Effect of Music on Infant Stress as Measured by Cortisol Levels and Premature Infant Pain Profile: A Systematic Review with Meta-analysis

Marissa Elizabeth L Lim, MD, MS* and Francisco Jr E Anacleto

Department of Physiology, College of Medicine, University of the Philippines, Manila, Philippines

*Corresponding author: Marissa Elizabeth L Lim, MD, MS, Department of Physiology, College of Medicine, University of the Philippines, Manila, Philippines, Tel: 09178651389

Abstract

Background: Hospitalized infants often experience pathological stress and pain with adverse short and long term effects on health outcomes. However, data regarding the benefits of music therapy for infant stress and pain are conflicting. This meta-analysis assessed music interventions for stress reduction among hospitalized infants as measured by changes in serum cortisol levels and pediatric Premature Infant Pain Profile (PIPP) Scores.

Methods: Seven databases were searched until May 2021 using the key terms ‘infant’, ‘music’, ‘pain’, ‘cortisol’, and ‘randomized trial’. Studies were selected for analysis according to the PRISMA guidelines. Review Manager 5.4 software was used for analysis. Two authors assessed the studies and extracted data. The assessment of the quality of studies was based on the risk of bias criteria of the Cochrane Collaboration. A meta-analysis was performed to analyze the changes in serum cortisol level reduction and PIPP Scores following music-based intervention.

Results: Six studies involving 424 infants were included in this review. The quality of the included studies was low to moderate, with most studies having an unclear risk of bias in allocation concealment, performance bias, and reporting bias. The analyzed data were heterogenous ($I^2 = 100$).

Conclusion: Music-based intervention did not have a significant effect on serum cortisol levels in premature infants. Although there was a trend supporting music therapy in terms of PIPP Scores, there was high heterogeneity among the studies. Additional research with better assessment technique, reproducibility of method and standardization of interventions is needed in order to perform similar studies across different groups.

Keywords
Infant, Music, Stress, Cortisol, Premature/Prematurity

Abbreviations
AOG: Age of Gestation; CENTRAL: Cochrane Central Register of Controlled Trials; CI: Confidence Interval; HPA: Hypothalamic-Pituitary-Adrenal; NICU: Neonatal Intensive Care Unit; PICC: Peripherally Inserted Central Catheter; PIPP: Premature Infant Pain Profile; PIPP-R: Premature Infant Pain Profile-Revised; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocol; PROSPERO: Prospective Register of Systematic Reviews; RCT: Randomized Controlled Trial; RevMan: Review Manager; SMD: Standard Mean Difference; SNS: Sympathetic Nervous System

Introduction

In medicine and physiology, stress is defined as reaction to a stimulus which could either be a positive challenge or a negative threat [1]. The negative aspect is often emphasized - stress is characterized as an experience brought about by an interaction of a person with his environment that brings about discomfort [2]. Consequently, stress responses can be categorized as either as a physiological arousal or an emotional response [3,4].

Worldwide, about 15 million premature infants are born each year [5]. These infants are often at-risk for pathologic processes owing to genetic abnormalities and as a result of an impaired peripartum and intrauterine environment [6]. As a result, many of these patients require several invasive procedures in the neonatal intensive care unit (NICU) aimed at improving their survival. On average, hospitalized infants experience...
an average of 10-14 painful daily events in their first 2 weeks including surgical and postoperative procedures [7]. Because pain is associated with actual or potential tissue damage, it can be an uncomfortable experience for both the senses and emotion even in newborns [8].

Pain brings about neurophysiological, behavioral and physiological responses that instinctively protect the body from harm. When an infant experiences pain, damaging impulses are transmitted from the organ affected to the spinal cord via nociceptors initiating a spinal reflex reaction. This manifests as both limbs withdrawing from the painful stimuli [9]. The damaging impulses pass through the brainstem resulting in physiologic changes, including facial grimacing and vocalizations calling the attention of the caregiver. These signals then go through the thalamus to other portions of the cortex leading to the initiation of various processes. Physiological stability may be disturbed for many hours after the painful event observed as heart rate changes, oxygen desaturations, and temporary cessation of breathing [9,10]. Pain stimulates the hypothalamic-pituitary-adrenal (HPA) axis inducing a response from the endocrine system, which is exhibited as an increase in cortisol levels [11]. Subjects less than 3-months-old were found to have higher salivary cortisol levels from baseline when undergoing interventions inducing pain such as heel lance, inoculation, and similar interventions [12]. In addition to cortisol, infant pain evaluation measures include cuff blood pressure measurement, vital sign monitoring, facial expression, crying, resting positions, fussiness or consolability, body movements and sleeplessness. However, the validity and accuracy of these techniques may be affected by confounders and may have high variability owing to different subjective methods used in various settings [8]. Healthcare providers must therefore be well-trained in assessing neonates to adequately manage pain. One of the commonly used preterm pain scales is the Premature Infant Pain Profile (PIPP). Stevens, et al. designed the PIPP in 1996, later updated to the PIPP-revised (PIPP-R) scale, to assess pain in infants as it has higher validity and reliability than other pain scales [13]. It is a 7-indicator composite scale that evaluates parameters such as heart rate, oxygenation, alertness, eye squeeze, brow bulge, and nasolabial furrow. A limitation of PIPP is that it has been validated only in infants < 28 weeks age of gestation and that it is time consuming for clinical documentation purposes, especially in the emergency department [14,15].

There is growing attention on music intervention as an alternative or adjunct in the management of stress and pain. Music interventions are easy to apply, inexpensive, and have no to minimal side effects. New research in neuroscience has shed light on how music therapy can help in relaxation and promotion of wellbeing. Primarily, music has been shown to lower neural arousal that is elevated during stressful situations [16]. Listening to music, composing music/playing musical instruments and singing have been associated with downregulated physiologic parameters such as a decrease in cortisol levels, lower heart rate and decrease in blood pressure. A likely relationship exists between music and the downregulation of stress reactions of the HPA indicating its parasympathetic mechanism of nervous system activity [17]. Data regarding music therapy for infant stress and pain are few and conflicting. A few systematic reviews have highlighted equivocal results on the efficacy of various types of music therapy on preterm infants citing variations in study quality, age groups, outcome measures and timing of the interventions as challenges [18,19].

This systematic review and meta-analysis aims to determine the association between the effect of music exposure on relief of infant stress as manifested by change in cortisol levels and secondary outcomes of change in PIPP scores.

Methods

Study design

This meta-analysis included randomized controlled trials with participating infants aged 0-12 months investigating auditory music intervention (Figure 1). This illustrates how stress is brought about by pain. Stress affects the HPA axis and brings about outcomes such as increased cortisol, sympathetic nervous system stimulation, pain and long-term depression and anxiety. Music is thought to inhibit the stress reaction causing decreased cortisol level, parasympathetic nervous system stimulation, less pain and eventually improved well-being. The protocol was performed following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocol (PRISMA) 2020 statement [20]. The present work is registered under: 1) Prospective Register of Systematic Reviews (PROSPERO) (registry ID: CRD42021254148); 2) Philippine Health Research Registry (registry ID: PHRR210510-003550), and 3) In the University of the Philippines Manila Research Grants Administration Office (ref. no. RGAO-2021-0463).

Inclusion and exclusion criteria

Studies were included if they: 1) Included infants from 0-12 months; 2) Had intervention of music compared with no intervention; 3) With outcomes measuring cortisol levels; 4) Usage of PIPP scores; 5) Randomized controlled trials. Studies were excluded if they: 1) Were in a language other than English, and 2) Did not provide adequate information to calculate effect sizes to be included in the analysis.

Primary and secondary outcomes

Primary outcome was change in serum cortisol, and secondary outcome was change in PIPP.
Search Strategy and selection process

Seven electronic databases were searched systematically for publications until May 2021 including PubMed, Cochrane Central Register of Controlled Trials (CENTRAL), DOAJ, Herdin, Trip Database, Scopus and Google Scholar. Search terms were related to the population (infant), type of intervention (music), and outcome of interest (cortisol, pain, stress,) and study design (randomized controlled trials [RCTs], quasi-experimental). In addition to searching the database, relevant key journals were searched. Significant researches were referenced with the electronic database to identify additional studies for inclusion. The bibliography of included researches was examined to further identify additional papers for inclusion. Conference proceedings were not included in the study. There were attempts to contact experts in the field, but no new material was added as a result of the communication.

Two authors independently assessed titles for possible inclusion. A consensus with a third party was reached when disagreements arose. The abstracts of eligible articles were then retrieved and screened for inclusion.

Data collection, extraction and analysis

A data extraction form was used to record demographics, study design and outcome measures to evaluate data from the included studies, based on the Cochrane Collaboration Recommendation. Data were extracted and analyzed by two authors.

Two reviewers separately critically appraised each study based on the revised Cochrane Risk of Bias tool for randomized controlled trials [21]. The quality assessment included seven characteristics: Random sequence generation, allocation concealment, blinding of participants or personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other bias. Each characteristic was rated low, high, or unclear in the risk of bias. Review Manager (RevMan) software 5.4 was used to provide the risk of bias graph and summary.

Authors were contacted by email when missing data were not available. Data provided in medians were transformed to means using the formulas of SP Hozo, B Djulbegovic and I Hozo (2005), Luo, et al. (2018) and Wan, et al. (2014). Graphical data were extracted to numerical data utilizing the online WebPlotDigitizer.

Data synthesis was done using the software Review Manager version 5.4. Continuous variables were reported as standard mean difference (SMD) with 95% confidence interval (95% CI). Data were analyzed using a random effects model with generic inverse variance methods due to variability. Chi-square test and the

<table>
<thead>
<tr>
<th>Author (Year)</th>
<th>Country</th>
<th>Method</th>
<th>Subjects</th>
<th>Intervention</th>
<th>Outcome</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barandouzi</td>
<td>Iran</td>
<td>Double blinded RCT</td>
<td>Neonates</td>
<td>10 minutes recorded Brahms Lullaby prior venipuncture with or without sucrose vs. control</td>
<td>Change in PIPP score</td>
<td>120</td>
</tr>
<tr>
<td>Bergomi</td>
<td>Italy</td>
<td>RCT</td>
<td>Premature infants</td>
<td>Recorded Mozart music for 18 minutes maximum vs. glucose vs. control prior to heel lancet</td>
<td>Change in PIPP score</td>
<td>35</td>
</tr>
<tr>
<td>Qiu</td>
<td>China</td>
<td>RCT</td>
<td>Preterm neonates</td>
<td>Recorded Smart Baby Lullaby from 5 minutes before until 30 minutes after varied painful procedures in combination with touch intervention vs. control</td>
<td>Change in serum cortisol and PIPP score</td>
<td>62</td>
</tr>
<tr>
<td>Shukla</td>
<td>India</td>
<td>RCT</td>
<td>Neonates age 26-36 months AOG</td>
<td>Recorded Indian classical flute music with or without kangaroo care 10 minutes before until 5 minutes post heel prick procedure vs. control</td>
<td>Change in PIPP score</td>
<td>200</td>
</tr>
<tr>
<td>Tang</td>
<td>China</td>
<td>Random, data-analysts-blinded, controlled trial, two parallel arms</td>
<td>Hospitalized infant</td>
<td>Chinese lullabies 10 minutes before until 10 minutes after PICC vs. control</td>
<td>Change in serum cortisol and PIPP score</td>
<td>60</td>
</tr>
<tr>
<td>Tekgündüz</td>
<td>Turkey</td>
<td>Double-blind, randomized controlled design</td>
<td>Preterm infants</td>
<td>Lullabies played during the tracheal tube removal and reininsertion vs. glucose vs. control, time not specified</td>
<td>Change in PIPP score</td>
<td>106</td>
</tr>
</tbody>
</table>

AOG: Age of Gestation; PICC: Peripherally Inserted Central Catheter; PIPP: Pediatric Infant Pain Profile Scores; RCT: Randomized Controlled Trial
heterogeneity statistic, $I^2$, were used to evaluate the statistical heterogeneity between treatment effects across the trials. Low, moderate, and high risk was assigned to $I^2$ values of 25-50%, 51-75%, and 76-100% respectively [21]. A funnel plot for analysis of publication bias was not done due to the small study number.

Results

Search results

The database search resulted in 2,011 titles. After duplicates were removed, 1,860 titles remained. Titles were reviewed and abstracts ($n = 141$) were screened. All titles that could not be that clearly excluded had the corresponding abstracts retrieved. After abstracts were read and assessed for eligibility, full-text reading ($n = 12$) was done. Six full-text articles were included in the meta-analysis (Table 1). Two articles were included for cortisol levels, and six for meta-analysis of the PIPP (Figure 2).

Description of studies

The studies were published from 2014-2019. Two of the studies were performed in China, and one study each was performed in Iran, Italy, India and Turkey. All the interventions were randomized controlled design. Sample sizes ranged from 60-200. Infants gestational ages ranged from 26-36 months age of gestation. Music interventions varied from lullabies, classical music and native music. Timing of music intervention varied in relation to the different pain-producing procedures. The pain-producing procedures were varied, including tracheal tube removal, insertion of central line, heel prick, lancet and venipuncture. Music exposure time ranged from 10 to 35 minutes. There were differences in time points of measure of the cortisol levels, as well as the PIPP scores. Outcomes were measured before and after the music intervention and all studies were performed in the hospital setting.

Methodological quality assessment

Risk of bias assessments is shown in Figure 3. There was unclear assessment of the risk of bias in most of the studies. For random sequence generation of included studies, two trials failed to mention if randomization was performed [22,23] two studies randomized without a stated method [11,22], while two studies detailed the use of computer randomization [24,25]. Allocation concealment was not clearly addressed for most studies. Only Shukla mentioned the use of opaque envelopes for concealment [25]. Performance bias was not addressed in all studies likely because music intervention is difficult to conceal from the personnel and the participants of the study. Detection bias was only clearly addressed by Bergomi [22] and Tang [24] who mentioned that assessments were made by researchers blinded to the study treatment. In the study done by Shukla [25], the PIPP score was graded based on a muted video so that the presence or absence of music intervention would be blinded during the assessment. Attrition bias was addressed by the two studies, namely by Barandouzi and by Qiu as the dropout of participants for various reasons was identified [11,26]. Two studies, one by Shukla and on by Tang, mentioned that all the participants were analyzed and no dropouts were noted [24,25]. Reporting bias was addressed by two studies, by Shukla and by Barandouzi with complete data reported [25,26].

Figure 1: The conceptual framework.
Figure 2: PRISMA flow chart of included studies.
424 infants (Figure 5). The random-effects model showed that PIPP Score change was significant ($p < 0.00001$) favouring music therapy, but with very high heterogeneity ($I^2 = 98\%$).

**Discussion**

Six studies with 242 infants were included in the systematic review. There was very high heterogeneity in all studies included in this analysis. There was no significant difference in the change of serum cortisol levels between groups receiving music therapy as opposed to those who did not. One possible reason for
this is blood extraction is a painful process that likely contributed to the increase in cortisol levels. There was a significant difference in the change in the PIPP Scores, but results were weak due to the very high heterogeneity of studies. This phenomenon could be explained by the inherent stress that is present by being in the neonatal intensive care unit (NICU) at baseline even before painful procedures are performed and long after the effects have worn off.

No clear masking or soundproofing of the music intervention to limit exposure in the treatment group was specified in all studies, which could lead to a leak of the music intervention to the control group.

The increased heterogeneity may be due to factors such as study design inconsistencies, sample size differences, individual variability, diverse type and frequency of treatment sessions. There was lack of standardization of the type of music across all the studies. There were differences in the auditory inputs presented to the infants, dependent on the country of origin and cultural differences. These differences may be significant as there is evidence that fetuses and neonates are able to distinguish between different types of music based on reactions such as heart rate accelerations, increased motor response, and synchronization of the music to sucking, among others [27].

The heterogeneity may also be explained by the differences in pain-inducing and stress evoking procedures among the studies, which ranged from heel prickling, central line insertions, venipunctures and laryngeal tube changes. The heterogeneity could not be resolved by subgroup analysis due to the limited number of studies analyzed. Some systematic reviews and/or meta-analyses studied music interventions that improve clinical outcomes. These include the following methods: pre-recorded music, one-to-one music sessions, or group sessions with a music therapist. In all of these reviews, the music interventions had positive influence on vital signs of the participants [28-30]. More recently, an infant meta-analysis was done by the group of Yue in 2020. They studied the effect of music therapy on 1,099 preterm infants admitted at the NICU. Their results showed that music may greatly improve the heart rate, respiratory rate, stress level, and oral feeding of preterm infants [31].

To the best of the authors’ knowledge, there are no other reviews that examined the effect of music on neuroendocrine parameters in infants. Many of the existing trials have inherent differences in study design and protocol implementation and do not outline the reason for the choice of music, music duration or intensity. The range of music sessions were from one to 3 sessions. Length of music exposure ranged from 10 to 30 minutes. Other reviews reflected this inconsistency as well. The short duration and intensity of music interventions may have limited benefits on infant well-being. Studies evaluating longer exposures may be explored.

In the future, it would be ideal for music-based intervention to have clear guidelines. This would make the replicability of the studies less difficult for both research and clinical settings. Theories behind the selection of music intervention, cultural differences, environment of musical delivery, musical content music type and selection may be included in the guidelines.

Due to the complex nature of infant stress, it is often overlooked as an aspect of care. In previous years, several standardized, reliable and valid infant pain assessment tools have been proposed. However, there is a lack of standardization and utility among all the different infant scoring systems for pain and stress and the relative convenience of testing remains to be addressed and may be the subject of future research. Other methods evaluating stress and pain may also be looked into by future studies.

This study has several limitations. First, data were not fully available and some studies were difficult to obtain. Second, the heterogeneity of music intervention makes it difficult to draw conclusive results. In addition, this heterogeneity was not resolved as subgroup analysis was not performed due to the small study and sample size number of the component studies. Given that these studies were conducted in different countries of the world, significant variations in antenatal care and musical exposure is to be anticipated. Finally, most of the researches reported on short-term effects. The lasting effects of music therapy were not discussed in this paper. As a result of these limitations, the meta-analysis failed to answer the questions about the prolonged effect of music therapy on infants. Caution is necessary in drawing conclusions on the applicability of the results.

Conclusion

Evidence supporting musical intervention in the alleviation of stress and/or pain in infants remains inconclusive due to the high heterogeneity of currently available clinical studies. Further investigations focusing on better assessment techniques, reproducibility of method and standardization of intervention are necessary in order to perform similar studies across different and gather more data. Determining the best type, dose, and method of music-based intervention may be good research topics in the future. New studies are recommended to include the intervention duration and frequency. Longer follow-up and inclusion of other physiological measures of distress may be included in futures studies.

Acknowledgments

The authors would like to thank Dr. Elizabeth S. Montemayor, the University of the Philippines College
of Medicine. Department of Physiology headed by Dr. Leticia T. Ibanez. We also would like to extend our gratitude to Dr. Clarissa Lim Velayo, Dr. Mark Anthony S. Sandoval, Dr. Myna B. Sedurante and Dr. Teresita Joy P. Evangelista for their guidance.

**Conflict of Interest**

The authors declare no conflict of interest.

**Funding**

Self-funded by the study authors.

**Statement of Equal Authors’ Contribution**

All authors have contributed equally.

**Registration**

The protocol was registered in the following: 1) PROSPERO (registry ID: CRD42021254148); 2) Philippine Health Research Registry (registry ID: PHRHR210510-003550) and 3) In the University of the Philippines Manila Research Grants Administration Office (ref. no. RGAO-2021-0463).

**References**

