Cochrane Library

Cochrane Database of Systematic Reviews | Review - Intervention

Musical and vocal interventions to improve neurodevelopmental outcomes for preterm infants

Friederike B Haslbeck, Katharina Mueller, Tanja Karen, Joanne Loewy, Joerg J Meerpohl, Dirk Bassler Authors' declarations of interest

Version published: 07 September 2023 Version history
https://doi.org/10.1002/14651858.CD013472.pub2

Collapse all Expand all

Abstract

Available in **English** | Español | فارسی | Français | ภาษาไทย | 简体中文

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.

PICOs

Plain language summary

Available in English | Deutsch | Español | فارسی | Français | 日本語 | Nederlands | ภาษาไทย | 简体中文 | 繁體中文

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.

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.

Summary of findings

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

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
[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.

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]

[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.

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]

[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.

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]

[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.

PARENTAL ANXIETY: STAI-SKD assessed with: State Anxiety Inventory Scale from: 6 to 80 (lower = favourable)	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 interventions 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).

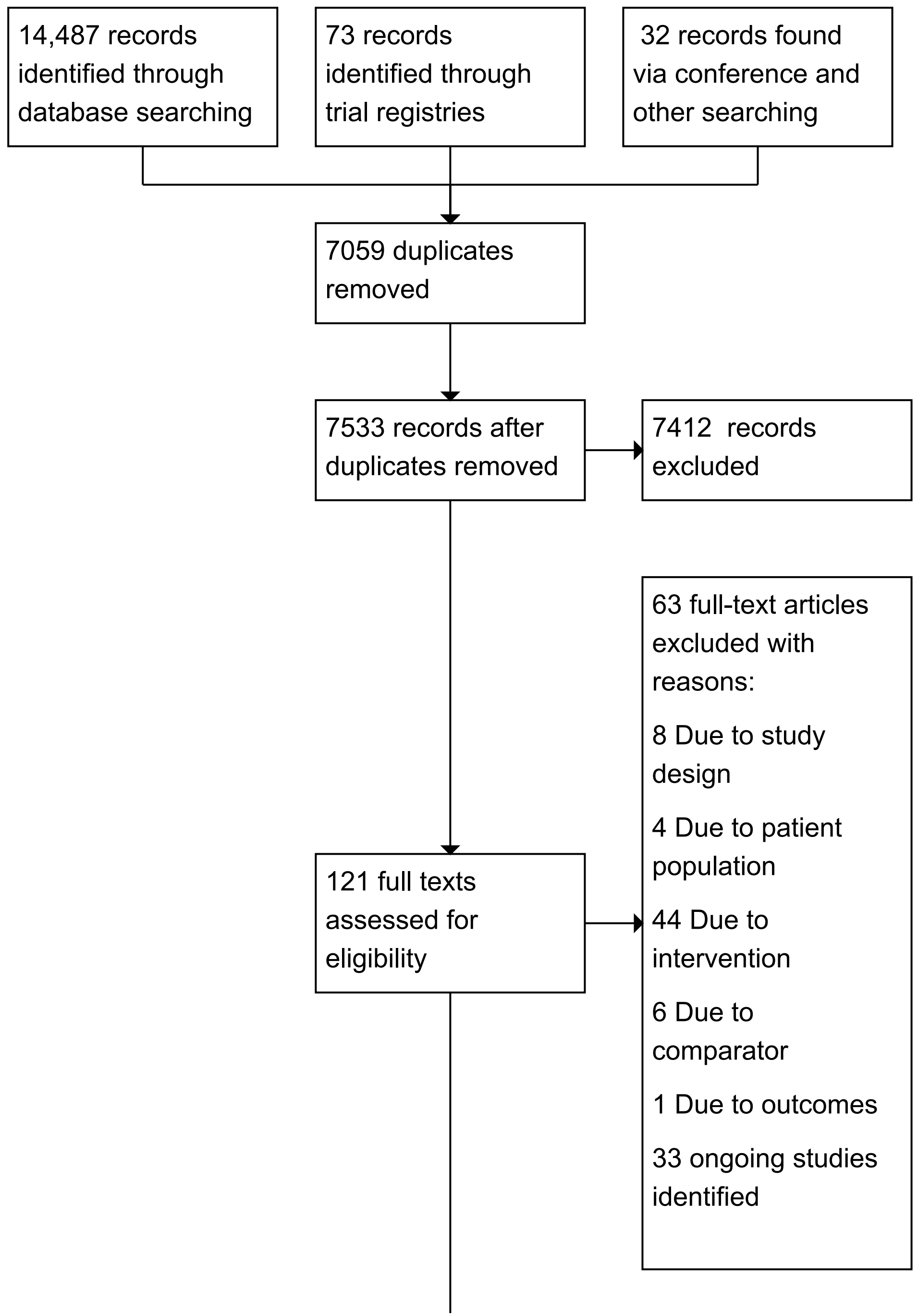
The following databases were searched in 2019 and strategies are presented in Appendix 10.

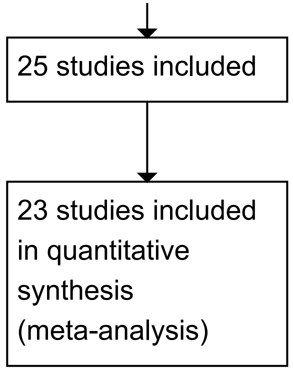
- Cochrane Central Register of Controlled Trials (CENTRAL 2019, Issue #11) in the Cochrane Library;
- Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Daily and Versions(R) (1946 to 15 November 2019);
- CINAHL (1981 to 15 November 2019);
- PsycINFO (1806 to 15 November 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

Open in figure viewer


14,487 records identified through database searching
73 records identified through trial registries
32 records found via conference and other searching
7059 duplicates removed
7533 records after duplicates removed
7412 records excluded
63 full-text articles excluded with reasons:
8 Due to study design
4 Due to patient population
44 Due to intervention
6 Due to comparator
1 Due to outcomes
33 ongoing studies identified
121 full texts assessed for eligibility



Study flow diagram

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 Table 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).

Table 1. Summary characteristics of included studies primary outcomes **Open in table viewer**

Study	Design	Country	Infant/parent••	Infant category	Intervention	Control condition	SPO_2	Infant develop-ment	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
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
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

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.

Table 2. Summary first music/vocal intervention characteristics of included studies **Open in table viewer**

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
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

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
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 **Open in table viewer**

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
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 **Open in table viewer**

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

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

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

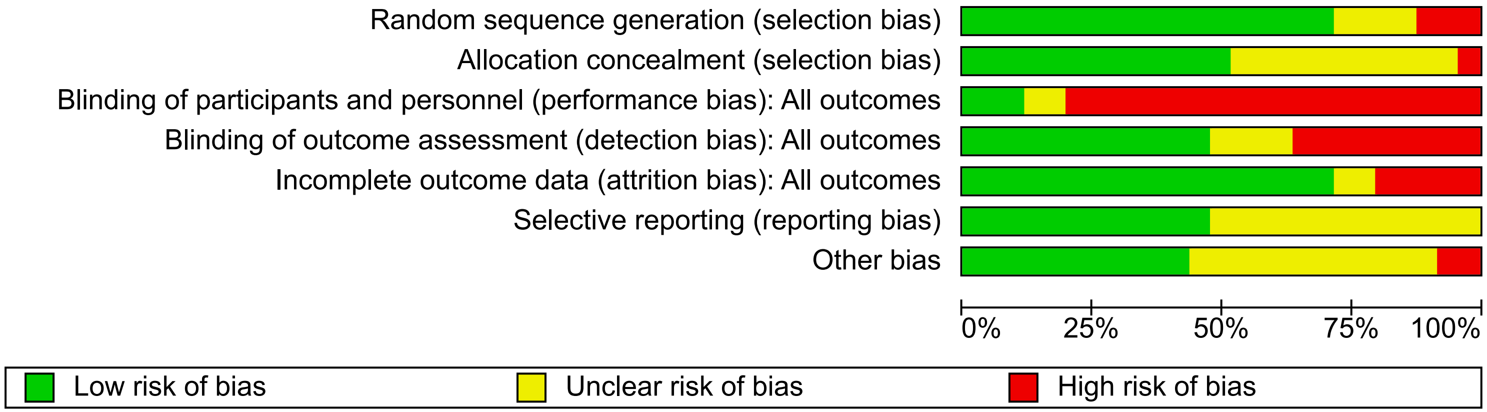
There are 33 ongoing studies. For further details of those studies, please see the Characteristics of ongoing studies.

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

Open in figure viewer



Risk of bias graph

Figure 3

Open in figure viewer

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias): All outcomes	Blinding of outcome assessment (detection bias): All outcomes	Incomplete outcome data (attrition bias): All outcomes	Selective reporting (reporting bias)	Other bias
Calabro 2003	+	+	−	+	+	?	?
Caparros-Gonzalez 2018	+	+	+	+	+	+	?
Cevasco 2008	?	+	?	−	−	?	−
Epstein 2021	+	+	−	+	+	+	+
Ettenberger 2014	+	+	−	−	−	+	?
Farhat 2010	−	?	−	−	?	?	?
Haslbeck 2021	+	+	−	+	+	+	+
Jabraeili 2016	+	?	+	+	+	?	?
Johnston 2007	+	?	−	+	−	?	+

Kehl 2021	+	+	−	+	+	+	+
Kraft 2021	+	+	−	−	−	+	−
Kucuk Alemdar 2020	+	?	−	?	+	?	+
Lafferty 2021	+	?	?	+	+	+	?
Lejeune 2019	?	?	+	+	−	+	+
Liao 2021	+	+	−	?	+	?	+
Loewy 2013	+	+	−	+	+	+	?
Menke 2021	+	+	−	−	+	+	+
Nakhwa 2017	−	−	−	?	+	?	?
Namjoo 2021	−	?	−	?	+	+	+
Portugal 2017	+	?	−	−	+	?	?
Tandoi 2015	?	?	−	−	+	?	?
Vastani 2017	+	?	−	+	+	?	+
White-Traut 1988	?	?	−	−	?	?	+
Wirth 2016	+	+	−	−	+	?	?
Yu 2021	+	+	−	+	+	+	?

Risk of bias summary

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 Table 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).

Table 5. Summary characteristics of included studies secondary outcomes **Open in table viewer**

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									
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

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

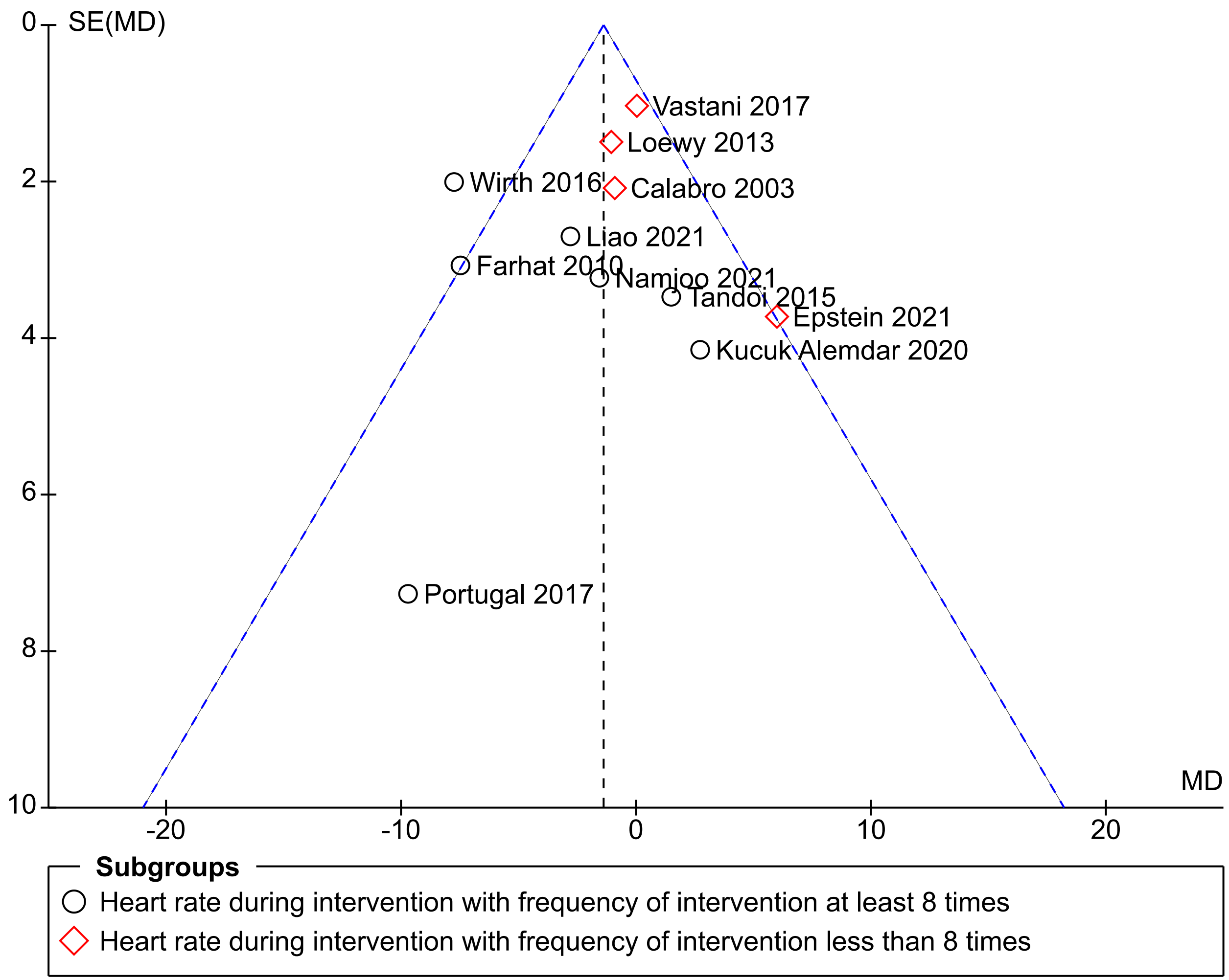
We identified 14 studies that reported measuring heart rate by pulse oximetry or electrocardiogram. In one of these studies (Ettenberger 2014), outcome data were not sufficiently reported. After receiving the requested data from the authors the amount of missing data did not allow us to include the study 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

Open in figure viewer



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 Table 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 Table 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 Table 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 Table 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).