart rate after the intervention compared to standard care after heel lance (MD 1.11, 95% CI -3.45 to 5.67; P = 0.63; 200 infants, 2 studies; Analysis 1.8). Heterogeneity was high (Chi^2 = 3.41, df = 1 (P = 0.06); I^2 = 71%) and, after removing the study with high attrition bias (Johnston 2007), the remaining study (Yu 2021) showed no favourable effect of the music intervention on heart rate compared to standard care (MD -2.6, 95% CI -7.72 to 3.60; P = 0.48; 64 infants, 1 study; Analysis 1.8).

Heart rate variability

One study (Epstein 2021), of 35 infants with severe brain injury (grade 3 or 4 intraventricular haemorrhage or periventricular leukomalacia), measured heart rate variability by low-frequency power (ms^2/Hz); high-frequency power (ms^2/Hz); and low frequency/high-frequency ratio, reflecting the balance between sympathetic and parasympathetic tone. A higher mean ± standard deviation (SD) LF/HF ratio (1.8 ± 0.7 vs. 1.1 ± 0.25, P = 0.01), were reported in maternal skin-to-skin care combined with maternal singing during music therapy as compared to maternal skin-to-skin-care alone.

Behavioural outcomes: behavioural state (Als)

Two studies (Epstein 2021; Tandoi 2015) provided data on behavioural outcomes measured with the behavioural numerical scales for neonates by Als 2005 (Table 5). We found no effect of music and vocal interventions compared to standard care (MD -0.12, 95% CI -0.52 to 0.27; P = 0.54; 69 infants, 2 studies; Analysis 1.11). Heterogeneity was very high (Chi^2 = 12.93, df = 1 (P = 0.0003); I^2 = 92%; Analysis 1.11) and, due to the small number of studies, subgroup analysis and sensitivity analysis were not indicated.

Other behavioural outcomes in the infants were reported by four small-scale studies (Calabro 2003; Kraft 2021; Nakhwa 2017; White-Traut 1988), measuring with different outcome scales which will be reported narratively: Calabro 2003 and colleagues, showed no significant changes in organised/disorganised infant behaviour between recorded sedative music and standard care group analysing 17 infants. Nakhwa 2017 provided data of 36 infants on the infant motor performance score showing a significant beneficial effect in favour of the music group (music group: MD 21.16 ± 0.5145; control group: MD 20.78 ± 0.4278; P < 0.0001). They reported significantly improved Infant Neurological International Battery (INFANIB) scores in favour of the music group (MD 65.11 ± 1.568) compared to the control group (MD 63.22 ± 2.756) with a P-value of 0.0163. Kraft 2021 provided data of 21 infants on the Neonatal Infant Stressor Scale showing no significant differences between the music and control groups. White-Traut 1988 showed no significant changes between the mother's live talking/singing group compared to standard care measured with the Nursing Child Assessment Feeding Scale (NCAFS) in 22 infants.

Hospitalisation

Three studies (Cevasco 2008; Ettenberger 2014; Menke 2021), provided data on oxygen hospitalisation in days (Table 5). The evidence suggested no effect of the intervention (MD -1.57, 95% CI -7.64 to 4.50; P = 0.61; 89 infants, 3 studies; Analysis 1.12) compared to standard care with wide CIs both favouring music and the control group. Heterogeneity was very high (Chi^2 = 9.78, df = 2 (P = 0.008); I^2 = 80%; Analysis 1.12) and all studies displayed high risk of bias.

Adverse effects

None of the included studies reported adverse effects from the music and vocal intervention including severe apnoea during the intervention requiring stimulation by the neonatal care team. However, only in two included studies (Lafferty 2021; Liao 2021) were adverse effects measured as explicit outcomes of interest.

Weight gain

Four studies (Cevasco 2008; Ettenberger 2014; Farhat 2010; Menke 2021), provided data on weight gain in grams per day (Table 5). We found no effect of the intervention (MD 3.88, 95% CI -1.61 to 9.38; P = 0.17; 137 infants, 4 studies; Analysis 1.13) compared to standard care with wide CIs both favouring the music group more than the control group. Heterogeneity was moderate (Chi^2 = 6.19, df = 3 (P = 0.10); I^2 = 52%; Analysis 1.13) and all studies displayed high risk of bias.

Long-term outcomes in preterm infants

No study provided data on neurodevelopment in preterm infants at five years of age assessed by standardised follow-up examinations.

Outcomes in parents

Parental well-being

Postnatal depression

Two music therapy studies (Kehl 2021; Menke 2021) provided data on parental depression measures with the Edinburgh Postnatal Depression Scale (EPDS) (Table 5). Music therapy may not reduce parental depression compared to standard care with the CIs ranging from no effect to a favourable effect (MD 0.50, 95% CI -1.80 to 2.81; P = 0.67; 67 participants; 2 studies; low-certainty evidence; Analysis 1.14). Heterogeneity between studies was moderate (Chi^2 = 2.05, df = 1 (P = 0.15); I^2 = 51%; Analysis 1.14).

State anxiety

Three music therapy studies (Kehl 2021; Kraft 2021; Menke 2021), provided data on parental state anxiety measured with the Parental State Anxiety Inventory (Table 5). The evidence is very uncertain about the effect of music therapy on parental state anxiety compared to standard care with the CIs ranging from no effect to a favourable effect (MD -0.15, 95% CI -2.72 to 2.41; P = 0.91; 87 parents, 3 studies; very low-certainty evidence; Analysis 1.15). Heterogeneity between studies was very low (Chi^2 = 1.11, df = 2 (P = 0.57); I^2 = 0%; Analysis 1.15).

Perception and stress outcomes

We identified three more small-scale studies (Cevasco 2008; Kehl 2021; Menke 2021), that reported measuring parental well-being after the whole intervention period with other outcomes (see Table 5) which we report narratively. Cevasco 2008 assessed maternal well-being in 16 mothers by measuring with the Parental Perception

Inventory (PPI) showing no significant differences between the music and control groups. Kehl 2021 provided data on parental stress measured by the Parental Stressor Scale: Neonatal Intensive Care Unit (PSS: NICU) in 32 parents identifying no significant differences between the intervention and the standard groups after the whole intervention period. Menke 2021 provided data from 30 parents on parental stress measured by the Parental Stress Questionnaire: Stress (PSQ) with no evidence of a difference between the intervention and control groups after the whole intervention period.

Attachment/bonding

We identified one music therapy study (Kehl 2021), with 32 parents providing data on the Pictorial Representation of Attachment Measurement (PRAM). There was no significant difference between music therapy and standard care after the intervention period. Another music therapy study (Ettenberger 2014) provided data on the Mother-Infant Bonding Scale with 16 mothers displaying no significant differences between the groups after the last intervention (Table 5).

DISCUSSION

Summary of main results

Our review evaluated the effectiveness of music and vocal interventions compared to standard care to improve neurodevelopmental outcomes in preterm infants and the well-being of parents. We included 25 RCTs, involving 1532 randomised infants and 691 parents. The music and vocal interventions varied widely in intervention type, delivery, frequency, and duration across studies mainly characterised by calm, soft, and musical parameters in lullaby style, often integrating the sang mother's voice live or recorded delivered by neonatal staff, parents, or music therapists. The included studies compared the intervention to standard care in extremely to late-preterm infants.

We found no substantial effects on our primary outcomes in the infants and parents. Music and vocal interventions do not increase oxygen saturation in preterm infants during the intervention compared to standard care and probably not post-intervention either. The evidence suggests that the intervention does not increase infant development. The certainty of the evidence was rated as 'low' due to imprecision since data on infant development were only available from two trials involving 69 infants in total meaning that our confidence in these results is limited (Summary of findings 1). Music therapy compared to standard care may not reduce our primary outcome of parental state-trait anxiety, measured with the State-Trait Anxiety Inventory. However, the certainty of the evidence was rated again as 'low' meaning that these results should be considered with caution.

A substantial effect in favour of music and vocal interventions was retained in one of our secondary outcomes. Music and vocal interventions likely result in a substantial small-to-medium reduction in heart rate in preterm infants compared to standard care during the intervention, with moderate certainty, meaning we are moderately confident in the effect estimate. Greater reduction of the heart rate was associated with the music/vocal intervention with at least eight sessions of music therapy. Post-intervention, this effect was even greater with a medium-to-large substantial effect in favour of the music and vocal interventions on heart rate, with high certainty, meaning we are highly certain that there is evidence of a beneficial effect of music/vocal interventions on heart rate in the post-intervention period (see Summary of findings 1). However, this beneficial effect on the heart rate is of questionable clinical significance.

Music and vocal interventions compared to standard care probably do not reduce respiratory rate. The evidence is uncertain of any effect on heart rate variability, behavioural outcomes, hospitalisation, weight gain, neurodevelopmental outcomes at five years, and further parental well-being and attachment outcomes. We found no certain evidence of any effect of musical and vocal interventions during and after painful procedures. For most of these outcomes, the certainty of the evidence was 'low' meaning that our confidence in these results is limited. We could not identify reported adverse effects of musical and vocal interventions on preterm infants and their parents. Study numbers were not sufficient to evaluate any possible differences between any of the intervention types, duration, frequency, or gestational age in the infants.

Overall completeness and applicability of evidence

Our review results suggest that the evidence base for the use of musical and vocal interventions in neonatal care is emerging, but still limited in number while displaying high heterogeneity in the intervention types, frequency, and duration. Physiological short-term outcomes in the infants comparing recorded music and vocal interventions with standard care inherent in a short time frame received the most attention. One reason for this may be that these outcomes and interventions are most feasible for research purposes since physiological data, such as oxygen saturation, respiratory rate, and heart rate are assessed in preterm infants during standard routine care anyway. Music therapy trials with individualised family-integrating live music are still limited in number whereas music medicine trials with recorded interventions received more attention. It is possible that a recorded intervention appears more feasible for research since only a recorded delivery can be double-blinded, e.g. by a sham recording in the control group. However, in the last decades, more and more music therapy services have been applied in clinical neonatal practice around the world (Bieleninik 2016; Haslbeck 2012; Loewy 2013; Standley 2012) and the first randomised controlled trials were conducted on live family-centred music therapy including parental outcomes (Haslbeck 2021; Kraft 2021; Loewy 2013; Menke 2021). Family-centred or family-integrated care programmes are highly endorsed by several organisations including the American Academy of Pediatrics (AAP) and the European Foundation for the Care of Newborn Infants (EFCNI) (Mushtaq 2019). Therefore, it would be important to evaluate further the possible effects of integrating parents as primary caregivers of their infant with the approach of family-centred music therapy on infant and parental short- and long-term outcomes in big-scale trials.

Notably, clinically relevant long-term developmental outcomes in preterm infants, such as the Bayley Scales of Infants and Toddler Development at two years of age were only assessed by two small-scale trials, and five years neurodevelopmental outcomes are completely missing, meaning that the current evidence is too limited to guarantee certainty of the evidence of musical and vocal interventions on clinically relevant outcomes in preterm infants. Moreover, the outcomes addressed in the included studies cover clinically pathological-oriented outcomes. However, resource-oriented outcomes that assess quality of life, resilience, empowerment, self-efficacy, and mental health may enhance the

relevance of the review in the future since particularly music therapy approaches are based on health- and resource-oriented approaches.

It was not possible to determine if certain types, frequencies, durations, and delivery modes of musical and vocal interventions have a differential impact on certain outcomes since the provided interventions varied widely in all aspects. Neither was it possible to distinguish between certain participant characteristics, such as birth weight categories or gender. Including clinically stable preterm infants without severe malformation or congenital diseases, intraventricular haemorrhages (IVH) (III-IV) and periventricular leukomalacia (PVL) received the most attention. However, there was one exceptional small-scale study (Epstein 2021) that included only preterm infants who developed IVH grades 3 or 4 or PVL diagnosed by brain ultrasound demonstrating unstable physiological responses during maternal singing with music therapy. Therefore, it would be important to evaluate further if music and vocal interventions may have a different effect on preterm infants with severe brain injury. Notably, most studies which included parents were only able to include mothers or fewer fathers than mothers. Thus, we were are not certain if the current evidence is applicable to fathers as well.

Our review results may reflect moderate generalisability since the review evidence comes from neonatal intensive care settings from around the world with cultural variations in treatment practices, musical choice, neonatal standard care, and parental attitudes towards the intervention. Additionally, an interdisciplinary variety was given by researchers from various disciplines (e.g. medicine, nursing, psychology, and music therapy) that conducted the studies. Moreover, the relevance of our findings will vary from healthcare systems which value music and music therapy to healthcare systems which decline or do not have the resources for complementary medicine approaches. However, parental lullaby singing is known in almost every culture and available to everyone so that health equity should play no substantial role in empowering parents to sing for their preterm infant.

Quality of the evidence

We found moderate-to-high-certainty evidence for physiological outcomes in the infants in this review. The certainty of the evidence was rated as 'low' to 'very low' for other outcomes in the infants and parents (see Summary of findings 1) which means that further research is likely to change the effect estimates and our confidence that they are precise. These outcome results should therefore be considered with caution. The methodological quality of included studies varied. As anticipated in our Cochrane protocol (Haslbeck 2019), most studies did not blind the intervention for participants and personnel due to the nature of the intervention itself, except for three studies with recorded music which provided a scam intervention for the control group (Caparros-Gonzalez 2018; Jabraeili 2016; Lejeune 2019). Selection bias was mostly low risk for adequate random sequence generation, but mainly unclear or low for allocation concealment (Figure 3). Many studies demonstrated a high risk of detection bias or insufficient details to determine if detection bias occurred. The overall likelihood of all other biases was low or unclear.

Our assessments of the certainty of the evidence mainly reflect concerns about the risk of bias and imprecision due to wide CIs and small sample sizes. Although we judged the studies to be at varying risks of bias overall, the evidence for our physiological outcomes is drawn from studies at low risk of bias demonstrating moderate-to-high certainty. We have demonstrated a high certainty of the evidence, particularly for the substantial effect in favour of music and vocal interventions on the infants' heart rates. One primary outcome, oxygen saturation, demonstrated a high certainty of the evidence while the two other primary outcomes of long-term infant development and change in parental anxiety reflected low-to-very low certainty of the evidence. We downgraded the quality of evidence for these two main outcomes, due mainly to imprecision (Summary of findings 1).

Potential biases in the review process

We performed an extensive search of databases and additional sources and applied no restrictions concerning nationality or language within the search process; and relied additionally on an existing international network of leading researchers in the field. Thus, we consider that the probability that we have missed an eligible trial is low. There was no evidence of publication bias either. Study investigators were contacted directly to request missing data and most authors replied and delivered the missing trial details.

We conducted study selection, data extraction, and risk of bias assessments in duplicate and independently, and we reached a consensus by discussing any discrepancies. Some published trial reports did not provide enough details to extract outcomes and adequately assess risks of bias. Although we contacted the authors of the trials to request missing data, we could not avoid some bias assessments in the review process due to incomplete reporting of trial details or results, or both. Furthermore, we found 38 ongoing studies, meaning that incorporating these studies in a future update may alter the conclusions of the current review.

Agreements and disagreements with other studies or reviews

The findings of our review are partially consistent with other meta-analyses of music interventions for preterm infants and their parents. Our identified substantial beneficial effect of music and vocal interventions on heart rate is consistent with the recent meta-analysis results of Yue 2021 which found that the heart rate in the music group was significantly higher than in the control group (MD = −3.21; 95% CI = −5.22 to −1.19; P = 0.002). Our findings are further in line with the meta-analysis results of Standley 2012 suggesting a beneficial effect on heart rate and the results of Mohan 2021 that provided an overview of systematic reviews demonstrating a beneficial heart rate effect as well. In contrast, our results are inconsistent with the review of Bieleninik 2016 which found no substantial effect of music therapy on the infants' heart rate, but a large effect favouring music therapy for infant respiratory rate similar to Yue 2021 which we cannot confirm by our meta-analysis.

Our results that music and vocal interventions do not increase oxygen saturation are consistent with the meta-analysis of Bieleninik 2016, Mohan 2021 and Yu 2021 but inconsistent with the review of Standley 2012 which suggests a significant difference in favour of music therapy. Interestingly Van Dokkum 2021 identified two distinct reactions in oxygenations to music therapy. These individually different reactions in the infants to music therapy may explain our identified lack of evidence of an effect on a group-average level and may need further exploration for possible clinical relevance.

Notably, Bieleninik 2016 and Yu 2021 reported a significant positive influence of music therapy on maternal anxiety in contrast to our findings. However, we included only anxiety outcomes that had been assessed after the whole intervention or control period in contrast to the included studies in these two reviews (Bieleninik 2016; Yu 2021) evaluating the immediate short-term effect on maternal anxiety directly after the music intervention. These contrary results at different time points are in line with the recently published multi-centre study by Gaden 2022, which found no significant differences after the music therapy intervention period at discharge between the music therapy and the control groups. Interestingly, the two studies (Kehl 2021; Menke 2021) which assessed maternal anxiety over the whole music therapy period in neonatal care at various time points showed a significant decrease in parental anxiety levels over the first weeks in neonatal care that was only apparent in the music therapy group and not in the control group. Similarly, Kraft 2021 suggested that early provision of family-centred music therapy may accelerate the reduction of maternal anxiety particularly in the first stressful weeks in the intensive care unit. Moreover, maternal anxiety levels were strongly related to infant stress, as indicated by Ettenberger 2014; and Kraft 2021 which claimed that paternal anxiety is heavily dependent on particular stress factors during the neonatal intensive care period and, thereby, should be considered as possible confounders in future studies.

Music and vocal interventions may result in little to no difference in infant development. To our knowledge, to date, only two small-scale studies (Haslbeck 2021; Lejeune 2019) assessed possible long-term effects at two years, and the available data may be insufficient in amount to demonstrate efficacy. Interestingly, both studies reported in further publications (Haslbeck 2020a; Lordier 2019) improved brain functional connectivity measured by resting-state functional magnet resonance imaging at term-equivalent age suggesting neurodevelopmental maturation effects. These pilot results indicate that further research is needed with big-scale studies correlating short-term brain imaging outcomes with neurodevelopmental long-term outcomes. Notably, a multi-centre study (Ghetti 2019) is ongoing assessing whether music therapy is superior to standard care by measuring Bayley Scale Scores at two years of age and one of the small-scale brain imaging studies (Haslbeck 2018) will examine neurodevelopment in preterm infants at five years to provide further insights of a possible influence of music therapy on long-term neurodevelopment in preterm infants.

Due to a further lack of evidence, we cannot yet determine a dose-, type- or frequency-response relationship for musical and vocal interventions, but our finding of a greater beneficial effect on heart rate with music concurs with the findings of Haslbeck 2020a, demonstrating a dose-dependent effect in favour of music therapy indicating further research in this direction.

Moreover, other reviews of various types of musical and vocal interventions for preterm infants (and their caregivers) (Anderson 2018; Hartling 2009; Haslbeck 2012; Hodges 2010; Mohan 2021; Standley 2012; Tramo 2011; Van der Heijden 2016; Yue 2021) show much higher heterogeneity amongst studies than in our meta-analysis. One reason may be that we restricted the inclusion of studies to a minimum duration of at least five minutes and a minimum frequency of at least three times. In addition, we divided time points into during and after the intervention and restricted parental well-being and attachment outcomes to a time point after the whole intervention period which was not reported in the other reviews (Mohan 2021).

AUTHORS' CONCLUSIONS

Implications for practice

Music and vocal interventions do not increase oxygen saturation during, and probably not after, the intervention compared to standard care. The evidence suggests that the intervention does not increase infant development (BSID) or parental state-trait anxiety. However, we have high confidence in our findings that music and vocal interventions results in a reduction in heart rate in preterm infants, particularly post-intervention. Although the reduction of four heartbeats per minute may not be clinically relevant, it may be meaningful, as the intervention may contribute to overall physiological stabilisation in preterm infants. It remains unknown if the intervention has any further beneficial effects on the infants and their parents. Further evidence suggests that the intervention has no adverse effects and is associated with more beneficial than harmful influences in general.

Based on our current review findings, it is not possible to express explicit implications for clinical practice yet, nor to implement music and vocal interventions in general, nor determine whether specific types, durations, frequencies, and modes of music or voice may be superior to others. However, the central role of the voice emerged, whether sung live, recorded, or with an accompanying instrument, as argued in the literature, to be the most attractive and meaningful auditory stimulus for a baby (Filippa 2017). It is strongly recommended that close observation of the immediate effect of any music and vocal interventions in neonatal care be undertaken in order to not harm, overstimulate or overwhelm this vulnerable group. Music and voice should never be played unobserved and with too high a duration, volume, and frequency. Moreover, identical vocal and music interventions may induce various effects depending upon the individual health condition, developmental stage, and age of the infant and upon the well-being and cultural background of the parents, so caution and precise observation, sensitivity, and overall responsiveness are indicated for best clinical decision-making with any kind of musical provision.

Implications for research

Further examination of music and vocal interventions for preterm infants and their parents with larger power, fewer risks of bias, and more sensitive and clinically relevant outcomes are needed. Research may be improved by: reducing risks of bias, e.g. allocation concealment and blinding of the outcomes assessors while decreasing heterogeneity within and between studies; evaluating the impact of type, length, and frequency of treatment; carefully considering possible confounders such as infant and parental stress on parental anxiety, and evaluating long-term clinically relevant outcomes.

Decreasing risks of bias and imprecision

It is valuable to increase precision by conducting studies which include power calculations and bigger sample sizes. Allocation concealment should be conducted, and detection bias should be avoided particularly when blinding participants and personnel is not possible due to the nature of the intervention itself. In general, future investigations need to pay more attention to reporting the

risk of bias in more detail accordingly to the Consort guidelines (Hopewell 2008).

Decreasing heterogeneity

Future research should take the high level of heterogeneity in the intervention into account and may entail comparisons between various types, duration, and frequencies while continuing to examine music and vocal interventions compared to other interventions or standard care. Since preterm infants differ in sensory maturation depending on their gestational age at birth and clinically relevant baseline characteristics, future trials might entail more precise reporting of these demographic data. Future trials may investigate possible differences in age, diagnoses, and gender in the infants as well as in culture, socioeconomics, and gender of the parents.

Considering possible confounders

In this review, music and voice were mostly delivered during routine care and possible other auditory influences were hardly reported. Parental well-being was assessed at different time points or without assessing the general mental health status of the parents which could have influenced parental well-being outcomes. Therefore, future trials should consider assessing possible confounders for outcomes such as additional stressors, diagnoses, challenges, and socio-emotional circumstances.

Increasing relevant outcomes measures

Future trials should consider investigating long-term outcomes in infants and their parents. Despite clinically relevant outcomes, future research should include outcomes that are relevant for the users themselves, and should consult with parents during trial development to guarantee that outcomes are pertinent and of value for the infants, parents, and families (for example, quality of life and sensitive long-term outcomes, such as executive functioning in the infants).

Parental advice should be obtained on study feasibility, intervention implementation, the parental informed consent procedure, and dissemination of results. In addition, neurological outcomes should be integrated to gain deeper insight into possible mechanisms of music and voice for the infants, the parents, and the bonding/attachment process. Brain imaging techniques, such as MRI, EEG, NIRS, and neuropsychological outcomes seem to be promising to deepen our understanding of music processing and the possible advantages of music and voice on socio-emotional development (Chorna 2019). Moreover, implementation science should play a central role in future investigations to increase the adoption and dissemination of research findings into evidence-based practice.

ACKNOWLEDGEMENTS

The methods section of this review is based on a standard template used by Cochrane Neonatal.

We would like to thank Cochrane Neonatal: Fiona Russell, Jane Cracknell and Michelle Fiander, Managing Editors; and Roger Soll and Bill McGuire, Co-coordinating Editors, who provided editorial and administrative support. We thank Anne Lethaby for copy-editing.

We thank Information Specialists, Carol Friesen and Michelle Fiander for developing and running searches in 2019 and 2021, respectively.

We thank Mohan Pammi, Professor, Section of Neonatology, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, USA, and Alice R. Rumbold, SAHMRI Women and Kids, South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia, for peer reviewing the manuscript.

REFERENCES

References to studies included in this review

Calabro 2003 *{published and unpublished data}*

Calabro J, Wolfe R, Shoemark H. The effects of recorded sedative music on the physiology and behaviour of premature infants with a respiratory disorder. *Australian Journal of Music Therapy* 2003;**14**:19.

Caparros-Gonzalez 2018 *{published and unpublished data}*

Caparros-Gonzalez RA, de la Torre-Luque A, Diaz-Piedra C, Vico FJ, Buela-Casal G. Listening to relaxing music improves physiological responses in premature infants: a randomized controlled trial. *Advances in Neonatal Care* 2018;**18**(1):58-69. [DOI: 10.1097/ANC.0000000000000448] [PMID: 29045255]

Cevasco 2008 *{published and unpublished data}*

Cevasco AM. The effects of mothers' singing on full-term and preterm infants and maternal emotional responses. *Journal of Music Therapy* 2008;**45**(3):273-306. [DOI: 10.1093/jmt/45.3.273] [PMID: 18959452]

Epstein 2021 *{published and unpublished data}*

Epstein S, Bauer S, Levkovitz Stern O, Litmanovitz I, Elefant C, et al. Preterm infants with severe brain injury demonstrate unstable physiological responses during maternal singing with music therapy: a randomised controlled study. *European Journal of Pediatrics* 2021;**180**(5):1403-12. [DOI: 10.1007/s00431-020-03890-3] [PMID: 33244709]

Ettenberger 2014 *{published and unpublished data}*

Ettenberger M, Odell-Miller H, Rojas Cardenas C, Torres Serrano S, Parker M, Camargo Llanos SM. Music therapy with premature infants and their caregivers in Colombia: a mixed methods pilot study including a randomized trial. *Voices: a world forum for music therapy* 2014;**14**(2):26.

Farhat 2010 *{published and unpublished data}*

Farhat A, Amiri R, Karbandi S, Esmaily H, Mohammadzadeh A. The effect of listening to lullaby music on physiologic response and weight gain of premature infants. *Journal of Neonatal-perinatal Medicine* 2010;**3**(2):103-7. [DOI: 10.3233/NPM-2010-0101]

Haslbeck 2021 *{published data only}*

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. *Frontiers in Pediatrics* 2021;**9**:660393. [DOI: 10.3389/fped.2021.660393] [PMID: 34222141]

Jabraeili 2016 *{published data only}*

Jabraeili M, Sabet T, MustafaGharebaghi M, Asghari Jafarabadi M, Arshadi M. The effect of recorded mum's lullaby and Brahms' lullaby on oxygen saturation in preterm infants: a randomised double-blind clinical trial. *Journal of Caring Sciences* 2016;**5**(1):85-93. [DOI: 10.15171/jcs.2016.009] [PMID: 26989669]

Johnston 2007 *{published and unpublished data}*

Johnston CC, Filion F, Nuyt AM. Recorded maternal voice for preterm neonates undergoing heel lance. *Advances in Neonatal Care* 2007;**7**(5):258-66. [DOI: 10.1097/01.ANC.0000296634.26669.13] [PMID: 18049153]

Kehl 2021 *{published and unpublished data}*

Kehl SM, La Marca-Ghaemmaghami P, Haller M, Pichler-Stachl E, Bucher HU, Bassler D, et al. Creative music therapy with premature infants and their parents: a mixed-method pilot study on parents' anxiety, stress and depressive symptoms and parent–infant attachment. *International Journal of Environmental Research and Public Health* 2021;**18**(1):1-20. [DOI: 10.3390/ijerph18010265] [PMID: 33396496]

Kraft 2021 *{published and unpublished data}*

ISRCTN94562698. Live music for preterm infants [Musical intervention for extremely preterm infants during NICU stay: a feasibility pilot study]. isrctn.com/ISRCTN94562698 (first received 29 January 2019). [CENTRAL: CN-01950138]

Kraft KE, Jaschke AC, Ravensbergen AG, Feenstra-Weeli A, van Goor ME, de Kroon ML, et al. Maternal anxiety, infant stress, and the role of live-performed music therapy during NICU stay in The Netherlands. *International Journal of Environmental Research and Public Health* 2021;**18**(13):02. [DOI: 10.3390/ijerph18137077] [PMID: 34281014]

Kucuk Alemdar 2020 *{published data only}*

Kucuk Alemdar D, Inal S. The effect of individualised developmental care practices in preterm infants. *Journal of complementary Medicine Research* 2020;**27**(2):97-104. [DOI: 10.1159/000504357] [PMID: 31751994]

Lafferty 2021 *{published data only}*

Lafferty MA, Mackley A, Green P, Ottenthal D, Locke R, Guillen U. Can Mozart improve weight gain and development of feeding skills in premature infants? A randomized trial. *American Journal of Perinatology* 2021;**22**:22. [DOI: 10.1055/s-0041-1731279] [PMID: 34157772]

Lejeune 2019 *{published data only}*

Lejeune F, Lordier L, Pittet MP, Schoenhals L, Grandjean, Huppi PS, et al. Effects of an early postnatal music intervention on cognitive and emotional development in preterm children at 12 and 24 months: preliminary findings. *Frontiers in Psychology* 2019;**10**(101550902):494. [PMID: 30890993]

Liao 2021 *{published data only}*

Liao J, Liu G, Xie N, Wang S, Wu T, Lin Y, et al. Mothers' voices and white noise on premature infants' physiological reactions in a neonatal intensive care unit: A multi-arm randomized controlled trial. *International Journal of Nursing Studies* 2021;**119**:103934. [DOI: 10.1016/j.ijnurstu.2021.103934] [PMID: 33975075]

Loewy 2013 *{published and unpublished data}*

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):902-18. [DOI: 10.1542/peds.2012-1367] [PMID: 23589814]

Menke 2021 *{published and unpublished data}*

Menke BM, Hass J, Diener C, Poschl J. Family-centered music therapy—empowering premature infants and their primary caregivers through music: results of a pilot study. *PLOS One* 2021;**16**(5):18. [DOI: 10.1371/journal.pone.0250071] [PMID: 33989286]

Nakhwa 2017 *{published data only}*

Nakhwa PK, Mandar M, Shrikhande DY, Shrikhande S, Rokade P. Efficacy of music therapy in improvement of neuromotor development in preterm infants. *Revista Romana de Kinetoterapie {Romanian Journal of Physical Therapy}* 2017;**23**(40):5-11.

Namjoo 2021 *{published data only}*

Namjoo R, Mehdipour-Rabori R, Bagherian B, Nematollahi M. Comparing the effectiveness of mother's live lullaby and recorded lullaby on physiological responses and sleep of preterm infants: a clinical trial study. *Journal of Complementary and Integrative Medicine* 2021;**24**:24. [DOI: 10.1515/jcim-2020-0507] [PMID: 34022122]

Portugal 2017 *{published and unpublished data}*

Portugal CM, Sá LO, Guimarães MH. Cardiorespiratory effects of maternal sounds in infants bom between 26 and 33 weeks of gestation. *Revista de Enfermagem Referencia* 2017;**4**(12):55-63. [DOI: 10.12707/RIV16075]

Tandoi 2015 *{published data only}*

Tandoi F, Francescato G, Pagani A, Buzzetti G, Negri E, Agosti M. "The Original Sound": a new non-pharmacological approach to the postnatal stress management of preterm infants. *Journal of Maternal-fetal & Neonatal Medicine* 2015;**28**(16):1934-8. [DOI: 10.3109/14767058.2014.973392] [PMID: 25283851]

Vastani 2017 *{published data only}*

Vastani MA, Zahedpasha Y, Amiri SR, Khafri S, Farhadi R. The effect of experience on recognition of mother's voice in preterm infants. *Nursing and Midwifery Studies* 2017;**6**(1):7. [DOI: 10.17795/nmsjournal40964]

White-Traut 1988 *{published data only}*

White-Traut RC, Nelson MN. Maternally administered tactile, auditory, visual, and vestibular stimulation: relationship to later interactions between mothers and premature infants. *Research in Nursing & Health* 1988;**11**(1):31-9. [DOI: 10.1002/nur.4770110106] [PMID: 3347763]

Wirth 2016 *{published and unpublished data}*

Wirth L, Dorn F, Wege M, Zemlin M, Lemmer B, Gorbey S, et al. Effects of standardized acoustic stimulation in premature infants: a randomized controlled trial. *Journal of Perinatology* 2016;**36**(6):486-92. [DOI: 10.1038/jp.2016.1] [PMID: 26890554]

Yu 2021 *{published data only}*

Yu WC, Chiang MC, Lin KC, Chang CC, Lin KH, Chen CW. Effects of maternal voice on pain and mother-Infant bonding in premature infants in Taiwan: a randomized controlled trial. *Journal of Pediatric Nursing* 2021;**30**:30. [DOI: 10.1016/j.pedn.2021.09.022] [PMID: 34602338]

References to studies excluded from this review

Alipour 2013 *{published data only}*

Alipour Z, Eskandari N, Ahmari Tehran H, Eshagh Hossaini SK, Sangi S. Effects of music on physiological and behavioral responses of premature infants: a randomized controlled trial. *Complementary Therapies in Clinical Practice* 2013;**19**(3):128-32. [DOI: 10.1016/j.ctcp.2013.02.007] [PMID: 23890458]

Amini 2013 *{published data only}*

Amini E, Rafiei P, Zarei K, Gohari M, Hamidi M. Effect of lullaby and classical music on physiologic stability of hospitalized preterm infants: a randomized trial. *Journal of Neonatal-Perinatal Medicine* 2013;**6**(4):295-301. [DOI: 10.3233/NPM-1371313] [PMID: 24441085]

Arnon 2006 *{published data only}*

Arnon S, Shapsa A, Forman L, Regev R, Bauer S, Litmanovitz I, et al. Live music is beneficial to preterm infants in the neonatal intensive care unit environment. *Birth (Berkeley, Calif.)* 2006;**33**(2):131-6. [DOI: 10.1111/j.0730-7659.2006.00090.x] [PMID: 16732778]

Arnon 2014 *{published data only}*

Arnon S, Diamant C, Bauer S, Regev R, Sirota G, Litmanovitz I. Maternal singing during kangaroo care led to autonomic stability in preterm infants and reduced maternal anxiety. *Acta Paediatrica* 2014;**103**(10):1039-44. [DOI: 10.1111/apa.12744] [PMID: 25039678]

Auto 2013 *{published data only}*

Auto FM, Amancio OM, de Córdoba Lanza F. The effect of music on weight gain of preterm infants older than 32 weeks: a randomized clinical trial. *Revista Paulista de Pediatria* 2013;**33**(4):e293-9. [DOI: 10.1590/0103-058231369512]

Badr 2017 *{published data only}*

Badr LK, Demerjian T, Daaboul T, Abbas H, Zeineddine MH, Charafeddine L. Preterm infants exhibited less pain during a heel stick when they were played the same music their mothers listened to during pregnancy. *Acta Paediatrica* 2017;**106**(3):438-45. [DOI: 10.1111/apa.13666] [PMID: 27883227]

Barandouzi 2020 *{published data only}*

Barandouzi ZA, Keshavarz M, Montazeri A, Ashayeri H, Rajaei Z. Comparison of the analgesic effect of oral sucrose and/or music in preterm neonates: a double-blind randomized clinical trial. *Complementary Therapies in Medicine* 2020;**48**:102271. [DOI: 10.1016/j.ctim.2019.102271] [PMID: 31987233]

Bellieni 2001 *{published data only}*

Bellieni CV, Buonocore G, Nenci A, Franci N, Cordelli DM, Bagnoli F. Sensorial saturation: an effective analgesic tool for heel-prick in preterm infants: a prospective randomized trial. *Biology of the Neonate* 2001;**80**(1):15-8. [DOI: 10.1159/000047113] [PMID: 11474143]

Bergomi 2014 *{published data only}*

Bergomi P, Chieppi M, Maini A, Mugnos T, Spotti D, Tzialla C, et al. Nonpharmacological techniques to reduce pain in preterm infants who receive heel-lance procedure: a randomized controlled trial. *Research and Theory for Nursing Practice* 2014;**28**(4):335-48. [DOI: 10.1891/1541-6577.28.4.335] [PMID: 25577862]

Castano 2021 *{published data only}*

Castano D, Alzate EA, Murillo AF, Vera C, Villada MG, Giraldo SJ, et al. Effect of musical exposure on cerebral oxygen saturation,neurobehavior and sleep-wake cycle in premature infants hospitalized in a high complex institution. Randomized, triple-blind and controlled trial. *Pediatrics* 2021;**147**(3):676-7. [DOI: 10.1542/peds.147.3-MeetingAbstract.676]

Cavaiuolo 2015 *{published data only}*

Cavaiuolo C, Casani A, Di Manso G, Orfeo L. Effect of Mozart music on heel prick pain in preterm infants: a pilot randomized controlled trial. *Journal of Pediatric and Neonatal individualized Medicine* 2015;**4**(1):5.

Cevasco 2005 *{published data only}*

Cevasco AM, Grant RE. Effects of the pacifier activated lullaby on weight gain of premature infants. *Journal of Music Therapy* 2005;**42**(2):123-39. [DOI: 10.1093/jmt/42.2.123] [PMID: 15913390]

Chirico 2017 *{published data only}*

Chirico G, Cabano R, Villa G, Bigogno A, Ardesi M, Dioni E. Randomised study showed that recorded maternal voices reduced pain in preterm infants undergoing heel lance procedures in a neonatal intensive care unit. *Acta Paediatrica* 2017;**106**(10):1564-8. [DOI: 10.1111/apa.13944] [PMID: 28580602]

Chorna 2018 *{published data only}*

Chorna OD, Hamm EL, Shrivastava H, Maitre NL. Feasibility of event-related potential (ERP) biomarker use to study effects of mother's voice exposure on speech sound differentiation of preterm infants. *Developmental Neuropsychology* 2018;**43**(2):123-34. [DOI: 10.1080/87565641.2018.1433671] [PMID: 29436854]

Corrigan 2020 *{published data only}*

Corrigan MJ, Keeler JR, Miller HD, Ben Khallouq BA, Fowler SB. Music therapy and retinopathy of prematurity screening: using recorded maternal singing and heartbeat for post exam recovery. *Journal of Perinatology* 2020;**40**(12):1780-8. [DOI: 10.1038/s41372-020-0719-9] [PMID: 32681063]

Corrigan 2021 *{published data only}*

Corrigan M, Keeler J, Miller H, Naylor C, Diaz A. Music therapy and family-integrated care in the NICU: using heartbeat-music interventions to promote mother-infant bonding. *Advances in Neonatal Care* 2021;**16**:16. [DOI: 10.1097/ANC.0000000000000910] [PMID: 34138791]

Dalili 2020 *{published data only}*

Dalili H, Zaker Z, Keihanidoust Z, Farahani Z, Shariat M. Comparison of neuro-developmental status in preterm neonates with and without family based interventions. *World Journal of Advanced Research and Reviews* 2020;**08**(02):56-63. [DOI: 10.30574/wjarr.2020.8.2.0357]

IRCT201705079568N17. Comparison of neuro-developmental status of preterm neonates with and without family based interventions [Influence of family based intervention on neuro-developmental status of preterm neonates]. en.irct.ir/trial/10142 (first received 02 October 2017). [CENTRAL: CN-01890699]

Dearn 2014 *{published data only}*

Dearn T, Shoemark H. The effect of maternal presence on premature infant response to recorded music. *Journal of Obstetric, Gynecologic, and Neonatal Nursing* 2014;**43**(3):341-50. [DOI: 10.1111/1552-6909.12303] [PMID: 24707819]

Detmer 2020 *{published data only}*

Detmer MR, Evans K, Shina E, Walker K, DeLoach D, Malowitz JR. Multimodal neurologic enhancement improves preterm infants' developmental outcomes: a longitudinal pilot study. *Neonatal Network* 2020;**39**(1):16-23. [DOI: 10.1891/0730-0832.39.1.16] [PMID: 31919289]

Doheny 2012 *{published data only}*

Doheny L, Hurwitz S, Insoft R, Ringer S, Lahav A. Exposure to biological maternal sounds improves cardiorespiratory regulation in extremely preterm infants. *Journal of Maternal-Fetal & Neonatal Medicine* 2012;**25**(9):1591-4. [DOI: 10.3109/14767058.2011.648237] [PMID: 22185623]

Döra 2021 *{published data only}*

Döra Ö, Büyük ET. Effect of white noise and lullabies on pain and vital signs in invasive interventions applied to premature babies. *Pain Management Nursing* 2021;**28**:28. [DOI: 10.1016/j.pmn.2021.05.005] [PMID: 34210600]

Edraki 2017 *{published data only}*

Edraki M, Beheshtipour N, Razavi SM, Zahadatpour Z. Studying the effect of lullaby on some physiological factors. Pharmacophore 2017;**8**(6S):e-1173552. [URL: pharmacophorejournal.com/article/studying-the-effect-of-lullaby-on-some-of-physiological-factors-due-to-endtracheal-tube-suction-of-preterm-infants-confined-to-bed-in-intensive-care-unit-icu]

Efendi 2018 *{published data only}*

Efendi D, Caswini N, Rustina Y, Iskandar RA. Combination of mother therapeutic touch (MTT) and maternal voice stimulus (MVS) therapies stabilise sleep and physiological function in preterm infants receiving minor invasive procedures. *Journal of Neonatal Nursing* 2018;**24**(6):318-24. [DOI: 10.1016/j.jnn.2018.08.001]

Eskandari 2012 *{published data only}*

Eskandari N, Keshavars M, Ashayeri H, Jahdi F, Hosseini AF. Quran recitation: short-term effects and related factors in preterm newborns. *Research Journal of Medical Sciences* 2012;**6**(3):148-53. [HTTPS: www.medwelljournals.com/abstract/?doi=rjmsci.2012.148.153]

Filippa 2021 *{published data only}*

Filippa M, Monaci MG, Spagnuolo C, Serravalle P, Daniele R, Grandjean D. Maternal speech decreases pain scores and increases oxytocin levels in preterm infants during painful procedures. *Scientific Reports* 2021;**11**(1):17301. [DOI: 10.1038/s41598-021-96840-4] [PMID: 34453088]

Garunkstiene 2014 *{published data only}*

Garunkstiene R, Buinauskiene J, Uloziene I, Markuniene E. Controlled trial of live versus recorded lullabies in preterm infants. *Nordic Journal of Music Therapy* 2014;**23**(1):71-88. [DOI: 10.1080/08098131.2013.809783]

Holsti 2019 *{published data only}*

Holsti L, MacLean K, Oberlander T, Synnes A, Brant R. Calmer: a robot for managing acute pain effectively in preterm infants in the neonatal intensive care unit. *Pain Reports* 2019;**4**(2):e727. [DOI: 10.1097/PR9.0000000000000727] [PMID: 31041426]

Jie 2017 *{published data only}*

Jie Q, Yun-fei J, Fang L, Qian-hong T, Hui R, Rui C, et al. Effect of combined music and touch intervention on pain response and β-endorphin and cortisol concentrations in late preterm infants. *BMC Pediatrics* 2017;**17**:1-7. [DOI: 10.1186/s12887-016-0755-y] [PMID: 28125980]

Johnston 2009 *{published data only}*

Johnston CC, Filion F, Campbell-Yeo M, Goulet C, Bell L, McNaughton K, et al. Enhanced kangaroo mother care for heel lance in preterm neonates: a crossover trial. *Journal of Perinatology* 2009;**29**(1):51-6. [DOI: 10.1038/jp.2008.113] [PMID: 18769382]

Kahraman 2020 *{published data only}*

Kahraman A, Gumus M, Akar M, Sipahi M, Bal Yilmaz H, Basbakkal Z. The effects of auditory interventions on pain and comfort in premature newborns in the neonatal intensive care unit; a randomised controlled trial. *Intensive and Critical Care Nursing* 2020;**61**:102904. [DOI: 10.1016/j.iccn.2020.102904] [PMID: 32653359]

Kanagasabai 2013 *{published data only}*

Kanagasabai PS, Mohan D, Lewis LE, Kamath A, Rao BK. Effect of multisensory stimulation on neuromotor development in preterm infants. *Indian Journal of Pediatrics* 2013;**80**(6):460-4. [DOI: 10.1007/s12098-012-0945-z] [PMID: 23288407]

Karadag 2021 *{published data only}*

Karadag OE, Kerimoglu Yildiz G, Akdogan R, Yildiz S, Hakyemez Toptan H. The effect of simulative heartbeat nest used in preterm new-borns on vital signs, pain, and comfort in Turkey: a randomized controlled study. *Journal of Pediatric Nursing* 2021;**23**:23. [DOI: 10.1016/j.pedn.2021.10.003] [PMID: 34702595]

Karimi 2020 *{published data only}*

Karimi A, Zabihi A, Zahed Pasha, Y, Jafarian Amiri SR, Hajian K, Moudi S. The effect of music on the anxiety of mothers of infants admitted to the neonatal ntensive Care Unit. *Journal Babol Uniersity of Medical Sciences* 2020;**22**:348–54. [DOI: 10.22088/jbums.22.1.348]

Keith 2009 *{published data only}*

Keith DR, Russell K, Weaver BS. The effects of music listening on inconsolable crying in premature infants. *Journal of Music Therapy* 2009;**46**(3):191-203. [DOI: 10.1093/jmt/46.3.191] [PMID: 19757875]

Kucuk Alemdar 2018 *{published data only}*

Kucuk Alemdar D, Guducu TufekcI F. Effects of maternal heart sounds on pain and comfort during aspiration in preterm infants. *Japan Journal of Nursing Science* 2018;**15**(4):330-9. [DOI: 10.1111/jjns.12202] [PMID: 29345047]

Kurdahi Badr 2017 *{published data only}*

Kurdahi Badr L, Demerjian T, Daaboul T, Abbas H, Hasan Zeineddine M, Charafeddine L. Preterm infants exhibited less pain during a heel stick when they were played the same music their mothers listened to during pregnancy. *Acta Paediatrica* 2017;**106**(3):438-45. [DOI: 10.1111/apa.13666] [PMID: 27883227]

Majidipour 2018 *{published data only}*

Majidipour N, Nirouzad F, Madmoli Y, Sarrafzade S, Kalani L, Aghababaeian H, et al. The effect of Holy Quran recitation on the physiological responses of premature infants during phlebotomy: a randomized clinical trial. *International Journal of Pediatrics-Masshad* 2018;**6**(7):7869-81. [DOI: 10.22038/ijp.2017.24203.2038]

Picciolini 2014 *{published data only}*

Picciolini O, Porro M, Meazza A, Gianni ML, Rivoli C, Lucco G, et al. Early exposure to maternal voice: effects on preterm infants development. *Early Human Development* 2014;**90**(6):287-92. [DOI: 10.1016/j.earlhumdev.2014.03.003] [PMID: 24661448]

Pouraboli 2019 *{published data only}*

Pouraboli B, Rayyani M, Anari Mahlegha D, Hosseini F, Loghmani L. Lullaby effect with mother's voice on respiratory rate and the speed of its return to the pre-suction state in intubated preterm infants, during tracheal tube suction Kerman, Afzali Pour hospital 2016. *Electronic Journal of General Medicine* 2019;**16**(1):1-9. [DOI: 10.29333/ejgm/93471]

Qolizadeh 2019 *{published data only}*

Qolizadeh A, Myaneh ZT, Rashvand F. Investigating the effect of listening to the Holy Quran on the physiological responses of neonates admitted to neonatal intensive care units: a pilot study. *Advances in Integrative Medicine* 2019;**6**(4):159-62. [DOI: 10.1016/j.aimed.2018.08.004]

Ramezani 2020 *{published data only}*

Ramezani Z, Pasha YZ, Jafarian Amiri SR, Hajiahmadi M, Pirnia B. Comparison of the effectiveness of warm compress and music on the pain caused by heel blood sampling in neonates: a single-blind, prospective, randomized controlled trial. *Crescent Journal of Medical and Biological Sciences* 2020;**7**(1):138-9. [PDF: www.cjmb.org/uploads/pdf/pdf_CJMB_359.pdf]

Ranger 2018 *{published data only}*

Ranger A, Helmert E, Bott TS, Ostermann T, Als H, Bassler D, et al. Physiological and emotional effects of pentatonic live music played for preterm neonates and their mothers in the

newborn intensive care unit: a randomized controlled trial. *Complementary Therapies in Medicine* 2018;**41**:240-6. [DOI: 10.1016/j.ctim.2018.07.009] [PMID: 30477847]

Sarhangi 2021 *{published data only}*

Sarhangi F, Azarmnejad E, Javadi M, Tadrisi, SD, Rejeh N, Vaismoradi M. The effect of the mother's heartbeat sound on physiological parameters and pain intensity after blood sampling in neonates in the intensive care unit: a randomized controlled clinical trial. *Journal of Neonatal Nursing* 2021;**27**(2):123-8. [DOI: 10.1016/j.jnn.2020.07.006]

Schlez 2011 *{published data only}*

Schlez A, Litmanovitz I, Bauer S, Dolfin T, Regev R, Arnon S. Combining kangaroo care and live harp music therapy in the neonatal intensive care unit setting. *Israel Medical Association Journal* 2011;**13**(6):354-8. [PMID: 21809733]

Shabani 2016 *{published data only}*

Shabani F, Nayeri ND, Karimi R, Zarei K, Chehrazi M. Effects of music therapy on pain responses induced by blood sampling in premature infants: a randomized cross-over trial. *Iranian Journal of Nursing and Midwifery Research* 2016;**21**(4):391-6. [DOI: 10.4103/1735-9066.185581] [PMID: 27563323]

Shah 2017 *{published data only}*

Shah SR, Kadage S, Sinn J. Trial of music, sucrose, and combination therapy for pain relief during heel prick procedures in neonates. *Journal of Pediatrics* 2017;**190**:153-8. [DOI: 10.1016/j.jpeds.2017.08.003] [PMID: 29144240]

Shukla 2018 *{published data only}*

Shukla VV, Bansal S, Nimbalkar A, Chapla A, Phatak A, Patel D, et al. Pain control interventions in preterm neonates: a randomized controlled trial. *Indian Pediatrics* 2018;**55**(4):292-6. [DOI: 10.1007/s13312-018-1270-z] [PMID: 29428919]

Standley 1995 *{published data only}*

Standley JM, Moore RS. Therapeutic effects of music and mother's voice on premature infants. *Pediatric Nursing* 1995;**21**(6):509-12, 574. [PMID: 8700604]

Standley 2000 *{published data only}*

Standley JM. The effect of contingent music to increase non-nutritive sucking of premature infants. *Pediatric Nursing* 2000;**26**(5):493-5, 498-9. [PMID: 12026338]

Standley 2003 *{published data only}*

Standley JM. The effect of music-reinforced nonnutritive sucking on feeding rate of premature infants. *Journal of Pediatric Nursing* 2003;**18**(3):169-73. [DOI: 10.1053/jpdn.2003.34] [PMID: 12796858]

Standley 2010 *{published data only}*

Standley JM, Cassidy J, Grant R, Cevasco A, Szuch C, Nguyen J, et al. The effect of music reinforcement for non-nutritive sucking on nipple feeding of premature infants. *Pediatric Nursing* 2010;**36**(3):138-45. [PMID: 20687305]

Stokes 2018 *{published data only}*

Stokes A, Agthe AG, El Metwally D. Music exposure and maturation of late preterm sleep-wake cycles: a randomised crossover trial. *Acta Paediatrica* 2018;**107**(4):582-6. [DOI: 10.1111/apa.14079] [PMID: 28929527]

Tang 2018 *{published data only}*

Tang L, Wang H, Liu Q, Wang F, Wang M, Sun J, et al. Effect of music intervention on pain responses in premature infants undergoing placement procedures of peripherally inserted central venous catheter: a randomized controlled trial. *European Journal of Integrative Medicine* 2018;**19**:105-9. [DOI: 10.1016/j.eujim.2018.03.006]

Tekgunduz 2019 *{published data only}*

Tekgunduz KS, Polat S, Gurol A, Apay SE. Oral glucose and listening to lullaby to decrease pain in preterm infants supported with NCPAP: a randomized controlled trial. *Pain Management Nursing* 2019;**20**(1):54-61. [DOI: 10.1016/j.pmn.2018.04.008] [PMID: 29776872]

Uematsu 2019 *{published data only}*

Uematsu H, Sobue I. Effect of music (Brahms lullaby) and non-nutritive sucking on heel lance in preterm infants: a randomized controlled crossover trial. Journal of Paediatrics and Child Health 2019;**24**(1):e33-9. [DOI: 10.1093/pch/pxy072] [PMID: 30792607]

Ullsten 2017 *{published data only}*

Ullsten A, Hugoson P, Forsberg M, Forzelius L, Klassbo M, Olsson E, et al. Efficacy of live lullaby singing during procedural pain in preterm and term neonates. *Music and Medicine* 2017;**9**(2):73-85. [DOI: 10.11124/jbisrir-2012-428] [PMID: 27820525]

Vahdati 2017 *{published data only}*

Vahdati M, Mohammadizadeh M, Talakoub S. Effect of kangaroo care combined with music on the mother-premature neonate attachment: a randomized controlled trial. *Iranian Journal of Nursing and Midwifery Research* 2017;**22**(5):403-7. [DOI: 10.4103/ijnmr.IJNMR_50_16] [PMID: 29033998]

Walworth 2012 *{published data only}*

Walworth D, Standley JM, Robertson A, Smith A, Swedberg O, Peyton JJ. Effects of neurodevelopmental stimulation on premature infants in neonatal intensive care: randomized controlled trial. *Journal of Neonatal Nursing* 2012;**18**(6):210-6. [DOI: 10.1016/j.jnn.2012.01.001]

White-Traut 2004 *{published data only}*

White-Traut RC, Nelson MN, Silvestri JM, Patel M, Berbaum M, Gu GG, et al. Developmental patterns of physiological response to a multisensory intervention in extremely premature and high-risk infants. *Journal of Obstetric, Gynecologic & Neonatal Nursing* 2004;**33**(2):266-75. [DOI: 10.1177/0884217504263289] [PMID: 15095806]

Wu 2020 *{published data only}*

Wu HP, Yang L, Lan HY, Peng HF, Chang YC, Jeng MJ, et al. Effects of combined use of mother's breast milk, heartbeat sounds, and non-nutritive sucking on preterm infants' behavioral stress

during venipuncture: a randomized controlled trial. *Journal of Nursing Scholarship* 2020;**52**(5):467-75. [DOI: 10.1111/jnu.12571] [PMID: 32564489]

Yildiz 2011 *{published data only}*

Yildiz A, Arikan D. The effects of giving pacifiers to premature infants and making them listen to lullabies on their transition period for total oral feeding and sucking success. *Journal of Clinical Nursing* 2012;**21**(5-6):644-56. [DOI: 10.1111/j.1365-2702.2010.03634.x] [PMID: 21668549]

Yilmaz 2021 *{published data only}*

Yilmaz G, Kucuk Alemdar D. The effect of supportive nursing interventions on reducing stress levels of mothers of infants in the neonatal intensive care unit: a randomized controlled trial. *Clinical Nursing Research* 2021;**31**(5):941-51. [DOI: 10.1177/10547738211047359] [PMID: 34555952]

Zeraati 2018 *{published data only}*

Zeraati H, Nasimi F, Rezaeian A, Shahinfar J, Ghorban Zade M. Effect of multi-sensory stimulation on neuromuscular development of premature infants: a randomized clinical trial. *Iranian Journal of Child Neurology* 2018;**12**(3):32-9. [DOI: 10.22037/ijcn.v12i3.16134] [PMID: 30026767]

References to ongoing studies

Aita 2021 *{published data only}*

Aita M, Heon M, Lavallee A, De Clifford Faugere G, Altit G, Le May S, et al. Nurturing and quiet intervention (NeuroN-QI) on preterm infants' neurodevelopment and maternal stress and anxiety: a pilot randomized clinical trial protocol. *Journal of Advanced Nursing* 2021;**77**(7):3192-203. [DOI: 10.1111/jan.14819] [TRIAL REGISTRATION: NCT04593095]

Bonjorn Juarez 2020 *{published data only}*

Bonjorn Juarez M, Manrique Pons M, Grau Alcon L, Martinez-Momblan MA, Alonso-Fernandez S. Reduction of visual and auditory stimuli to reduce pain during venipuncture in premature infants. Study protocol for a randomized controlled trial. *Journal of Advanced Nursing* 2020;**76**(4):1077-81. [DOI: 10.1111/jan.14300] [TRIAL REGISTRATION: NCT04041635]

Brignoni-Perez 2021 *{published data only}*

Brignoni-Perez E, Morales MC, Marchman VA, Scala M, Feldman HM, Yeom K, et al. Listening to Mom in the NICU: effects of increased maternal speech exposure on language outcomes and white matter development in infants born very preterm. *Trials* 2021;**22**(1):444. [DOI: 10.1186/s13063-021-05385-4] [TRIAL REGISTRATION: NCT04193579]

ChiCTR1900026017 *{published data only}*

ChiCTR1900026017. A randomized controlled trial for investigation of the associations between a music program and neonatal health outcomes. trialsearch.who.int/Trial2.aspx?TrialID=ChiCTR1900026017 (first received 18 September 2019). [CENTRAL: CN-02065137]

CTRI/2012/11/003117 *{published data only}*

* CTRI/2012/11/003117. A clinical trial to study the effectiveness of instrumental Indian classical music on serum cortisol concentrations of preterm infants on assisted ventilation admitted to a tertiary level neonatal intensive care unit (NICU) [Effectiveness of Instrumental Indian Classical music on serum cortisol concentrations of preterm infants on assisted ventilation admitted to a tertiary level neonatal intensive care unit (NICU)- a prospective, randomized controlled trial]. trialsearch.who.int/Trial2.aspx?TrialID=CTRI/2012/11/003117 (first received 16 November 2012). [CENTRAL: CN-01854686]

CTRI/2016/06/007028 *{published data only}*

CTRI/2016/06/007028. Comparison of music therapy with kangaroo mother care for pain reduction in premature babies [Factorial design randomised control trial to compare music therapy with kangaroo mother care for pain in premature neonates]. trialsearch.who.int/Trial2.aspx?TrialID=CTRI/2016/06/007028 (first received 22 June 1016). [CENTRAL: CN-01818320]

CTRI/2017/04/008395 *{published data only}*

CTRI/2017/04/008395. Effect of music therapy on breast milk production in mothers practicing kangaroo mother care [Effect of music therapy on breast milk production in KMC mothers]. trialsearch.who.int/Trial2.aspx?TrialID=CTRI/2017/04/008395 (first received 21 April 2017). [CENTRAL: CN-01894530]

CTRI/2021/06/033970 *{published data only}*

CTRI/2021/06/033970. A study to check whether music combined with standard kangaroo mother care help in improving neurodevelopmental outcome of preterm infants [Does music combined with kangaroo mother care improve the neurodevelopmental outcome of preterm infants compared to standard kangaroo mother care?- A randomised controlled trial]. trialsearch.who.int/Trial2.aspx?TrialID=CTRI/2021/06/033970 (first received 03 June 2021). [CENTRAL: CN-02327617]

CTRI/2021/07/035287 *{published data only}*

CTRI/2021/07/035287. Vestibular stimulation for babies born before 34 weeks of gestational age [Effects of vestibular stimulation on neurodevelopment in preterm infants]. trialsearch.who.int/Trial2.aspx?TrialID=CTRI/2021/07/035287 (first received 30 July 2021). [CENTRAL: CN-02328316]

D'Souza 2017 *{published data only}*

D'Souza SR, Lewis LE, Kumar V, Ramesh BY, Purkayastha J, Kamath S. Effectiveness of Indian classical music on developmental responses of preterm infants - a randomized controlled trial protocol. *Medico Legal Update* 2017;**17**(2):243-7. [PDF: https://medicolegalupdate.org/scripts/MLU%20July-Decr%202017.pdf]

DRKS00018806 *{published data only}*

DRKS00018806. Music and maternal voice: investigating the impact on premature babies outcomes [Music and maternal voice: investigating the impact on premature babies outcomes - MAMBO study]. trialsearch.who.int/Trial2.aspx?TrialID=DRKS00018806 (first received 22 October 2019). [CENTRAL: CN-02067360]

DRKS00025753 *{published data only}*

DRKS00025753. Does music therapy influence the neurological development of premature infants born less than 32 weeks of gestation? trialsearch.who.int/Trial2.aspx?TrialID=DRKS00025753 (first received 05 July 2021). [CENTRAL: CN-02328823]

Filippa 2021a *{published data only}*

Filippa M, Della Casa E, D'Amico R, Picciolini O, Lunardi C, Sansavini A, et al. Effects of early vocal contact in the neonatal intensive care unit: study protocol for a multi-centre, randomised clinical trial. *International Journal of Environmental Research and Public Health* 2021;**18**(8):08. [DOI: 10.3390/ijerph18083915] [TRIAL REGISTRATION: NCT04759573]

Ghetti 2019 *{published data only}*

Ghetti C, Bieleninik L, Hysing M, Kvestad I, Assmus J, Romeo R, et al. Longitudinal study of music therapy's effectiveness for premature infants and their caregivers (LongSTEP): protocol for an international randomised trial CNO. *BMJ Open* 2019;**9**(8):e025062. [DOI: 10.1136/bmjopen-2018-025062] [TRIAL REGISTRATION: NCT03564184]

IRCT2014101819566N2 *{published data only}*

IRCT2014101819566N2. Comparison of the effect of Koran (yasin) and lullaby music listening on physiological indicators of premature neonates admitted in Motahari hospital of Jahrom. trialsearch.who.int/Trial2.aspx?TrialID=IRCT2014101819566N2 (first received 8 November 2014). [CENTRAL: CN-01837693]

IRCT20170620034653N3 *{published data only}*

IRCT20170620034653N3. The effect of the Quran's voice on the formation of oral nutrition in premature infants admitted to the intensive care unit [The effect Voice of the Holy Quran on the duration of transition from gavage to first oral feeding and vital signe at premature newborns hospitalized in neonatal intensive care unit]. trialsearch.who.int/Trial2.aspx?TrialID=IRCT20170620034653N3 (first received 19 June 2018). [CENTRAL: CN-01908773]

IRCT20180526039844N1 *{published data only}*

IRCT20180526039844N1. Effect of voice and tune in pain of neonates [Comparison of the effectiveness of relaxing voice and tune voices using voice pillow on pain intensity of heel stick in hospitalized neonates]. trialsearch.who.int/Trial2.aspx?TrialID=IRCT20180526039844N1 (first received 02 July 2018). [CENTRAL: CN-01905961]

NCT02471482 *{published data only}*

NCT02471482. Effect of Mozart music on cerebral oxygenation and behavioural response of premature infants [Effect of Mozart music on cerebral oxygenation and behavioural response of premature infants between 28-32 weeks of corrected gestation]. clinicaltrials.gov/ct2/show/NCT02471482 (first received 07 November 2014). [CENTRAL: CN-01552873]

NCT02948491 *{published data only}*

NCT02948491. Effectiveness of music therapy on the development of preterm neonates (EMT-DRNP) [Effectiveness of music therapy on the intra-hospital evolution and neurodevelopment of preterm neonates treated in neonatal intensive care units in the Bucaramanga metropolitan area]. clinicaltrials.gov/ct2/show/NCT02948491 (first received 28 October 2016). [CENTRAL: CN-01588910]

NCT03688386 *{published data only}*

NCT03688386. A language intervention study of preterm infants [Bridging connections: a language intervention study of preterm infants in the NICU]. clinicaltrials.gov/ct2/show/NCT03688386 (first received 18 June 2018). [CENTRAL: CN-01663692]

NCT03795454 *{published data only}*

NCT03795454. Can singing kangaroo improve outcome of preterm Infants [Singing kangaroos - infant care With combined auditory intervention and kangaroo care]. clinicaltrialsgov/show/nct03795454 (first received 30 November 2019). [CENTRAL: CN-01795746]

NCT03830580 *{published data only}*

NCT03830580. Benefit of singing in the care of premature children undergoing screening for retinopathy of prematurity in the neonatology and neonatal resuscitation unit of the Dijon university hospital. clinicaltrialsgov/show/nct03830580 (first received 25 January 2019). [CENTRAL: CN-01796265]

NCT04314440 *{published data only}*

NCT04314440. Cognitive processing in preterm infants and NICU music therapy. clinicaltrials.gov/show/NCT04314440 (first received 7 March 2016). [CENTRAL: CN-02089202]

NCT04335240 *{published data only}*

NCT04335240. Premature family music therapy intervention (PFMI) [Premature family music therapy intervention (PFMI): an Italian protocol to support parenting, attachment bond and preterm development]. clinicaltrials.gov/show/NCT04335240 (first received 01 April 2020). [CENTRAL: CN-02091393]

NCT04559620 *{published data only}*

NCT04559620. Mother's recorded voice for preterm infants. https://clinicaltrials.gov/show/NCT04559620 (first received 30 July 2020).

NCT04565210 *{published data only}*

NCT04565210. Effects of oriental music on preterm infants [Effects of oriental music on preterm Infants: a randomized controlled trial. (OMNIA Trial)]. clinicaltrials.gov/show/NCT04565210 (first received 07 September 2020). [CENTRAL: CN-02181470]

NCT04759170 *{published data only}*

NCT04759170. The effects of recorded receptive music therapy on oral nutrition and the well-being of the italian premature baby [The effects of recorded receptive music therapy on oral nutrition and the well-being of the Italian premature baby: prospective randomized controlled study mom, dad and music therapist singing]. clinicaltrials.gov/show/NCT04759170 (first received 28 January 2021). [CENTRAL: CN-02249068]

NCT04759573 *{published data only}*

NCT04759573. Effects of early vocal contact (EVC) in the neonatal intensive care unit [Effects of early vocal contact (EVC)

in the neonatal intensive care unit: a multi-centre, randomized clinical trial]. clinicaltrials.gov/show/NCT04759573 (first received 06 February 2021). [CENTRAL: CN-02236469]

NCT04886648 *{published data only}*

NCT04886648. The effect of mother's voice and lullaby on preterm infants' physiological parameters, stress and sleeping-waking state. clinicaltrials.gov/show/NCT04886648 (first received 05 April 2021). [CENTRAL: CN-02257455]

Neel 2019 *{published data only}*

Neel ML, Yoder P, Matusz PJ, Murray MM, Miller A, Burkhardt S, et al. Randomized controlled trial protocol to improve multisensory neural processing, language and motor outcomes in preterm infants. *BMC Pediatrics* 2019;**19**(1):81. [CENTRAL: NCT03232931] [DOI: 10.1186/s12887-019-1455-1]

Pavel 2019 *{published data only}*

Pavel AM, Dhais FA, Howard C, O'Toole JM, Pavlidis E, Finn D, et al. The effect of music therapy on the electroencephalogram (EEG) and heart rate variability (HRV) of premature infants during routine painful procedures. *Archives of Disease in Childhood* 2019;**104**:A135. [DOI: 10.1136/archdischild-2019-epa.311]

Purdy 2011 *{published data only}*

Purdy IB, Anderson MS, Brown W, Kehoe P, Melwak MA, Garg M. A randomized controlled trial of NICU music: impact on preterm infant rest and growth. *Communicating Nursing Research* 2011;**44**:445.

Tosun 2014 *{published data only}*

Tosun O, Erdem E, Elmali F, Kurtoglu S. The effect of aromatherapy, music therapy and vibration applications on neonatal stress and behaviours. *Archives of Disease in Childhood* 2014;**99**:A82-3. [DOI: 10.1136/archdischild-2014-307384.220]

Additional references

Abbott 2002

Abbott A. Music, maestro, please! *Nature* 2002;**416**(6876):12-4. [PMID: 11882864]

Aita 2021

Aita M, De Clifford FG, Lavallée A, Feeley N, Stremler R, Rioux É, et al. Effectiveness of interventions on early neurodevelopment of preterm infants: a systematic review and meta-analysis. *BMC Pediatrics* 2021;**21**(1):210. [DOI: 10.1186/s12887-021-02559-6] [PMID: 33926417]

Als 1986

Als H, Lawhon G Brown E, Gibes R, Duffy FH, McAnulty G, et al. Individualized behavioral and environmental care for the very low birth weight preterm infant at high risk for bronchopulmonary dysplasia: neonatal intensive care unit and developmental outcome. *Pediatrics* 1985;**78**(6):1123-32. [PMID: https://pubmed.ncbi.nlm.nih.gov/3786036/]

Als 2005

Als H, Butler S, Kosta S, McAnulty G. The assessment of preterm infants' behavior (APIB): furthering the understanding and measurement of neurodevelopmental competence in preterm and full-term infants. *JAMA* 2005;**11**(1):94-102. [PMID: 15856436]

Anderson 2018

Anderson DE, Patel AD. Infants born preterm, stress, and neurodevelopment in the neonatal intensive care unit: might music have an impact? *Developmental Medicine & Child Neurology* 2018;**60**(3):256-66. [DOI: 10.1111/dmcn.13663] [PMID: 29363098]

Bieleninik 2016

Bieleninik L, Ghetti C, Gold C. Music therapy for preterm infants and their parents: a meta-analysis. *Pediatrics* 2016;**138**(3):e20160971. [PMID: 27561729]

Bradt 2015

Bradt J, Potvin N, Kesslick A, Shim M, Radl D, Schriver E, et al. The impact of music therapy versus music medicine on psychological outcomes and pain in cancer patients: a mixed methods study. *Support Care Cancer* 2015;**23**(5):1261-71. [DOI: 10.1007/s00520-014-2478-7] [PMID: 25322972]

Castrillón 2005

Castrillón Moreno, DA, Borrero Copete, PE. Validación del inventario de ansiedad estado-rasgo (STAIC) en niñosescolarizados entre los 8 y 15 años. *Acta Colombiana de Psicología* 2005;**13**:79-90. [HTTPS://: actacolombianapsicologia.ucatolica.edu.co/article/view/433]

Chawanpaiboon 2019

Chawanpaiboon S, Vogel JP, Moller AB, Lumbiganon P, Petzold M, Hogan D, et al. Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *Lancet Global Health* 2019;**7**(1):e37-e46. [DOI: 10.1016/S2214-109X(18)30451-0] [PMID: 30389451]

Chorna 2019

Chorna O, Filippa M, De Almeida JS, Lordier L, Monaci MG, Hüppi P, et al. Review article neuroprocessing mechanisms of music during fetal and neonatal development: a role in neuroplasticity and neurodevelopment. *Neural Plasticity* 2019;**2019**(3972918):0.1155/2019/3972918. [DOI: 10.1155/2019/3972918] [PMID: 31015828]

Clark-Gambelunghe 2015

Clark-Gambelunghe MB, Clark DA. Sensory development. *Pediatric Clinics of North America* 2015;**62**(2):367-84. [PMID: 25836703]

Covidence [Computer program]

Covidence. Version accessed before November 2021. Melbourne, Australia: Veritas Health Innovation, assessed before November 2021. Available at covidence.org.

Deeks 2017

Deeks JJ, Higgins JP, Altman DG, editor(s). Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JP, Churchill R, Chandler J, Cumpston MS, editor(s). Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0 (updated June 2017). The Cochrane Collaboration, 2017. training.cochrane.org/handbook.

Dileo 1999

Dileo C. A classification model for music and medicine. In: Applications of Music in Medicine. National Association of Music Therapy, 1999:1-6. [ISBN: 1884914004]

Egger 1997

Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**(7109):629-34. [DOI: 10.1136/bmj.315.7109.629] [PMID: 9310563]

Filippa 2017

Filippa M, Kuhn P, Westrup B. Early Vocal Contact and Preterm Infant Brain Development. Bridging the Gaps Between Research and Practice. Cham, Switzerland: Springer International Publishing, 2017. [DOI: 10.1007/978-3-319-65077-7]

Flacking 2007

Flacking R, Ewald U, Starrin B. "I wanted to do a good job": experiences of 'becoming a mother' and breastfeeding in mothers of very preterm infants after discharge from a neonatal unit. *Social Science & Medicine* 2007;**64**:2405-16. [PMID: 17428597]

Forcada-Guex 2006

Forcada-Guex M, Pierrehumbert B, Borghini A, Moessinger A, Muller-Nix C. Early dyadic patterns of mother-infant interactions and outcomes of prematurity at 18 months. *Pediatrics* 2006;**118**(1):107-14. [PMID: 16818525]

Gaden 2022

Gaden TS, Ghetti C, Kvestad I, Bieleninik Ł, Stordal AS, Assmus J, 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] [PMID: 34988583]

Ghetti 2019

Ghetti C, Bieleninik Ł, Hysing M, Kvestad I, Assmus J, Romeo R, 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] [PMID: 31481362]

GRADEproGDT [Computer program]

GRADEproGDT. Version accessed 9 June 2019. Hamilton (ON): McMaster University (developed by Evidence Prime), 2015.

Hanson-Abromeit 2008

Hanson-Abromeit D, Shoemark H, Loewy J. Music therapy with pediatric units: newborn intensive care unit (NICU). In: Hanson-Abromeit D, Colwell C, editors(s). Medical Music Therapy for Pediatrics in Hospital Settings. Using Music to Support Medical Interventions. Silver Spring (MD): American Music Therapy Association, 2008:15-69.

Hartling 2009

Hartling L, Shaik MS, Tjosvold L, Leicht R, Liang Y, Kumar M. Music for medical indications in the neonatal period: a systematic review of randomised controlled trials. *Archives of Disease in Childhood. Fetal and Neonatal Edition.* 2009;**94**:F349-54. [DOI: 10.1136/adc.2008.148411] [PMID: 19477913]

Haslbeck 2012

Haslbeck FB. Music therapy for premature infants and their parents: an integrative review. *Nordic Journal of Music Therapy* 2012;**21**(3):203-26. [DOI: 10.1080/08098131.2011.648653] [PMID: 27763631]

Haslbeck 2018

Haslbeck FB, Stegemann T. The effect of music therapy on infants born preterm. *Developmental Medicine & Child Neurology* 2018;**60**(3):217. [DOI: 10.1111/dmcn.13677] [PMID: 29369335]

Haslbeck 2020

Haslbeck FB, Bassler D. Clinical practice protocol of creative music therapy for preterm infants and their parents in the neonatal intensive care unit. Journal of Visualized Experiments 2020;**7**(155). [DOI: 10.3791/60412] [PMID: 31984968]

Haslbeck 2020a

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: Clinical* 2020;**25**(January):102171. [PMID: 6974781]

Heim 1999

Heim C, Nemeroff CB. The impact of early adverse experiences on brain systems involved in the pathophysiology of anxiety and affective disorders. *Biological Psychiatry* 1999;**46**(11):1509-22. [DOI: 10.1016/S0006-3223(99)00224-3] [PMID: 10599479]

Hepper 1994

Hepper PG, Shahidullah BS. Development of fetal hearing. *Archives of Disease in Childhood* 1994;**71**(2):81-7. [PMID: 1061088]

Higgins 2017

Higgins JP, Altman DG, Sterne JA, editor(s). Chapter 8: Assessing risk of bias in included studies. In: Higgins JP, Churchill R, Chandler J, Cumpston MS, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0 (updated June 2017). Cochrane, 2017. Available from training.cochrane.org/handbook/archive/v5.2.

Higgins 2019

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). The Cochrane Collaboration, 2019. (available at www:training.cochrane.org/handbook).

Hodges 2010

Hodges AL, Wilson LL. Effects of music therapy on preterm infants in the neonatal intensive care unit. *Journal of Alternative Therapies in Health and Medicine* 2010;**16**(5):72-3. [PMID: 20882735]

Hoffenkamp 2015

Hoffenkamp H, Tooten A, Hall R, Braeken J, Eliens M, Vingerhoets AD, et al. Effectiveness of hospital-based video interaction guidance on parental interactive behavior, bonding, and stress after preterm birth: a randomized controlled trial. *Journal of Consulting and Clinical Psycholgy* 2015;**83**(2):416-29. [DOI: 10.1037/a0038401] [PMID: 25486375]

Hopewell 2008

Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al, CONSORT Group. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 2008;**371**(9609):281-3. [PMID: 18221781]

Huotilainen 2010

Huotilainen M, Näätänen R. Auditory perception and early brain development. www.child-encyclopedia.com/brain/according-experts/auditory-perception-and-early-brain-development (accessed prior to 27 June 2019).

Inder 2005

Inder TE, Warfield SK, Wang H, Huppi PS, Volpe JJ. Abnormal cerebral structure is present at term in premature infants. *Pediatrics* 2005;**115**(2):286-94. [DOI: 10.1542/peds.2004-0326] [PMID: 15687434]

Jobe 2014

Jobe AH. Sensory deprivation in private rooms in the NICU. *Journal of Pediatrics* 2014;**164**(1):1-3. [DOI: 10.1016/j.jpeds.2013.10.049]

Johnson 2008

Johnson S, Wolke D, Marlow N. Developmental assessment of preterm infants at 2 years: validity of parent reports. *Developmental Medicine and Child Neurology* 2008;**50**(1):58-62. [DOI: 10.1111/j.1469-8749.2007.02010.x] [PMID: 18173632]

Jotzo 2005

Jotzo M, Poets CF. Helping parents cope with the trauma of premature birth: an evaluation of a trauma-preventive psychological intervention. *Pediatrics* 2005;**115**(4):915-9. [DOI: 10.1542/peds.2004-0370] [PMID: 15805364]

Koelsch 2014

Koelsch S. Brain correlates of music-evoked emotions. *Nature Reviews. Neuroscience* 2014;**15**(3):170-80. [DOI: 10.1038/nrn3666] [PMID: 24552785]

Korja 2011

Korja R, Latva R, Lehtonen L. The effects of preterm birth on mother–infant interaction and attachment during the infant's first two years. *Acta Obstetricia et Gynecologica Scandinavica* 2011;**91**:164-73. [DOI: 10.1111/j.1600-0412.2011.01304.x] [PMID: 22007730]

Krueger 2010

Krueger C. Exposure to maternal voice in preterm infants: a review. *Journal of Advances in Neonatal Care* 2010;**10**(1):13-8. [DOI: 10.1097/ANC.0b013e3181cc3c69] [PMID: 20150775]

Kuhn 2013

Kuhn P, Zores C, Langlet C, Escande B, Astruc D, Dufour A. Moderate acoustic changes can disrupt the sleep of very preterm infants in their incubators. *Acta Paediatrica* 2013;**102**(10):949-54. [DOI: 10.1111/apa.12330] [PMID: 23800026]

Lahav 2014

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. *Frontiers in Neuroscience* 2014;**8**(DEC):1-8. [DOI: 10.3389/fnins.2014.00381] [PMID: 25538543]

Lakshmanan 2021

Lakshmanan A, Song AY, Belfort MB, Yieh L, Dukhovny D, Friedlich PS, et al. The financial burden experienced by families of preterm infants after NICU discharge. *Journal of Perinatology* 2022;**42**(2):223–30. [DOI: 10.1038/s41372-021-01213-4] [PMID: 34561556]

Lancet Child Adolescent Health 2019

Lancet Child & Adolescent Health. Putting the family at the centre of newborn health. Lancet Child and Adolescent Health 2019;**3**(1):1-1. [PMID: 30501959]

Loewy 2013

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;**31**(5):902-18. [DOI: 10.1542/peds.2012-1367] [PMID: 23589814]

Loewy 2015

Loewy, J. NICU music therapy: song of kin as critical lullaby in research and practice. *Annals of the New York Academy of Sciences* 2015;**1337**:178-85. [DOI: 10.1111/nyas.12648] [PMID: 25773633]

Lordier 2019

Lordier L, Meskaldji DE, Grouiller F, Pittet MP, Vollenweider A, Vasung L, et al. Music in premature infants enhances high-level cognitive brain networks. *Proceedings of the National Academy of Sciences* 2019;**116**(24):12103-8. [DOI: 10.1073/pnas.1817536116] [PMID: 31138687]

Malouf 2022

Malouf R, Harrison S, Burton HL, Gale C, Stein A, Franck LS, et al. Prevalence of anxiety and post-traumatic stress (PTS) among the parents of babies admitted to neonatal units: A systematic review and meta-analysis. *EClinicalMedicine* 2022;**43**:101233. [DOI: 10.1016/j.eclinm.2021.101233] [PMID: 34993425]

McMahon 2012

McMahon E, Wintermark P, Lahav A. Auditory brain development in premature infants: the importance of early experience. *Annals of the New York Academy of Sciences* 2012;**1252**:17-24. [DOI: 10.1111/j.1749-6632.2012.06445.x] [PMID: 22524335]

Melchers 2009

Melchers P, Preuss U. Kaufman Assessment Battery for Children. German edition. Frankfrut am Main (Germany): Pearson Assessment, 2009.

Ment 2008

Ment LR, Vohr BR. Preterm birth and the developing brain. *Lancet Neurology* 2008;**7**(5):378-9. [DOI: 10.1016/S1474-4422(08)70073-5.] [PMID: 18420149]

Mohan 2021

Mohan A, Gokulakrishnan G, El-Saie A, Brickley A, Hagan J, Pammi M. Music therapy for preterm neonates in the neonatal intensive care unit: an overview of systematic reviews. *Acta Paediatrica* 2021;**110**(12):3180-200. [DOI: 10.1111/apa.16055] [PMID: 34346112]

Mondanaro 2016

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

Monson 2018

Monson BB, Eaton-Rosen Z, Kapur K, Liebenthal E, Brownell A, Smyser CD, et al. Differential rates of perinatal maturation of human primary and nonprimary auditory cortex. *eNeuro* 2018;**5**(1):1-12. [PMID: 29354680]

Moon 2013

Moon CM, Lagercrantz H, Kuhl PK. Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paediatrica, International Journal of Paediatrics* 2013;**102**(2):156-60. [DOI: 10.1111/apa.12098] [EMBASE: 000313593700023] [PMID: 23173548]

Mushtaq 2019

Mushtaq A, Kazi F. Family-centered care in the NICU. *Lancet Child & Adolescent Health* 2019;**3**(5):295-6. [DOI: 10.1016/S2352-4642(19)30089-6]

Park & Demakis 2017

Park SE, Demakis GJ. Wechsler preschool and primary scale of intelligence. In: Zeigler-Hill V, Shackelford T, editors(s). Encyclopedia of Personality and Individual Differences. Cham: Springer International Publishing, 2017:1-4.

Park 2014

Park M, Kohlrausch A, de Bruijn W, de Jager P, Simons K. Analysis of the soundscape in an intensive care unit based on the annotation of an audio recording. *Journal of the Acoustical Society of America* 2014;**135**(4):1875-86. [DOI: 10.1121/1.4868367] [PMID: 25234986]

Partanen 2013

Partanen E, Kujala T, Tervaniemi M, Huotilainen M. Prenatal music exposure induces long-term neural effects. *PLOS One* 2013;**8**(10):e78946. [DOI: 10.1371/journal.pone.0078946] [PMID: 24205353]

Perani 2010

Perani D, Saccuman MC, Scifo P, Spada D, Andreolli G, Rovelli R, et al. Functional specializations for music processing in the human newborn brain. *Proceedings of the Natlonal Academy of Sciences of the USA* 2010;**107**(10):4758-63. [DOI: 10.1073/pnas.0909074107] [PMID: 20176953]

Philbin 2017

Philbin MK. The sound environments and auditory perceptions of the fetus preterm newborn. In: Filippa M, Kuhn P, Westrup B, editors(s). Early Vocal Contact and Preterm Infant Brain Development. Cham, Switzerland: Springer International Publishing, 2017:91-112. [DOI: 10.1007/978-3-319-65077-7_6]

Pierrat 2017

Pierrat V, Marchand-Martin L, Arnaud C, Kaminski M, Resche-Rigon M, Lebeaux C, et al. Neurodevelopmental outcome at 2 years for preterm children born at 22 to 34 weeks' gestation in France in 2011: EPIPAGE-2 cohort study. *BMJ* 2017;**358**:1-13. [DOI: 10.1136/bmj.j3448] [PMID: 28814566]

Pineda 2014

Pineda RG, Neil J, Dierker D, Smyser CD, Wallendorf M, Kidokoro H, et al. Alterations in brain structure and neurodevelopmental outcome in preterm infants hospitalized in different neonatal intensive care unit environments. *Journal of Pediatrics* 2014;**164**(1):52-60.e2. [DOI: 10.1016/j.jpeds.2013.08.047] [PMID: 24139564]

RevMan Web 2022 [Computer program]

Review Manager Web (RevMan Web). Version 4.12.0. Copenhagen: The Cochrane Collaboration, 2022. Available at revman.cochrane.org.

Rickard 2005

Rickard NS, Toukhsati SR, Field SE. The effect of music on cognitive performance: insight from neurobiological and animal studies. *Behavioral Cognitive Neuroscience Reviews* 2005;**4**(4):235-61. [DOI: 10.1177/1534582305285869] [PMID: 16585799]

Roque 2017

Roque AF, Lasiuk GC, Radünz V, Hegadoren K. Scoping review of the mental health of parents of infants in the NICU. *Journal of Obstetric Gynecologic and Neonatal Nursing* 2017;**46**(4):576-87. [DOI: 10.1016/j.jogn.2017.02.005] [PMID: 28506679]

Rossetti 2013

Rossetti A, Canga B. Environmental music therapy: rationale for 'multi-individual' music psychotherapy in modulation of the pain experience. In: Mondanaro JF, Sara AG, editors(s). Music and Medicine: Integrative Models in the Treatment of Pain. New York, NY: Satchnote Press: The Louis Armstrong Center for Music and Medicine, 2013:468.

Ruiz 2018

Ruiz N, Piskernik B, Witting A, Fuiko R, Ahnert L. Parent-child attachment in children born preterm and at term: a multigroup analysis. *PLoS ONE* 2018;**13**(8):1-14. [DOI: 10.1371/journal.pone.0202972] [PMID: 30161170]

Rücker 2012

Rücker G. Network meta-analysis, electrical networks and graph theory. *Research Synthesis Methods* 2012;**3**(4):312-24. [DOI: 10.1002/jrsm.1058] [PMID: 26053424]

Rücker 2014

Rücker G, Schwarzer G. Reduce dimension or reduce weights? Comparing two approaches to multi-arm studies in network meta-analysis. *Statistics in Medicine* 2014;**33**(25):4353-69. [DOI: 10.1002/sim.6236] [PMID: 24942211]

Rücker 2016

Rücker G, Schwarzer G. Automated drawing of network plots in network meta-analysis. *Research Synthesis Methods* 2016;**7**(1):94-107. [DOI: 10.1002/jrsm.1143] [PMID: 26060934]

Sacks 2007

Sacks O. Tales of Music and the Brain. New York (NY): Vintage Books, 2007.

Schulz 1994

Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA* 1994;**272**(2):125-8. [PMID: 8015122]

Schulz 2000

Schulz KF. Assessing allocation concealment and blinding in randomised controlled trials: why bother? *Evidence-based Nursing* 2001;**4**(1):4-6. [PMID: 11707879]

Schünemann 2013

Schünemann H, Brożek J, Guyatt G, Oxman A, editor(s). Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach (updated October 2013). GRADE Working Group, 2013. Available from gdt.guidelinedevelopment.org/app/handbook/handbook.html.

Shoemark 2015

Shoemark H, Hanson-Abromeit D, Stewart L. Constructing optimal experience for the hospitalized newborn through neuro-based music therapy. *Frontiers in Human Neuroscience* 2015;**9**(487):1-5. [DOI: 10.3389/fnhum.2015.00487] [PMID: 26388762]

Spielberger 1983

Spielberger C, Gorsuch R, Lushene R, Vagg P, Jacobs G. Manual for the State-Trait Anxiety Inventory. Palo Alto (CA): Consulting Psychologists Press, 1983.

Standley 2012

Standley J. Music therapy research in the NICU: an updated meta-analysis. *Neonatal Network* 2012;**31**(5):311-6. [DOI: 10.1891/0730-0832.31.5.311] [PMID: 22908052]

Stoll 2015

Stoll BJ, Hansen NI, Bell EF, Walsh MC, Carlo WA, Shankaran S, et al. Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993–2012. *JAMA* 2015;**314**(10):1039-51. [DOI: 10.1001/jama.2015.10244] [PMID: 26348753]

Symington 2006

Symington A, Pinelli J. Developmental care for promoting development and preventing morbidity in preterm infants. *Cochrane Database of Systematic Reviews* 2006, Issue 2. Art. No: CD001814. [DOI: 10.1002/14651858.CD001814.pub2] [PMID: 16625548]

Tramo 2011

Tramo MJ, Lense M, van Ness C, Kagan J, Settle MD, Cronin JH. Effects of music on physiological and behavioral indices of acute pain and stress in premature infants: clinical trial and literature review. *Music and Medicine* 2011;**3**(2):72-83. [DOI: 10.1177/1943862111400613] [EMBASE: 69665973]

Twilhaar 2018

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. *Archives of Disease in Childhood: Fetal and Neonatal Edition* 2018;**103**(4):322-30. [DOI: 10.1136/archdischild-2017-312916] [PMID: 28847871]

Van der Heijden 2016

Van der Heijden MJ, Oliai Araghi S, Jeekel J, Reiss IK, Hunink MG, Van Dijk M. Do hospitalized premature infants benefit from music interventions? A systematic review of randomized controlled trials. *PLOS One* 2016;**11**(9):e0161848. [DOI: 10.1371/journal.pone.0161848] [PMID: 27606900]

Van Dokkum 2021

Van Dokkum NH, Kooi EM, Berhane B, Ravensbergen A-G, Hakvoort L, Jaschke AC, et al. Neonatal music therapy and cerebral oxygenation in extremely and very preterm infants: a pilot study. *Music and Medicine* 2021;**13**(2):91-8. [DOI: 10.47513/mmd.v13i2.813]

Wachman 2011

Wachman EM, Lahav A. The effects of noise on preterm infants in the NICU. *Archives of Disease in Childhood. Fetal and Neonatal Edition* 2011;**96**(4):F305-9. [DOI: 10.1136/adc.2009.182014] [PMID: 20547580]

Woodward 2006

Woodward LJ, Anderson PJ, Austin NC, Howard K, Inder TE. Neonatal MRI to predict neurodevelopmental outcomes in preterm infants. *New England Journal of Medicine* 2006;**355**(7):685-94. [DOI: 10.1056/NEJMoa053792] [PMID: 16914704]

Yan 2003

Yan J. Canadian Association of Neuroscience Review: development and plasticity of the auditory cortex. *Canadian Journal of Neurological Sciences* 2003;**30**(3):189-200. [PMID: 12945940]

Yue 2021

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. *Journal of Advanced Nursing* 2021;**77**(2):635-52. [DOI: 10.1111/jan.14630] [PMID: 33200833]

References to other published versions of this review

Haslbeck 2019

Haslbeck FB, Karen T, Loewy J, Meerpohl JJ, Bassler D. Musical and vocal interventions to improve neurodevelopmental outcomes for preterm infants (Protocol). *Cochrane Database of Systematic Reviews* 2019, Issue 11. Art. No: CD013472. [DOI: 10.1002/14651858.CD013472]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Calabro 2003

Study characteristics	
Methods	RCT, parallel group
Participants	Preterm infants GA 34, diagnosed with chronic lung disorder or oxygen dependency, or both, availability of 4 consecutive days, no substance addiction ***Pretreatment:*** *data not shown: only reported that groups were balanced for weight, type of oxygen, diagnosis, therapy, average heart, and oxygen saturation rates, sex* ***N infants randomised:*** *22* ***N infants analysed:*** *17* ***Reasons for dropout:*** *1 infant discharged* ***Reasons for exclusion:*** *4 infants showed hearing impairment* ***Sample size calculation:*** *sample size of 20 participants was calculated for the primary outcome of oxygen saturation level. Ten participants per group were required to detect a 5% difference in oxygen saturation (effect size of 1.3) at a significance level of 0.05 and a power of 80%.*
Interventions	**Intervention characteristics** Control: usual NICU environmental sounds Recorded sedative music • *Intervention type*: recorded sedative music: Compact Disc (quote:) "music for dreaming": instrumental (strings, flute and harp) lullabies Brahms Lullaby and Sandman • *Dose*: 10-minute silence, 20-minute music from the 2 tracks, 15-minute silence after (total 45 minutes) • *Frequency*: once per day for 4 days • *Mode of delivery*: in isolette or crib by nurse researcher via mini-speakers • *dB level*: 60-70 decibels
Outcomes	*Heart rate* • **Outcome type**: continuous outcome • **Reporting**: partially reported • **Unit of measure**: beats per minute • **Direction**: lower is better • **Data value**: change from baseline • **Notes**: no significant results: changes from baseline period to: a) music/ silence period: difference in mean change (95% CI): -0.9 (-5.0 to 3.2); P = 0.64; b) post-observation: difference in mean change (95% CI): -0.1. (-7.8 to 7.6); P = 0.98 *Respiratory rate* • **Outcome type**: continuous outcome

Calabro 2003 *(Continued)*

- **Reporting**: partially reported
- **Unit of measure**: beats per minute
- **Direction**: lower is better
- **Data value**: change from baseline
- **Notes**: no significant results: changes from baseline period to: a) music/ silence period: difference in mean change (95% CI): 1.8 (-2.4 to 6.0); P = 0.38; b) post-observation: difference in mean change (95% CI): 0.1. (-4.3 to 4.5); P = 0.96

Oxygen saturation

- **Outcome type**: continuous outcome
- **Reporting**: partially reported
- **Unit of measure**: %
- **Direction**: lower is better
- **Data value**: change from baseline
- **Notes**: no significant results: changes from baseline period to: a) music/ silence period: difference in mean change (95% CI): 0.7 (-0.9 to 2.3); P = 0.36; b) post-observation: difference in mean change (95% CI): 0.4 (-1.3 to 2.0); P = 0.62

Organised behaviours

- **Outcome type**: Continuous outcome
- **Reporting**: Partially reported
- **Scale**: Physiological and Behavioural Assessment From (adapted from Als 1986)
- **Range**: 0 - 11
- **Unit of measure**: per minute
- **Data value**: change from baseline
- **Notes**: no significant results:changes from baseline period to: a) music/silence period: difference in mean change (95% CI): 0.0 (-0.2 to 0.2); P = 0.78 b) post-observation: difference in mean change (95% CI): 0.1. (-0.2 to 0.3); P = 0.63

Disorganised behaviour

- **Outcome type**: continuous outcome
- **Reporting**: partially reported
- **Scale**: Physiological and Behavioural Assessment From (adapted from Als 1986)
- **Range**: 0-11
- **Unit of measure**: per 10 minutes
- **Data value**: change from baseline
- **Notes**: no significant results: changes from baseline period to: a) music/ silence period: difference in mean change (95% CI): 0.1 (-0.9 to 1.1); P = 0.83) post-observation: difference in mean change (95% CI): -0.2. (-1.7 to 1.3); P = 0.74

Identification	**Sponsorship source:** supported by Ronald McDonald House **Declarations of interest:** not stated **Country:** Australia **Setting:** Tertiary level NICU and special care nursery **Comments:** single-centre study **Authors name:** Jacinta Calabro **Institution:** 1) Monash Medical Centre, Melbourne; 2) Royals Children's Hospital Melbourne **Email:** jacinta@musictherapyonline.org **Address:** 1) 246 Clayton Roas, Clayton VIC31682) 50 Flemington Rd, Parkville Vic 3052, Australia

Calabro 2003 *(Continued)*

	Website: 1) monashhealth.org; 2) www.rch.org.au; 3) http://tlcmusic.com.au/about
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random sequence generation reported
Allocation concealment (selection bias)	Low risk	No information given for allocation concealment
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Blinding of outcome assessment reported
Incomplete outcome data (attrition bias) All outcomes	Low risk	Data reporting in the results were not shown in the table but delivered by the author later on.
Selective reporting (reporting bias)	Unclear risk	Selective reporting unclear
Other bias	Unclear risk	Other bias unclear

Caparros-Gonzalez 2018

Study characteristics

Methods	**Study design:** randomised controlled trial **Study grouping:** parallel group
Participants	**Baseline characteristics** Recorded sedative instrumental music • *Number participants*: 9 • *Corrected age, wk as mean (SD)*: 32.96 (1.84) • *Gestational age at intervention (weeks) as mean (SD)*: 33.77 (1.56) • *Gender: % males (n) as mean (SD)*: 56.25 (5) • *Birth weight (g) as mean (SD)*: 1615 (329.28) • *Apgar test score*: 7.10 (2.22) Control: silence • *Number participants*: 8 • *Corrected age, wk as mean (SD)*: 33.04 (1.74) • *Gestational age at intervention (weeks) as mean (SD)*: 34.75 (1.98)

Caparros-Gonzalez 2018 *(Continued)*

- *Gender: % males (n) as mean (SD)*: 69.23 (5)
- *Birth weight (g) as mean (SD)*: 1719.92 (434.39)
- *Apgar test score*: 7.85 (1.95) 5 minutes

Overall

- *No. participants*: 17
- *Corrected age, wk as mean (SD)*: P alue 0.27
- *Gestational age at intervention (weeks) as mean (SD)*: P = 0.91
- *Gender: % males (n) as mean (SD)*: P = 0.47
- *Birth weight (g) as mean (SD)*: P = 0.47
- *Apgar test score*: P = 0.34

Inclusion criteria: the inclusion criteria were premature infants who had a satisfactory hearing screening test through a brain stem auditory evoked potentials test at 38 weeks and were in the incubator while in the NICU.

Exclusion criteria: premature infants were excluded if they were being treated with sedatives, analgesia medications, caffeine, or steroids; if they were supported by ventilation, continuous positive airway pressure, or nasal cannula; if they had any auditory impairment, malformation, or disease; if they were exposed to any potentially central nervous system depressor prenatally; or if they were diagnosed with a heart, lung, or congenital disease.

Pretreatment: there were no significant differences between groups.

N infants analysed: 17

N infants randomised: 22

Reasons for dropouts: 5 participants were discharged from the incubator before finishing the intervention. These participants were excluded from the data analysis, as major protocol violations happened (they received only the adaptation session of the intervention).

Reasons for exclusion: 5 participants were discharged from the incubator before finishing the intervention. These participants were excluded from the data analysis, as major protocol violations happened (they received only the adaptation session of the intervention).

Sample size calculation: was conducted. We estimated that a sample size of 8 subjects was needed to obtain a good effect size of the relaxing music therapy intervention ($f = 0.41$; $\alpha = 0.05$, $1 - \beta = 0.80$), assuming a correlation between the levels of the within-subjects variable = 0.50.

Interventions	**Intervention characteristics** Recorded sedative instrumental music • *Dose*: 10-minute baseline period (pre-session measurement), followed by a 20-minute music/silence period. Finally, there was a 2-minute post-session measurement • *Frequency*: 8 sessions across 3 consecutive days; the first session was considered an adaptation session to music devices and was not included in the analyses • *Delivery*: by blinded nurse in incubator via speaker • *dB level*: 30 • *Genre*: a relaxing tune composed and synthesised by Melo-mics computer system with no human intervention. Slow tempo (up to 52-54 beats per minute), string and wind instruments, highly predictable melody, low tone, lack of dissonances, and at a volume below 50 decibels. Control: Silence • *Dose*: 10-minute baseline period (pre-session measurement), followed by a 20-minute music/silence period. Finally, there was a 2-minute post-session measurement • *Frequency*: 8 sessions across 3 consecutive days; the first session was considered an adaptation session to music devices and was not included in the analyses

Caparros-Gonzalez 2018 *(Continued)*

- *Delivery*: by blinded nurse in incubator via speaker
- *dB level*: 0 dB

Outcomes	*Heart rate* • **Outcome type**: continuous outcome • **Reporting**: fully reported • **Unit of measure**: beats per minute • **Direction**: lower is better • **Data value**: endpoint • **Notes**: endpoint: post-intervention *Respiratory rate* • **Outcome type**: continuous outcome • **Reporting**: partially reported • **Unit of measure**: breaths per minute • **Direction**: lower is better • **Data value**: endpoint • **Notes**: endpoint: post-intervention *Oxygen saturation* • **Outcome type**: continuous outcome • **Reporting**: fully reported • **Unit of measure**: in percentage • **Direction**: higher is better • **Data value**: endpoint • **Notes**: endpoint: post-intervention
Identification	**Sponsorship source:** The study was funded by Spanish Ministry of Science and Innovation grant (INNPACTO IPT300000-2010-10). R.C.G. is supported by the I + D Project "PSI2015-348 63494-P" of the Spanish Ministry of Science and Innovation. A.T.-L. is supported by an FPI grant from the Spanish Ministry and Competitiveness (BES-2013-064257). C.D.-P. is supported by a UGR Postdoctoral Fellowship (2013 University of Granada. **Declarations of interest:** None of the authors has biomedical financial interests or potential conflicts of interest. **Country:** Spain **Setting:** NICU **Comments:** **Authors name:** Rafael A. Caparros-Gonzalez **Institution:** Mind, Brain, and Behaviour Research Center (CIMCYC), Campus de Cartuja s/n, 18011, Universidad de Granada, Granada, Spain **Email:** rcg477@correo.ugr.es; rafawolfy@gmail.com **Address:** Mind, Brain, and Behavior Research Center (CIMCYC), Campus de Cartuja s/n, 18011, Universidad de Granada, Granada, Spain **Website:**
Notes	

Risk of bias

Caparros-Gonzalez 2018 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomly allocated
Allocation concealment (selection bias)	Low risk	To ensure allocation concealment, a 3rd researcher was responsible for this task. This researcher stored the experimental music tracks on a secure digital card, depending on the randomised matching. When the participant's code was matched to the experimental condition, the relaxing music track was recorded; otherwise, the track consisted of silence.
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Double-blinded
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Double-blinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	5 participants were discharged from the incubator before finishing the intervention. These participants were excluded from the data analysis, as major protocol violations happened (they received only the adaptation session of the intervention). Therefore, results from 17 infants with gestational age at the time of delivery ranging from 32 to 36 weeks (M = 32.33; SD = 1.79) were analysed. According to an updated meta-analysis, we estimated that a sample size of 8 subjects was needed to obtain a good effect size of the relaxing music therapy intervention (f = 0.41; α = 0.05, 1 − β = 0.80), assuming a correlation between the levels of the within-subjects variable = 0.50.
Selective reporting (reporting bias)	Low risk	Low
Other bias	Unclear risk	Unclear

Cevasco 2008

Study characteristics	
Methods	**Study design:** randomised controlled trial **Study grouping:** parallel group
Participants	**Baseline characteristics** Control • *Gestational age at birth (weeks) mean (SD)*: 32.46 (1.57) • *Birth weight (g) mean (SD)*: 1762.80 (379.28) • *Gender (male/female)*: 6/4 • *Race (white/black)* : 6/4 • *Number of infants*: 10 • *Number of mothers*: 8 • *Age of mothers as mean (SD)*: 23.50 (5.37) • *Children per mother*: 1.75 • *First-time mothers as mean (SD)*: 5 (0.83)

Cevasco 2008 *(Continued)*

- *Race (white/black/others)*: 4/4/0
- *Marital status (married/single)*: 2/6

Intervention A

- *Gestational age at birth (weeks) mean (SD):* 31.73 (1.80)
- *Birth weight (g) mean (SD):* 1551.7 (352.36)
- *Gender (male/female)*: 6/4
- *Race (white/black)*: 6/4
- *Number of infants*: 10
- *Number of mothers*: 8
- *Age of mothers as mean (SD)*: 28.13 (6.53)
- *Number of children per mother*: 1.63
- *First-time mothers as mean (SD)*: 6 (1.19)
- *Race (white/ black/others)*: 6/2/0
- *Marital status (married/single)*: 6/2

Overall

- *Gender (male/female)*: 12/8
- *Race (white/black)*: 12/8
- *Number of infants*: 20
- *Number of mothers*: 16
- *Marital status (married/single)*: significant difference

Inclusion criteria: born prior to 36 weeks' corrected gestational age (CGA), and low birth weight infants (LBW), weighing less than 2500 grams at birth, hospitalised in a Level II (intermediate care for premature infants to grow and gain feeding skills) and Level III NICU (high-risk care for infants needing advanced treatment to sustain life) infant born at or later than 28 and before 36 weeks' CGA, infant evinced toleration of auditory stimulation between 30-32 weeks CGA.

Exclusion criteria: severe abnormalities which affected the ability to listen to music (including significant neurological disorders such as periventricular leukomalacia or intraventricular haemorrhage), mother or infant tested positive for illegal drugs, the infant had disease necessitating quarantine, the infant on a ventilator or continuous positive airway pressure past 30 weeks' CGA.

Pretreatment: analysis of demographic information for preterm infants revealed that there was no significant difference between groups and birth CGA, t (18) = -0.960, P > 0.05. The age range for infants in the experimental group was from 28 weeks to 34 weeks and 4 days CGA and 30 weeks to 34 weeks and 6 days in the control group. Also, there was no significant difference in birth weight between groups, t (18) = 0.214, P > 0.05. Birth weight ranged from 954 to 2080 grams for the experimental group and from 1200 to 2535 grams for the control group.

N infants analysed: 20 (n = 10 experimental group, n = 10 control group)

N infants randomised: 25 (n = 13 experimental group; n = 11 control group)

Reasons for dropouts: 1 mother in the experimental and 1 in the control group did not complete a phone or mail survey.

Reasons for exclusion: 2 mothers of twins in the experimental group were dropped from the study due to early discharge and only 1 or 2 days of music intervention and 1 infant in the control group became extremely sick and participation in the study was discontinued.

N mothers randomised: 21 (n = 11 experimental group, n = 10 control group)

Mothers analysed: 16 (n = 8 experimental group, n = 8 control group)

Interventions	**Intervention characteristics** Control

Cevasco 2008 *(Continued)*

- *Intervention type*: standard care

Intervention A

- *Intervention type*: recorded lullabies sang by mother (lullabies and children songs with guitar accompaniment by researcher from provided list picked by mother)
- *Dose*: 20 minutes per day
- *Frequency*: 3-5 times per week until discharge
- *Mode of delivery*: recorded, via speaker in isolette or open crib with infants lying on their backs or sides
- *dB level*: 65 db

Outcomes	*Adapted Parental Perception Inventory* • **Outcome type**: continuous outcome • **Reporting**: partially reported • **Scale**: Adapted Parental Perception Inventory • **Range**: 0-13 • **Direction**: lower is better • **Data value**: endpoint • **Notes**: the Parental Perception Inventory (PPI), which was administered post-test only, indicated experimental preterm mothers scored the highest on the PPI, indicating less adjustment to their baby and lifestyle changes compared to control preterm mothers as well as experimental and control full-term mothers. *Weight gain* • **Outcome type**: continuous outcome • **Reporting**: partially reported • **Unit of measure**: kg per day • **Direction**: higher is better • **Notes**: mean weight gained per day was 17.70 grams for infants in the experimental group and 16.95 for infants in the control group. A t-test indicated that there was no difference between groups for weight gained, t (18) = -0.214, P > 0.05 *Hospitalisation* • **Outcome type**: continuous outcome • **Reporting**: partially reported • **Unit of measure**: in days • **Direction**: lower is better • **Notes**: number of days in hospital, t (18) = 0.53, P > 0.05. Infants in the experimental group left the hospital 2 days sooner than those in the control group, at an average of 15.6 days compared to 17.8 days.
Identification	**Sponsorship source:** not mentioned **Declarations of interest:** not stated **Country:** USA **Setting:** Newborn Nursery, NICU Level I, II, III **Comments:** single centre **Authors name:** Andrea M. Cevasco **Institution:** University of Alabama **Email:** acevasco@ua.edu

Cevasco 2008 *(Continued)*

Address: Tuscaloosa, AL 35487, Vereinigte Staaten

Website: https://music.ua.edu/people/andrea-cevasco-trotter/

Notes

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Random sequence generation unclear
Allocation concealment (selection bias)	Low risk	Allocation concealment was done
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Unclear if participants and personnel were blinded
Blinding of outcome assessment (detection bias) All outcomes	High risk	No blinding of outcome assessment
Incomplete outcome data (attrition bias) All outcomes	High risk	All data only for the remaining participants after dropouts and exclusions
Selective reporting (reporting bias)	Unclear risk	Unclear selective reporting
Other bias	High risk	Per-protocol analysis and not intention-to-treat Number of randomised PI reported was inconsistent/controversial: in abstract: was 20 and later intent was 25 (with 11 + 13 = 24 and not 25!)

Epstein 2021

Study characteristics

Methods

Study design: randomised controlled trial

Study grouping: cross-over

Participants

Baseline characteristics

SSC mother and infant

- *Gestational age at birth (weeks): n.a.*
- *Gestational age at intervention (weeks) mean ± SD*: 31 ± 2.7
- *Birth weight (g): n.a.*
- *Gender n (%) male*: *n.a.*
- *Gestational age at endpoint (weeks) mean ± SD*: 31 ± 2.7

SSC & live mothers' singing in lullaby style (song chosen by mother) accompanied by a music therapist with voice and or guitar

Epstein 2021 *(Continued)*

- *Gestational age at birth (weeks): n.a.*
- *Gestational age at intervention (weeks) mean ± SD*: 32 ± 3.5
- *Birth weight (g): n.a.*
- *Gender n (%) male*: n.a.
- *Gestational age at endpoint (weeks) mean ± SD:* 32 ± 3.5

Overall

- *Gestational age at birth (weeks) mean ± SD*: 27 ± 2.5
- *Birth weight (g) mean ± SD*: 823 ± 213
- *Gender n (%) male*: 21 (60%)
- *Intraventricular haemorrhage grade 3 count (percentage)*: 16 (45)
- *Intraventricular haemorrhage grade 4 count (percentage)*: 10 (29)
- *Periventricular leukomalacia median count (percentage)*: 9 (26)
- *Education mothers median (interquartile range)*: 12 (10–20)

Inclusion criteria: All preterm infants born before 32 weeks' gestation, admitted to the NICU during 2014–2018 who developed IVH grades 3 or 4 and/or PVL diagnosed by brain ultrasound. Inclusion after their clinical state stabilised, regardless of their corrected age at study entry. A clinically stable infant at the time of enrolment was defined as one without supplemental oxygen demands during enrolment, no significant apnoea requiring stimulation during the 3 days before enrolment, with or without methylxanthines, and no documented seizure episode or anticonvulsive treatment a week before the study enrolment. All infants had their hearing confirmed by distortion product otoacoustic emissions. Due to the national parental rights of birth and preterm birth in Israel, where fathers are entitled to very short paternity leave, only mothers who were able to attend all the sessions participated in the study.

Exclusion criteria: if it became necessary for medical reasons to discontinue an infant's participation in the study for a period shorter than 2 weeks, participation recommenced when the criteria for clinical stability were again met.

Pretreatment: no group differences possibly emissions due to within-subject cross-over design

N infants analysed: 35

N infants randomised: 40

Reasons for dropouts: no dropouts

Reasons for exclusion: analysed (n = 35). Excluded from analysis (n = 5): did not complete all sessions = 3; prolonged illness between sessions = 2

Sample size calculation: the minimal clinically significant difference for LF/HF power in infants is unknown. The power calculation for the primary outcome was informed by a previous study [18] that found mean LF/HF power values of 1 (SD = 0.8) during MT combined with SSC, compared to 3.4 (SD = 1.6) during SSC alone. Therefore, the study was powered to detect a substantial effect on LF/HF (d = 0.80). In an individually randomised trial, with a 2-sided 5% significance level and 80% power, a sample size of 35 infants was required.

Interventions	**Intervention characteristics** SSC mother and infant • *Intervention type*: SSC as standard care procedure • *Dose*: sessions began with 10 minutes of SSC alone (baseline phase), followed by either SSC for 20 minutes alone or the intervention of combined SSC and maternal singing during MT (intervention phase). The sessions ended with another 10 minutes of SSC alone (recovery phase). • *Frequency*: each therapy session began 30 minutes after completion of feeding, 2 to 3 times per week. All SSC or maternal singing, or both, during MT sessions were offered about 3–5 days after the previous session, according to mothers' ability and presence in the NICU and the infants' length of hospitalisation. • *Mode of delivery*: SSC by mother without music

Epstein 2021 *(Continued)*

- *dB level*: not reported

SSC & live mothers` singing in lullaby style (song chosen by mother) accompanied by a music therapist with voice and or guitar

- *Intervention type*: a certified music therapist delivered the intervention. During the intervention, the mother was guided by the therapist to sing her chosen songs/lullabies using a repetitive, soothing tone, softly, simply, and with a slow tempo. According to the mother's choice, the music therapist provided vocal or instrumental support for the mother using a nylon string guitar. As both the mother and music therapist were women, the vocal range during the intervention was the average female vocal range (G3 to C5). The music therapist helped the mother attune the live music to the infant's breathing patterns, movements, vocals, or other viewed reactions. According to the infant's reactions, the mother and music therapist decided whether to continue with the same rhythm, to change the tempo and vocals, or pause. After the singing, the session ended with a few moments of silence and reorganisation of breathing. Then, the mother was encouraged to reflect on her experience.
- *Dose*: sessions began with 10 minutes of SSC alone (baseline phase), followed by either SSC for 20 minutes alone or the intervention of combined SSC and maternal singing during MT (intervention phase). The sessions ended with another 10 minutes of SSC alone (recovery phase).
- *Frequency*: each therapy session began 30 minutes after completion of feeding, 2 to 3 times per week. All SSC or maternal singing, or both, during MT sessions were offered about 3–5 days after the previous session, according to mothers' ability and presence in the NICU and the infants' length of hospitalisation.
- *Mode of delivery*: live by mother in SSC and accompanied by music therapist
- *dB level*: not reported

Outcomes

Heart rate

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Unit of measure**: beats per minute
- **Direction**: lower is better
- **Data value**: endpoint

Respiratory rate

- **Outcome type**: continuous outcome
- **Unit of measure**: number per minute
- **Direction**: lower is better
- **Data value**: endpoint

Oxygen saturation

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Direction**: higher is better
- **Data value**: endpoint

State-Trait-Anxiety-Inventory (STAI)

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Scale**: 20 statements, each item scored from 1 to 4
- **Range**: 20-80
- **Unit of measure**: 1 point
- **Direction**: lower is better
- **Data value**: endpoint
- **Notes**: this is a validated, self-report questionnaire with 20 statements describing present anxiety symptoms. Each item is scored on a scale from 1 (not at all) to 4 (very much), leading to total scores ranging from 20 to 80, with higher scores indicating more anxiety. A cut-off point of 39–40 has been

Epstein 2021 *(Continued)*

suggested to detect clinically significant symptoms of anxiety. A reliable, validated Hebrew version was used. After each session, mothers were asked to complete the STAI questionnaire and return it in a sealed envelope.

Heart rate mothers

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Unit of measure**: beats per minute
- **Direction**: lower is better
- **Data value**: endpoint

Oxygen saturation mothers

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Direction**: higher is better
- **Data value**: endpoint

HRV: low-frequency power

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Unit of measure**: ms^2/Hz
- **Data value**: endpoint
- **Notes**: heart rate variability* (measured by low-frequency power; high-frequency power; low-frequency/high-frequency ratio reflecting the balance between sympathetic and parasympathetic tone)

HRV: high-frequency power

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Data value**: endpoint

HRV: LF/HF ratio

- **Outcome type**: continuous outcome
- **Direction**: lower is better
- **Data value**: endpoint
- **Notes**: low-frequency/high-frequency (LF/HF) power. The assumption underlying the LF/HF ratio is that LF power may be generated by the sympathetic nervous system, while HF power is produced by the parasympathetic nervous system. In this model, a low LF/HF ratio reflects parasympathetic dominance. This is seen when we conserve energy and engage in tend-and-befriend behaviours. In contrast, a high LF/HF ratio indicates sympathetic dominance, which occurs when we engage in fight-or-flight behaviours or parasympathetic withdrawal.

Behavioural state (Als)

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Scale**: 7 behaviours
- **Direction**: lower is better
- **Data value**: endpoint
- **Notes**: the behavioural state of the preterm infants during intervention sessions was evaluated using the criteria of Als, based on the NIDCAP manual for naturalistic observation and the Brazelton Neonatal Behavioural Assessment Scale. It is used to assess and differentiate between 7 behavioural states: (1) deep sleep, with obligatory regular breathing, relaxed facial expression; (2) diffuse light sleep with eyes closed, low activity level, respirations are regular, mild sucking; (3) diffusely drowsy, fussy, and/or little discharge of vocalisation, whimpers, facial grimacing; (4) hyper-alertness, eyes wide open, giv-

Epstein 2021 *(Continued)*

	ing the impression of panic, fear or being overwhelmed; (5) actively awake and aroused, may also be clearly fussy without crying robustly; (6) highly aroused, agitated, upset and/or crying; and (7) removal from the state continuum, a prolonged respiratory pause beyond 8 s. Lower behavioural states indicate higher sleep quality with fewer signs of discomfort and, hence, greater stability. The music therapy was provided by a trained, NIDCAP-certified therapist. The research assistant was familiar with physiological parameter recordings and the Behavioural Scale and recorded this information every 5 minutes during each session.
Identification	**Sponsorship source:** not reported **Declaration of interest:** not stated **Country:** Israel **Setting:** Level III NICU **Comments:** **Authors name:** Shulamit Epstein **Institution:** School of Creative Arts Therapies, University of Haifa, Haifa, Israel **Email:** harnon@netvision.net.il **Address:** Department of Neonatology, Meir Medical Center, 59 Tchernichovsky St., 44281 Kfar Saba, Israel **Website:** https://hw.haifa.ac.il/en/departments/artschool
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random sequence generation reported
Allocation concealment (selection bias)	Low risk	Sealed, opaque envelope
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	Low risk	The outcome assessor was blinded to treatment allocation and arranged physiological data of infants and mothers, STAI score, and behavioural scale according to the sequence of the intervention, without knowing whether the *intervention was SSC or maternal singing with MT.*
Incomplete outcome data (attrition bias) All outcomes	Low risk	3 did not complete all sessions, and 2 were excluded due to prolonged illnesses during their hospitalisations that interrupted the timing between study sessions for more than 2 weeks. Therefore, 35 preterm infants were included in the final analysis.
Selective reporting (reporting bias)	Low risk	No selective reporting detected
Other bias	Low risk	No other bias detected

Ettenberger 2014

Study characteristics	
Methods	**Study design:** randomised controlled trial **Study grouping:** parallel group
Participants	**Baseline characteristics** Control • *Gestational age at birth (weeks) mean (SD):* 33.50 (0.84) • *Cephalic perimeter at birth (cm) mean (SD)*: 30.17 (1.17) • *Birth weight (g) mean (SD):* 1890.83 (257.34) • *Age of mother (years) mean (SD)*: 22.50 (5.17) • *Number of children n (%)*: 1 (16.7) • *Education university n (%):* 1 (16.7) • *Social Security status paying n (%)*: 0 (0.0) • *Number of participants*: 6 Live music therapy with infants and parents in kangaroo-care • *Gestational age at birth (weeks) mean (SD):* 32.12 (1.55) • *Cephalic perimeter at birth (cm) mean (SD)*: 29.19 (1.69) • *Birth weight (g) mean (SD):* 1598.38 (388.30) • *Age of mother (years) mean (SD)*: 23.12 (6.73) • *Number of children n (%)*: 2 (25.0) • *Education university n (%)*: 3 (37.5) • *Social Security status paying n (%)*: 6 (75.0) • *Number of participants*: 8 Live music therapy with infant • *Gestational age at birth (weeks) mean (SD)*: 31.60 (2.07) • *Cephalic perimeter at birth (cm) mean (SD)*: 30.30 (1.35) • *Birth weight (g) mean (SD):* 1642.00 (284.99) • *Age of mother (years) mean (SD)*: 26.00 (7.18) • *Number of children n (%)*: 4 (80.0) • *Education university n (%)*: 2 (40.0) • *Social Security status n (%)*: 2 (40.0) • *Number of participants*: 5 Overall • *Gestational age at birth (weeks) mean (SD):* not available • *Cephalic perimeter at birth (cm)*: not available • *Birth weight (g) mean (SD):* not available • *Age of mother (years)*: not available • *Number of children n (%):* not available • *Education university n (%)*: not available • *Social Security status paying n (%)*: not available • *Number of participants*: 19 **Inclusion criteria:** signed informed consent by the parents, medically stable infants born between the 30th and 37th week of gestation (without the requirement of inotropic support or mechanical ventilation and absence of frequent episodes of apnoea), and having initiated kangaroo-care.

Ettenberger 2014 *(Continued)*

Exclusion criteria: parents who declined to participate in the study, medically unstable infants, infants who suffered from congenital malformations, and if surgery was scheduled for the week after the initial music therapy session. In the case of eligible preterm infants who were the result of multiple births (twins, triplets, etc.), all neonates were allocated to the same intervention group. To avoid potential secondary effects and reduce the possibilities of over-stimulation through the music therapy interventions, the following procedure was established: If a preterm infant showed signs of hypersensitivity towards the intervention (unusual agitated behaviour, excessive alteration of the vital signals), the music therapy intervention would be stopped. During a second intervention, a member of the medical staff would accompany the music therapist. If the same signs mentioned above occurred again and if the member of the medical staff interpreted these signs as a potential hypersensitivity towards the intervention, the preterm infant would be excluded from the study.

Pretreatment: the 3 groups did not differ significantly in terms of sociodemographic or baseline data. Just one statistically significant association was found between the groups in terms of Social Security Status (paying or beneficiary) P = 0.01.

N infants analysed: 19 (8 infants in IG1, 5 infants in IG2, and 6 infants in CG)

N infants randomised: 30

Reasons for dropouts: no

Reasons for exclusion: medical complications (n = 5), short hospitalisation (n = 5), or formal study requirements (n = 1)

Sample size calculation: N sample size calculation due to the pilot study design

Parents randomised: 27

Mothers analysed: 18 (8 mothers in intervention group 1, 5 mothers in intervention group 2, and 5 mothers in control group)

Interventions

Intervention characteristics

Control

- *Intervention type*: standard NICU environmental sounds
- *Dose*: standard NICU environmental sounds
- *Frequency*: standard NICU environmental sounds
- *Mode of delivery*: standard NICU environmental sounds
- *dB level*: standard NICU environmental sounds

Live music therapy with infants and parents in kangaroo-care

- *Intervention type*: lullabies or children's songs together with the parents, or other songs that had a positive meaning for the caregivers. If possible, the songs were chosen by the caregivers rather than proposed by the therapist. Other musical styles proposed by the parents were pop songs, sung prayers, Christian songs, or songs that caregivers invented for their babies. These were slowed down in tempo or adapted to a "lullaby style" (3/4 times, 60-80 bpm), accompanied by simple chord progressions. Normally, the music therapist accompanied the singing with guitar playing (classical guitar with nylon strings). In case a caregiver did not mention any specific music, the music therapist proposed lullabies or children's songs that are familiar to most of the population in Colombia or offered a receptive session. In one case, a mother opted to listen to improvised music while she read a children's story to her baby.
- *Dose*: 13.7 (8-25) minutes with parents
- *Frequency*: 2-4 sessions over two weeks
- *Mode of delivery*: live by trained music therapist/ parents
- *dB level*: not available

Live music therapy with infant

Ettenberger 2014 *(Continued)*

- *Intervention type*: Gentle entrained humming with instrumental accompaniment (guitar, harp, ocean drum, gato-box)
- *Dose*: 7.2 (5-10) minutes with the infant alone
- *Frequency*: 2-4 sessions over 2 weeks
- *Mode of delivery*: live by trained music therapist
- *dB level*: not available

Outcomes

Weight gain

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Unit of measure**: g per day
- **Direction**: higher is better
- **Data value**: endpoint
- **Notes**: a favourable effect towards an increased weight gain was observed for both intervention groups compared to the control group. No statistically significant differences were found for any of the parameters related to weight gain, but the weight gain per day during the intervention period for intervention group 1 was noted to be clinically significant (mean 26.39 g/day for intervention group 1 vs. 3.98 g/day for control group).

Length of hospitalisation

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Unit of measure**: days
- **Direction**: lower is better
- **Data value**: endpoint
- **Notes**: no statistically significant differences between the groups were found in terms of the total length of hospitalisation, days in the hospital from the first intervention to hospital discharge, and days from the last intervention to hospital discharge. Analysis of variance was performed by adjusting the gestational age of the three groups since the infants in both intervention groups were younger in comparison to the infants in the control group. 3.83 days shorter hospitalisation was observed for IG2 when the length of hospitalisation was adjusted for gestational age, although this difference was not statistically significant.

State-Trait-Anxiety-Inventory for Children (STAI-C, Colombian adaption from Castrillón 2005

- **Outcome type**: continuous outcome
- **Reporting**: partially reported
- **Scale**: 3-point Likert scale
- **Range**: 18 statements
- **Data value**: change from baseline
- **Notes**: handed out before the first intervention, after the last intervention, and during the follow-up. The STAI-C (adapted, shorter version of the STAI) consists of a single form with 18 statements rated on a 3-point Likert scale, which is divided into 6 factors. Two factors measure state anxiety and four factors measure trait anxiety. The raw points obtained are converted into percentiles and percentiles between 1-15 and 85-100 are reported to be clinically significant (Castrillón 2005). A statistically significant difference was found for Factor 2 before the first intervention and for Factor 6 after the last intervention.

Mother-to-infant bonding

- **Outcome type**: continuous outcome
- **Reporting**: partially reported
- **Scale**: 4-point Likert scale
- **Unit of measure**: 8 adjectives
- **Data value**: change from baseline

Ettenberger 2014 *(Continued)*

- **Notes**: the analysis of the MIBS did not show any statistically significant differences between the groups at baseline (before the first intervention) compared to the data obtained after the last intervention. MIBS was handed out before the first intervention, after the last intervention, and during the follow-up. The MIBS consists of 8 adjectives regarding how mothers feel towards their baby at the moment rated on a 4-point Likert scale (loving, resentful, neutral or felt nothing, joyful, dislike, protective, disappointed, aggressive). The MIBS were translated into Spanish by the music therapist who is fluent in both Spanish and English and was controlled by a native Spanish-speaking colleague who is fluent in English.

Heart rate and oxygen saturation: were recorded before, during, and after the intervention. However, the lack of most of the data did not allow for analysis.

Identification	**Sponsorship source:** not mentioned **Declarations of interest:** not stated **Country:** Colombia **Setting:** "Centro Policlínico del Olaya", a hospital located in one of the poorer neighbourhoods of Bogotá, the capital of Colombia. The NICU offers 28 beds distributed on two floors for both intensive and acute care and intermediate care for the more stable infants. **Comments:** not available **Authors name:** Mark Ettenberger **Institution:** Anglia Ruskin University; Universidad National Colombia **Email:** mark.ettenberger@gmx.at **Address:** Not available **Website:** www.sono.la
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomisation took place in the form of assigning the participants according to a set of randomised numbers generated by computer software.
Allocation concealment (selection bias)	Low risk	A member of the administrative staff was the only person with this set of numbers and at the time of giving written consent to participate in the study, neither the music therapist nor the participants knew to which group the participants would be assigned.
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	High risk	No blinding of outcome assessors
Incomplete outcome data (attrition bias) All outcomes	High risk	Data from just 8 out of 19 follow-up participants

Ettenberger 2014 *(Continued)*

Selective reporting (reporting bias)	Low risk	Broad variety of outcome measures were used for this study to set the path for reliable outcome measures that will be used in a larger study that is planned to start soon. However, more outcome measures add also complexity to the data recollection and analysis process.
Other bias	Unclear risk	Unclear

Farhat 2010

Study characteristics	
Methods	**Study design:** randomised controlled trial **Study grouping:** parallel group
Participants	**Baseline characteristics** Control • *Gestational age at birth (weeks) mean ± SD*: 30.5 ± 1.7 • *Birth weight (g) mean ± SD*: 1298 ± 133 • *Gender n (%) male*: 11 (48) • *Birth head circumference (cm) (g) mean ± SD*: 27.7 ± 1.5 Recorded lullaby music • *Gestational age at birth (weeks) mean ± SD*: 30.5 ± 1.7 • *Birth weight (g) mean ± SD*: 1279 +/- 172 • *Gender n (%) male*: 12 (54) • *Birth head circumference (cm) (g) mean ± SD*: 28.7 ± 1.2 **Inclusion criteria:** Adjusted gestational age younger or equal to 34 weeks; birth weight between 1000–1500 grams; postnatal age of 4 days; and clinically stable neonates (not supported by ventilator assistance or oxygen therapy) **Exclusion criteria:** Diagnosed hearing impairment; congenital anomalies; maternal drug addiction; and use of oxygen **Pretreatment:** The two groups did not differ in demographic and clinical characteristics. **N infants analysed:** 44 (22/22) **N infants randomised:** 44 (22/22) **Reasons for dropouts:** not reported **Reasons for exclusion:** not reported **Sample size calculation:** Sample size was calculated by considering the mean and standard deviation (SD) of previous studies. The sample size was calculated for all variables including SPO_2, heart rate (HR), respiratory rate (RR), and weight gain and a sample of 40 infants (20 in each group) was considered adequate.
Interventions	**Intervention characteristics** Control • *Intervention type*: routine NICU care • *Dose*: NICU standard care

Farhat 2010 *(Continued)*

- *Frequency*: NICU standard care
- *Mode of delivery*: NICU standard care
- *dB level*: NICU standard care

Recorded lullaby music

- *Intervention type*: commercially recorded lullabies sung by Iranian female vocalists; played by a Motion Pictures Expert Group player at 60–65 dB inside the isolette
- *Dose*: 20 minutes of music (10 minutes of observation before and after)
- *Frequency*: every day for 8 days
- *Mode of delivery*: via headphones inside the isolette
- *dB level*: 60-65

Outcomes	*Heart rate* • **Outcome type**: Continuous outcome • **Reporting**: Partially reported • **Unit of measure**: Beats per minute • **Direction**: Lower is better • **Data value**: Change from baseline • **Notes**: Throughout the entire study, changes in the HR in the second (intervention) and third phase (post-intervention) did not differ between the two groups (P = 0.24 and P = 0.32, respectively). *Respiratory rate* • **Outcome type**: Continuous outcome • **Reporting**: Partially reported • **Direction**: Lower is better • **Data value**: Change from baseline • **Notes**: Throughout the entire study, changes in the RR during the second phase (intervention) were different between the two groups (P = 0.017). During the third phase (post-intervention), changes in the RR were not different between the two groups (P = 0.94). The respiratory rate decreased in the control group between day 1 and day 4, followed by an increase after day 4. In the music group, RR increased on days 2, 3 and 6; it decreased on days 4 and 5, and it remained steady on days 7 and 8. *Oxygen saturation* • **Outcome type**: Continuous outcome • **Direction**: Higher is better • **Data value**: Change from baseline • **Notes**: For the duration of the intervention, the changes in SPO_2 in the intervention period were higher in the music group than in the control group (P = 0.001). Post-intervention, the changes in the music group were still significant when compared to the control group (P = 0.019). When compared to the control group, more stable SPO_2 (less fluctuation) was observed in the music group on days 7 and 8 (P = 0.05 and P = 0.022 respectively). *Weight gain* • **Outcome type**: Continuous outcome • **Reporting**: Partially reported • **Direction**: Higher is better • **Data value**: Change from baseline • **Notes**: Daily weight gain did not differ between the music and control groups (36.79 ± 31.83 and 18.63 ± 39.05, respectively, P = 0.093).
Identification	**Sponsorship source:** Mashhad University of Medical Science **Declarations of interest:** No financial interested to disclose

Farhat 2010 *(Continued)*

	Country: Iran
	Setting: NICU
	Comments: no comments
	Authors name: Rana Amiri
	Institution: Neonatal Research Center, NICU, Emamreza Hospital, Mashhad, Iran
	Email: amirir1@mums.ac.ir
	Address: Neonatal Research Center, NICU, Emamreza Hospital, Mashhad, Iran
	Website: Not available
	Tel: Tel.: +98 5112709006
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	No sequence generation described
Allocation concealment (selection bias)	Unclear risk	Unclear if allocation concealment happened
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	High risk	No blinding of outcome assessment
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not described
Selective reporting (reporting bias)	Unclear risk	Unclear
Other bias	Unclear risk	Unclear

Haslbeck 2021

Study characteristics

Methods	**Study design:** Randomised controlled trial
	Study grouping: Parallel group
Participants	**Baseline characteristics**
	Control

Haslbeck 2021 *(Continued)*

- *Gestational age at birth (weeks) mean ± SD:* 27.8 ± 2.3
- *Gestational age at intervention (weeks) mean ± SD:* 25.3 ± 2.6
- *Birth weight Z-score (g) mean ± SD*: 1063 ± 358
- *Gender n (%) female*: 21 (47)
- *Z-score of birth weight (g) mean ± SD*: 1063 ± 358
- *Oxygen supplementation (days), mean ± SD*: 35.9 ± 29.5
- *Birth head circumference Z-score (cm) (g) mean ± SD*: 5.3 ± 2.6
- *Mechanically ventilation (days) mean ± SD*: 3.1 ± 4.0
- *Parental socioeconomic score (range: 2–12), mean ± SD*: 5.5 ± 2.6
- *Small for gestational age (i.e. birth weight < 10 percentile), n (%)*: 3 (6.7)
- *Retinopathy of prematurity, n (%)*: 3 (6.7)
- *Sepsis, n (%)*: 4 (9)
- *Bronchopulmonary dysplasia, n (%)*: 8 (9)
- *Intraventricular haemorrhage, n (%)*: 7 (8)
- *Duration of hospitalisation (days), mean ± SD*: 57.4 ± 31.4

Family-centred live creative music therapy

- *Gestational age at birth (weeks) mean ± SD*: 28.1 ± 1.8
- *Gestational age at intervention (weeks) mean ± SD*: 25.8 ± 2.0
- *Birth weight Z-score (g) mean ± SD*: 1060 ± 287
- *Gender n (%) female*: 13 (35)
- *Z-score of birth weight (g) mean ± SD*: 1060 ± 287
- *Oxygen supplementation (days), mean ± SD*: 42.0 ± 33.8
- *Birth head circumference Z-score (cm) (g) mean ± SD*: 25.8 ± 2.0
- *Mechanically ventilation (days) mean ± SD*: 2.6 ± 4.4
- *Parental socioeconomic score (range: 2–12), mean ± SD*: 5.3 ± 2.4
- *Small for gestational age (i.e. birth weight < 10 percentile), n (%)*: 3 (8.1)
- *Retinopathy of prematurity, n (%)*: 0 (0)
- *Sepsis, n (%)*: 6 (16)
- *Bronchopulmonary dysplasia, n (%)*: 7 (8)
- *Intraventricular haemorrhage, n (%)*: 4 (5)
- *Duration of hospitalisation (days), mean ± SD*: 59.1 ± 27.9

Inclusion criteria: Infants were eligible if they were born prematurely (gestational age < 32 weeks), reached the chronological age ≥ seven days of life, and did not require invasive cardiovascular and respiratory support at recruitment.

Exclusion criteria: Diagnosed with a major congenital anomaly/genetically defined syndrome, congenital malformation adversely affecting life expectancy/neurodevelopment, intraventricular haemorrhage (≥ Volpe grade 3), or when the infants were admitted for palliative care

Pretreatment: Groups neither differed in neonatal parameters nor baseline demographic characteristics. Compared to infants lost to follow-up, infants assessed at age 2 years had similar baseline characteristics except for lower gestational age (P < 0.001), longer supplemental oxygen therapy (P = 0.001), and longer hospital stay (P < 0.001).

N infants analysed: 56

N infants randomised: 82

Reasons for dropouts: 26 dropouts (n = 13 in both the CMT and standard care groups) discontinuation resulted from moving abroad (n = 2) or for unknown reasons (n = 24).

Reasons for exclusion: reasons for exclusion for per-protocol analysis: not received CMT stimulation

Haslbeck 2021 *(Continued)*

	Sample size calculation: The sample size for this pilot trial was set at 60 participants per pilot trial design recommendations (40–42).
Interventions	**Intervention characteristics** Control • *Intervention type*: standard care • *Dose*: n.a. • *Frequency*: n.a. • *Mode of delivery*: n.a. • *dB level*: n.a. Family-centred live creative music therapy • *Intervention type*: During CMT (infant only), the music therapist started the session by touching the infants, e.g. at the head and feet, to "welcome" the infant and entrain the breathing rhythm. After observation, the music therapist would hum smoothly and tailor vocalisations in line with the infant's breathing rhythm, facial expression, and gesture. Throughout the session, a melody progressively develops according to the infant behavioural state. For example, the therapist would use sedating musical parameters (e.g. calm, repetitive humming) to soothe an agitated infant. Conversely, the therapist would use activating musical parameters (up-rising melodies) to stimulate a limp infant. In all cases, parental musical preferences would be integrated into the improvisation – see clinical CMT protocol for details. At the end of the session, the music therapist would smoothly fade out the humming smoothly and gently remove her hands from the infant. • *Dose*: 20 minutes • *Frequency*: 2-3 per week until discharge • *Mode of delivery*: live at the bedside or with parents in kangaroo-care • *dB level*: not standardised
Outcomes	*Bayley Scale cognitive composition score* • **Outcome type**: Continuous outcome • **Reporting**: Fully reported • **Range**: 0-200 • **Direction**: Higher is better • **Data value**: Endpoint • **Notes**: The cognitive composite score includes evaluation of sensorimotor development, object relatedness, memory, exploration, manipulation, concept formation, and memory. *Bayley's language composition score* • **Outcome type**: Continuous outcome • **Reporting**: Fully reported • **Range**: 0-200 • **Direction**: Higher is better • **Data value**: Endpoint *Bayley motor composition score* • **Outcome type**: Continuous outcome • **Reporting**: Fully reported • **Range**: 0-200 • **Direction**: Higher is better • **Data value**: Endpoint
Identification	**Sponsorship source:** Foundation of Anna Müller Grocholski; Stiftung für Neonatologie; Vontobel Foundation

Haslbeck 2021 *(Continued)*

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

Country: Switzerland

Setting: NICU III level

Comments:

Authors name: Friederike B Haslbeck

Institution: Department of Neonatology, Newborn Research Zurich, University Hospital Zurich, Zurich, Switzerland

Email: friederike.haslbeck@usz.ch

Address: Frauenklinikstrasse 10, 8091 Zurich

Website: neonatologie.usz.ch

Notes

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random sequence generation reported
Allocation concealment (selection bias)	Low risk	Allocation concealment reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Blinding of outcome assessment reported
Incomplete outcome data (attrition bias) All outcomes	Low risk	No incomplete outcome data
Selective reporting (reporting bias)	Low risk	No selective reporting detected
Other bias	Low risk	No other bias detected

Jabraeili 2016

Study characteristics

Methods

Study design: Randomised controlled trial

Study grouping: Parallel group

Participants

Baseline characteristics

Jabraeili 2016 *(Continued)*

Control

- *Gestational age at birth (weeks) mean (SD):* 31.80 (1.83)
- *Postnatal age at intervention in days*: 3 days
- *Birth weight (g) mean (SD):* 690.47 (436.88)
- *Gender n (%) male*: 7 (33.3)
- *Apgar score, 1 minute mean (SD)*: 7.47 (1.53)
- *Apgar score, 5 minutes mean (SD)*: 8.90 (0.94)
- *Number*: 20

Recorded mum's lullaby

- *Gestational age at birth (weeks) mean (SD)*: 31.45 (1.63)
- *Postnatal age at intervention in days*: 3 days
- *Birth weight (g) mean (SD)*: 792.50 (464.21)
- *Gender n (%) male*: 10 (50)
- *Apgar score, 1 minute mean (SD)*: 7.75 (.91)
- *Apgar score, 5 minutes mean (SD)*: 9.05 (0.75)
- *Number*: 21

Recorded Brahms` lullaby

- *Gestational age at birth (weeks) mean (SD)*: 32.16 (1.31)
- *Postnatal age at intervention in days*: 3 days
- *Birth weight (g) mean (SD)*: 762.40 (390.01)
- *Gender n (%) male*: 16 (64.0)
- *Apgar score, 1 minute mean (SD)*: 7.64 (1.22)
- *Apgar score, 5 minutes mean (SD)*: 9.00 (0.91)
- *Number*: 25

Inclusion criteria: 29-34 weeks of gestation, postnatal age of 3 days, weight under 2800 grams

Exclusion criteria: Congenital heart diseases and acute infections need for ventilators, receiving oxygen through continuous positive airway pressure, or receiving hypnotics as well as sedative drugs and traditional treatments during the intervention were excluded from the study.

Pretreatment: No significant difference between the three groups in baseline data including infant gender, birth weight, weight in intervention days, gestational age, postnatal age, 1-minute, and 5-minute Apgar scores, mother's age, and neonate's race

N infants analysed: 66 (Mom's Lullaby Group = 21 + Brahms Lullaby Group = 25 + Control Group = 20)

N infants randomised: 75

Reasons for dropouts: Five infants in the mum's lullaby group, and 4 infants in the control group were excluded. Therefore, the study was completed with 66 premature infants.

Reasons for exclusion: 3 infants were excluded because their parents refused to take part in the study.

Sample size calculation: Reviewing a meta-analysis and previously published studies that evaluated the effects of patterned sound in the NICU on preterm infants and utilising "Gpower" software, the sample size was determined considering quantitative variables based on (effect size: 0.42, α = 0.05, power: 0.8). This calculation referred to 20 neonates for each group. Considering an attrition rate of 10-15% and also the Automated Auditory Brainstem Response results, a total of 25 subjects were included in each of the three groups.

Interventions	**Intervention characteristics** Recorded mum's lullaby

Jabraeili 2016 *(Continued)*

- *Intervention type*: Maternal voice in this study refers to a taped recording of a 15-minute session of maternal voice modified in low 65-70 dB, 1000 hertz. The correctness of intensity and frequency was checked and determined by the researcher. A CD recording of the mother reciting a folkloric (Turkish or Persian lullaby) based on the dominant mum's language was used in the experimental component of this study. The researcher recorded mum's sound in a quiet place by means of a digital recorder. The recording material was copied onto a CD and then transferred to an iPod touch and saved as an Mp3 file. The verbalised lullaby took approximately five minutes to recite and repeated three times and was not a common verse.
- *Dose*: 15 mins
- *Frequency*: 3 consecutive days
- *Mode of delivery*: via earphones
- *dB level*: 65-70 dB

Recorded Brahms` lullaby

- *Intervention type*: Brahms' lullaby in this study refers to a taped recording of a 15-minute session of Brahms' lullaby in low 65 dB, 1000 Hertz that as a sound file was kept and saved on an iPod; the infants in the control group received routine auditory stimulation.
- *Dose*: 15 minutes (10 before, 15 music, 20 after = 45 minutes data collection)
- *Frequency*: 3 consecutive days
- *Mode of delivery*: via earphones
- *dB level*: 65 dB

Outcomes

Oxygen saturation

- **Outcome type**: Continuous outcome
- **Direction**: Higher is better
- **Notes**: Neonate oxygen saturation did not change significantly in 10 minutes before the sound (P = 0.63) or during the 15 minutes of sound (P = 0.08); however, neonate oxygen saturation increased significantly in the 15 minutes after providing the sound conditions (P = 0.02). At the 45 minutes point, there was a significant difference in oxygen saturation levels as a function of Brahms' lullaby.

Identification

Sponsorship source: Tabriz University of Medical Science (Project number: 359)

Declaration of interest: The authors declared no conflict of interest in this study.

Country: Iran

Setting: NICU

Comments: no comments

Authors name: Tahmineh Sabet (MSc)

Institution: Department of Pediatric Nursing, Faculty of Nursing and Midwifery, Tabriz University of Medical Sciences, Tabriz, Iran

Email: sabet@sums.ac.ir

Address: Department of Pediatrics, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran

Website: not available

Notes

Risk of bias

Bias	Authors' judgement	Support for judgement

Jabraeili 2016 *(Continued)*

Random sequence generation (selection bias)	Low risk	Mum's lullaby, and control groups. Randomised block design by a random assignment software was used for producing random numbers.
Allocation concealment (selection bias)	Unclear risk	Unclear
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Double-blind
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Double-blind
Incomplete outcome data (attrition bias) All outcomes	Low risk	5 infants in the mum's lullaby group, and 4 infants in the control group were excluded. Therefore, the study completed with 66 premature infants.
Selective reporting (reporting bias)	Unclear risk	Unclear
Other bias	Unclear risk	Unclear

Johnston 2007

Study characteristics	
Methods	**Study design:** randomised controlled trial **Study grouping:** cross-over
Participants	**Baseline characteristics** Control • *Gestational age at birth (weeks): n.a.* • *Postnatal age at intervention (days) mean (SD)*: 6.6 days (2.5) • *Birth weight (g): n.a.* • *Gender: n.a.* • *Apgar score at 5 minutes: n.a.* Recorded mothers voice • *Gestational age at birth (weeks): n.a.* • *Postnatal age at intervention (days) mean (SD)*: 6.9 days (3.1) • *Birth weight (g): n.a.* • *Gender: n.a.* • *Apgar score at 5 minutes: n.a.* Overall • *Gestational age at birth (weeks) mean (SD):* 33.1 (0.89) • *Postnatal age at intervention (days): n.a.* • *Birth weight (g) mean (SD):* 1985 (312) • *Gender n (%) male*: 57

Johnston 2007 *(Continued)*

- *Apgar score at 5 minutes mean (SD)*: 8.5 (1.1)

Inclusion criteria: between 32 weeks 0 days and 35 weeks 6 days gestational age (GA) according to ultrasound or last menstrual period. All infants had Apgar scores at 5 minutes greater than 3.

Exclusion criteria: congenital anomalies, intraventricular haemorrhage (IVH) greater than a grade 2, periventricular leukomalacia (PVL), requiring surgery, receiving analgesic therapy

N infants analysed: 20

N infants randomised: 65

Reasons for dropouts: 19 had no blood work ordered and another 7 did not have a second session in the time they were eligible to actually participate in the study before being discharged from the NICU. There was equipment failure with either the camera or the pulse oximeter for 19, leaving a final sample of 20 with completed data sets at 2 points in time.

Reasons for exclusion: not reported

Sample size calculation: sample size estimates based on previous studies with differences in PIPP scores of 2 points and a standard deviation of 2 points, setting alpha at 0.05 and a power of 0.80, indicated 44 infants would be needed for separate groups, 22 for a repeated measures design; for facial actions of 22% and standard deviation of 30%, a sample size of 46 infants in a randomised controlled trial but only 20 in a repeated measures design would be indicated. Using heart rate differences of 15 with a standard deviation of 17 to 20 (from previous studies), a sample size of 27 was needed for a repeated measures design. Thus, a conservative sample of 27 was sought, allowing for a 20% loss.

Interventions

Intervention characteristics

Recorded mothers voice

- *Intervention type*: recording of mothers voice talking to her baby, either singing, saying nursery rhymes, or talking "baby talk" in a sing-song voice, filtered through a music editing software program (CoolEdit, AdobeSystems Inc., San Jose, Calif.), in which it was low-pass filtered with an attenuation of −96 decibels (dB)/octave above the normal cut-off frequency of 500 hertz that emulated the effect of the mother's voice travelling through amniotic fluid. This filtered talk was made into a recording, which was repeated to constitute 10 minutes duration.
- *Dose*: 10 minutes
- *Frequency*: 3 times a day during quiet periods for 28 hours. After 48 hours, the next instance in which the infant required blood sampling via heel lance for clinical purposes was done according to the predetermined random assignment. With at least 24 hours between sessions, the subsequent heel lance was done again according to the random assignment.
- *Mode of delivery*: this recording was then played to the infant at other sites on a small cassette player with small speakers placed in the isolette.
- *dB level*: 60-65 dB

Outcomes

Heart rate

- **Outcome type**: continuous outcome
- **Reporting**: partially reported
- **Unit of measure**: bpm
- **Direction**: lower is better
- **Data value**: endpoint
- **Notes**: mean heart rate was 159 bpm for the control condition and 162 bpm for the voice condition. For heart rate, the differences were consistent in that in the second session there were significantly higher heart rate levels across many phases, including baseline and warming. In fact, after the procedure, the heart rate was higher in the voice condition than in the control condition.

Oxygen saturation

- **Outcome type**: continuous outcome

Johnston 2007 *(Continued)*

- **Notes**: the Apgar score at 5 minutes was a significant covariate for oxygen saturation levels, with babies with higher Apgar scores having higher oxygen saturation levels during all phases and across both conditions, with Pearson correlations ranging from 0.485 to 0.584. After the procedure, oxygen saturation was lower in the voice than in the control condition.

PIPP

- **Outcome type**: continuous outcome
- **Reporting**: partially reported
- **Data value**: endpoint
- **Notes**: there were no differences between the 2 conditions of voice or control on the composite score of the PIPP or of its individual components at any point in time, other than there was a lower oxygen saturation in the voice condition during the final phase after the heel lance before return to baseline (94.1 vs 96.2, P < 0.01). The means were almost identical, for example, PIPP scores at 30 seconds following lance were 10.15 for voice and 10.95 for control, facial actions at the 30 seconds from heel lance were 52% for voice condition and 56% for the control condition.

Identification	**Sponsorship source:** Canadian Institutes of Health Research MOP-38074 and in part by McGill University James McGill Chair Fund **Declaration of interest:** not reported **Country:** Canada **Setting:** 2 university-affiliated NICUs were the sites for this study. Both were level 3 NICUs. **Comments:** no comment **Authors name:** Celeste Johnston **Institution:** School of Nursing, McGill University, Montreal, Quebec, Canada H3A 2A7 **Email:** celeste.johnston@mcgill.ca **Address:** School of Nursing, McGill University, 3506 University Street, Montreal, Quebec, Canada H3A 2A7 **Website:** http://childkindinternational.org/celeste-johnston-rn-phd
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	A computer-generated list of random ordering for the 2 conditions of 20 infants was produced for each site before recruitment began.
Allocation concealment (selection bias)	Unclear risk	At the time of enrolment, that is, at the moment the mother gave consent, the infant was assigned the next sequenced study number on the list, which then determined the first session.
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Faces were scored by research assistants who did not know the purpose of the study.

Johnston 2007 *(Continued)*

Incomplete outcome data (attrition bias) All outcomes	High risk	65 consented; 20 complete data
Selective reporting (reporting bias)	Unclear risk	No study protocol or trial registration found
Other bias	Low risk	A washout period of at least 24 hours lessened the possibility of a cumulative effect from one condition to the next.

Kehl 2021

Study characteristics	
Methods	**Study design:** randomised controlled trial **Study grouping:** parallel group **Qualitative data:** parental interview on parental experience of music therapy
Participants	**Baseline characteristics** Control • *Maternal age: median (range)*: 30.00 (27.00–35.00) • *Married (% (n))*: 58 (n = 7) • *Cohabitation (% (n))*: 100 (n = 12) • *Maternal nationality Swiss (% (n))*: 50 (n = 3) • *Maternal educational qualification: apprenticeship (% (n))*: 50 (n = 3) • *Maternal educational qualification: University of Applied Sciences degree (% (n))*: • *Maternal educational qualification: university degree (% (n))*: 50 (n = 3) • *Paternal nationality: Swiss (% (n))*: 67 (n = 4) • *Paternal language: German (% (n)*: 100 (n = 6) • *Paternal educational qualification: apprenticeship (% (n)):* 50 (n = 3) • *Primigravida (% (n))*: 67 (n = 4) • *Primiparous (% (n))*: 67 (n = 4) • *Caesarean section (% (n))*: 100 (n = 6) • *Male infants (% (n))*: 50 (n = 4) • *Gestational age at birth (weeks) (median (range))*: 28.29 (25.00–29.00) • *Birth weight (g) (median (range))*: 785.00 (600–1310) • *Birth size (cm) (median (range))*: 34.50 (32.00–39.0) • *Apgar score (10 min) (median (range))*: 9 (6–10) Live family-centred Creative Music Therapy • *Maternal age: median (range)*: 31.50 (26.00–43.00) • *Married (% (n))*: 50 (n = 10) • *Cohabitation (% (n))*: 100 (n = 20) • *Maternal nationality Swiss (% (n))*: 90 (n = 9) • *Maternal educational qualification: apprenticeship (% (n))*: 40 (n = 4) • *Maternal educational qualification: University of Applied Sciences degree (% (n))*: 30 (n = 3) • *Maternal educational qualification: university degree (% (n))*: 30 (n = 3) • *Paternal nationality: Swiss (% (n))*: 70 (n = 7) • *Paternal language: German (% (n)*: 90 (n = 9)

Kehl 2021 *(Continued)*

- *Paternal educational qualification: apprenticeship (% (n)):* 60 (n = 6)
- *Primigravida (% (n))*: 80 (n = 8)
- *Primiparous (% (n))*: 80 (n = 8)
- *Caesarean section (% (n))*: 100 (n = 10)
- *Male infants (% (n))*: 82 (n = 9)
- *Gestational age at birth (weeks) (median (range))*: 28.43 (24.86–31.43)
- *Birth weight (g) (median (range))*: 1070 (680–1590)
- *Birth size (cm) (median (range))*: 38.00 (33.00–43.00)
- *Apgar score (10 min) (median (range))*: 10 (7–10)

Inclusion criteria: the participants were mothers and fathers of prematurely born infants recruited after birth at the Department of Neonatology of the University Hospital Zurich, Switzerland. Parents of clinically stable infants born before 32 weeks of gestation and at a chronological age of 7 days of life were eligible for study participation.

Exclusion criteria: exclusion criteria were parents of premature infants with a genetically defined syndrome, severe congenital malformation adversely affecting life expectancy and/or neurodevelopment, high-grade intraventricular haemorrhage, and/or cystic white matter lesions. Parents of premature infants admitted to palliative care were also excluded from the study. Lack of sufficient knowledge of the German language was a further defined exclusion criterion.

Pretreatment: no significant differences in sociodemographic, pregnancy- and birth-related characteristics (P > 0.05 for all variables)

N parents analysed: 32

N parents randomised: 46

Reasons for dropouts: relocation to another hospital

Reasons for exclusion: not reported

Sample size calculation: no sample size calculation due to the pilot design

Interventions

Intervention characteristics

Control

- *Intervention type*: standard care

Live family-centred Creative Music Therapy

- *Intervention type*: in the presence of the parents (or 1 parent) who were in skin-to-skin contact with their premature infant. If the parents could not be present (due to a conflicting schedule, such as attending a meeting with the paediatricians and nurses or caring for a sibling at home), the therapeutic session was conducted at the bedside with the premature infant alone, using initial and therapeutic touch as well as the music therapist's voice. After a period of observation, the music therapist started to hum or sing according to the breathing rhythm, facial expression, and gestures of the premature infant. Whenever the parents could participate in the therapeutic sessions, a vibroacoustic monochord was used to accompany the singing. It was placed on 1 parent's elbow so that the vibrations could be felt. A monochord is a wooden, single-stringed instrument manufactured for therapeutic purposes. When touched, its vibroacoustic stimulation can serve to relax or even stimulate. During the sessions, the parents were involved in the therapy and decided each time, whether they wanted to relax, participate in singing or just observe their child's reaction to the music. Before the therapy sequence started, the parents were asked about their musical wishes, so their musical and cultural backgrounds became part became of the therapy and the sessions were shaped by the music therapist and the parents together. As a previous study has shown, parents often started to use their own voice during the therapeutic process, became more confident in this respect, and aware of how important their singing, humming, or even speaking is for their child. After the session and when appropriate, the music therapist provided explanations and feedback to the parents and shared her perceptions of the preterm infant's reaction to the music. This was done in order to support the parent–infant interaction and to strengthen the parents in their parental role.

Kehl 2021 *(Continued)*

- *Dose*: 20 minutes
- *Frequency*: 2-3 x per week
- *Mode of delivery*: live, infant-directed, parents integrated
- *dB level*: live, individualised

Outcomes

State-Trait Anxiety Inventory (STAI-t)

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Scale**: 4
- **Direction**: lower is better
- **Data value**: endpoint
- **Notes**: the 2 groups did not differ significantly in trait anxiety levels at time point 1, time point 2, and time point 3 (P > 0.05 for all time points). Concerning the course of trait anxiety across time, the reduction from time point 2 to time point 3 failed to reach the conservative significance level in the music group. In the control group, there were no significant changes, neither from time point 1 to time point 2 nor from time point 2 to time point 3.

State-Anxiety Inventory (STAI-SKD)

- **Outcome type**: continuous outcome
- **Scale**: 4
- **Direction**: lower is better
- **Data value**: endpoint
- **Notes**: the 2 groups did not differ significantly in state anxiety levels for all time points. However, there was a significant reduction in state anxiety levels from time point 1 to time point 2 within the music therapy group, while in the control group, there were no significant changes.

Depressive symptoms (EPDS)

- **Outcome type**: continuous outcome
- **Reporting**: partially reported
- **Data value**: endpoint
- **Notes**: At time point 1, the extent of depressive symptoms (EPDS) in the music group was not significantly different from the extent of depressive symptoms in the control group. There were no significant differences between the two groups at time point 2 and time point 3 either (P > 0.05 for all). However, within the music group, there was a significant reduction in depressive symptoms from time point 2 to time point 3. There were no significant changes across time in the control group.

Parental Stressor Scale: Neonatal Intensive Care Unit (PSS: NICU)

- **Outcome type**: continuous outcome
- **Scale**: Likert scale
- **Range**: 5
- **Direction**: lower is better
- **Data value**: endpoint
- **Notes**: there were no significant differences between the 2 groups at time point 1 and time point 2 regarding the Sights and Sounds subscale of the PSS: NICU (P > 0.05 for both). Within the music therapy group, there were no significant differences from time point 1 to time point 2. However, within the control group, parental stress regarding the Sights and Sounds subscale increased from time point 1 to time point 2.

Pictorial Representation of Attachment Measure (PRAM)

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Unit of measure**: cm
- **Direction**: lower is better

Kehl 2021 *(Continued)*

- **Data value**: endpoint
- **Notes**: there was no significant difference between the 2 groups at time point 1 regarding the PRAM Self–Baby Distance. However, at time point 2, the music group exhibited a significantly shorter distance compared to the control group. At time point 3, there were no significant group differences any more. Within the music group, the PRAM Self–Baby Distance decreased significantly from time point 1 to time point 2. There were no differences in the PRAM Self–Baby Distance from time point 2 to time point 3. There were no significant differences within the control group across all time points.Concerning the PRAM Self–Baby Emotional Distance, again there were no significant differences between the 2 groups at time point 1 and time point 3 (P > 0.05). Similarly, at time point 2, the MTG did not exhibit a significantly shorter distance. There were no significant differences corrected from time point 1 to time point 2 and from time point 2 to time point 3 within both groups. At time point 1, the music therapy group displayed a significantly shorter distance in the PRAM Self–Baby Physical Distance. There were no significant group differences at time point 2 and 3 (P > 0.05). However, in the music therapy group, there was a significant decrease in distance from time point 1 (Mdn = 4.20 cm) to time point 2 (Mdn = 3.50 cm), z = −2.36, P = 0.016, r = −0.399, and again from T2 to T3 (Mdn = 2.80 cm), z = −2.34, P = 0.016, r = −0.435, at a Bonferroni corrected P value of 0.025. No such changes across time were apparent in the control group.

Identification	**Sponsorship source:** Anna Müller Grocholski Stiftung für Neonatologie; Vontobel Foundation **Declaration of interest:** the authors declared no conflict of interest. **Country:** Switzerland **Setting:** NICU Level III **Comments:** **Authors name:** Selina M. Kehl **Institution:** Department of Neonatology, University Hospital Zurich **Email:** Selina.kehl@kispi.uzh.ch, corresponding author: Friederike Haslbeck friederike.haslbeck@usz.ch **Address:** Frauenklinikstrasse 10, 8091 Zurich **Website:**
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random sequence generation reported
Allocation concealment (selection bias)	Low risk	Allocation concealment reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Blinding of outcome assessment reported

Kehl 2021 *(Continued)*

Incomplete outcome data (attrition bias) All outcomes	Low risk	No incomplete outcome data detected
Selective reporting (reporting bias)	Low risk	No selective reporting detected
Other bias	Low risk	No other bias detected

Kraft 2021

Study characteristics	
Methods	**Study design:** randomised controlled trial **Study grouping:** cross-over
Participants	**Baseline characteristics** Control • *Gestational age at birth (weeks) median (range):* 26 (25–28) • *Birth weight (g) median (range):* 810 (734–965) • *Gender n (%) male*: 3 (30) • *Mechanically ventilated n (%)*: 9 (90) • *Sepsis, culture-proven n (%)*: 3 (30) • *IVH grade III + IV n (%)*: 0 (0) • *PVL grade I + II n (%)*: 0 (0) • *First pregnancy n (%)*: 6 (60) • *Previously premature childbirth (< 32 weeks of gestation) n (%)*: 0 (0) • *Social and/or psychological problems n (%)*: 1 (10) Live RBL music therapy • *Gestational age at birth (weeks) median (range):* 26 (25–28) • *Birth weight (g) median (range):* 900 (703–1045) • *Gender n (%) male*: 6 (46) • *Mechanically ventilated n (%)*: 9 (69) • *Sepsis, culture-proven n (%)*: 2 (15) • *IVH grade III + IV n (%)*: 3 (23) • *PVL grade I + II n (%)*: 1 (8) • *First pregnancy n (%)*: 6 (46) • *Previously premature childbirth (< 32 weeks of gestation) n (%)*: 3 (23) • *Social and/or psychological problems n (%)*: 0 (0) **Inclusion criteria:** parents of infants born with a gestational age of fewer than 30 weeks **Exclusion criteria:** infants whose parents did not speak or understand Dutch were not eligible for inclusion. **Pretreatment:** characteristics of the infants in both groups were comparable, except for a significantly higher number of infants with haemodynamically significant persistent ductus arteriosus in the waiting-list group. 2 infants in the music therapy group died during the study period, including a female twin. **N infants analysed:** 21

Kraft 2021 *(Continued)*

N infants randomised: 59

Reasons for dropouts: not reported

Reasons for exclusion: questionnaires not completed by parents

Sample size calculation: not reported

Interventions	**Intervention characteristics** Live music therapy • *Intervention type*: live music therapy tailored to their behavioural state. Infants in quiet sleep received calmly played music based on entrainment to their breathing patterns. If they were in active sleep, quiet wakefulness, or active states, the improvisation focused on muscle tension and breathing patterns. The musical interventions comprised an acoustic guitar, OceanDisc, and voice. These were based on the 'rhythm, breath, and lullaby' method. An Ocean Disc is a round drum with metal balls that produces sounds as the therapist moves. The Ocean Disc was the primary instrument in the first 2 sessions. When the guitar was used, the therapist used a song chosen by the parents that were converted into a lullaby (song-of-kin), or the lullaby "Twinkle Twinkle Little Star". From the third session onwards, song-of-kin was incorporated into the sessions. The music therapist specifically searched for signs of overstimulation in infants. Parents were actively involved and received extensive information on music therapy before the first session. They were mainly present during the sessions and were provided with their song-of-kin if they had chosen one. • *Dose*: 15 minutes • *Frequency*: 6 x • *Mode of delivery*: live by a trained music therapist • *dB level*: not reported
Outcomes	*State-Anxiety Inventory (STAI-SKD)* • **Outcome type**: continuous outcome • **Reporting**: partially reported • **Scale**: 4-point Likert scale • **Range**: 20-80 • **Direction**: lower is better • **Data value**: change from baseline • **Notes**: at the last time of inclusion, the mean (SD) state and trait anxiety were comparable in both groups (state anxiety 45.2 (9.3) vs. 40.9 (9.2), P = 0.31, Likert and trait anxiety 39.3 (6.1) vs. 38.7 (9.4), P = 0.85 for music therapy group and waiting-list groups, respectively) *State-Trait Anxiety Inventory (STAI-T)* • **Outcome type**: continuous outcome • **Reporting**: partially reported • **Scale**: 4-point Likert Scale • **Range**: 20-80 • **Direction**: lower is better • **Data value**: change from baseline • **Notes**: trait anxiety 39.3 (6.1) vs. 38.7 (9.4), P = 0.85 for LPMT and waiting-list groups, respectively *Neonatal Infant Stressor Scale (NISS)* • **Outcome type**: continuous outcome • **Reporting**: partially reported • **Range**: 2-5 • **Direction**: lower is better • **Data value**: endpoint

Kraft 2021 *(Continued)*

- **Notes**: the Neonatal Infant Stressor Scale (NISS), a validated quantitative measure of neonatal stress, was used to measure stress exposure in infants. The NISS score was developed to quantify the stressful experiences of preterm neonates. It assigns points to various care practices felt by the neonatal care team to be stressful and classifies these care practices and specific medical events as "acute" or "chronic" stressors. Events are rated on a scale of 2 to 5, in which 2 is a little stressful, 3 implies moderately stressful, 4 is very stressful, and a score of 5 implies an extremely stressful event. For example, for acute items, nappy changes are considered to be moderately stressful, 1 attempt at gaining intravenous access is very stressful, and multiple attempts to gain intravenous access are extremely stressful. Regarding chronic items, having a nasogastric tube in situ is considered as being a little stressful, fasting for surgery is moderately stressful, and having a systemic infection is very stressful. The NISS does not measure stress experienced by infants, but recently, cumulative stress exposure has been associated with cortisol measurements, which have proven to be one of the most precise physiological signs of reactions to stress. Therefore, we believe that the NISS is a valid measure of infant stress. During NICU admission, NISS data were abstracted from the medical records of the infants, and the mean NISS score was calculated daily. In the STRONG study, an average NISS score was calculated over days 1 (day of birth) to 14. In the LPMT trial, the average NISS scores were calculated over the total days of the music therapy group or waiting-list period. In both groups, the daily NISS scores declined gradually over time, which did not significantly differ between the music therapy and waiting-list group.

Identification	**Sponsorship source:** Junior Scientific Masterclass of the University of Groningen. **Declaration of interest:** the authors declared no conflict of interest. **Country:** the Netherlands **Setting:** NICU Level III **Comments:** **Authors name:** Karianne Kraft **Institution:** Division of Neonatology, Department of Pediatrics, Beatrix Children's Hospital, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713GZ Groningen, The Netherlands **Email:** k.e.kraft@umcg.nl **Address:** Beatrix Children's Hospital, University MedicalCenter Groningen, University of Groningen, Hanzeplein 1, 9713GZ Groningen, The Netherlands **Website:** **Tel:** +31-50-361-4215
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Judgement comment: assignment to treatment allocation was done through a web portal hosted by the study data centre at the TCC, at the UMCG. The randomisation schedule was computer-generated, using the method of randomised blocks. This trial was thereby protected from selection bias by using concealed randomisation.
Allocation concealment (selection bias)	Low risk	Judgement comment: assignment to treatment allocation was done through a web portal hosted by the study data centre at the TCC, at the UMCG. The randomisation schedule was computer-generated, using the method of randomised blocks. This trial was thereby protected from selection bias by using concealed randomisation.

Kraft 2021 *(Continued)*

Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	High risk	Judgement comment: outcome assessors were not blinded. They knew which children were in the music therapy group and which weren't. Parents were not blinded as well.
Incomplete outcome data (attrition bias) All outcomes	High risk	The low response after randomisation could have caused bias. Stressed parents may not have felt the opportunity to complete the questionnaires, as they were occupied with anxiety and worries about their infant's well-being.
Selective reporting (reporting bias)	Low risk	No selective reporting detected
Other bias	High risk	We only collected limited background information on the mothers. It would have been interesting to investigate the relation of, e.g. age, maternal socioeconomic status, and marital status on anxiety as well.

Kucuk Alemdar 2020

Study characteristics	
Methods	**Study design:** randomised controlled trial **Study grouping:** parallel group
Participants	**Baseline characteristics** Control • *Gestational age at birth (weeks) mean ± SD :* 30.25 ± 0.50 • *Birth weight (g) mean ± SD*:: 1503.80 ± 194.86 • *Gender n (%) male*: 15 (46.9) • *Birth head circumference (cm) (g) mean ± SD*: 29.28 ± 0.85 • *Neonatal Therapeutic Intervention Scoring System mean ± SD*: 32.66 ± 6.88 Recorded mothers spoken voice • *Gestational age at birth (weeks) mean ± SD*: 30.06 ± 0.63 • *Birth weight (g) mean ± SD*: 1460.50 ± 133.36 • *Gender n (%) male*: 17 (56.7) • *Birth head circumference (cm) (g) mean ± SD*: 29.46 ± 0.81 • *Neonatal Therapeutic Intervention Scoring System mean ± SD*: 32.19 ± 7.81 **Inclusion criteria:** being born between the 30th and 34th week of gestation, having a birth weight > 1000 g, having a cardiorespiratory condition after admission to the NICU, having an Apgar mean score of > 6, having spontaneous respiration at birth, having no congenital malformation potentially causing asphyxia or otherwise affecting respiration, and having no cranial bleeding or hyperbilirubinemia that could have led to blood abnormalities **Exclusion criteria:** having chromosomal anomalies, necrotising enterocolitis, craniofacial malformation, bronchopulmonary dysplasia, respiratory distress syndrome, neonatal seizures, need for mechanical ventilation, periventricular leukomalacia, intracranial haemorrhage, or culture-positive sepsis or meningitis at screening

Kucuk Alemdar 2020 *(Continued)*

Pretreatment: no statistically significant difference was found between groups in terms of gender, gestational age, birth weight, birth height, head circumference, Apgar score at the first and fifth minutes.

N infants analysed: 1) breast milk odour group; n = 30; 2) maternal voice group; n = 30; 3) incubator cloth group; n = 31; 4) control group; n = 32

N infants randomised: N = 136; 1) breast milk odour group; n = 33; 2) maternal voice group; n = 34; 3) incubator cloth group; n = 35; 4) control group; n = 34

Reasons for dropouts: clinically unstable; inadequate breast milk

Reasons for exclusion: not meeting inclusion criteria

Sample size calculation: power analysis was conducted, and the sample size was determined as 123 with a power of 0.90, a significance level of 0.05, and a medium effect size.

Interventions

Intervention characteristics

Control

- *Intervention type*: routine care; treatment, feeding, and hygiene practice
- *Dose*: n.a.
- *Frequency*: n.a.
- *Mode of delivery*: n.a.
- *dB level*: 46-80

Recorded mother's spoken voice

- *Intervention type*: mothers were recorded expressing their thoughts and feelings and anything they wanted to say using a voice recorder
- *Dose*: 30 minutes
- *Frequency*: 3 times per day
- *Mode of delivery*: when awake recording played inside the incubator
- *dB level*: 45

Outcomes

Heart rate

- **Outcome type**: continuous outcome
- **Direction**: lower is better
- **Data value**: endpoint

Respiratory rate

- **Outcome type**: continuous outcome
- **Direction**: lower is better
- **Data value**: endpoint

Oxygen saturation

- **Outcome type**: continuous outcome
- **Direction**: higher is better
- **Data value**: endpoint

Identification

Sponsorship source: The Scientific Research Projects (BAP) of Management Unit of Giresun University (SAĞ-BAP-A-200515-23)

Declaration of interest: the authors declared no conflict of interest.

Country: Turkey

Kucuk Alemdar 2020 *(Continued)*

Setting: level II NICU

Authors name: Dilek Küçük Alemdar

Institution: 1) Giresun University, Health Sciences Faculty, Department of Nursing, Giresun, Turkey; 2) Istanbul University, Health Sciences Faculty, Department of Midwifery, Istanbul, Turkey

Email: dilek.alemdar@giresun.edu.tr; dilek.alemdar@giresun.edu.tr

Address: Dilek Küçük Alemdar, PhD, RN, Assistant Professor Giresun University, Health Science Faculty, Department of Nursing Yenimahalle TR–28340 Piraziz-Giresun (Turkey)

Website: https://www.giresun.edu.tr/en

Notes

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer program (www.randomization.org/block design)
Allocation concealment (selection bias)	Unclear risk	Unclear
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Unclear
Incomplete outcome data (attrition bias) All outcomes	Low risk	In every group, some infants (2-4) had to be excluded due to being "clinically unstable".
Selective reporting (reporting bias)	Unclear risk	Unclear
Other bias	Low risk	No other bias detected

Lafferty 2021

Study characteristics

Methods

Study design: randomised controlled trial

Study grouping: parallel group

Participants

Baseline characteristics

Control

- *Gestational age at birth (weeks) mean ± SD*: 29.7 ± 1.1
- *Birth weight (g) mean ± SD*: 1267 ± 205

Lafferty 2021 *(Continued)*

- *Gender n (%) female*: 8 (42)
- *Ventilated n (%)*: 1 (5)
- *CPAP n (%)*: 11 (58)
- *Room air n (%)*: 2 (11)

Recorded Mozart's double piano sonata

- *Gestational age at birth (weeks) mean ± SD*: 30.0 ± 1.0
- *Birth weight (g) mean ± SD*: 1391 ± 408
- *Gender n (%) female*: 8 (38)
- *Ventilated n (%)*: 1 (5)
- *CPAP n (%)*: 9 (43)
- *Room air n (%)*: 3 (14)

Inclusion criteria: (1) neonates between 280/7 and 316/7 weeks of gestational age; (2) parents equal to or older than 18 years of age whose primary language was English; (3) and the intention of providing breast milk as primary source of nutrition

Exclusion criteria: congenital anomalies affecting hearing or oral feeding and the need for high-frequency ventilatory support. Demographic information was collected on each infant including gestational age, birth weight, sex, time to full enteral feeds, and highest mode of respiratory support at the time of the study.

Pretreatment: no significant differences between the two groups

N infants analysed: 40

N infants randomised: 40

Reasons for dropouts: no

Reasons for exclusion: no

Sample size calculation: the target enrolment number was calculated based on population data that the average time to regain birth weight for LBW infants is 10 days. To detect a 3-day difference in time to regain birth weight, with 80% power and two-sided α at 0.05, 16 babies were needed in each group. This number was rounded up to a target of 20 infants in each group to allow for dropouts.

Interventions	**Intervention characteristics** Control • *Intervention type*: no additional music • *Dose*: 24 minutes • *Frequency*: twice per day for a total of 14 days • *Mode of delivery*: music player in an incubator without playing music • *dB level*: n.a. Recorded Mozart's double piano sonata • *Intervention type*: Mozart's double piano sonata • *Dose*: 24 minutes • *Frequency*: twice per day for a total of 14 days • *Mode of delivery*: recorded music in incubator in between routine infant care times, during daytime or night-time hours • *dB level*: n.a.
Outcomes	*Weight gain per day* • **Outcome type**: continuous outcome

Lafferty 2021 *(Continued)*

- **Reporting**: partially reported
- **Unit of measure**: grams
- **Direction**: higher is better
- **Notes**: exposure to Mozart's double piano sonata for14 days after birth did not significantly improve the time to regain BW. The mean time to regain BW in the music group was 10.0 ± 2.8 days compared with 11.4 ± 3.6 days in the control group (P = 0.181).

Adverse effects

- **Outcome type**: dichotomous outcome
- **Reporting**: fully reported
- **Direction**: lower is better
- **Data value**: endpoint

Identification	**Sponsorship source:** none **Declaration of interest: t**he authors declared no conflict of interest. **Country:** USA **Setting:** NICU Level III **Authors name:** Margaret A. Lafferty **Institution:** DepartmendDepartment of Pediatrics, Christiana Care Health System **Email:** margaret.lafferty@gmail.com **Address:** Margaret Lafferty, MD, Department of Pediatrics, Thomas Jefferson University Hospital, 833 Chestnut StreetSuite 1210, Philadelphia, PA 19107
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Infants were randomised to music (intervention) or no music (control) exposure by using a random number generator.
Allocation concealment (selection bias)	Unclear risk	Unclear
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	To accomplish the blinded nature of this study, a music player was present in each infant's incubator in both groups (unclear if the nurse was blinded to group allocation since music player was placed in the incubator).
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Outcome assessors were blinded.
Incomplete outcome data (attrition bias) All outcomes	Low risk	No incomplete data reporting detected
Selective reporting (reporting bias)	Low risk	Compared with: https://clinicaltrials.gov/ct2/show/NCT02919358?term=NCT02919358&draw=2&rank=1, reported as planned

Lafferty 2021 *(Continued)*

Other bias	Unclear risk	Unclear

Lejeune 2019

Study characteristics	
Methods	**Study design:** randomised controlled trial **Study grouping:** parallel group
Participants	**Baseline characteristics** Control • *Number*: 14 • *Gestational age at intervention (weeks) mean (SD):* 29 (2.2) • *Birth weight (g) mean (SD):* 1207 (274) • *Gender, number (%) girls*: 7 (50) • *Small for gestational, number (%)*: 3 (21.4) • *Socioeconomic score mean (SD):* 6.07 (3.3) • *Birth head circumference (cm) (g) mean (SD):* 26.6 (2.2) • *Bronchopulmonary dysplasia late-onset mean (SD):* 5 (35.7) • *Intraventricular haemorrhage (grade I–II) n (%)*: 3 (21.4) • *Early and late onset sepsis n (%)*: 3 (21.4) • *Patent ductus arteriosus n (%)*: 1 (7.1) Recorded instrumental music • *Number* 13 • *Gestational age at intervention (weeks) mean (SD):* 29.14 (2.3) • *Birth weight (g) mean (SD):* 1217 (377) • *Gender, number n (%) girls*: 8 (61.5) • *Small for gestational n (%)*: 1 (7.7) • *Socioeconomic score mean (SD):* 6.38 (3.6) • *Birth head circumference (cm) (g) mean (SD):* 27.1 (3.1) • *Bronchopulmonary dysplasia late-onset mean (SD)*: 2 (15.4) • *Intraventricular haemorrhage (grade I–II) n (%)*: 4 (30.8) • *Early and late onset sepsis n (%)*: 0 • *Patent ductus arteriosus n (%)*: 1 (7.7) **Inclusion criteria:** gestational age at birth < 32 weeks **Exclusion criteria:** major brain lesions detected on early MRI, such as intraventricular haemorrhage grade III with or without apparent periventricular haemorrhagic infarction, or cystic periventricular leukomalacia **Pretreatment:** for the preterm group, there was only 1 significant difference for the family's socioeconomic status (SES), t(35) = 2.760, P = 0.009. The family's SES is a 12-point scale based on paternal occupation and maternal education (range from 2 (the highest SES) to 12 (the lowest SES)). The SES of the families of preterm children who dropped out (mean = 3.20, SD = 0.9) was higher than that of the families of preterm children who participated in the follow-up assessment (mean = 6.22, SD = 3.4) **N infants analysed:** 23 at 12 months (music: 13; control: 10), 17 at 24 months (music: 10; control: 7) **N infants randomised:** 39 (music: 20; control: 19)

Lejeune 2019 *(Continued)*

Reasons for dropouts: at 12 months: 2 refusals, 2 unreachable; at 24 months: 3 expired, 7 unreachable

Reasons for exclusion: refusal 3, unreachable 3, cerebral anomaly 1, moved 5

Sample size calculation: not reported

Interventions	**Intervention characteristics** Control • *Intervention type*: no music via headphones • *Dose*: 8 minutes • *Frequency*: from gestational age of 33 weeks until hospital discharge or term-equivalent age, 5 times per week • *Mode of delivery*: via open headphones • *dB level*: environmental NICU sounds Recorded instrumental music • *Intervention type*: instrumental calm music especially created by Andreas Vollenweider: the music was composed of background, bells, harp, and punji. Three tracks were created in collaboration with a nurse specialised in developmental care. Music was presented to the baby according to the state of wakefulness, following his biological rhythm: one was composed with the aim of helping the baby to wake up, one to maintain the child in a state of calm awakening, and the last one to help the baby to fall asleep. • *Dose*: 8 minutes • *Frequency*: 5 times per week, from gestational age of 33 weeks until hospital discharge or term-equivalent age • *Mode of delivery*: via headphones • *dB level*: 30-65 dB
Outcomes	*Bayley Scale: cognitive 24 months* • **Outcome type**: continuous outcome • **Reporting**: fully reported • **Scale**: Bayley scale • **Direction**: higher is better • **Data value**: endpoint *Bayley Scale: motor at 24 months* • **Outcome type**: continuous outcome • **Reporting**: fully reported • **Scale**: Bayley scale • **Direction**: higher is better • **Data value**: endpoint • **Notes**:no significant differences *Bayley Scale: language at 24 months* • **Outcome type**: continuous outcome • **Reporting**: fully reported • **Scale**: Bayley scale • **Direction**: higher is better • **Data value**: endpoint • **Notes**: no significant differences
Identification	**Sponsorship source:** Swiss NationalScience Foundation n32473B_135817/1 and the Fondation Prim'enfance

Lejeune 2019 *(Continued)*

Declaration of interest: the authors declared no conflict of interest.

Country: Switzerland

Setting: Neonatal units of the University Hospital of Geneva

Comments: no comments

Authors name: Fleur Lejeune

Institution: Child Clinical Neuropsychology Unit, Faculty of Psychology and Education Sciences, University of Geneva, Geneva, Switzerland

Email: Fleur.Lejeune@unige.ch

Address: Child Clinical Neuropsychology Unit, Faculty of Psychology and Education Sciences, University of Geneva, Geneva, Switzerland

Website: https://www.unige.ch/fapse/en/recherche/groupes/psycho/clinique/upcne/

Notes

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Preterm infants were randomised to either music intervention or control condition (without music).
Allocation concealment (selection bias)	Unclear risk	Unclear
Blinding of participants and personnel (performance bias) All outcomes	Low risk	The nurse determined the readiness for music exposure and chose the track based on a neonatal behavioural assessment scale.
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Subsequent analysis were done by trained psychologists or developmental paediatricians who were blind to the music group assignment.
Incomplete outcome data (attrition bias) All outcomes	High risk	Lost around half of randomised participants until analysis.
Selective reporting (reporting bias)	Low risk	No selective reporting found
Other bias	Low risk	2 coders scored the episodes independently for 16% of the sample after thorough training on the scoring methods. Inter-rater reliability was calculated using Pearson correlations.

Liao 2021

Study characteristics

Methods	**Study design:** randomised controlled trial

Liao 2021 *(Continued)*

	Study grouping: parallel group
Participants	**Baseline characteristics** Control • *Gestational age at birth (weeks) mean (SD):* 31.8 (1.7) • *Gestational age at intervention (weeks) mean (SD):* 33.5 (1.4) • *Birth weight (g) mean (SD):* 1053 (0.20) • *Gender n (%) male*: 18 (17.5) • *Weight at testing (g), mean (SD)*: 1063 (0.10) • *Apgar score at 1 minute, median (range)*: 10 (9–10) Recorded mothers singing voice • *Gestational age at birth (weeks) mean (SD):* 31.3 (1.8) • *Gestational age at intervention (weeks) mean (SD):* 33.7 (1.7) • *Birth weight (g) mean (SD):* 1043 (0.20) • *Gender n (%) male*: 12 (11.6) • *Weight at testing (g), mean (SD)*: 1062 (0.09) • *Apgar score at 1 min, median (range)*: 9 (8–10) **Inclusion criteria:** a) gestational age < 37 weeks with a birth weight of ≤ 1800 g; (b) newborn Apgar score at 1 minute > 5 points; (c) length of NICU stay ≥ 1 week; (d) infant being cared for in an incubator; (e) spontaneous breathing and stable haemodynamics; (f) absence of severe craniocerebral diseases, genetic metabolic diseases, heart diseases, and/or other serious illnesses; and (g) no history of surgery **Exclusion criteria:** (a) presence of congenital hearing loss and (b) use of sedatives or muscle relaxants. The study termination criteria were as follows: (a) parent request for withdrawal; (b) serious adverse reactions; and, (c) medically instability or death or discharge from the incubator within the period of study. Staff members were asked to continue all usual routines and care practices. **Pretreatment:** no significant differences were found in sexes, types of delivery, use of dexamethasone, Apgar scores at 1 minute, gestational ages, birth weights, and body weights at the time of study amongst the 3 groups (P > 0.05). **N infants analysed:** 103 **N infants randomised:** 103 **Reasons for dropouts:** in the routine care group, 5 infants were withdrawn due to the worsening of their conditions (n = 2) and discharge during the study period (n = 3). In the white noise group, 4 infants were withdrawn due to the worsening of their condition (n = 3) and discharge (n = 1). In the mothers' voice group, 3 infants were withdrawn due to the worsening of their condition (n = 3). **Reasons for exclusion:** **Sample size calculation:** the sample size was estimated using the free G-power programme (3.1.9.7), which calculates the sample size using the mean difference and standard deviation. The preliminary experiment was conducted on 15 preterm infants with 5 infants in each group. The means (M) and standard deviations (SD) of the sleep efficiency of premature infants with the mothers' voice, white noise, and routine care groups were as follows: M 1 = 0.456, M 2 = 0.578, M 3 = 0.400, SD 1 = 0.171, SD 2 = 0.173, and SD 3 = 0.174, respectively. With a calculated effect size of 0.43, a power of 0.90, an alpha of 0.05 (2-sided), and an attrition rate of 20% in the present study, the minimum total sample of preterm infants was 90, and 103 premature infants were randomised finally.
Interventions	**Intervention characteristics** Control

Liao 2021 *(Continued)*

- *Intervention type*: in the routine care group, preterm infants were cared for with feeding every 3 hours, performing the physiological measurements of heart rate and oxygen saturation by neonatal nurses, without any other voices interventions.

Recorded mothers' singing voice

- *Intervention type*: the mothers were invited to record their singing of the Chinese version of Lullaby (the same as Schubert's lullaby, Wiegenlied, D498) for about 5 minutes in a quiet environment on her smartphone or recorder. The mothers were encouraged to call the baby's name and speak in soft voices for comforting the newborn baby. Finally, the recorded songs were uploaded onto a streamlined digital audio software program to edit any background noise and were played on a loop for 20 minutes. The playing time and the doses of the mothers' voice intervention were the same as those administered to the white noise group.
- *Dose*: 20 minutes
- *Frequency*: 3 times a day for 4 consecutive days
- *Mode of delivery*: recording played in incubator
- *dB level*: 50-55 dB

Outcomes	*Heart rate* • **Outcome type**: continuous outcome • **Reporting**: fully reported • **Direction**: lower is better • **Data value**: endpoint *Oxygen saturation* • **Outcome type**: continuous outcome • **Direction**: higher is better • **Data value**: endpoint *Weight Gain* • **Outcome type**: continuous outcome • **Reporting**: fully reported • **Direction**: higher is better • **Data value**: change from baseline *Adverse effects* • **Outcome type**: dichotomous outcome • **Reporting**: fully reported • **Direction**: lower is better • **Data value**: endpoint
Identification	**Sponsorship source:** the study was supported by the Young Backbone Personnel Training Project that was funded by the Health Department of Fujian Province in China (Project No.:2014-ZQN-JC-26). **Declaration of interest:** the authors declared no conflict of interest. **Country:** China **Setting:** NICU Level III **Comments:** **Authors name:** Jinhua Liao **Institution:** School of Nursing, Fujian Medical University, China **Email:** hulu2886@sina.com (R. Hu), nurhhg@nus.edu.sg (H.-G. He)

Liao 2021 *(Continued)*

	Address: School of Nursing, Fujian Medical University, 1 Xue Yuan Road, University Town, Fuzhou, Fujian, China **Website:**
Notes	*Primary outcome (comparison of sleep-wake patterns)* 1-way ANOVA showed no significant differences in all sleep-wake patterns amongst the 3 groups (P > 0.05). Paired-Student *t*-test showed no significant differences between pre- and post-tests in all sleep-wake patterns in both the white noise and mothers' voice groups. Significant differences were only found between the pre-test and post-test evaluations in sleep efficiency, wake time after sleep onset, and average awakening time in the routine care group. Therefore, hypothesis 1 was partially supported. *Comparison of cortisol levels* No significant differences in cortisol levels were found in all the groups between the pre-test and post-test (P > 0.05). In addition, no significant differences were found amongst the 3 groups at pre-test and post-test evaluations (P = 0.218, P = 0.069, respectively) by the Kruskal–Wallis analysis.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Divided into 3 by a 3rd-party researcher who was blinded to the patients and did not participate in the clinical trial. Random sequence of 103 numbers was generated using Microsoft Excel
Allocation concealment (selection bias)	Low risk	The allocation was concealed using the closed-envelope method and was kept by the 3rd researcher.
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Blinding of outcome assessment unclear
Incomplete outcome data (attrition bias) All outcomes	Low risk	In the routine care group, 5 infants were withdrawn due to the worsening of their condition (n = 2) and discharge during the study period (n = 3). In the white noise group, 4 infants were withdrawn due to the worsening of their condition (n = 3) and discharge (n = 1). In the mothers' voice group, 3 infants were withdrawn due to the worsening of their condition (n = 3). Finally, 91 preterm infants completed the interventions and all the measurements.
Selective reporting (reporting bias)	Unclear risk	Unclear if selected reporting happened
Other bias	Low risk	No other bias detected

Loewy 2013

Study characteristics

Methods	**Study design:** randomised controlled trial

Loewy 2013 *(Continued)*

	Study grouping: cross-over
Participants	**Baseline characteristics** Overall • *Gestational age at birth (weeks) mean (SD)*: 29.57 (2.89) • *Age at intervention (days) median (minimum/maximum)*: 22 (1, 140) • *Birth weight (g) mean ± SD*: 1321.22 ± 495.32 • *Gender female n (%)*: 152 (56%) • *Apgar score at birth mean ± SD*: 6.81 ± 2.08 • *Apgar score at 5 minutes mean ± SD*: 8.11 ± 1.21 • *Respiratory distress/Respiratory distress syndrome (frequency (%)*: 153 (88%) • *Clinical sepsis (frequency (%)*: 87 (32%) • *SGA (frequency %)*: 53 (19%) **Inclusion criteria:** infants ≥ 32 weeks' premature with respiratory distress syndrome, clinical sepsis, and/or SGA (small for gestational age) **Exclusion criteria:** no specific exclusion criteria were mentioned. **Pretreatment:** not reported, since within-subject **N infants analysed:** 272 **N infants randomised:** 284 **Reasons for dropouts:** early discharge (n = 4); data collection availability was not possible (n = 8) **Reasons for exclusion:** study was originally designed to take place at 8 hospitals for the duration of 2 years; 8 sites were approved by their hospitals' institutional review boards, but 2 sites were discontinued after the study mentioned year 1, due to music therapist position changes. Data collection took place at a total of 11 hospitals where the participation of new sites was co-ordinated with the investigatory team. On average, each site maintained a total of 30 babies during the 2½ -year study period, excluding the discontinued site. **Sample size calculation:** an initial sample estimate revealed that 240 infants allowed for 80% power to detect a small size (Cohen) of 0.18. The final sample of 272 allowed for 84% power to detect an effect size of this magnitude.
Interventions	**Intervention characteristics** Control • *Intervention type*: standard care • *Frequency*: 3 per week (either morning or afternoon) during the afternoon period of 2 weeks • *Mode of delivery*: no explicit aural stimulation presented Lullaby song-of-kin • *Intervention type*: Lullaby "song-of-kin" • *Dose*: 6 interventions: 10 minutes of intervention • *Frequency*: each treatment was given 2 times per week over the course of the 2-week study period. The presentation of the treatments was varied by day of the week within each week and by the time of day and randomised (either morning or the afternoon) across the 2 weeks or 3 days per week during a period of 2 weeks for a total of 6 interventions. • *Mode of delivery*: provided live and delivered through the portholes of the incubators, isolette, or at bassinet side at the infants' midline to encourage foetal positioning • *dB level*: 55-65 Ocean Disc

Loewy 2013 *(Continued)*

- *Intervention type*: continuous relaxing instrumental sound similar to wave sounds
- *Dose*: 6 interventions: 10 minutes of intervention
- *Frequency*: 3 days per week during a period of 2 weeks for a total of 6 interventions
- *Mode of delivery*: provided live and delivered through the portholes of the incubators, isolette, or at the bassinet side at the infants' midline to encourage foetal positioning
- *dB level*: 55-65

Gato Box

- *Intervention type*: the gato-box is a small rectangular tuned musical instrument that is used to provide an entrained rhythm in soft timbre meant to simulate a heartbeat sound that the neonate would hear in the womb
- *Dose*: 6 interventions: 10 minutes of intervention
- *Frequency*: 3 days per week during a period of 2 weeks for a total of 6 interventions
- *Mode of delivery*: provided live and delivered through the portholes of the incubators, isolette, or at the bassinet side at the infants' midline to encourage foetal positioning
- *dB level*: 55-65

Outcomes

Heart rate

- **Outcome type**: continuous outcome
- **Reporting**: partially reported
- **Unit of measure**: beats per minute
- **Direction**: lower is better
- **Data value**: change from baseline
- **Notes**: all 3 interventions showed a significant effect of the intervention over time (before, during, and after) on HR. Lower HRs were recorded during the intervention only for lullaby and gato-box post-intervention. There was a greater pre–post intervention response for lullaby than for gato-box. For the ocean disc, the HR decreased significantly after the intervention.

Respiratory rate

- **Outcome type**: continuous outcome
- **Reporting**: partially reported
- **Direction**: lower is better
- **Data value**: change from baseline
- **Notes**: there appeared to be a trend for RR for the ocean disc with higher RRs during intervention and after intervention, whereas the P = for lullaby was 0.47 and for the gato-box was 0.71.

Oxygen saturation

- **Outcome type**: continuous outcome
- **Reporting**: partially reported
- **Direction**: higher is better
- **Data value**: change from baseline
- **Notes**: there was a significant difference in oxygen saturation levels as a function of lullaby type with"Twinkle" showing higher levels than the song-of-kin.

Identification

Sponsorship source: Heather on Earth Music Foundation

Declaration of interest: the authors indicated they had no financial relationships relevant to this article to disclose.

Country: USA

Setting: multi-centre study (11 hospitals/NICUs)

Comments: the study was originally designed to take place at 8 hospitals for the duration of 2 years; 8 sites were approved by their hospitals' institutional review boards (IRBs), but 2 sites were discontinued

Loewy 2013 *(Continued)*

after year 1, due to music therapist position changes. Data collection took place at a total of 11 hospitals where the participation of new sites was co-ordinated with the investigatory team. On average, each site maintained a total of 30 babies during the 2½-year study period, excluding the discontinued site.

Authors name: Joanne Loewy

Institution: The Louis Armstrong Center for music and Medicine, NICU, and Department of Biostatistics, Beth Israel Medical Center, New York, New York

Email: jloewy@chpnet.org

Address: The Louis Armstrong Center for music and Medicine, Beth Israel Medical Center, 6 Silver 21 1st Ave and 16th St, New York, NY 1003

Website: https://www.mountsinai.org/locations/music-therapy

Notes

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Allocation sequencing was generated by the biostatistician, who had no contact with the participants. Randomisation of the sequence of presentation was generated by computer to allow for separate sequences for morning and afternoon.
Allocation concealment (selection bias)	Low risk	Allocation sequencing was generated by the biostatistician, who had no contact with the participants.
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Outcome assessors were blinded.
Incomplete outcome data (attrition bias) All outcomes	Low risk	Some infants were omitted because of early discharge (n = 4) or data collection availability was not possible (n = 8).
Selective reporting (reporting bias)	Low risk	Selective outcome reporting bias not detected
Other bias	Unclear risk	Unclear

Menke 2021

Study characteristics

Methods

Study design: randomised controlled trial

Study grouping: parallel group

Menke 2021 *(Continued)*

Participants	**Baseline characteristics** Control • *Gestational age at birth (days) mean (SD):* 187.46 (14.80) • *Length mean (SD)*: 33.81 (4.32) • *Birth weight (g) mean (SD):* 858.85 (334.97) • *Gender n (%) female*: 8 (30.8) • *Mothers age mean (SD):* 34.96 (7.51) • *Birth head circumference (cm) (g) mean (SD):* 24.01 (2.72) Live interactive improvised music therapy • *Gestational age at birth (days) mean (SD):* 196.29 (8.36) • *Length mean (SD):* 36.75 (3.86) • *Birth weight (g) mean (SD):* 1072.08 (346.47) • *Gender n (%) female*: 12 (50) • *Mothers age mean (SD):* 33.50 (5.48) • *Birth head circumference (cm) (g) mean (SD):* 25.40 (2.45) **Inclusion criteria:** gestational age at birth younger or equal to 30 weeks, chronological age older or equal to 10 days of life, clinical stability (at the time of inclusion) as well as sufficient German language skills and informed consent of the parents **Exclusion criteria:** genetically defined syndrome, neurological diseases (e.g. high-grade intraventricular haemorrhage, periventricular leukomalacia), severe sepsis (necrotising enterocolitis), clinically relevant malformations as well as malformations that potentially impaired the development of 2 children up to the age of 2 years. Preterm infants with the need for palliative care were also excluded from the study. **Pretreatment:** not reported **N infants analysed:** 50 **N infants randomised:** 65 **Reasons for dropouts:** **Reasons for exclusion:** change of study design during the course of recruiting. 2 infants had to be excluded from further analysis because of serious medical complications. 1 family withdrew from the study because of a misunderstanding regarding the procedure of the study. **Sample size calculation:** the target group of the study was premature infants and their parents or primary caregivers at the beginning and the end of the hospitalisation of the infant. According to a power analysis (single outcome, 2-sample t-test for independent samples), N = 104 parent-infant pairs would have been needed to obtain a power of 80% assuming medium-sized effects. However, this sample size was not feasible given the planned duration of the recruitment process of the pilot study of a maximum of 1 year. Based on the birth rate of the University Children's Hospital in Heidelberg within the last 12 months, we aimed to initially recruit N = 65 parent-infant pairs and, due to expected dropouts during the course of the study, to include N = 50 pairs in the final data analysis.
Interventions	**Intervention characteristics** Live interactive improvised music therapy • *Intervention type*: live family-centred music therapy • *Dose*: 30 minutes • *Frequency*: 2 x per week • *Mode of delivery*: live at bedside or in skin-to-skin care with parents

Menke 2021 *(Continued)*

Outcomes

Weight gain

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Unit of measure**: g per day
- **Direction**: higher is better
- **Data value**: endpoint

Duration of hospitalisation in days

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Direction**: lower is better
- **Data value**: endpoint

Parental Stress Questionnaire: Stress (PSQ)

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Scale**: 4-point Likert scale
- **Direction**: lower is better

State-Anxiety Inventory (STAI-SKD)

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Scale**: 4-point Likert scale
- **Range**: 20-80
- **Direction**: lower is better
- **Data value**: endpoint
- **Notes**: MATERNAL outcomes

Edinburgh Postnatal Depression Scale (EPDS)

- **Outcome type**: continuous outcome
- **Reporting**: partially reported
- **Scale**: 5-point Likert scale
- **Range**: 0-30
- **Direction**: lower is better
- **Data value**: endpoint
- **Notes**: MATERNAL outcomes

Parental Competences Questionnaire

- **Outcome type**: continuous outcome
- **Reporting**: partially reported
- **Scale**: 6-point scale
- **Range**: 8-48
- **Direction**: lower is better
- **Data value**: endpoint

State-Trait Anxiety Inventory (STAI-T)

- **Outcome type**: continuous outcome
- **Reporting**: partially reported
- **Scale**: 4-point Likert scale
- **Range**: 20-80

Menke 2021 *(Continued)*

- **Direction**: lower is better
- **Data value**: endpoint

Parental Stress Questionnaire: Resources (PSQ)

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Scale**: 4-point Likert Scale
- **Range**: 6-24
- **Direction**: higher is better
- **Data value**: endpoint

Identification	**Sponsorship source:** no funding **Declaration of interest:** not stated **Country:** Germany **Setting:** NICU Level III **Comments:** **Authors name:** Barbara Menke **Institution:** Department of Neonatology, University Children's Hospital, Heidelberg, Germany **Email:** Barbara.Menke@srh.de **Address:** Im Neuenheimer Feld 43069120 Heidelberg **Website:** https://www.klinikum.uni-heidelberg.de/zentrum-fuer-kinder-und-jugendmedizin/iv-neonatologie
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Using a random sequence, participants were randomly assigned to the treatment or the control group.
Allocation concealment (selection bias)	Low risk	Sealed, opaque envelopes
Blinding of participants and personnel (performance bias) All outcomes	High risk	Judgement comment: in the present study, blinding of the participating families could not be ensured because standard care with the music therapy intervention was tested against standard care without the music therapy intervention. It follows, as in almost all studies of the effectiveness of psychosocial interventions, that the music therapist was not blinded to group assignment. For organisational reasons, it was also not possible to ensure that the observers were fully blinded. The use of objective measurement data, such as body measurements at child discharge, reduces the risk of intentional bias in data collection.
Blinding of outcome assessment (detection bias) All outcomes	High risk	No blinding of outcome assessors
Incomplete outcome data (attrition bias)	Low risk	No incomplete outcome data detected

Menke 2021 *(Continued)*
All outcomes

Selective reporting (reporting bias)	Low risk	Judgement comment: all outcomes were collected and reported as planned in the study protocol and approved by the Ethics Committee of the Medical Faculty, Heidelberg University (Study ID: S-044/2016). At the beginning of the study, data were collected in a pre-post follow-up design with a fixed number of music therapy interventions, independent of the duration of hospitalisation. This initial design was not well accepted by many families who expected interventions throughout the entire hospital stay. Thus, the design was changed during the course of the study to a pure pre-post design. The first 9 data sets were collected according to the initial design. These data sets were excluded from further analysis, as the 2 designs differed in schedule, rendering the data sets incomparable to the rest. Only complete data sets were included in data analysis, leading to the exclusion of 3 data sets containing missing data (1 data set in the treatment group and 2 in the control group). A total of 50 parent-infant pairs remained as the final cohort to analyse physiological development at discharge. 47 mothers and 30 fathers completed the questionnaires on parental stress factors and were included in the respective analysis.
Other bias	Low risk	Judgement comment: both parents were invited to participate in music therapy. However, in practice, it was not always possible that both parents were present during music therapy interventions. For example, some families alternated their visits to the ward in order to be able to care for siblings at home. For this reason, it was not possible to ensure that both parents participated fully in the intervention in every case during the study period. This limits the conclusions that can be drawn from the music therapy intervention about the parental stress factors, especially for the fathers. A strict separation of experimental and control groups could not always be guaranteed despite the greatest efforts for organisational reasons in the course of the units. It could happen that families of the experimental and control groups temporarily shared a room. In this situation, which was unfavourable for the study, care was taken to ensure that the parents of the control group were not in the room during the music therapy interventions. However, this did not completely prevent the children in the control group from listening to the music. The music therapy intervention was investigated as a supplement to standard care. This was defined by an implemented care concept. However, a uniform definition of standard care is difficult. For further investigations, it should be examined whether further parameters, such as the use of psychological counselling, should also be recorded and thus controlled.

Nakhwa 2017

Study characteristics	
Methods	**Study design:** randomised controlled trial **Study grouping:** parallel group
Participants	**Baseline characteristics** Control developmental programme • *Gestational age at birth (weeks) mean (SD):* 33.33 ± 2.808 • *Gender ratio (female:male)*: 11:07 Developmental programme + recorded music • *Gestational age at birth (weeks) mean (SD):* 33 ± 2.828

Nakhwa 2017 *(Continued)*

- *Gender ratio (female:male)*: 9:9

Overall

- *Gestational age at birth (weeks) mean (SD):* 33
- *Gender ratio (female:male)*: 18:16

Inclusion criteria: preterm infants clinically diagnosed by a paediatrician

Exclusion criteria: hypoxic ischaemic encephalopathy, any congenital anomalies

Pretreatment: not reported

N infants analysed: n = 36 (n = 18 controlled group; n = 18 experimental group)

N infants randomised: 40 (20 + 20)

Reasons for dropouts: There were 2 dropouts in each group because of early discharge.

Reasons for exclusion: not reported

Sample size calculation: not reported

Interventions	**Intervention characteristics** Control developmental programme • *Intervention type*: developmental programme involved therapeutic positioning, joint proprioception, tactile facilitation, and vestibular facilitation • *Dose*: 30 minutes • *Frequency*: 3 times per week for 3 weeks • *Mode of delivery*: delivered during quiet alert state of the infant, 30 minutes before feeding Developmental programme + recorded music • *Intervention type*: developmental programme involved therapeutic positioning, joint proprioception, tactile facilitation and vestibular facilitation + soft lullaby • *Dose*: 30 minutes • *Frequency*: 3 times per week for 3 weeks • *Mode of delivery*: delivered during quiet alert state of infant, 30 minutes before feeding + lullaby via speaker • *dB level*: 30-40 dB
Outcomes	*Infant Motor Performance score* • **Outcome type**: continuous outcome • **Reporting**: partially reported • **Direction**: higher is better • **Data value**: change from baseline • **Notes**: the intervention group's mean value of the test of infant motor performance score was 21.167 with a standard deviation of 0.5145. The control group mean group P value of the test of infant motor performance score was 20.78 with a standard deviation of 0.4278. Using the “paired t-test” the P value was < 0.0001 which shows an extremely significant difference. The t-value was 2.715 with a degree of freedom of 34. *Infant Neurological International Battery Score* • **Outcome type**: continuous outcome • **Direction**: higher is better • **Data value**: change from baseline

Nakhwa 2017 *(Continued)*

- **Notes**: the intervention group mean value of the infant neurological international battery score was 65.11 with a standard deviation of 1.568. The control group mean value of the infant neurological international battery score was 63.22 with standard deviation of 2.756. The P value was 0.0163 which shows an extremely significant difference. The t-value was 2.528 with a degree of freedom of 34.

Identification	**Sponsorship source:** not reported **Declaration of interest:** not stated **Country:** India **Setting:** Single centre, NICU of Pravara Rural Hospital, Loni **Comments:** identification only partially reported **Authors name:** Priyanka K. Nakhwa **Institution:** Paediatric Physiotherapy, Pravara Rural Hospital, Loni (Bk) **Address:** Paediatric Physiotherapy, Pravara Rural Hospital, Loni. **Website:** https://www.pravara.com/rmc-paediatrics.html
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Simple random sampling
Allocation concealment (selection bias)	High risk	Inadequate concealment of allocation prior to assignment
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Unclear if outcome assessors were blinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	No incomplete outcome data detected
Selective reporting (reporting bias)	Unclear risk	Unclear if selective reporting happened
Other bias	Unclear risk	Unclear if other bias occurred

Namjoo 2021

Study characteristics

Methods	**Study design:** randomised controlled trial

Namjoo 2021 *(Continued)*

Study grouping: parallel group

Participants	**Baseline characteristics** • *Gestational age at birth (weeks) mean ± SD:* 34.30 ± 1.95 • *Gestational age at intervention (weeks): n.a.* • *Birth weight (g) mean ± SD:* 2368.3 ± 631.96 • *Gender n (%) male*: 20 (66.7) • *Height at birth mean ± SD:* 44.51 ± 4.23 • *Apgar minute 1 mean ± SD:* 8.43 ± 1.45 • *Apgar minute 5 mean ± SD:* 9.6 ± 1.22 Recorded Persian lullaby • *Gestational age at birth (weeks) mean ± SD:* 34.53 ± 1.87 • *Gestational age at intervention (weeks): n.a.* • *Birth weight (g) mean ± SD:* 2525.4 ± 577.91 • *Gender n (%) male*: 17 (56.7) • *Height at birth mean ± SD:* 44.98 ± 3.32 • *Apgar minute 1 mean ± SD:* 8.73 ± 0.44 • *Apgar minute 5 mean ± SD:* 9.76 ± 0.43 Live mother's lullaby • *Gestational age at birth (weeks) mean ± SD:* 34.03 ± 2.52 • *Gestational age at intervention (weeks): n.a.* • *Birth weight (g) mean ± SD:* 2286.6 ± 901.127 • *Gender n (%) male*: 18 (60) • *Height at birth mean ± SD:* 43.61 ± 6.01 • *Apgar minute 1 mean ± SD:* 8.10 ± 1.06 • *Apgar minute 5 mean ± SD:* 9.23 ± 0.93 **Inclusion criteria:** infants aged over 28 weeks of gestation at birth **Exclusion criteria:** congenital malformations (heart disease, hydrocephalus, etc.), taking any sedatives for up to 3 hours before the intervention, mechanically ventilated during the intervention, hearing impairment (hearing health was confirmed by the startle reflex and the doctor's opinion), suffering from hypoxic ischaemic encephalopathy **Pretreatment:** the results showed no statistically significant difference between the three groups in demographic characteristics of the mothers and infants (P > 0.05), and the 3 groups were homogeneous in terms of demographic variables **N infants analysed:** 90 **N infants randomised:** 90 **Reasons for dropouts:** no **Reasons for exclusion:** no **Sample size calculation:** the study sample size was 81 preterm infants (27 in each group) and was powered 90% using a pilot study. However, regarding the probability of dropouts and an increase in the power of our study, 96 samples were assigned to the intervention and control groups, respectively (total 96 samples).
Interventions	**Intervention characteristics** Control

Namjoo 2021 *(Continued)*

- *Intervention type*: standard care with infants lying in mothers` arm
- *Dose*: 20 minutes
- *Frequency*: once a day for 14 days

Recorded Persian lullaby

- *Intervention type*: the recorded voice of a strange woman singing in Persian. The recorded songs were Persian lullabies released by Pouya Publications, which were approved by child psychologists and authorised by the Ministry of Culture and Islamic Guidance. In addition, the lullabies were prepared in line with Iranian cultural norms.
- *Dose*: 20 minutes
- *Frequency*: once a day for 14 days
- *Mode of delivery*: recording played to infant while laying in the mother's arm using headphones
- *dB level*: 45

Live mother's lullaby

- *Intervention type*: The infants in the second group were placed in the mother's arms, and the mother sang lullabies to the infant.
- *Dose*: 20 minutes
- *Frequency*: once a day for 14 days
- *Mode of delivery*: live sang by mother
- *dB level*: n.a.

Outcomes	*Heart rate* • **Outcome type**: continuous outcome • **Reporting**: fully reported • **Direction**: lower is better • **Data value**: endpoint *Oxygen saturation* • **Outcome type**: continuous outcome • **Reporting**: fully reported • **Direction**: higher is better • **Data value**: endpoint
Identification	**Sponsorship source:** Kerman University of Medical Sciences (IR.KMU.REC.97000409) **Declaration of interest:** the funding organisation(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication. **Country:** Iran **Setting:** NICU Level III **Comments:** **Authors name:** Razyeh Namjoo **Institution:** Nursing in Neonatal Intensive Care, Nursing Research Center, Kerman University of Medical Sciences, Kerman, Iran **Email:** m.nematolahi@kmu.ac.ir **Address:** Nursing Research Center, Kerman University of Medical Sciences, Kerman, Iran, Phone:+98 9132416944, Fax:+3431325218

Namjoo 2021 *(Continued)*

	Website:
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	"The participants were assigned randomly to the groups. To this end, one infant was placed in the mother's live lullaby group, one in the recorded lullaby group, and the third infant in the control group. The same procedure was repeated until all participants were assigned."
Allocation concealment (selection bias)	Unclear risk	Unclear if allocation concealment happened
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Unclear if outcome assessors were blinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	No incomplete outcome data detected
Selective reporting (reporting bias)	Low risk	No selective reporting detected
Other bias	Low risk	No other bias detected

Portugal 2017

Study characteristics

Methods	**Study design:** randomised controlled trial **Study grouping:** parallel group
Participants	**Baseline characteristics** Control • *Gestational age at birth (weeks) mean ± SD:* 29.4 ± 2.0 • *Birth weight (g)mean ± SD:* 1.128 ± 462 • *Gender n (%) male*: 1 (11.1) • *Apgar 1 minute mean ± SD:* 6.4 ± 2.1 • *Number*: 9 • *Apgar 5 minute mean ± SD:* 8.0 ± 1.0 • *Standard otoacoustic emissions n (%)*: 9 (100) • *Sepsis n (%)*: 4 (44.4) • *Chronic pulmonary disease, n (%)*: 1 (11.1) • *Intraventricular haemorrhage (grade I/II), n (%)*: 2 (22.2)

Portugal 2017 *(Continued)*

- *Necrotising enterocolitis, n (%)*: 0 (0)
- *Days in the study mean ± SD:* 37.6 ± 20.6

Recorded mother's voice and heartbeats

- *Gestational age at birth (weeks) mean ± SD:* 28.7 ± 1.6
- *Birth weight (g) mean ± SD:* 1.153 ± 334
- *Gender n (%) male*: 3 (33.3)
- *Apgar 1 minute mean ± SD:* 7.0 ± 1.2
- *Number:* 9
- *Apgar 5 minute mean ± SD:* 8.0 ± 1.2
- *Standard otoacoustic emissions n (%)*: 9 (100)
- *Sepsis n (%)*: 5 (55.6)
- *Chronic pulmonary disease, n (%)*: 1 (11.1)
- *Intraventricular haemorrhage (grade I/II), n (%)*: 4 (44.4)
- *Necrotising enterocolitis, n (%)*: 0 (0)
- *Days in the study mean ± SD:* 37.3 ± 21.2

Inclusion criteria: newborns, admitted to the Neonatology Unit of the Hospital of São João, with gestational age between ≥ 26 and ≤ 33 weeks confirmed by ultrasound

Exclusion criteria: newborn infants with major congenital or chromosomal abnormalities (Down syndrome, Turner syndrome, and Klinefelter syndrome), congenital infections, significant brain damage, with prenatal diagnosis (e.g. neonatal asphyxia), uncontrolled maternal disease, maternal history of tobacco, alcohol, and illicit drug consumption, history of significant maternal nutrient deprivation or malnutrition

Pretreatment: no statistically significant differences were found between the groups regarding their characterisation.

N infants analysed: 18

N infants randomised: 18

Reasons for dropouts: not available

Reasons for exclusion: 7 of the parents declined to participate, and the remaining 67 were excluded for various reasons, from linguistic barriers, logistical difficulties (unavailability of monitors, unavailability of recording material, improvement works at the unit)

Sample size calculation: not stated

Interventions

Intervention characteristics

Control

- *Intervention type*: standard care with environmental sounds of NICU
- *Dose*: 45 minutes
- *Frequency*: 4x/day for ~7 weeks
- *Mode of delivery*: standard care with environmental sounds of NICU
- *dB level*: not recorded

Recorded mother's voice and heartbeats

- *Intervention type*: mother’s recorded voice (spoken and sung) with heartbeats
- *Dose*: 45 minutes after feeding
- *Frequency*: 4 times a day until discharge or infant being placed in the cradle
- *Mode of delivery*: recorded via speakers placed inside the incubator, baby placed in the nest
- *dB level*: 60-65 dB

Portugal 2017 *(Continued)*

Outcomes	*Heart rate* • **Outcome type**: continuous outcome • **Reporting**: fully reported • **Unit of measure**: beats per minute • **Direction**: lower is better • **Data value**: endpoint *Respiratory rate* • **Outcome type**: continuous outcome • **Reporting**: fully reported • **Unit of measure**: breath per minute • **Direction**: lower is better • **Data value**: endpoint *Oxygen saturation* • **Outcome type**: continuous outcome • **Reporting**: fully reported • **Unit of measure**: % • **Direction**: higher is better • **Data value**: endpoint
Identification	**Sponsorship source:** not stated **Declaration of interest:** not stated **Country:** Portugal **Setting:** NICU **Comments:** not available **Authors name:** Crisanta Maria Gomes da Silva Leopoldo **Institution:** Santa Maria Health School, Porto **Email:** crisanta-portugal@sapo.pt **Address:** Praceta Fernão de Magalhães, 101, 10.º, Vila Nova de Gaia, 4049-024, Porto, Portugal **Website:** not available
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random sequence generation reported
Allocation concealment (selection bias)	Unclear risk	Unclear allocation concealment
Blinding of participants and personnel (performance bias)	High risk	No blinding of participants and personnel

Portugal 2017 *(Continued)*

All outcomes		
Blinding of outcome assessment (detection bias) All outcomes	High risk	No blinding of outcome assessment
Incomplete outcome data (attrition bias) All outcomes	Low risk	Complete outcome data
Selective reporting (reporting bias)	Unclear risk	Unclear selective reporting
Other bias	Unclear risk	Unclear if there was other bias

Tandoi 2015

Study characteristics	
Methods	**Study design:** randomised controlled trial **Study grouping:** parallel group
Participants	**Baseline characteristics** Control • *Cardiac frequency, mean:* 137.7 • *Respiratory frequency, mean*: 43.1 • *Saturation, mean*: 96.2 • *Gender (%) male*: 59 (n = 10) • *Apgar 5 (mean)*: 9.1 • *Weight, mean grams*: 2107 • *Behavioural score (mean)*: 2.3 • N: 17 Recorded womb sounds • *Cardiac frequency, mean*: 133.8 • *Respiratory frequency, mean*: 41.4 • *Saturation, mean*: 96.6 • *Gender (%) male*: 53 (n = 9) • *Apgar 5 (mean)*: 9.4 • *Weight, mean grams*: 2048 • *Behavioural score (mean)*: 2.1 • N: 17 **Inclusion criteria:** 1. birth within the previous 24 hours; 2. gestational age greater than or equal to 32 weeks and less than 37 weeks; 3. birth weight greater than or equal to 1500 g; 4. written informed consent of a parent **Exclusion criteria:** major disease **Pretreatment:** 2 completely overlapping populations and therefore comparable. Both groups were comparable for the type of deliveries (50% versus 55% caesarean section) and type of feeding (60% versus 55% breastfeeding, respectively)

Tandoi 2015 *(Continued)*

N infants analysed: 34

N infants randomised: 34

Reasons for dropouts: no dropouts

Reasons for exclusion: no exclusions

Sample size calculation: This gave our study an 80% probability of detecting a difference of 15.0 bpm between the two groups with a significance of 5% two-way, assuming a standard deviation of the response of 15.0 bpm. Similar sample sizes have been found to be consistent with the assessment of one standard deviation difference amongst groups, in a recent review on this topic.

Interventions

Intervention characteristics

Control

- *Intervention type*: usual environmental noise
- *Dose*: 30 min
- *Frequency*: daily for the first 10 day of life

Recorded womb sounds

- *Intervention type*: TOS is a musical track recording obtained with several specific audio equipment. The recording has been done using special contact microphones in contact (membrane microphones), Doppler sensors, and micro-vaginal microphones strictly in contact with the womb. Using hydrophones placed in water tanks and in contact with the womb of pregnant women, we have recorded sounds like a foetal heartbeat, breathing, blood flow and other visceral sounds, potentially recreating those internal sources that the foetus hears during the gestational period. In addition, TOS includes a series of sounds that the foetus experienced during intrauterine life such as female and male voices, footsteps, electrical noises, traffic noises, TV and radio output, music, wind, chirping, ocean waves, traffic, TV and radio output recorded from inside a sheath with characteristics similar to the uterine wall and filtered by calibrating the different sequences to reproduce what the foetus hears from external sources. TOS is faithfully replicating all intrauterine songs emphasising all high frequencies and filtering the song output up to 70/80 decibels. With this method, we obtained a rhythm track that plays the music and melodies that help to create the first audio-sensory memory of the foetus.
- *Dose*: approx. 30 minutes
- *Frequency*: 3 times a day on the first 3 days of life
- Each infant concluded the trial at discharge from the NICU and in any case after 10 days of hospitalisation
- *Mode of delivery*: sound diffusion has been done using 2 speakers and an MP3 player placed opposite the baby's head with a limited volume (maximum 70/80 decibels) and fixed upside down reproducing the sound within the cradle.
- *dB level*: 55-70

Outcomes

Heart rate

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Direction**: lower is better
- **Data value**: change from baseline
- **Notes**: heart rate in the treated group was positively correlated with TOS exposure, showing a significant reduction on day 2.

Respiratory rate

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Unit of measure**: breaths per minute
- **Direction**: lower is better

Tandoi 2015 *(Continued)*

- **Data value**: change from baseline
- **Notes**: not significant, a decrease in RR on day 2 in the TOS group, in agreement with the reduction of HR.

Oxygen saturation

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Unit of measure**: transcutaneous O_2 saturation
- **Direction**: higher is better
- **Data value**: change from baseline

Behavioural scale (APIB)

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Scale**: infant behaviour assessment scale (APIB)
- **Range**: 1-7
- **Direction**: lower is better
- **Data value**: change from baseline
- **Notes**: overall, this type of treatment, according to behavioural parameters, involves a ''calming'' effect about 3 times greater than in the treated population.

Identification	**Sponsorship source:** not reported **Declaration of interest:** The authors reported no conflicts of interest. **Country:** Italy **Setting:** NICU **Comments:** no comments **Authors name:** Gaia Francescato **Institution:** Neonatology and NICU, Department of Clinical Sciences and Community Health, Fondazione IRCCS Ca ` Granda – Ospedale Maggiore Policlinico, University of Milan  **Email:** gaia.francescato@mangiagalli.it or gaia.francescato@yahoo.it  **Address:** Gaia Francescato, Neonatology, Department of Clinical Sciences and Community Health, Fondazione IRCCS Ca` Granda – OspedaleMaggiore Policlinico, University of Milan, via Della Commenda 1220122, Milano
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	The participant was assigned on the basis of a list of randomisation to 1 of the 2 groups in the study.
Allocation concealment (selection bias)	Unclear risk	Unclear
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel

Tandoi 2015 *(Continued)*

Blinding of outcome assessment (detection bias) All outcomes	High risk	No blinding of outcome assessment
Incomplete outcome data (attrition bias) All outcomes	Low risk	36 participants enrolled and analysed
Selective reporting (reporting bias)	Unclear risk	36 participants enrolled and analysed
Other bias	Unclear risk	Unclear

Vastani 2017

Study characteristics	
Methods	**Study design:** randomised controlled trial **Study grouping:** parallel group
Participants	**Baseline characteristics** Control • *Gestational age at birth (weeks) mean ± SD:* 32.7 ± 2.17 • *Chronic age (days) mean ± SD:* 8.85 ± 5.62 • *Birth weight (g) mean ± SD:* 1847 ± 659.53 • *Gender male n (%)*: 9 (45) • *Mother's education level (university graduate) n (%):* 13 (65) • *Five-minute Apgar mean ± SD* 9.64 ± 0.78 • *Mother's age mean ± SD:* 28.95 ± 5.03 • N: 20 Live mother's voice • *Gestational age at birth (weeks) mean ± SD:* 32.6 ± 1.98 • *Chronic age (days) mean ± SD:* 7.45 ± 4.12 • *Birth weight (g) mean ± SD:* 1800 ± 302.56 • *Gender male n (%):* 10 (50) • *Mother's education level (university graduate) n (%):* 12 (60) • *Five-minute Apgar mean ± SD:* 9.91 ± 0.28 • *Mother's age mean ± SD:* 31.20 ± 5.64 • N: 20 Live strangers voice • *Gestational age at birth (weeks) mean ± SD:* 33.0 ± 2.15 • *Chronic age (days) mean ± SD:* 6.95 ± 4.09 • *Birth weight (g) mean ± SD:* 1771 ± 352.97 • *Gender male n (%):* 7 (35) • *Mother's education level (university graduate) n (%):* 11 (55) • *Five-minute Apgar mean ± SD:* 9.73 ± 0.59 • *Mother's age mean ± SD:* 30.55 ± 5.31 • N: 20

Vastani 2017 *(Continued)*

Inclusion criteria: gestational age of 28 to less than 37 weeks, birth weight of between 1000 and 2500 g, at least 3 days passed from the birth, being kept in incubators, having no congenital anomalies, physiological stability, using no respiratory devices such as nasal cannula or positive-pressure ventilation delivered through nasal/face mask and mechanical ventilator, receiving no muscle relaxants, undergoing no surgical operation, using no pacifier or feeding only with breast milk, healthy auditory system.

Exclusion criteria: intracranial haemorrhage, periventricular leukomalacia, cerebral palsy, necrotising enterocolitis, patent ductus arteriosus, sepsis, chronic lung disease, anaemia, addiction to alcohol, narcotic drugs and psychotropic substances in mother. Occurrence of seizure during the intervention, the mother's decision to withdraw from the study, the use of sedatives by the mother on the intervention day, the incidence of any problem in recording the physiological responses, and the infant's need for any medical or nursing intervention or any contact during the study.

Pretreatment: No significant differences were found between the three groups in terms of baseline demographics and clinical characteristics.

N infants analysed: 60

N infants randomised: 66

Reasons for dropouts: not reported

Reasons for exclusion: the data of 6 infants were omitted from analysis as they were outliers (the data showed excessive increase or decrease in heart rates insofar as it seemed that the pulse oximeter probes placed on the infant's legs were dislocated).

Sample size calculation: sample size was calculated using a pilot study on 2 groups of 5 infants. 1 group was the silent group and the other group received the mother's voice stimulation similar to the main study. The post-intervention mean of the neonatal heart rate was 1487.50 in the group that received the other's voice and 1435.75 in the silent group. Then by using an online software (i.e. https://www.statstodo.com/SSizAOV_Pgm.php), and considering β = 0.20, and α = 0.05, the number of groups in the analysis = 3, the largest difference between the 2 means = 5 and the expected background standard deviation = 6, the sample size for each of the three groups was estimated to be 22 infants.

Interventions

Intervention characteristics

Control

- *Intervention type*: silence but with the mother sitting next to incubator
- *Dose*: 30 minutes: 10 minutes pre, 10 minutes intervention, 10 minutes post
- *Frequency*: once a day for 3 consecutive days
- *Mode of delivery*: live next to incubator
- *dB level*: ambient noise level < 45 dB

Live mother's voice

- *Intervention type*: mother sings lullaby: God Night Kid (produced by the national radio of Iran) gently, rhythmically, and steadily
- *Dose*: 30 minutes (10 minutes silence, 10 minutes singing, 10 minutes silence)
- *Frequency*: once a day for three consecutive days
- *Mode of delivery*: live singing next to the incubator
- *dB level*: 60-70 dB

Live strangers voice

- *Intervention type*: nurse sings lullaby: God Night Kid (produced by the national radio of Iran) gently, rhythmically, and steadily
- *Dose*: 30 minutes (10 minutes silence, 10 minutes singing, 10 minutes silence)
- *Frequency*: once a day for 3 consecutive days
- *Mode of delivery*: live singing next to the incubator
- *dB level*: 60-70 dB

Vastani 2017 *(Continued)*

Outcomes	*Heart rate* • **Outcome type**: continuous outcome • **Reporting**: fully reported • **Direction**: lower is better • **Data value**: change from baseline • **Notes**: no significant difference was found between the mean heart rates of the 3 groups neither before (P > 0.588), nor 10 minutes after the intervention (P > 0.345). However, a significant difference was observed amongst the three groups during the intervention (P = 0.016). The maximum heart rate occurred in the mother's voice group and the minimum in the stranger's voice group. The Tukey's post-hoc test showed that the observed statistic difference was between the mother's voice group and the stranger's voice group (P = 0.012). In the repeated measures' analysis of variance, the results of Mauchly's sphericity test showed that the assumption of sphericity was violated (P < 0.001). As multivariate statistics do not need to establish the sphericity assumption, Greenhouse-Geisser correction was used for performing tests on the variable heart rate on 3 consecutive times within the 3 groups. Only in the mother's voice group, a statistically significant difference was observed in the 3 measurement times (P < 0.001); the infants' heart rates increased significantly during the intervention compared to their heart rates before and after the intervention. However, no significant differences were found in the within-subjects comparisons neither in the silent group (P > 0.711), nor in the stranger voice group (P > 0.399). Moreover, an interaction was found between the groups and the time in the silent and the mother's voice groups (P < 0.001).
Identification	**Sponsorship source:** not reported **Declaration of interest:** none declared **Country:** Iran **Setting:** NICU **Comments:** no comment **Authors name:** Seyedeh Roghayeh Jafarian Amiri, **Institution:** Department of Nursing, Babol University of Medical Sciences, Babol, IR Iran **Email:** Mozhganahmadivastani@gmail.com **Address:** Iran, Mazandaran, BabolGanjafrooz Street, University of Medical Sciences **Website:** http://en.mubabol.ac.ir/page/nuesing-midwife.htm
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	The samples were randomly allocated into the 3 groups of 22 (silent, mother's voice and stranger's voice) by computer randomisation.
Allocation concealment (selection bias)	Unclear risk	Unclear
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel

Vastani 2017 *(Continued)*

Blinding of outcome assessment (detection bias) All outcomes	Low risk	Moreover, the patient monitoring system was put in an automatic mode. The monitoring system and the digital scale used to measure the variables for all the infants were the same.
Incomplete outcome data (attrition bias) All outcomes	Low risk	No incomplete outcome data detected
Selective reporting (reporting bias)	Unclear risk	Unclear
Other bias	Low risk	An inter-rater reliability method was used to evaluate the reliability of the instruments (the digital scale and the patient monitoring system). To evaluate the reliability of the digital scale, 2 colleagues were trained and then were asked to weigh 10 infants and record the results in 2 distinct charts. Then the correlation coefficient was calculated, and the result was approximately equal to 1 (r = 0.97) for the digital scale. Moreover, the patient monitoring system was put in an automatic mode. The monitoring system and the digital scale used to measure the variables for all the infants were the same.

White-Traut 1988

Study characteristics	
Methods	**Study design:** randomised controlled trial **Study grouping:** parallel group
Participants	**Baseline characteristics** Routine care • *Gestational age at birth (weeks) mean*: 33.18 • *Postnatal Complications Scale mean*: 6.91 • *Birth weight (g) mean:* 1864.55 • *Gender male/female*: 5/6 • *Age mother mean*: 20.82 • *SES-Socioeconomic status mean*: 2.09 • *Education (years) mother mean*: 12.73 • *Race (black/white)*: 10/1 Live talking mother • *Gestational age at birth (weeks) mean*: 34.18 • *Postnatal Complications Scale mean*: 8.18 • *Birth weight (g) mean:* 1956.18 • *Gender male/female*: 5/6 • *Age mother mean*: 21.82 • *SES-Socioeconomic status mean*: 1.45 • *Education (years) mother mean*: 12.73 • *Race (black/white)*: 10/1 RISS mother • *Gestational age at birth (weeks) mean*: 33.91 • *Postnatal Complications Scale mean*: 7.27

White-Traut 1988 *(Continued)*

- *Birth weight (g)mean:* 1949.09
- *Gender male/female*: 6/5
- *Age mother mean*: 22.18
- *SES-Socioeconomic status mean*: 1.73
- *Education (years) mother mean*: 12.45
- *Race (black/white)*: 11/0

Inclusion criteria: maternal inclusion criteria included the ability to speak and understand English, vaginal delivery of the infant, and an age of at least 16 years. Infant inclusion criteria included gestational age between 28 and 35 weeks, birth weight appropriate for gestational age, and absence of ventilatory assistance by 24 hours after delivery.

Exclusion criteria: infants suspected or diagnosed with chromosomal abnormalities were excluded. Women who experienced multiple births were not eligible to participate.

Pretreatment: there were no significant differences amongst the groups on either maternal or infant variables prior to the intervention.

N infants analysed: n = 33; n = 11 routine care, n =11 talking, n = 11 RISS

N infants randomised: 33 mother-infant pairs

Reasons for dropouts: not reported

Reasons for exclusion: not reported

Sample size calculation: not reported

Interventions

Intervention characteristics

Routine care

- *Intervention type*: routine nursery care; including a consistent feeding schedule of feedings every 3 or 4 hours, handling for procedures, and periodic (occasional) opportunities for non-nutritive sucking. Parental visiting was encouraged, and open visiting hours were in effect. Family members and staff often placed visual stimuli in the incubator or bassinet. Primary nursing was in effect. In addition, mothers assigned to the routine care group received a didactic discussion on infant clothing for the premature infant.

Live talking mother

- *Intervention type*: unstructured talking treatment; to talk or sing to their infants for 15 minutes at the specified time intervals. Mothers administered the talking treatment according to a uniform protocol in a controlled setting (infant in the incubator or under the infant warmer). Group received a didactic discussion on infant clothing for the premature infant.
- *Dose*: 15 minutes
- *Frequency*: during the following post-birth periods: 24 to 36, 37 to 48, 49 to 60, and 61 to 72 hours after birth
- *Mode of delivery*: live by mother: Mothers administered the talking treatment according to a uniform protocol in a controlled setting (infant in the incubator or under the infant warmer). 3 mothers insisted on holding their infants for the talking treatment, and some other mothers engaged in some touching of their infants.
- *dB level*: not reported

RISS mother

- *Intervention type*: RISS infant massage technique; the technique provided tactile contact, vestibular motion, auditory stimulation and eye-to-eye contact (visual stimulation) and promoted active interaction between mother and infant. These mothers did not receive the presentation on infant clothing.
- *Dose*: as a sequential cephalocaudal progression of massaging the infant's body for 10 minutes followed by rocking (vestibular stimulation) for 5 minutes
- *Frequency*: during the following post-birth periods: 24 to 36, 37 to 48, 49 to 60, and 61 to 72 hours

White-Traut 1988 *(Continued)*

- *Mode of delivery*: live by mother: administered as a sequential cephalocaudal progression of massaging the infant's body for 10 minutes followed by rocking (vestibular stimulation) for 5 minutes
- *dB level*: not reported

Outcomes	*The Nursing Child Assessment Feeding Scale* • **Outcome type**: continuous outcome • **Reporting**: partially reported • **Scale**: the Nursing Child Assessment Feeding Scale • **Range**: 0-26 • **Direction**: higher is better • **Data value**: endpoint • **Notes**: 1-way analysis of variance amongst the 3 groups was computed for the total infant behaviour raw score. Significant differences were found. A Scheffé analysis did not reveal significant differences between the groups, indicating that infant behaviour was significantly altered by the maternally administered intervention, but not different between the Talking and RISS Groups. Based on the Scheffé results, individual group comparisons were not examined for these data. The 2 subscales of infant behaviour were analysed separately by 1-way analyses of variance. Significant differences were not identified for the clarity of infant cues subscale or the infant's responsiveness to his/her parent subscale. *Maternal behaviour* • **Outcome type**: continuous outcome • **Reporting**: partially reported • **Scale**: Maternal Behaviour Scores • **Range**: 0-50 • **Direction**: higher is better • **Data value**: endpoint • **Notes**: There were significant differences amongst the 3 groups on the total maternal behaviour raw score. A follow-up Scheffé analysis indicated significant differences between the Talking Group and the RISS Group for maternal behaviour. Separate analyses of the 4 individual subscales of maternal behaviour also were conducted. Significant differences via a 1-way analysis of variance were identified for the sensitivity to infant cues' subscale and the cognitive growth fostering subscale.
Identification	**Sponsorship source:** supported in part by a grant from the Schweppes Foundation **Country:** USA **Setting:** single centre; Special Care nursery, Department of Pediatrics, St. Luke's Medical Center, Chicago **Comments:** no comment **Authors name:** Rosemary C. White-Traut **Institution:** College of Nursing at the University of Illinois-Chicago **Email:** rwt@uic.edu **Address:** Dr. Rosemary White-Traut, University of Illinois- Chicago, 845 S. Damen Street, Chicago, IL 60612 **Website:** https://today.uic.edu/nurse-guides-new-moms-of-preemies & https://today.uic.edu/nurse-guides-new-moms-of-preemies
Notes	
Risk of bias	

White-Traut 1988 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Unclear
Allocation concealment (selection bias)	Unclear risk	Unclear
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	High risk	No blinding of outcome assessment
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	No dropouts or exclusions reported
Selective reporting (reporting bias)	Unclear risk	Unclear
Other bias	Low risk	To obtain certification, an inter-rater agreement score of 85% or greater had to be maintained for five observations of parent and infant behaviour scores. Reliability in scoring the feeding interactions was determined prior to assessments of feeding interactions and during the data collection period.

Wirth 2016

Study characteristics	
Methods	**Study design:** randomised controlled trial **Study grouping:** parallel group
Participants	**Baseline characteristics** Control • *Gestational age at birth (weeks) median (range)*: 33.4 (31.3-35.0) • *Age at start of intervention (days) median (range)*: 4 (1-9) • *Birth weight (g) mean (range)*: 1990 (995–2495) • *Gender n (%) male*: 9 (57) • *Apgar score (5 min) median (range)*: 8 (5-10) • *SGA n (%)*: 4 (19) • *Prenatal steroids n (%)*: 10 (48) • *Associated disorder n (%)*: 0 (0) • *Continuous positive airway pressure n (%)*: 3 (15) Recorded lullabies • *Gestational age at birth (weeks) median (range)*: 33.3 (30.0-36.3) • *Age at start of intervention (days) median (range)*: 5 (2-9) • *Birth weight (g) median (range)*: 1732 (870–3050)

Wirth 2016 *(Continued)*

- *Gender n (%) male*: 12 (60)
- *Apgar score (5 min) median (range)*: 8 (3-10)
- *SGA n (%)*: 4 (20)
- *Prenatal steroids n (%)*: 9 (45)
- *Associated disorder n (%)*: 2 (10)
- *Continuous positive airway pressure n (%)*: 3 (15)

Recorded mothers voice

- *Gestational age at birth (weeks) median (range)*: 32.0 (30.0-35.0)
- *Age at start of intervention (days) median (range)*: 6 (2-9)
- *Birth weight (g) median (range)*: 1685 (740–2895)
- *Gender n (%) male*: 12 (60)
- *Apgar score (5 min) median (range)*: 8 (5-10)
- *SGA n (%)*: 6 (30)
- *Prenatal steroids n (%)*: 12 (60)
- *Associated disorder n (%)*: 3 (15)
- *Continuous positive airway pressure n (%)*: 3 (15)

Inclusion criteria: preterm infants in a stable condition with a gestational age of 30 to 36 completed weeks and a postnatal age of fewer than 10 days

Exclusion criteria: (a) lack of parental consent, (b) expected hospitalisation of fewer than 14 further days, (c) artificial ventilation, but no continuous positive airway pressure, (d) treatment with catecholamines, (e) cardiac anomalies or diseases, (f) congenital malformations, and (g) hearing loss, confirmed by auditory brainstem-evoked potentials

Pretreatment: the 3 groups did not differ significantly with respect to the infants' baseline characteristics, except postnatal age at the start of the study.

N infants analysed: 61

N infants randomised: 62 preterm infants, 20 were randomly assigned to the 'Lullabies group', 20 to the 'Maternal voice group' and 22 to the 'control group'.

Reasons for dropouts: 9 infants dropped out: in 1 infant of the control group, parents withdrew consent, 1 infant died from necrotising enterocolitis and 7 infants were discharged before completion of the study period. Because of technical problems, data collection in 2 infants was incomplete.

Reasons for exclusion: 46 infants were excluded for the following reasons: (a) lack of parental consent (n = 24), (b) expected short hospitalisation (n = 20), (c) admission at # 10 postnatal days (n = 1) and (d) congenital malformation (n = 1).

Sample size calculation: no sample size calculation reported. For this pilot study, an empirical number of 20 infants per group was assessed.

Interventions	**Intervention characteristics**

Control

- *Intervention type*: standard NICU environmental sounds

Recorded lullabies

- *Intervention type*: recorded lullabies in a standardised manner using a commercially available CD with classical melodies sung by females/opera singer ('Wiegenlieder Vol. 1', Carus-Verlag, Stuttgart 2009)
- *Dose*: for 30 minutes between 2000 and 2100 hours, about 30 to 60 minutes after feeding and in the absence of their parents
- *Frequency*: on 14 consecutive days
- *Mode of delivery*: via speaker, 30 to 60 minutes after feeding and in the absence of their parents, in the incubator or the cot

Wirth 2016 *(Continued)*

- *dB level*: 55-65

Recorded mothers voice

- *Intervention type*: recorded mothers reading of a chapter from the German book 'Der kleine Prinz' (English:'The Litte Prince') by Antoine de Saint-Exupéry (Karl Rauch Verlag, Düsseldorf 1950 and 1998). After reviewing the quality of the recording, the reading was transcribed to a CD.
- *Dose*: for 30 minutes between 2000 and 2100 hours, about 30 to 60 minutes after feeding and in the absence of their parents
- *Frequency*: on 14 consecutive days
- *Mode of delivery*: via speaker, 30 to 60 minutes after feeding and in the absence of their parents, in the incubator or the cot
- *dB level*: 55-65

Outcomes	*Heart rate* • **Outcome type**: continuous outcome • **Reporting**: partially reported • **Direction**: lower is better • **Data value**: change from baseline • **Notes**: whereas heart rate did not significantly change during the daily study period in the control group, it significantly decreased in both intervention groups during the interventions from baseline before the intervention. A significant difference between both intervention groups was not observed. Heart rate was significantly lower after stimulation than before stimulation in both intervention groups. No significant difference was observed between the two intervention groups. The decrease in heart rate during stimulation and after stimulation was related to gestational age and was more pronounced in more mature infants. A significant correlation of the change in heart rate and the infants' postnatal age at entering the study could not be observed. *Respiratory rate* • **Outcome type**: continuous outcome • **Direction**: lower is better • **Data value**: change from baseline • **Notes**: respiratory rate decreased during and after acoustic stimulation as well: however, a significant difference between the 2 intervention groups was not found.
Identification	**Sponsorship source:** KIM e.V., Marburg, Germany, financial support **Country:** Germany **Setting:** single centre, NICU or newborn special care unit, Children's Hospital, Philipps University Marburg, Germany **Comments:** no comments **Authors name:** Lara Wirth **Institution:** Children's Hospital, Philipps University Marburg, Marburg, Germany **Email:** Lara.Wirth@gmx.de **Address:** L Wirth, Children's Hospital, Philipps University Marburg, Baldingerstrasse, Marburg D-35033, Germany **Website:** https://www.ukgm.de/ugm_2/deu/umr_lpl/27708.html
Notes	
Risk of bias	

Wirth 2016 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Use of the randomisation program R 2.15.0. Block randomisation in a random block of 6 (2 lullabies, 2 maternal voice, 2 control). Safekeeping the randomisation results was done by sealed and consecutive numbered envelopes, together with the pseudonym code. Allocation of the included participants according to the numbering of the envelopes
Allocation concealment (selection bias)	Low risk	Infants were randomly assigned to 1 of 2 intervention groups or a control group by opening consecutive sealed envelopes.
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	High risk	No blinding of outcome assessment
Incomplete outcome data (attrition bias) All outcomes	Low risk	All analyses were performed on the intention-to-treat population.
Selective reporting (reporting bias)	Unclear risk	Unclear
Other bias	Unclear risk	Unclear

Yu 2021

Study characteristics	
Methods	**Study design:** randomised controlled trial **Study grouping:** parallel group
Participants	**Baseline characteristics** Control heel stick procedure • *Gestational age at birth (weeks) mean/± SD*: 33.16 ± 5.09 • *Birth weight (g) mean/± SD*: 1860.31 ± 474.84 • *Gender n (%) male*: 14 (37.5%) • *AP1 mean/± SD*: 8.13 ± 0.83 • *AP5 mean/± SD*: 9.22 ± 0.55 • *Neonatal Medical Index mean/± SD*: 2.78 ± 1.64 • *Mother age mean/± SD*: 33.16 ± 5.09 • *Parity mean/± SD*: 1.59 ± 0.76 • *Education Junior n (%)*: 1 (3.1%) • *Education University n (%):* 20 (62.5%) • *Experience of preterm Yes n (%)*: 5 (15.63%) Recorded maternal voice before, during & after heel stick procedure • *Gestational age at birth (weeks) mean/± SD*: 31.97 ± 5.67

Yu 2021 *(Continued)*

- *Birth weight (g) mean/± SD*: 861.53 ± 543.5
- *Gender n (%) male*: 17 (53.1%)
- *Neonatal Medical Index mean/± SD*: 3.09 ± 1.35
- *Mother age mean/± SD*: 31.97 ± 5.66
- *Parity mean/± SD*: 1.5 ± 0.57
- *Education Junior n (%):* 2 (6.3%)
- *Education University n (%)*: 20 (62.5%)
- *Experience of preterm Yes n (%)*: 2 (6.25%)

Overall

- *Gestational age at birth (weeks) mean/± SD:* 32.56 ± 5.37
- *Birth weight (g) mean/± SD:* 1860.92 ± 506.26
- *Gender n (%) male*: 31 (48.4%)
- *Neonatal Medical Index mean/± SD:* 2.94 ± 1.5
- *Mother age mean/± SD:* 32.56 ± 5.37
- *Parity mean/± SD:* 1.55 ± 0.67
- *Education Junior n (%):* 3 (4.7%)
- *Education University n (%):* 40 (62.5%)
- *Experience of preterm Yes n (%)*: 7 (10.9%)

Inclusion criteria: all infants with a gestational age of less than 37 weeks who received routine care every 8 hours to determine their blood glucose level through the heel sticks and who needed to be in an incubator on the 4th postnatal day were eligible to participate in this study.

Exclusion criteria: congenital or chromosomal abnormalities, had congenital or acquired infections, were connected to a high-frequency oscillatory ventilator, or were administered a sedative or inotropic agents. Premature infants were also excluded if their mothers smoked, consumed alcohol, or used illegal drugs during pregnancy or if the mothers were transferred to an intensive care unit rather than the postnatal ward (recordings were created only in the postnatal ward).

Pretreatment: no significant differences were found between the control and intervention groups in terms of gender; length of gestation; birth body weight; Appearance, Pulse, Grimace, Activity, and Respiration score; Neonatal Medical Index; or mother's age, education level, or preterm birth experience (P > 0.05). These results indicated that the intervention and control groups were similar in terms of the studied variables.

N infants analysed: 64

N infants randomised: 64

Reasons for dropouts: during the study, 1 infant in the intervention group was dressed and removed from the incubator on the 3rd day of the study, and 2 infants in the control group did not have to undergo the heel sticks on the third day. This data loss, however, had a negligible effect on the data analysis. All the infants passed the hearing test before discharge, and none of them had congenital deafness. The 2 groups, each consisting of 32 preterm infants, were included in the data analysis.

Reasons for exclusion: no

Sample size calculation: the sample size was calculated using G*Power analysis for repeated measures. The estimated median effect size was 0.15 with an α value of 0.05 and power of 0.8.

Interventions	**Intervention characteristics** Control heel stick procedure • *Intervention type*: the infants in both groups received containment and non-nutritive suckling after the procedure if they cried for > 1 minute. • *Frequency*: 3 consecutive days

Yu 2021 *(Continued)*

Recorded maternal voice before, during & after heel stick procedure

- *Intervention type*: the content of the maternal voice recordings was the mother's reading of a children's book presenting the refections of a premature infant. In this text, premature infants are compared with immature persimmons who are nurtured by their family through patience and affection while undergoing self-exploration and passing through all difficulties to eventually become sweet and mature fruits. The mother also said some words that they wanted their baby to hear. The recordings were edited into a 13-min-long audio file by using the music editing software Sound Organiser 2. The infants in both groups received containment and non-nutritive sucking after the procedure if they cried for > 1 minute.
- *Dose*: 13 minutes
- *Frequency*: once a day for 3 consecutive days
- *Mode of delivery*: 3 minutes before a heel stick procedure and lasting until the procedure completion, 13 minutes in total. During playback in the incubator, a decibel meter was used to ensure that the maternal voice intensity was between 50 and 60 dB so that the sound would not harm the infant. During these 3 days, the research processes for the intervention and control groups were recorded using a camera, and the videos were uploaded to YouTube within 24 hours of recording, with the video URLs sent to the mother using a private link. The purpose of providing the mothers with the videos was to facilitate mother–infant bonding and to offer the mothers a visualisation of the study process. A Google questionnaire was administered to the mother on the 7th postnatal day.
- *dB level*: 50-60 dB

Outcomes

Heart rate

- **Outcome type**: continuous outcome
- **Direction**: lower is better
- **Data value**: endpoint

Respiratory rate

- **Outcome type**: continuous outcome
- **Direction**: lower is better
- **Data value**: endpoint

Oxygen saturation

- **Outcome type**: continuous outcome
- **Direction**: higher is better
- **Data value**: endpoint

Pain status control

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Scale**: 2
- **Range**: 0-7
- **Direction**: lower is better
- **Data value**: endpoint

Mother–Infant Bonding Inventory

- **Outcome type**: continuous outcome
- **Reporting**: fully reported
- **Scale**: 6-point Likert scale
- **Range**: 19-114
- **Direction**: lower is better
- **Data value**: endpoint
- **Notes**: The Mother–Infant Bonding Inventory was used to measure the mother's thoughts, feelings, and commitment towards her baby. The scale consists of four factors: proximity, parental adjustment,

Yu 2021 *(Continued)*

commitment, and confidence of reciprocity, which have 7, 6, 6 and 6 items, respectively (25 items in total). Each item was scored using a 6-point Likert scale, with the scores 1 –6 corresponding to strongly disagree, disagree, slightly disagree, slightly agree, agree, and strongly agree, respectively. The parental adjustment subscale consisted of negative items, whereas the other 3 subscales consisted of positive items, with higher scores indicating better bonding. The parental adjustment subscale was left unanswered when the hospitalised preterm infants had not yet been in close contact with or received care from its mother. The other 3 subscales comprised 19 items, and the total score varied between 19 and 114. The Cronbach's α of the aforementioned inventory was 0.88–0.89.

Identification	**Sponsorship source:** Research Support Scheme of the Chang Gung Memorial Hospital, Linkou, Taiwan, grant no. CMRP20190320210 **Country:** Taiwan **Setting:** NICU **Comments:** **Authors name:** Wan-Chen Yu **Institution:** Department of Nursing, Chang Gung Memorial Hospital, Taoyuan, Taiwan **Email:** chiwenchen@nycu.edu.tw (C.-W. Chen), redmapleyu@gmail.com(W.-C. Yu), newborntw@gmail.com (M.-C. Chiang), kuanchia@nycu.edu.tw (K.-C. Lin), m22096@cgmh.org.tw (C.-C. Chang), d22038@cgmh.org.tw (K.-H. Lin) **Address:** Department of Nursing, Chang Gung Memorial Hospital, Taoyuan, Taiwan **Website:**
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Field block randomisation corresponding to individual preterm infants
Allocation concealment (selection bias)	Low risk	Allocation concealment was used to randomly divide the enrolled infants into two groups.
Blinding of participants and personnel (performance bias) All outcomes	High risk	No blinding of participants and personnel
Blinding of outcome assessment (detection bias) All outcomes	Low risk	The outcome assessment of the premature infants was double-blinded.
Incomplete outcome data (attrition bias) All outcomes	Low risk	No incomplete outcome data detected
Selective reporting (reporting bias)	Low risk	No selective reporting detected
Other bias	Unclear risk	The purpose of providing the mothers with the videos was to facilitate mother–infant bonding and to offer the mothers a visualisation of the study process. A Google questionnaire was administered to the mother on the 7th

Yu 2021 *(Continued)*	postnatal day. This questionnaire contained the Mother–Infant Bonding Inventory. Finally, the Mother–Infant Bonding Inventory data were collected by the researcher.

ANOVA: analysis of variance**; APIB:** Assessment of Preterm Infants' Behavior; **Bpm**: beats per minute;**BW**: birth weight; **CD**: Compact Disc; **CG:** control group**; CGA:** corrected gestational age;**CI**: confidence interval; ; **cm**: centimetre; **CMT**: creative music therapy; **CPAP**: continuous positive airway pressure; **dB level**: decibel level; **EPDS**: Edinburgh Postnatal Depression Scale; **g**: gram;**GA**: gestational age; **HF**: high-frequency; **HR**: heart rate; **HRV**: heart rate variability;**IGI**: intervention group 1; **IG2**: intervention group 2; **IVH**: intraventricular haemorrhage; **kg**: kilogram; **LBW**: low birth weight infants; **LF**: low-frequency; **M**: means; **Mdn:** median; **MIBS**: mother-to-infant-bonding scale; **MRI**: magnetic resonance imaging ; **MT:** music therapy; **MTG**: music therapy group; **n**: number; **n.a**.: not available; **NICU**: neonatal intensive care unit; **NIDCAP**: Newborn Individualised Developmental Care and Assessment Program; **NISS**: Neonatal Infant Stressor Scale;**PCQ**: Parental Competences Questionnaire; **O_2**:Oxygen; **PIPP**: Premature Infant Pain Profile; **PPI**: Parental Perception Inventory; **PRAM**: Pictorial Representation of Attachment Measure;**PSS**: Parental Stressor Scale; **PSQ**: Parental Stress Questionnaire; **PVL**: periventricular leukomalacia; **RBL**: Rhythm Breath and Lullaby; **RCT**: randomised controlled trial; **RISS**: massage, talking, eye contact and rocking;**RR**: respiratory rate; **SD**: standard deviation; **SES**: socioeconomic status; **SGA**: Small for gestational age;**SKD:** State Scale; **SPO_2**: oxygen saturation; **SSC**: skin-to-skin-contact: **STAI**: State-Trait Anxiety Inventory; **STAI-SKD**: State-Anxiety Inventor ; **TOS**: musical track recording obtained with several specific audio equipment; **UMCG**:University Medical Center Groningen ; **URL**: Uniform Resource Locator; **wk**: weeks

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Alipour 2013	Due to intervention (frequency of intervention only once)
Amini 2013	Due to intervention (frequency of intervention only twice)
Arnon 2006	Due to intervention (frequency of intervention only once)
Arnon 2014	Due to intervention (frequency of intervention only once)
Auto 2013	Due to comparator (no standard care comparison)
Badr 2017	Due to intervention (frequency of intervention only once)
Barandouzi 2020	Due to intervention (frequency of Intervention only once)
Bellieni 2001	Due to comparator (additional intervention to music/voice provided; not in both compared arms)
Bergomi 2014	Due to intervention (frequency of intervention only once)
Castano 2021	Due to intervention (frequency of intervention only once)
Cavaiuolo 2015	Due to intervention (frequency of intervention only once)
Cevasco 2005	Due to intervention (no continuous music duration comprising at least 5 minutes)
Chirico 2017	Due to study design
Chorna 2018	Due to study design
Corrigan 2020	Due to intervention (frequency of intervention only once)
Corrigan 2021	Due to intervention (frequency of intervention only once)

Study	Reason for exclusion
Dalili 2020	Due to intervention (combined with other interventions)
Dearn 2014	Due to intervention (frequency of intervention only once)
Detmer 2020	Due to study design
Doheny 2012	Due to study design
Döra 2021	Due to intervention (frequency of intervention only once)
Edraki 2017	Due to intervention (freqeuncy of intervention only once)
Efendi 2018	Due to intervention (frequency of intervention only twice)
Eskandari 2012	Due to intervention (frequency of intervention only once)
Filippa 2021	Due to intervention (frequency of intervention only once)
Garunkstiene 2014	Due to intervention (frequency of intervention only once)
Holsti 2019	Due to intervention (frequency of intervention only once)
Jie 2017	Due to intervention (combined with other intervention)
Johnston 2009	Due to intervention (frequency of intervention only once)
Kahraman 2020	Due to intervention (frequency of intervention only once)
Kanagasabai 2013	Due to intervention (combined with other interventions)
Karadag 2021	Due to intervention (frequency of intervention only once)
Karimi 2020	Due to patient population (music for mothers was not in relation to their infant)
Keith 2009	Due to patient population
Kucuk Alemdar 2018	Due to intervention (frequency of intervention only once)
Kurdahi Badr 2017	Due to intervention (frequency of intervention only once)
Majidipour 2018	Due to intervention (frequency of intervention only once)
Picciolini 2014	Due to study design
Pouraboli 2019	Due to intervention (frequency of intervention only once)
Qolizadeh 2019	Due to intervention (frequency of intervention only once)
Ramezani 2020	Due to intervention (frequency of intervention only once)
Ranger 2018	Due to intervention (frequency of intervention only once)
Sarhangi 2021	Due to patient population
Schlez 2011	Due to intervention (frequency of intervention only once)

Study	Reason for exclusion
Shabani 2016	Due to intervention (frequency of intervention only once)
Shah 2017	Due to comparator (additional intervention to music/voice provided not in both compared arms)
Shukla 2018	Due to intervention (frequency of intervention only once)
Standley 1995	Due to study design
Standley 2000	Due to study design
Standley 2003	Relevant outcomes were not available because they had not been measured.
Standley 2010	Due to intervention (no continuous music duration comprising at least 5 minutes)
Stokes 2018	Due to intervention (frequency of intervention only twice)
Tang 2018	Due to intervention (frequency of intervention only once)
Tekgunduz 2019	Due to intervention (frequency of intervention only once)
Uematsu 2019	Due to comparator (additional intervention to music/voice provided not in both compared arms)
Ullsten 2017	Due to intervention (frequency of intervention only once)
Vahdati 2017	Due to intervention (frequency of intervention only once)
Walworth 2012	Due to comparator (additional intervention to music/voice provided not in both compared arms)
White-Traut 2004	Due to intervention (combined with other interventions)
Wu 2020	Due to intervention (frequency of intervention only once)
Yildiz 2011	Due to study design
Yilmaz 2021	Due to patient population
Zeraati 2018	Due to comparator (additional intervention to music/voice provided not in both compared arms)

Characteristics of ongoing studies *[ordered by study ID]*

Aita 2021

Study name	Nurturing and quiet intervention on preterm infants' neurodevelopment and maternal stress and anxiety: protocol of a pilot randomised clinical trial
Methods	RCT
Participants	26 weeks to 32 weeks
Interventions	SSC session lasting 2 hours during the day 4 times/week including a 15-minute of auditory stimulation with maternal voice and controlled levels of NICU light and noise followed by a 1-hour quiet period where infants will rest in their incubator/crib with olfactory stimulation and where the control of light and noise levels will be continued.

Aita 2021 *(Continued)*

Outcomes	Feasibility and acceptability of the nurturing and quiet intervention and the study procedures as assessed by a self-completed questionnaire (mothers) and a logbook (time frame: 1 year) Questionnaire completed by mothers - each question treated separately (no total score) Log book completed by a research assistant
Starting date	2020
Contact information	Contact: Marilyn Aita, PhD 5143436111 ext 51473 marilyn.aita@umontreal.ca
Notes	

Bonjorn Juarez 2020

Study name	Reduction of visual and auditory stimuli to reduce pain during venipuncture in premature infants. A randomised controlled trial
Methods	RCT
Participants	• Premature babies born between 32 and 36 weeks of gestation (both included)
Interventions	• Procedure: stimuli reduction Phototherapy goggles and earmuffs will be placed 3 minutes before the venipuncture (leaving the patient in resting state after the manipulation) and will be maintained during the procedure. Monitor alarms and devices will be silenced and will remain silenced and noise in the unit will be minimised during the procedure. • Procedure: usual care Babies in the control group will receive physical contention with administration of sucrose two minutes before carrying out the venipuncture procedure (usual care). Venipuncture will be performed with 22G extraction needles, or peripheric venous catheter. During the puncture, the eyes will not be covered, and monitor alarms and devices not be silenced.
Outcomes	Change in the Premature Infant Pain Profile score (time frame: from baseline pain determination just before venipuncture to 30 seconds after venipuncture) Calculation of pain response using a validated instrument (Premature Infant Pain Profile). It varies from "0" (no pain) to 21 (maximum pain response)
Starting date	2019
Contact information	Contact: Sergio Alonso-Fernández, RN 0034934978437 salonso.germanstrias@gencat.cat Contact: Maria Bonjorn-Juarez, RN 0034678116280 mariabonjornjuarez@gmail.com

Bonjorn Juarez 2020 *(Continued)*

Notes	

Brignoni-Perez 2021

Study name	Listening to mom in the NICU: effects of increased maternal speech exposure on language outcomes and white matter development in infants born very preterm
Methods	RCT
Participants	24-31 weeks 6/7days' gestational age
Interventions	Increased maternal speech exposure, accomplished by playing audio recordings of each baby's own mother reading a children's book via an iPod placed in their crib/incubator
Outcomes	Measures of expressive and receptive language skills, obtained from a parent questionnaire collected at 12-18 months
Starting date	
Contact information	
Notes	

ChiCTR1900026017

Study name	A randomised controlled trial for investigation of the associations between a music programme and neonatal health outcomes
Methods	Interventional study
Participants	28–37 weeks
Interventions	1: normal music; 2: soft music
Outcomes	Body weight gain
Starting date	2019
Contact information	Cheng Shuk Man Address: Ward 6HK, Department of Paediatrics, 6/F, Prince of Wales Hospital, 30-32, Ngan Shing Street, Shatin No postcode Telephone: +852 35054179 Email: dickygabriel@yahoo.com.hk Affiliation:

ChiCTR1900026017 *(Continued)*

	Prince of Wales Hospital, Chinese University of Hong Kong
Notes	

CTRI/2012/11/003117

Study name	A clinical trial to study the effectiveness of instrumental Indian classical music on serum cortisol concentrations of preterm infants on assisted ventilation admitted to a tertiary level neonatal intensive care unit (NICU)
Methods	RCT
Participants	Preterm infants
Interventions	Music refers to the playing of recorded Indian classical instrumental music, which has soft pitch (shruthi shuddi-frequency) of sound waves or acoustical energy (Hz), set at a volume of 55-60 dB (decibels) of sound pressure level which is the comfortable hearing threshold for preterm infants which will be checked and preset for the headphones, amplifiers and music player by using a Sound Level Meter (Hand Held Analyser type 2250, Bruel and Kjaer, Denmark) which is calibrated using a Sound Level Calibrator 4231 (Class 1) (Bruel and Kjaer, Denmark) using a weighted scale and played to the preterm neonate using headphones alternating with 30 minutes of quiet period for every 25 minutes 40 seconds in the first music session and 40 minutes 20 seconds in the second music session, i.e. a total of 1 hour/day during the late evening for a total of 2 weeks, starting from 3rd day of life after collecting basal level of serum cortisol
Outcomes	1. The main outcome variable will be to determine the effectiveness of instrumental Indian classical music on serum cortisol concentrations of very preterm infants on assisted ventilation in the 1st and 2nd week of life. 2. To find the effectiveness on the developmental responses as measured by weight gain, head circumference and vital signs
Starting date	
Contact information	
Notes	Recruitment started in 2012 and was estimated for three years but no further publication found.

CTRI/2016/06/007028

Study name	Comparison of music therapy with kangaroo mother care for pain reduction in premature babies
Methods	Randomised controlled trial
Participants	Premature infants from gestational age of 26-36 weeks
Interventions	Intervention1: Music therapy: flute music Intervention 2: Kangaroo Mother Care: skin-to-skin contact between the mother and the baby during the procedure Control Intervention 1: Music therapy: soft flute music 55 dB during the procedure
Outcomes	Pain assessment by measuring Premature Infant pain Profile score time point: (1) Score the behavioural state before the potentially painful event by observing the infant for 15 seconds.

CTRI/2016/06/007028 *(Continued)*

	(2) Record the baseline heart rate and oxygen saturation. (3) Observe the infant for 30 seconds immediately following the painful event. Score physiologic and facial changes seen during this time and record immediately.
Starting date	2016
Contact information	Somashekhar M Nimbalkar Address: Department of Paediatrics, Pramukhswami medical College, Karamsad, 388325 Anand, GUJARAT India Telephone: 09825087842 Email: somu_somu@yahoo.com Affiliation: Department of Paediatrics, Pramukh swami Medical College
Notes	

CTRI/2017/04/008395

Study name	Effect of music therapy on breast milk production in mothers practising Kangaroo Mother Care
Methods	Randomised controlled trial
Participants	Mothers who delivered between 28-32 weeks of gestation
Interventions	Intervention 1: Music therapy: women in the Kangaroo Mother Care ward will listen to music prior to and during breast milk expression
Outcomes	Breast milk volume expressed Time point: Day 5 of intervention
Starting date	2017
Contact information	Dr Latha Sashi Address: 4-1-1230, Bogulkunta, Near Abids 500001 Hyderabad, ANDHRA PRADESH India Telephone: 9848135621 Email: drlatha_s@fernandezhospital.com Affiliation:

CTRI/2017/04/008395 *(Continued)*

	Fernandez Hospital
Notes	

CTRI/2021/06/033970

Study name	A study to check whether music combined with standard Kangaroo Mother Care help in improving neurodevelopmental outcome of preterm infants
Methods	Randomised controlled trial
Participants	Preterm neonates (< 37 weeks) with birth weight between 2 and 2.5 kg
Interventions	Intervention 1: KMC plus instrumental music: mother infant dyads who will be practising KMC with instrumental music for 5 days with play time being a minimum of 90 minutes per day. Instrumental music selected is the Instrumental version of a popular lullaby song (Omaan Thingal Kidavo) Control Intervention 1: KMC: mother infant dyads who will be practising KMC alone for 5 days
Outcomes	Neurodevelopmental outcome using Developmental Assessment Scale in Indian Infants. Time point: Neurodevelopmental outcome will be measured at 6 months of age.
Starting date	2021
Contact information	Name: Dr Adhisivam B Address: Dept of Neonatology, JIPMER, Puducherry 605006 Pondicherry, PONDICHERRY India Telephone: 9488822113 Email: adhisivam1975@yahoo.vo.uk Affiliation: JIPMER Pondicherry
Notes	

CTRI/2021/07/035287

Study name	Effects of vestibular stimulation on neurodevelopment in preterm infants
Methods	Randomised controlled trial
Participants	28-36 weeks of gestation
Interventions	Intervention 1: (experimental): 1. Tactile stimulation daily, each session 5 minutes, for 4 weeks Intervention 2: Promotor stimulation daily, each session 5 minutes, for 4 weeks

CTRI/2021/07/035287 *(Continued)*

	Intervention 3: Auditory stimulation: daily, each session 5 minutes, for 4 weeks Intervention 4: Vestibular stimulation: daily, each session 10 minutes, for 4 weeks Intervention 5: Positioning: daily each session 5 minutes, for 4 weeks Control Intervention 1: (control): 1. Tactile stimulation daily thrice, each session 7 minutes for 4 weeks Control Intervention 2: Promotor stimulation: daily, each session 7 minutes for 4 weeks Control Intervention 3: Auditory stimulation: daily, each session 8 minutes for 4 weeks Control Intervention 4: Positioning: daily, each session 8 minutes for 4 weeks Control Intervention 5: Tactile stimulation: daily, each session 7 minutes for 4 weeks
Outcomes	1. Neurological maturity 2. Weight gain 3. Physiological parameters (oxygen saturation, heart rate, respiratory rate). Time point: 0 week and after 4 weeks
Starting date	2021
Contact information	Dr Kumaresan A Address: 5th floor Department of Neuro physiotherapy Saveetha college of physiotherapy, Saveetha Institute of Medical and Technical Sciences, Saveetha nagar Thandalam Chennai 602105 Kancheepuram, TAMIL NADU India Telephone: 07299934070 Email: kresh49@gmail.com Affiliation: Saveetha college of physiotherapy, Saveetha Institute of Medical and Technical Sciences
Notes	

D'Souza 2017

Study name	Effectiveness of Indian classical music on developmental responses of preterm infants - a randomized controlled trial protocol
Methods	Randomised controlled trial
Participants	Preterm infants who are admitted to the NICU. A sample size of 132 preterm infants who are admitted to the NICU, will be assessed for eligibility to be included in the study. The parent(s) of the preterm infants will be approached for informed proxy consent. A study assistant will manage the allocation of participants and the provision of the intervention independently. The investigators who perform the outcome assessment will be blinded to the allocation of participants to the intervention and control arm of the study.
Interventions	Indian classical music
Outcomes	Outcome measures will be assessed at baseline (day 3 of life). Following the administration of the intervention in the experimental group, the outcome measures will be assessed after completion

D'Souza 2017 *(Continued)*

	of 1 week of intervention (day 10 of life) and after completion of 2 weeks of intervention (day 17 of life).
Starting date	Not available
Contact information	D'Souza SRB: Department of Obstetrical and Gynaecological Nursing, Manipal College of Nursing, Manipal Department of Paediatrics, Kasturba Medical College, Manipal
Notes	Only abstract found, no trial entry

DRKS00018806

Study name	Music and maternal voice: investigating the impact on premature babies outcomes
Methods	RCT
Participants	27 + 0/7 and 34 + 6/7 gestation age
Interventions	Intervention 1: Lullabies sung by own mother Intervention 2: Lullabies sung by strange women Intervention Intervention 3: No intervention
Outcomes	Evaluation of the effect of the intervention (strange woman's voice and mothers voice) on vital parameters (heart frequency, blood pressure, oxygen saturation, breathing) of the premature babies - evaluation of the effect of the intervention (strange woman's voice and mothers voice) on weight gain of the premature babies - evaluation of the effect of the intervention (strange woman's voice and mothers voice) on cortisol levels of the premature babies (by means of saliva samples)
Starting date	2019
Contact information	Name: Nora K. Schaal Address: Universitätsstr. 1 40225 Düsseldorf Germany Telephone: 0211 8113863 Email: nora.schaal@hhu.de Affiliation: Institut für Experimentelle Psychologie Heinrich-Heine-Universität
Notes	

DRKS00025753

Study name	Does music therapy influence the neurological development of premature infants born less than 32 weeks of gestation?
Methods	RCT
Participants	Inclusion criteria: born with a gestational age less than 32 weeks
Interventions	Intervention 1: Music therapy for premature infants, born with a gestational age less than 32 weeks Intervention 2: No music therapy for premature infants, born with a gestational age less than 32 weeks
Outcomes	The end point of the investigations is the influence of music therapy on the neurological development of very premature infants. This parameter is checked by measuring the vital signs during the music therapy interventions and a questionnaire to evaluate the music therapy at discharge; re-examination takes place when the child is 24 months old using the Bayley III scale and the questionnaire Austrian Communicative Developmental Inventory.
Starting date	2018
Contact information	Susann Kobus Address: Hufelandstraße 55 45147 Essen Germany Telephone: 020172386267 Email: susann.kobus@uk-essen.de Affiliation: Universitätsklinikum Essen
Notes	

Filippa 2021a

Study name	Effects of early vocal contact in the neonatal intensive care unit: study protocol for a multi-centre, randomised clinical trial
Methods	RCT
Participants	25–32 weeks and 6 days' gestational age
Interventions	EVC will be performed by mothers 3 times per week for 2 weeks, more than 1 hour after afternoon feeding. It will begin when the newborns are in an active sleep state, in calm awake state or in active awake state, but not in deep sleep or crying.
Outcomes	During the interventions, time and frequency analysis of HRV will be applied. The change from baseline will be calculated across the 3 conditions (i.e. singing, speaking, and control).
Starting date	

Filippa 2021a *(Continued)*

Contact information	
Notes	

Ghetti 2019

Study name	Longitudinal study of music therapy's effectiveness for premature infants and their caregivers (LongSTEP): protocol for an international randomised trial
Methods	RCT
Participants	Preterm infants
Interventions	Music therapy focusing on parental singing specifically tailored to infant responses, will be delivered during NICU and/or during a post-discharge 6-month period.
Outcomes	Changes in mother-infant bonding at 6-month corrected age, as measured by the Postpartum Bonding Questionnaire
Starting date	
Contact information	
Notes	

IRCT2014101819566N2

Study name	Comparison of the effect of Koran (yasin) and lullaby music listening on physiological indicators of premature neonates admitted to Motahari hospital of Jahrom
Methods	RCT
Participants	Preterm newborns admitted to NICU
Interventions	1) 20 minutes of koran 1-83 Yasin-Tartil male sound 2) 20 minutes of lullaby by male sound 3) ambient sound
Outcomes	O_2 saturation, heart rate
Starting date	2014
Contact information	mohsenhojat@yahoo.com
Notes	

IRCT20170620034653N3

Study name	The effect of the Quran's voice on the formation of oral nutrition in premature infants admitted to the intensive care unit
Methods	RCT
Participants	Gestational age less than 32 weeks
Interventions	Intervention(s) Intervention 1: Intervention group: The Voice of the Holy Quran was performed at 45 dB for 20 minutes at three times before feeding at first, middle and end of the evening shift, and after that, the nutrition begins. This is done until the baby reaches her first full oral feeding. Intervention 2: Control group: Premature babies who do not hear the Quran's voice
Outcomes	Vital Signs - duration of the passage from feeding to the tube to oral feeding-weighing. Time point: Physiological responses including heart rate, respiratory rate, arterial oxygen saturation, mean arterial pressure and temperature were recorded in three stages, every 10 minutes; immediately before the intervention, during the intervention (minutes 10 and 20 after intervention) And 10 minutes after the end of the intervention. Method of measurement: using a monitoring device
Starting date	2018
Contact information	Zahra Tayebi Myaneh Address: Qazvin University of Medical Sciences, Shahid Bahonar Blvd, Qazvin. 00000000 Qazvin Iran Telephone: +98 28 3224 8614 Email: z.tayebi@qums.ac.ir Affiliation: Qazvin University of Medical Sciences
Notes	

IRCT20180526039844N1

Study name	Effect of voice and tune in pain of neonates
Methods	RCT
Participants	Baby has a foetal age of 34 weeks and older
Interventions	Intervention 1: In the 1st intervention (voice group), the baby is sleeping on an audio pillow and a sedative vocal with specific sound intensity (35-40 dB) is distributed to the baby. Then, after 3 minutes, a study worker takes blood from the baby and is filmed by a researcher. The 2nd co-worker examines films for the baby. Intervention 2: Intervention group (tune group): In the second intervention, the baby is sleeping on an audio pillow and a sedative tune with specific intensity (40-35 dB) is distributed to the baby. Then, 3 minutes later, a worker takes blood from the newborn and is filmed by a researcher. The 2nd co-worker examines films for the baby.

IRCT20180526039844N1 *(Continued)*

	Control Intervention: The baby is sleeping on an audio pillow without any intervention/ audio input; after 3 minutes, a worker takes blood from the newborn and is filmed by a researcher. The 2nd co-worker examines films for the baby.
Outcomes	Intensity of pain. Time point: 3 minutes after start of intervention. Method of measurement: Neonatal Infants Pain Scale.
Starting date	2018
Contact information	Manijeh Rabiei Address: Esfahan University of Medical Sciences, Hezar Jerib street 8174673461 Isfahan Iran (Islamic Republic of)Telephone:+98 31 1668 7153Email:mrabiey1356@gmail.comAffiliation: Esfahan University of Medical Sciences
Notes	

NCT02471482

Study name	Effect of Mozart music on cerebral oxygenation and behavioural response of premature infants
Methods	RCT
Participants	Babies with corrected age of 28-32 (28 + 0 to 32 + 0) weeks of gestation, more than 7 days old
Interventions	Experimental: music first group Mozart music lullaby for 30 minutes with recording of cerebral oxygenation (by using Near infrared spectroscopy) and vital signs (respiratory rate, heart rate, oxygen saturations and frequency of apnoeic episodes continuously) followed by 10 minutes of "washout" period and then period of no music for next 30 minutes (with same variables recorded as outlined for music session) This cycle is repeated every 6 hours for 24 hours. In addition, the behavioural response of the baby is observed during the study period by video recording which will be in 'mute' video mode and then reviewed by our developmental staff.
Outcomes	Change in cerebral oxygenation values with and without Mozart music as assessed by near infrared spectroscopy (time frame: 1 hour after feeding; period of music (assessed up to 30 minutes), washout period (assessed in next 10 minutes), period of no music (assessed in next 30 minutes)). Same cycle would be repeated every 6 hours for 24 hours only).
Starting date	2015
Contact information	Faiza Yasin Neonatal Department, Coombe Women and Infants University Hospital, Dublin-8, Ireland
Notes	

NCT02948491

Study name	Effectiveness of music therapy on the development of preterm neonates (EMT-DRNP)
Methods	Randomised parallel study
Participants	90 participants

NCT02948491 *(Continued)*

Interventions	1. Recording of the voice of the mother singing the song to the child (to choose this option, the song must be sung to the child during pregnancy at least 2 to 3 times per week). 2. The mother's favourite song, obtained externally (to choose this option, the mother must bring the song that she listened to frequently, at least 5 times per week). 3. Pre-determined lullaby (this option refers only to the lullaby melody or Brahms lullaby, edited to obtain stable volumes and normalisation of the audio, with the effect of progressively appearing and fading for the first and last 10 seconds of the intervention with music therapy, so there are no drastic acoustic changes in the intervention. Audacity editing software will be used.)
Outcomes	The Bayley-III Cognitive Scale
Starting date	
Contact information	
Notes	Recruitment status unknown

NCT03688386

Study name	A language intervention study of preterm infants
Methods	RCT
Participants	23 weeks to 31 weeks
Interventions	Behavioural: reading intervention Written packet with 3 lessons. The 1st lesson includes how to begin to read and talk to their baby. The 2nd lesson includes reading or talking about the day using parent talk. The 3rd lesson includes continuing to engage with the baby through interactive reading. Behavioural: infant bonding Written packet with 3 lessons. The 1st lesson includes skin-to-skin and learning about their baby. The 2nd lesson includes learning readiness cues and participation in feeding times. The 3rd lesson includes developing routines and playing games. Behavioural: LENA recording The LENA device provides 16 hours of language recordings placed inside an infant vest. The recordings are uploaded to a computer which analyses total adult word counts, infant vocalisations, conversational turns, background noise, and silence. Behavioural: LENA linguistic feedback LENA recordings of adult word counts, infant vocalisations, and conversational turns will be provided in printed form after each recording with review of each recording and progress over time.
Outcomes	1. LENA Counts (time frame: change from 32 week language counts to 36 week language counts) Adult word counts, conversational turns, and infant vocalisations from each recording
Starting date	2018
Contact information	Betty Vohr, MDWomen and Infants Hospital of Rhode Island
Notes	

NCT03795454

Study name	Singing kangaroo-infant care with combined auditory intervention and kangaroo care
Methods	Randomised, sequential assignment
Participants	140 participants
Interventions	The parent is singing during the skin-to-skin sessions; music therapist gives the instructions.
Outcomes	Bayley Scales of Infant and Toddler Development, Third Edition (in Finnish in Finland and in Swedish in Sweden)
Starting date	May 3, 2013
Contact information	Vineta Fellman, University of Helsinki
Notes	

NCT03830580

Study name	Benefit of singing in the care of premature children undergoing screening for retinopathy of prematurity in the neonatology and neonatal Resuscitation Unit of the Dijon University Hospital (Voix Chantée)
Methods	Clinical trial
Participants	Infant born prematurely before the end of 31 weeks of amenorrhoea or with a birth weight of less than 1250 g, or both
Interventions	A nurse or assistant trained in singing with an opera singer sings a lullaby She sits next to the incubator and sings from the time the sucrose is administered (2 minutes before the fundus exams is performed) until the end of the examination - taking into account the child's reaction
Outcomes	Acute pain assessed by the Premature Infant Pain Profile scale (time frame: during the examination) The children's faces and bodies will be filmed - without sound - during the examination. The nurse who examines the fundus will start and finish the video recording.
Starting date	2019
Contact information	Contact: Anne-Cécile CHARY-TARDY 03.80.29.33.62 annececile.charytardy@chu-dijon.fr
Notes	

NCT04314440

Study name	Cognitive processing in preterm infants and NICU music therapy

NCT04314440 *(Continued)*

Methods	Randomised parallel group
Participants	102 participants
Interventions	After the initial warm-up session, music therapy will be delivered 3 times a week for 15-20 minutes. All babies in the experimental group will be exposed to 3 weeks of music therapy sessions that are a total of 9 music therapy sessions prior to discharge from the hospital. The baby will be placed at the bassinet 15 minutes before music therapy begins. The baby will be in the bassinet during music therapy and 15 minutes post-music therapy to allow measurement of physiological indicators. If parents are present during music therapy they will not engage in kangaroo care during the music therapy session or when physiological measurements are being taken pre- and post-music therapy, but can do so at other times.
Outcomes	1. Time looking at a familiar stimulus (time frame: 18-24 months) Average difference in rate of habituation as measured by the time looking at a familiar stimulus in milliseconds between experimental and control groups Secondary outcome measures : 1. Physiological measure (time frame: 15 minutes, before, during, after music therapy, 6 days a week) Average difference in heart rate before music therapy to after music therapy from before to during and from during to after music therapy 1. Physiological measure (time frame: 15 minutes, before, during, after music therapy, 6 days a week) Average difference in breathing rate before music therapy to after music therapy from before to during and from during to after music therapy 1. Physiological measure (time frame: 15 minutes, before, during, after music therapy, 6 days a week < 9 Average difference in oxygen saturation before music therapy to after music therapy from before to during and from during to after music therapy 1. Behavioural measure (time frame: before to after music therapy 6 days a week) Difference in sleep between experimental and control group 1. Behavioural measure (time frame: before to after music therapy 6 days a week) Difference in apnoea between experimental and control group 1. Behavioural measure (time frame: before to after music therapy 6 days a week) Difference in bradycardia (number and severity) between experimental and control groups 1. Behavioural measure (time frame: before to after music therapy 6 days a week) Difference in de-saturation degrees between experimental and control groups
Starting date	https://clinicaltrials.gov/ct2/show/NCT04314440
Contact information	Saskatchewan Health Authority - Regina Area
Notes	

NCT04335240

Study name	Premature family music therapy intervention: an Italian protocol to support parenting, attachment bond and preterm development
Methods	Prospective randomised controlled trial parallel group
Participants	90 infants and their parents
Interventions	This protocol is a family-centred care music therapy intervention. The methodologies will provide early intervention from the first days of hospitalisation in NICU and consist of music therapy sessions both "active" (parental chant, live music and lullaby) and "receptive" (listening to recorded tracks). The music therapy accompanies the newborn and the parents during the hospitalisation and focuses its attention on the emotional-relational care, according to the different needs that they will develop over time. Therapy sessions will be performed starting from 3 times per week, on 3 different days of the week, during the entire hospitalisation of the enrolled infants. After discharge, music therapy treatment will be performed once a week until 12 months of corrected age.
Outcomes	Primary outcome measures : 1. HRV analysis (time frame: through test completion, starting from 5 minutes before the test until 5 minutes after test ended) 2. Oxygen's saturation (time frame: through test completion, starting from 5 minutes before the test until 5 minutes after test ended) Monitor oxygen saturation by medical monitor in Neonatal Intensive Care and Neonatology Secondary outcome measures : 1. Stress level of the parents (time frame: at 6 and 12 months of corrected age) Parenting Stress Index-Short Form at 6-12 months of corrected age, State-Trait Anxiety Scale at 6-12 months of corrected age, Beck Depression Inventory at 6-12 months of corrected age 1. Neurobehavioral and neurological development of the child [time frame: at 6-12 months of corrected age] Temperament detected by the test Italian Questionnaires of Temperament. 1. Neurobehavioral and neurological development of the child (time frame: at 1, 3, 6, 9, 12 months of corrected age) General movements. Using digital recording and analysis 1. Neurobehavioral and neurological development of the child (time frame: 12 months of corrected age) Griffiths Scale
Starting date	April 30, 2021
Contact information	Barbara Sgobbi, Ospedale di Circolo - Fondazione Macchi
Notes	https://clinicaltrials.gov/ct2/show/NCT04335240

NCT04559620

Study name	Mother providing recorded voice to her preterm infant in incubator improves her own grief and emotional status

NCT04559620 *(Continued)*

Methods	Randomised parallel group
Participants	40 infants and their mothers
Interventions	Maternal voice (song/words) played in the incubator for 30 minutess x 4/day x 7 days
Outcomes	Primary outcome measures : 1. Maternal response to Depression, Anxiety, and Stress Scale (time frame: baseline to 23 days) Secondary outcome measures : 1. Incidence of apnoeas, bradycardias and desaturation episodes (time frame: up to 23 days) Frequency of these events is measured 1. Feelings questionnaire (time frame: up to 23 days) (1-10 questionnaire where higher score = better outcome)
Starting date	September 23, 2020
Contact information	Drager
Notes	https://clinicaltrials.gov/ct2/show/NCT04559620

NCT04565210

Study name	Effects of oriental music on preterm infants: a randomised controlled trial
Methods	Randomised parallel group
Participants	102 preterm infants
Interventions	Experimental: oriental music Infants assigned to this group will be exposed to oriental music. Active comparator: western music Infants assigned to this group will be exposed to western music. Placebo comparator: silence/control Infants assigned to this group will be exposed to the same protocol but using a track of silence.
Outcomes	Primary outcome measures: 1. Heart rate variability (time frame: 3 years) It consists of changes in the time intervals between consecutive heartbeats called inter-beat intervals. Secondary outcome measures: 1. Mean respiratory rate (time frame: 3 years) The respiratory rate will be retrieved from bedside monitors. 1. Oxygen saturation (time frame: 3 years)

NCT04565210 *(Continued)*

	The oxygen saturation will be retrieved from bedside monitors. 1. Behavioural state (time frame: 3 years) The behavioural score will be assessed using a 7-point score by a certified nurse.
Starting date	September 25, 2020
Contact information	American University of Beirut Medical Center
Notes	https://clinicaltrials.gov/ct2/show/NCT04565210

NCT04759170

Study name	The effects of recorded receptive music therapy on oral nutrition and the well-being of the Italian premature baby: prospective randomised controlled study
Methods	Randomised parallel group
Participants	40
Interventions	1. Mother recorded receptive music therapy intervention. The effects of recorded receptive music therapy on oral nutrition and the well-being of the Italian premature baby Other: music therapy The effects of receptive music therapy on oral nutrition and the well-being of the Italian premature baby 2. Father recorded receptive music therapy intervention. The effects of recorded receptive music therapy on oral nutrition and the well-being of the Italian premature baby Other: music therapy The effects of receptive music therapy on oral nutrition and the well-being of the Italian premature baby 3. Music therapist recorded receptive music therapy intervention The effects of recorded receptive music therapy on oral nutrition and the well-being of the Italian premature baby
Outcomes	Primary outcome measures: 1. Clinical stability by measuring Heart Rate Variability (HRV) (time frame: 8 days) Evaluate and monitor the tolerance of recorded receptive music therapy on clinical stability parameter HRV Secondary outcome measures: 1. Effects of positive reinforcement carried out through the use of recorded receptive music therapy (time frame: through study completion, an average of 20 weeks) Evaluation of the achievement of exclusive oral sucking

NCT04759170 *(Continued)*

	1. Effects of recorded receptive music therapy on positive reinforcement on the acquisition of food skills (time frame: through study completion, an average of 20 weeks) Evaluation of the variation in the speed of meal intake between the beginning and the end of the intervention 1. Effects of the use of recorded receptive music therapy on the acquisition in skill of preterm infants. (time frame: through study completion, an average of 20 weeks) Evaluation of the weight gain of the newborn in the 24 hours following stimulation 1. Effects of positive reinforcement carried out through the use of recorded receptive music therapy on the acquisition of food skills and the growth of preterm infants (time frame: through study completion, an average of 20 weeks) Assessment of the growth rate from the last day of treatment to discharge
Starting date	February 18, 2021
Contact information	Barbara Sgobbi, Ospedale di Circolo - Fondazione Macchi
Notes	https://clinicaltrials.gov/ct2/show/NCT04759170

NCT04759573

Study name	Effects of early vocal contact (EVC) in the neonatal intensive care unit: a multi-centre, randomised clinical trial
Methods	Randomised parallel group
Participants	80 infants and their parents
Interventions	Experimental: early vocal contact The EVC will take place in the hospital room while infants are in their individual incubators or open cribs. In the intervention group, mothers will be asked to speak and sing to their infants continuously over a 10-minute period for each type of intervention (20 minutes in total). Mothers will be asked to talk in their native language and to sing familiar songs, while observing their infant's reactions. The order of the 2 vocalisations, speaking and singing, will be reversed in the next intervention. Early vocal contact will be performed by mothers three times a week for 2 weeks, more than 1 hour after afternoon feeding. It will begin when the newborns are in an active sleep state, in calm awake state or in active awake state, but not in deep sleep or crying. Preterm infants will be enrolled from 25 + 0 to 32 + 6 weeks of gestational age, following the established inclusion criteria.
Outcomes	Primary outcome measures: 1. Change in heart rate variability (time frame: pre intervention (baseline), during the intervention and immediately after the intervention) Heart rate is the number of heartbeats per minute. Secondary outcome measures: 1. Change in general movement assessment (time frame: pre-intervention (baseline)) The general movement quality from video recording will be scored according to the Ferrari optimality score. 2 blinded coders will attribute a single final score for each infant at each time point. For each item, a description of optimal performance is given and scored with "2" (e.g. cramped

NCT04759573 *(Continued)*

components are absent). Less optimal performance is scored with "1" (e.g. cramped components are occasionally present); non-optimal performance is scored with "0" (e.g. cramped components are predominately present).

Adding the scores of each item within a category ("neck and trunk", "upper extremity" and "lower extremity") plus the score for "sequence" gives the GM optimality score with a minimum value of 5 and a maximum value of 42, indicating optimal performance. The minimum score (the worst performance) is 5.

1. Change in general movement assessment (time frame: after the intervention, at term equivalent age)

The general movement quality from video recording will be scored according to the Ferrari optimality score. 2 blinded coders will attribute a single final score for each infant at each time point. For each item, a description of optimal performance is given and scored with "2" (e.g. cramped components are absent). Less optimal performance is scored with "1" (e.g. cramped components are occasionally present); non-optimal performance is scored with "0" (e.g. cramped components are predominately present).

Adding the scores of each item within a category ("neck and trunk", "upper extremity" and "lower extremity") plus the score for "sequence" gives the GM optimality score with a minimum value of 5 and a maximum value of 42, indicating optimal performance. The minimum score (the worst performance) is 5.

1. Change in general movement assessment (time frame: at 3 months)

The general movement quality from video recording will be scored according to the Ferrari optimality score. 2 blinded coders will attribute a single final score for each infant at each time point. For each item a description of optimal performance is given and scored with "2" (e.g. cramped components are absent). Less optimal performance is scored with "1" (e.g. cramped components are occasionally present); non-optimal performance is scored with "0" (e.g. cramped components are predominately present).

Adding the scores of each item within a category ("neck and trunk", "upper extremity" and "lower extremity") plus the score for "sequence" gives the GM optimality score with a minimum value of 5 and a maximum value of 42, indicating optimal performance. The minimum score (the worst performance) is 5.

1. The Griffiths Mental Development Scales (GMDS) (time frame: at 6 months corrected age)

The GMDS will be assessed with mean values in 4 subscales (Locomotor, Per-Social, Hear/Speech, Hand/Eye). A composite final Performance score will also be assessed for each participant at each time point. The mean values will be compared between the intervention and control groups. Scores range from 0 to 109, with better results with higher values.

1. The Griffiths Mental Development Scales (GMDS) (time frame: at 12 months corrected age)

The GMDS will be assessed with mean values in 4 subscales (Locomotor, Per-Social, Hear/Speech, Hand/Eye). A composite final Performance score will also be assessed for each participant at each time point. The mean values will be compared between the intervention and control groups. Scores range from 0 to 109, with better results with higher values.

1. MacArthur-Bates Communicative Development Inventories (time frame: at 12 months (Gestures and Words Form) and 24 months (Words and Sentences Form) corrected age)

Each child, at each time point, will receive a final score for each questionnaire, measured as a discrete numeric value; the mean values will be compared in the intervention and control groups. The minimum score is 0 and the maximum is 429, with higher scores for better performance.

1. MacArthur-Bates Communicative Development Inventories (time frame: at 12 months corrected age)

NCT04759573 *(Continued)*

	Each child, at each time point, will receive a final score for each questionnaire, measured as a discrete numeric value; the mean values will be compared in the intervention and control groups. The minimum score is 0 and the maximum is 429, with higher scores for better performance. 1. MacArthur-Bates Communicative Development Inventories (time frame: at 24 months corrected age) Each child, at each time point, will receive a final score for each questionnaire, measured as a discrete numeric value; the mean values will be compared in the intervention and control groups. The minimum score is 0 and the maximum is 429, with higher scores for better performance. 1. Parole in Gioco test (time frame: at 24 months corrected age) Linguistic test for assessing lexical comprehension and production for early childhood. The minimum score is 0 and the maximum is 60, with higher scores for better performance. 1. Change in Parental Stressor Scale (PSS-NICU) (time Frame: pre-intervention) The PSS-NICU aims at assessing the parental perception of stressors derived from the physical and psychosocial environment of the NICU across three domains: (i) their parental role, (ii) their infant's behaviour and appearance, and (iii) the sights and sounds in the NICU. For each domain, a mean score will be assessed, and a final composite stress score will be calculated from the mean values of the single scores. Each mother, at each time point, will receive a final score for the single questionnaire (range 0-10). The minimum score is 0 and the maximum is 156, with lower scores for better mental health levels. 1. Change in Parental Stressor Scale (time frame: at hospital discharge) The PSS-NICU aims at assessing the parental perception of stressors derived from the physical and psychosocial environment of the NICU across three domains: (i) their parental role, (ii) their infant's behaviour and appearance, and (iii) the sights and sounds in the NICU. For each domain, a mean score will be assessed, and a final composite stress score will be calculated from the mean values of the single scores. Each mother, at each time point, will receive a final score for the single questionnaire (range 0-10). The minimum score is 0 and the maximum is 156, with lower scores for better mental health levels. 1. Maternal presence in the NICU (time frame: at hospital discharge) Time that the mothers spend in the NICU (hours) using maternal self-report forms will be filled out after each visit to the NICU.
Starting date	February 18, 2021
Contact information	Elisa Della Casa Muttini
Notes	https://clinicaltrials.gov/ct2/show/NCT04759573

NCT04886648

Study name	The effect of mother's voice and lullaby on preterm infants`physiological parameters, stress and sleeping-waking state
Methods	Randomised parallel group
Participants	90
Interventions	Experimental: mother's voice Premature baby group with mother's voice application.

NCT04886648 *(Continued)*

	Behavioural: music therapy Experimental: lullaby Premature baby group with lullaby application Behavioural: music therapy No intervention: control Premature baby group with no application
Outcomes	Primary outcome measures: 1. Newborn Stress Evaluation Form (NSEF) (time frame: NSEF was measured on days 1 through 5, which had a 5-day application period) NSEF: no stress indicator '0', mild stress indicators '1', moderate stress indicators '2', and severe stress indicators scored as '3'. Secondary outcome measures: 1. Newborn Sleepıng-Wakıng State Evaluation Form (NSWEF) (time frame: NSWEF was measured on days 1 through 5, which had a 5-day application period.) The evaluation of the newborn status was carried out under the main headings of sleep and waking behaviours. Sleep behaviour; deep sleep, light sleep and drowsy. Waking behaviour is grouped under the headings of awake (extremely awake and eyelids awake), active awake and crying. As a result of the evaluation, it was decided that the behaviour of the newborn were in an organised or disorganised range, and the form was marked and scoring was created for the conditions starting from the deep sleep state to disorganised crying. In the evaluation of the data, the interpretation was made according to the state score of the newborn. It was interpreted that as the YUUDF score decreased, the sleep state of the newborns increased, and as the YUUDF score increased, the wakefulness and crying status of the newborns increased. Other outcome measures: 1. Heart rate (time frame: heart rate was measured on days 1 through 5, which had a 5-day application period) During the study, heart rate was recorded with the help of a pulse oximeter. 1. Respiratory rate (time frame: respiratory rate was measured on days 1 through 5, which had a 5-day application period) During the study, the number of respirations of preterm newborns was counted. 1. Oxygen saturation (time frame: oxygen saturation was measured on days 1 through 5, which had a 5-day application period) During the study, oxygen saturation was recorded with the help of a pulse oximeter.
Starting date	May 14, 2021
Contact information	Dilek Derïnce Eryuruk, TC Erciyes University
Notes	https://clinicaltrials.gov/ct2/show/NCT04886648

Neel 2019

Study name	Randomised controlled trial protocol to improve multisensory neural processing, language and motor outcomes in preterm infants
Methods	RCT
Participants	Preterm infants
Interventions	1) Exposure to recorded parent's voice 2) Skin-to-skin holding
Outcomes	The primary outcome is multisensory response and the secondary outcome is neurodevelopmental outcomes, including sensory adaptation, motor, language, tactile processing, and speech sound differentiation.
Starting date	
Contact information	
Notes	

Pavel 2019

Study name	The effect of music therapy on the electroencephalogram (EEG) and heart rate variability (HRV) of premature infants during routine painful procedures
Methods	Randomised cross-over study
Participants	Newborns delivered before 32 weeks' gestational age
Interventions	Either sucrose or sucrose and music therapy (Brahms' lullaby) during routine venepuncture
Outcomes	EEG and HRV analysis
Starting date	
Contact information	
Notes	

Purdy 2011

Study name	A randomised controlled trial of NICU music: impact on preterm infant rest and growth
Methods	Randomised controlled blinded multi-centre study
Participants	28 to < 37 weeks' gestational age at birth
Interventions	Structured music intervention offered 2 x daily for 7 days
Outcomes	Rest, growth, serum insulin and insulin-like growth factor-1

Purdy 2011 *(Continued)*

Starting date	
Contact information	
Notes	

Tosun 2014

Study name	The effect of aromatherapy, music therapy and vibration applications on neonatal stress and behaviours
Methods	RCT
Participants	Preterm infants
Interventions	Control, aromatherapy, music therapy and vibration application
Outcomes	Data were collected with a questionnaire form, Brazelton Newborn Behavioural Assessment Scale and Newborn Stress Evaluation Form: on the 1st, 3rd and 5th days, applied both pre- and post-intervention; application continued one session/day for 5 days.
Starting date	
Contact information	
Notes	

dB: decibel level; **EEG**: electroencephalogram ; **EVC**: Early Vocal Contact;**GMDS**: Griffiths Mental Development Scales; **HRV**: Heart rate variability; **KMC**: Kangaroo-mother care; **LENA**: language recordings: **NICU**: neonatal intensive care unit; **NSEF**: Newborn Stress Evaluation Form; **NSWEF**: Newborn Sleeping-Waking State Evaluation Form; **O_2**: Oxygen; **PSS**: Parental Stressor Scale; **RCT**: Randomised controlled trial; **SSC**: Skin-to-skin contact

DATA AND ANALYSES

Comparison 1. Music versus control

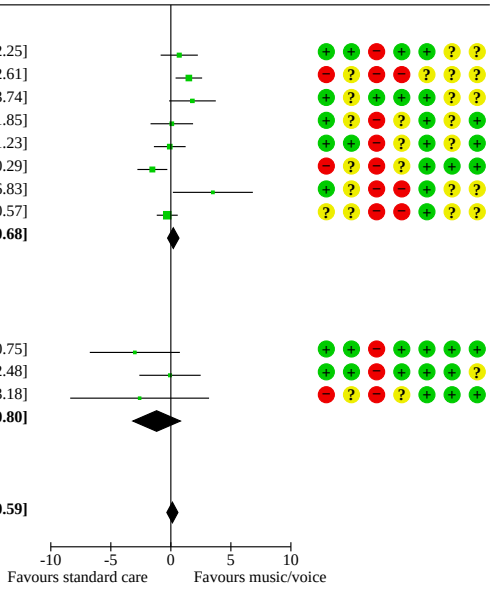
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.1 Oxygen saturation during intervention	10	958	Mean Difference (IV, Fixed, 95% CI)	0.13 [-0.33, 0.59]
1.1.1 Oxygen saturation during intervention recorded music	8	355	Mean Difference (IV, Fixed, 95% CI)	0.20 [-0.27, 0.68]
1.1.2 Oxygen saturation during intervention live music	3	603	Mean Difference (IV, Fixed, 95% CI)	-1.19 [-3.17, 0.80]
1.2 Oxygen saturation post-intervention	7	800	Mean Difference (IV, Fixed, 95% CI)	0.63 [-0.01, 1.26]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.2.1 Oxygen saturation postintervention recorded music	6	232	Mean Difference (IV, Fixed, 95% CI)	0.63 [-0.08, 1.35]
1.2.2 Oxygen saturation postintervention live music	2	568	Mean Difference (IV, Fixed, 95% CI)	0.61 [-0.75, 1.97]
1.3 Oxygen saturation after heel lance	2	100	Mean Difference (IV, Fixed, 95% CI)	0.75 [-0.02, 1.51]
1.4 Infant Development: Bayley Scales of Infant and Toddler Development	2		Mean Difference (IV, Fixed, 95% CI)	Subtotals only
1.4.1 Bayley Scale cognitive composition score 24 months	2	69	Mean Difference (IV, Fixed, 95% CI)	0.35 [-4.85, 5.55]
1.4.2 Bayley Scale motor composition score 24 months	2	69	Mean Difference (IV, Fixed, 95% CI)	-0.17 [-5.45, 5.11]
1.4.3 Bayley Scale language composition score 24 months	2	69	Mean Difference (IV, Fixed, 95% CI)	0.38 [-5.45, 6.21]
1.5 Parental anxiety: STAI-T	4	97	Mean Difference (IV, Fixed, 95% CI)	-1.12 [-3.20, 0.96]
1.6 Heart rate during intervention	11	1014	Mean Difference (IV, Fixed, 95% CI)	-1.38 [-2.63, -0.12]
1.6.1 Heart rate during intervention with frequency of intervention at least 8 times	7	379	Mean Difference (IV, Fixed, 95% CI)	-4.21 [-6.48, -1.94]
1.6.2 Heart rate during intervention with frequency of intervention less than 8 times	4	635	Mean Difference (IV, Fixed, 95% CI)	-0.12 [-1.63, 1.39]
1.7 Heart rate post-intervention period	9	903	Mean Difference (IV, Fixed, 95% CI)	-3.80 [-5.05, -2.55]
1.7.1 Heart rate postintervention recorded music	7	271	Mean Difference (IV, Fixed, 95% CI)	-6.36 [-8.89, -3.83]
1.7.2 Heart rate postintervention live music	3	632	Mean Difference (IV, Fixed, 95% CI)	-2.98 [-4.42, -1.54]
1.8 Heart rate after heel lance	2	100	Mean Difference (IV, Fixed, 95% CI)	1.11 [-3.45, 5.67]
1.9 Respiratory rate during intervention	7	750	Mean Difference (IV, Fixed, 95% CI)	0.42 [-1.05, 1.90]
1.10 Respiratory rate post-intervention	5	636	Mean Difference (IV, Fixed, 95% CI)	0.51 [-1.57, 2.58]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.11 Behavioural outcomes: Behavioural state (Als)	2	69	Mean Difference (IV, Fixed, 95% CI)	-0.12 [-0.52, 0.27]
1.12 Hospitalisation in days	3	89	Mean Difference (IV, Fixed, 95% CI)	-1.57 [-7.64, 4.50]
1.13 Weight gain	4	137	Mean Difference (IV, Random, 95% CI)	3.88 [-1.61, 9.38]
1.14 Postnatal depression: EPDS	2	67	Mean Difference (IV, Fixed, 95% CI)	0.50 [-1.80, 2.81]
1.15 Parental state anxiety: STAI-SKD	3	87	Mean Difference (IV, Fixed, 95% CI)	-0.15 [-2.72, 2.41]

Analysis 1.1. Comparison 1: Music versus control, Outcome 1: Oxygen saturation during intervention

Study or Subgroup	Music/voice Mean	SD	Total	Standard care Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.1.1 Oxygen saturation during intervention recorded music								
Calabro 2003 (1)	1.1	2.09	8	0.4	0.84	9	8.9%	0.70 [-0.85 , 2.25]
Farhat 2010	95.88	1.84	22	94.38	1.9	22	17.5%	1.50 [0.39 , 2.61]
Jabraeili 2016	94.56	2.25	46	92.76	4.18	20	5.6%	1.80 [-0.14 , 3.74]
Kucuk Alemdar 2020	95.26	3.56	30	95.18	3.55	32	6.8%	0.08 [-1.69 , 1.85]
Liao 2021	96.61	2.78	34	96.7	2.82	35	12.2%	-0.09 [-1.41 , 1.23]
Namjoo 2021	92.56	1.35	30	94.1	2.27	15	13.7%	-1.54 [-2.79 , -0.29]
Portugal 2017	98.6	2.2	9	95.1	4.6	9	1.9%	3.50 [0.17 , 6.83]
Tandoi 2015	97.9	1.3	17	98.2	1.3	17	27.9%	-0.30 [-1.17 , 0.57]
Subtotal (95% CI)			**196**			**159**	**94.6%**	**0.20 [-0.27 , 0.68]**
Heterogeneity: Chi² = 21.04, df = 7 (P = 0.004); I² = 67%								
Test for overall effect: Z = 0.84 (P = 0.40)								
1.1.2 Oxygen saturation during intervention live music								
Epstein 2021	91	7	18	94	4	17	1.5%	-3.00 [-6.75 , 0.75]
Loewy 2013	94.03	14.81	261	94.1	14.98	262	3.3%	-0.07 [-2.62 , 2.48]
Namjoo 2021	91.5	15.84	30	94.1	2.27	15	0.6%	-2.60 [-8.38 , 3.18]
Subtotal (95% CI)			**309**			**294**	**5.4%**	**-1.19 [-3.17 , 0.80]**
Heterogeneity: Chi² = 1.86, df = 2 (P = 0.39); I² = 0%								
Test for overall effect: Z = 1.17 (P = 0.24)								
Total (95% CI)			**505**			**453**	**100.0%**	**0.13 [-0.33 , 0.59]**
Heterogeneity: Chi² = 24.68, df = 10 (P = 0.006); I² = 59%								
Test for overall effect: Z = 0.54 (P = 0.59)								
Test for subgroup differences: Chi² = 1.78, df = 1 (P = 0.18), I² = 43.9%								

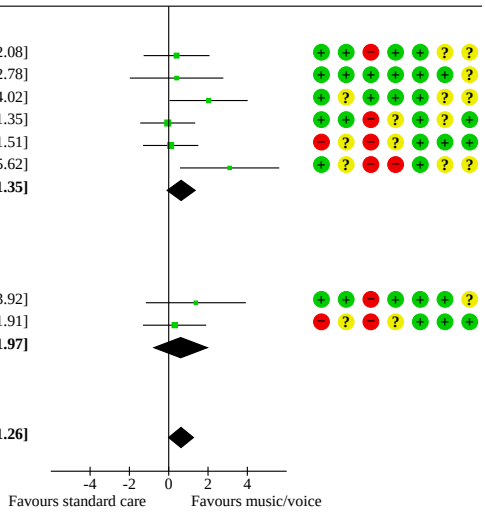


Footnotes
(1) Data value: Change from baseline

Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 1.2. Comparison 1: Music versus control, Outcome 2: Oxygen saturation post-intervention

Study or Subgroup	Music/voice Mean	SD	Total	Standard care Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.2.1 Oxygen saturation postintervention recorded music								
Calabro 2003 (1)	0.5	2.23	8	0.1	0.99	9	14.2%	0.40 [-1.28 , 2.08]
Caparros-Gonzalez 2018	96.73	1.83	8	96.33	3.08	9	7.1%	0.40 [-1.98 , 2.78]
Jabraeili 2016	94.36	2.47	46	92.33	4.23	20	10.1%	2.03 [0.04 , 4.02]
Liao 2021	96.49	2.94	34	96.54	2.98	35	20.5%	-0.05 [-1.45 , 1.35]
Namjoo 2021	93.6	1.69	30	93.5	2.51	15	20.2%	0.10 [-1.31 , 1.51]
Portugal 2017	98.8	1.4	9	95.7	3.6	9	6.3%	3.10 [0.58 , 5.62]
Subtotal (95% CI)			**135**			**97**	**78.4%**	**0.63 [-0.08 , 1.35]**
Heterogeneity: Chi² = 7.15, df = 5 (P = 0.21); I² = 30%								
Test for overall effect: Z = 1.73 (P = 0.08)								
1.2.2 Oxygen saturation postintervention live music								
Loewy 2013	94.88	13.52	261	93.5	16.06	262	6.2%	1.38 [-1.16 , 3.92]
Namjoo 2021	93.8	2.75	30	93.5	2.51	15	15.5%	0.30 [-1.31 , 1.91]
Subtotal (95% CI)			**291**			**277**	**21.6%**	**0.61 [-0.75 , 1.97]**
Heterogeneity: Chi² = 0.49, df = 1 (P = 0.48); I² = 0%								
Test for overall effect: Z = 0.88 (P = 0.38)								
Total (95% CI)			**426**			**374**	**100.0%**	**0.63 [-0.01 , 1.26]**
Heterogeneity: Chi² = 7.65, df = 7 (P = 0.36); I² = 8%								
Test for overall effect: Z = 1.94 (P = 0.05)								
Test for subgroup differences: Chi² = 0.00, df = 1 (P = 0.98), I² = 0%								

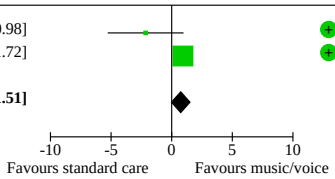


Footnotes
(1) Data value: Change from baseline

Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

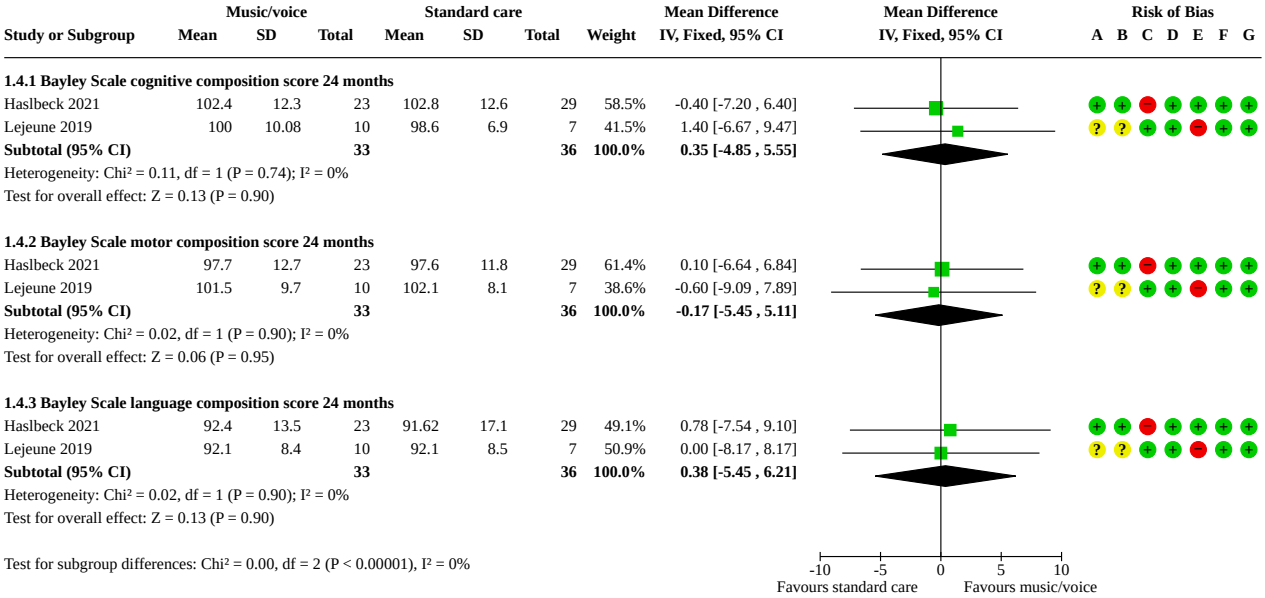
Analysis 1.3. Comparison 1: Music versus control, Outcome 3: Oxygen saturation after heel lance

Study or Subgroup	Music/voice Mean	SD	Total	Standard care Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Johnston 2007	94.9	4.9	11	97.05	3	25	6.0%	-2.15 [-5.28 , 0.98]
Yu 2021	98.34	1.41	32	97.41	1.79	32	94.0%	0.93 [0.14 , 1.72]
Total (95% CI)			**43**			**57**	**100.0%**	**0.75 [-0.02 , 1.51]**
Heterogeneity: Chi² = 3.51, df = 1 (P = 0.06); I² = 71%								
Test for overall effect: Z = 1.91 (P = 0.06)								
Test for subgroup differences: Not applicable								



Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 1.4. Comparison 1: Music versus control, Outcome 4: Infant Development: Bayley Scales of Infant and Toddler Development

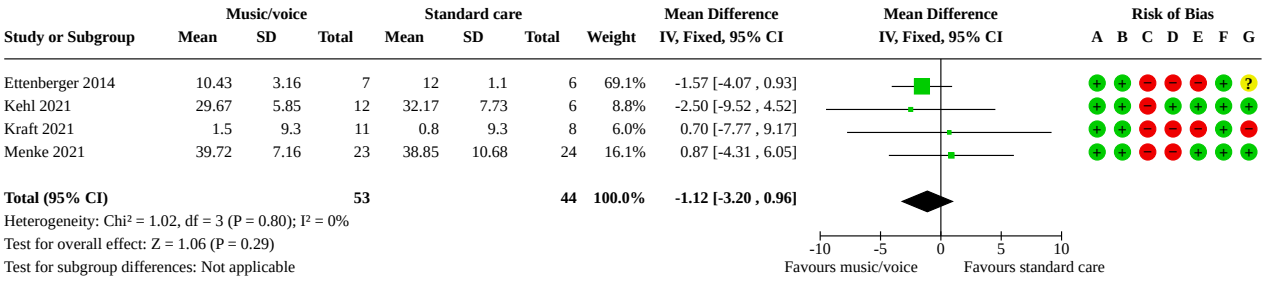


Study or Subgroup	Music/voice Mean	SD	Total	Standard care Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.4.1 Bayley Scale cognitive composition score 24 months								
Haslbeck 2021	102.4	12.3	23	102.8	12.6	29	58.5%	-0.40 [-7.20 , 6.40]
Lejeune 2019	100	10.08	10	98.6	6.9	7	41.5%	1.40 [-6.67 , 9.47]
Subtotal (95% CI)			**33**			**36**	**100.0%**	**0.35 [-4.85 , 5.55]**
Heterogeneity: Chi² = 0.11, df = 1 (P = 0.74); I² = 0%								
Test for overall effect: Z = 0.13 (P = 0.90)								
1.4.2 Bayley Scale motor composition score 24 months								
Haslbeck 2021	97.7	12.7	23	97.6	11.8	29	61.4%	0.10 [-6.64 , 6.84]
Lejeune 2019	101.5	9.7	10	102.1	8.1	7	38.6%	-0.60 [-9.09 , 7.89]
Subtotal (95% CI)			**33**			**36**	**100.0%**	**-0.17 [-5.45 , 5.11]**
Heterogeneity: Chi² = 0.02, df = 1 (P = 0.90); I² = 0%								
Test for overall effect: Z = 0.06 (P = 0.95)								
1.4.3 Bayley Scale language composition score 24 months								
Haslbeck 2021	92.4	13.5	23	91.62	17.1	29	49.1%	0.78 [-7.54 , 9.10]
Lejeune 2019	92.1	8.4	10	92.1	8.5	7	50.9%	0.00 [-8.17 , 8.17]
Subtotal (95% CI)			**33**			**36**	**100.0%**	**0.38 [-5.45 , 6.21]**
Heterogeneity: Chi² = 0.02, df = 1 (P = 0.90); I² = 0%								
Test for overall effect: Z = 0.13 (P = 0.90)								

Test for subgroup differences: Chi² = 0.00, df = 2 (P < 0.00001), I² = 0%

Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 1.5. Comparison 1: Music versus control, Outcome 5: Parental anxiety: STAI-T

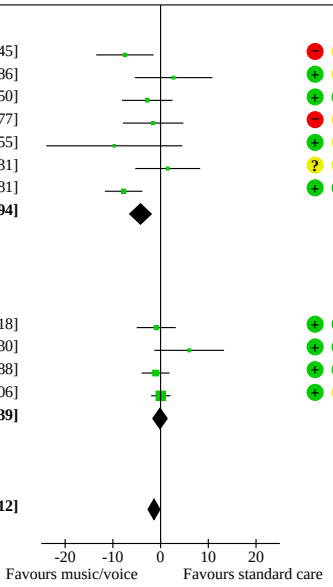


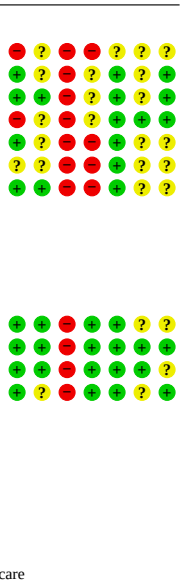
Study or Subgroup	Music/voice Mean	SD	Total	Standard care Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Ettenberger 2014	10.43	3.16	7	12	1.1	6	69.1%	-1.57 [-4.07 , 0.93]
Kehl 2021	29.67	5.85	12	32.17	7.73	6	8.8%	-2.50 [-9.52 , 4.52]
Kraft 2021	1.5	9.3	11	0.8	9.3	8	6.0%	0.70 [-7.77 , 9.17]
Menke 2021	39.72	7.16	23	38.85	10.68	24	16.1%	0.87 [-4.31 , 6.05]
Total (95% CI)			**53**			**44**	**100.0%**	**-1.12 [-3.20 , 0.96]**
Heterogeneity: Chi² = 1.02, df = 3 (P = 0.80); I² = 0%								
Test for overall effect: Z = 1.06 (P = 0.29)								
Test for subgroup differences: Not applicable								

Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 1.6. Comparison 1: Music versus control, Outcome 6: Heart rate during intervention

Study or Subgroup	Music/voice Mean	SD	Total	Standard care Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.6.1 Heart rate during intervention with frequency of intervention at least 8 times								
Farhat 2010	140.39	10.44	22	147.86	9.93	22	4.3%	-7.47 [-13.49 , -1.45]
Kucuk Alemdar 2020	140.17	16.58	30	137.44	16.06	32	2.4%	2.73 [-5.40 , 10.86]
Liao 2021	146.89	11.3	35	149.68	11.3	35	5.6%	-2.79 [-8.08 , 2.50]
Namjoo 2021	144.44	17.7	60	146	12.5	30	3.9%	-1.56 [-7.89 , 4.77]
Portugal 2017	153.3	16.8	9	163	13.9	9	0.8%	-9.70 [-23.95 , 4.55]
Tandoi 2015	132.9	9.4	17	131.4	10.8	17	3.4%	1.50 [-5.31 , 8.31]
Wirth 2016 (1)	-4.87	7.53	40	2.87	7.4	21	10.2%	-7.74 [-11.67 , -3.81]
Subtotal (95% CI)			**213**			**166**	**30.7%**	**-4.21 [-6.48 , -1.94]**
Heterogeneity: Chi² = 11.24, df = 6 (P = 0.08); I² = 47%								
Test for overall effect: Z = 3.64 (P = 0.0003)								
1.6.2 Heart rate during intervention with frequency of intervention less than 8 times								
Calabro 2003 (1)	-0.4	3.75	8	0.5	4.82	9	9.5%	-0.90 [-4.98 , 3.18]
Epstein 2021	140	12	18	134	10	17	3.0%	6.00 [-1.30 , 13.30]
Loewy 2013	157.76	16.86	261	158.81	17.32	262	18.4%	-1.05 [-3.98 , 1.88]
Vastani 2017	138.32	5.73	40	138.28	2.21	20	38.5%	0.04 [-1.98 , 2.06]
Subtotal (95% CI)			**327**			**308**	**69.3%**	**-0.12 [-1.63 , 1.39]**
Heterogeneity: Chi² = 3.25, df = 3 (P = 0.35); I² = 8%								
Test for overall effect: Z = 0.16 (P = 0.87)								
Total (95% CI)			**540**			**474**	**100.0%**	**-1.38 [-2.63 , -0.12]**
Heterogeneity: Chi² = 23.15, df = 10 (P = 0.01); I² = 57%								
Test for overall effect: Z = 2.15 (P = 0.03)								
Test for subgroup differences: Chi² = 8.66, df = 1 (P = 0.003), I² = 88.5%								





Footnotes

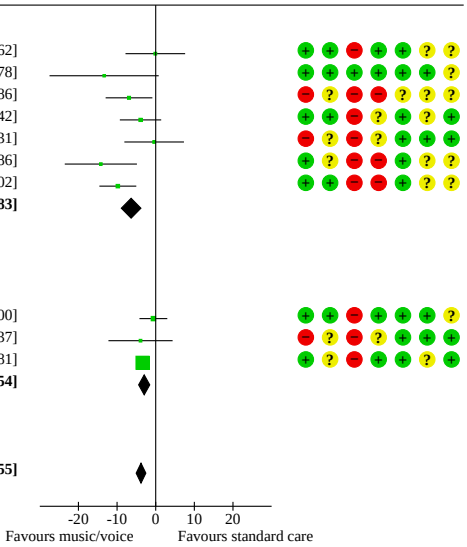
(1) Data value: Change from baseline

Risk of bias legend

(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 1.7. Comparison 1: Music versus control, Outcome 7: Heart rate post-intervention period

Study or Subgroup	Music/voice Mean	SD	Total	Standard care Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.7.1 Heart rate postintervention recorded music								
Calabro 2003 (1)	1	3.4	8	1.1	11.25	9	2.6%	-0.10 [-7.82 , 7.62]
Caparros-Gonzalez 2018	134.8	13.44	8	148.15	16.26	9	0.8%	-13.35 [-27.48 , 0.78]
Farhat 2010	141.1	10.37	22	148.01	10.1	22	4.3%	-6.91 [-12.96 , -0.86]
Liao 2021	147.33	11.25	34	151.26	11.43	35	5.5%	-3.93 [-9.28 , 1.42]
Namjoo 2021	145.33	13.53	30	145.73	11.86	15	2.6%	-0.40 [-8.11 , 7.31]
Portugal 2017	147.2	9.7	9	161.4	10.5	9	1.8%	-14.20 [-23.54 , -4.86]
Wirth 2016 (1)	-5.16	8.79	40	4.64	9.17	21	6.8%	-9.80 [-14.58 , -5.02]
Subtotal (95% CI)			**151**			**120**	**24.4%**	**-6.36 [-8.89 , -3.83]**
Heterogeneity: Chi² = 11.29, df = 6 (P = 0.08); I² = 47%								
Test for overall effect: Z = 4.93 (P < 0.00001)								
1.7.2 Heart rate postintervention live music								
Loewy 2013	157.17	21.78	261	157.77	20.36	266	12.0%	-0.60 [-4.20 , 3.00]
Namjoo 2021	141.8	16.03	30	145.73	11.86	15	2.3%	-3.93 [-12.23 , 4.37]
Vastani 2017	134.78	3.53	40	138.19	2.65	20	61.3%	-3.41 [-5.01 , -1.81]
Subtotal (95% CI)			**331**			**301**	**75.6%**	**-2.98 [-4.42 , -1.54]**
Heterogeneity: Chi² = 2.01, df = 2 (P = 0.37); I² = 0%								
Test for overall effect: Z = 4.06 (P < 0.0001)								
Total (95% CI)			**482**			**421**	**100.0%**	**-3.80 [-5.05 , -2.55]**
Heterogeneity: Chi² = 18.50, df = 9 (P = 0.03); I² = 51%								
Test for overall effect: Z = 5.97 (P < 0.00001)								
Test for subgroup differences: Chi² = 5.20, df = 1 (P = 0.02), I² = 80.8%								

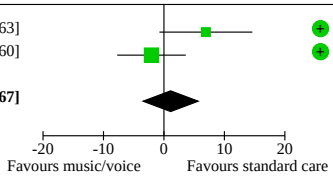


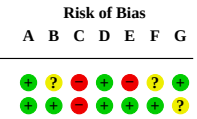
Footnotes
(1) Data value: Change from baseline

Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 1.8. Comparison 1: Music versus control, Outcome 8: Heart rate after heel lance

Study or Subgroup	Music/voice Mean	SD	Total	Standard care Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Johnston 2007	153.03	10.9	11	146.09	10.7	25	35.2%	6.94 [-0.75 , 14.63]
Yu 2021	147.38	11.21	32	149.44	11.9	32	64.8%	-2.06 [-7.72 , 3.60]
Total (95% CI)			**43**			**57**	**100.0%**	**1.11 [-3.45 , 5.67]**
Heterogeneity: Chi² = 3.41, df = 1 (P = 0.06); I² = 71%								
Test for overall effect: Z = 0.48 (P = 0.63)								
Test for subgroup differences: Not applicable								



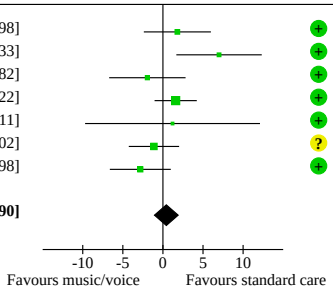


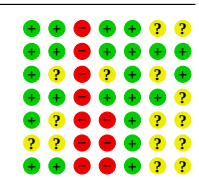
Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 1.9. Comparison 1: Music versus control, Outcome 9: Respiratory rate during intervention

Study or Subgroup	Music/voice Mean	SD	Total	Standard care Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Calabro 2003 (1)	1.3	5.48	8	-0.5	2.68	9	12.4%	1.80 [-2.38 , 5.98]
Epstein 2021	60	9	18	53	7	17	7.7%	7.00 [1.67 , 12.33]
Kucuk Alemdar 2020	44.4	8.96	30	46.34	10.17	32	9.6%	-1.94 [-6.70 , 2.82]
Loewy 2013	50.72	15.42	261	49.13	15.29	262	31.4%	1.59 [-1.04 , 4.22]
Portugal 2017	49.8	11.5	9	48.6	12.1	9	1.8%	1.20 [-9.71 , 12.11]
Tandoi 2015	40.08	5.4	17	41.2	3.8	17	22.1%	-1.12 [-4.26 , 2.02]
Wirth 2016 (1)	-1.4	8.48	40	1.42	6.42	21	15.0%	-2.82 [-6.62 , 0.98]
Total (95% CI)			**383**			**367**	**100.0%**	**0.42 [-1.05 , 1.90]**

Heterogeneity: Chi² = 11.72, df = 6 (P = 0.07); I² = 49%
Test for overall effect: Z = 0.56 (P = 0.57)
Test for subgroup differences: Not applicable





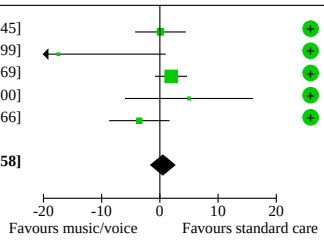
Footnotes
(1) Data value: Change from baseline

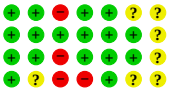
Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 1.10. Comparison 1: Music versus control, Outcome 10: Respiratory rate post-intervention

Study or Subgroup	Music/voice Mean	SD	Total	Standard care Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Calabro 2003 (1)	0.7	4.98	8	0.6	4.06	9	22.7%	0.10 [-4.25 , 4.45]
Caparros-Gonzalez 2018	42.86	17.21	8	60.33	21.57	9	1.3%	-17.47 [-35.93 , 0.99]
Loewy 2013	51.11	15.81	261	49.18	16.38	262	56.5%	1.93 [-0.83 , 4.69]
Portugal 2017	53.5	11.8	9	48.5	12	9	3.6%	5.00 [-6.00 , 16.00]
Wirth 2016 (1)	-2.58	9.89	40	0.94	9.77	21	16.0%	-3.52 [-8.70 , 1.66]
Total (95% CI)			**326**			**310**	**100.0%**	**0.51 [-1.57 , 2.58]**

Heterogeneity: Chi² = 7.66, df = 4 (P = 0.10); I² = 48%
Test for overall effect: Z = 0.48 (P = 0.63)
Test for subgroup differences: Not applicable





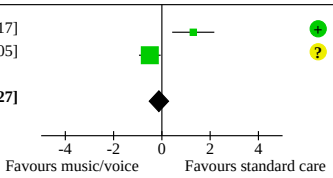
Footnotes
(1) Data value: Change from baseline

Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 1.11. Comparison 1: Music versus control, Outcome 11: Behavioural outcomes: Behavioural state (Als)

Study or Subgroup	Music/voice Mean	SD	Total	Standard care Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Epstein 2021	3.4	1.7	18	2.1	0.8	17	20.9%	1.30 [0.43 , 2.17]
Tandoi 2015	1.6	0.5	17	2.1	0.8	17	79.1%	-0.50 [-0.95 , -0.05]
Total (95% CI)			**35**			**34**	**100.0%**	**-0.12 [-0.52 , 0.27]**

Heterogeneity: Chi² = 12.93, df = 1 (P = 0.0003); I² = 92%
Test for overall effect: Z = 0.61 (P = 0.54)
Test for subgroup differences: Not applicable

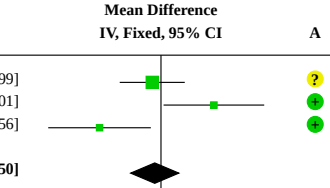


Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 1.12. Comparison 1: Music versus control, Outcome 12: Hospitalisation in days

Study or Subgroup	Music/voice Mean	SD	Total	Standard care Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Cevasco 2008	15.6	9.3	10	17.8	9.39	10	55.0%	-2.20 [-10.39 , 5.99]
Ettenberger 2014	28	20.92	13	14.67	7	6	23.0%	13.33 [0.65 , 26.01]
Menke 2021	70.79	21.12	24	86.27	25.44	26	22.1%	-15.48 [-28.40 , -2.56]
Total (95% CI)			**47**			**42**	**100.0%**	**-1.57 [-7.64 , 4.50]**

Heterogeneity: Chi² = 9.78, df = 2 (P = 0.008); I² = 80%
Test for overall effect: Z = 0.51 (P = 0.61)
Test for subgroup differences: Not applicable

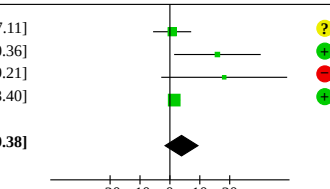


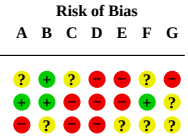
Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 1.13. Comparison 1: Music versus control, Outcome 13: Weight gain

Study or Subgroup	Music/voice Mean	SD	Total	Standard care Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Cevasco 2008	17.7	6.75	13	16.95	8.8	11	31.5%	0.75 [-5.61 , 7.11]
Ettenberger 2014	19.87	17.15	13	3.98	13.83	6	11.4%	15.89 [1.42 , 30.36]
Farhat 2010	36.79	31.83	22	18.63	39.05	22	6.1%	18.16 [-2.89 , 39.21]
Menke 2021	21.71	4.01	24	20.27	2.95	26	51.0%	1.44 [-0.52 , 3.40]
Total (95% CI)			**72**			**65**	**100.0%**	**3.88 [-1.61 , 9.38]**

Heterogeneity: Tau² = 14.38; Chi² = 6.19, df = 3 (P = 0.10); I² = 52%
Test for overall effect: Z = 1.39 (P = 0.17)
Test for subgroup differences: Not applicable



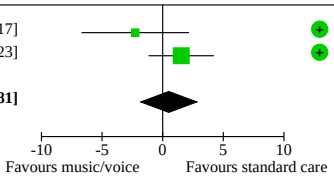


Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 1.14. Comparison 1: Music versus control, Outcome 14: Postnatal depression: EPDS

Study or Subgroup	Music Mean	Music SD	Music Total	Standard care Mean	Standard care SD	Standard care Total	Weight	Mean Difference IV, Fixed, 95% CI
Kehl 2021	5.57	5.14	14	7.83	4.4	6	27.1%	-2.26 [-6.69 , 2.17]
Menke 2021	9.61	4.73	23	8.08	4.71	24	72.9%	1.53 [-1.17 , 4.23]
Total (95% CI)			**37**			**30**	**100.0%**	**0.50 [-1.80 , 2.81]**

Heterogeneity: Chi² = 2.05, df = 1 (P = 0.15); I² = 51%
Test for overall effect: Z = 0.43 (P = 0.67)
Test for subgroup differences: Not applicable

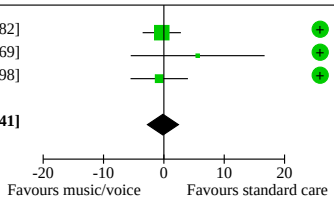


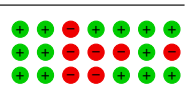
Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Analysis 1.15. Comparison 1: Music versus control, Outcome 15: Parental state anxiety: STAI-SKD

Study or Subgroup	Music/voice Mean	Music/voice SD	Music/voice Total	Standard care Mean	Standard care SD	Standard care Total	Weight	Mean Difference IV, Fixed, 95% CI
Kehl 2021	8.15	2.73	13	8.5	3.5	6	65.5%	-0.35 [-3.52 , 2.82]
Kraft 2021	5.2	9.6	12	-0.4	14.8	9	5.4%	5.60 [-5.49 , 16.69]
Menke 2021	43.02	7.4	23	43.79	9.17	24	29.1%	-0.77 [-5.52 , 3.98]
Total (95% CI)			**48**			**39**	**100.0%**	**-0.15 [-2.72 , 2.41]**

Heterogeneity: Chi² = 1.11, df = 2 (P = 0.57); I² = 0%
Test for overall effect: Z = 0.12 (P = 0.91)
Test for subgroup differences: Not applicable





Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

ADDITIONAL TABLES

Table 1. Summary characteristics of included studies primary outcomes

Study	Design	Country	Infant/parent••	Infant category	Intervention	Control condition	SPO_2	Infant development	State-Trait-anxiety
Calabro 2003	Parallel	Australia	22/0	n.a.	Recorded music	Standard care	x		
Caparros-Gonzalez 2018	Parallel	Spain	22/0	Moderately preterm infant	Recorded music	Standard care	x		
Cevasco 2008	Parallel	USA	25/21	Moderately preterm infant	Recorded music	Standard care			
Portugal 2017	Parallel	Portugal	18/0	Very preterm infant	Recorded auditory stimulation	Standard care	x		
Epstein 2021	Cross-over	Israel	40/40	Extremely preterm infant	Live music	Standard care	x		Narratively reported
Ettenberger 2014	Parallel	Colombia	30/27	Moderately preterm infant	Live music/live music + kangaroo	Standard care	n.a.		x
Farhat 2010	Parallel	Iran	44/0	Very preterm infant	Recorded music	Standard care	x		
Haslbeck 2021	Parallel	Switzerland	82/0	Extremely preterm infant	Live music	Standard care		x	
Jabraeili 2016	Parallel	Iran	75/0	Extremely preterm infant	Recorded music	Standard care	x		
Johnston 2007	Cross-over	Canada	65/0	Moderately preterm infant	Recorded auditory stimulation during heal lance	Standard care heal lance	x		
Kehl 2021*	Parallel	Switzerland	0/46	Extremely preterm infant	Live music	Standard care			x
Kraft 2021	Cross-over	Netherlands	59/23	Extremely preterm infant	Live music	Standard care			x

Table 1. Summary characteristics of included studies primary outcomes *(Continued)*

Kucuk Alemdar 2020	Parallel	Turkey	136/0	Very preterm infant	Recorded mothers spoken voice/breast mild odour/incubator cover	Standard care	x			
Lafferty 2021	Parallel	USA	40/0	Very preterm infant	Recorded music	Standard care				
Lejeune 2019	Parallel	Switzerland	39/0	Very preterm infant	Recorded music	Standard care		x		
Liao 2021	Parallel	China	103/0	Very preterm infant	Recorded music/white noise	Standard care	x			
Loewy 2013	Cross-over	USA	284/284	Very preterm infant	Live music	Standard care	x			
Menke 2021	Parallel	Germany	65/65	Extremely preterm infant	Live music	Standard care				x
Nakhwa 2017	Parallel	India	40/0	Not available	Recorded music & developmental care	Control + developmental program				
Namjoo 2021	Parallel	Iran	90/90	Moderately preterm infant	Recorded music/live music	Standard care	x			
Tandoi 2015	Parallel	Italy	34/0	Moderately preterm infant	Recorded music	Standard care	x			
Vastani 2017	Parallel	Iran	60/0	Moderately preterm infant	Live music	Standard care				
White-Traut 1988	Parallel	USA	33/33	Moderately preterm infant	Live music / massage, talking, eye contact and rocking mother	Standard care				
Wirth 2016	Parallel	Germany	62/62	Moderately preterm infant	Recorded auditory stimulation/recorded music	Standard care				
Yu 2021	Parallel	Taiwan	64/0	Moderately preterm infant	Recorded auditory stimulation during heal lance	Standard care heal lance	x			

Table 1. Summary characteristics of included studies primary outcomes *(Continued)*

Total	**1109/691**

*Kehl 2021 is the same trial as Haslbeck 2020 with two different publications - for a) Haslbeck 2020: infant long-term outcomes and b) Kehl 2021: parental outcomes
**recruited
n.a.: not available
SPO_2: Oxygen saturation

Table 2. Summary first music/vocal intervention characteristics of included studies

Study	Intervention 1: Type & delivery (MM/MT): live, infant-directed, entrained/recorded or standardised	Musical selection by parents/random, unidentified or unknown	Spoken voice/sung voice/music without voice	Womb sound/ rhythmic sound	Intervention alone/combined with SSC/laying in mother's arm/during pain	Frequency & duration/dose (min)	5-10 min/ > 10 min	3-7 times/ ≥ 8 times
Calabro 2003	MM: recorded sedative instrumental lullabies (strings, flute & harps: Brahms & Sandman)	random	music		alone	4 x over 4 consecutive days (daily), 60-70 dB/20 min	> 10	3-7
Caparros-Gonzalez 2018	MM: recorded sedative instrumental music, composed by artificial intelligence	random	music		alone	8 x (7: first not for analysis) over three consecutive days, 30 dB/20 min	> 10	≥ 8
Cevasco 2008	MT: recorded mothers singing voice accompanied by guitar playing of music therapist/lullabies picked by mother from list or Brahms lullaby	parent	spoken voice		alone	3-5 per week until discharge; 65 dB/20 min	> 10	≥ 8
Portugal 2017	MM: recorded mothers voice (spoken and sung) and her heartbeats	parent	spoken & sung voice	rhythmic sound	alone	4 x the day until moved to cradle or discharged; 60-65 dB/45 min	> 10	≥ 8
Epstein 2021	MT: live, infant-directed, entrained vocal & instrumental music (parents preferred)	parent	sung voice		SSC	3 sessions over two weeks; dB: n.a./20 min	> 10	3-7

Table 2. Summary first music/vocal intervention characteristics of included studies *(Continued)*

Ettenberger 2014	MT: SSC & live, infant-directed entrained mothers` singing in lullaby style (song chosen by mother) accompanied by a music therapist with voice and or guitar	parent	sung voice		SSC	17 sessions/13.7 min (range 8-25)	> 10	≥ 8
Farhat 2010	MM: commercially recorded lullabies sung by Iranian female vocalists	random	sung voice		SSC	8 x for 8 consecutive days (daily); 60-65 dB/20 min	> 10	≥ 8
Haslbeck 2021	MT: live, infant-directed, entrained singing accompanied by monochord, parents integrated, individualised therapy with or without parents, parental musical preferences integrated into the improvisation	parent	sung voice	womb sound	alone & SSC	8 to 30 x (2-3 per week until discharge)/20 min	> 10	≥ 8
Kehl 2021	MT: live, infant-directed, entrained singing accompanied by monochord, parents integrated, individualised therapy with or without parents, parental musical preferences integrated into the improvisation	parent	sung voice	womb sound	alone & SSC	8 to 30 x (2-3 per week until discharge)/20 min	> 10	≥ 8
Kraft 2021	MT: live, infant-directed, entrained vocal & instrumental music (parents preferred)	parent	sung voice	womb sound	alone	6 x over two weeks/15 min	> 10	3-7
Kucuk Alemdar 2020	MM: recorded mothers' voice expressing their thoughts and feelings and anything they wanted to say	parent	spoken voice		(alone)	3 x a day until discharge/30 min	> 10	≥ 8
Lafferty 2021	MM: recorded Mozart's double piano sonata	random	music		alone	2 x the day for 14 days/24 min	> 10	≥ 8
Jabraeili 2016	MM: recorded Brahms lullaby	random	sung voice		alone	3 x for 3 consecutive days (daily); 65 dB/15 min	> 10	3-7
Johnston 2007	MM: mother's sung or spoken filtered voice during pain	random	spoken & sung voice		during pain	6 x over 2 days (3 daily); 60-65 dB/10 min	5-10	3-7

Table 2. Summary first music/vocal intervention characteristics of included studies *(Continued)*

Lejeune 2019	MM: recorded instrumental especially composed calm music	random	music		alone	5 per week until discharge; 30-65 dB/8 min	5-10	≥ 8
Liao 2021	MM: recorded mother's sung Chinese version of Schubert's Lullaby	random	sung voice		alone	3 x a day for 4 consecutive days/20 min	> 10	≥ 8
Loewy 2013	MT: live lullaby singing (preferred song of parents) with guitar	parent	sung voice		alone	6 x over two consecutive weeks; 55-65 dB/10 min	5-10	3-7
Menke 2021	MT: live, infant-directed, entrained singing accompanied by monochord, parents integrated, individualised therapy with or without parents, parental musical preferences integrated into the improvisation	parent	sung voice	womb sound	alone & SSC	at least 6 times (2 x per week until discharge)/20-30 min	> 10	≥ 8
Nakhwa 2017	MM: recorded lullaby	random	sung voice		alone	9 x over 3 weeks; 30-40 dB/30 min	> 10	≥ 8
Namjoo 2021	MM: recorded Persian lullaby sang by strange woman, played to infant while laying in the mother's arm using headphones	random	sung voice		arm	once a day for 14 days/20 min	> 10	≥ 8
Tandoi 2015	MM: recorded womb sounds & voices	parent	spoken voice	womb & rhythmic sound	alone	10 x in the first 10 days (daily); 55-70 dB/20-30 min	> 10	≥ 8
Vastani 2017	MM: live mothers` singing of standardised lullaby	random	sung voice		alone	3 x for 3 consecutive days (daily)/10 min	5-10	3-7
White-Traut 1988	MM: live talking or singing mother	parent	spoken & sung voice		alone & arm	6 x over three consecutive days; dB: n.a./15 min	> 10	≥ 8
Wirth 2016	MM: recorded reading mother's voice	random	spoken & sung voice		alone	14 x over 14 consecutive days (daily); 55-65 DB/ 30 min	> 10	≥ 8

Table 2. Summary first music/vocal intervention characteristics of included studies *(Continued)*

Yu 2021	MM: recorded mother's spoken voice reading standardised story and adding personal individual words	random & parent			pain	once a day for 3 consecutive days/13 min	> 10	3-7
Total	**MM: 17** **MT: 7*** **Recorded: 16** **Live: 8***	**Parent: 12*** **Random: 13**	**Spoken voice: 7** **Sung voice: 17*** **Music: 4**	**Rhythmic sound: 2** **Womb sound: 4***	**Alone: 19*** **SSC: 4*** **Arm: 2** **Pain: 2**	**Frequency: 3 - n.a.** **Min: 13-45**	**5-10 min: 3** **> 10 min: 20***	**3-7 times: 8** **≥ 8 times: 16***

*minus 1 since Kehl 2021 is an additional publication of the same intervention study of Haslbeck 2021
dB: decibel level
MM: music medicine
MT: music therapy
Min: minute
n.a.: not available
SSC: Skin-to-Skin Care

Table 3. Summary second music/vocal intervention characteristics of included studies

Study	Intervention 2: type & delivery (MM/MT): live, infant-directed, entrained/recorded or standardised	Musical selection by parents/random, unidentified or unknown	Spoken voice/sung voice/music without voice	Womb sound/ Rhythmic sound/ Breathing sound	Music alone/ Combined with SSC/ Laying mother's arm/ Combined with RISS	Frequency/ Dose (minutes)	5-10 minutes/ > 10 minutes	3-7 times/ ≥ 8 times
Ettenberger 2014	MT: Live entrained vocal & instrumental music (parents preferred); dB n.a.	parent	spoken voice		SSC	2-4 x over two weeks/13.7 min (8-25)	> 10	3-7
Kucuk Alemdar 2020	MM: Live Mum's picked sang lullaby; 65 dB	parent	spoken voice		SSC	Per day for 3 consecutive days/15 min	> 10	3-7

Table 3. Summary second music/vocal intervention characteristics of included studies *(Continued)*

Loewy 2013	MT: Live entrained ocean disc sound; 55-65 dB	random	music	womb sound	alone	6 x over two consecutive weeks/10 min	5-10	≥ 8
Namjoo 2021	MM: Live singing lullabies by mother, baby placed in her arms	parent	spoken voice		arm	once a day for 14 days/20 min	> 10	≥ 8
Vastani 2017	MM: Live nurses` singing of standardised lullaby; 60-70 dB	random	spoken voice		alone	3 x for 3 consecutive days (daily)/30 min	> 10	3-7
White-Traut 1988	MM: Live tactile, vestibular motion, auditory & visual stimulation; dB: n.a.	parent	spoken & sung voice		RISS	6 x over three consecutive days/15 min	> 10	≥ 8
Wirth 2016	MM: Recorded sung lullabies; 55-65	random	spoken voice		alone	14 x over 14 consecutive days (daily)/30 min	> 10	≥ 8
Total	**7 x intervention 2** **MT: 2** **MM: 5** **Live: 6** **Recorded: 1**	**Parent: 5** **Random: 3**	**Spoken voice: 1** **Sung voice: 6**	**Womb sound: 1**	**Alone: 3** **SSC: 2** **Arm: 1** **RISS: 1**	**Frequency: 3 - n.a.** **Min: 14-30**	**5 -10 min:** **1** **> 10 min: 6**	**3-7 times: 3** **≥ 8 times: 4**

dB: decibel level
MM: music medicine
MT: music therapy
Min: minute
n.a.: not available
RISS: massage, talking, eye contact, and rocking
SSC: Skin-to-Skin Care

Table 4. Summary third music/vocal intervention characteristics of included studies

Study	Intervention 3	Intervention type & delivery (MT), live, infant-directed, entrained	Musical selection by parents/random, unidentified or unknown	Spoken voice/sung voice/music without voice	Womb sounds/ rhythmic sounds	Music intervention alone/combined with skin-to skin	Frequency/dose (minutes)	5-10 minutes/ > 10 minutes	Between 3-7 times

Table 4. Summary third music/vocal intervention characteristics of included studies *(Continued)*

Loewy 2013	live music	MT: Live entrained gato box; 55-65 dB	By music therapist	only music	rhythmic sound	alone	6 times over two consecutive weeks/10 min	5-10 min	3-7 times

dB: decibel level
MT: music therapy
Min: minute

Table 5. Summary characteristics of included studies secondary outcomes

Study	Secondary outcome infant heart rate	Secondary outcome infant respiratory rate	Secondary outcome infant heart rate variabilty	Secondary outcome infant behavioural outcomes (Als)	Secondary outcome infant hospitalisation (days)	Secondary outcome infant adverse effects	Secondary outcome infant weight gain (kg/day)	Secondary outcome parents' well-being	Secondary outcome parents attachment
Calabro 2003	x	x							
Caparros-Gonzalez 2018	x	x							
Cevasco 2008					x		x	x (Parental Adaption Inventory)	
Portugal 2017	x	x							
Epstein 2021	x	x	narratively reported	x				x (STAI directly after intervention)	
Ettenberger 2014	no results reported				x		x	x (STAI-T directly after intervention)	x (mother-to-infant bonding)
Farhat 2010	x	x					x		
Haslbeck 2021									

Table 5. Summary characteristics of included studies secondary outcomes *(Continued)*

Jabraeili 2016									
Johnston 2007	x								
Kehl 2021								x (EPDS; PSS; STAI-T)	x (PRAM)
Kraft 2021									
Kucuk Alemdar 2020	x	x							
Lafferty 2021						x			
Lejeune 2019									
Liao 2021	x					x	narratively reported		
Loewy 2013	x	x						x (parental stress)	
Menke 2021					x		x	x (EPDS; PSQ; PCQ; STAI-T)	
Nakhwa 2017									
Namjoo 2021	x								
Tandoi 2015	x	x		x					
Vastani 2017	x								
White-Traut 1988									
Wirth 2016	x	x							
Yu 2021	x	x during pain							
Total	**14**	**10**	**1**	**2**	**3**	**2**	**4 (5)**	**(6)**	**(2)**

EPDS: Edinburgh Postnatal Depression Scale
n: number
n.a.: not available

PCQ: Parental Competences Questionnaire
PRAM: Pictorial Representation of Attachment Measure
PSS: Parental Stressor Scale
PSQ: Parental Stress Questionnaire
RISS: massage, talking, eye contact and rocking
STAI-T: State-Trait Anxiety Inventory

APPENDICES

Appendix 1. Cochrane CENTRAL strategy 2021

	Cochrane CENTRAL via CRS	
	Search date: November 1, 2021	
#	Terms	Results
1	MESH DESCRIPTOR Music Therapy EXPLODE ALL AND CENTRAL:TARGET	914
2	MESH DESCRIPTOR Music EXPLODE ALL AND CENTRAL:TARGET	722
3	MESH DESCRIPTOR Singing EXPLODE ALL AND CENTRAL:TARGET	56
4	MESH DESCRIPTOR Acoustic Stimulation EXPLODE ALL AND CENTRAL:TARGET	1154
5	MESH DESCRIPTOR Auditory Perception EXPLODE ALL AND CENTRAL:TARGET	1863
6	MESH DESCRIPTOR Speech Perception EXPLODE ALL AND CENTRAL:TARGET	667
7	MESH DESCRIPTOR Sound EXPLODE ALL AND CENTRAL:TARGET	730
8	MESH DESCRIPTOR Voice EXPLODE ALL AND CENTRAL:TARGET	404
9	(acoustic and intrauterine) AND CENTRAL:TARGET	10
10	((intrauterine or womb or heart* or breathing or maternal or mother* or parent*) and sound*) AND CENTRAL:TARGET	1584
11	(music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart ADJ3 beat*) or lullab*):ti,ab,kw AND CENTRAL:TARGET	14,407
12	((auditory or acoustic) and (stimul* or cue*)):ti,ab,kw AND CENTRAL:TARGET	3249
13	((speech or language or auditory) and (perception or exposure)):ti,ab,kw AND CENTRAL:TARGET	2226
14	#13 OR #12 OR #11 OR #10 OR #9 OR #8 OR #7 OR #6 OR #5 OR #4 OR #3 OR #2 OR #1	22,002
15	MESH DESCRIPTOR Infant, Newborn EXPLODE ALL AND CENTRAL:TARGET	17,083
16	MESH DESCRIPTOR Intensive Care, Neonatal EXPLODE ALL AND CENTRAL:TARGET	353
17	MESH DESCRIPTOR Intensive Care Units, Neonatal EXPLODE ALL AND CENTRAL:TARGET	841
18	(baby* or babies or infant or infants or infant? or infantile or infancy or low birth weight or low birthweight or neonat* or newborn* or new born or new borns or newly born or premature or prematures or prematurity or preterm or	71,944

(Continued)

	preterms or pre term or preemie or preemies or premies or premie or VLBW or LBW or ELBW or NICU):ti,ab,kw AND CENTRAL:TARGET	
19	#15 OR #16 OR #17 OR #18	75,013
20	#14 AND #19	1624

Appendix 2. MEDLINE strategy 2021

	Ovid MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations, Daily and Versions(R) 1946 to October 29, 2021	
	Search date: November 1, 2021	
#	Searches	Results
1	Music Therapy/	3912
2	Music/	15,172
3	Singing/	1074
4	Acoustic Stimulation/	45,149
5	auditory perception/ or auditory threshold/ or auditory fatigue/ or loudness perception/ or perceptual masking/ or pitch perception/ or pitch discrimination/ or sound localization/ or speech perception/ or timbre perception/ or voice recognition/	80,655
6	voice/ or voice quality/	14,378
7	(music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart adj3 beat*) or lullab*).mp.	147,967
8	((auditory or acoustic) and (stimul* or cue*)).mp.	81,587
9	((speech or language or auditory) and (perception or exposure)).mp.	88,836
10	(acoustic and intrauterine).mp.	100
11	((intrauterine or womb or heart* or breathing or maternal or mother* or parent*) and sound*).mp.	13,901
12	or/1-11 [Music: Sensitive set to combine with RCT Filter]	292,178
13	Music/ or Music Therapy/ or Singing/	19,087
14	(music* or singing or song or songs or sing? or lullab*).ti,ab,kw,kf.	36,408
15	((vocal* or speaking or talking) adj2 (intervention? or parent* or father? or mother?)).ti,ab,kw,kf.	1060

(Continued)

16	or/13-15 [Precise Music set to search for related Systematic Reviews]	41,979
17	exp infant, newborn/ or Intensive Care, Neonatal/ or Intensive Care Units, Neonatal/	64,0467
18	(baby* or babies or infant or infants or infant? or infantile or infancy or low birth weight or low birthweight or neonat* or newborn* or new born or new borns or newly born or premature or prematures or prematurity or preterm or preterms or pre term or preemie or preemies or premies or premie or VLBW or LBW or ELBW or NICU).ti,ab,kw,kf.	947,761
19	or/17-18 [Filter: Neonatal Population 2021--MEDLINE]	1,226,768
20	randomized controlled trial.pt.	548,945
21	controlled clinical trial.pt.	94,499
22	(randomized or randomised).ti,ab.	696,149
23	placebo.ab.	222,732
24	drug therapy.fs.	2,395,933
25	randomly.ab.	368,952
26	trial.ab.	574,229
27	groups.ab.	2,266,078
28	(quasirandom* or quasi-random*).ti,ab.	5245
29	exp animals/ not humans/	4,908,071
30	(or/20-28) not 29 [RCT Filter-Based on Cochrane- Box 6.4.c: Cochrane Highly Sensitive Search Strategy]	4,521,264
31	meta-analysis/ or "systematic review"/ or network meta-analysis/ [/ finds same as.pt. syntax]	246,324
32	((systematic* adj3 (review* or overview*)) or (methodologic* adj3 (review* or overview*))).ti,ab,kf,kw.	245,404
33	((integrative adj3 (review* or overview*)) or (collaborative adj3 (review* or overview*)) or (pool* adj3 analy*)).ti,ab,kf,kw.	32,232
34	(data synthes* or data extraction* or data abstraction*).ti,ab,kf,kw.	32,619
35	(hand search* or handsearch*).ti,ab,kf,kw.	10,128
36	(mantel haenszel or peto or der simonian or dersimonian or fixed effect* or latin square*).ti,ab,kf,kw.	30,350
37	meta-analysis as topic/ or network meta-analysis/	23,318
38	(met analy* or metanaly* or meta regression* or metaregression*).ti,ab,kf,kw.	11,984
39	(medline or cochrane or pubmed or medlars or embase or cinahl).ab.	266,504

(Continued)

40	(cochrane or systematic review?).jw.	18,833
41	or/31-40 [SR filter-Medline; based on CADTHhttps://www.cadth.ca/strings-attached-cadths-database-search-filters]	478,430
42	12 and 19 and 30 [RCT Results]	2948
43	16 and 19 and 41 [SR Results]	73

Appendix 3. Embase strategy 2021

	Embase (OVID) 1974 to 2021 October 29	
	Search date: November 1, 2021	
#	Query	Results
1	(MH "Music Therapy")	6269
2	(MH "Music")	10,960
3	(MH "Singing")	3542
4	(MH "Voice+") OR (MH "Voice Perception")	8256
5	TI ((music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart and beat*) or lullab* or ((auditory or acoustic) and (stimul* or cue*)) or ((speech or language) and (perception or exposure) and (intrauterine OR womb)) or (acoustic and intrauterine) or ((intrauterine OR womb OR heart* OR breathing OR maternal OR mother* OR parent*) and sound*))) OR AB ((music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart and beat*) or lullab* or ((auditory or acoustic) and (stimul* or cue*)) or ((speech or language) and (perception or exposure) and (intrauterine OR womb)) or (acoustic and intrauterine) or ((intrauterine OR womb OR heart* OR breathing OR maternal OR mother* OR parent*) and sound*)))	69,137
6	S1 OR S2 OR S3 OR S4 OR S5	77,501
7	AB (medline or cochrane or pubmed or medlars or embase OR CINAHL)	104,870
8	(TI met analy* or metanaly* or meta regression* or metaregression*)) OR (AB met analy* or metanaly* or meta regression* or metaregression*))	4256
9	AB (mantel haenszel or peto or der simonian or dersimonian or fixed effect* or latin square*)	8724
10	AB (hand search* or handsearch*)	4686
11	(TI (data synthes* or data extraction* or data abstraction*)) OR (AB (data synthes* or data extraction* or data abstraction*))	12,413

(Continued)

12	(TI ((systematic* N3 (review* or overview*)) or (methodologic* N3 (review* or overview*)))) OR (AB ((systematic* N3 (review* or overview*)) or (methodologic* N3 (review* or overview*))))	124,389
13	S7 OR S8 OR S8 OR S9 OR S10 OR S11 OR S12	178,941
14	(MH "Single-Blind Studies") OR (MH "Triple-Blind Studies") OR (MH "Randomized Controlled Trials+") OR (MH "Double-Blind Studies") OR (MH "Clinical Trials+")	326,454
15	TI (quasirandom* or quasi-random*) OR AB (quasirandom* or quasi-random*)	2144
16	AB (trial OR placebo* or random*) OR SU random*	570,644
17	S14 OR S15 OR S16	672,944
18	(MH "Animal Studies") NOT (MH "Human")	117,833
19	S17 NOT S18	656,117
20	((MH "Infant, Newborn+") OR (MH "Infant, Large for Gestational Age") OR (MH "Infant, Low Birth Weight+") OR (MH "Infant, Postmature") OR (MH "Infant, Premature")) OR ((MH "Intensive Care, Neonatal+") OR (MH "Intensive Care Units, Neonatal"))	151,986
21	TI ((baby* or babies or infant or infants or infant? or infantile or infancy or low birth weight or low birthweight or neonat* or newborn* or new born or new borns or newly born or premature or prematures or prematurity or preterm or preterms or pre term or preemie or preemies or premies or premie or VLBW or LBW or ELBW or NICU)) OR AB ((baby* or babies or infant or infants or infant? or infantile or infancy or low birth weight or low birthweight or neonat* or newborn* or new born or new borns or newly born or premature or prematures or prematurity or preterm or preterms or pre term or preemie or preemies or premies or premie or VLBW or LBW or ELBW or NICU)) OR SU ((baby* or babies or infant or infants or infant? or infantile or infancy or low birth weight or low birthweight or neonat* or newborn* or new born or new borns or newly born or premature or prematures or prematurity or preterm or preterms or pre term or preemie or preemies or premies or premie or VLBW or LBW or ELBW or NICU))	530,443
22	S20 OR S21	530,546
23	S6 AND S19 AND S22	900
24	(TI (music* or melod* or song* or sing or sings or singing or sung or singer* or chant* or whistl* or lullab*)) OR (AB (music* or melod* or song* or sing or sings or singing or sung or singer* or chant* or whistl* or lullab*))	21,388
25	S13 AND S22 AND S24	63
26	S23 OR S25	937

Appendix 4. CINAHL strategy 2021

	CINAHL EbscoHost (1980-)	
	Search date: November 12, 2021	
1	(MH "Music Therapy")	6269
2	(MH "Music")	10,960
3	(MH "Singing")	3542
4	(MH "Voice+") OR (MH "Voice Perception")	8256
5	TI ((music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart and beat*) or lullab* or ((auditory or acoustic) and (stimul* or cue*)) or ((speech or language) and (perception or exposure) and (intrauterine OR womb)) or (acoustic and intrauterine) or ((intrauterine OR womb OR heart* OR breathing OR maternal OR mother* OR parent*) and sound*))) OR AB ((music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart and beat*) or lullab* or ((auditory or acoustic) and (stimul* or cue*)) or ((speech or language) and (perception or exposure) and (intrauterine OR womb)) or (acoustic and intrauterine) or ((intrauterine OR womb OR heart* OR breathing OR maternal OR mother* OR parent*) and sound*)))	69,137
6	S1 OR S2 OR S3 OR S4 OR S5	77,501
7	AB (medline or cochrane or pubmed or medlars or embase OR CINAHL)	104,870
8	(TI met analy* or metanaly* or meta regression* or metaregression*)) OR (AB met analy* or metanaly* or meta regression* or metaregression*))	4256
9	AB (mantel haenszel or peto or der simonian or dersimonian or fixed effect* or latin square*)	8724
10	AB (hand search* or handsearch*)	4686
11	(TI (data synthes* or data extraction* or data abstraction*)) OR (AB (data synthes* or data extraction* or data abstraction*))	12,413
12	(TI ((systematic* N3 (review* or overview*)) or (methodologic* N3 (review* or overview*)))) OR (AB ((systematic* N3 (review* or overview*)) or (methodologic* N3 (review* or overview*))))	124,389
13	S7 OR S8 OR S8 OR S9 OR S10 OR S11 OR S12	178,941
14	(MH "Single-Blind Studies") OR (MH "Triple-Blind Studies") OR (MH "Randomized Controlled Trials+") OR (MH "Double-Blind Studies") OR (MH "Clinical Trials+")	326,454
15	TI (quasirandom* or quasi-random*) OR AB (quasirandom* or quasi-random*)	2144
16	AB (trial OR placebo* or random*) OR SU random*	570,644
17	S14 OR S15 OR S16	672,944
18	(MH "Animal Studies") NOT (MH "Human")	117,833

(Continued)

19	S17 NOT S18	656,117
20	((MH "Infant, Newborn+") OR (MH "Infant, Large for Gestational Age") OR (MH "Infant, Low Birth Weight+") OR (MH "Infant, Postmature") OR (MH "Infant, Premature")) OR ((MH "Intensive Care, Neonatal+") OR (MH "Intensive Care Units, Neonatal"))	151,986
21	TI ((baby* or babies or infant or infants or infant? or infantile or infancy or low birth weight or low birthweight or neonat* or newborn* or new born or new borns or newly born or premature or prematures or prematurity or preterm or preterms or pre term or preemie or preemies or premies or premie or VLBW or LBW or ELBW or NICU)) OR AB ((baby* or babies or infant or infants or infant? or infantile or infancy or low birth weight or low birthweight or neonat* or newborn* or new born or new borns or newly born or premature or prematures or prematurity or preterm or preterms or pre term or preemie or preemies or premies or premie or VLBW or LBW or ELBW or NICU)) OR SU ((baby* or babies or infant or infants or infant? or infantile or infancy or low birth weight or low birthweight or neonat* or newborn* or new born or new borns or newly born or premature or prematures or prematurity or preterm or preterms or pre term or preemie or preemies or premies or premie or VLBW or LBW or ELBW or NICU))	530,443
22	S20 OR S21	530,546
23	S6 AND S19 AND S22	900
24	(TI (music* or melod* or song* or sing or sings or singing or sung or singer* or chant* or whistl* or lullab*)) OR (AB (music* or melod* or song* or sing or sings or singing or sung or singer* or chant* or whistl* or lullab*))	21,388
25	S13 AND S22 AND S24	63
26	S23 OR S25	937

Appendix 5. PsycINFO strategy 2021

APA PsycInfo (OVID) <1806 to November Week 2 2021>
Search date: November 12, 2021
1 exp Music/ or exp Music Therapy/ or exp Music Perception/ (25,833)
2 singing/ (1401)
3 rhythm/ or auditory stimulation/ or auditory perception/ (39670)
4 voice/ (5025)
5 (music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or lullab*).ti,ab,id. (110,104)
6 ((heartbeat* or (heart adj3 beat*)) adj4 (mother* or father* or parent* or feeding or rocking)).ti,ab,id. (20)

(Continued)

7 ((speech or language or auditory) adj4 (perception or exposure)).ti,ab,id. (14,525)
8 ((auditory or acoustic) adj4 (stimul* or cue? or cuing)).ti,ab,id. (18,412)
9 ((intrauterine or womb or heart or heartbeat* or breathing or maternal or mother* or father? or parent*) adj4 sound?).ti,ab,id. (487)
10 or/1-9 [Music] (161,110)
11 (music* or singing or song or songs or sing? or lullab*).ti,ab,id. (48,691)
12 ((vocal* or speaking or talking) adj2 (intervention? or parent* or father? or mother?)).ti,ab,id. (1312)
13 or/1-4,11-12 [Music: less sensitive set to combine with SR Filter] (90,639)
14 premature birth/ or birth/ or birth weight/ (16,009)
15 neonatal development/ or neonatal period/ (3849)
16 neonatal intensive care/ (1760)
17 (baby* or babies or infant or infants or infant? or infantile or infancy or low birth weight or low birthweight or neonat* or newborn* or new born or new borns or newly born or premature or prematures or prematurity or preterm or preterms or pre term or preemie or preemies or premies or premie or VLBW or LBW or ELBW or NICU).ti,ab,id. (137,836)
18 or/14-17 [Neonate] (143300)
19 (randomized controlled trial or controlled clinical trial or randomized or randomised or placebo or clinical trial? or randomly).ti,ab,id. (202,439)
20 randomized controlled trials/ or clinical trials/ or randomized clinical trials/ (13,006)
21 or/19-20 [RCT Filter] (203,414)
22 "systematic review"/ (641)
23 meta analysis/ (5084)
24 "literature review"/ (22,928)
25 "literature review"/ and (Medline or embase or cinahl or pubmed).ab. (336)
26 ((systematic* adj3 (review* or overview*)) or ((methodologic* or scoping or overview?) adj3 (review* or overview*)) or (metaanal* or meta-analy*)).ti,ab,id. (155,348)
27 or/22-26 [SR Filter] (176,414)
28 10 and 18 and 21 [Music (sensitive) AND Neonate AND RCT Filter] (243)
29 13 and 18 and 27 [Music (precise) AND Neonate AND SR Filter] (96)
30 or/28-29 [All results PsycINFO] (329)

Appendix 6. Web of Science (WoS) strategy 2021

	Web of Science (WoS) Core Collection: Science Citation Index Expanded (1982-); Social Sciences Citation Index (1982-); Ars & Humanities Citation Indes (1982-); Emerging Sources Citation Index (2015-). Additonal WoS databases: KCI (Korean Journal Database (1980-); Russian Science Citation Index (2005-); SciELO Citation Index (2005-)	Core Collection	Korean	Russian	Scielo
	Search date: November 12, 2021				
#	Search Terms				
1	(((TI=((music* or melod* or voice* or song* or sing or sings or singing or sung or singer* or chant* or whistl* or "maternal speech"))) OR TI=(((heart or heartbeat or voice) NEAR/4 (mother* or father* or parent*)))) OR TI=(((womb or intrauterine) NEAR/4 (voice?))))	210,640	17,540	4153	8240
2	TS=(music* or melod* or voice* or song* or sing or sings or singing or sung or singer* or chant* or whistl* or "maternal speech") OR AK=(music* or melod* or voice* or song* or sing or sings or singing or sung or singer* or chant* or whistl* or "maternal speech")	353,922	57,832	4342	8482
3	TS=((heart or heartbeat or voice) NEAR/4 (mother* or father* or parent*)) OR AK=((heart or heartbeat or voice) NEAR/4 (mother* or father* or parent*))	2275	218	27	67
4	TS=((womb or intrauterine) NEAR/4 (voice OR voices))	5	0	0	0
5	#1 OR #2 OR #3 OR #4	355,077	57,914	4363	8482
6	TI= (neonate or neonates or baby or babies or preterm* or preemie or preemies or "new born" or "new borns" or newborn* or infancy or infant or infants) OR AK= (neonate or neonates or baby or babies or preterm* or preemie or preemies or "new born" or "new borns" or newborn* or infancy or infant or infants)	321,654	7128	7508	15,402
7	TS= (neonate or neonates or baby or babies or preterm* or preemie or preemies or "new born" or "new borns" or newborn* or infancy or infant or infants)	654,144	14,682	7812	
8	#6 OR #7	654,144	14,682	7812	15,402
9	TI=(randomized or randomised or random or trial) OR TS=(randomised or randomized or trial)	2,008,499	26,165	13,664	42,151

(Continued)

10	#5 AND #8 AND #9	378	9	0	9
11	TI=("systematic review" OR metaanaly* or "meta-analy*") OR TS=("systematic review" OR metaanaly* or "meta-analy*") OR AK=("systematic review" OR metaanaly* or "meta-analy*")	544,369		1234	6659
12	#5 AND #8 AND #11	101	2	0	2
13	#10 OR #12	437	11	0	11
		Total: 449			
	Note: in KCI, Russian Science Citation Indexm and SciELO, tersm were searched in title, abstract and topic (ti,ab,ts)				

Appendix 7. RILM strategy 2021

	RILM Abstracts of Music Literature (EBSCOhost)	
	Search date November 12, 2021	
#	Search terms	Results
1	TI ((music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart and beat*) or lullab* or ((auditory or acoustic) and (stimul* or cue*)) or ((speech or language) and (perception or exposure) and (intrauterine OR womb)) or (acoustic and intrauterine) or ((intrauterine OR womb OR heart* OR breathing OR maternal OR mother* OR parent*) and sound*))) OR AB (((music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart and beat*) or lullab* or ((auditory or acoustic) and (stimul* or cue*)) or ((speech or language) and (perception or exposure) and (intrauterine OR womb)) or (acoustic and intrauterine) or ((intrauterine OR womb OR heart* OR breathing OR maternal OR mother* OR parent*) NEAR/4 sound*)) OR SW "therapy—music therapy"	675,364
2	(TI ((baby* or babies or infant or infants or infant? or infantile or infancy or low birth weight or low birthweight or neonat* or newborn* or new born or new borns or newly born or premature or prematures or prematurity or preterm or preterms or pre term or preemie or preemies or premies or premie or VLBW or LBW or ELBW or NICU)) OR AB ((baby* or babies or infant or infants or infant? or infantile or infancy or low birth weight or low birthweight or neonat* or newborn* or new born or new borns or newly born or premature or prematures or prematurity or preterm or preterms or pre term or preemie or preemies or premies or premie or VLBW or LBW or ELBW or NICU))) OR SU ((baby* or babies or infant or infants or infant? or infantile or infancy or low birth weight or low birthweight or neonat* or newborn* or new born or new borns or newly born or premature or prematures or prematurity or preterm or preterms or pre term or preemie or preemies or premies or premie or VLBW or LBW or ELBW or NICU)))	3612

(Continued)

3	TI (randomized or randomised or trial or study) or AB (randomly) OR AB ((random* NEAR/3 (trial or controlled or study))	43,221
4	(S1 AND S2 and S3)	156
5	systematic review or meta-analysis	293
6	TI ((systematic or scoping or overview) NEAR/2 review) OR AB ((systematic or scoping or overview) NEAR/2 review) or TI (meta-analy* or metaanaly*) OR AB (meta-analy* or metaanly*)	156
7	SU systematic review	7
8	SU S6 OR S7	161
9	SU S1 AND S2 AND S8	1
10	S4 OR S9	157

Appendix 8. ERIC strategy 2021

ERIC via ProQuest	
Search date: November 12, 2021	
(music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart and beat*) or lullab* or ((auditory or acoustic) and (stimul* or cue*)) or ((speech or language) and (perception or exposure) and (intrauterine OR womb)) or (acoustic and intrauterine) or ((intrauterine OR womb OR heart* OR breathing OR maternal OR mother* OR parent*) and sound*)) AND (newborn* or "new born" or "new borns" or "newly born" or baby* or babies or premature or prematurity or preterm or "pre term" or "low birth weight" or "low birthweight" or VLBW or LBW or infant* or infancy or neonat*) AND (randomized controlled trial OR clinical trial OR randomized OR randomised OR placebo OR randomly)	21

Appendix 9. Trial register search strategies 2021

Date	Source	Search terms	#
12-Nov-21	ISRCTN	Music AND neonate (age group)	3
12-Nov-21	ICTRP	(music OR lullaby OR singing or song or songs) AND neonate	17
12-Nov-21	clinicaltrials.gov	(music AND (neonate or newborn)	53
	Subtotal		73

Appendix 10. Search strategies 2019

The RCT filters have been created using Cochrane's highly sensitive search strategies for identifying randomised trials (Higgins 2019). The neonatal filters were created and tested by the Cochrane Neonatal Information Specialist.

CENTRAL via CRS Web:

Date searched: 15 November 2019
Terms:
1MESH DESCRIPTOR Music Therapy EXPLODE ALL AND CENTRAL:TARGET
2MESH DESCRIPTOR Music EXPLODE ALL AND CENTRAL:TARGET
3MESH DESCRIPTOR Singing EXPLODE ALL AND CENTRAL:TARGET
4MESH DESCRIPTOR Acoustic Stimulation EXPLODE ALL AND CENTRAL:TARGET
5MESH DESCRIPTOR Auditory Perception EXPLODE ALL AND CENTRAL:TARGET
6MESH DESCRIPTOR Speech Perception EXPLODE ALL AND CENTRAL:TARGET
7MESH DESCRIPTOR Sound EXPLODE ALL AND CENTRAL:TARGET
8MESH DESCRIPTOR Voice EXPLODE ALL AND CENTRAL:TARGET
9(music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart ADJ3 beat*) or lullab*) AND CENTRAL:TARGET
10((auditory or acoustic) and (stimul* or cue*)) AND CENTRAL:TARGET
11((speech or language or auditory) and (perception or exposure)) AND CENTRAL:TARGET
12(acoustic and intrauterine) AND CENTRAL:TARGET
13((intrauterine or womb or heart* or breathing or maternal or mother* or parent*) and sound*) AND CENTRAL:TARGET
14#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 AND CENTRAL:TARGET
15MESH DESCRIPTOR Infant, Newborn EXPLODE ALL AND CENTRAL:TARGET
16infant or infants or infant's or "infant s" or infantile or infancy or newborn* or "new born" or "new borns" or "newly born" or neonat* or baby* or babies or premature or prematures or prematurity or preterm or preterms or "pre term" or premies or "low birth weight" or "low birthweight" or VLBW or LBW or ELBW or NICU AND CENTRAL:TARGET
17#16 OR #15 AND CENTRAL:TARGET
18#14 AND #17 AND CENTRAL:TARGET

Results: 1633

MEDLINE via Ovid:

Date ranges: 1946 to 15 November 2019
Terms:
1. exp Music Therapy/ or exp Music/
2. exp Singing/
3. exp Acoustic Stimulation/
4. exp Auditory Perception/ or exp Speech Perception/
5. exp Sound/
6. exp Voice/
7. (music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart adj3 beat*) or lullab*).mp.
8. ((auditory or acoustic) and (stimul* or cue*)).mp.
9. ((speech or language or auditory) and (perception or exposure)).mp.
10. (acoustic and intrauterine).mp.
11. ((intrauterine or womb or heart* or breathing or maternal or mother* or parent*) and sound*).mp.
12. or/1-11
13. exp infant, newborn/
14. (newborn* or new born or new borns or newly born or baby* or babies or premature or prematurity or preterm or pre term or low birth weight or low birthweight or VLBW or LBW or infant or infants or 'infant s' or infant's or infantile or infancy or neonat*).ti,ab.
15. 13 or 14
16. randomized controlled trial.pt.
17. controlled clinical trial.pt.
18. randomized.ab.
19. placebo.ab.
20. drug therapy.fs.
21. randomly.ab.
22. trial.ab.
23. groups.ab.
24. or/16-23
25. exp animals/ not humans.sh.

26. 24 not 25
27. 15 and 26
28. randomi?ed.ti,ab.
29. randomly.ti,ab.
30. trial.ti,ab.
31. groups.ti,ab.
32. ((single or doubl* or tripl* or treb*) and (blind* or mask*)).ti,ab.
33. placebo*.ti,ab.
34. 28 or 29 or 30 or 31 or 32 or 33
35. 14 and 34
36. limit 35 to yr="2018 -Current"
37. 27 or 36
38. 12 and 37

Results: 2567

CINAHL via EBSCOhost:

Date ranges: 1981 to 15 November 2019
Terms:
(music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart and beat*) or lullab* or ((auditory or acoustic) and (stimul* or cue*)) or ((speech or language) and (perception or exposure) and (intrauterine OR womb)) or (acoustic and intrauterine) or ((intrauterine OR womb OR heart* OR breathing OR maternal OR mother* OR parent*) and sound*)) AND
(infant or infants or infant's or infantile or infancy or newborn* or "new born" or "new borns" or "newly born" or neonat* or baby* or babies or premature or prematures or prematurity or preterm or preterms or "pre term" or premies or "low birth weight" or "low birthweight" or VLBW or LBW) AND
(randomized controlled trial OR controlled clinical trial OR randomized OR randomised OR placebo OR clinical trials as topic OR randomly OR trial OR PT clinical trial)

Results: 730

PsycINFO:

Date ranges: 1806 to 15 November 2019
Terms:
(music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart and beat*) or lullab* or ((auditory or acoustic) and (stimul* or cue*)) or ((speech or language) and (perception or exposure)) or (acoustic and intrauterine) or ((intrauterine OR womb OR heart* OR breathing OR maternal OR mother* OR parent*) and sound*)) AND
(newborn* or "new born" or "new borns" or "newly born" or baby* or babies or premature or prematurity or preterm or "pre term" or "low birth weight" or "low birthweight" or VLBW or LBW or infant* or infancy or neonat*) AND
(randomized controlled trial OR controlled clinical trial OR randomized OR randomised OR placebo OR clinical trials as topic OR randomly OR trial)

Results: 562

Web of Science via Clarivate Analytics:

Date ranges: 1982 to 15 November 2019
Terms:
(music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart and beat*) or lullab* or ((auditory or acoustic) and (stimul* or cue*)) or ((speech or language) and (perception or exposure) and (intrauterine OR womb)) or (acoustic and intrauterine) or ((intrauterine OR womb OR heart* OR breathing OR maternal OR mother* OR parent*) and sound*)) AND
(newborn* or "new born" or "new borns" or "newly born" or baby* or babies or premature or prematurity or preterm or "pre term" or "low birth weight" or "low birthweight" or VLBW or LBW or infant* or infancy or neonat*) AND
(randomized controlled trial OR controlled clinical trial OR randomized OR randomised OR placebo OR randomly)

Results: 789

RILM Abstracts of Music Literature via EBSCOhost:

Date ranges: 1967 to 15 November 2019
Terms:

(music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart and beat*) or lullab* or ((auditory or acoustic) and (stimul* or cue*)) or ((speech or language) and (perception or exposure) and (intrauterine OR womb)) or (acoustic and intrauterine) or ((intrauterine OR womb OR heart* OR breathing OR maternal OR mother* OR parent*) and sound*)) AND
(newborn* or "new born" or "new borns" or "newly born" or baby* or babies or premature or prematurity or preterm or "pre term" or "low birth weight" or "low birthweight" or VLBW or LBW or infant* or infancy or neonat*) AND
(randomized controlled trial OR clinical trial OR randomized OR randomised OR placebo OR randomly)

Results: 33

ProQuest Dissertations & Theses A&I via ProQuest:

Date ranges: 1637 to 15 November 2019
Terms:
(music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart and beat*) or lullab* or ((auditory or acoustic) and (stimul* or cue*)) or ((speech or language) and (perception or exposure) and (intrauterine OR womb)) or (acoustic and intrauterine) or ((intrauterine OR womb OR heart* OR breathing OR maternal OR mother* OR parent*) and sound*)) AND
(newborn* or "new born" or "new borns" or "newly born" or baby* or babies or premature or prematurity or preterm or "pre term" or "low birth weight" or "low birthweight" or VLBW or LBW or infant* or infancy or neonat*) AND
(randomized controlled trial OR clinical trial OR randomized OR randomised OR placebo OR randomly)

Results: 100

ERIC via ProQuest:

Date ranges: 1966 to 15 November 2019
Terms:
(music* or melod* or voice* or song* or vocal* or sing or sings or singing or sung or singer* or chant* or whistl* or heartbeat* or (heart and beat*) or lullab* or ((auditory or acoustic) and (stimul* or cue*)) or ((speech or language) and (perception or exposure) and (intrauterine OR womb)) or (acoustic and intrauterine) or ((intrauterine OR womb OR heart* OR breathing OR maternal OR mother* OR parent*) and sound*)) AND
(newborn* or "new born" or "new borns" or "newly born" or baby* or babies or premature or prematurity or preterm or "pre term" or "low birth weight" or "low birthweight" or VLBW or LBW or infant* or infancy or neonat*) AND
(randomized controlled trial OR clinical trial OR randomized OR randomised OR placebo OR randomly)

Results: 17

ISRCTN:

Date searched (if no date restrictions):
Terms:
Interventions: Music* AND Participant age range: Neonate
Interventions: Auditory stimulation AND Participant age range: Neonate
Interventions: Voice* AND Participant age range: Neonate
Interventions: Song* AND Participant age range: Neonate
Interventions: singing AND Participant age range: Neonate
Interventions: Womb sounds AND Participant age range: Neonate

Results: 0

Appendix 11. Risk of bias tool

We used the standard methods of Cochrane and Cochrane Neonatal to assess the methodological quality of the trials. For each trial, we sought information regarding the method of randomisation, blinding and reporting of all outcomes of all the infants enrolled in the trial. We assessed each criterion as being at a low, high, or unclear risk of bias. Two review authors separately assessed each study. We resolved any disagreement by discussion. We added this information to the Characteristics of included studies table. We evaluated the following issues and entered the findings into the Risk of bias tables.

1. Sequence generation (checking for possible selection bias). Was the allocation sequence adequately generated?

For each included study, we categorised the method used to generate the allocation sequence as:

- low risk (any truly random process e.g. random number table; computer random number generator);
- high risk (any non-random process e.g. odd or even date of birth; hospital or clinic record number); or

- unclear risk.

2. Allocation concealment (checking for possible selection bias). Was allocation adequately concealed?

For each included study, we categorised the method used to conceal the allocation sequence as:

- low risk (e.g. telephone or central randomisation; consecutively numbered sealed opaque envelopes);
- high risk (open random allocation; unsealed or non-opaque envelopes, alternation; date of birth); or
- unclear risk

3. Blinding of participants and personnel (checking for possible performance bias). Was knowledge of the allocated intervention adequately prevented during the study?

For each included study, we categorised the methods used to blind study participants and personnel from knowledge of which intervention a participant received. Blinding was assessed separately for different outcomes or class of outcomes. We categorised the methods as:

- low risk, high risk or unclear risk for participants; and
- low risk, high risk or unclear risk for personnel.

4. Blinding of outcome assessment (checking for possible detection bias). Was knowledge of the allocated intervention adequately prevented at the time of outcome assessment?

For each included study, we categorised the methods used to blind outcome assessment. Blinding was assessed separately for different outcomes or class of outcomes. We categorised the methods as:

- low risk for outcome assessors;
- high risk for outcome assessors; or
- unclear risk for outcome assessors.

5. Incomplete outcome data (checking for possible attrition bias through withdrawals, dropouts, protocol deviations). Were incomplete outcome data adequately addressed?

For each included study and for each outcome, we described the completeness of data including attrition and exclusions from the analysis. We noted whether attrition and exclusions were reported, the numbers included in the analysis at each stage (compared with the total randomised participants), reasons for attrition or exclusion where reported, and whether missing data were balanced across groups or were related to outcomes. Where sufficient information was reported or supplied by the trial authors, we re-included missing data in the analyses. We categorised the methods as:

- low risk (< 20% missing data);
- high risk (≥ 20% missing data); or
- unclear risk.

6. Selective reporting bias. Are reports of the study free of suggestion of selective outcome reporting?

For each included study, we described how we investigated the possibility of selective outcome reporting bias and what we found. For studies in which study protocols were published in advance, we compared prespecified outcomes versus outcomes eventually reported in the published results. If the study protocol was not published in advance, we contacted study authors to gain access to the study protocol. We assessed the methods as:

- low risk (where it is clear that all the study's prespecified outcomes and all expected outcomes of interest to the review have been reported);
- high risk (where not all the study's prespecified outcomes have been reported; one or more reported primary outcomes were not prespecified outcomes of interest and are reported incompletely and so cannot be used; study fails to include results of a key outcome that would have been expected to have been reported); or
- unclear risk.

7. Other sources of bias. Was the study apparently free of other problems that could put it at a high risk of bias?

For each included study, we described any important concerns we had about other possible sources of bias (for example, whether there was a potential source of bias related to the specific study design or whether the trial was stopped early due to some data-dependent process). We assessed whether each study was free of other problems that could put it at risk of bias as:

- low risk;
- high risk; or

- unclear risk.

If needed, we planned to explore the impact of the level of bias through undertaking sensitivity analyses.

HISTORY

Protocol first published: Issue 11, 2019

CONTRIBUTIONS OF AUTHORS

The first author wrote the manuscript, and all other authors reviewed and edited the review. The first author led the data screening and extraction process. FH and KM screened abstracts and full texts. FH, KM, and TK contributed to data extraction. JL contributed to musical details in the data extraction of included studies. All analyses and write-ups were done by the first author.

DECLARATIONS OF INTEREST

FH works as a music therapist and senior researcher at the University Hospital Zurich, Switzerland. FH was the project leader of a study included in this review (Haslbeck 2021), and so did not determine the overall study inclusion and exclusion criteria; make study eligibility decisions about, extract data from, carry out the risk of bias assessment for, or perform GRADE assessments for that study. TK and KM assessed this study instead.

KM has no conflicts of interest to declare.

TK has no conflicts of interest to declare.

JL works as Director, Louis Armstrong Center for Music and Medicine, New York City, USA. She has published opinions in medical journals related to the subject of this review. JL was an author on a study included in this review (Loewy 2013), and so did not determine the overall study inclusion and exclusion criteria; make study eligibility decisions about, extract data from, carry out the risk of bias assessment for, or perform GRADE assessments for that study.

JM is Director of the German Cochrane Centre. JM determined the final overall inclusion and exclusion criteria.

DB works as the director of the Department of Neonatology at the University Hospital Zurich, Switzerland. DB was a co-author on a study included in this review (Haslbeck 2021), and so did not determine the overall study inclusion and exclusion criteria; make study eligibility decisions about, extract data from, carry out the risk of bias assessment for, or perform GRADE assessments for that study. TK and KM assessed this study instead.

SOURCES OF SUPPORT

Internal sources

- University Hospital Zurich and University Zurich, Switzerland

  Paid the salary of Friederike Haslbeck. Funding from Vontobel-Stitftung.

External sources

- Vermont Oxford Network, USA

  Cochrane Neonatal Reviews are produced with support from Vermont Oxford Network, a worldwide collaboration of health professionals dedicated to providing evidence-based care of the highest quality for newborn infants and their families.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We made the following changes to the protocol (Haslbeck 2019).

- Edited the search strategies;
- Excluded studies published as only as abstracts.