healthcare

MDPI

Systematic Review

Music During Electronic Fetal Monitoring in Healthy Pregnancies: A Systematic Review and Meta-Analysis

Ezgi Keskin Gölçek [1], Eslem Altıntaş [2,*], İlkay Güngör Satılmış [3] and Özden Tandoğan [4]

[1] Department of Medical Services and Techniques, Vocational School of Health Services, Demiroğlu Bilim University, 34570 Silivri, Türkiye; ezgi.keskingolcek@demiroglu.bilim.edu.tr
[2] Department of Nursing, Faculty of Health Sciences, Acibadem Mehmet Ali Aydinlar University, 34752 İstanbul, Türkiye
[3] Department of Women's Health and Diseases Nursing, Florence Nightingale Faculty of Nursing, Istanbul University-Cerrahpaşa, 34381 Istanbul, Türkiye; ilkay.gungorsatilmis@iuc.edu.tr
[4] Department of Nursing, Faculty of Health Sciences, Istanbul Arel University, 34010 Istanbul, Türkiye; ozdentandogan@arel.edu.tr
* Correspondence: eslem.altintas@acibadem.edu.tr or kuruleslem@gmail.com

Abstract

Background/Objectives: Non-reactive non-stress tests (NSTs) in healthy pregnancies are often caused by fetal sleep rather than by fetal compromise, and music has been proposed as a simple means of increasing fetal reactivity. This systematic review and meta-analysis evaluated whether listening to music during electronic fetal monitoring changes fetal movements, baseline fetal heart rate (FHR), FHR accelerations and NST reactivity in healthy pregnant women. **Methods:** The review followed PRISMA 2020 and was registered in PROSPERO (CRD42024481527). Four databases (PubMed, Web of Science, Scopus and ScienceDirect) were searched from inception to June 2024 (date of last search) for randomized, quasi-randomized, non-randomized or crossover-controlled studies, published in English or Turkish, comparing music during fetal monitoring with no music. Methodological quality was appraised with the JBI tools and the certainty of evidence with GRADE. Random-effects (restricted maximum likelihood) meta-analyses used Hedges' g, mean differences (MD) and risk ratios (RR); the primary analyses were restricted to uncomplicated pregnancies and were stratified by study design. **Results:** Twelve studies (1747 women) were included; eight contributed to the primary meta-analyses, three were used in sensitivity analyses only, and one was synthesized narratively. Music—mostly slow-tempo classical, instrumental or culturally familiar pieces, self-selected by the woman in four studies and delivered through headphones or a personal player for 10–30 min with playback volume adjusted for comfort where reported—increased fetal movements (6 studies, $n = 549$; $g = 0.69$, 95% CI 0.30–1.07; $I^2 = 74\%$) and FHR accelerations (7 studies, $n = 688$; $g = 0.60$, 95% CI 0.24–0.97; $I^2 = 78\%$) and modestly increased the proportion of reactive NSTs (5 studies, $n = 687$; RR = 1.29, 95% CI 1.04–1.59; $I^2 = 90\%$), without changing baseline FHR (8 studies, $n = 909$; MD = 1.3 bpm, 95% CI −0.5 to 3.2). Heterogeneity was high, prediction intervals included no effect, the reactivity effect was not robust to the Hartung–Knapp adjustment or to restriction to randomized trials, and the certainty of evidence was low (very low for reactivity). **Conclusions:** Music during the NST is a promising, low-risk adjunct that may increase fetal activity and accelerations without altering baseline FHR, but the evidence comes from small, heterogeneous and partly non-randomized studies identified in four databases and two languages; one study reported a shorter time to a reactive result with music, but no study assessed total test duration, downstream clinical management (repeat testing, additional investigations) or perinatal outcomes. Adequately powered randomized trials with standardized music protocols are needed before music can be recommended for routine use.

Check for updates

Academic Editor: Daniele Giansanti

Received: 28 September 2026
Revised: 24 October 2026
Accepted: 26 October 2026
Published: 18 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: music; music intervention; cardiotocography; fetal monitoring; non-stress test; fetal movement; fetal heart rate; pregnancy; antenatal care

1. Introduction

Assessment of fetal well-being is a core component of antenatal care, because it aims to identify fetuses at risk of intrauterine death or neonatal compromise and to allow timely intervention [1]. The burden that this care seeks to reduce remains considerable: in 2023, approximately 2.3 million newborns died in the first month of life, and almost 2 million babies were stillborn at 28 weeks or more of gestation, most of them in low- and middle-income countries [2,3].

Electronic fetal monitoring (EFM), or cardiotocography, records the fetal heart rate (FHR) together with uterine activity and is one of the most frequently used obstetric assessment methods [4]. Its interpretation is based on the baseline FHR, variability, accelerations and decelerations; the baseline is the mean FHR, rounded to increments of 5 beats per minute (bpm), over a 10 min window, and a normal baseline ranges from 110 to 160 bpm [5]. In the antepartum period, EFM is performed as the non-stress test (NST), in which a reactive result—two or more accelerations of at least 15 bpm lasting at least 15 s within 20 min (10 bpm for 10 s before 32 weeks)—is reassuring [1]. However, the NST has a high false-positive rate: non-reactive results are frequently caused by fetal sleep cycles rather than compromise, leading to a longer total test duration, repeat testing, additional investigations and maternal anxiety [6,7]. Various methods have therefore been proposed to increase fetal reactivity during the test, including maternal glucose intake, manual stimulation and vibroacoustic stimulation [6,7].

Maternal psychological state is a further factor that may influence NST findings. Maternal stress and anxiety during pregnancy have been associated with adverse obstetric outcomes and with changes in the fetal hypothalamic–pituitary–adrenal axis, possibly through placental transfer of cortisol and altered uteroplacental blood flow [8,9]. Maternal stress is also reflected, non-invasively, in fetal heart rate variability [10], and induced maternal relaxation has been shown to modify fetal heart rate and movement patterns [11]. Because the NST itself can be a source of anxiety for pregnant women, non-pharmacological interventions that promote relaxation during the test are of clinical interest [12].

Music has been used for therapeutic purposes since antiquity and is increasingly applied in obstetrics as a non-invasive, low-cost intervention [13]. A distinction is made between music therapy, which is delivered by a trained music therapist within a therapeutic relationship, and music medicine, in which pre-recorded music is offered to patients by other healthcare professionals [14]; the interventions evaluated during fetal monitoring belong to the latter category and are referred to here as music interventions. Music interventions reduce stress and anxiety during pregnancy [15–17], and the World Health Organization lists music among relaxation techniques that may support women during labour [18]. From the third trimester onwards, the fetus perceives and responds to external sounds [19,20], and music exposure has been associated with changes in fetal heart rate and movements [21] as well as in neonatal behaviour [22]. Music may therefore act on the fetus both indirectly, through maternal relaxation, and directly, as an auditory stimulus; the fetal response is likely to depend on the characteristics of the stimulus (type, tempo and sound level) and on whether the sound reaches the fetus through the abdominal wall or acts mainly on the mother [19–21].

Several primary studies have examined the effect of music on NST parameters, with conflicting results: some reported increased fetal movements, accelerations or reactivity,

while others found no effect on baseline FHR or accelerations [12,23,24]. Previous systematic reviews on music during pregnancy have mainly focused on maternal anxiety or labour outcomes, have combined low- and high-risk pregnancies, or have synthesized fetal outcomes narratively [15,25,26]. To the best of our knowledge, no meta-analysis has quantitatively pooled NST outcomes—including accelerations and test reactivity—in healthy pregnancies. The present review considers music as an intervention delivered during EFM with the aim of improving the parameters of the test itself, rather than as a maternal anxiety intervention; maternal relaxation is regarded as one possible mechanism of fetal effects and not as an outcome. The objective was to evaluate the effect of music during electronic fetal monitoring on fetal movements and baseline FHR (primary outcomes) and on FHR accelerations and NST reactivity (secondary outcomes) in healthy pregnant women.

2. Materials and Methods

2.1. Protocol and Registration

The review was prospectively registered in the International Prospective Register of Systematic Reviews (PROSPERO; CRD42024481527) and is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [27]. Differences from the registered protocol are reported in Section 2.9.

2.2. Eligibility Criteria

Eligibility was defined using the PICOS framework. Population: pregnant women with uncomplicated pregnancies—without known maternal or fetal complications (e.g., preeclampsia, gestational diabetes requiring treatment, fetal growth restriction, multiple pregnancy, premature rupture of membranes)—undergoing antepartum EFM/NST. Studies whose samples consisted of, or explicitly included, women with a maternal disease or with a specific clinical indication for testing (e.g., reduced fetal movements) were classified as mixed-risk samples; they were retained in the systematic review but were excluded from the primary meta-analyses and used only in sensitivity analyses (Section 2.8). Intervention: listening to music during fetal monitoring, delivered to the mother (e.g., via headphones or speakers) or to the fetus (via abdominal speakers), irrespective of whether the authors framed the music as fetal stimulation or as maternal relaxation. Comparison: fetal monitoring without music. Outcomes: the primary outcomes, as registered, were fetal movements and baseline FHR; the number of FHR accelerations and the NST result (reactive/non-reactive) were secondary outcomes added during the review (Section 2.9). Three related measures were not pre-specified outcomes and were summarized narratively where reported: the time to a reactive result (the time from the start of the recording until the reactivity criteria were met), the total test duration (the total recording time, including any prolongation of a non-reactive test) and downstream clinical management (e.g., repeat testing or additional investigations). Study design: randomized controlled trials (including crossover trials) and quasi-randomized or non-randomized controlled studies. Studies published in English or Turkish with accessible full texts were included. Reviews, case reports, cross-sectional studies, conference abstracts, book chapters and studies restricted to high-risk pregnancies were excluded.

Studies in which music was combined with another active intervention that could independently affect NST results (e.g., food intake or video) were retained in the systematic review but were included in the meta-analyses only as a sensitivity analysis. Studies without a concurrent or within-subject no-music comparison, or without extractable outcome data, were synthesized narratively.

2.3. Information Sources and Search Strategy

PubMed, Web of Science, Scopus and ScienceDirect were searched from inception to June 2024 without date restrictions. The search combined terms related to music and complementary therapies (music, music therapy, complementary medicine, traditional and complementary medicine, alternative medicine) with terms related to pregnancy (pregnancy, pregnant women) and to fetal outcomes (fetal monitoring, fetal heart rate, fetal movement, vital signs, anthropometric measurements), using Boolean operators; in addition, a hand search was performed. The strategy was deliberately broad because music interventions in pregnancy are indexed inconsistently, which explains the large number of records retrieved (Section 3.1).

2.4. Study Selection

Records were imported into EndNote (version 9.0). Before formal screening, two reviewers (E.K.G. and E.A.), working together, removed in EndNote the duplicate records and the records that clearly did not meet the eligibility criteria on the basis of their titles (e.g., records not concerning pregnancy or music, or non-eligible publication types); all records removed at this step were then checked independently by a third reviewer (Ö.T.) to confirm that no potentially eligible record had been removed. Because duplicate removal and title-based exclusion were performed in a single step, the number of duplicates was not recorded separately and could not be reconstructed afterwards; the two categories are therefore reported as a combined total (Section 3.1 and Figure 1). Two reviewers (E.A. and E.K.G.) then independently screened the remaining titles and abstracts and assessed full texts against the eligibility criteria. Disagreements were resolved by discussion with the other authors.

2.5. Data Extraction

Two reviewers (E.K.G. and E.A.) independently extracted data using an Excel form developed by the authors, including author, year, country, design, setting, sample size, gestational age, type, timing and duration of music, delivery device, comparator, NST duration, method of counting fetal movements and outcomes. The two extraction forms were compared, and discrepancies were resolved by discussion and, when necessary, by consulting a third reviewer (İ.G.S.). For continuous outcomes, the number of participants, means and standard deviations (SDs) at the end of the intervention were extracted; for NST reactivity, the number of reactive tests per group was extracted. All extracted values were re-checked against the full texts before analysis; study authors were not contacted for additional data.

The following decision rules were applied a priori or, where indicated, during data extraction. (i) For multi-arm studies, only the comparison of a music arm with a no-music arm was used, so that no control group was counted twice; when two music arms were compared with one control group (Brillo et al.), the music arms were combined using the Cochrane formula [28]. (ii) For the randomized crossover study (Denizli et al.), the change in fetal movements differed between the AB and BA sequences, indicating a period or carry-over effect; in this situation, the Cochrane Handbook recommends using only first-period data, which converts the trial into a parallel-group comparison and avoids carry-over bias at the cost of precision [28,29]. First-period data were therefore used in the primary analysis, and both periods—analyzed as a parallel comparison, which is conservative because the within-woman pairing is ignored [28]—were used in a sensitivity analysis. (iii) When a study reported only a t statistic (Fathi et al.), the standardized mean difference was derived from the t value with equal group sizes. (iv) Post-intervention values were used (e.g., baseline FHR at the end of the NST in Estrella-Juárez et al.). (v) When fetal movements

were recorded both by the mother and by the device, device-recorded values were used. (vi) "Dubious" NSTs were classified as non-reactive.

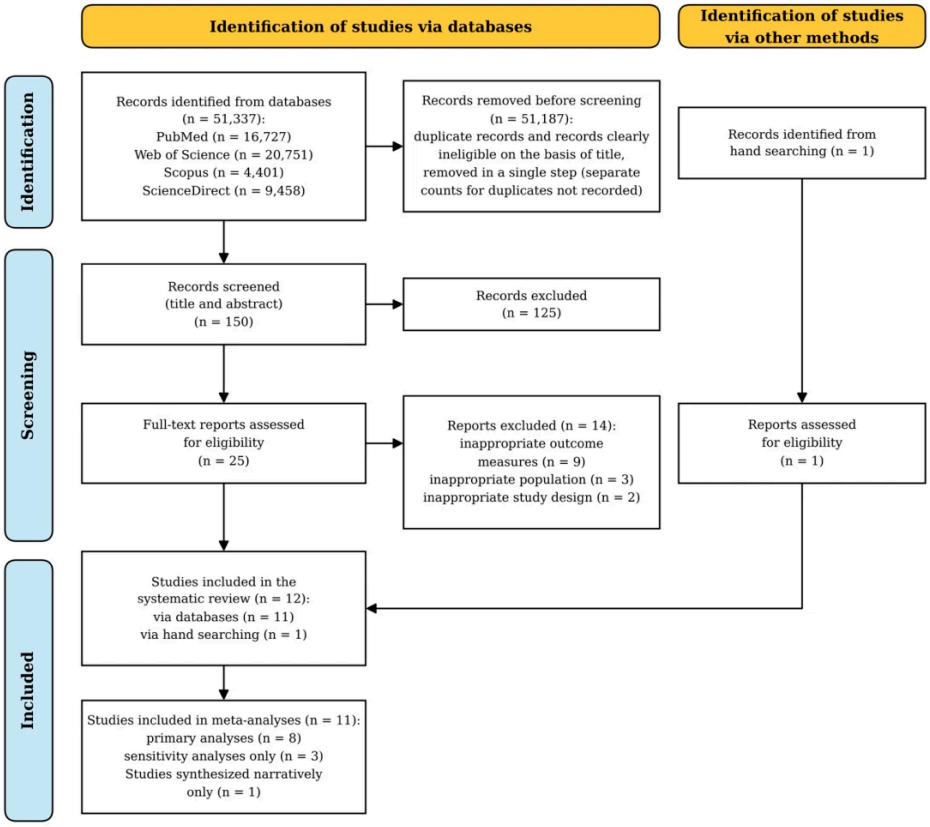


Figure 1. PRISMA 2020 flow diagram of study selection. Duplicate records and records that were clearly ineligible on the basis of title were removed in a single step before screening, so their separate numbers are not available (Section 2.4). PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

2.6. Methodological Quality

Two reviewers independently appraised each study using the revised JBI critical appraisal tools for randomized controlled trials [30] and for quasi-experimental studies [31]; disagreements were resolved by consensus. Studies were classified by design on the basis of the allocation method actually reported, not the label used by the authors; studies that allocated participants by day of the week or by participant-selected envelopes/cards were classified as quasi-randomized. The JBI ratings were used to inform the risk-of-bias domain of GRADE and the interpretation of the design-stratified analyses (Sections 2.7 and 2.9).

2.7. Effect Measures and Data Synthesis

Because fetal movements and accelerations were counted using different methods and time windows, continuous outcomes were pooled as standardized mean differences using Hedges' g, which corrects for small-sample bias [32]; baseline FHR, measured in the same unit in all studies, was additionally pooled as a mean difference (MD, bpm). NST reactivity was pooled as a risk ratio (RR) of a reactive test. Positive g or MD values and RR > 1 indicate higher values in the music group. Effect sizes of 0.2, 0.5 and 0.8 were interpreted as small, moderate and large, respectively.

Given the expected clinical and methodological heterogeneity, random-effects models were used, with between-study variance (τ^2) estimated by restricted maximum likelihood (REML) [33]. Heterogeneity was assessed with Cochran's Q test ($p < 0.10$) and the I^2

statistic (25%, 50% and 75% indicating low, moderate and high heterogeneity) [34], and 95% prediction intervals were calculated to describe the expected range of effects in a new setting [35]. Because the number of studies was small, the Hartung–Knapp–Sidik–Jonkman (HKSJ) adjustment was applied to every pooled estimate as a sensitivity analysis [36]. A meta-analysis was performed when at least two studies reported the same outcome.

A pre-specified analysis hierarchy was followed. The primary analysis for each outcome pooled the studies of uncomplicated pregnancies irrespective of design, because only two to five randomized studies were available per outcome. Because pooling randomized and non-randomized studies can overstate intervention effects [28], (a) all forest plots were stratified by design (randomized studies, including the randomized crossover trial, versus quasi-/non-randomized studies), with separate subgroup estimates and a test for subgroup differences; (b) the estimate from randomized studies only was reported alongside the overall estimate for every outcome; and (c) an effect was interpreted as robust only if its direction and statistical significance were maintained in the randomized-only, HKSJ-adjusted and leave-one-out analyses. Secondary analyses comprised the sensitivity analyses described in Section 2.8. Conclusions were based primarily on the two registered primary outcomes (fetal movements and baseline FHR); the secondary outcomes (accelerations and NST reactivity) were interpreted in the light of these.

2.8. Subgroup, Sensitivity and Reporting-Bias Analyses

Subgroup analysis was performed according to study design (randomized vs. quasi-/non-randomized). Sensitivity analyses comprised (i) restriction to randomized studies; (ii) the HKSJ adjustment; (iii) leave-one-out analysis; (iv) an extended-population analysis adding the two mixed-risk samples (Fathi et al., whose sample included women with diabetes or hypertension, and Khodakarami et al., whose sample consisted of women presenting with reduced fetal movements); (v) addition of the multicomponent study by Timur Taşhan et al.; and (vi) use of both crossover periods for Denizli et al. Meta-regression was not attempted because fewer than ten studies were available per outcome [28]. Small-study effects were explored with funnel plots and Egger's regression test [37]; because fewer than ten studies were available for every outcome, these tests have low power and were interpreted as exploratory [38].

Analyses were performed in Python 3.13.17 (NumPy 2.5.3, SciPy 1.18.1 and Matplotlib 3.11.2) using scripts implementing standard formulas [28,33]; study-level effect sizes were cross-checked against Comprehensive Meta-Analysis (version 4). The extracted data (File S1) and the complete, annotated analysis code, including the computational specifications of all effect-size transformations and models (File S2), are provided in the Supplementary Materials so that every result can be reproduced independently.

2.9. Certainty of Evidence and Changes to the Protocol

The certainty of evidence for each outcome was assessed using the GRADE approach, considering risk of bias, inconsistency, indirectness, imprecision and publication bias [39]. Because the body of evidence combined randomized and non-randomized studies, the risk-of-bias domain was rated with reference to the JBI appraisal and to the design-stratified analyses. Inconsistency was rated as serious when I^2 exceeded 50% and the 95% prediction interval included no effect [40]. Imprecision was judged from the width of the 95% confidence interval in relation to the thresholds for small, moderate and large effects, from the optimal information size (a rule of thumb of at least 400 participants for continuous outcomes) and from whether the HKSJ-adjusted and randomized-only confidence intervals included no effect [41]. Outcome-specific reasons for each judgement are given as footnotes to the summary-of-findings table. Compared with the registered protocol, quasi-

randomized, non-randomized and crossover-controlled studies were included in addition to randomized trials; FHR accelerations and NST reactivity were added as secondary outcomes during full-text review, when it became apparent that most studies reported them, and before the quantitative synthesis was performed; and GRADE assessment was added.

3. Results

3.1. Study Selection

The database searches, which were deliberately broad (Section 2.3), identified 51,337 records (PubMed 16,727; Web of Science 20,751; Scopus 4401; ScienceDirect 9458), and one additional record was identified by hand searching. Before screening, 51,187 records were removed in EndNote in a single step because they were duplicates or clearly did not meet the eligibility criteria on the basis of their titles (e.g., records not concerning pregnancy or music, or non-eligible publication types); this step was performed by two reviewers working together, and all removed records were checked independently by a third reviewer (Section 2.4). Because duplicates and title-based exclusions were removed together, their separate numbers were not recorded and cannot be reported (Figure 1). The remaining 150 records were screened by title and abstract, 125 were excluded, and 25 full-text articles were assessed for eligibility. Fourteen articles were excluded because of inappropriate outcome measures ($n = 9$), population ($n = 3$) or study design ($n = 2$), leaving 11 eligible studies from the database search. The study identified by hand searching also met the eligibility criteria; thus, 12 studies were included in the systematic review (Figure 1). These studies are listed in Table 1 [12,23,24,42–50]. Eight studies of uncomplicated pregnancies contributed to the primary meta-analyses, three studies (two mixed-risk samples and one multicomponent trial) were used in sensitivity analyses only, and one study was synthesized narratively.

Table 1. Characteristics of the included studies.

Main Results (Music vs. Control)	Outcomes Used	Comparator	Music Intervention	Participants (Music/Control); Gestational Age	Design	Study, Country
FM 8.9 ± 4.7 vs. 5.9 ± 3.9 *; FHR 134.1 ± 7.2 vs. 130.3 ± 5.7 *; ACC 5.7 ± 2.1 vs. 4.5 ± 2.0 *; dubious NST 1 vs. 5; time to a reactive result 13.4 ± 5.2 vs. 15.6 ± 6.1 min *	FM, FHR, ACC, reactivity	Blank cassette	Self-selected slow-tempo classical, Turkish art or folk music via headphones throughout 30 min NST	201 (96/105); low-risk; 36 wk (mean 37.7 wk)	RCT	Kafali 2011 [12], Türkiye
2nd 10 min: FHR 138.0 ± 9.1 vs. 138.7 ± 8.9; ACC 1.1 ± 1.2 vs. 0.9 ± 1.1 (within-group increase $p = 0.013$)	FHR, ACC	No music	Maternal arm: instrumental music (Paganini, Chopin) via headphones during 2nd 10 min of NST; fetal acoustic-stimulation arm not used	213 (71/71/71); 37–41 wk	RCT, 3 arms	Khoshkholgh 2016 [24], Iran

Table 1. *Cont.*

Main Results (Music vs. Control)	Outcomes Used	Comparator	Music Intervention	Participants (Music/Control); Gestational Age	Design	Study, Country
FM 3.02 ± 1.81 vs. 2.06 ± 0.90 *; FHR 135.0 ± 8.8 vs. 137.1 ± 8.2; ACC 2.14 ± 1.12 vs. 1.29 ± 1.09 *; reactive 30/48 vs. 16/48 *	FM, FHR, ACC, reactivity	Routine NST	Self-selected from playlist (classical, Turkish art/folk, Rehavi makam, Kurdish, hymns) via MP3 during 15–20 min NST	96 (48/48); ≥33 wk; multiparous	Non-randomized (weekday allocation)	Şimşek Küçükkelepçe 2018 [42], Türkiye
FM 2.52 ± 0.70 vs. 2.48 ± 0.58; FHR 136.5 ± 8.9 vs. 135.1 ± 7.1; ACC 8.9 ± 3.9 vs. 4.3 ± 3.2 *; reactive 49/50 vs. 33/50 *	FM, FHR, ACC, reactivity	Routine NST	Self-selected Turkish classical, folk or classical music via MP3 during 20 min NST	100 (50/50); 32–41 wk; multiparous	Non-randomized (weekday allocation)	Erkun Dolker 2019 [43], Türkiye
First period: FM 14.4 ± 3.0 vs. 9.7 ± 3.8; FHR 133.4 ± 6.1 vs. 135.4 ± 7.2; ACC 6.4 ± 2.4 vs. 4.1 ± 1.5; both periods: FM 11.1 ± 4.6 vs. 9.2 ± 3.4 *	FM, FHR, ACC (first period)	Same women without music	Tchaikovsky (Sleeping Beauty, Swan Lake) during 20 min NST; 10 min interval between periods	50 (25/25); 34–37 wk	Randomized AB/BA crossover	Denizli 2019 [44], Türkiye
Device-recorded FM 3.31 ± 1.26 vs. 1.20 ± 1.39 *; FHR 136.8 ± 9.3 vs. 139.3 ± 10.4; reactive 79/80 vs. 31/80 *	FM, FHR, reactivity (mixed-risk sample; sensitivity analyses only)	No music	Non-verbal music via headphones for 30 min, starting 10 min before a 20 min NST	160 (80/80); nulliparous; 37–40 wk; reduced fetal movements	RCT	Khodakarami 2020 [45], Iran
Mean FHR before/after reported without SD; not poolable	Narrative only	No music	Classical music, natural sounds or murottal, single 30 min session	40 (10 per group)	Quasi-experimental, pre–post, 4 groups	Sumaningsih 2020 [46], Indonesia
Music phase: FHR 139.9 vs. 138.9; ACC 5.1 vs. 4.3; FM 8.8 vs. 8.2 (NS); ACC and FM increased during music at the 4th NST in group A	FM, FHR, ACC (A + B vs. C)	No music	Music via abdominal speakers (~10 min) during four NSTs; group A also pre-listening from 33 wk	30 (10/10/10); enrolled 28–32 wk; NSTs at 37–38 wk	RCT (pilot), 3 arms	Brillo 2021 [47], Italy
FM 9.08 ± 3.84 vs. 5.22 ± 2.66 *; FHR 144.4 ± 7.8 vs. 141.5 ± 6.7; ACC 4.45 ± 1.75 vs. 4.55 ± 2.04; reactive 35/36 vs. 34/36	FM, FHR, ACC, reactivity	Routine NST	Self-selected genre (Turkish pop, hymns, classical, other) via MP3 during NST	74 (37/37), 72 analysed; ≥32 wk	Quasi-randomized (weekday allocation)	Soylu 2022 [23], Türkiye

Table 1. *Cont.*

Main Results (Music vs. Control)	Outcomes Used	Comparator	Music Intervention	Participants (Music/Control); Gestational Age	Design	Study, Country
Video vs. cake: FM 3.09 ± 0.73 vs. 3.05 ± 0.83; ACC 3.11 ± 0.73 vs. 3.06 ± 0.83; FHR 133.6 ± 9.2 vs. 132.2 ± 8.8; reactive 112/120 vs. 116/120 (control 88/120)	Sensitivity analysis only (multicomponent intervention)	Cake and juice only; no intervention	Motivational video with relaxing music plus cake and juice during 15–20-min NST	360 (120/120/120); ≥32 wk; multiparous	RCT, 3 arms	Timur Taşhan 2022 [48], Türkiye
End-of-NST FHR 145.3 ± 7.3 vs. 141.0 ± 8.9 *; reactive 99/104 vs. 72/114 *	FHR, reactivity	Routine care	Relaxing instrumental (string) music via an individual listening device for 20 min during NST and in labour; VR arm not used	343 (music 104, VR 124, control 115; 114 analysed); nulliparous; ≥37 wk	RCT, 3 arms	Estrella-Juárez 2023 [49], Spain
FHR $t = 0.01$ (NS); FM $t = 2.68$ *; ACC $t = 6.77$ *	FM, FHR, ACC (t statistics; mixed-risk sample; sensitivity analyses only)	Routine NST	Recorded music via headset during 20 min NST	80 (40/40); 32–42 wk; sample included diabetes and hypertension	Quasi-randomized (participant-selected colour card)	Fathi 2023 [50], Iran

Values are mean ± SD unless stated otherwise. * Statistically significant difference ($p < 0.05$) as reported by the authors. Studies with mixed-risk samples [45,50] and the multicomponent trial [48] were used in sensitivity analyses only; the study by Sumaningsih et al. [46] was synthesized narratively. ACC, accelerations; FHR, baseline fetal heart rate (bpm); FM, fetal movements; NS, not significant; NST, non-stress test; RCT, randomized controlled trial; SD, standard deviation; VR, virtual reality; wk, weeks of gestation.

3.2. Characteristics of the Included Studies

The characteristics of the included studies are summarized in Table 1. The studies were published between 2011 and 2023 and were conducted in Türkiye ($n = 6$), Iran ($n = 3$), Italy, Spain and Indonesia ($n = 1$ each). Six were parallel-group randomized controlled trials [12,24,45,47–49], one was a randomized crossover trial [44] and five were quasi-randomized or non-randomized controlled studies [23,42,43,46,50]. Two studies described as randomized allocated participants by day of the week [23] or by a colour card selected by the participant [50] and were therefore classified as quasi-randomized. In total, 1747 pregnant women were included (range 30–360 per study). Most NSTs were performed at or after 32 weeks of gestation; in the study by Brillo et al., women were enrolled at 28–32 weeks, but the NSTs were performed at 37–38 weeks [47]. Two studies had mixed-risk samples: Khodakarami et al. enrolled nulliparous women presenting with reduced fetal movements [45], and the sample of Fathi et al. included women with diabetes or hypertension [50]; in accordance with the eligibility criteria, these two studies were analyzed in sensitivity analyses only.

The music interventions were heterogeneous (Table 1). The music comprised Western classical music (e.g., Tchaikovsky, Paganini, Chopin) [24,44], slow-tempo or relaxing non-verbal instrumental music [45,49], Turkish classical, art or folk music and other cultur-

ally familiar genres [12,42,43], religious music (hymns, murottal) [23,42,46] and natural sounds [46]; in one study the type of recorded music was not specified [50]. In four studies, participants selected music from a playlist or brought their own music [12,23,42,43]; in the remaining studies, researchers selected the music. Music was delivered to the mother through headphones or a headset in four studies [12,24,45,50], through an MP3 player in three [23,42,43] and through an individual listening device in one [49], and directly to the fetus through abdominal speakers in one [47]; the delivery device was not specified in three studies [44,46,48]. Playback volume was rarely reported; where specified, it was adjusted to participants' comfort. The duration of exposure ranged from about 10 min [24,47] to 30 min [12,23,45,46]; in most studies, music was played throughout the NST, in one study only during the second 10 min of the test [24], and in one study, it started 10 min before the test [45]. Fetal movements were counted by the mother in several studies and by the monitor in others [45], over counting windows that coincided with NST durations of 15–30 min; accelerations were counted on the trace. In one study, music was combined with a motivational video and a snack [48], and one study reported only descriptive pre–post FHR values without standard deviations [46]; these two studies were not included in the primary meta-analyses.

3.3. Methodological Quality

In the JBI appraisal, the main limitations were related to allocation: five studies did not use random allocation, and allocation concealment was rarely described in the randomized trials. Blinding of participants was not feasible, and fetal movements were counted by the mother in several studies, which may introduce performance and detection bias; outcome assessors interpreting the NST traces were reported to be blinded to group allocation in only three studies [24,47,50]. In the crossover study, the change in fetal movements differed by sequence, indicating a period or carry-over effect [44]. Several studies had small samples, and one pilot trial included only ten women per group [47]. Because these limitations apply mainly to the quasi-/non-randomized studies and to maternally counted fetal movements, all pooled estimates are presented stratified by design (Figures 2–5) and the randomized-only estimates are reported for every outcome (Table 2); the JBI findings also informed the risk-of-bias domain of GRADE (Table 3).

Table 2. Primary meta-analyses of music during electronic fetal monitoring compared with no music in uncomplicated pregnancies.

Extended Population: Estimate (95% CI); *k*	**Randomized Studies Only: Estimate (95% CI); *k***	**HKSJ 95% CI (LCL to UCL)**	**95% PI (LCL to UCL)**	**I^2 (τ^2)**	***p*-Value**	**Pooled Estimate (95% CI: LCL to UCL)**	***k* (*n*)**	**Outcome**
0.80 (0.43 to 1.17); 8	0.77 (0.21 to 1.34); 3	0.17 to 1.20	−0.58 to 1.95	74% (0.17)	<0.001	g 0.69 (0.30 to 1.07)	6 (549)	Fetal movements
0.11 (−0.11 to 0.34); 10	0.23 (−0.12 to 0.58); 5	−0.11 to 0.47	−0.59 to 0.96	70% (0.08)	0.151	g 0.18 (−0.07 to 0.43)	8 (909)	Baseline FHR
0.87 (−0.95 to 2.70); 9	1.58 (−0.89 to 4.06); 5	−0.80 to 3.46	−4.31 to 6.97	68% (4.45)	0.151	MD 1.33 (−0.49 to 3.15)	8 (909)	Baseline FHR (bpm)

Table 2. *Cont.*

Extended Population: Estimate (95% CI); *k*	Randomized Studies Only: Estimate (95% CI); *k*	HKSJ 95% CI (LCL to UCL)	95% PI (LCL to UCL)	I^2 (τ^2)	*p*-Value	Pooled Estimate (95% CI: LCL to UCL)	*k* (*n*)	Outcome
0.72 (0.33 to 1.10); 8	0.54 (0.15 to 0.92); 4	0.15 to 1.05	−0.59 to 1.80	78% (0.18)	0.001	g 0.60 (0.24 to 0.97)	7 (688)	Accelera-tions
1.46 (1.10 to 1.93); 6	1.24 (0.86 to 1.79); 2	0.95 to 1.75	0.59 to 2.80	90% (0.05)	0.018	RR 1.29 (1.04 to 1.59)	5 (687)	Reactive NST

Primary analyses: random-effects model (REML) restricted to studies of uncomplicated pregnancies, pooling randomized and quasi-/non-randomized studies. *p*-values are two-sided (*z* test); statistical significance was set at $p < 0.05$. CI, confidence interval; FHR, fetal heart rate; g, Hedges' *g*; HKSJ, Hartung–Knapp–Sidik–Jonkman adjustment (*t* distribution with $k - 1$ degrees of freedom); *k*, number of studies; LCL, lower confidence limit; MD, mean difference (bpm); *n*, number of participants; NST, non-stress test; PI, prediction interval; REML, restricted maximum likelihood; RR, risk ratio of a reactive NST; UCL, upper confidence limit. Extended population: primary analysis set plus the two mixed-risk samples (Khodakarami et al. [45] and Fathi et al. [50]; for the MD analysis only Khodakarami et al., because Fathi et al. reported *t* statistics only).

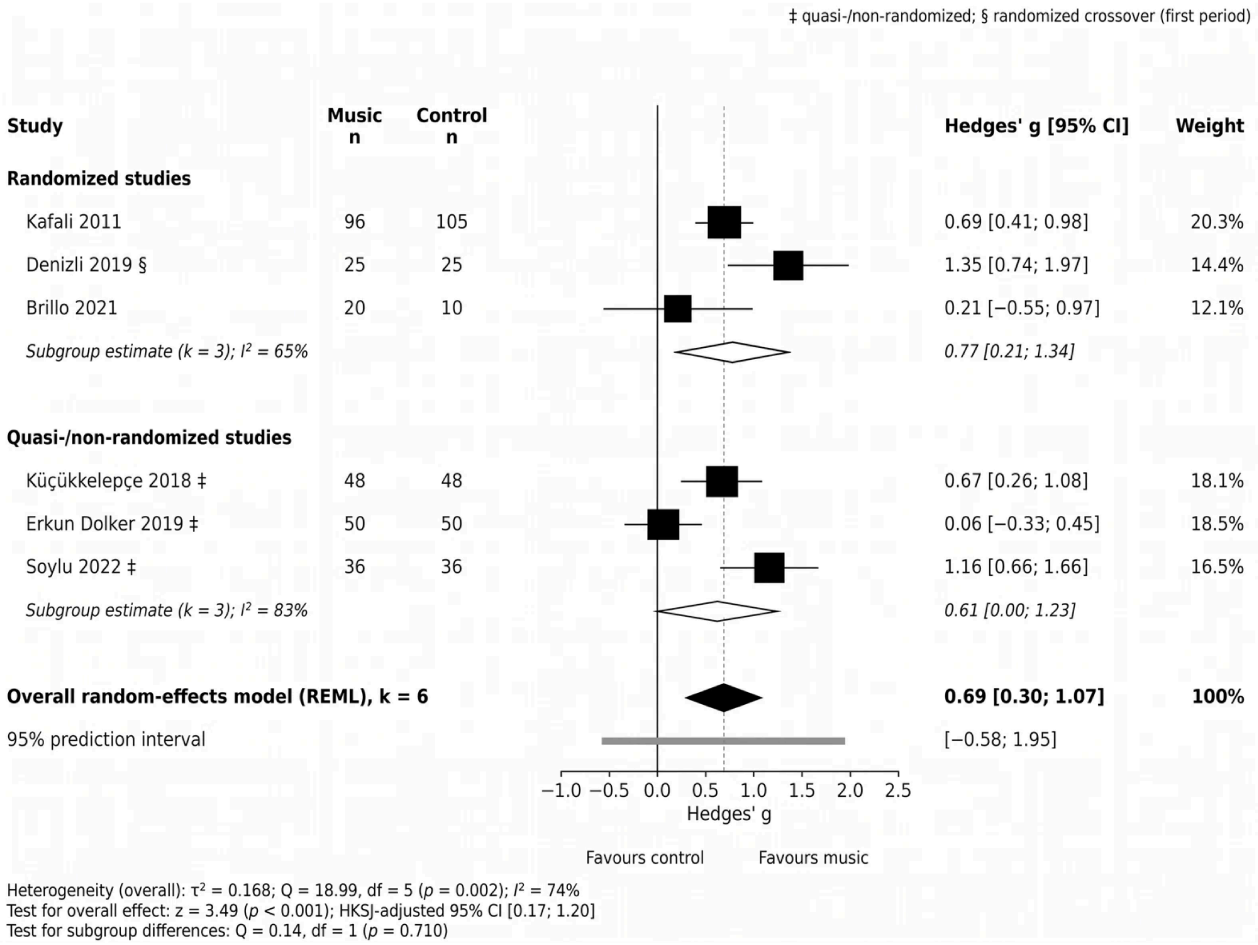


Figure 2. Forest plot of the effect of music on the number of fetal movements in uncomplicated pregnancies (Hedges' *g*, random-effects REML model), stratified by study design. CI, confidence interval; HKSJ, Hartung–Knapp–Sidik–Jonkman; REML, restricted maximum likelihood. Kafali 2011 [12], Denizli 2019 [44], Brillo 2021 [47], Küçükkelepçe 2018 [42], Erkun Dolker 2019 [43], Soylu 2022 [23].

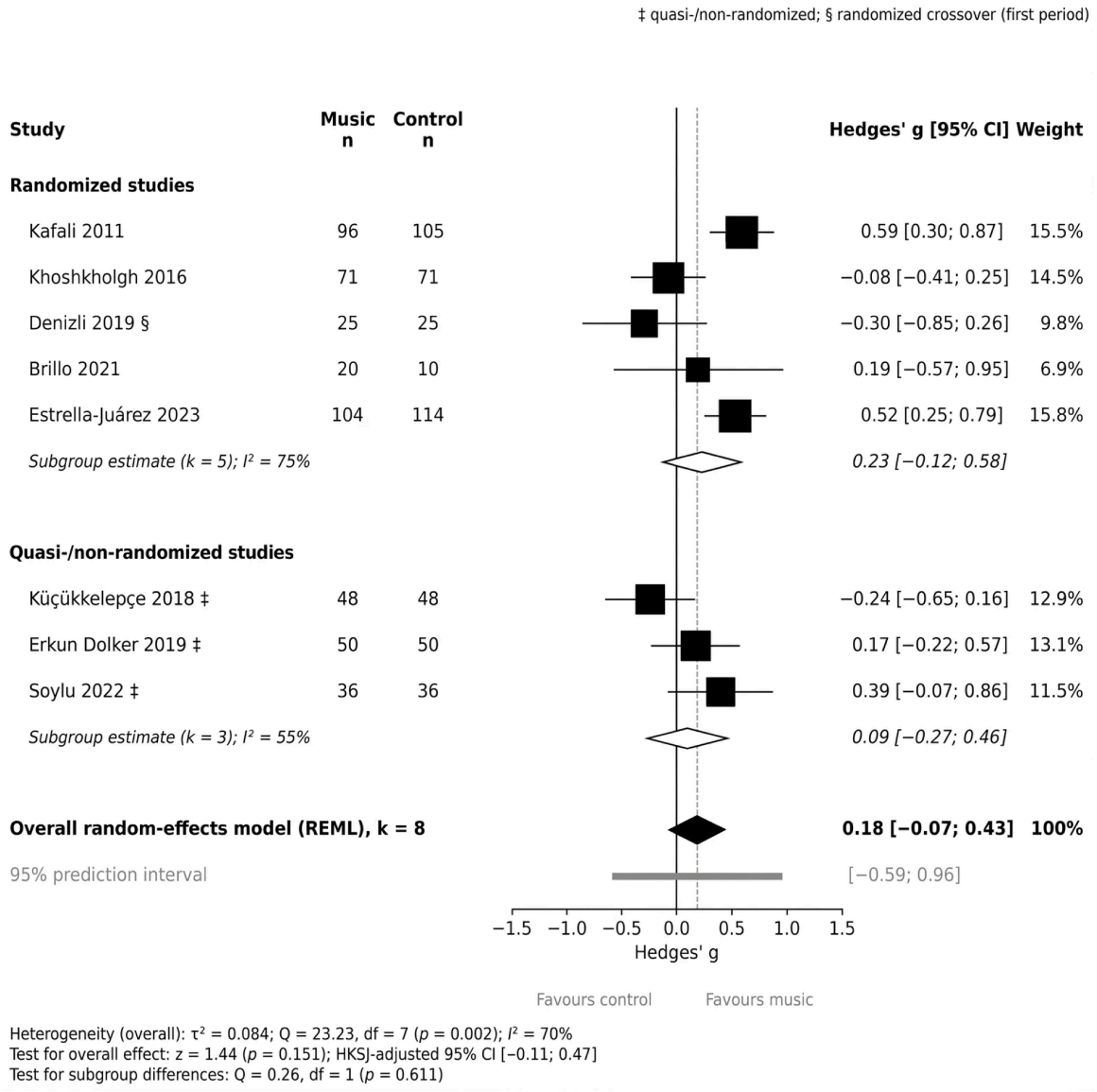


Figure 3. Forest plot of the effect of music on baseline fetal heart rate in uncomplicated pregnancies (Hedges' *g*, random-effects REML model), stratified by study design. CI, confidence interval; HKSJ, Hartung–Knapp–Sidik–Jonkman; REML, restricted maximum likelihood. Kafali 2011 [12], Khoshkholgh 2016 [24], Denizli 2019 [44], Brillo 2021 [47], Estrella-Juárez 2023 [49], Küçükkelepçe 2018 [42], Erkun Dolker 2019 [43], Soylu 2022 [23].

Table 3. Subgroup and sensitivity analyses of the primary meta-analyses.

Egger's Test: Intercept (SE); *p*-Value	Leave-One-out Range of Estimates	Both Crossover Periods: Estimate (95% CI)	Multicomponent Study Added: Estimate (95% CI)	Subgroup Difference *p*-Value	Quasi-/Non-Randomized Studies: *k*; Estimate (95% CI)	Randomized Studies: *k*; Estimate (95% CI)	Outcome
1.05 (2.90); 0.736	0.57 to 0.82	0.56 (0.26 to 0.85)	0.58 (0.20 to 0.96)	0.710	3; 0.61 (0.00 to 1.23)	3; 0.77 (0.21 to 1.34)	Fetal movements
−3.03 (2.07); 0.194	0.11 to 0.25	—	0.18 (−0.03 to 0.40)	0.611	3; 0.09 (−0.27 to 0.46)	5; 0.23 (−0.12 to 0.58)	Baseline FHR (g)
−4.67 (1.57); 0.025	0.83 to 1.82	—	1.38 (−0.20 to 2.95)	0.668	3; 0.76 (−2.09 to 3.60)	5; 1.58 (−0.89 to 4.06)	Baseline FHR (MD, bpm)

Table 3. *Cont.*

Egger's Test: Intercept (SE); *p*-Value	Leave-One-out Range of Estimates	Both Crossover Periods: Estimate (95% CI)	Multicomponent Study Added: Estimate (95% CI)	Subgroup Difference *p*-Value	Quasi-/Non-Randomized Studies: *k*; Estimate (95% CI)	Randomized Studies: *k*; Estimate (95% CI)	Outcome
1.21 (3.01); 0.704	0.48 to 0.71	—	0.53 (0.18 to 0.87)	0.759	3; 0.67 (−0.08 to 1.42)	4; 0.54 (0.15 to 0.92)	Accelerations
3.97 (1.48); 0.074	1.23 to 1.38	—	1.22 (1.01 to 1.48)	0.736	3; 1.35 (0.97 to 1.89)	2; 1.24 (0.86 to 1.79)	Reactive NST (RR)

All analyses in studies of uncomplicated pregnancies (primary analysis set) unless stated otherwise. Subgroup estimates are from separate random-effects (REML) models; the test for subgroup differences is a *Q* test with one degree of freedom. "Both crossover periods" applies to fetal movements only, the only outcome for which both-period data were reported by Denizli et al. [44]. Egger's test: regression of the standardized effect on precision with a two-sided *t*-test of the intercept, interpreted as exploratory because fewer than ten studies were available. Statistical significance was set at $p < 0.05$. —, not applicable; CI, confidence interval (lower to upper confidence limit); *k*, number of studies; MD, mean difference (bpm); RR, risk ratio; SE, standard error.

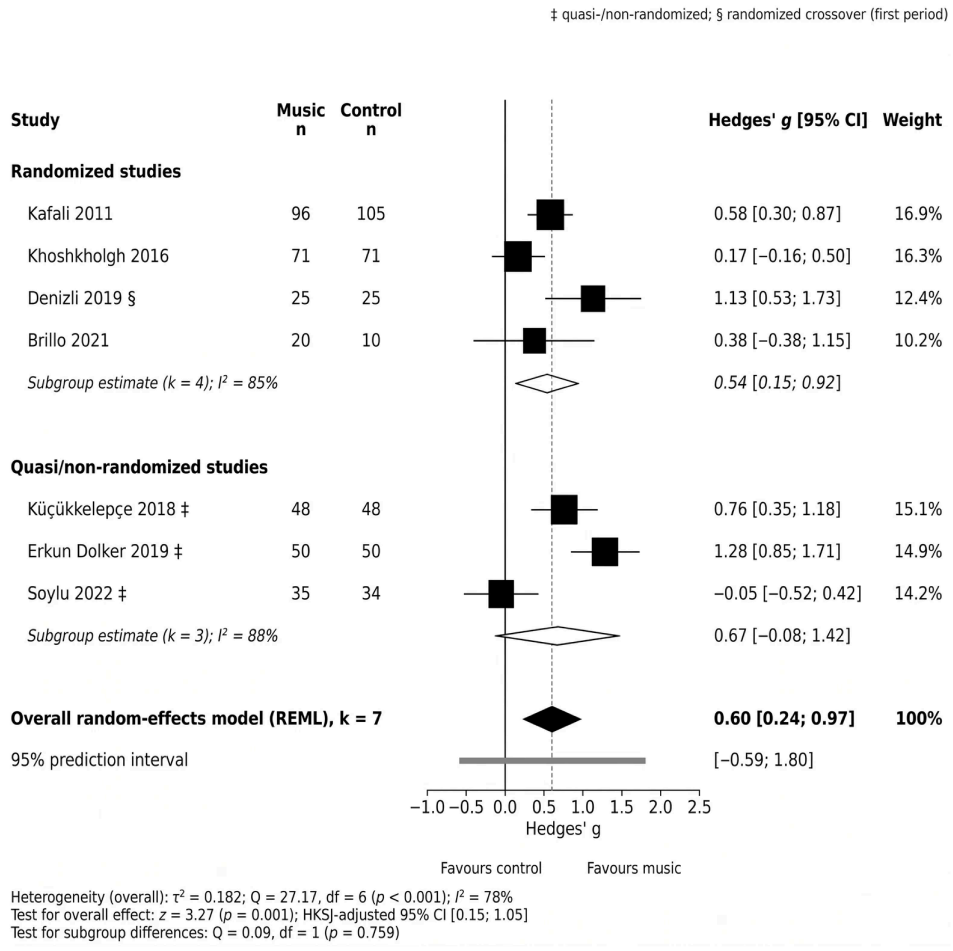


Figure 4. Forest plot of the effect of music on the number of fetal heart rate accelerations in uncomplicated pregnancies (Hedges' *g*, random-effects REML model), stratified by study design. CI, confidence interval; HKSJ, Hartung–Knapp–Sidik–Jonkman; REML, restricted maximum likelihood. Kafali 2011 [12], Khoshkholgh 2016 [24], Denizli 2019 [44], Brillo 2021 [47], Küçükkelepçe 2018 [42], Erkun Dolker 2019 [43], Soylu 2022 [23].

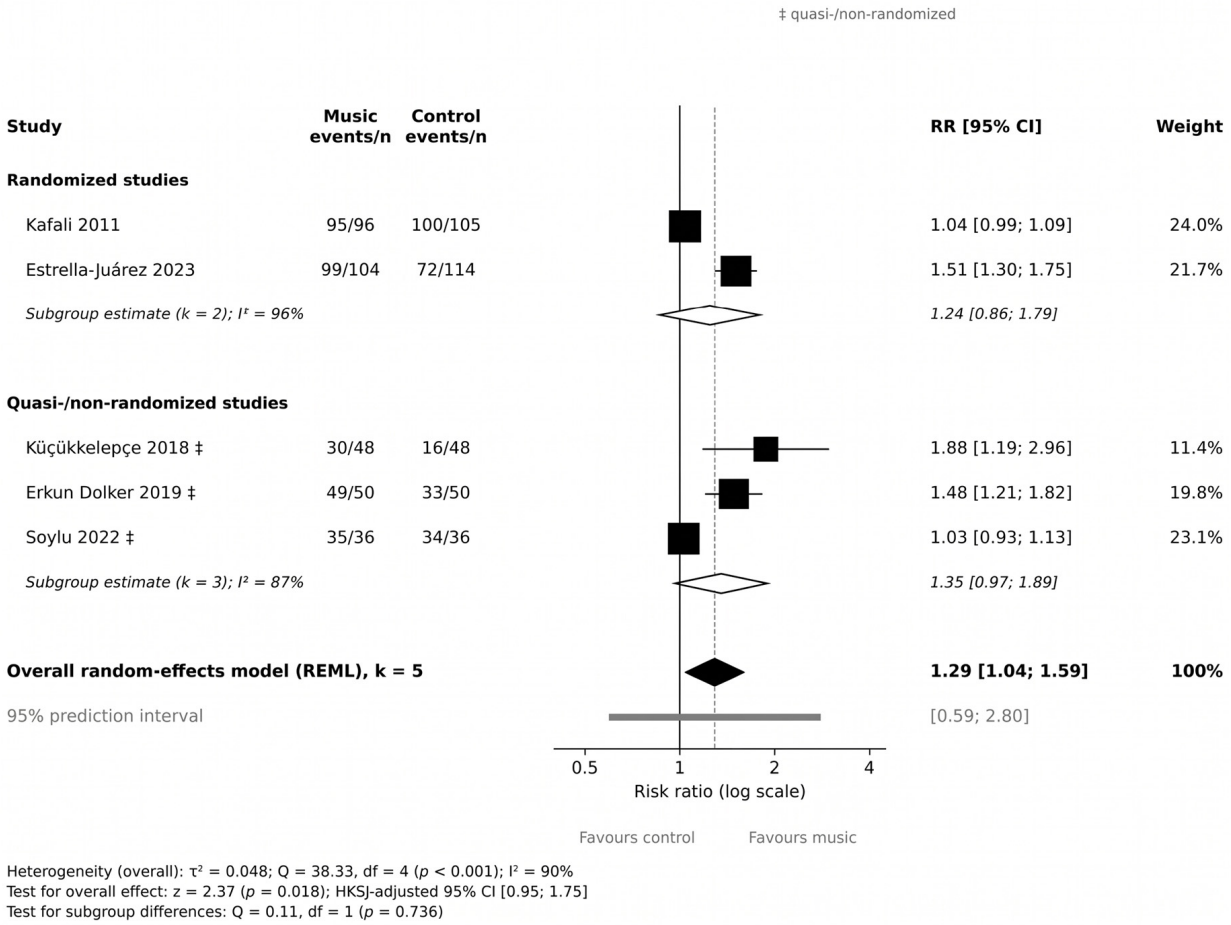


Figure 5. Forest plot of the effect of music on the proportion of reactive non-stress tests in uncomplicated pregnancies (risk ratio, random-effects REML model), stratified by study design. CI, confidence interval; HKSJ, Hartung–Knapp–Sidik–Jonkman; REML, restricted maximum likelihood. Kafali 2011 [12], Estrella-Juárez 2023 [49], Küçükkelepçe 2018 [42], Erkun Dolker 2019 [43], Soylu 2022 [23].

3.4. Narrative Synthesis

3.4.1. Fetal Movements

Nine studies reported fetal movements. The number of movements was significantly higher with music in six studies [12,23,42,44,45,50], whereas no overall between-group difference was found in two studies [43,47]; in the pilot trial, movements differed between groups only at the last of four NSTs [47]. In the multicomponent study, both the cake-and-juice group and the video-with-music group had more movements than the control group, but the two intervention groups did not differ from each other [48].

3.4.2. Baseline FHR

All twelve studies reported baseline FHR. Most found no significant difference between the music and control groups [23,24,42–45,47,48,50]. Two studies reported a significantly higher baseline FHR in the music group, with differences of about 4 bpm [12,49]. In the quasi-experimental study, FHR changes after a single music session were reported descriptively and could not be evaluated statistically [46].

Squares represent individual study effect estimates, with their sizes proportional to study weights. Horizontal lines indicate 95% confidence intervals. Diamonds represent pooled effect estimates, with their widths indicating 95% confidence intervals.

3.4.3. Accelerations and Decelerations

Nine studies reported accelerations. The number of accelerations was significantly higher with music in five studies [12,42,43,48,50], whereas no significant between-group difference was found in three studies [23,24,44]; in one of these trials, accelerations increased significantly within the maternal-music group from the first to the second 10 min of the test [24]. In the pilot trial, accelerations increased during music only in fetuses that had been exposed to the same music before the test [47]. Decelerations were rare, and only one study reported fewer decelerations with music [43].

3.4.4. NST Reactivity

Seven studies reported NST results. The proportion of reactive tests was higher with music in five studies [42,43,45,48,49], whereas it was similar in two studies in which almost all tests were reactive in both groups [12,23]. One of these trials also reported the time to a reactive result (the "minimum procedure time" needed to fulfil the reactivity criteria), which was shorter with music (13.4 ± 5.2 vs. 15.6 ± 6.1 min) [12]. No study reported total NST duration, repeat testing, additional investigations, other downstream clinical management or perinatal outcomes.

3.5. Meta-Analysis

The primary meta-analyses (uncomplicated pregnancies) are summarized in Table 2 and Figures 2–5, the subgroup and sensitivity analyses in Table 4 and the certainty of evidence in Table 4. For all outcomes, the direction of effect was similar in randomized and quasi-/non-randomized studies, with no statistically significant subgroup differences (Table 4).

3.5.1. Fetal Movements

Six studies of uncomplicated pregnancies ($n = 549$) were pooled. Music increased the number of fetal movements ($g = 0.69$, 95% CI 0.30 to 1.07; $p < 0.001$), a moderate-to-large effect (Figure 2). Heterogeneity was substantial ($I^2 = 74\%$; $\tau^2 = 0.17$; $Q = 19.0$, $p = 0.002$), and the 95% prediction interval (−0.58 to 1.95) included no effect. The effect remained statistically significant after the HKSJ adjustment (95% CI 0.17 to 1.20; $p = 0.019$) and in the three randomized studies ($g = 0.77$, 95% CI 0.21 to 1.34), and it was of similar size in the three quasi-/non-randomized studies ($g = 0.61$, 95% CI 0.00 to 1.23; subgroup difference $p = 0.71$). Adding the two mixed-risk samples increased the estimate ($g = 0.80$, 95% CI 0.43 to 1.17; $k = 8$), largely because of the large effect in women presenting with reduced fetal movements [45]; adding the multicomponent study ($g = 0.58$, 95% CI 0.20 to 0.96), using both crossover periods ($g = 0.56$, 95% CI 0.26 to 0.85) and omitting each study in turn (g range 0.57 to 0.82) did not change the conclusion (Table 3).

3.5.2. Baseline FHR

Eight studies ($n = 909$) were pooled. Music did not significantly change baseline FHR ($g = 0.18$, 95% CI −0.07 to 0.43; $p = 0.15$; $I^2 = 70\%$; Figure 3). Expressed in beats per minute, the mean difference was 1.3 bpm (95% CI −0.5 to 3.2; $p = 0.15$; $I^2 = 68\%$; Figure S1), and even the 95% prediction interval (−4.3 to 7.0 bpm) remained within the range of normal variation of the baseline. The result was non-significant in all subgroup and sensitivity analyses, including the extended population ($k = 10$; $g = 0.11$, 95% CI −0.11 to 0.34), and the leave-one-out estimates ranged from 0.11 to 0.25 (Tables 2 and 3).

Table 4. Certainty of evidence (GRADE) for the primary analyses: music during electronic fetal monitoring compared with no music in healthy pregnant women.

Certainty	Publication Bias	Imprecision	Indirectness	Inconsistency	Risk of Bias	Randomized Studies Only (95% CI)	Effect (95% CI: LCL to UCL)	Studies (Participants)	Outcome
⊕⊕○○ Low	Undetected [d]	Not serious [c]	Not serious	Serious [b]	Serious [a]	g 0.77 (0.21 to 1.34)	g 0.69 (0.30 to 1.07)	6 (549)	Fetal movements
⊕⊕○○ Low	Undetected [d,e]	Not serious [f]	Not serious	Serious [b]	Serious [a]	MD 1.6 bpm (−0.9 to 4.1)	MD 1.3 bpm (−0.5 to 3.2)	8 (909)	Baseline FHR
⊕⊕○○ Low	Undetected [d]	Not serious [c]	Not serious	Serious [b]	Serious [a]	g 0.54 (0.15 to 0.92)	g 0.60 (0.24 to 0.97)	7 (688)	Accelerations
⊕○○○ Very low	Undetected [d,h]	Serious [g]	Not serious	Serious [b]	Serious [a]	RR 1.24 (0.86 to 1.79)	RR 1.29 (1.04 to 1.59)	5 (687)	Reactive NST

Evidence from randomized and quasi-/non-randomized studies of uncomplicated pregnancies. CI, confidence interval (lower to upper confidence limit); FHR, fetal heart rate; GRADE, Grading of Recommendations Assessment, Development and Evaluation; MD, mean difference; NST, non-stress test; RR, risk ratio. [a] Non-random allocation in three of the pooled studies, allocation concealment rarely described in the randomized trials, no blinding of participants and maternally counted fetal movements in several studies. [b] $I^2 > 50\%$ and 95% prediction interval including no effect. [c] Confidence interval excludes no effect in the overall, HKSJ-adjusted and randomized-only analyses and the pooled sample exceeds 400 participants; not downgraded, although the interval spans small to large effects. [d] Funnel plots without marked asymmetry and non-significant Egger's test; the tests have low power with fewer than ten studies. [e] For baseline FHR expressed as MD, Egger's test was significant, but the asymmetry was driven by the two largest studies rather than by small studies. [f] Confidence interval (−0.5 to 3.2 bpm) and prediction interval (−4.3 to 7.0 bpm) exclude differences large enough to alter the interpretation of the tracing (the baseline is rounded to 5 bpm; normal range 110–160 bpm). [g] HKSJ-adjusted (0.95 to 1.75) and randomized-only (0.86 to 1.79) confidence intervals include no effect; the lower confidence limit of the overall estimate (1.04) corresponds to a trivial effect; optimal information size not met. [h] Egger's test borderline ($p = 0.074$), with larger effects in smaller studies; not downgraded because of the small number of studies, but small-study effects cannot be excluded.

3.5.3. Accelerations

Seven studies ($n = 688$) were pooled. Music increased the number of FHR accelerations ($g = 0.60$, 95% CI 0.24 to 0.97; $p = 0.001$), a moderate effect (Figure 4). Heterogeneity was high ($I^2 = 78\%$; $\tau^2 = 0.18$), and the 95% prediction interval (−0.59 to 1.80) included no effect. The effect remained significant after the HKSJ adjustment (95% CI 0.15 to 1.05; $p = 0.017$) and in the four randomized studies ($g = 0.54$, 95% CI 0.15 to 0.92), whereas the estimate from the three quasi-/non-randomized studies was similar in size but imprecise ($g = 0.67$, 95% CI −0.08 to 1.42; subgroup difference $p = 0.76$). The conclusion was unchanged in the extended population ($g = 0.72$, 95% CI 0.33 to 1.10; $k = 8$), after adding the multicomponent study ($g = 0.53$, 95% CI 0.18 to 0.87) and in leave-one-out analyses (g range 0.48 to 0.71).

3.5.4. NST Reactivity

Five studies ($n = 687$) were pooled. Women who listened to music were somewhat more likely to have a reactive NST (RR = 1.29, 95% CI 1.04 to 1.59; $p = 0.018$; Figure 5), but this result was fragile. Heterogeneity was very high ($I^2 = 90\%$), mainly because reactivity in the control groups ranged from 33% to 95% and two studies had reactivity rates close to 100% in both groups [12,23]; the 95% prediction interval (0.59 to 2.80) included no effect. The effect was no longer statistically significant after the HKSJ adjustment (95% CI 0.95 to 1.75; $p = 0.081$) or when the analysis was restricted to the two randomized trials (RR = 1.24, 95% CI 0.86 to 1.79), and the estimate from the three quasi-/non-randomized studies was likewise imprecise (RR = 1.35, 95% CI 0.97 to 1.89; subgroup difference $p = 0.74$). Adding the mixed-risk study of women with reduced fetal movements [45] increased the estimate (RR = 1.46, 95% CI 1.10 to 1.93; $k = 6$), adding the multicomponent study gave RR = 1.22 (95% CI 1.01 to 1.48), and the leave-one-out estimates ranged from 1.23 to 1.38, with confidence intervals that included or touched 1 when any one of three studies was omitted [42,43,49].

3.6. Publication Bias

Funnel plots for the four outcomes are shown in Figure 6. Visual inspection did not suggest marked asymmetry for fetal movements or accelerations, and Egger's test was not significant for fetal movements (intercept 1.05, $p = 0.74$), baseline FHR expressed as Hedges' g (intercept −3.03, $p = 0.19$), or accelerations (intercept 1.21, $p = 0.70$). For baseline FHR expressed as MD, Egger's test was significant ($p = 0.025$), driven by the two largest studies, which reported a higher FHR in the music group [12,49]; for NST reactivity, the test was borderline (intercept 3.97, $p = 0.074$), reflecting larger effects in the smaller studies. Because fewer than ten studies were available for every outcome, these tests have low power, and small-study effects cannot be excluded, particularly for NST reactivity.

3.7. Certainty of Evidence

The certainty of evidence was low for fetal movements, baseline FHR and accelerations and very low for NST reactivity (Table 4). All outcomes were downgraded for risk of bias (non-random allocation in several studies, unclear allocation concealment, no blinding of participants and maternally counted fetal movements) and for inconsistency (substantial heterogeneity with prediction intervals including no effect). NST reactivity was additionally downgraded for imprecision, because the HKSJ-adjusted and randomized-only confidence intervals included no effect and the optimal information size was not reached. Outcome-specific explanations for every judgement are given in the footnotes to Table 4.

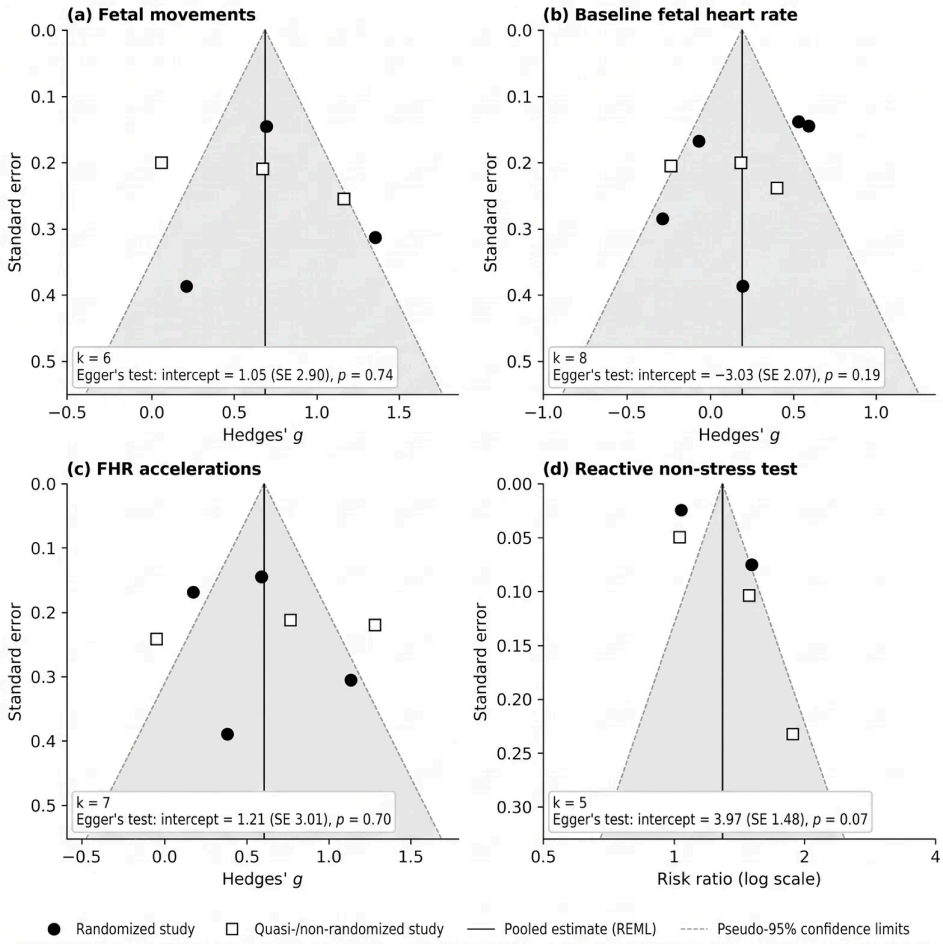


Figure 6. Funnel plots for (**a**) fetal movements, (**b**) baseline fetal heart rate, (**c**) FHR accelerations and (**d**) reactive non-stress tests in the primary analyses (uncomplicated pregnancies). Filled circles, randomized studies; open squares, quasi-/non-randomized studies; solid line, pooled random-effects estimate; dashed lines, pseudo-95% confidence limits. The results of Egger's test are shown in each panel. FHR, fetal heart rate; REML, restricted maximum likelihood.

4. Discussion

This systematic review of twelve controlled studies, with primary meta-analyses restricted to eight studies of uncomplicated pregnancies, found that listening to music during electronic fetal monitoring was associated with more fetal movements and more FHR accelerations, with a small and fragile increase in the proportion of reactive NSTs, and with no change in baseline FHR. The effects on fetal movements and accelerations were moderate, similar in direction in randomized and quasi-/non-randomized studies, and robust to the HKSJ adjustment and to leave-one-out analyses, but heterogeneity was substantial, the prediction intervals included no effect, and the certainty of evidence was low; for NST reactivity it was very low. These findings characterize music as a promising, low-risk intervention that warrants confirmation, not as an intervention of established clinical benefit.

The increase in fetal movements is consistent with most primary studies [12,23,42,44] and with previous narrative reviews reporting that music increases fetal activity [25,26]. The included studies were not designed to identify the mechanism of this effect, and two hypotheses can be proposed. First, music might act indirectly through maternal relaxation: music reduced maternal anxiety during the NST in several of the included studies [12,23,49], and maternal stress and induced relaxation have been shown to modify fetal behaviour and heart rate patterns [10,11]. Second, music might act directly as an auditory stimulus, because the fetus perceives and responds to external sounds and music from the third

trimester onwards [19–21]; this hypothesis is compatible with the pilot trial in which music was delivered only to the fetus [47] and with the fetal acoustic-stimulation arm of the Khoshkholgh trial [24], and it would make music analogous to vibroacoustic stimulation, which increases fetal reactivity [7]. Neither hypothesis has been tested directly by the included studies, and the two are not mutually exclusive. The substantial heterogeneity may partly reflect how movements were counted: in several studies, movements were recorded by the mother, whose perception may itself change when she is relaxed, whereas in others, they were recorded by the monitor; the counting windows and NST durations also differed. The largest single effect was observed in women presenting with reduced fetal movements [45], in whom baseline activity was low; this study was excluded from the primary analysis, and its addition increased the pooled estimate.

Baseline FHR did not change with music. The pooled mean difference of about 1 bpm, with a confidence interval excluding differences larger than about 3 bpm, is clinically negligible, and even the prediction interval remained within the normal range of variation of the baseline. Two studies reported a higher baseline FHR of about 4 bpm in the music group [12,49], whereas others reported similar or slightly lower values [42,44]. Differences in music tempo, timing of measurement and whether FHR was measured during or after exposure may explain these inconsistencies. The absence of a change in baseline FHR suggests that music does not shift the baseline in a way that could affect the interpretation of the tracing, although this conclusion rests on low-certainty evidence.

Music increased the number of accelerations, and because accelerations are closely linked to fetal movements, this finding is consistent with the increase in fetal activity. The increase in the proportion of reactive tests was smaller and less robust. One trial reported a shorter time to a reactive result with music (13.4 vs. 15.6 min) [12], a finding that is plausible given the increase in accelerations but that rests on a single study. The time to a reactive result should, however, be distinguished from the total test duration, which also depends on the testing protocol and on how non-reactive tests are managed, and from downstream clinical management: no included study assessed the total test duration, the need for repeat testing or additional investigations, or any other aspect of clinical management, and no conclusion about clinical management can therefore be drawn. Similarly, although an acceleration induced by stimulation is generally regarded as reflecting an intact fetal autonomic response, as accepted for vibroacoustic stimulation [1,7], the diagnostic safety of music—that is, whether it could produce a reactive test in a compromised fetus—has not been studied, and no included study evaluated perinatal outcomes.

The effect on NST reactivity should be interpreted with particular caution. Heterogeneity was very high because reactivity in the control groups varied from 33% to 95%: in populations where almost all tests are reactive, music cannot improve reactivity [12,23], whereas the largest effects were seen where control-group reactivity was low [42,49] and, in the extended population, in women with reduced fetal movements [45]. The pooled estimate lost statistical significance after the HKSJ adjustment, when the analysis was restricted to the two randomized trials and when single studies were omitted, and the certainty of evidence was very low. Any benefit of music on reactivity is therefore unproven and, if present, is likely to be confined to settings with high rates of non-reactive tests.

Compared with previous reviews, which combined low- and high-risk pregnancies or synthesized fetal outcomes narratively [15,25,26], the present review focused on healthy pregnancies, followed a pre-specified analysis hierarchy, and quantitatively pooled four NST outcomes with design-stratified estimates. All outcome data were re-extracted from the original articles, and design-specific issues—multi-arm trials, a crossover design and studies reporting only test statistics—were handled according to Cochrane recommendations [28,29].

4.1. Strengths and Limitations

The strengths of this review include prospective registration, reporting according to PRISMA 2020, independent title/abstract and full-text screening and independent data extraction by two reviewers, verification of all extracted data against the full texts, a pre-specified analysis hierarchy with design-stratified estimates, the use of REML random-effects models with prediction intervals and HKSJ-adjusted sensitivity analyses, GRADE assessment with outcome-specific justifications, and provision of the extracted data and the complete analysis code.

Several limitations should be considered, relating both to the included studies and to the review methods. First, the number of studies was small (two to five randomized studies per outcome), and randomized and quasi-/non-randomized studies were pooled; although the design-stratified estimates were similar and the randomized-only estimates are reported throughout, confounding in the non-randomized studies cannot be excluded, and the pooled estimates should not be regarded as more precise than the randomized evidence alone. Second, heterogeneity was substantial for all outcomes, reflecting differences in music type (self-selected vs. researcher-selected), delivery (maternal vs. fetal), duration (10–30 min), gestational age, NST duration and, importantly, outcome measurement: fetal movements and accelerations were counted with different methods and over different time windows, which limits the comparability of the pooled standardized effects, and the small number of studies precluded meta-regression. Third, blinding was not possible, and fetal movements were often maternally perceived, which may have inflated the effect. Fourth, for the crossover trial, only first-period data were used in the primary analysis; this avoids carry-over bias but discards half of the data, and the both-period sensitivity analysis gave a smaller estimate. Fifth, two outcomes (accelerations and NST reactivity) and the GRADE assessment were added after registration, and the JBI tools were used rather than the more detailed Cochrane RoB 2 and ROBINS-I tools. Sixth, most studies were conducted in Türkiye and Iran, which may limit generalizability. Seventh, the search was restricted to four databases and to articles published in English or Turkish, grey literature and trial registries were not searched, study authors were not contacted for missing data and studies published after June 2024 were not included; relevant studies may therefore have been missed. In addition, duplicate records and records that were clearly ineligible on the basis of title were removed in a single step before formal screening, so that the number of duplicates cannot be reported separately; this pre-screening step was performed by two reviewers working together, with an independent check of all removed records by a third reviewer, rather than independently in duplicate. Finally, reporting bias could not be reliably assessed because fewer than ten studies were available per outcome, and Egger's test was borderline for NST reactivity.

4.2. Implications for Practice and Research

Given its low cost and acceptability, music is a reasonable candidate intervention for further evaluation during the NST, and women who wish to listen to music of their own choice at a comfortable volume during the test need not be discouraged from doing so. However, the current evidence does not support recommending music as a means of improving test performance, and music should not replace established approaches to a non-reactive test. Future research should consist of adequately powered randomized trials with concealed allocation, blinded outcome assessment and computerized analysis of the tracing, standardized music protocols (type, tempo, sound level, duration and timing), device-recorded fetal movements, and clinically relevant outcomes such as the time to a reactive result, the total test duration, repeat testing, further investigations and perinatal

outcomes. Studies in high-risk pregnancies, with explicit attention to diagnostic safety, are also needed before music can be considered in these populations.

5. Conclusions

Low-certainty evidence from six to eight controlled studies of uncomplicated pregnancies suggests that listening to music during electronic fetal monitoring may increase fetal movements ($g = 0.69$) and FHR accelerations ($g = 0.60$) without changing baseline FHR (MD = 1.3 bpm), and very-low-certainty evidence suggests a small, non-robust increase in NST reactivity (RR = 1.29). Because the effects were heterogeneous, derived partly from non-randomized studies and, for reactivity, sensitive to the analytical approach, and because the search was limited to four databases and two languages and—apart from a shorter time to a reactive result reported by one study—no study assessed total test duration, downstream clinical management or perinatal outcomes, music should be regarded as a promising, low-risk adjunct that requires confirmation rather than as an established intervention. Adequately powered randomized trials with standardized music protocols and clinically relevant outcomes are needed before recommendations for practice can be made.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/healthcare14223959/s1, Figure S1: Forest plot of the effect of music on baseline fetal heart rate, mean difference in beats per minute; Table S1: PRISMA 2020 checklist; File S1: Data extraction workbook; File S2: Analysis code and computational specifications.

Author Contributions: Conceptualization, İ.G.S.; methodology, E.K.G. and İ.G.S.; investigation, E.K.G. and E.A.; data curation, E.K.G. and E.A.; formal analysis, Ö.T.; writing—original draft preparation, E.K.G., İ.G.S. and Ö.T.; writing—review and editing, E.K.G., E.A., İ.G.S. and Ö.T.; supervision, İ.G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study is a systematic review and meta-analysis of previously published studies and did not involve new data collection from human participants.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data analyzed in this review were extracted from published articles and are provided in the article and its Supplementary Materials (data extraction workbook, File S1; analysis code, File S2).

Acknowledgments: An earlier version of this review was presented as an oral presentation at the 3rd International and 4th National Women's Health Nursing Congress, 1–3 May 2025. AI assistance was used for language editing and revision drafting. The authors reviewed and edited all AI-assisted text and take full responsibility for the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACC	accelerations
CI	confidence interval
EFM	electronic fetal monitoring
FHR	fetal heart rate
FM	fetal movements
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HKSJ	Hartung–Knapp–Sidik–Jonkman

JBI	Joanna Briggs Institute
LCL	lower confidence limit
MD	mean difference
NST	non-stress test
PI	prediction interval
PICOS	Population, Intervention, Comparison, Outcomes, Study design
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO	International Prospective Register of Systematic Reviews
RCT	randomized controlled trial
REML	restricted maximum likelihood
RR	risk ratio
SD	standard deviation
UCL	upper confidence limit
VR	virtual reality

References

1. American College of Obstetricians and Gynecologists. Antepartum fetal surveillance: ACOG Practice Bulletin, Number 229. *Obstet. Gynecol.* **2021**, *137*, e116–e127. [CrossRef]
2. United Nations Inter-Agency Group for Child Mortality Estimation (UN IGME). *Levels and Trends in Child Mortality: Report 2024*; United Nations Children's Fund: New York, NY, USA, 2025.
3. United Nations Inter-Agency Group for Child Mortality Estimation (UN IGME). *Standing Up for Stillbirth: Current Estimates and Key Interventions*; United Nations Children's Fund: New York, NY, USA, 2025.
4. Aktaş, S.; Osmanağaoğlu, M.A. Application of intrapartum electronic fetal monitoring and responsibilities of healthcare professionals. *NWSA Life Sci.* **2017**, *12*, 14–29.
5. American College of Obstetricians and Gynecologists. ACOG Practice Bulletin No. 106: Intrapartum fetal heart rate monitoring: Nomenclature, interpretation, and general management principles. *Obstet. Gynecol.* **2009**, *114*, 192–202. [CrossRef]
6. Esin, S. Factors that increase reactivity during fetal nonstress testing. *Curr. Opin. Obstet. Gynecol.* **2014**, *26*, 61–66. [CrossRef]
7. Tan, K.H.; Smyth, R.M.D.; Wei, X. Fetal vibroacoustic stimulation for facilitation of tests of fetal wellbeing. *Cochrane Database Syst. Rev.* **2013**, *2013*, CD002963. [CrossRef]
8. Talge, N.M.; Neal, C.; Glover, V. Antenatal maternal stress and long-term effects on child neurodevelopment: How and why? *J. Child Psychol. Psychiatry* **2007**, *48*, 245–261. [CrossRef]
9. Sarkar, P.; Bergman, K.; O'Connor, T.G.; Glover, V. Maternal antenatal anxiety and amniotic fluid cortisol and testosterone: Possible implications for foetal programming. *J. Neuroendocrinol.* **2008**, *20*, 489–496. [CrossRef]
10. Lobmaier, S.M.; Müller, A.; Zelgert, C.; Shen, C.; Su, P.C.; Schmidt, G.; Haller, B.; Berg, G.; Fabre, B.; Weyrich, J.; et al. Fetal heart rate variability responsiveness to maternal stress, non-invasively detected from maternal transabdominal ECG. *Arch. Gynecol. Obstet.* **2020**, *301*, 405–414. [CrossRef]
11. DiPietro, J.A.; Costigan, K.A.; Nelson, P.; Gurewitsch, E.D.; Laudenslager, M.L. Fetal responses to induced maternal relaxation during pregnancy. *Biol. Psychol.* **2008**, *77*, 11–19. [CrossRef]
12. Kafali, H.; Derbent, A.; Keskin, E.; Simavli, S.; Gözdemir, E. Effect of maternal anxiety and music on fetal movements and fetal heart rate patterns. *J. Matern. Fetal Neonatal Med.* **2011**, *24*, 461–464. [CrossRef]
13. Arabin, B. Music during pregnancy. *Ultrasound Obstet. Gynecol.* **2002**, *20*, 425–430. [CrossRef]
14. Bradt, J.; Potvin, N.; Kesslick, A.; Shim, M.; Radl, D.; Schriver, E.; Gracely, E.J.; Komarnicky-Kocher, L.T. The impact of music therapy versus music medicine on psychological outcomes and pain in cancer patients: A mixed methods study. *Support. Care Cancer* **2015**, *23*, 1261–1271. [CrossRef]
15. Corbijn van Willenswaard, K.; Lynn, F.; McNeill, J.; McQueen, K.; Dennis, C.L.; Lobel, M.; Alderdice, F. Music interventions to reduce stress and anxiety in pregnancy: A systematic review and meta-analysis. *BMC Psychiatry* **2017**, *17*, 271. [CrossRef]
16. Yang, M.; Li, L.; Zhu, H.; Alexander, I.M.; Liu, S.; Zhou, W.; Ren, X. Music therapy to relieve anxiety in pregnant women on bedrest: A randomized, controlled trial. *MCN Am. J. Matern. Child Nurs.* **2009**, *34*, 316–323. [CrossRef]
17. Wulff, V.; Hepp, P.; Wolf, O.T.; Balan, P.; Hagenbeck, C.; Fehm, T.; Schaal, N.K. The effects of a music and singing intervention during pregnancy on maternal well-being and mother–infant bonding: A randomised, controlled study. *Arch. Gynecol. Obstet.* **2021**, *303*, 69–83. [CrossRef]
18. World Health Organization. *WHO Recommendations: Intrapartum Care for a Positive Childbirth Experience*; World Health Organization: Geneva, Switzerland, 2018.
19. Hepper, P.G.; Shahidullah, B.S. Development of fetal hearing. *Arch. Dis. Child. Fetal Neonatal Ed.* **1994**, *71*, F81–F87. [CrossRef]
20. Kisilevsky, B.S.; Hains, S.M.J.; Jacquet, A.Y.; Granier-Deferre, C.; Lecanuet, J.P. Maturation of fetal responses to music. *Dev. Sci.* **2004**, *7*, 550–559. [CrossRef]

21. Al-Qahtani, N.H. Foetal response to music and voice. *Aust. N. Z. J. Obstet. Gynaecol.* **2005**, *45*, 414–417. [CrossRef]
22. Arya, R.; Chansoria, M.; Konanki, R.; Tiwari, D.K. Maternal music exposure during pregnancy influences neonatal behaviour: An open-label randomized controlled trial. *Int. J. Pediatr.* **2012**, *2012*, 901812. [CrossRef]
23. Soylu, N.; Bülbül, T.; Müderris, İ.İ. The effect of music on fetal well-being and anxiety levels and vital signs of pregnant women during non-stress test: Turkey sample. *Health Care Women Int.* **2022**, *43*, 499–517. [CrossRef]
24. Khoshkholgh, R.; Keshavarz, T.; Moshfeghy, Z.; Akbarzadeh, M.; Asadi, N.; Zare, N. Comparison of the effects of two auditory methods by mother and fetus on the results of non-stress test (baseline fetal heart rate and number of accelerations) in pregnant women: A randomized controlled trial. *J. Fam. Reprod. Health* **2016**, *10*, 27–34.
25. Ji, C.; Zhao, J.; Nie, Q.; Wang, S. The role and outcomes of music therapy during pregnancy: A systematic review of randomized controlled trials. *J. Psychosom. Obstet. Gynaecol.* **2024**, *45*, 2291635. [CrossRef]
26. Dogan-Gangal, A.; Akin, B.; Kocoglu-Tanyer, D. Effect of music-based interventions on maternal and fetal parameters during pregnancy: A systematic review. *Complement. Med. Res.* **2024**, *31*, 540–550. [CrossRef]
27. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef]
28. Higgins, J.P.T.; Thomas, J.; Chandler, J.; Cumpston, M.; Li, T.; Page, M.J.; Welch, V.A. (Eds.) *Cochrane Handbook for Systematic Reviews of Interventions*; Version 6.4 (updated August 2023); Cochrane: London, UK, 2023. Available online: https://www.training.cochrane.org/handbook (accessed on 25 October 2026).
29. Elbourne, D.R.; Altman, D.G.; Higgins, J.P.T.; Curtin, F.; Worthington, H.V.; Vail, A. Meta-analyses involving cross-over trials: Methodological issues. *Int. J. Epidemiol.* **2002**, *31*, 140–149. [CrossRef]
30. Barker, T.H.; Stone, J.C.; Sears, K.; Klugar, M.; Tufanaru, C.; Leonardi-Bee, J.; Aromataris, E.; Munn, Z. The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. *JBI Evid. Synth.* **2023**, *21*, 494–506. [CrossRef]
31. Barker, T.H.; Habibi, N.; Aromataris, E.; Stone, J.C.; Leonardi-Bee, J.; Sears, K.; Hasanoff, S.; Klugar, M.; Tufanaru, C.; Moola, S.; et al. The revised JBI critical appraisal tool for the assessment of risk of bias for quasi-experimental studies. *JBI Evid. Synth.* **2024**, *22*, 378–388. [CrossRef]
32. Hedges, L.V.; Olkin, I. *Statistical Methods for Meta-Analysis*; Academic Press: Orlando, FL, USA, 1985.
33. Borenstein, M.; Hedges, L.V.; Higgins, J.P.T.; Rothstein, H.R. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res. Synth. Methods* **2010**, *1*, 97–111. [CrossRef]
34. Higgins, J.P.T.; Thompson, S.G.; Deeks, J.J.; Altman, D.G. Measuring inconsistency in meta-analyses. *BMJ* **2003**, *327*, 557–560. [CrossRef]
35. Riley, R.D.; Higgins, J.P.T.; Deeks, J.J. Interpretation of random effects meta-analyses. *BMJ* **2011**, *342*, d549. [CrossRef]
36. IntHout, J.; Ioannidis, J.P.A.; Borm, G.F. The Hartung–Knapp–Sidik–Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian–Laird method. *BMC Med. Res. Methodol.* **2014**, *14*, 25. [CrossRef]
37. Egger, M.; Davey Smith, G.; Schneider, M.; Minder, C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* **1997**, *315*, 629–634. [CrossRef]
38. Sterne, J.A.C.; Sutton, A.J.; Ioannidis, J.P.A.; Terrin, N.; Jones, D.R.; Lau, J.; Carpenter, J.; Rücker, G.; Harbord, R.M.; Schmid, C.H.; et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* **2011**, *343*, d4002. [CrossRef]
39. Guyatt, G.H.; Oxman, A.D.; Vist, G.E.; Kunz, R.; Falck-Ytter, Y.; Alonso-Coello, P.; Schünemann, H.J. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **2008**, *336*, 924–926. [CrossRef]
40. Guyatt, G.H.; Oxman, A.D.; Kunz, R.; Woodcock, J.; Brozek, J.; Helfand, M.; Alonso-Coello, P.; Glasziou, P.; Jaeschke, R.; Akl, E.A.; et al. GRADE guidelines: 7. Rating the quality of evidence—Inconsistency. *J. Clin. Epidemiol.* **2011**, *64*, 1294–1302. [CrossRef]
41. Guyatt, G.H.; Oxman, A.D.; Kunz, R.; Brozek, J.; Alonso-Coello, P.; Rind, D.; Devereaux, P.J.; Montori, V.M.; Freyschuss, B.; Vist, G.; et al. GRADE guidelines 6. Rating the quality of evidence—Imprecision. *J. Clin. Epidemiol.* **2011**, *64*, 1283–1293. [CrossRef]
42. Şimşek Küçükkelepçe, D.; Timur Taşhan, S. The effect of music on the results of a non-stress test: A non-randomized controlled clinical trial. *Eur. J. Integr. Med.* **2018**, *18*, 8–12. [CrossRef]
43. Erkun Dolker, H.; Basar, F. The effect of music on the non-stress test and maternal anxiety. *Complement. Ther. Clin. Pract.* **2019**, *35*, 259–264. [CrossRef]
44. Denizli, R.; Çiçekli, N.; Gülyaşar, G.; Aboalhasan, Y.; Aygün, T.; Farisoğulları, N.; Sakin, Ö. Assessment of the effect of music on non-stress test. *SDÜ Tıp Fak. Derg.* **2019**, *26*, 464–470. (In Turkish)
45. Khodakarami, B.; Janesari Ladani, M.; Kazemi, F.; Aghababaei, S. The effect of music on fetus movement during non-stress test. *Avicenna J. Nurs. Midwifery Care* **2020**, *28*, 1–8. [CrossRef]
46. Sumaningsih, R.; Rahayu, T.P.; Santosa, B.J. Effects of classical music, natural and murottal music on fetal well-being. *Health Notions* **2020**, *4*, 222–225. [CrossRef]

47. Brillo, E.; Tosto, V.; Ceccagnoli, A.; Nikolova, N.; Pinzaglia, V.; Bordoni, F.; Spano, F.; Bini, V.; Giardina, I.; Di Renzo, G.C. The effect of prenatal exposure to music on fetal movements and fetal heart rate: A pilot study. *J. Matern. Fetal Neonatal Med.* **2021**, *34*, 2274–2282. [CrossRef]
48. Timur Taşhan, S.; İnci Coşkun, E.; Nacar, G.; Erci, B. The effect of motivational video and nutrition on the non-stress test: A randomised controlled clinical trial. *J. Obstet. Gynaecol.* **2022**, *42*, 2144–2150. [CrossRef]
49. Estrella-Juarez, F.; Requena-Mullor, M.; Garcia-Gonzalez, J.; Lopez-Villen, A.; Alarcon-Rodriguez, R. Effect of virtual reality and music therapy on the physiologic parameters of pregnant women and fetuses and on anxiety levels: A randomized controlled trial. *J. Midwifery Womens Health* **2023**, *68*, 35–43. [CrossRef]
50. Fathi, L.; Shakarami, A.; Amraei, K.; Yari, F.; Behzadvand, A. Effects of music therapy on the fetal outcomes of non-stress test and maternal anxiety. *Neuropsychiatr. Enfance Adolesc.* **2023**, *71*, 316–319. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.